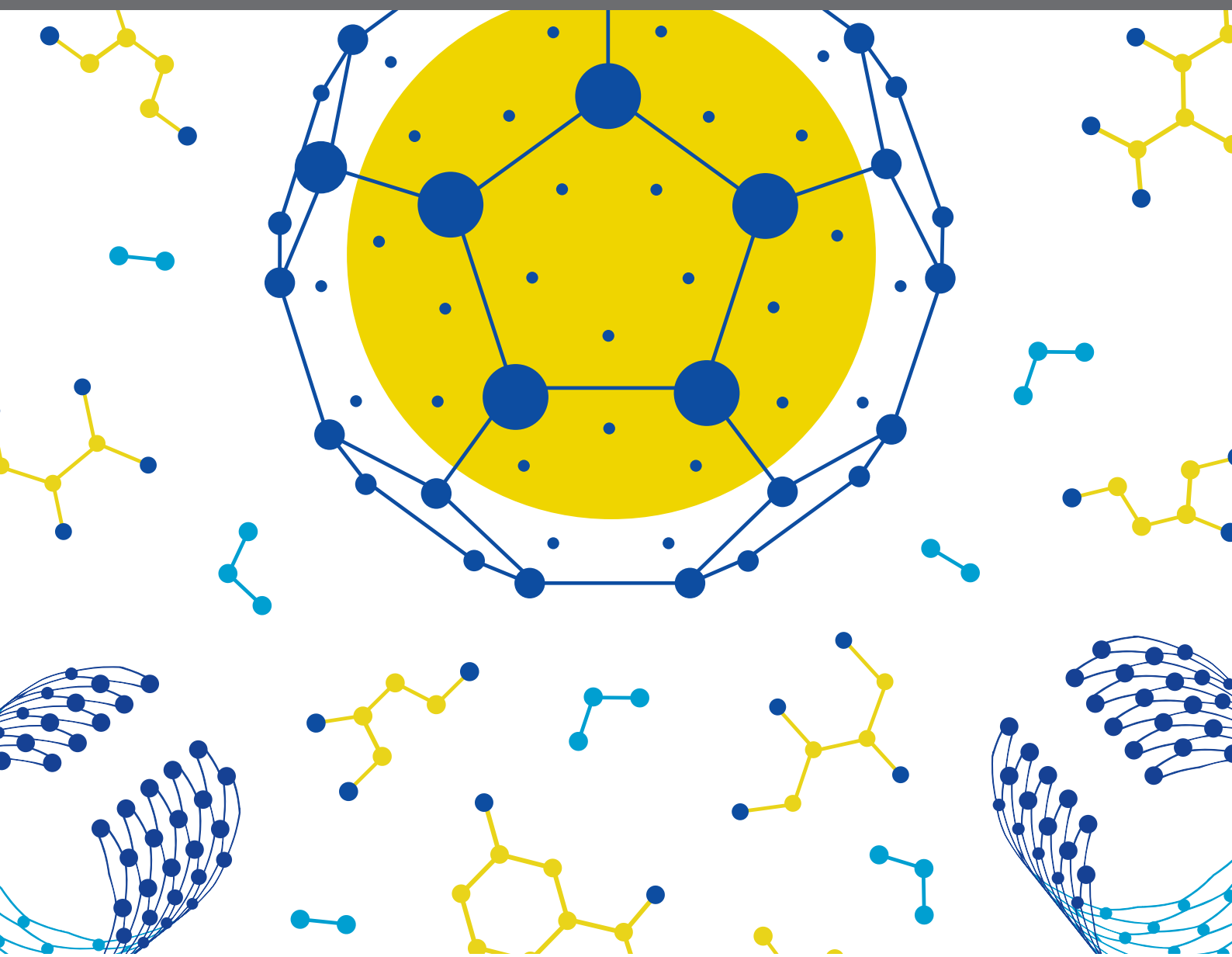
MEMRISTIVE NEUROMORPHICS: MATERIALS, DEVICES, CIRCUITS, ARCHITECTURES, ALGORITHMS AND THEIR CO-DESIGN

EDITED BY: Huanglong Li, J. Joshua Yang and Hongsik Jeong
PUBLISHED IN: Frontiers in Nanotechnology and Frontiers in Neuroscience



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88974-460-2
DOI 10.3389/978-2-88974-460-2

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

MEMRISTIVE NEUROMORPHICS: MATERIALS, DEVICES, CIRCUITS, ARCHITECTURES, ALGORITHMS AND THEIR CO-DESIGN

Topic Editors:
Huanglong Li, Tsinghua University, China
J. Joshua Yang, University of Southern California, United States
Hongsik Jeong, Ulsan National Institute of Science and Technology, South Korea

Citation: Li, H., Yang, J. J., Jeong, H., eds. (2022). Memristive Neuromorphics: Materials, Devices, Circuits, Architectures, Algorithms and Their Co-Design. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-460-2

Table of Contents

frontiers in Nanotechnology

ORIGINAL RESEARCH
published: 30 October 2020
doi: 10.3389/fnano.2020.592684

Investigation on the Stabilizing Effect of Titanium in HfO_2-Based Resistive Switching Devices With Tungsten Electrode

Vittorio Fra[1,2], *Elmira Shahrabi*[2], *Yusuf Leblebici*[2] *and Carlo Ricciardi*[1]*

[1] Department of Applied Science and Technology (DiSAT), Politecnico di Torino, Turin, Italy, [2] Microelectronic Systems Laboratory (LSM), Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland

Resistive switching (RS) devices, also referred to as resistive random access memories (ReRAMs), rely on a working principle based on the change of electrical resistance following proper external electrical stimuli. Since the demonstration of the first resistive memory based on a binary transition metal oxide (TMO) enclosed in a metal–insulator–metal (MIM) structure, this class of devices has been considered a key player for simple and low-cost memories. However, successful large-scale integration with standard complementary metal–oxide–semiconductor (CMOS) technologies still needs systematic investigations. In this work, we examine the beneficial effect titanium has when employed as a buffer layer between CMOS-compatible materials like hafnium dioxide and tungsten. Hindering the tungsten oxidation, Ti provides RS stabilization and allows getting faster responses from the devices. Through an extensive comparative study, the effect of both thickness and composition of Ti-based buffer layers is investigated. The reported results show how titanium can be effectively employed to stabilize and tailor the RS behavior of the devices, and they may open the way to the definition of new design rules for ReRAM–CMOS integration. Moreover, the gradual switching and the response speed tunability observed employing titanium might also extend the domain of interest of these results to brain-inspired computing applications.

Keywords: resistive switching, ReRAM, tungsten, titanium, buffer layer

OPEN ACCESS

Edited by:
J. Joshua Yang,
University of Southern California, Los Angeles, United States

Reviewed by:
Massood Atashbar,
Western Michigan University, United States
Abhay A. Sagade,
SRM Institute of Science and Technology, India

***Correspondence:**
Carlo Ricciardi
carlo.ricciardi@polito.it

Specialty section:
This article was submitted to Nanodevices, a section of the journal Frontiers in Nanotechnology

Received: *07 August 2020*
Accepted: *29 September 2020*
Published: *30 October 2020*

Citation:
Fra V, Shahrabi E, Leblebici Y and Ricciardi C (2020) Investigation on the Stabilizing Effect of Titanium in HfO_2-Based Resistive Switching Devices With Tungsten Electrode. Front. Nanotechnol. 2:592684. doi: 10.3389/fnano.2020.592684

INTRODUCTION

Devices with tunable electrical resistance find application in information and communication technologies (ICTs) since the end of the 19th century, when the so-called coherer was employed as receiver in Marconi's wireless telegraph (Marconi, 1899) thanks to the possibility of changing, and retaining, its electrical conductivity upon external stimuli. Some decades later, in the 1960s, attention started focusing on oxide materials with similar properties (Gibbons and Beadle, 1964; Lamb and Rundle, 1967), opening the way for the wide class of devices nowadays identified as resistive memories. Also referred to as resistive random access memories (ReRAMs) or oxide RAMs (OxRAMs), these resistive switching (RS) devices typically rely on a simple metal–insulator–metal (MIM) structure composed of two metallic electrodes enclosing an insulating oxide layer (Waser and Aono, 2007), but similar stacks without metals have been demonstrated too (Yen et al., 2019). As for the coherer, their working principle is based on the change of electrical resistance

as a response to proper external electrical stimuli. The condition of low conductivity is defined as high-resistance state (HRS), and it can be turned into a more conductive low-resistance state (LRS) through the so-called SET process. The opposite transition, resulting in a resistance increase, namely the transition from LRS to HRS, is instead named RESET. When both the state transitions occur with the same polarity, RS devices are classified as unipolar, while they are defined as bipolar if SET and RESET require opposite polarities (Ielmini and Waser, 2016). In most cases, before exhibiting successful switching between these states, RS devices require the so-called forming process, which gives the first transition of the pristine device to a highly conductive state. Since the demonstration of the first resistive memory based on a binary transition metal oxide (TMO) (Baek et al., 2005), this class of devices has been considered a key player for simple and low-cost memories able to compete with the market-leading technologies (Wong et al., 2012; Meena et al., 2014). Such a perspective translated into an unceasing driving force for research efforts to continuously improve features like low power consumption, high density, fast switching, high endurance, long retention, and compatibility with complementary metal–oxide–semiconductor (CMOS) technologies (Cai et al., 2019; Tang et al., 2019; Xia and Yang, 2019; Wang et al., 2020). In seeking to fulfill these requirements, many studies have been carried out on subjects ranging from the physical behavior to the hardware implementation. As a result, it is now well-established that both interface-type (Celano et al., 2017; Govoreanu et al., 2017) and filamentary-type (Joshua Yang et al., 2009; Lee et al., 2009; Celano et al., 2014) resistive switching exist, and it is widely accepted that the formation of a conductive filament involves ion motion within the insulating layer of the MIM structure (Valov, 2014; Sun et al., 2019; Wang et al., 2020). In-memory computing systems have been shown (Zidan et al., 2018) and brain-inspired functionalities have been demonstrated (Xia and Yang, 2019). Despite these outstanding findings, successful large-scale integration with standard CMOS technologies is only just at the beginning and still needs further systematic investigations able to provide new design rules. In this context, many materials have been studied for both the insulating layer and the electrodes. Silver and copper have been employed in the so-called electrochemical metallization (ECM) memory cells, where they work as electrochemically active electrodes to release cations for metallic filament formation upon electromigration through the "I" layer (Valov et al., 2011). Platinum and titanium nitride have been shown to be suitable for inert electrodes (Tappertzhofen et al., 2014), while oxidizing metals like tungsten, titanium, hafnium, and tantalum have been studied as electrodes in valence change memory (VCM) devices (Chen et al., 2013; Lin et al., 2013; Shahrabi et al., 2019) and many oxides have been tested as an insulating layer. Among them, resounding success has been achieved by HfO_2 (Chen et al., 2009), Ta_2O_5 (Kim et al., 2016), TaO_x (Yang et al., 2010), TiO_2 (Chen et al., 2017), and ZnO (Conti et al., 2019). Moreover, it has been pointed out by different works that the whole material stack of each ReRAM cell, and not only the single layers, is the ultimate responsible for the device performances (Gilmer et al., 2011; Walczyk et al., 2012; Chen et al., 2013; Kim et al., 2016; Rahaman et al., 2017; Singh et al., 2018; Ambrosi et al., 2019; Kindsmüller et al., 2019; Lee et al., 2019; Shahrabi et al., 2019). Particularly, in the framework of VCM devices, a key role is played by the interaction between the metal oxide in the "I" layer and the oxidizing electrode. Such devices, indeed, rely on the formation and rupture of a conductive filament resulting from local valence changes of the metal within the oxide, which, in turn, results from the migration of O^{2-} ions and the subsequent formation of oxygen vacancies ($V_{\ddot{O}}$) under the action of an applied voltage (Celano et al., 2016). The motion of these species strongly depends on the oxygen exchange between the oxide film and the oxidizing electrode and can be described by the reaction:

$$M\,(\text{bulk}) + \text{TMO} \rightleftharpoons MO_x + \text{TMO}_{1-x} + x \cdot V_{\ddot{O}} + 2x{\cdot}e^{-} \quad (1)$$

where M is the oxidizing electrode and TMO is the oxide in the "I" layer.

In view of the upcoming CMOS integration, tungsten turns out to be a feasible choice for the oxidizing electrode due to its already established employment for vertical interconnect accesses (VIAs). However, when used in direct contact with an oxide, its multiple and metastable oxide forms introduce relevant instability in the memory cell performances, so that the insertion of a buffer layer becomes necessary (Shahrabi et al., 2019). In order to efficiently mitigate the effect of the non-stable oxides tungsten can form, a suitable candidate to play this role is titanium. Thanks to the lower energy it requires for reaction (1) with respect to tungsten (Guo and Robertson, 2014; Kim et al., 2016), titanium can indeed extract oxygen from the "I" layer more effectively, and so hinder the formation of metastable tungsten oxides. Furthermore, in the perspective of possible future applications and integrations, the strength of titanium as a suitable candidate for buffer layers comes from its capability to allow gradual RESET transitions for multiple resistance levels tuning (Shahrabi et al., 2019).

In this work, a systematic study of the effect of titanium-based buffer layers enclosed between a tungsten electrode and an oxide layer is carried out on RS devices exhibiting hafnium oxide (HfO_2) as the insulating layer of the MIM structure and platinum as the inert electrode. The role played by Ti in modulating the interaction between the oxidizing electrode and the oxide is investigated through an extensive, comparative investigation of devices with buffer layers having different thicknesses and different compositions. Devices without a buffer layer, namely with the tungsten electrode in direct contact with the HfO_2 film, are also tested and kept as performance references. A clear effect of thickness is observed in both static and dynamic operations, with lower and tunable forming, SET and RESET voltages, better endurance, and faster response achieved through a thicker Ti-based buffer layer. Especially, with respect to devices without any buffer layer, the early HRS failure is fixed and pulses down to three orders of magnitude shorter can be employed. These results, coupled with gradual RESET transitions, make the Ti buffer/W electrode stack a versatile candidate for CMOS-compatible ReRAM cells to be employed in brain-inspired applications.

MATERIALS AND METHODS

Device Fabrication

For our devices, a cross-point geometry was adopted, with VIA openings defining the active region of the ReRAM cells. Using a standard 4-in. Si wafer with a 500-nm-thick SiO_2 layer as a substrate, platinum electrodes were first defined, starting with sputtering deposition of a 5-nm-thick titanium adhesion layer and a 125-nm-thick Pt film by a Pfeiffer Spider 600. Patterning was then performed through photolithography and dry etching, carried out with an STS Multiplex ICP etcher. Afterwards, in order to assure electrical isolation between the electrodes, a 100-nm-thick low thermal oxide (LTO) was deposited at 425°C by means of low-pressure chemical vapor deposition (LPCVD). Once the Pt electrodes were patterned and isolated, VIA openings of different sizes were defined across the LTO passivation layer performing photolithography and buffer oxide etch (BOE). Thereafter, HfO_2 and the Ti-based buffer layers were deposited, the latter with thickness varying sample by sample (1, 3, and 5 nm) and the former always 5 nm thick. Concerning the oxide, atomic layer deposition (ALD) at 200°C was performed by means of a BENQ TFS200, while the buffer layers were deposited by room temperature sputtering, with an Alliance Concept DP650, employing two different targets: pure titanium (99.9995%) and mixed titanium–tungsten (99.99% of purity with 10% in weight of Ti). By means of the same sputtering tool, the tungsten electrode and a titanium nitride capping layer were then deposited, with thicknesses of 60 and 15 nm, respectively. Finally, to pattern the electrode and define the arrays of cross-point cells, photolithography and dry etching were performed, employing again the STS Multiplex ICP dry etcher.

Device Characterization

The device characterization was carried out through electrical tests in three different configurations, all of them performed in air at room temperature. DC sweeping mode was first adopted to evaluate the forming voltage and to inspect the cycling operation. To this aim, a parameter analyzer (Agilent B1500) was employed, applying voltage ramps at the tungsten electrode and keeping grounded the platinum one. During these measurements, a compliance current, I_{cc}, intended to prevent irreversible damages to the devices, was imposed through the internal modules of the characterization tool. Pulse measurements were instead performed to test the dynamic behavior in terms of endurance, response speed, and retention. In this case, since parameter analyzers generally suffer from pure accuracy in current limitation due to a certain delay with respect to the characteristic times of forming and SET processes (Tirano et al., 2011; Nafria et al., 2017), an external n-channel transistor (n-MOSFET) was used to control the compliance current. The device under test was connected in series to the drain of the transistor (bit line), while the source (source line) was grounded and voltages were applied at the gate (word line) to adjust the current limitation. Additionally, conductive atomic force microscopy (C-AFM), by an Asylum Research Cipher VRS, was employed to investigate the forming process directly probing the HfO_2 layer on top of the W/Ti buffer/HfO_2 stacks. Full-platinum AFM tips from Rocky Mountain Nanotechnologies were used as the top electrode in order to reproduce the same MIM structure as for the cross-point cells characterized by means of the parameter analyzer.

RESULTS

All the different stacks employed for the tested devices are summarized in **Table 1**. A 3D sketch of the device structure is then reported in **Figure 1A**, while the field emission scanning electron microscope (FESEM) image in **Figure 1B** shows the actual geometry with a top view of a single ReRAM cell. The micrograph in **Figure 1C** highlights the periodic arrangement of the devices adopted on each sample, with the different VIA diameters of 1.5, 2, 3, 5, and 10 μm. A topography image acquired by AFM in contact mode is also reported in **Figure 1D**, where a 10-μm VIA is shown.

In order to carry out a complete performance analysis suited to compare the material stacks and investigate the effect of the Ti-based buffer layers, 25 devices for each sample were first subjected to a systematic DC characterization made of forming and cycling steps. Pristine devices underwent positive voltage sweeps from 0 to 7 V with a compliance current of 1 mA; then, bipolar voltage ramps ranging from −1.5 V to 3 V were applied to the same devices to test the cycling behavior. Bipolar resistive switching, with SET and RESET occurring in positive and negative polarity, respectively, was observed for all the devices regardless of the material stack. The latter, conversely, turned out to play a role in the definition of the device performance. First of all, an impact of the titanium-based buffer layers on the forming process was observed, with a decrease of the forming voltage ($V_{FORMING}$) for thicker buffer layers (**Figure 2A**). Particularly, as presented

TABLE 1 | Material stacks of all the tested devices.

Sample name	Inert electrode		Oxide layer		Buffer layer		Oxidizing electrode	
	Material	Thickness (nm)	Material	Thickness (nm)	Material	Thickness (nm)	Material	Thickness (nm)
noBuffer	Pt	125	HfO_2	5	–	–	W	60
mixBuffer	Pt	125	HfO_2	5	W:Ti 10%	3	W	60
Buffer1	Pt	125	HfO_2	5	Ti	1	W	60
Buffer3	Pt	125	HfO_2	5	Ti	3	W	60
Buffer5	Pt	125	HfO_2	5	Ti	5	W	60

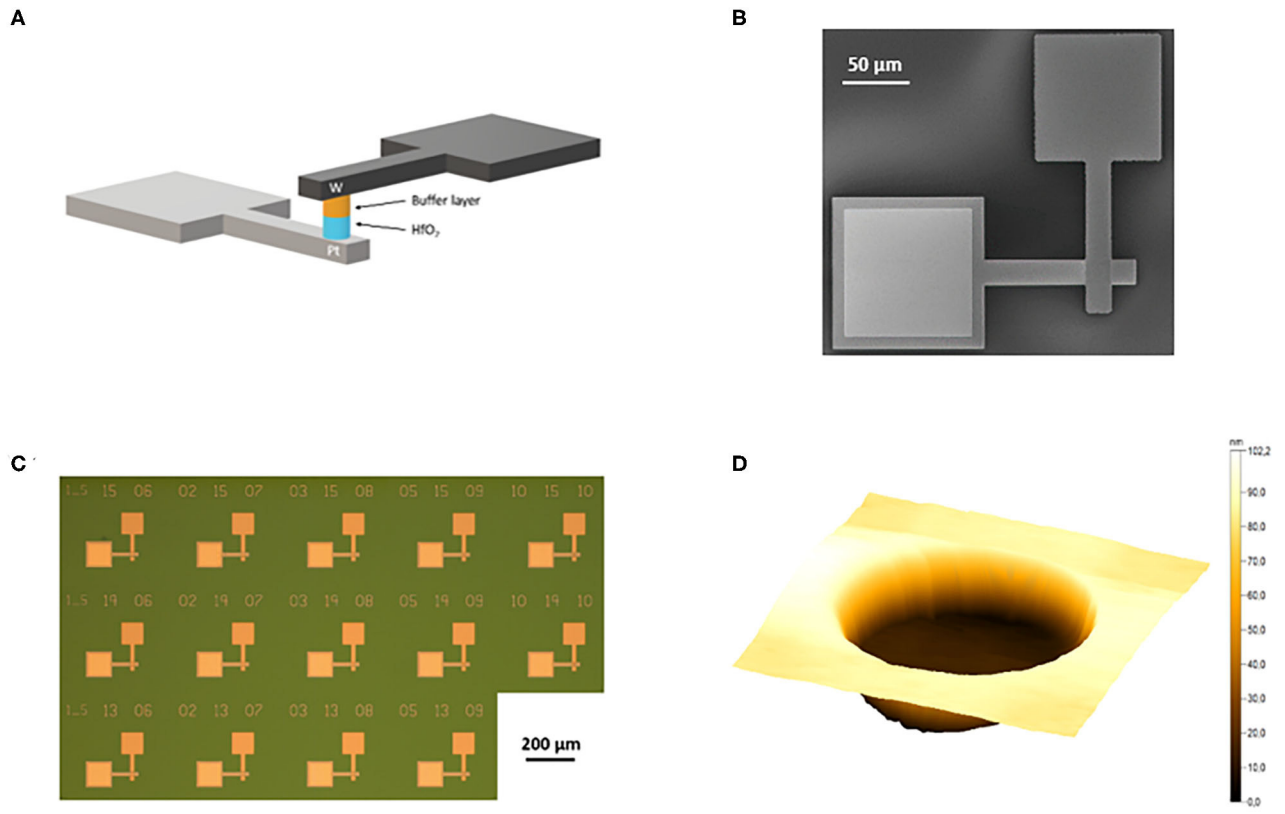


FIGURE 1 | The cross-point geometry of the tested devices is shown through a 3D sketch **(A)** and a field emission scanning electron microscope (FESEM) image in *top view* **(B)**. In the first case, the schematic, not to scale, focuses on a single device emphasizing the material stack, while in the second picture the actual geometry is shown. The periodic arrangement of the resistive random access memory (ReRAM) cells with all the different vertical interconnect access (VIA) dimensions is then highlighted by a micrograph **(C)**. In **(D)**, a single VIA with a diameter of 10 µm is shown by a topography image obtained with contact mode atomic force microscopy (AFM).

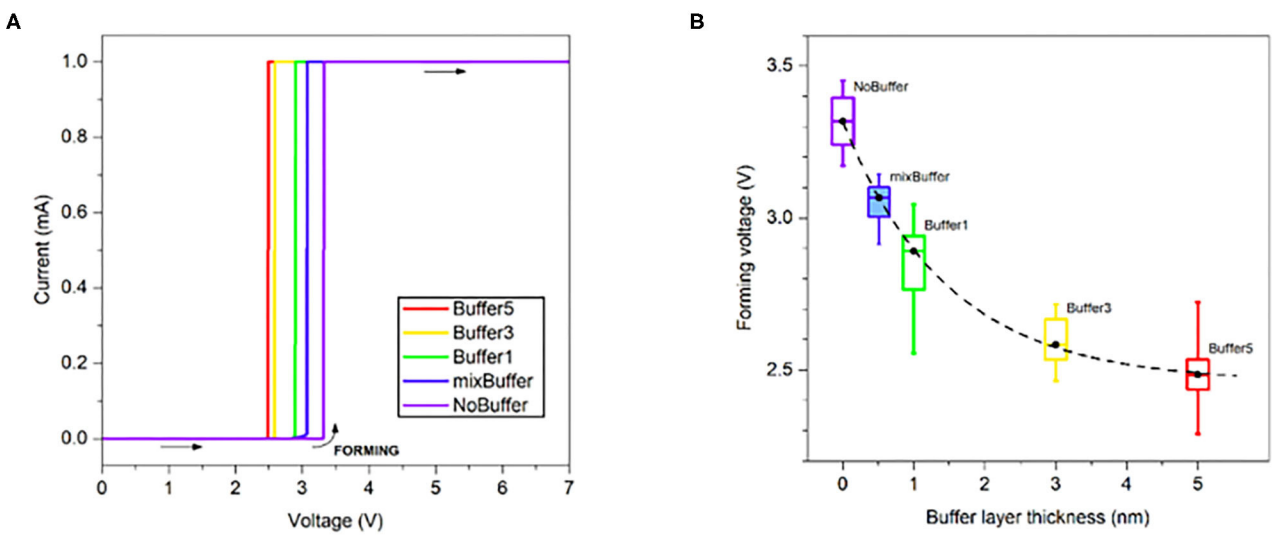


FIGURE 2 | The median values of $V_{FORMING}$ for each material stack show a dependence of the forming process on the buffer layer. In **(A)**, the corresponding *I–V* characteristics are shown, while in **(B)** their exponential decay as a function of the Ti-based buffer layer thickness is presented.

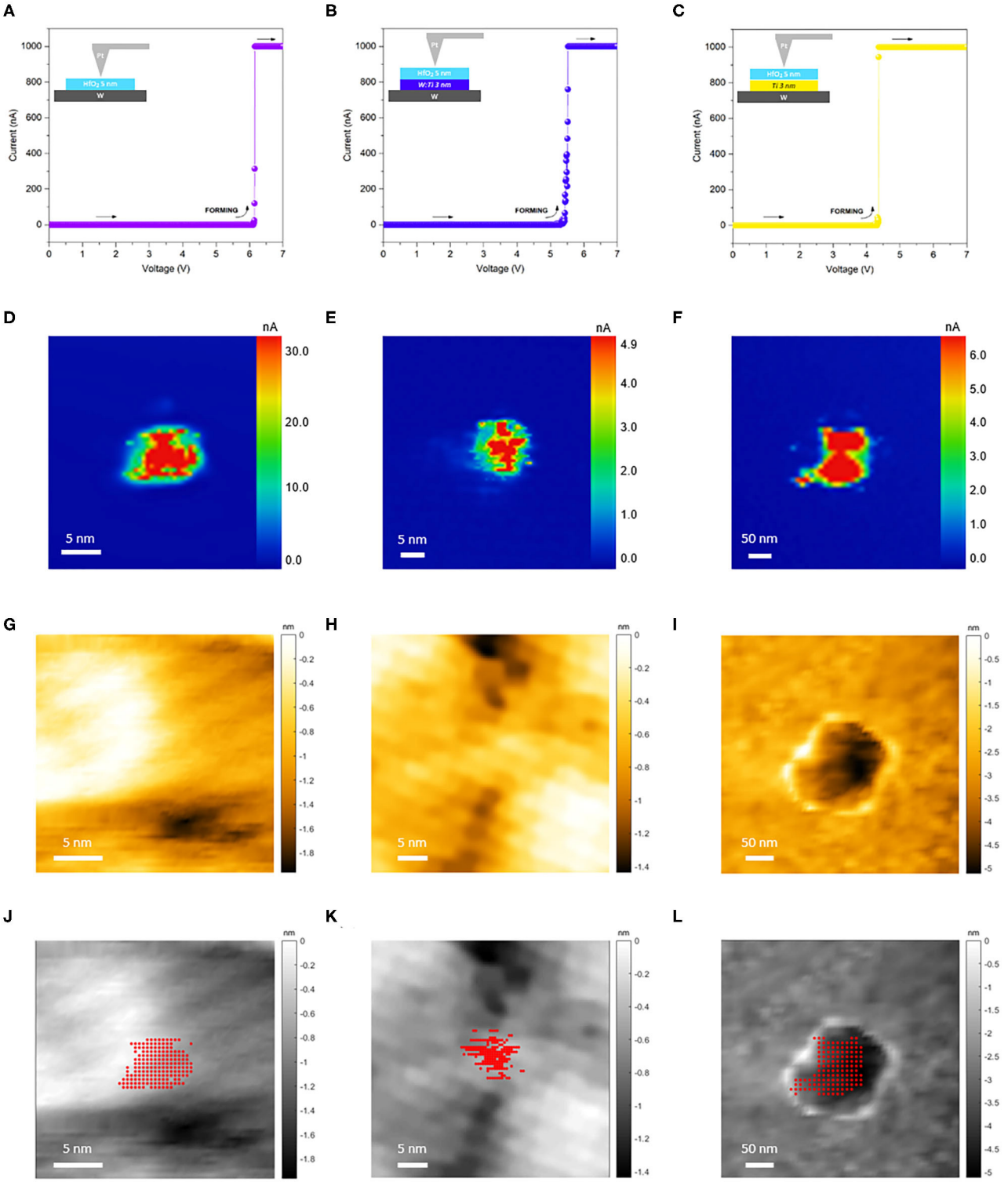


FIGURE 3 | Investigation of the forming process was performed with conductive atomic force microscopy (C-AFM) directly probing the oxide layer by means of full-platinum tips. The resulting *I–V* characteristics are reported in **(A–C)**, where $V_{FORMING}$ is reported to decrease with increasing amount of titanium in the buffer layer. In **(D–F)**, the current maps acquired after the forming process are reported, with conductive spots clearly shown for all the tested material stacks. The topography images obtained at the same time with the current maps are then reported in **(G–I)**, showing the presence of morphological changes in the case of the Buffer3 samples **(I)**. In **(J–L)**, the superposition of topographical and electrical images is presented.

by the box plot in **Figure 2B**, such a reduction turned out to be well-described by an exponential decay of the median values of $V_{FORMING}$ for the samples with pure Ti buffer layers. The same curve was then employed, the other way around, to define an effective thickness of the mixed buffer layer, which came out to be about 0.5 nm. Such an effective thickness, smaller than the real one of 3 nm, clarifies that the key player in the reduction of the forming voltage is not properly the thickness of the buffer layer but rather the presence, and the amount, of titanium between the hafnium dioxide film and the tungsten electrode.

Consistent results were shown by the C-AFM characterization too. As reported in **Figures 3A–C**, three different stacks were investigated, namely NoBuffer, mixBuffer, and Buffer3, with the structure sketched in the insets of **Figures 3A–C**. As is clear from those pictures, the same MIM geometry as the one schematized in **Figure 1A** was reproduced thanks to the full-platinum AFM tip, which played the role of the inert electrode. By selecting such triplet of stacks, the key points of the previous analysis were further investigated. Indeed, with this set of devices, two main comparisons were possible, namely (i) the case with or without the buffer layer and (ii) the case of pure or mixed titanium with a fixed thickness. For each sample investigated by means of C-AFM, forming was induced first and current maps were produced afterwards. For both the measurements, the platinum tip was kept grounded and voltages were applied at the tungsten electrode. To achieve forming, voltage sweeps from 0 to 7 V were employed as for the analysis carried out through the parameter analyzer, with a current compliance set at 1 μA. Current maps were instead produced applying fixed voltages. Similarly to **Figures 2A**, **3A–C** show a reduction of $V_{FORMING}$ depending on the presence of titanium between the oxidizing electrode and the oxide layer. Furthermore, in accordance with **Figure 2B**, such a reduction turned out to be related to the amount of titanium in the buffer layer rather than to the physical thickness of the Ti-based layer only. Indeed, both the buffer layers employed for the mixBuffer and Buffer3 samples are 3 nm thick, but their compositions differ from one another, with the mixBuffer sample exhibiting a tungsten film with 10% in weight of titanium instead of a pure Ti film. The current maps, produced after forming was induced, are then reported in **Figures 3D–F**. They show a nanometer-sized conductive spot for each sample, which strongly suggests a filamentary nature of the resistive switching in the tested devices. The topography images, acquired simultaneously with the creation of the current maps, are reported too (**Figures 3G–I**), and a superposition of the current maps on the topography images is presented in **Figures 3J–L** as the result of a point-by-point analysis of the electrical conduction. Particularly, in the case of Buffer3 samples, the presence of morphological changes is reported (**Figure 3I**), and **Figure 3L** highlights that such modifications turn out to perfectly match with the conductive spot found in the current map.

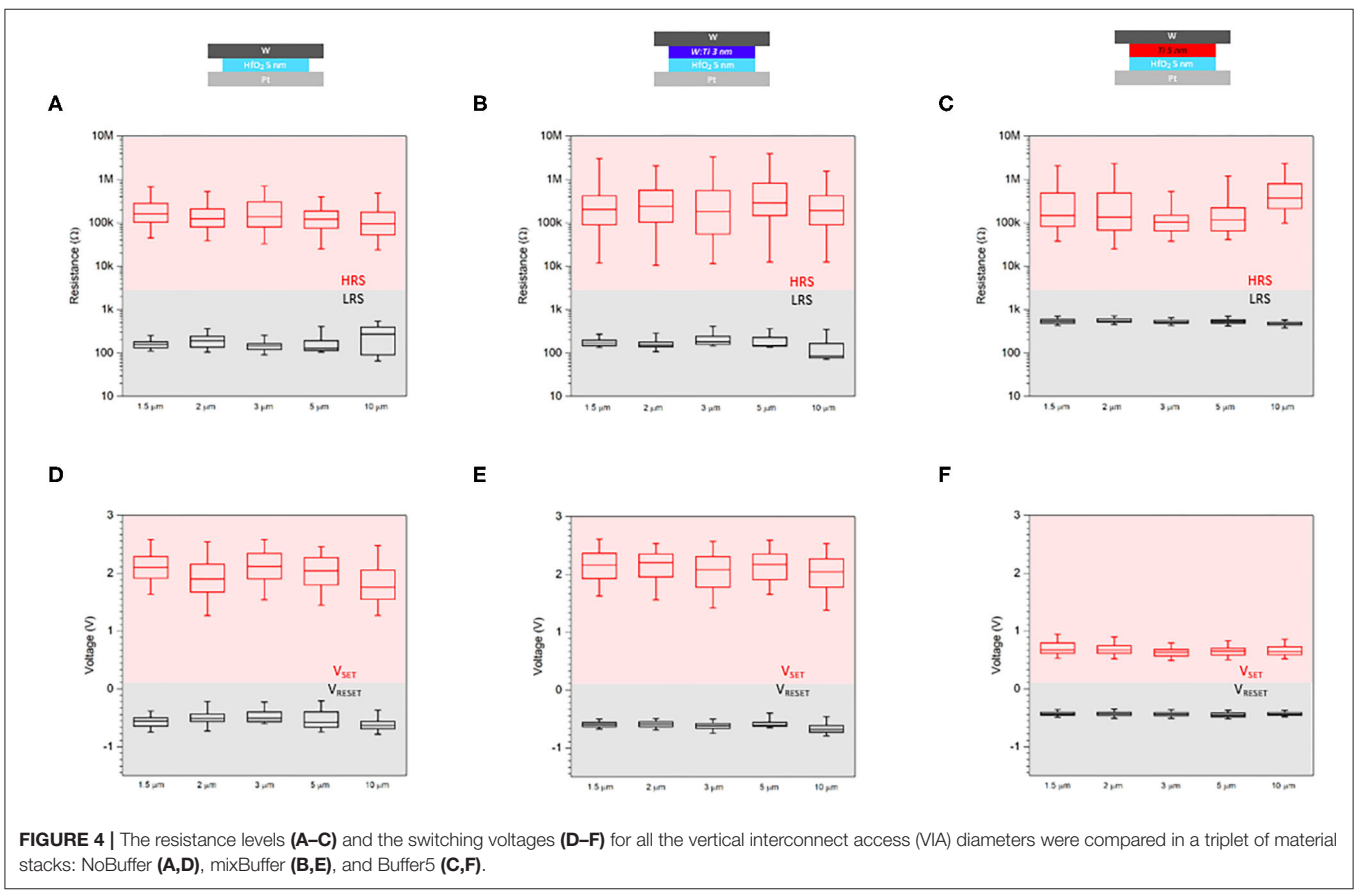


FIGURE 4 | The resistance levels **(A–C)** and the switching voltages **(D–F)** for all the vertical interconnect access (VIA) diameters were compared in a triplet of material stacks: NoBuffer **(A,D)**, mixBuffer **(B,E)**, and Buffer5 **(C,F)**.

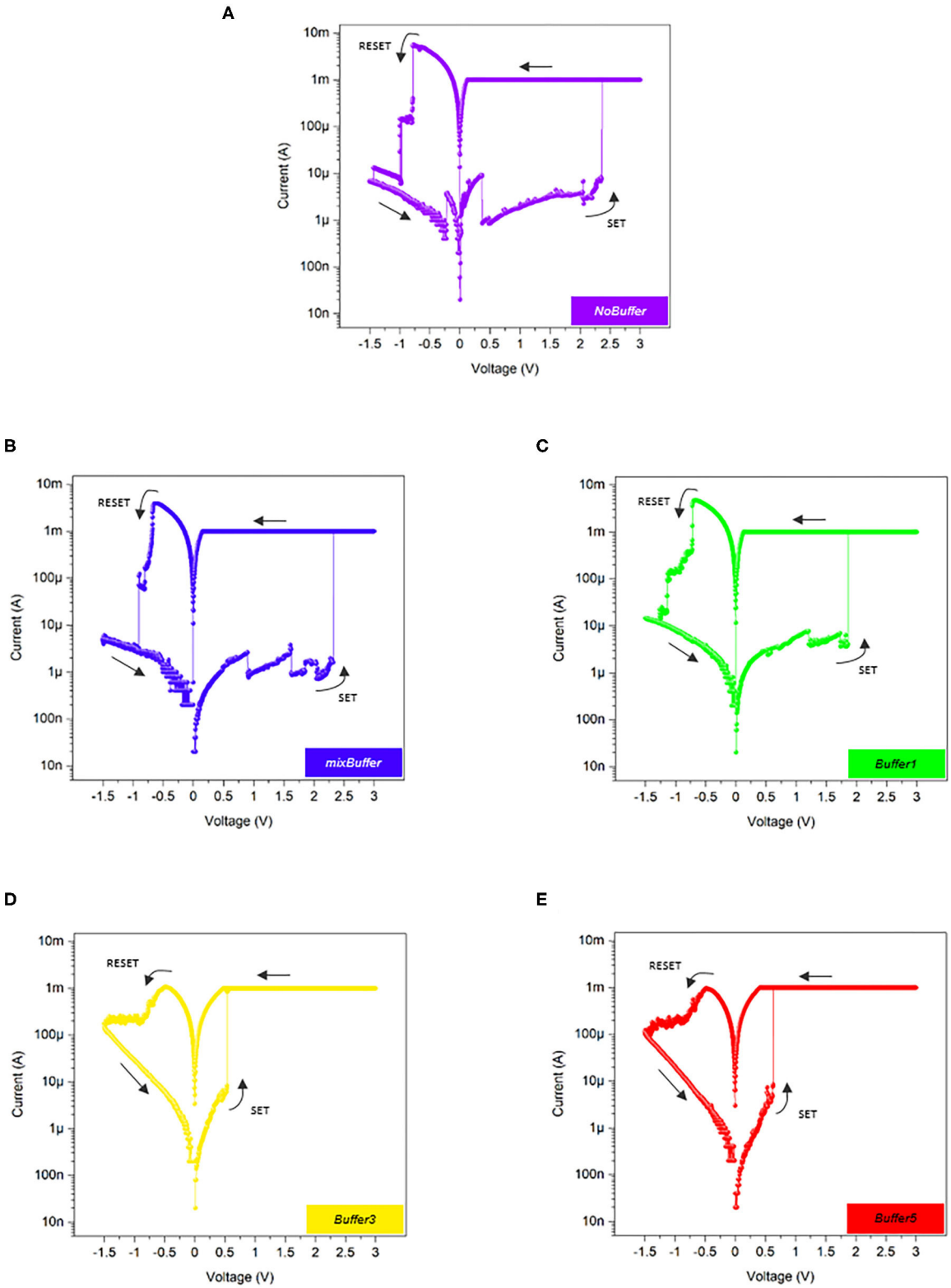


FIGURE 5 | Applying bipolar voltage sweeps, the switching behavior of the tested device was investigated through the resulting *I–V* characteristics. The reported graphs are representative curves for each material stack. In the NoBuffer samples **(A)**, current fluctuations occur before the SET process and RESET takes place abruptly. A similar behavior is shown by the mixBuffer **(B)** and Buffer1 **(C)** samples also. In the Buffer3 **(D)** and Buffer5 **(E)** samples, instead, the current fluctuations in the high-resistance state (HRS) in positive polarity are not observed; a lower V_{SET} is found and gradual RESET occurs as it is flagged by a smooth current decrease for negative voltages close to V_{RESET}.

Even though the C-AFM analysis clearly reported the presence, in all the tested material stacks, of conductive spots at the HfO_2/Pt interface, suggesting resistive switching of filamentary type, a statistical DC characterization was performed to exclude a dependence of the RS on the device area as a consequence of interfacial effects at the W/HfO_2 or buffer layer/HfO_2 interface. By means of the parameter analyzer, bipolar voltage sweeps in the range −1.5 to 3 V were applied on 25 devices for the NoBuffer, mixBuffer, and Buffer5 samples. As summarized by the box plots in **Figure 4**, such characterization revealed that RS parameters like the resistance levels, V_{SET}, and V_{RESET} are independent of the device area since no correlation was found between these quantities and the diameter of the VIAs (1.5, 2, 3, 5, and 10 μm). Therefore, we can conclude that the observed resistive switching can be truly ascribed to a filamentary mechanism.

Through the same DC characterization, that is to say applying consecutive cycles of bipolar voltage sweeps 0 V → −1.5 V → 3 V → 0 V, the switching behavior of the devices was investigated. As is reported in **Figure 5**, where *I*–*V* characteristics representative of a typical cycle for each material stack are shown, two different behaviors can be highlighted in the DC operation regime for Ti-based buffer layers thinner or thicker than 3 nm, respectively. More in detail, starting from **Figure 5A**, which reports the cycling behavior of the devices with no buffer layer, clear current fluctuations can be appreciated in the HRS for positive voltages. Interpreted from a different, but complementary, perspective, **Figure 5A** shows that the devices without a buffer layer exhibit some instability during the SET process, with sharp transitions from HRS to LRS occurring at relatively high voltages only after quick, repeated current variations. A similar behavior can be observed in the case of the mixBuffer (**Figure 5B**) and Buffer1 (**Figure 5C**) samples too, while a clear change occurs in the Buffer3 (**Figure 5D**) and Buffer5 (**Figure 5E**) samples. The latter two, indeed, still exhibit abrupt switching from HRS to LRS, but the sharp transition takes place at lower voltages and the *I*–*V* characteristics in HRS in positive polarity are much more stable, with no fluctuations. Moreover, as is graphically summarized in **Figure 6A**, which reports the statistical variations of V_{SET} and V_{RESET} obtained from the DC characterization, such improved stability in the device operation is coupled to a significantly reduced device-to-device variability.

The second major result arising from the insertion of a Ti-based buffer layer, which becomes apparent for Buffer3 and Buffer5 samples as for the HRS stability above-mentioned, involves the opposite polarity and the opposite transition. In **Figures 5D,E**, indeed, a fairly different behavior in the transition from LRS to HRS can be appreciated, with a gradual resistance variation instead of an abrupt switch. Interestingly, such change does not reflect into an increased device stability or reduced device-to-device variability. As reported in **Figure 6A**, indeed, differently from V_{SET}, the RESET voltage does not significantly vary neither from a device to another nor from a sample to another. Similar observations can be made for the resistance values also, whose statistical analysis is reported in **Figure 6B**. In this case, the effect of Ti-based buffer layers as a stabilizer can be appreciated looking at the variability of the LRS in the different samples. As is clear from the box plot, the resistance value of the highly conductive state is significantly more stable in Buffer3 and Buffer5 samples, while a relevant device-to-device variability affects the devices based on the other material stacks.

The statistical parameters resulting from the DC characterization performed on a total of 125 devices are summarized in **Table 2**.

Based on reaction (1), both the current fluctuations before SET occurs and the gradual RESET can be interpreted referring to oxygen exchanges, which, in turn, involve the oxidizing activity

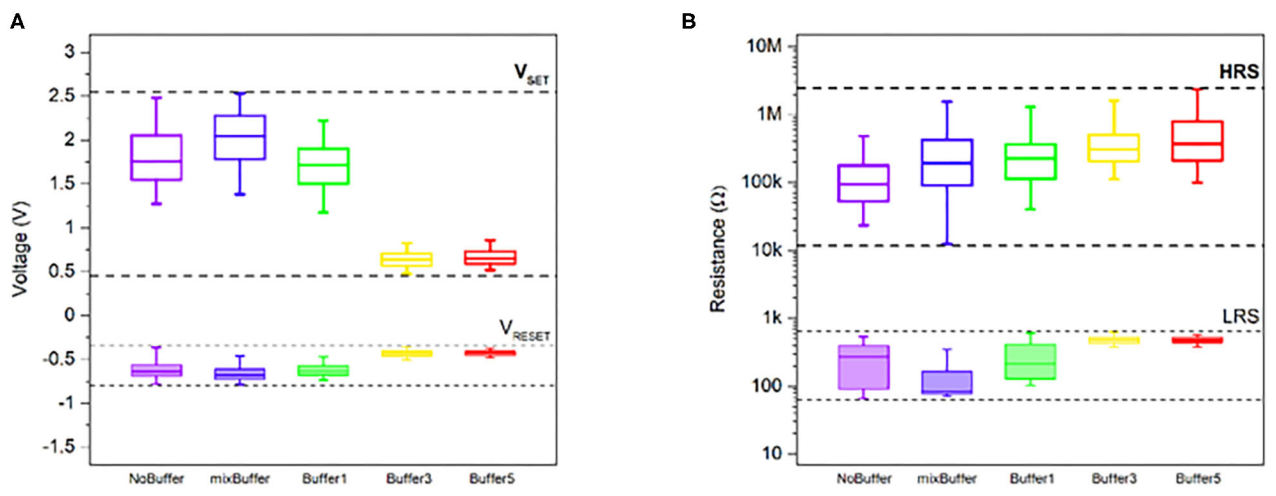


FIGURE 6 | The statistical analysis of the DC characterization performed on 25 devices for each material stack is presented by means of *box plots*. In **(A)**, the switching voltages are reported, while the resistance levels are shown in **(B)**. In both cases, for the sake of clarity, *dashed lines* are employed to distinguish the different ranges of data.

TABLE 2 | Median values and standard deviations from the statistical DC characterization.

Sample name	$V_{FORMING}$ (V)		V_{SET} (V)		V_{RESET} (V)		HRS (kΩ)		LRS (Ω)	
	Median	SD	Median	SD	Median	SD	Median	SD	Median	SD
noBuffer	3.31	0.09	1.76	0.38	−0.64	0.12	94.48	154.61	272.55	171.05
mixBuffer	3.07	0.08	2.05	0.34	−0.68	0.10	192.86	3711.84	84.86	84.86
Buffer1	2.89	0.15	1.72	0.32	−0.63	0.09	226.78	427.42	214.86	169.08
Buffer3	2.58	0.09	0.64	0.10	−0.43	0.04	309.41	499.79	483.00	77.28
Buffer5	2.49	0.12	0.65	0.10	−0.43	0.03	368.39	1011.09	474.82	69.09

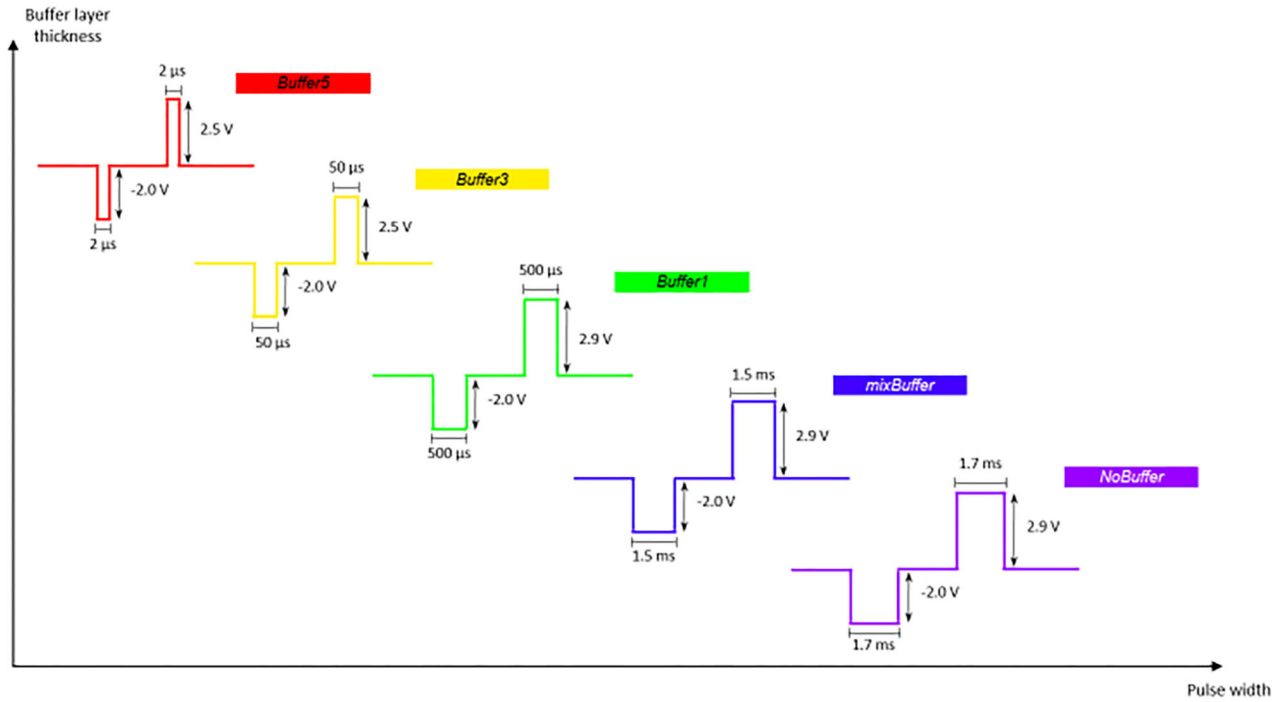


FIGURE 7 | The insertion of buffer layers with increasing amount of titanium translates into a faster response of the devices, as shown by the pulse width reduction achieved with thicker Ti-based buffer layers.

of the layers in contact with the hafnium dioxide. Since, as already mentioned, one of these layers is always made of platinum, which is inert, the two phenomena must be related to the buffer layer, or to the tungsten electrode when the former is not present. In this view, the interpretation of the observed behavior in the DC regime can be traced back to the different oxidizing characteristics of titanium and tungsten. Current fluctuations may be related to the multiple, metastable oxides tungsten can form before reaching the stable WO_3 (Lassner and Schubert, 1999; Shahrabi et al., 2019) since the emergence of such fluctuations can be appreciated in the NoBuffer, mixBuffer, and Buffer1 samples only. In the Buffer3 and Buffer5 samples, indeed, the thickness of the buffer layers is such that a large enough amount of titanium is present to effectively hinder the formation of metastable tungsten oxides (Shahrabi et al., 2019). On the other hand, concerning the transition from an abrupt to a gradual RESET, similar arguments hold, and the smoother resistance change can be again ascribed to the oxidizing behavior of the buffer layer. The gradual transition, indeed, takes place in the Buffer3 and Buffer5 samples only, that is to say, once more, only in those devices with a large enough amount of titanium between the tungsten electrode and the hafnium dioxide.

A further confirmation of the stabilizing effect given by the titanium buffer layer was then found with pulse tests aimed at investigating the endurance of the devices, namely their cycling reliability. For each material stack, an initial optimization procedure was first performed on the pulse parameters in order to find the best combinations of pulse width and pulse amplitude. As is shown in **Figure 7**, pulses were optimized for both SET and RESET since, as shown by the DC characterization, the bipolar RS of the tested devices is not symmetrical. Specifically, $|V_{SET}|$ turned out to be higher than $|V_{RESET}|$, while identical pulse widths were used in both polarities. Finally, a delay of 200 ms was always employed between a pulse and the following one.

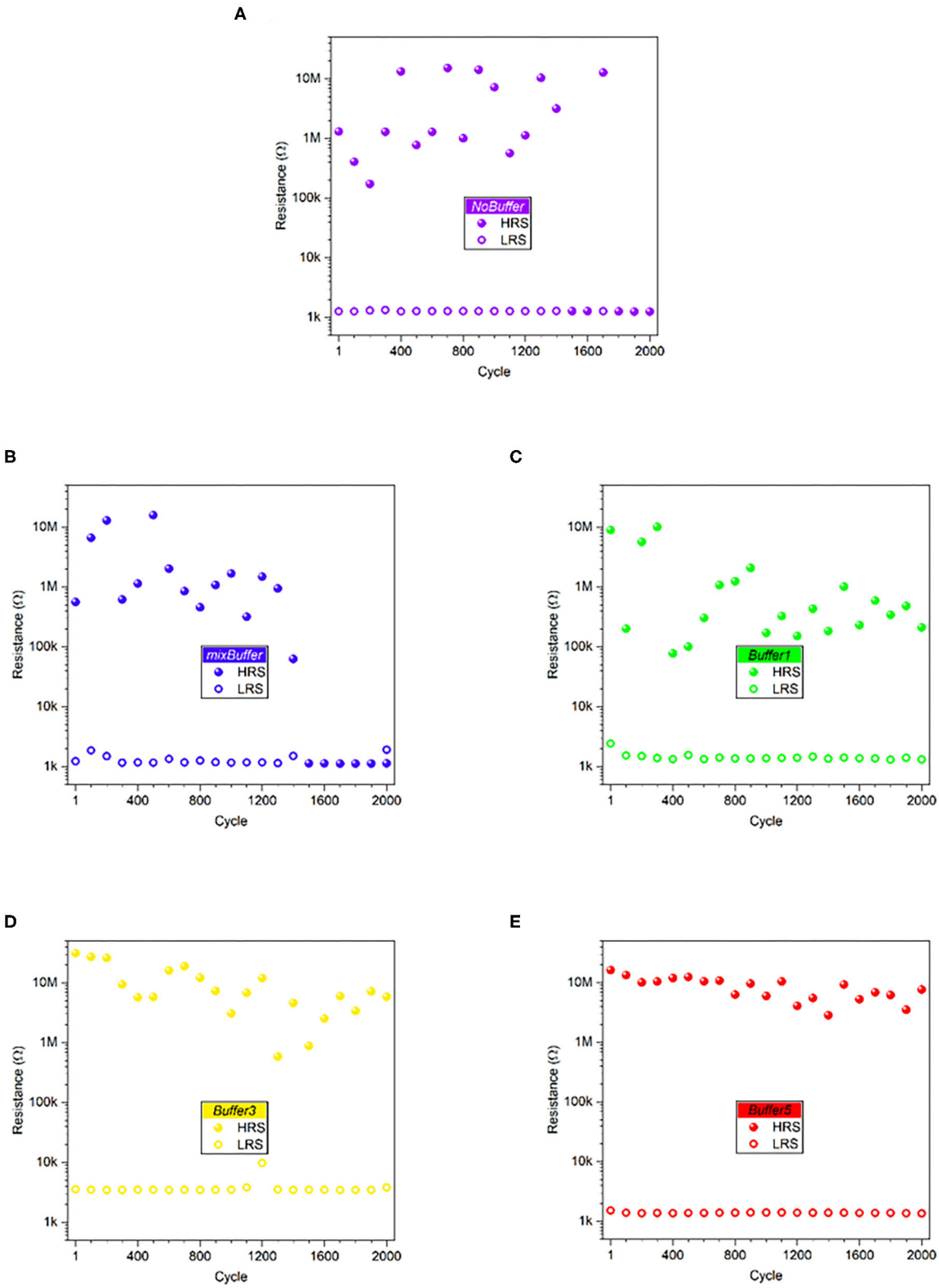


FIGURE 8 | Endurance tests were performed with a fixed sequence of 2,000 SET–RESET pulse pairs. The NoBuffer **(A)** and mixBuffer **(B)** samples showed high-resistance state (HRS) failure after about 1,500 cycles, while the devices from the other material stacks **(C–E)** successfully completed the fixed-length test. Additionally, a stabilization of the HRS was observed for the Buffer1 **(C)**, Buffer3 **(D)**, and Buffer5 **(E)** samples.

TABLE 3 | Summary of resistance values during endurance and retention measurements.

Sample name	Endurance				Retention			
	HRS		LRS		HRS		LRS	
	Mean (MΩ)	Relative uncertainty (%)	Mean (kΩ)	Relative uncertainty (%)	Mean (MΩ)	Relative uncertainty (%)	Mean (kΩ)	Relative uncertainty (%)
noBuffer	4.73	119	1.27	1.4	13.22	30	1.47	4.8
mixBuffer	3.10	156	1.27	18	1.26	2.0	1.00	2.3
Buffer1	1.61	180	1.44	16	0.93	2.9	1.35	2.3
Buffer3	10.11	87	3.81	36	32.26	3.3	2.52	3.9
Buffer5	8.52	41	1.38	2.3	17.02	5.1	1.27	6.3

HRS, high-resistance state; LRS, low-resistance state.

Figure 7 clearly shows that such an optimization revealed a key impact of titanium on the dynamical operation regime of the tested devices. Indeed, besides the reduction of pulse amplitude needed for successful RS, which was already pointed out with the DC characterization, pure Ti buffer layers turned out to lead to a pulse width reduction down to three orders of magnitude with respect to the devices without a buffer layer.

Once the pulse parameter optimization was completed, a common test procedure was defined and adopted for all the material stacks, so that a clear performance comparison among the different samples was possible. Specifically, all the devices subjected to the endurance test were subjected to 2,000 SET–RESET pulse pairs aimed at continuously switching between HRS and LRS. The results of this characterization, reported in **Figure 8**, show how, besides improving the device stability, pure Ti buffer layers also have a beneficial effect on the endurance itself. **Figures 8A,B**, indeed, reveal that the NoBuffer and mixBuffer samples suffered for HRS failures preventing them from reaching the common test length of 2,000 cycles. Particularly, the devices from both material stacks were not able to overcome 1,500 cycles. Conversely, the Buffer1, Buffer3, and Buffer5 samples were all able to reach the fixed benchmark of 2,000 cycles, thus demonstrating an improvement of about 30%. Moreover, as already mentioned, the stability of the devices significantly improved, as is highlighted in **Figures 8C–E** by the much less scattered data as the amount of titanium in the buffer layers increases. A quantitative evaluation of the data dispersion can be made through the relative uncertainty (**Table 3**), which, in the case of HRS, turns out to be smaller than 100% for the Buffer3 and Buffer5 samples only. In more detail, such samples provide relative uncertainties of 87 and 41%, respectively, while 119% is found for the NoBuffer samples, 156% in the case of the mixBuffer ones, and 180% for the devices coming from the Buffer1 samples. The relative uncertainties are instead much smaller for the LRS during the endurance tests, and they do not show any trend related to the material stack.

To complete the set of electrical characterizations, retention tests were performed on new samples to compare the capability of the different material stacks of preserving each resistance state. A summary of the mean values, together with their relative uncertainties, for both HRS and LRS during pulse operations is presented in **Table 3**. As for the endurance, a common benchmark was set for the retention tests too, and 2×10^4 s was adopted as the fixed length for the measurements in order to define a standard procedure for all the samples. The results are shown in **Figure 9**, where a good stability is reported for all the material stacks. In this case, no significant difference can be appreciated depending on the buffer layer, with all the devices able to reach the fixed value of 2×10^4 s and small relative uncertainties on the resistance values, in both HRS and LRS. As a consequence, the retention tests, which rely on the stability of the conductive filament rather than on its formation and rupture, seem to suggest that the Ti-based buffer layers actively play a role only when oxygen exchanges, as in reaction (1), occur, while they remain silent otherwise. The performance tunability and improvement titanium allows to achieve can hence be directly related to the stabilization of the interactions between the oxidizing electrode and the oxide layer. Compared to tungsten, titanium indeed requires a much lower energy to create oxygen vacancies in HfO_2, and this significantly hinders the slower tungsten oxidation (Kim et al., 2016). As a consequence, titanium efficiently mitigates the fluctuations induced by the formation of metastable tungsten oxides (Lassner and Schubert, 1999) and allows faster responses from the devices.

DISCUSSION

With this work, we have shown how titanium can be employed, as a buffer layer, to stabilize and tune the RS performances of ReRAM cells based on CMOS-compatible materials like HfO_2 and tungsten. With an extensive, systematic approach, 125 devices with different material stacks have been tested. Investigating different thicknesses and compositions of the Ti-based films, a dependence of the device performances on the buffer layer properties was found, and the amount of titanium in the buffer layer turned out to play a key role. The presented results can be ascribed to the different oxidizing characteristics of titanium and tungsten. The latter, indeed, suffers from a relatively slower oxidation process, producing a variety of metastable oxides, responsible for the RS instability which clearly appears in both DC switching and pulse operations. Employing buffer

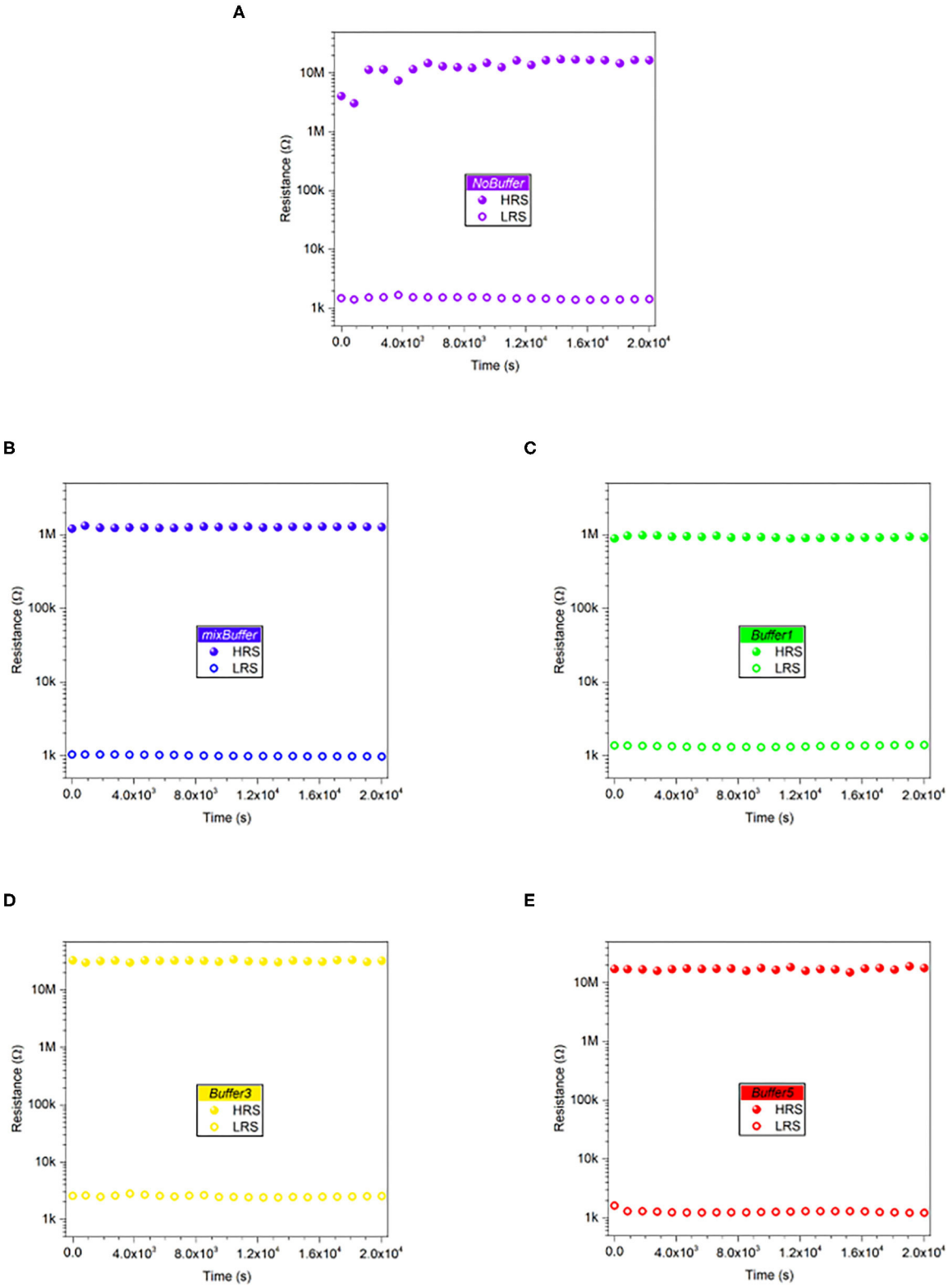


FIGURE 9 | (A–E) Retention measurements, performed for a fixed time of 2×10^4 s, highlighted that a good resistance state stability is provided by all the material stacks, with no significant contribution given by titanium.

layers with a high enough amount of titanium, relevant changes in the device performances have been reported. More in detail, the response speed has been shown to significantly increase according to the pulse width reduction of three orders of magnitude; an improvement of about 30% has been achieved in terms of endurance performance, and an increased stability of the resistance states, especially the HRS, has been obtained in the dynamic operation regime. In light of these results, the Ti buffer/W stack turns out to be a suitable choice for CMOS-compatible ReRAM cells that have to solve reliability issues coming from tungsten electrodes. Furthermore, the possibility of tuning the device performances according to the Ti-based buffer layer properties may open the way to the definition of new design rules for ReRAM integration with standard CMOS technology.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

VF and ES contributed to the design and fabrication of the devices. VF performed device characterization. VF and CR wrote and revised the manuscript. CR and YL helped with supervision. All authors contributed to the article and approved the submitted version.

FUNDING

This research was partially supported by Compagnia di San Paolo through the project for internationalization of research of Politecnico di Torino.

ACKNOWLEDGMENTS

The help and the technical support given by the staff of the Center of Micro-nano Technology (CMi) of EPFL are gratefully acknowledged. A hearthfelt thank goes to P. Charpilloz for his supervision during AFM measurements.

REFERENCES

Ambrosi, E., Bricalli, A., Laudato, M., and Ielmini, D. (2019). Impact of oxide and electrode materials on the switching characteristics of oxide ReRAM devices. *Faraday Discuss.* 213, 87–98. doi: 10.1039/C8FD00106E

Baek, I. G., Lee, M. S., Sco, S., Lee, M. J., Seo, D. H., Suh, D.-S., et al. (2005). "Highly scalable non-volatile resistive memory using simple binary oxide driven by asymmetric unipolar voltage pulses," in *IEDM Technical Digest. IEEE International Electron Devices Meeting, 2004.* (San Francisco, CA), 587–590.

Cai, F., Correll, J. M., Lee, S. H., Lim, Y., Bothra, V., Zhang, Z., et al. (2019). A fully integrated reprogrammable memristor–CMOS system for efficient multiply–accumulate operations. *Nat. Electron* 2, 290–299. doi: 10.1038/s41928-019-0270-x

Celano, U., Fantini, A., Degraeve, R., Jurczak, M., Goux, L., and vandervorst, W. (2016). Scalability of valence change memory: from devices to tip-induced filaments. *AIP Adv.* 6:085009. doi: 10.1063/1.4961150

Celano, U., Gastaldi, C., Govoreanu, B., Richard, O., Bender, H., Goux, L., et al. (2017). Evidences of areal switching in vacancy-modulated conductive oxide (VMCO) memory. *Microelectron. Eng.* 178, 122–124. doi: 10.1016/j.mee.2017.04.046

Celano, U., Goux, L., Belmonte, A., Opsomer, K., Franquet, A., Schulze, A., et al. (2014). Three-dimensional observation of the conductive filament in nanoscaled resistive memory devices. *Nano Lett.* 14, 2401–2406. doi: 10.1021/nl500049g

Chen, Y., Li, L., Yin, X., Yerramilli, A., Shen, Y., Song, Y., et al. (2017). Resistive switching characteristics of flexible TiO2 thin film fabricated by deep ultraviolet photochemical solution method. *IEEE Electron Device Lett.* 38, 1528–1531. doi: 10.1109/LED.2017.2756444

Chen, Y. S., Lee, H. Y., Chen, P. S., Gu, P. Y., Chen, C. W., Lin, W. P., et al. (2009). "Highly scalable hafnium oxide memory with improvements of resistive distribution and read disturb immunity," in *Tech Dig Int Electron Devices Meet IEDM* (Baltimore, MD), 105–108. doi: 10.1109/IEDM.2009.5424411

Chen, Y. Y., Goux, L., Clima, S., Govoreanu, B., Degraeve, R., Kar, G. S., et al. (2013). Endurance/retention trade-off on HfO2/Metal Cap 1T1R Bipolar RRAM. *IEEE Trans. Electron Device* 60, 1114–1121. doi: 10.1109/TED.2013.2241064

Conti, D., Laurenti, M., Porro, S., Giovinazzo, C., Bianco, S., Fra, V., et al. (2019). Resistive switching in sub-micrometric ZnO polycrystalline films. *Nanotechnology* 30:065707. doi: 10.1088/1361-6528/aaf261

Gibbons, J. F., and Beadle, W. E. (1964). Switching properties of thin NiO films. *Solid State Electron* 7, 785–790. doi: 10.1016/0038-1101(64)90131-5

Gilmer, D. C., Bersuker, G., Park, H. Y., Park, C., Butcher, B., Wang, W., et al. (2011). "Effects of RRAM stack configuration on forming voltage and current overshoot," in *2011 3rd IEEE Int Mem Work IMW* (Monterey, CA), 1–4. doi: 10.1109/IMW.2011.5873225

Govoreanu, B., Redolfi, A., Zhang, L., Adelmann, C., Popovici, M., Clima, S., et al. (2017). "Vacancy-modulated conductive oxide resistive RAM (VMCO-RRAM): an area-scalable switching current, self-compliant, highly nonlinear and wide on/off-window resistive switching cell," in *2013 IEEE International Electron Devices Meeting* (Washington, DC: IEEE), 10.2.1–10.2.4.

Guo, Y., and Robertson, J. (2014). Materials selection for oxide-based resistive random access memories. *Appl. Phys. Lett.* 105:223516. doi: 10.1063/1.4903470

Ielmini, D., and Waser, R. (2016). *Resistive Switching*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. doi: 10.1002/9783527680870

Joshua Yang, J., Miao, F., Pickett, M. D., Ohlberg, D. A. A., Stewart, D. R., Lau, C. N., et al. (2009). The mechanism of electroforming of metal oxide memristive switches. *Nanotechnology* 20:215201. doi: 10.1088/0957-4484/20/21/215201

Kim, W., Menzel, S., Wouters, D. J., Guo, Y., Robertson, J., Roesgen, B., et al. (2016). Impact of oxygen exchange reaction at the ohmic interface in Ta 2 O 5 -based ReRAM devices. *Nanoscale* 8, 17774–17781. doi: 10.1039/C6NR03810G

Kindsmüller, A., Meledin, A., Mayer, J., Waser, R., and Wouters, D. J. (2019). On the role of the metal oxide/reactive electrode interface during the forming procedure of valence change ReRAM devices. *Nanoscale* 11, 18201–18208. doi: 10.1039/C9NR06624A

Lamb, D. R., and Rundle, P. C. (1967). A non-filamentary switching action in thermally grown silicon dioxide films. *Br. J. Appl. Phys.* 18, 29–32. doi: 10.1088/0508-3443/18/1/306

Lassner, E., and Schubert, W.-D. (1999). *Tungsten*. Boston, MA: Springer US. doi: 10.1007/978-1-4615-4907-9

Lee, J. H., Wu, C., Sung, S., An, H., and Kim, T. W. (2019). Highly flexible and stable resistive switching devices based on WS2 nanosheets:poly(methylmethacrylate) nanocomposites. *Sci. Rep.* 9:19316. doi: 10.1038/s41598-019-55637-2

Lee, M.-J., Kang, B. S., Ahn, S.-E., Kim, K. H., Lee, C. B., Kim, C. J., et al. (2009). Electrical manipulation of nanofilaments in transition-metal oxides for resistance-based memory. *Nano Lett.* 9, 1476–1481. doi: 10.1021/nl803387q

Lin, Y. S., Zeng, F., Tang, S. G., Liu, H. Y., Chen, C., Gao, S., et al. (2013). Resistive switching mechanisms relating to oxygen vacancies migration in both interfaces in Ti/HfOx/Pt memory devices. *J. Appl. Phys.* 113:064510. doi: 10.1063/1.4791695

Marconi, G. (1899). Wireless telegraphy. *J. Inst. Electr. Eng.* 28, 273–290. doi: 10.1049/jiee-1.1899.0010

Meena, J. S., Sze, S. M., Chand, U., and Tseng, T. Y. (2014). Overview of emerging nonvolatile memory technologies. *Nanoscale Res. Lett.* 9:526. doi: 10.1186/1556-276X-9-526

Nafria, M., Rodriguez, R., Porti, M., Martin-Martinez, J., Crespo-Yepes, A., Claramunt, S., et al. (2017). (Invited) advanced measurement techniques for the characterization of ReRAM devices. *ECS Trans.* 79, 139–148. doi: 10.1149/07901.0139ecst

Rahaman, S. Z., Lin, Y.-D., Lee, H.-Y., Chen, Y.-S., Chen, P.-S., Chen, W.-S., et al. (2017). The role of Ti buffer layer thickness on the resistive switching properties of hafnium oxide-based resistive switching memories. *Langmuir* 33, 4654–4665. doi: 10.1021/acs.langmuir.7b00479

Shahrabi, E., Giovinazzo, C., Hadad, M., LaGrange, T., Ramos, M., Ricciardi, C., et al. (2019). Switching kinetics control of W-based ReRAM cells in transient operation by interface engineering. *Adv Electron Mater.* 5:1800835. doi: 10.1002/aelm.201800835

Singh, A. K., Blonkowski, S., and Kogelschatz, M. (2018). Resistive switching study in HfO2 based resistive memories by conductive atomic force microscopy in vacuum. *J. Appl. Phys.* 124, 014501. doi: 10.1063/1.5025143

Sun, Y. M., Song, C., Yin, J., Qiao, L. L., Wang, R., Wang, Z. Y., et al. (2019). Modulating metallic conductive filaments via bilayer oxides in resistive switching memory. *Appl. Phys. Lett.* 114:193502. doi: 10.1063/1.5098382

Tang, J., Yuan, F., Shen, X., Wang, Z., Rao, M., He, Y., et al. (2019). Bridging biological and artificial neural networks with emerging neuromorphic devices: fundamentals, progress, and challenges. *Adv. Mater.* 31:1902761. doi: 10.1002/adma.201902761

Tappertzhofen, S., Waser, R., and Valov, I. (2014). Impact of the counter-electrode material on redox processes in resistive switching memories. *ChemElectroChem* 1, 1287–1292. doi: 10.1002/celc.201402106

Tirano, S., Perniola, L., Buckley, J., Cluzel, J., Jousseaume, V., Muller, C., et al. (2011). Accurate analysis of parasitic current overshoot during forming operation in RRAMs. *Microelectron. Eng.* 88, 1129–1132. doi: 10.1016/j.mee.2011.03.062

Valov, I. (2014). Redox-based resistive switching memories (ReRAMs): Electrochemical systems at the atomic scale. *ChemElectroChem* 1, 26–36. doi: 10.1002/celc.201300165

Valov, I., Waser, R., Jameson, J. R., and Kozicki, M. N. (2011). Electrochemical metallization memories—fundamentals, applications, prospects. *Nanotechnology* 22:254003. doi: 10.1088/0957-4484/22/28/289502

Walczyk, D., Bertaud, T., Sowinska, M., Lukosius, M., Schubert, M. A., Fox, A., et al. (2012). "Resistive switching behavior in TiN/HfO2/Ti/TiN devices," in *2012 International Semiconductor Conference Dresden-Grenoble (ISCDG)* (Grenoble: IEEE), 143–146. doi: 10.1109/ISCDG.2012.6360035

Wang, Z., Wu, H., Burr, G. W., Hwang, C. S., Wang, K. L., Xia, Q., et al. (2020). Resistive switching materials for information processing. *Nat. Rev. Mater.* 5, 173–195. doi: 10.1038/s41578-019-0159-3

Waser, R., and Aono, M. (2007). Nanoionics-based resistive switching memories. *Nat. Mater.* 6, 833–840. doi: 10.1038/nmat2023

Wong, H. S. P., Lee, H. Y., Yu, S., Chen, Y. S., Wu, Y., Chen, P. S., et al. (2012). Metal-oxide RRAM. *Proc. IEEE* 100, 1951–1970. doi: 10.1109/JPROC.2012.2190369

Xia, Q., and Yang, J. J. (2019). Memristive crossbar arrays for brain-inspired computing. *Nat. Mater.* 18, 309–323. doi: 10.1038/s41563-019-0291-x

Yang, J. J., Zhang, M. X., Strachan, J. P., Miao, F., Pickett, M. D., Kelley, R. D., et al. (2010). High switching endurance in TaOx memristive devices. *Appl. Phys. Lett.* 97, 6–9. doi: 10.1063/1.3524521

Yen, T. J., Gismatulin, A., Volodin, V., Gritsenko, V., and Chin, A. (2019). All nonmetal resistive random access memory. *Sci. Rep* 9:6144. doi: 10.1038/s41598-019-42706-9

Zidan, M. A., Jeong, Y. J., Lee, J., Chen, B., Huang, S., Kushner, M. J., et al. (2018). A general memristor-based partial differential equation solver. *Nat Electron* 1, 411–420. doi: 10.1038/s41928-018-0100-6

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.

frontiers in Nanotechnology

REVIEW
published: 24 March 2021
doi: 10.3389/fnano.2021.645995

Advances in Memristor-Based Neural Networks

Weilin Xu[1,2,3*†], *Jingjuan Wang*[1†] *and Xiaobing Yan*[1,4*]

[1] *Key Laboratory of Brain-Like Neuromorphic Devices and Systems of Hebei Province, College of Electron and Information Engineering, Hebei University, Baoding, China,* [2] *Guangxi Key Laboratory of Precision Navigation Technology and Application, Guilin University of Electronic Technology, Guilin, China,* [3] *Electrical and Computer Engineering Department, Southern Illinois University Carbondale, Carbondale, IL, United States,* [4] *Department of Materials Science and Engineering, National University of Singapore, Singapore, Singapore*

The rapid development of artificial intelligence (AI), big data analytics, cloud computing, and Internet of Things applications expect the emerging memristor devices and their hardware systems to solve massive data calculation with low power consumption and small chip area. This paper provides an overview of memristor device characteristics, models, synapse circuits, and neural network applications, especially for artificial neural networks and spiking neural networks. It also provides research summaries, comparisons, limitations, challenges, and future work opportunities.

Keywords: memristor, integrated circuit, artificial neural network, spiking neural network, artificial intelligence

OPEN ACCESS

Edited by:
J. Joshua Yang, University of Southern California, United States

Reviewed by:
Rivu Midya, University of Massachusetts Amherst, United States
Zhongrui Wang, The University of Hong Kong, Hong Kong

***Correspondence:**
Weilin Xu
xwl@guet.edu.cn
Xiaobing Yan
xiaobing_yan@126.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Nanodevices, a section of the journal Frontiers in Nanotechnology

Received: *24 December 2020*
Accepted: *03 March 2021*
Published: *24 March 2021*

Citation:
Xu W, Wang J and Yan X (2021) Advances in Memristor-Based Neural Networks. Front. Nanotechnol. 3:645995. doi: 10.3389/fnano.2021.645995

INTRODUCTION

Resistance, capacitance and inductance are the three basic circuit components in passive circuit theory. In 1971, Professor Leon O. Chua of the University of California at Berkeley first described a basic circuit that relates flux to charge, called the missing fourth memristor element, and was successfully found by a team led by Stanley Williams at HP Labs in 2008 (Chua, 1971; Strukov et al., 2008). As a non-linear two-terminal passive electrical component, studies have shown that the conductance of a memristor is tunable by adjusting the amplitude, direction, or duration of its terminal voltages. Memristors have shown various outstanding properties, such as good compatibility with CMOS technology, small device area for high-density on-chip integration, non-volatility, fast speed, low power dissipation, and high scalability (Lee et al., 2008; Waser et al., 2009; Akinaga and Shima, 2010; Wong et al., 2012; Yang et al., 2013; Choi et al., 2014; Sun et al., 2020; Wang et al., 2020; Zhang et al., 2020). Thus, although memristors took many years to transform from a purely theoretical derivation into a feasible implementation, these devices have been widely used in applications such as machine learning and neuromorphic computing, as well as non-volatile random-access memory (Alibart et al., 2013; Liu et al., 2013; Sarwar et al., 2013; Fackenthal et al., 2014; Prezioso et al., 2015; Midya et al., 2017; Yan et al., 2017, 2019b,d; Ambrogio et al., 2018; Krestinskaya et al., 2018; Li C. et al., 2018, Li et al., 2019; Wang et al., 2018a, 2019a,b; Upadhyay et al., 2020). Furthermore, thanks to its powerful computing and storage capability, a memristor is a promising device for processing tremendous data and increasing the data processing efficiency in neural networks for artificial intelligence (AI) applications (Jeong and Shi, 2018).

This article intends to analyze the memristor theory, models, circuits, and important applications in neural networks. The contents of this paper are organized as follows. Section Memristor Characteristics and Models introduces the memristor theory and models. Section Memristor-Based Neural Networks presents its applications in the second-generation neural networks, namely artificial neural networks (ANNs) and the third-generation neural networks, namely spiking neural networks (SNNs). Section Summary is the conclusions and future research direction.

MEMRISTOR CHARACTERISTICS AND MODELS

The relationship between the physical quantities (namely charge q, voltage v, flux φ, and current i) and basic circuit elements (namely resistor R, capacitor C, inductor L, and memristor M) is shown in **Figure 1A** (Chua, 1971). Specifically, C defined as a linear relationship between voltage v and electric charge q ($C = dq/dv$), L is defined as a relationship between magnetic flux φ and current i ($L = d\varphi/di$), R is defined as a relationship between voltage v and current i ($R = dv/di$). The missing link between the electric charge and flux is defined as the memristor M and its differential equation is $M = d\varphi/dq$ or $G = dq/d\varphi$. **Figure 1B** shows the current-voltage characteristics of the memristor, where the pinched hysteresis loop is its fundamental identifier (Yan et al., 2018c). As a basic element, the memristor I–V curve cannot be obtained using R, C, and L. According to the shape of the pinched curve, it can be roughly classified into a digital type memristor or an analog type memristor. The resistance of a digital memristor exhibits an abrupt change at higher resistance ratios. The high-resistance and low-resistance states in a digital memristor have a long retention period, making it ideal for memory and logic operations. An analog memristor exhibits a gradual change in resistance. Therefore, it is more suitable for analog circuits and hardware-based multi-state neuromorphic system applications.

Memristor device technology and modeling research are the cornerstones of system applications. As shown in **Figure 2**, top-level system applications (brain-machine interface, face or picture recognition, autonomous driving, IoT edge computing, big data analytics, and cloud computing) are built on the device technology and modeling. Memristor-based analog, digital, and memory circuits play a key role in the link between device materials and system applications. The main usage for bi-stable memristors is binary switches, binary memory, and digital logic circuits, while multi-state memristors are used as multi-bit memories, reconfigurable analog circuits, and neuromorphic circuits.

Since the HP labs verified the nanoscale physical implementation, the physical behavior models of memristors have received a lot of attention. Accuracy, convergence, and

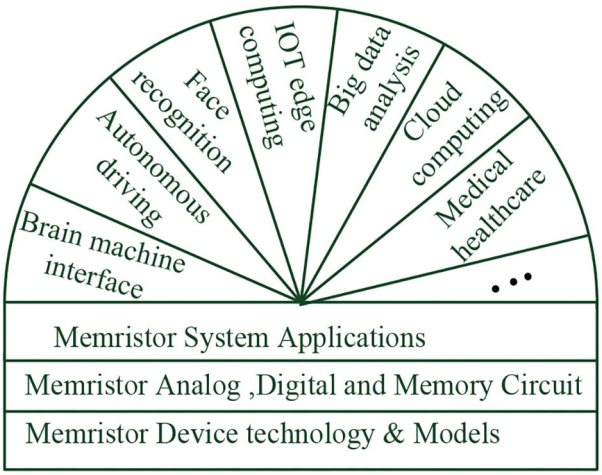


FIGURE 2 | Memristor research and applications.

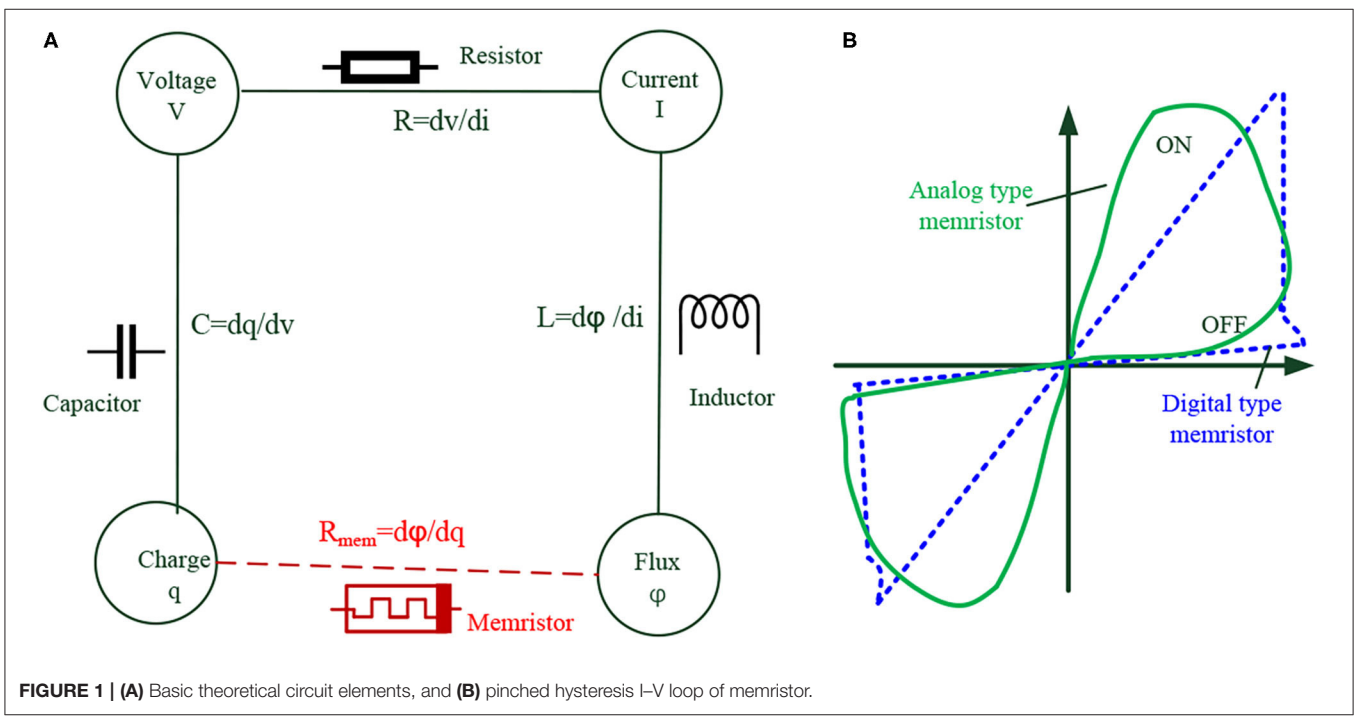


FIGURE 1 | (A) Basic theoretical circuit elements, and **(B)** pinched hysteresis I–V loop of memristor.

TABLE 1 | Classic memristor models.

Models	Linear ion drift	Non-linear ion drift	Simmons tunnel barrier	TEAM
I–V characteristic	$v(t) = \left(Ron\frac{w(t)}{D} + Roff\left(1 - \frac{w(t)}{D}\right)\right) i(t)$	$i(t) = w(t)^n \beta \sinh(\alpha v(t)) + \chi[\exp(\gamma v(t)) - 1]$	$v(t) = \left(Ron + \frac{Roff - Ron}{Woff - Won}(w - won)\right) i(t)$	$v(t) = R_{on} \cdot \frac{\lambda}{e^{woff - won}}(w - won)$
State variable $\frac{dw(t)}{dt}$	$\left(u_v \frac{R_{on}}{D}\right) i(t)$	$a \cdot f(w) \boldsymbol{v(t)}^m$	$\begin{cases} coff - \sinh\left(\frac{i}{ioff}\right) \exp\left[-\exp\left(\frac{w - aoff}{wc} - \frac{\lvert i\rvert}{b}\right) - \frac{w}{wc}\right], & i > 0 \\ coff - \sinh\left(\frac{i}{ioff}\right) \exp\left[-\exp\left(\frac{w - aoff}{wc} - \frac{\lvert i\rvert}{b}\right) - \frac{w}{wc}\right], & i < 0 \end{cases}$	$\begin{cases} koff\left(\frac{i(t)}{ioff} - 1\right)^{aoff}, & 0 < ioff < i \\ kon\left(\frac{i(t)}{ion} - 1\right)^{aon}, & 0 < ioff < i \\ 0, & otherwise \end{cases}$
Interval	$0 \le w \le D$	$0 \le w \le 1$	$aoff \le w \le aon$	$aon \le w \le aoff$
Control mechanism	Current controlled	Voltage-controlled	Current-controlled	Current-controlled
Accuracy	Lowest accuracy	Low accuracy	Highest accuracy	Sufficient accuracy
Thershold exists	No	No	Pracitcally exists	Yes
Linearity	linear	No-linear	No-linear	No-linear

computational efficiency are the most important factors in memristor models. These behavior models are expected to be simple, intuitive, better understood, and closed form. Up to date, various models have been developed, each with its unique advantages and shortcomings. The models listed in **Table 1** are the most popular models, including a linear ion drift memristor model, a non-linear ion drift memristor model, a Simmons tunnel barrier memristor model, a threshold adaptive memristor model (TEAM) (Simmons, 1963; Strukov et al., 2008; Biolek et al., 2009; Pickett et al., 2009; Kvatinsky et al., 2012). In the linear ion drift memristor model, D and u_v represent the full length and device mobility of a memristor film, respectively. ω(t) is a dynamic state variable whose value is limited between 0 and D, taking into account the size of the physical device. The low turn-on resistance R_{on} is the full doped resistance when dynamic variable ω(t) is equal to D. The high turn-off resistance R_{off} is a fully undoped resistance when ω(t) is equal to 0. Besides, a window function multiplied by a state variable is needed to nullify the derivative and provide a non-linear transition for the physical boundary simulation. Several window functions have been presented for modeling memristors such as Biolek, Strukov, Joglekar, and Prodromakis window functions (Strukov et al., 2008; Biolek et al., 2009; Joglekar and Wolf, 2009; Strukov and Williams, 2009; Prodromakis et al., 2011). As the first memristor model, the linear ion drift model shows the features of simple, intuitive, and better understood. However, the state variable ω modulation in nano-scale devices is not a linear process, and the memristor experimental results show non-linear I–V characteristics. The non-linear ion drift model provides a better description of non-linear ionic transport and higher accuracy by experimentally fitting the parameters n, β, α, and χ (Biolek et al., 2009). But more physical reaction kinetics still need to be considered. The Simmons tunnel barrier model consists of a resistor in series with an electron tunnel barrier, which provides a more detailed representation of non-linear and asymmetrical features (Simmons, 1963; Pickett et al., 2009). There are nine fitting parameters in this segmentation model, which makes the mathematical model very complicated and computationally inefficient. The TEAM model can be thought of as a simplified version of the Simmons tunnel barrier model (Kvatinsky et al., 2012). However, all of the above models suffer from smoothing problems or mathematical ill-posedness issues, and they cannot provide robust and predictable simulation results in DC, AC, transient analysis, not to mention complicated circuit analysis such as noise analysis and periodic steady-state analysis (Wang and Roychowdhury, 2016). Therefore, in the face of transistor-level circuit design simulation, circuit designers usually have to replace the actual memristor with an emulator (Yang et al., 2019). The emulator is a complex CMOS circuit used to simulate some performance aspect of a special memristor. An emulator is not a true model, and it is very different from the real memristor model (Yang et al., 2014). Thus, it is urgent to establish a complete memristor model. Correct bias definition and right physical characteristics in SPICE or Verilog-a model are important for complex memristor circuit design. Otherwise, non-physical predictions will confuse circuit engineers in physical chip design.

MEMRISTOR-BASED NEURAL NETWORKS

Neuron Biological Mechanisms and Memristive Synapse

The human brain can solve complex tasks, such as image recognition and data classification, more efficiently than traditional computers. The reason why a brain excels in complicated functions is the large number of neurons and synapses that process information in parallel. As shown in **Figure 3**, when an electrical signal is transmitted between two neurons via axon and synapse, the joint strength or weight is adjusted by the synapse. There are approximately 100 billion neurons in an entire human brain, each with about 10,000 synapses. Pre-synaptic and post-synaptic neurons transfer and receive the signal of excitatory and inhibitory

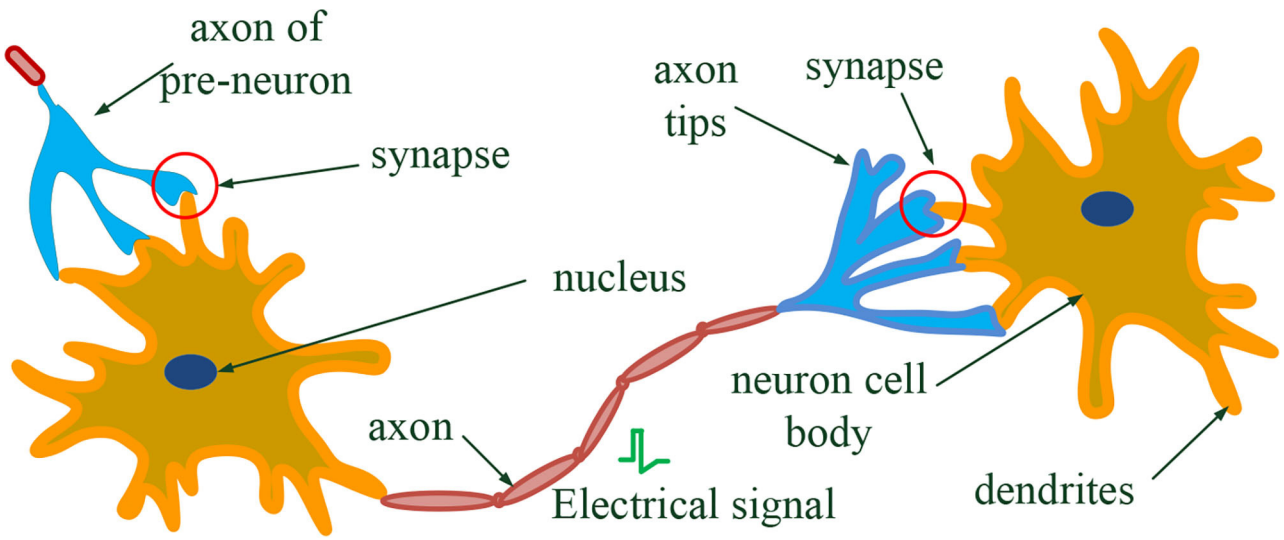


FIGURE 3 | Schematic of two interconnected neurons by synapses.

post-synaptic potentials by updating synaptic weights. Long-term potentiation (LTP) and long-term depression (LTD) are important mechanisms in a biological nervous system, which indicates a deep-rooted transformation in the connection strengths between neurons. According to the interval between pre-synaptic and post-synaptic action potentials or spikes, the phenomenon of synaptic weight modification is known as spike-timing-dependent plasticity (STDP) (Yan et al., 2018a, 2019c). Due to scalability, low power operation, non-volatile features, and small on-chip area, memristors are good candidates for artificial synaptic devices to mimicking the LTP, LTD, and STDP behaviors (Jo et al., 2010; Ohno et al., 2011; Kim et al., 2015; Wang et al., 2017; Yan et al., 2017).

There are some key requirements for memristive devices in neural network applications. For example, a wide range of resistance is required to enable sufficient resistance states; devices are required to have low resistance fluctuations and low device-to-device variability; a higher absolute resistance is required for low power dissipation; and high durability is required for reprogramming and training (Choi et al., 2018; Yan et al., 2018b, 2019a; Xia and Yang, 2019). A concern with device stability is resistance drift, which occurs over time or with the environment. Resistance drift causes undesirable changes in synapse weight and blurs different resistance states, ultimately affecting the accuracy of neural network computation (Xia and Yang, 2019). To deal with this drift challenge, improvements can be made in three aspects: (1) material device engineering, (2) circuit design, and (3) system design (Alibart et al., 2012; Choi et al., 2018; Jiang et al., 2018; Lastras-Montaño and Cheng, 2018; Yan et al., 2018b, 2019a; Zhao et al., 2020). For example, as for the domain of material engineering, threading dislocations can be used to control programming variation and enhance switching uniformity (Choi et al., 2018). In terms of circuit-level design, a module of two series memristors and a transistor with the smallest size can be used, thus, the resistance ratio of the memristor can be encoded to compensate for the resistance drift (Lastras-Montaño and Cheng, 2018). For the system-design level, device deviation can be reduced by protocols, such as closed loop peripheral circuit with a write-verify function (Alibart et al., 2012). In order to obtain linear and symmetric weight update in LTP and LTD for efficient neural network training, optimized programming pulses can be used to excite memristors with either fixed-amplitude or fixed-width voltage pulses (Jiang et al., 2018; Zhao et al., 2020). Note it is inevitable to increase energy consumption if the memristor resistance value is changed through complex programmable pulses.

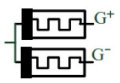
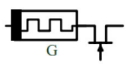
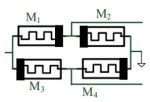
The comparison of different memristive synapse circuit structures is shown in **Table 2** (Kim et al., 2011a; Wang et al., 2014; Prezioso et al., 2015; Hong et al., 2019; Krestinskaya et al., 2019). Single memristor synapse (1M) crossbar arrays in neural networks have the lowest complexity and low power dissipation. However, it suffers from sneak path problems and complex peripheral switch circuits. Synapses with two memristors (2M) have a more flexible weight range and better symmetric LTP and LTD, but the corresponding chip area will be doubled. A synapse with one memristor and one transistor (1M-1T) has the advantage of solving the sneak path problem, but it also occupies a large area in the large-scale integration of neural networks. A bridge synapse architecture with four memristors (4M) provides a bidirectional programming mechanism with a voltage input voltage output. Due to the significant on-chip area overhead, the 1M-1T and 4M synapses may not be applicable for large-scale neural networks.

Memristor-Based ANNs

The basic operations of classical hardware ANNs include multiplication, addition, and activation, which are accomplished by CMOS circuits such as GPUs. The weights are typically saved

in SRAM or DRAM. Despite the scalability of CMOS circuits, they are still not enough for ANN applications. Furthermore, the SRAM cell size are too big to be integrated at high density. DRAM needs to be refreshed periodically to prevent data decay. Whether it is SRAM or DRAM, it often needs to interact with CMOS cores. No matter SRAM or DRAM, the data needs to be fetched by to the cache and register files of the digital processors before processing and returned through the same databus, leading to significant speed limit and large energy consumption, which is the main challenge for deep learning and big data applications (Xia and Yang, 2019). Nowadays, ANNs feature for large number of computational parameters stored in memory compared to classical computation. For example, a two-layer 784-800-10 fully-connected deep neural network in the MNIST dataset has 635,200 interconnections. A state of the art keep neural network like Visual Geometry Group (VGG) has a few millions of parameters. These factors pose a huge challenge to the implementation of ANN hardware. The memristor's non-volatility, lower power consumption, lower parasitic capacitance, and reconfigureable resistance states, high speed, and adaptability lead to a key role in ANN applications (Xia and Yang, 2019). An ANN is an information processing model which are derived from mathematical optimization. A typical ANN architecture and its memristor crossbar are shown in **Figure 4**. The system usually consists of three layers: an input layer, a middle layer or a hidden layer, and an output layer. The connected units or nodes are neurons which are usually series by weighted-sum module and activation function module. Neurons also perform tasks of decoding, control, and signal routing. Due to its powerful signal processing capability, CMOS analog and digital logic circuits are the best candidates for neurons hardware implementation. In **Figure 4**, arrow or connecting lines represent synapses, and their weights represent the connection strengths between two neurons. Assume the weight modulation matrix W_{ij} in a memristor synapse crossbar is a M × N dimensinal matrix, where i(i = 1, 2, …, N) and j(i = 1, 2, …, M) are the index numbers of the output and input ports of the memristor crossbar. W_{ij} between pre-neuron input vector X_j and post-neuron output vector Y_i is

TABLE 2 | Comparison of different structure memristive synapse circuit.

Synapses	Structure	Area(F2)	Weight	Weight range	Other features
1M		≈4	G	+	Lower power consumption; least complex; sneak path problem in neural network array
2M		≈8	G^+-G^-	+,0, –	Better symmetric between LTP and LTD; complex post-synaptic neurons
1M-1T		≈24	G	+	Solution for sneak path problem with transistor switch; biggest size; transistor non-ideal effect
4M		≈16	$\frac{M_2}{M_1+M_2} - \frac{M_4}{M_3+M_4}$	+,0, –	Voltage input voltage output; Bidirectional programming; bigger size

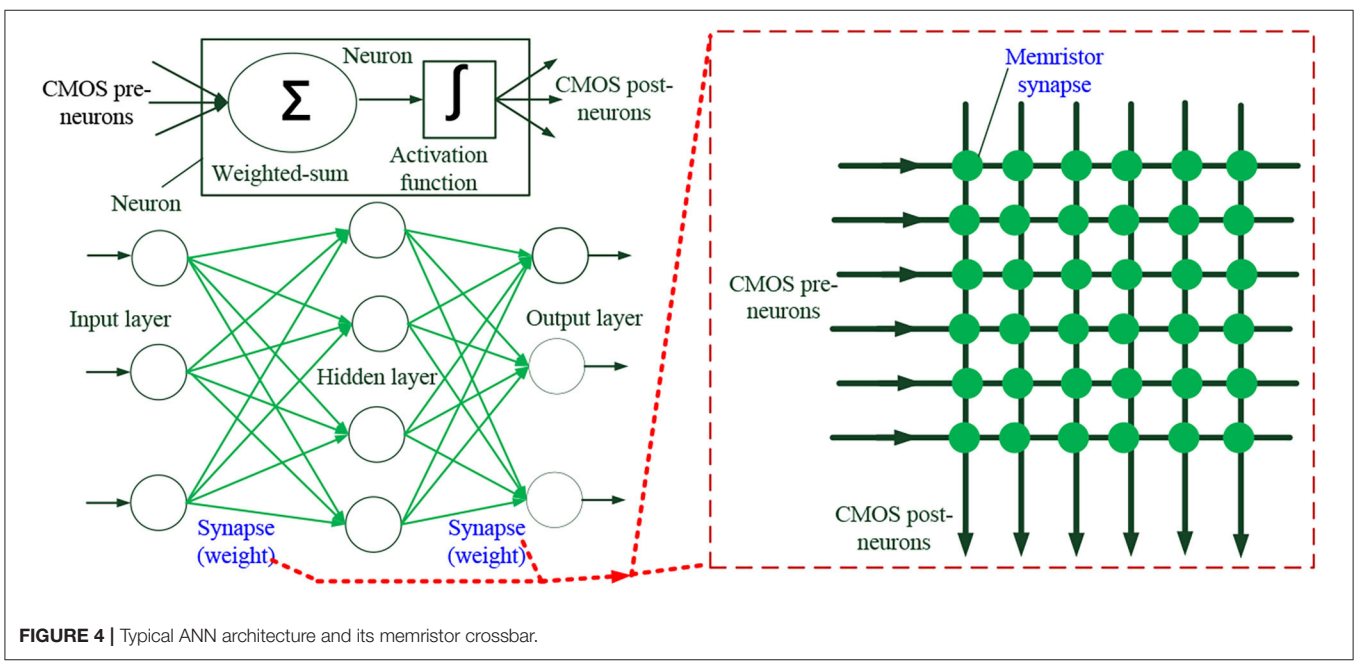


FIGURE 4 | Typical ANN architecture and its memristor crossbar.

TABLE 3 | Typical architectures of Memristive ANNs.

TYPES	Architecture	Layers properties	Applications	Challenges
SLP/MLP	Input layer+hidden layer+output layer	Sigmoid, tanh, etc., activation; Full-connections	Simple pattern classification; Hand-written letter recognition	Power dissipation in deep ANN; Overfitting; non-ideal memristor; Scalability
CNN	Input layer+ Convolution layer+ ReLu layer+Pooling +Fully-connected and output layer	Convolution; Pooling	Image classification; Face recognition; Video analysis	
CeNN	Cell array with templates; 1-D, 2-D, or 3-D	Dissipative non-linear cells; Lyapunov function; Neighborhood communication	Image filtering; Signal processing; moving object detection	Convergence and mulitistability in non-symmetric networks; non-ideal memristor
RNN	Fully recurrent; Elman; Jordan; gated recurrent unit; long short-term memory	Temporal dynamic behavior; directed graph along a temporal sequence; LSTM	Speech recognition; Machine translation; Video processing	Hard to train for long term dependencies; non-ideal memristor

TABLE 4 | ANNs learning accuracy improvement by mitigating memristor non-ideal effects.

Level	Strategies	Tradeoffs
Device materials	Optimizing redox reaction at the metal/oxide interface (Lee et al., 2015), Threading dislocations technology (Tanikawa et al., 2018), Heating element, selectively enhanced filament expansion stage (Jeong et al., 2015)	Manufacturing cost; Power consumption; On-chip area; Peripheral circuit complexity; Algorithm efficiency
Circuits	Hybrid CMOS-memristor Neuromorphic Synapse, 1R+1M1R for better device symmetry (Woo and Yu, 2018), Non-identical pulse excitation (Park et al., 2013; Chang et al., 2017), Bipolar-pulse-training (Li et al., 2016), Spike edge shape design (Li S. J. et al., 2018)	
Architectures	Multiple memristors cell for high redundancy (Chen et al., 2015), Pseudo-crossbar array, peripheral circuit compensation (Chen et al., 2015)	
Algorithms	co-optimization between memristors and ANN algorithms (Li et al., 2016)	

a matrix-vector multiplication operation, expressed as Equation (1) (Jeong and Shi, 2018).

$$Y_i = \Sigma W_{ij} \cdot X_j \tag{1}$$

$$\Delta w_{ij} = r\left[\frac{\partial\left(y - y^*\right)^2}{\partial w_{ij}}\right] \tag{2}$$

The matrix W can be continuously adjusted until the difference between the output value y and the target value y^* is minimized. The Equation (2) shows the synaptic weight tunning process with the gradient of output error $(y–y^*)^2$ under a training rate (Huang et al., 2018). Therefore, a memristor crossbar is equal to a CMOS adder plus a CMOS multiplier and an SRAM (Jeong and Shi, 2018), because data are computed, stored, and regenerated on the same local device (i.e., a memristor itself). Besides, a crossbar can be vertically integrated into three dimensions (Seok et al., 2014; Lin et al., 2020; Luo et al., 2020). In this way, it saves much chip area and power consumption. Due to the memristor synapse update and save weight data on itself, the memory wall problem with von Neumann bottleneck is solved.

Researchers have developed various topologies and learning algorithms for software-based or hardware-based ANNs. **Table 3** provides a comparison of typical memristive ANNs, including single-layer perceptron (SLP) or multi-layer perceptron (MLP), CNN, cellular neural network (CeNN), and recurrent neural network (RNN). SLP and MLP are classic neural networks with well-known learning rules such as Hebbian learning, backpropagation. Although a lot of ANN studies have been verified by simulations or small-scale implementation, a single-layer neural network with 128 × 64 1M-1T Ta/HfO_2 memristor array has been experimentally demonstrated with an image recognition accuracy of 89.9% for the MNIST dataset (Hu et al., 2018). CNNs (referred to as space-invariant or shift-invariant ANNs) are regularized versions of MLP. Their hidden layers usually contain multiple complex activation functions, and perform convolution or regional maximum value operations. Researchers have demonstrated an over 70% of accuracy in human behavior video recognition with a memristor-based 3D CNN (Liu et al., 2020). It should be emphasized that this verification is only a software simulation result, while the on-chip hardware demonstration is still very challenging, especially for deep CNNs (Wang et al., 2019a; Luo et al., 2020; Yao et al., 2020). CeNN is a massively parallel computing neural network, whose communication features are limited to between adjacent cell neurons. The cells are dissipative non-linear continuous-time or discrete-time processing units. Due to their dynamic processing capability and flexibility, CeNNs are promising candidates for real-time high frame rate processing or multi-target motion detection. For example, a CeNN with 4M memristive bridge circuit synapse has been proposed for image processing (Duan et al., 2014). Unlike classic feed forward ANNs, RNNs have a feedback connection that enables temporal dynamic behavior.

Therefore, it is suitable for speech recognition applications. Long short-term memory (LSTM) is a kind of useful RNN structure for deep learning. Hardware implementation of LSTM networks based on memristors have been reported (Smagulova et al., 2018; Li et al., 2019; Tsai et al., 2019; Wang et al., 2019a).

Due to atomic-level random defects and variability in the conductance modulation process, non-ideal memristor characteristics are the main causes of learning accuracy loss in ANNs. This phenomenon is manifested in the following aspects of memristor: asymmetric non-linear weight change between potentiation and depression, limited ON/OFF weight ratio and device variation. **Table 4** shows the main strategies for how to deal with these issues. One can mitigate the effects of non-ideal memristor characteristics on ANN accuracy from four levels: device materials, circuits, architectures, and algorithms. At device materials level, switching uniformity and analog on/off ratio can be enhanced by optimizing redox reaction at the metal/oxide interface, adopting threading dislocations technology or heating element (Jeong et al., 2015; Lee et al., 2015; Tanikawa et al., 2018). At circuits level, one can use customized excitation pulse or hybrid CMOS-memristor synapses to mitigate memristor non-ideal effects (Park et al., 2013; Li et al., 2016; Chang et al., 2017; Li S. J. et al., 2018; Woo and Yu, 2018). At architectures level, common techniques are multiple memristors cell for high redundancy, pseudo-crossbar array, and peripheral circuit compensation (Chen et al., 2015). Co-optimization between memristors and ANN algorithms is also reported (Li et al., 2016). However, it should be noted that implementation of these strategies inevitably brings side effects, such as high manufacturing cost, large power consumption, large chip area, complex peripheral circuits, or inefficient algorithm. For example, the non-identical pulse excitation or bipolar-pulse-training methods improve the linearity and symmetry of memristor synapses, but it increases the complexity of peripheral circuits, system power consumption, and chip area. Therefore, trade-offs and co-optimization need to be made at each design level to improve the learning accuracy of ANNs (Gi et al., 2018; Fu et al., 2019). **Figure 5** is a collaborative design example from bottom-level memristor devices to top-level training algorithms (Fu et al., 2019). The conductance response (CR) curve of memristors is first measured to obtain its non-linearity factor. Then, the CR curve is divided into piecewise linear segments to obtain their slope, and the pulse width of the excitation pulse is inversely proportional to the slope. These data are stored in memory for comparison and correction by memristor crossbars during the update. Through this method, the ANN recognition accuracy is finally improved.

The memristor-based ANN applications can be software, hardware or hybrid (Kozhevnikov and Krasilich, 2016). Software networks tend to be more accurate than their hardware counterparts because they do not have the analog element non-uniformity issues. However, hardware networks feature better speed and less power consumption due to non-von Neumann architectures (Kozhevnikov and Krasilich, 2016). In **Figure 6**, a deep neuromorphic accelerator ANN chip with 2.4 million Al_2O_3/TiO_2-xmemristors was designed and fabricated (Kataeva et al., 2019). This memristor chip consists of a 24 × 43 array with a 48 × 48 memristor crossbar at each intersection, which means its complexity is about 1,000 times higher than previous designs in the literature. This work is a good starting point for the operation of medium-scale memristor ANNs. Similar accelerators have appeared in the last 2 years (Cai et al., 2019; Chen W.-H. et al., 2019; Xue et al., 2020).

Memristive neural networks can be used to understand human emotion and simulate human operational abilities (Bishop, 1995). The well-known PavlTov associative memory experiment has been implemented in memristive ANNs with a novel weighted-input-feedback learning method (Ma et al., 2018). As more input signals, neurons, and memristor synapses, complex emotional processing will be achieved in further AI chips. Due to the material challenge and the lack of effective models, most of the demonstrations are limited to small-scale simulations for simple tasks. The shortcomings of memristors are mainly the non-linearity, asymmetry, and variability, which seriously affect the accuracy of ANNs. Moreover, the peripheral circuits and interface must provide superior energy efficiency and data throughput.

Memristor-Based SNN

Inspired by cognitive and computational methods of animal brains, the third-generation neural network, SNN, makes desirable properties of compact biological neurons mimic and remarkable cognitive performance. The most prominent feature of SNN is that it incorporates the concept of time into operations

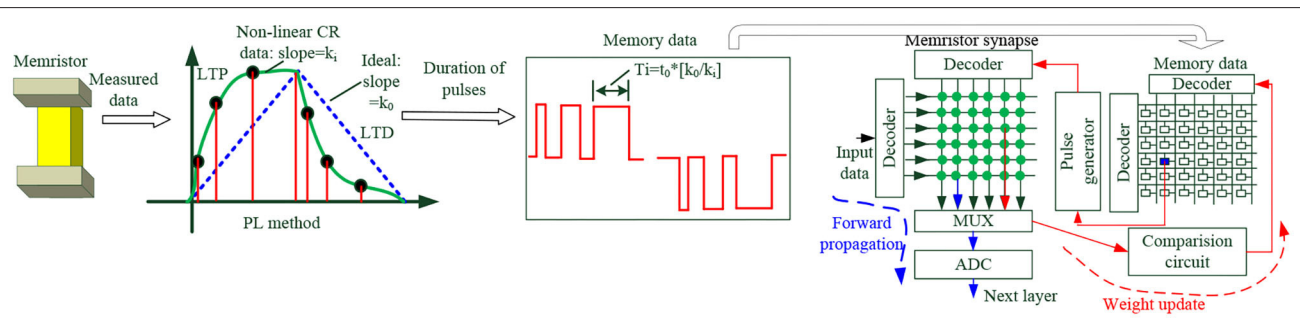


FIGURE 5 | Co-design from memristor non-ideal characteristics to the ANN algorithm (Fu et al., 2019).

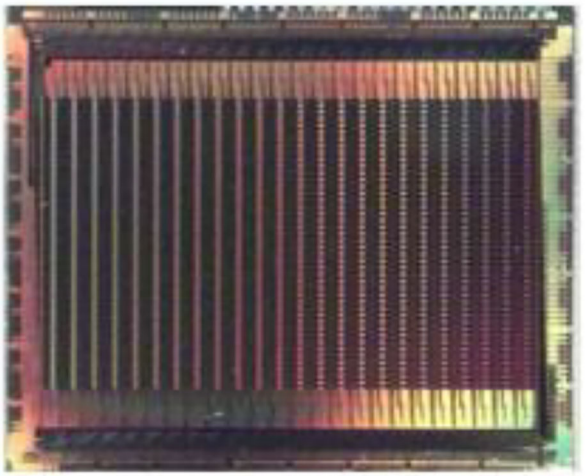


FIGURE 6 | A deep neuromorphic ANN chip with 2.4 million memristor devices (Kataeva et al., 2019).

TABLE 5 | Comparison of several memristor-based SNNs.

References	Neuron	Synapse	Learning rules	Size	Applications
Zheng and Mazumder (2018)	LIF	1M1R fixed-polarity memristor	STDP; Supervised learning	784-300-10	Handwritten digits recognition
Chen B. et al. (2019)	LIF	Lithium silicate memristor	STDP, Unsupervised learning, WTA	128-128-12	Motion-style recognition
Shukla and Ganguly (2018)	LIF	HfO_2 memristor	STDP; Supervised Hebbian	16-3	Classification problems, Fisher Iris dataset, etc.
Wu and Saxena (2018)	LIF	Stochastic binary memristor	STDP, Dendritic-inspired processing	1-4	Pattern Recognition
Chu et al. (2014)	LIF	$Pr_{0.7}Ca_{0.3}MnO_3$-memristor	STDP, Unsupervised learning	30-10	Visual Pattern Recognition
Volos et al. (2015)	H-R, FHN	Flux-controlled memristor	STDP	2	Chaotic oscillators; Neurodynamic behavior
Al-Shedivat et al. (2015)	SRM	Stochastic biolek's memristor model	STDP, WTA	1568-32	Handwritten digits recognition

with discrete values, while the input and output values of the second-generation ANNs are continuous. SNN can better leverage the strength of biological paradigm of information processing, thanks to the hardware emulation of synapses and neurons. ANN is calculated layer by layer, which is relatively simple. However, spike trains in SNN are relatively difficult to understand and efficient coding methods for these spike trains are not easy. These dynamic events driven spikes in SNN enhance the ability to process spatio-temporal or real-world sensory data, with fast adaptation and exponential memorization. The combination of spatio-temporal data allows SNN to process signals naturally and efficiently.

Neuron models, learning rules, and external stimulus coding are key research areas of SNN. The Hodgkin & Huxley (HH) model, leaky Integrate-and-Fire (LIF) model, spike response model (SRM), and Izhikevich model are the most common models of neurons (Hodgkin and Huxley, 1952; Chua, 2013; Ahmed et al., 2014; Pfeiffer and Pfeil, 2018; Wang and Yan, 2019; Zhao et al., 2019; Ojiugwo et al., 2020). The HH model is a continuous-time mathematical model based on conductance. Although this model is based on the study of squid, it is widely used in lower or higher organisms (even humans being). However, since complex non-linear differential equations are set with four variables, this model is difficult to achieve high accuracy. Chua established the memristor model of Hodgkin-Huxley neurons and proved that memristors can be applied to the imitation of complex neurobiology (Chua, 2013). The Izhikevich model integrates the bio-plasticity of HH model with simplicity and higher computational efficiency. The HH and Izhikevich models are calculated by differential equations, while the LIF and SRM models are computed by an integral method. SRM is an extended version of LIF, and the Izhikevich model can be considered as a simplified version of the Hodgkin-Huxley model. These mathematical models are the results of different degrees of customization, trade-offs and biological neural network optimization. **Table 5** shows a comparison of several memristor-based SNNs. It can be seen that these SNN studies are based on STDP learning rules and LIF neurons. Most of them are still in simple pattern recognition applications, only a few of which have hardware implementations.

The salient features of SNNs are as follows. First, biological neuron models (e.g., HH, LIF) are closer to biological neurons than neurons of ANN. Second, the transmitted information is time or frequency encoded discrete-time spikes, which can contain more information than traditional networks. Third, each neuron can work alone and enter a low power standby mode when there is no input signal. Since SNNs have been proven to be more powerful than ANNs in theory, it is natural to widely use SNNs. Since the spike training cannot be differentiated, the gradient descent method cannot be used to train SNNs without losing accurate temporal information. Another problem is that it takes a lot of computation to simulate SNNs on normal hardware, because it requires analog differential equations (Ojiugwo et al., 2020). Due to the complexity of SNNs, efficient learning rules that meet the characteristics of biological neural networks have not been discovered. This rule is required to model not only synaptic connectivity but also its growth and attenuation. Another challenge is the discontinuous nature of spike sequence, which makes many classic ANN learning rules unsuitable for SNNs, or can only be approximated, because the convergence problem is very serious. Meanwhile, many SNNs

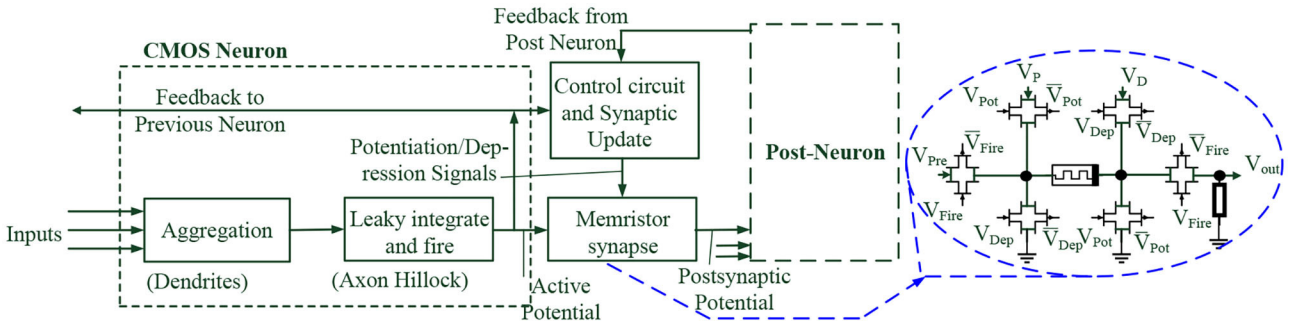


FIGURE 7 | CMOS neuron and memristor synapse weight update circuit (Sun, 2015).

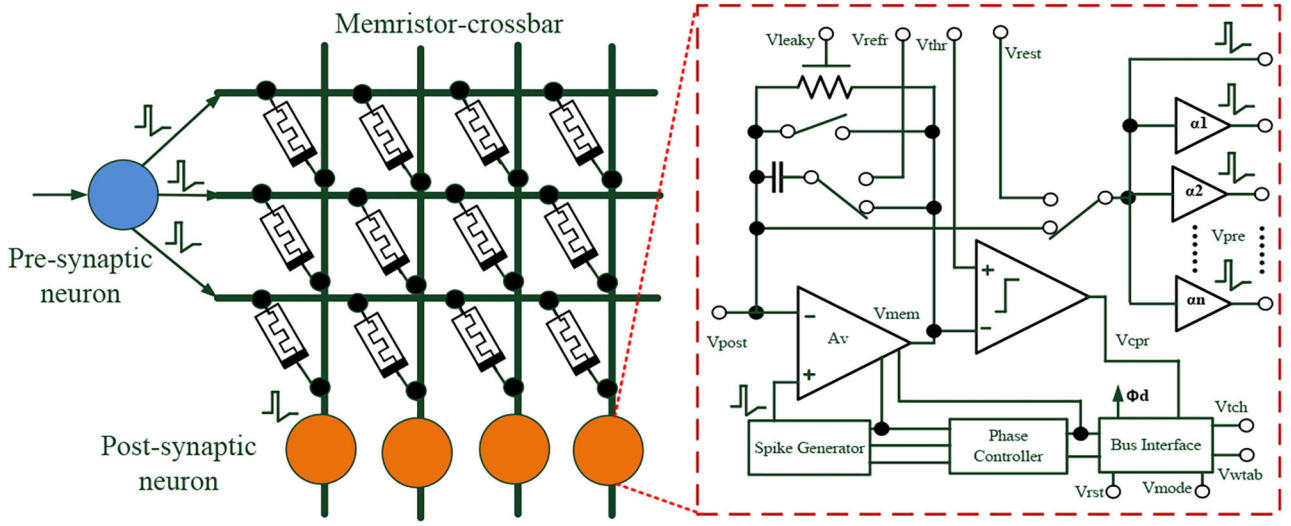


FIGURE 8 | CMOS-Memristor SNN (Wu and Saxena, 2018).

studies are limited to theoretical analysis and simulation of simple tasks rather than complex and intelligent tasks (e.g., multiple regression analysis, deductive and inductive reasoning, and their chip implementation) (Wang and Yan, 2019). Although the future of SNNs is still unclear, many researchers believe that SNNs will replace deep ANNs. The reason is that AI is essentially a biological brain mimicking process, and SNNs can provide a perfect mechanism for unsupervised learning.

As shown in **Figure 7**, a neural network is implemented with CMOS neurons, CMOS control circuits, and memristor synapses (Sun, 2015). The aggregation module, leaky integrate and fire module are equivalent to the role of dendrites and axon hillocks, respectively. Input neurons signals are temporally and spatially summed through a common-drain aggregation amplifier circuit. A memristor synapse gives the action potential signal a weight and its output signal, that is, a post-synaptic potential signal is transmitted to post-neurons. Using the action potential signal and feedback signals from post-neurons, the control circuit and synaptic update phase provide potentiation or depression signals to memristor synapses. According to the STDP learning rules, the transistor-level weight adjustment circuit is composed of a memristor device and CMOS transmission gates. The transmission gates are controlled by potentiation or depression signals. The system is very similar to the main features of biological neurons, which is useful for neuromorphic SNN hardware implementation. A more complete description of SNN circuits and system applications is shown in **Figure 8** (Wu and Saxena, 2018). The system consists of event-driven CMOS neurons, a competitive neural coding algorithm [i.e., winner take all (WTA) learning rule], and multi-bit memristor synapse array. A stochastic non-linear STDP learning rule with an exponential shaped window learning function is adopted to update memristor synapse weights *in situ*. The amplitude and additional temporal delay of the half rectangular half-triangular spike waveform can be adjusted for dendritic-inspired processing. This work demonstrates the feasibility and excellence of emerging memristor devices in neuromorphic applications, with low power consumption and compact on-chip area.

Despite the large on-chip area and power dissipation in CMOS implementation of synaptic circuits (Chicca et al., 2003; Seo et al.,

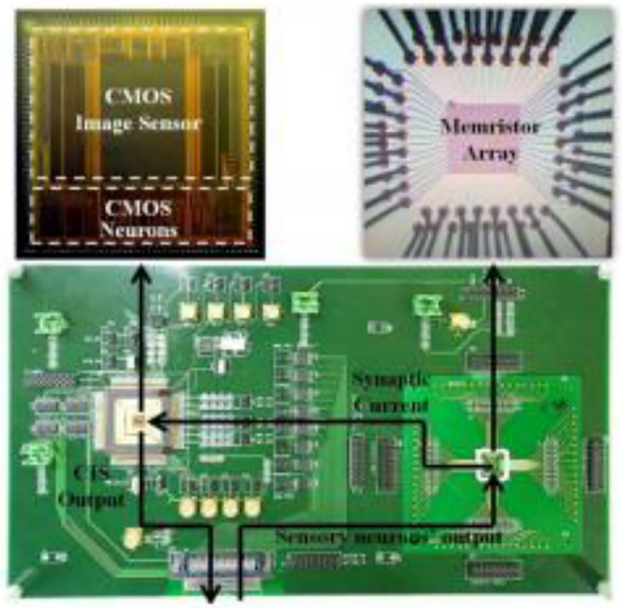


FIGURE 9 | A memristor synapse array micrograph for SNN Application (Chu et al., 2014).

TABLE 6 | Comparison between ANNs and SNNs.

	ANNs	SNNs
Generation	Second-generation NN	Third-generation NN
Biological brain mimicking	General	Better
Signal processing	Continuous multi-level value	Sparse and asynchronous binary time-domain coded spike signals. Event-driven discrete information processing
Energy efficiency	General	Better
Neurons and Synapses	Activation functions;	Hodgkin and Huxley, LIF, etc.
	Digital or analog memristor synapses	Analog memristor synapses
Classical algorithms	Error-backpropagation	SpikeProp, STDP
Chip design	In progress with some achievement.	Preliminary stage
	Near-term application goals	Long-term application goals

2011), Myonglae Chu adopted $Pr_{0.7}Ca_{0.3}MnO_3$-based memristor synaptic array and CMOS leaky IAF neurons in SNN. As shown in **Figure 9**, the SNN chip has been successfully developed for visual pattern recognition with modified STDP learning rules. The SNN hardware system includes 30 × 10 neurons and 300 memristor synapses. Although this hardware system only recognizes numbers 0–9, it is a good attempt, as most SNN studies have lingered around the software simulation phase (Kim et al., 2011b; Adhikari et al., 2012; Cantley et al., 2012). One can refer to literatures (Wang et al., 2018b; Ishii et al., 2019; Midya et al., 2019b) for more experimental memristor-SNN demos.

Comparison Between ANNs and SNNs

A comparison between ANNs and SNNs is shown in **Table 6** (Nenadic and Ghosh, 2001; Chaturvedi and Khurshid, 2011; Zhang et al., 2020). Traditional ANNs require layer-by-layer computation. Therefore, it is computationally intensive and has a relatively large power consumption. An SNN changes from a standby mode to a working mode, when a large nerve spike is coming with its spike threshold exceeding the membrane voltage. As a result, its system power consumption is relatively low.

SNNs with higher bio-similarity are expected to achieve higher energy efficiency than ANNs. But SNN hardware is harder to implement than ANN hardware. Thus, combining the advantages of ANN and SNN and using ANN-SNN converters to improve SNN performance is a valuable method, which has been experimentally demonstrated (Midya et al., 2019a). The first and second layers of a converter are ordinary ANN structures. The output signals of the second layer are converted to a spike sequence for a 32 × 1 1M-1T drift memristor synapse array at the third layer. This ANN-SNN converter may be a good way for SNN hardware implementation. Despite the enormous potential of SNNs, there is currently no fully satisfactory general learning rules and its computational capability has not been demonstrated. Most of these methods lack comparability and generality. Compared to ANNs, the study of dynamic devices and efficient algorithms in SNNs is very challenging. SNNs only need to compute the activated connections, rather than all connections at every time step in ANNs. However, the encoding and decoding of spikes is one of the challenges in SNN research. In fact, it needs further research in neuroscience. ANN is the recent target of memristors, and SNN is the long-term goal in the future.

For neural networks applications, ANN and SNN memristor grids have some common challenges, such as sneak path problems, IR-drop or ohmic drop, grid latency, and grid power dissipation, as shown as **Figure 10** (Zidan et al., 2013; Hu et al., 2014, 2018; Zhang et al., 2017). The large the size of the memristor array, the greater the effect of these parasitic capacitances and resistances. In Figure, the desired weight-update path is the dot-and-dash line, and the sneak path is the dotted line, which is an undesired parallel memristor path due to its relative resistance and non-gated memristor elements. This phenomenon leads to undesired weight changes and a reduction in the accuracy of neural networks. The basic solution for the sneak path is to add a series of connected gate-controlled MOS transistors to memristors as mentioned in **Table 2**. However, this method will lead to large on-chip synapse array and destroy the advantages of high-density integration of memristors. Grounding an unselected memristor array is another solution without the need to add synaptic area. But this approach leads to more power consumption. There are other techniques such as grounding line, floating line, additional bias,

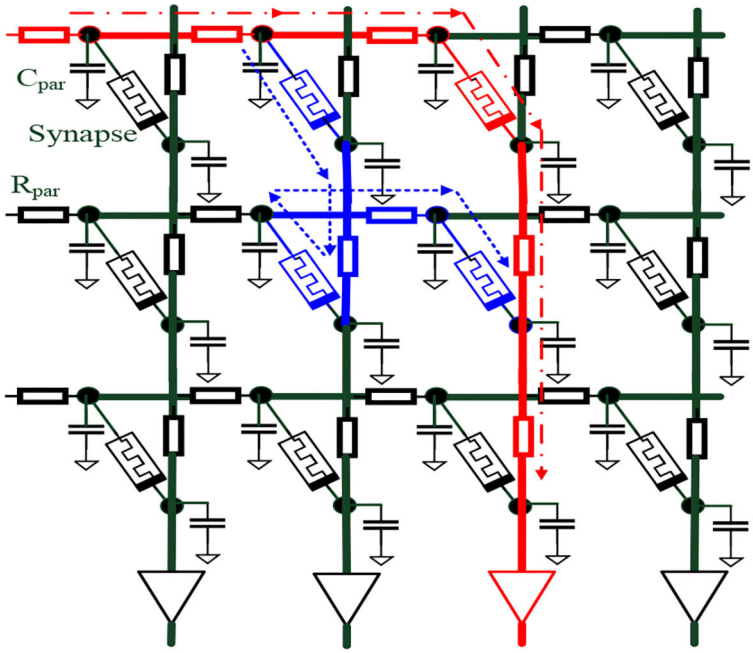


FIGURE 10 | Sneak path, IR-drop, latency, and energy in massive memristor grids of neural networks.

a non-unity aspect ratio of memristor arrays, three-electrode memristor devices. They may be welcome in memristor memory applications, but not necessarily in memristor-based neural network applications (Zidan et al., 2013). In neural network applications, the main concern for memristor arrays is whether the association between input and output signals is correct (Hu et al., 2014). This is one important difference compared to memristor memory applications. IR-drop, memristor grid latency, and power consumption are signal integrity effects caused by grid parasitic resistance Rpar and parasitic capacitance Cpar. These non-ideal factors affect the potential distribution, signal transmission, and ultimately affect the scale of memristor arrays. Similar to CMOS layout and routing techniques, large-scale memristors mesh can be divided into medium-sized modules with high-speed main signal paths for lower parasitic resistance, grid power consumption, and latency. It is worth noting that memristor process variations, gird IR-drop and noise can worsen the sneak path problem.

SUMMARY

The advantage of memristors in neural network applications is their fast processing time and energy efficiency in the computational process. At the device level, memristors have very low power dissipation and high on-chip density. At the architecture level, parallel computing is performed at the same location where data is stored, thereby avoiding frequent data movement and memory wall issues. Due to the quantum effect and non-ideal characteristics in the manufacturing of nanometer memristors, the robust performance of memristor neural networks still needs to be improved. Meanwhile, the adaptation range of various memristor models is limited and has not been fully explored in chip design. To date, there are no complete unified memristor models for chip designer. Furthermore, wire resistance, sneak path current, and half-select problems are also challenges for high-density integration of memristor crossbar arrays. Memristor neural network research involves engineering, biology, physics, algorithms, architecture, systems, circuits, equipment, and materials. There is still a long way to go for memristive neural networks, as most research remains in single devices or small-scale prototypes. However, with the marketing promotion of the IoT big data and AI, the breakthrough research of memristor-based ANN will be realized by the joint efforts of academia and industry.

AUTHOR CONTRIBUTIONS

WX drafted the manuscript, developed the concept, and conceived the experiments. JW revised the manuscript. XY drafted and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported by the National Natural Science Foundation of China (grant nos. 62064002, 61674050, and 61874158), the Project of Distinguished Young of Hebei Province (grant no. A2018201231), the Hundred Persons Plan of Hebei Province (grant nos. E2018050004 and E2018050003), the Supporting Plan for 100 Excellent Innovative Talents

in Colleges and Universities of Hebei Province (grant no. SLRC2019018), Special project of strategic leading science and technology of Chinese Academy of Sciences (grant no. XDB44000000-7), outstanding young scientific research and innovation team of Hebei University, Special support funds for national high level talents (041500120001 and 521000981426), Hebei University graduate innovation funding project in 2021 (grant no. HBU2021bs013), and the Foundation of Guangxi Key Laboratory of Precision Navigation Technology and Application, Guilin University of Electronic Technology (No. DH201908).

REFERENCES

Adhikari, S. P., Yang, C., Kim, H., and Chua, L. O. (2012). Memristor bridge synapse-based neural network and its learning. *IEEE. Trans. Neur. Netw. Learn. Syst.* 23, 1426–1435. doi: 10.1109/TNNLS.2012.2204770

Ahmed, F. Y., Yusob, B., and Hamed, H. N. A. (2014). Computing with spiking neuron networks: a review. *Int. J. Adv. Soft Comput. Appl.* 6. Available online at: https://www.researchgate.net/publication/262374523_Computing_with_Spiking_Neuron_Networks_A_Review

Akinaga, H., and Shima, H. (2010). Resistive random access memory (ReRAM) based on metal oxides. *Proc. IEEE* 98, 2237–2251. doi: 10.1109/JPROC.2010.2070830

Alibart, F., Gao, L., Hoskins, B. D., and Strukov, D. B. (2012). High precision tuning of state for memristive devices by adaptable variation-tolerant algorithm. *Nanotechnology* 23:075201. doi: 10.1088/0957-4484/23/7/075201

Alibart, F., Zamanidoost, E., and Strukov, D. B. (2013). Pattern classification by memristive crossbar circuits using *ex situ* and *in situ* training. *Nat. Commun.* 4:2072. doi: 10.1038/ncomms3072

Al-Shedivat, M., Naous, R., Cauwenberghs, G., and Salama, K. N. (2015). Memristors empower spiking neurons with stochasticity. *IEEE. J. Emerg. Select. Top. Circ. Syst.* 5, 242–253. doi: 10.1109/JETCAS.2015.2435512

Ambrogio, S., Narayanan, P., Tsai, H., Shelby, R. M., Boybat, I., Di Nolfo, C., et al. (2018). Equivalent-accuracy accelerated neural-network training using analogue memory. *Nature* 558, 60–67. doi: 10.1038/s41586-018-0180-5

Biolek, Z., Biolek, D., and Biolkova, V. (2009). SPICE model of memristor with nonlinear dopant drift. *Radioengineering* 18, 210–214. Available online at: https://www.researchgate.net/publication/26625012_SPICE_Model_of_Memristor_with_Nonlinear_Dopant_Drift

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. New York, NY: Oxford University Press.

Cai, F., Correll, J. M., Lee, S. H., Lim, Y., Bothra, V., Zhang, Z., et al. (2019). A fully integrated reprogrammable memristor–CMOS system for efficient multiply–accumulate operations. *Nat. Electron.* 2, 290–299. doi: 10.1038/s41928-019-0270-x

Cantley, K. D., Subramaniam, A., Stiegler, H. J., Chapman, R. A., and Vogel, E. M. (2012). Neural learning circuits utilizing nano-crystalline silicon transistors and memristors. *IEEE Trans. Neural Netw. Learn. Syst.* 23, 565–573. doi: 10.1109/TNNLS.2012.2184801

Chang, C. C., Chen, P. C., Chou, T., Wang, I. T., Hudec, B., Chang, C. C., et al. (2017). Mitigating asymmetric nonlinear weight update effects in hardware neural network based on analog resistive synapse. *IEEE. J. Emerg. Select. Top. Circ. Syst.* 8, 116–124. doi: 10.1109/JETCAS.2017.2771529

Chaturvedi, S., and Khurshid, A. A. (2011). "Review of spiking neural network architecture for feature extraction and dimensionality reduction," in *2011 Fourth International Conference on Emerging Trends in Engineering and Technology* (Port Louis), 317–322. doi: 10.1109/ICETET.2011.57

Chen, B., Yang, H., Zhuge, F., Li, Y., Chang, T. C., He, Y. H., et al. (2019). Optimal tuning of memristor conductance variation in spiking neural networks for online unsupervised learning. *IEEE. Trans. Electron. Dev.* 66, 2844–2849. doi: 10.1109/TED.2019.2907541

Chen, P. Y., Lin, B., Wang, I. T., Hou, T. H., Ye, J., Vrudhula, S., et al. (2015). "Mitigating effects of non-ideal synaptic device characteristics for on-chip learning," in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)* (Austin, TX), 194–199. doi: 10.1109/ICCAD.2015.7372570

Chen, W.-H., Dou, C., Li, K.-X., Lin, W.-Y., Li, P.-Y., Huang, J.-H., et al. (2019). CMOS-integrated memristive non-volatile computing-in-memory for AI edge processors. *Nat. Electron.* 2, 420–428. doi: 10.1038/s41928-019-0288-0

Chicca, E., Indiveri, G., and Douglas, R. (2003). "An adaptive silicon synapse," in *Proceedings of the 2003 International Symposium on Circuits and Systems (ISCAS'03)* (Bangkok), I–I. doi: 10.1109/ISCAS.2003.1205505

Choi, S., Tan, S. H., Li, Z., Kim, Y., Choi, C., Chen, P. Y., et al. (2018). SiGe epitaxial memory for neuromorphic computing with reproducible high performance based on engineered dislocations. *Nat. Mater.* 17, 335–340. doi: 10.1038/s41563-017-0001-5

Choi, S., Yang, Y., and Lu, W. (2014). Random telegraph noise and resistance switching analysis of oxide based resistive memory. *Nanoscale* 6, 400–404. doi: 10.1039/C3NR05016E

Chu, M., Kim, B., Park, S., Hwang, H., Jeon, M., Lee, B. H., et al. (2014). Neuromorphic hardware system for visual pattern recognition with memristor array and CMOS neuron. *IEEE. Trans. Ind. Electron.* 62, 2410–2419. doi: 10.1109/TIE.2014.2356439

Chua, L. (1971). Memristor-the missing circuit element. *IEEE Trans. Circ. Theor.* 18, 507–519. doi: 10.1109/TCT.1971.1083337

Chua, L. (2013). Memristor, Hodgkin–Huxley, and edge of Chaos. *Nanotechnology* 24:383001. doi: 10.1088/0957-4484/24/38/383001

Duan, S., Hu, X., Dong, Z., Wang, L., and Mazumder, P. (2014). Memristor-based cellular nonlinear/neural network: design, analysis, and applications. *IEEE Trans. Neural Netw. Learn. Syst.* 26, 1202–1213. doi: 10.1109/TNNLS.2014.2334701

Fackenthal, R., Kitagawa, M., Otsuka, W., Prall, K., Mills, D., Tsutsui, K., et al. (2014). "0.19.7 A 16Gb ReRAM with 200MB/s write and 1GB/s read in 27nm technology," in *IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)* (San Francisco, CA), 338–339. doi: 10.1109/ISSCC.2014.6757460

Fu, J., Liao, Z., Gong, N., and Wang, J. (2019). Mitigating nonlinear effect of memristive synaptic device for neuromorphic computing. *IEEE. J. Emerg. Select. Top. Circ. Syst.* 9, 377–387. doi: 10.1109/JETCAS.2019.2910749

Gi, S. G., Yeo, I., Chu, M., Moon, K., Hwang, H., and Lee, B. G. (2018). Modeling and system-level simulation for nonideal conductance response of synaptic devices. *IEEE. Trans. Electron. Dev.* 65, 3996–4003. doi: 10.1109/TED.2018.2858762

Hodgkin, A. L., and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117, 500–544. doi: 10.1113/jphysiol.1952.sp004764

Hong, Q., Zhao, L., and Wang, X. (2019). Novel circuit designs of memristor synapse and neuron. *Neurocomputing* 330, 11–16. doi: 10.1016/j.neucom.2018.11.043

Hu, M., Graves, C. E., Li, C., Li, Y., Ge, N., Montgomery, E., et al. (2018). Memristor-based analog computation and neural network classification with a dot product engine. *Adv. Mater.* 30:1705914. doi: 10.1002/adma.201705914

Hu, M., Li, H., Chen, Y., Wu, Q., Rose, G. S., and Linderman, R. W. (2014). Memristor crossbar-based neuromorphic computing system: a case study. *IEEE Trans. Neural Netw. Learn. Syst.* 25, 1864–1878. doi: 10.1109/TNNLS.2013.2296777

Huang, A., Zhang, X., Li, R., and Chi, Y. (2018). *Memristor Neural Network Design*. London: IntechOpen.

Ishii, M., Kim, S., Lewis, S., Okazaki, A., Okazawa, J., Ito, M., et al. (2019). "On-Chip Trainable 1.4M 6T2R PCM Synaptic Array with 1.6K Stochastic LIF Neurons for Spiking RBM," in *IEEE International Electron Devices Meeting (IEDM)* (San Francisco, CA), 7–11.

Jeong, H., and Shi, L. (2018). Memristor devices for neural networks. *J. Phys. D Appl. Phys.* 52:023003. doi: 10.1088/1361-6463/aae223

Jeong, Y., Kim, S., and Lu, W. D. (2015). Utilizing multiple state variables to improve the dynamic range of analog switching in a memristor. *Appl. Phys. Lett.* 107:173105. doi: 10.1063/1.4934818

Jiang, H., Yamada, K., Ren, Z., Kwok, T., Luo, F., Yang, Q., et al. (2018). "Pulse-width modulation based dot-product engine for neuromorphic computing system using memristor crossbar array," in *IEEE International Symposium on Circuits and Systems (ISCAS)* (Florence), 1–4. doi: 10.1109/ISCAS.2018.8351276

Jo, S. H., Chang, T., Ebong, I., Bhadviya, B. B., Mazumder, P., and Lu, W. (2010). Nanoscale memristor device as synapse in neuromorphic systems. *Nano Lett.* 10, 1297–1301. doi: 10.1021/nl904092h

Joglekar, Y. N., and Wolf, S. J. (2009). The elusive memristor: properties of basic electrical circuits. *Eur. J. Phys.* 30:661. doi: 10.1088/0143-0807/30/4/001

Kataeva, I., Ohtsuka, S., Nili, H., Kim, H., Isobe, Y., Yako, K., et al. (2019). "Towards the development of analog neuromorphic chip prototype with 2.4 M integrated memristors," in *IEEE International Symposium on Circuits and Systems (ISCAS)* (Sapporo), 1–5. doi: 10.1109/ISCAS.2019.8702125

Kim, H., Sah, M. P., Yang, C., Roska, T., and Chua, L. O. (2011a). Memristor bridge synapses. *Proc. IEEE* 100, 2061–2070. doi: 10.1109/JPROC.2011.2166749

Kim, H., Sah, M. P., Yang, C., Roska, T., and Chua, L. O. (2011b). Neural synaptic weighting with a pulse-based memristor circuit. *IEEE Trans. Circ. Syst. I Reg. Pap.* 59, 148–158. doi: 10.1109/TCSI.2011.2161360

Kim, S., Du, C., Sheridan, P., Ma, W., Choi, S., and Lu, W. D. (2015). Experimental demonstration of a second-order memristor and its ability to biorealistically implement synaptic plasticity. *Nano Lett.* 15, 2203–2211. doi: 10.1021/acs.nanolett.5b00697

Kozhevnikov, D. D., and Krasilich, N. V. (2016). Memristor-based hardware neural networks modelling review and framework concept. *Proc. Inst. Syst. Prog. RAS* 28, 243–258. doi: 10.15514/ISPRAS-2016-28(2)-16

Krestinskaya, O., James, A. P., and Chua, L. O. (2019). Neuromemristive circuits for edge computing: a review. *IEEE Trans. Neural Netw. Learn. Syst.* 31, 4–23. doi: 10.1109/TNNLS.2019.2899262

Krestinskaya, O., Salama, K. N., and James, A. P. (2018). Learning in memristive neural network architectures using analog backpropagation circuits. *IEEE. Trans. Circ. I* 66, 719–732. doi: 10.1109/TCSI.2018.2866510

Kvatinsky, S., Friedman, E. G., Kolodny, A., and Weiser, U. C. (2012). TEAM: threshold adaptive memristor model. *IEEE Trans. Circ. Syst. I Reg. Pap.* 60, 211–221. doi: 10.1109/TCSI.2012.2215714

Lastras-Montaño, M. A., and Cheng, K. T. (2018). Resistive random-access memory based on ratioed memristors. *Nat. Electron.* 1, 466–472. doi: 10.1038/s41928-018-0115-z

Lee, D., Park, J., Moon, K., Jang, J., Park, S., Chu, M., et al. (2015). "Oxide based nanoscale analog synapse device for neural signal recognition system," in *IEEE International Electron Devices Meeting (IEDM)* (Washington, DC), 4–7. doi: 10.1109/IEDM.2015.7409628

Lee, K. J., Cho, B. H., Cho, W. Y., Kang, S., Choi, B. G., Oh, H. R., et al. (2008). A 90 nm 1.8 V 512 Mb diode-switch PRAM with 266 MB/s read throughput. *IEEE. J. Solid State Circ.* 43, 150–162. doi: 10.1109/JSSC.2007.908001

Li, C., Belkin, D., Li, Y., Yan, P., Hu, M., Ge, N., et al. (2018). Efficient and self-adaptive *in-situ* learning in multilayer memristor neural networks. *Nat. Commun.* 9:2385. doi: 10.1038/s41467-018-04484-2

Li, C., Wang, Z., Rao, M., Belkin, D., Song, W., Jiang, H., et al. (2019). Long short-term memory networks in memristor crossbar arrays. *Nat. Mach. Intell.* 1, 49–57. doi: 10.1038/s42256-018-0001-4

Li, S., Wen, J., Chen, T., Xiong, L., Wang, J., and Fang, G. (2016). *In situ* synthesis of 3D CoS nanoflake/Ni (OH) 2 nanosheet nanocomposite structure as a candidate supercapacitor electrode. *Nanotechnology* 27:145401. doi: 10.1088/0957-4484/27/14/145401

Li, S. J., Dong, B. Y., Wang, B., Li, Y., Sun, H. J., He, Y. H., et al. (2018). Alleviating conductance nonlinearity via pulse shape designs in TaO x memristive synapses. *IEEE. Trans. Electron. Dev.* 66, 810–813. doi: 10.1109/TED.2018.2876065

Lin, P., Li, C., Wang, Z., Li, Y., Jiang, H., Song, W., et al. (2020). Three-dimensional memristor circuits as complex neural networks. *Nat. Electron.* 3, 225–232. doi: 10.1038/s41928-020-0397-9

Liu, J., Li, Z., Tang, Y., Hu, W., and Wu, J. (2020). 3D Convolutional Neural Network based on memristor for video recognition. *Pattern. Recogn. Lett.* 130, 116–124. doi: 10.1016/j.patrec.2018.12.005

Liu, T. Y., Yan, T. H., Scheuerlein, R., Chen, Y., Lee, J. K., Balakrishnan, G., et al. (2013). "A 130.7mm^2 2-layer 32Gb ReRAM memory device in 24nm technology," in *IEEE International Solid-State Circuits Conference Digest of Technical Papers* (San Francisco, CA), 210–211. doi: 10.1109/JSSC.2013.2280296

Luo, Q., Xu, X., Gong, T., Lv, H., Dong, D., Ma, H., et al. (2020). "8-Layers 3D vertical RRAM with excellent scalability towards storage class memory applications," in *IEEE International Electron Devices Meeting (IEDM)* (San Francisco, CA), 2.7.1–2.7. 4.

Ma, D., Wang, G., Han, C., Shen, Y., and Liang, Y. (2018). A memristive neural network model with associative memory for modeling affections. *IEEE Access.* 6, 61614–61622. doi: 10.1109/ACCESS.2018.2875433

Midya, R., Wang, Z., Asapu, S., Joshi, S., Li, Y., Zhuo, Y., et al. (2019a). Artificial neural network (ANN) to spiking neural network (SNN) converters based on diffusive memristors. *Adv. Electron. Mater.* 5:1900060. doi: 10.1002/aelm.201900060

Midya, R., Wang, Z., Asapu, S., Zhang, X., Rao, M., Song, W., et al. (2019b). Reservoir computing using diffusive memristors. *Adv. Intell. Syst.* 1:1900084. doi: 10.1002/aisy.201900084

Midya, R., Wang, Z., Zhang, J., Savel'ev, S. E., Li, C., Rao, M., et al. (2017). Anatomy of Ag/Hafnia-based selectors with 10^{10} nonlinearity. *Adv. Mater.* 29:1604457. doi: 10.1002/adma.201604457

Nenadic, Z., and Ghosh, B. K. (2001). "Computation with biological neurons," in *Proceedings of the 2001 American Control Conference (Cat. No. 01CH37148)* (Arlington, VA), 257–262. doi: 10.1109/ACC.2001.945552

Ohno, T., Hasegawa, T., Tsuruoka, T., Terabe, K., Gimzewski, J. K., and Aono, M. (2011). Short-term plasticity and long-term potentiation mimicked in single inorganic synapses. *Nat. Mater.* 10, 591–595. doi: 10.1038/nmat3054

Ojiugwo, C. N., Abdallah, A. B., and Thron, C. (2020). "Simulation of biological learning with spiking neural networks," in *Implementations and Applications of Machine Learning,* Vol. 782, eds S. A. Subair and C. Thron (Cham: Springer),207–227. doi: 10.1007/978-3-030-37830-1_9

Park, S., Sheri, A., Kim, J., Noh, J., Jang, J., Jeon, M., et al. (2013). "Neuromorphic speech systems using advanced ReRAM-based synapse," in *IEEE International Electron Devices Meeting* (Washington, DC), 25.6.1–25.6.4. doi: 10.1109/IEDM.2013.6724692

Pfeiffer, M., and Pfeil, T. (2018). Deep learning with spiking neurons: opportunities and challenges. *Front. Neurosci.* 12:774. doi: 10.3389/fnins.2018.00774

Pickett, M. D., Strukov, D. B., Borghetti, J. L., Yang, J. J., Snider, G. S., Stewart, D. R., et al. (2009). Switching dynamics in titanium dioxide memristive devices. *J. Appl. Phys.* 106:074508. doi: 10.1063/1.3236506

Prezioso, M., Merrikh-Bayat, F., Hoskins, B. D., Adam, G. C., Likharev, K. K., and Strukov, D. B. (2015). Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* 521, 61–64. doi: 10.1038/nature14441

Prodromakis, T., Peh, B. P., Papavassiliou, C., and Toumazou, C. (2011). A versatile memristor model with nonlinear dopant kinetics. *IEEE. Trans. Electron Dev.* 58, 3099–3105. doi: 10.1109/TED.2011.2158004

Sarwar, S. S., Saqueb, S. A. N., Quaiyum, F., and Rashid, A. H. U. (2013). Memristor-based nonvolatile random access memory: hybrid architecture for low power compact memory design. *IEEE Access* 1, 29–34. doi: 10.1109/ACCESS.2013.2259891

Seo, J. S., Brezzo, B., Liu, Y., Parker, B. D., Esser, S. K., Montoye, R. K., et al. (2011). "A 45nm CMOS neuromorphic chip with a scalable architecture for learning in networks of spiking neurons," in *IEEE Custom Integrated Circuits Conference (CICC)* (San Jose, CA), 1–4. doi: 10.1109/CICC.2011.6055293

Seok, J. Y., Song, S. J., Yoon, J. H., Yoon, K. J., Park, T. H., Kwon, D. E., et al. (2014). A review of three-dimensional resistive switching cross-bar array memories from the integration and materials property points of view. *Adv. Funct. Mater.* 24, 5316–5339. doi: 10.1002/adfm.201303520

Shukla, A., and Ganguly, U. (2018). An on-chip trainable and the clock-less spiking neural network with 1R memristive synapses. *IEEE. Trans. Biomed. Circ. Syst.* 12, 884–893. doi: 10.1109/TBCAS.2018.2831618

Simmons, J. G. (1963). Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film. *J. Appl. Phys.* 34, 1793–1803. doi: 10.1063/1.1702682

Smagulova, K., Krestinskaya, O., and James, A. P. (2018). A memristor-based long short term memory circuit. *Analog. Integr. Circ. Signal Process.* 95, 467–472. doi: 10.1007/s10470-018-1180-y

Strukov, D. B., Snider, G. S., Stewart, D. R., and Williams, R. S. (2008). The missing memristor found. *Nature* 453, 80–83. doi: 10.1038/nature06932

Strukov, D. B., and Williams, R. S. (2009). Exponential ionic drift: fast switching and low volatility of thin-film memristors. *Appl. Phys. A* 94, 515–519. doi: 10.1007/s00339-008-4975-3

Sun, J. (2015). *CMOS and Memristor Technologies for Neuromorphic Computing Applications.* Technical Report No. UCB/EECS-2015–S-2218, Electrical Engineering and Computer Sciences University of California at Berkeley.

Sun, K., Chen, J., and Yan, X. (2020). The future of memristors: materials engineering and neural networks. *Adv. Funct. Mater.* 31:2006773. doi: 10.1002/adfm.202006773

Tanikawa, T., Ohnishi, K., Kanoh, M., Mukai, T., and Matsuoka, T. (2018). Three-dimensional imaging of threading dislocations in GaN crystals using two-photon excitation photoluminescence. *Appl. Phys. Express*. 11:031004. doi: 10.7567/APEX.11.031004

Tsai, H., Ambrogio, S., Mackin, C., Narayanan, P., Shelby, R. M., Rocki, K., et al. (2019). "Inference of long-short term memory networks at software-equivalent accuracy using 2.5M analog phase change memory devices," in *Symposium on VLSI Technology* (Kyoto), 82–83. doi: 10.23919/VLSIT.2019.8776519

Upadhyay, N. K., Sun, W., Lin, P., Joshi, S., Midya, R., Zhang, X., et al. (2020). A memristor with low switching current and voltage for 1s1r integration and array operation. *Adv. Electron. Mater.* 6:1901411. doi: 10.1002/aelm.201901411

Volos, C. K., Kyprianidis, I. M., Stouboulos, I. N., Tlelo-Cuautle, E., and Vaidyanathan, S. (2015). Memristor: a new concept in synchronization of coupled neuromorphic circuits. *J. Eng. Sci. Technol. Rev.* 8, 157–173. doi: 10.25103/jestr.082.21

Wang, H., and Yan, X. (2019). Overview of resistive random access memory (RRAM): materials, filament mechanisms, performance optimization, and prospects. *Phys. Status. Solidi R.* 13:1900073. doi: 10.1002/pssr.201900073

Wang, T., and Roychowdhury, J. (2016). Well-posed models of memristive devices. *arXiv* preprint arXiv:1605.04897.

Wang, Z., Joshi, S., Savel'ev, S., Song, W., Midya, R., Li, Y., et al. (2018a). Fully memristive neural networks for pattern classification with unsupervised learning. *Nat Electron.* 1, 137–145. doi: 10.1038/s41928-018-0023-2

Wang, Z., Joshi, S., Saveliev, S. E., Jiang, H., Midya, R., Lin, P., et al. (2017). Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing. *Nat. Mater.* 16, 101–108. doi: 10.1038/nmat4756

Wang, Z., Li, C., Lin, P., Rao, M., Nie, Y., Song, W., et al. (2019a). *In situ* training of feed-forward and recurrent convolutional memristor networks. *Nat. Mach. Intell.* 1, 434–442. doi: 10.1038/s42256-019-0089-1

Wang, Z., Li, C., Song, W., Rao, M., Belkin, D., Li, Y., et al. (2019b). Reinforcement learning with analogue memristor arrays. *Nat. Electron.* 2, 115–124. doi: 10.1038/s41928-019-0221-6

Wang, Z., Rao, M., Han, J. W., Zhang, J., Lin, P., Li, Y., et al. (2018b). Capacitive neural network with neuro-transistors. *Nat. Commun.* 9:3208. doi: 10.1038/s41467-018-05677-5

Wang, Z., Wu, H., Burr, G. W., Hwang, C. S., Wang, K. L., Xia, Q., et al. (2020). Resistive switching materials for information processing. *Nat. Rev. Mater.* 5, 173–195. doi: 10.1038/s41578-019-0159-3

Wang, Z., Zhao, W., Kang, W., Zhang, Y., Klein, J. O., and Chappert, C. (2014). "Ferroelectric tunnel memristor-based neuromorphic network with 1T1R crossbar architecture," *International Joint Conference on Neural Networks (IJCNN)* (Beijing), 29–34. doi: 10.1109/IJCNN.2014.6889951

Waser, R., Dittmann, R., Staikov, G., and Szot, K. (2009). Redox-based resistive switching memories–nanoionic mechanisms, prospects, and challenges. *Adv. Mater.* 21, 2632–2663. doi: 10.1002/adma.200900375

Wong, H. S. P., Lee, H. Y., Yu, S., Chen, Y. S., Wu, Y., Chen, P. S., et al. (2012). Metal–oxide RRAM. *Proc. IEEE* 100, 1951–1970. doi: 10.1109/JPROC.2012.2190369

Woo, J., and Yu, S. (2018). Resistive memory-based analog synapse: the pursuit for linear and symmetric weight update. *IEEE Nanotechnol. Mag.* 12, 36–44. doi: 10.1109/MNANO.2018.2844902

Wu, X., and Saxena, V. (2018). Dendritic-inspired processing enables bio-plausible STDP in compound binary synapses. *IEEE. Trans. Nanotechnol.* 18, 149–159. doi: 10.1109/TNANO.2018.2871680

Xia, Q., and Yang, J. J. (2019). Memristive crossbar arrays for brain-inspired computing. *Nat. Mater.* 18, 309–323. doi: 10.1038/s41563-019-0291-x

Xue, C.-X., Chiu, Y.-C., Liu, T.-W., Huang, T.-Y., Liu, J.-S., Chang, T.-W., et al. (2020). A CMOS-integrated compute-in-memory macro based on resistive random-access memory for AI edge devices. *Nat. Electron.* 4, 81–90. doi: 10.1038/s41928-020-00505-5

Yan, X., Li, X., Zhou, Z., Zhao, J., Wang, H., Wang, J., et al. (2019a). Flexible transparent organic artificial synapse based on the tungsten/egg albumen/indium tin oxide/polyethylene terephthalate memristor. *ACS. Appl. Mater. Inter.* 11, 18654–18661. doi: 10.1021/acsami.9b04443

Yan, X., Pei, Y., Chen, H., Zhao, J., Zhou, Z., Wang, H., et al. (2019b). Self-assembled networked PbS distribution quantum dots for resistive switching and artificial synapse performance boost of memristors. *Adv. Mater.* 31:1805284. doi: 10.1002/adma.201805284

Yan, X., Wang, K., Zhao, J., Zhou, Z., Wang, H., Wang, J., et al. (2019c). A new memristor with 2D Ti3C2Tx MXene flakes as an artificial bio-synapse. *Small* 15, 1900107. doi: 10.1002/smll.201900107

Yan, X., Zhang, L., Chen, H., Li, X., Wang, J., Liu, Q., et al. (2018a). Graphene oxide quantum dots based memristors with progressive conduction tuning for artificial synaptic learning. *Adv. Funct. Mater.* 28:1803728. doi: 10.1002/adfm.201803728

Yan, X., Zhang, L., Yang, Y., Zhou, Z., Zhao, J., Zhang, Y., et al. (2017). Highly improved performance in Zr 0.5 Hf 0.5 O_2 films inserted with graphene oxide quantum dots layer for resistive switching non-volatile memory. *J. Mater. Chem. C.* 5, 11046–11052. doi: 10.1039/C7TC03037A

Yan, X., Zhao, J., Liu, S., Zhou, Z., Liu, Q., Chen, J., et al. (2018b). Memristor with Ag-cluster-doped TiO_2 films as artificial synapse for Neuroinspired computing. *Adv. Funct. Mater.* 28:1705320. doi: 10.1002/adfm.201705320

Yan, X., Zhao, Q., Chen, A. P., Zhao, J., Zhou, Z., Wang, J., et al. (2019d). Vacancy-induced synaptic behavior in 2D WS2 nanosheet–based memristor for low-power neuromorphic computing. *Small* 15:1901423. doi: 10.1002/smll.201901423

Yan, X., Zhou, Z., Zhao, J., Liu, Q., Wang, H., Yuan, G., et al. (2018c). Flexible memristors as electronic synapses for neuro-inspired computation based on scotch tape-exfoliated mica substrates. *Nano. Res.* 11, 1183–1192. doi: 10.1007/s12274-017-1781-2

Yang, C., Adhikari, S. P., and Kim, H. (2019). On learning with nonlinear memristor-based neural network and its replication. *IEEE. Trans. Circ. I.* 66, 3906–3916. doi: 10.1109/TCSI.2019.2914125

Yang, C., Choi, H., Park, S., Sah, M. P., Kim, H., and Chua, L. O. (2014). A memristor emulator as a replacement of a real memristor. *Semicond. Sci. Technol.* 30, 015007. doi: 10.1088/0268-1242/30/1/015007

Yang, J. J., Strukov, D. B., and Stewart, D. R. (2013). Memristive devices for computing. *Nat. Nanotechnol.* 8, 13–24. doi: 10.1038/nnano.2012.240

Yao, P., Wu, H., Gao, B., Tang, J., Zhang, Q., Zhang, W., et al. (2020). Fully hardware-implemented memristor convolutional neural network. *Nature* 577, 641–646. doi: 10.1038/s41586-020-1942-4

Zhang, Y., Wang, X., and Friedman, E. G. (2017). Memristor-based circuit design for multilayer neural networks. *IEEE. Trans. Circ. I.* 65, 677–686. doi: 10.1109/TCSI.2017.2729787

Zhang, Y., Wang, Z., Zhu, J., Yang, Y., Rao, M., Song, W., et al. (2020). Brain-inspired computing with memristors: challenges in devices, circuits, and systems. *Appl. Phys. Rev.* 7:011308. doi: 10.1063/1.5124027

Zhao, J., Zhou, Z., Zhang, Y., Wang, J., Zhang, L., Li, X., et al. (2019). An electronic synapse memristor device with conductance linearity using quantized conduction for neuroinspired computing. *J. Mater. Chem. C* 7, 1298–1306. doi: 10.1039/C8TC04395G

Zhao, Q., Xie, Z., Peng, Y. P., Wang, K., Wang, H., Li, X., et al. (2020). Current status and prospects of memristors based on novel 2D materials. *Mater. Horiz.* 7, 1495–1518. doi: 10.1039/C9MH02033K

Zheng, N., and Mazumder, P. (2018). Learning in memristor crossbar-based spiking neural networks through modulation of weight-dependent spike-timing-dependent plasticity. *IEEE. Trans. Nanotechnol.* 17, 520–532. doi: 10.1109/TNANO.2018.2821131

Zidan, M. A., Fahmy, H. A. H., Hussain, M. M., and Salama, K. N. (2013). Memristor-based memory: the sneak paths problem and solutions. *Microelectron. J.* 44, 176–183. doi: 10.1016/j.mejo.2012.10.001

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.

frontiers in Nanotechnology

ORIGINAL RESEARCH
published: 26 March 2021
doi: 10.3389/fnano.2021.649560

Reliable 2D Phase Transitions for Low-Noise and Long-Life Memory Programming

*Keyuan Ding, Tianci Li, Bin Chen and Feng Rao**

College of Materials Science and Engineering, Shenzhen University, Shenzhen, China

Extending cycling endurance and suppressing programming noise of phase-change random-access memory (PCRAM) are the key challenges with respect to the development of nonvolatile working memory and high-accuracy neuromorphic computing devices. However, the large-scale atomic migration along electrical pulse direction in the unconstrained three-dimensional phase transitions of the phase-change materials (PCMs) induces big resistance fluctuations upon repeated programming and renders the classic PCRAM devices into premature failure with limited cycling endurance. Previous efforts of superlattice-like and superlattice PCM schemes cannot effectively resolve such issues. In this work, we demonstrated that, through fine-tuning the sputtering techniques, a phase-change heterostructure (PCH) of Sb_2Te_3/$TiTe_2$ can be successfully constructed. In contrast to its superlattice-like counterpart with inferior crystal quality, the well-textured PCH architecture ensures the reliable (well-confined) two-dimensional phase transitions, promoting an ultralow-noise and long-life operation of the PCRAM devices. Our study thus provides a useful reference for better manufacturing the PCH architecture and further exploring the excellent device performances and other new physics.

Keywords: phase-change random-access memory, two-dimensional phase transitions, programming noise, cycling endurance, heterostructure

OPEN ACCESS

Edited by:
Huanglong Li,
Tsinghua University, China

Reviewed by:
Yuta Saito,
National Institute of Advanced Industrial Science and Technology (AIST), Japan
Xinglong Ji,
Tsinghua University, China

****Correspondence:***
Feng Rao
fengrao@szu.edu.cn

Specialty section:
This article was submitted to Nanodevices, a section of the journal Frontiers in Nanotechnology

Received: *05 January 2021*
Accepted: *02 February 2021*
Published: *26 March 2021*

Citation:
Ding K, Li T, Chen B and Rao F (2021) Reliable 2D Phase Transitions for Low-Noise and Long-Life Memory Programming. Front. Nanotechnol. 3:649560. doi: 10.3389/fnano.2021.649560

INTRODUCTION

With the development of the Internet of Things, the exponentially growing demands in data processing and storage have imposed critical requirements on the energy efficiency and computing speed for data-centric tasks. But in the current computing system based on classic von Neumann architecture, the constant data shuttling between a fast central processing unit (CPU) and other much slower program and storage memory units leads to significantly wasted working power and limited computing speed (Kestor et al., 2013; Wong and Salahuddin, 2015). Extensive studies on phase-change random-access memory (PCRAM) (Wuttig, 2005; Raoux et al., 2010) thus have been devoted to resolving the issues. One route is to renovate the von Neumann architecture by alleviating performance mismatch among hierarchical memories (Lam, 2010; Rao et al., 2015; Yu and Chen, 2016; Rao et al., 2017), such as Intel's Optane DC (Choe, 2017) chips bridged between volatile dynamic random-access memory (DRAM) and nonvolatile solid-state drive (SSD) flash memory. The other is to innovate non-von Neumann architecture by unifying processing with storage in PCRAM cells, such as neuromorphic computing of three paradigms (Kuzum et al., 2012; Tuma et al., 2016; Burr et al., 2017; Boybat et al., 2018; Ielmini and Wong, 2018; Le Gallo et al., 2018; Sebastian et al., 2018): in-memory computing, deep neural networks, and spiking neural networks.

Commercialized PCRAM device encodes digital information through reversible transformation between amorphous and crystalline phases of chalcogenide PCMs (Chen et al., 2019; Rao et al., 2019), e.g., $Ge_2Sb_2Te_5$ (GST). Typical (mushroom- or pillar-type) PCRAM devices execute three-dimensional (3D) phase transitions of GST film, where extensive cycles of high-energy/high-bias RESET (melting and amorphization) operation pulses trigger long-distance migrations of Sb (Ge) and Te elements in opposite (vertical) directions (Padilla et al., 2010; Xie et al., 2018), giving rise to the phase segregation and the formation of large voids near the bottom electrode. This degrades the performing reliability and limits the endurance of mass-produced GST devices to ~10^9–10^{12} cycles (Yu and Chen, 2016; Xie et al., 2018), obstructing the implementation of PCRAM as nonvolatile working memory (with > ~10^{16} cycles) for the deep modification of the von Neumann architecture (Wong and Salahuddin, 2015). Such a considerable deviation/variation in composition and microstructure during the unconstrained 3D phase transitions generates large fluctuations (noises) in programming resistance states, which poses crucial challenges to the accomplishments of high-accuracy and high-efficiency matrix-vector multiplications, unsupervised learning of temporal patterns, and other data-centric computational tasks (Burr et al., 2017; Sebastian et al., 2018).

Many efforts, e.g., superlattice-like (SLL) (Chong et al., 2006; Lu et al., 2012; Chia Tan et al., 2013) and superlattice (SL) (Simpson et al., 2011; Soeya et al., 2013; Takaura et al., 2014) PCM architectures, have been made to address the issue of limited endurance, attempting to tailor the 3D phase transitions into 2D fashion. However, both schemes encounter difficulties in maintaining a reliable 2D structural transformation upon repeated programming, because the RESET operations must be cautiously performed to avoid local overheating; otherwise the multilayers may melt together and then quench into a mixed amorphous phase (Simpson et al., 2011; Li et al., 2018), as the melting temperatures (T_m, being ~900–1,000 K) of the adopted PCMs in SLL or SL architectures are quite close (Chong et al., 2006; Simpson et al., 2011; Lu et al., 2012; Chia Tan et al., 2013; Soeya et al., 2013; Takaura et al., 2014). In addition, the growth condition must be tightly controlled to construct Ge(Sn)Te/Sb_2Te_3 SLs as Ge(Sn)Te is chemically reactive and may alloy into Ge(Sn)SbTe-like compounds easily during synthesis (Li et al., 2018). Inspired by the previous findings, we recently proposed a distinct approach to address the above issues by an innovative PCH design using Sb_2Te_3/$TiTe_2$ stackings (Ding et al., 2019). The relations between crystal quality of such stackings and the derived electrical performances have not been disclosed yet. In this work, we draw direct comparisons between Sb_2Te_3/$TiTe_2$ SLL and PCH (SL) cases. Note that here we only care about the sputtering technology that is commonly employed in mass production of PCRAM chips, rather than other techniques such as molecular beam epitaxy or chemical vapor deposition. We further reveal that, in contrast to the SLL case, only the well-textured (highly oriented) PCH architecture can guarantee a reliable 2D switching to inhibit large-scale (long-distance) atomic diffusion along electrical pulse direction, enabling substantially prolonged cycling endurance and suppressed programming noise.

EXPERIMENTAL SECTION

Film Preparation and Characterization

The Sb_2Te_3 and $TiTe_2$ films were deposited on a SiO_2/Si substrate by sputtering the respective pure target in ultra-high vacuum with a base pressure of <~1×10^{-8} Torr, and the deposition pressure was under ~4.7 mTorr. For *in situ* heating (at ~300°C) deposition of the Sb_2Te_3, $TiTe_2$, and PCH films, the deposition rate of Sb_2Te_3 and $TiTe_2$ sublayers was controlled to be ~0.5–1.0 nm/min. A ~5 nm thick Sb_2Te_3 seed layer was predeposited on the substrate before PCH film deposition. The ordinary Sb_2Te_3, $TiTe_2$, and SLL films were deposited at room temperature without the seed layer, and the deposition rate of Sb_2Te_3 and $TiTe_2$ sublayers was controlled to be ~5 nm/min. The deposited Sb_2Te_3 and $TiTe_2$ sublayers are ~5 nm and ~3 nm thick in PCH and SLL architectures. Half of the as-deposited films were postannealed at 300°C for 1 h. A ~10 nm thick SiO_2 capping layer was *in situ* grown on top of each film inside the vacuum chamber to avoid oxidation. The film compositions were confirmed by Axios X-ray fluorescence spectrum (PANalytical B.V. Netherlands). The surface morphology of the films was analyzed by field emission scanning electron microscope (SEM) ZEISS SUPRA 55. The X-ray diffraction (XRD) method was employed to characterize the crystal structures of the films (~100 nm in thickness) in the 2θ range of 5–60°, using Cu/K_α radiation with a scanning step of 0.02°.

Device Fabrication and Electrical Characterization

Mushroom-type SLL- and PCH-based PCRAM devices with tungsten bottom electrode contact of ~190 nm in diameter were fabricated using the 0.13 μm node complementary metal-oxide semiconductor technology. The thickness of the SLL and PCH films in the devices was controlled to be ~70 nm. The ~15 nm thick TiN and ~300 nm thick Al films were used as top electrode in all devices. The PCH film was deposited onto the bottom electrode of the PCRAM device under ~300°C, while SLL film in the PCRAM device was deposited at room temperature. The as-fabricated SLL-based device then was postannealed at 300°C for 1 h. All the electrical measurements on PCRAM devices were performed by using the Keithley 2400C source meter (measuring device/film resistance) and the Tektronix AWG5002B/AWG5202 pulse generator (generating voltage pulse with a minimum width of ~6 ns).

Ab Initio Simulations

All the ab initio simulations were carried out with Vienna Ab initio Simulations Package (VASP) (Kresse and Hafner, 1993). The Perdew-Burke-Ernzerhof (PBE) functionals (Perdew et al., 1996) and the projector augmented wave (PAW) pseudopotentials (Blöchl, 1994) were used for VASP. The 233-atom hexagonal supercells of SLL and PCH architectures were

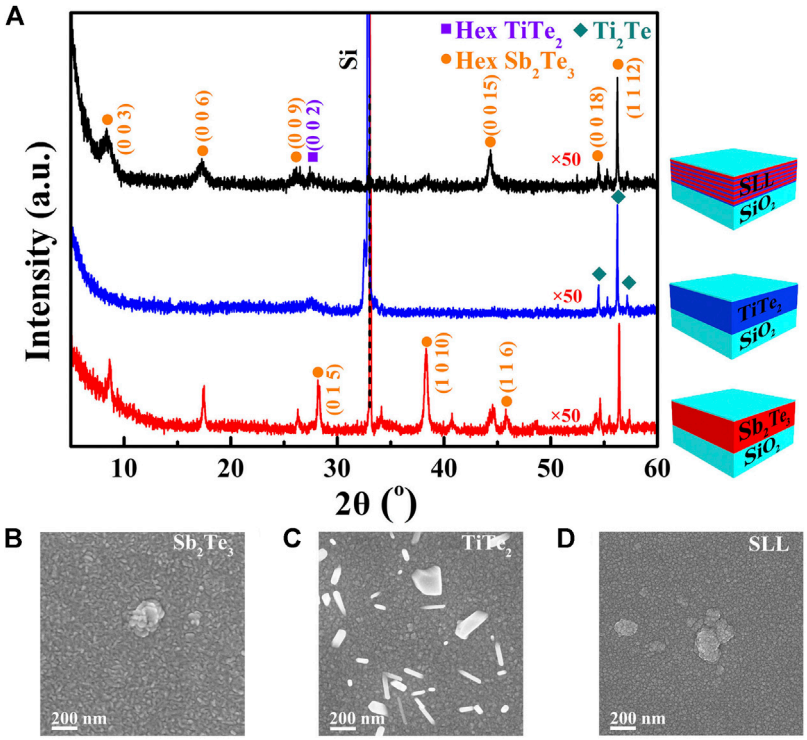


FIGURE 1 | (A) XRD curves of ~100 nm thick Sb_2Te_3, $TiTe_2$, and SLL films sandwiched between SiO_2 layers (see the corresponding sketch). The films were all deposited at room temperature and then postannealed at 300°C for 1 h, mainly showing the hexagonal (Hex) lattice configuration. For the sake of better observing the weak diffraction peaks, the diffraction intensity is multiplied by 50 for all the curves. **(B–D)** SEM images of the postannealed Sb_2Te_3, $TiTe_2$, and SLL films, respectively.

simulated with periodic boundary conditions by NVT density functional theory- (DFT-) based molecular dynamics (DFMD). The energy cutoff is 180 eV and the time step is 3 femtoseconds.

RESULTS AND DISCUSSION

Crystal Orientation and Morphology

Regarding the SLL films synthesized at room temperature, the sublayers are usually amorphous or have poor crystallinity (Chong et al., 2006; Lu et al., 2012; Chia Tan et al., 2013). Similar situation was observed for the $Sb_2Te_3/TiTe_2$ SLL samples, as well as the pure Sb_2Te_3 and $TiTe_2$ films grown on SiO_2 substrates (**Supplementary Figure S1**). We employed strong postannealing actions (at 300°C for 1 h) on the as-deposited samples to promote crystallization (**Figure 1A**). Both the Sb_2Te_3 and $TiTe_2$ films possess a hexagonal lattice configuration with multiple different (random) crystal orientations (**Figure 1A**), showing polycrystalline morphology with quite small grain size (< ~20 nm) (**Figures 1B,C**). One can find many crystal ribbons and particles that may belong to the Ti-rich titanium tellurides precipitate from the annealed $TiTe_2$ film (**Figure 1C**), which shall form rough interfaces between the sublayers inside the SLL film. It is clear that the postannealed $Sb_2Te_3/TiTe_2$ SLL film inherits almost all the crystal orientations from its subunits (**Figure 1A**), as well as an unsmoothed surface where some crystal grains aggregated into small islands (**Figure 1D**).

In stark contrast to the inferior crystalline qualities of the as-deposited and postannealed SLL films, the diffraction intensity, crystal orientation, and surface morphology of the samples fabricated by using *in situ* heating (at ~300°C) and slow growth (~0.5–1.0 nm/min) technique (see Materials and Methods) are significantly optimized (**Figure 2**). A ~5 nm thick hexagonal Sb_2Te_3 seed layer was pregrown on the SiO_2 substrate before the film deposition (Saito et al., 2016; Zhou et al., 2016), acting as a crystallization template to assist epitaxial-like crystal growth. All the *in situ* heating samples have much higher diffraction intensities than those of the postannealed ones, denoting the complete crystallinity for the formers (**Supplementary Figure S2**). It is also worth noting that only the strong (0 0 *l*) diffraction peaks appear for the *in situ* heating Sb_2Te_3 sample (**Figure 2A**), whereas the most prominent diffraction peaks of the postannealed Sb_2Te_3, i.e., (0 1 5) and (1 0 10) in **Figure 1A**, become invisible. This identifies that the degree of *c*-axis orientation of Sb_2Te_3 crystal is greatly improved by the optimized growth technique. The *in situ* heating Sb_2Te_3 sample has quite larger hexagonal grains (>~150–200 nm), with the (0 0 *l*) plane parallel to the substrate surface (**Figure 2B**). The *in situ* heating $TiTe_2$ sample is also well oriented along *c*-axis (**Figure 2A**), exhibiting a uniform surface morphology (**Figure 2C**), without any big segregated crystals of non-(0 0 *l*) orientations. On this basis, we then alternately deposited the

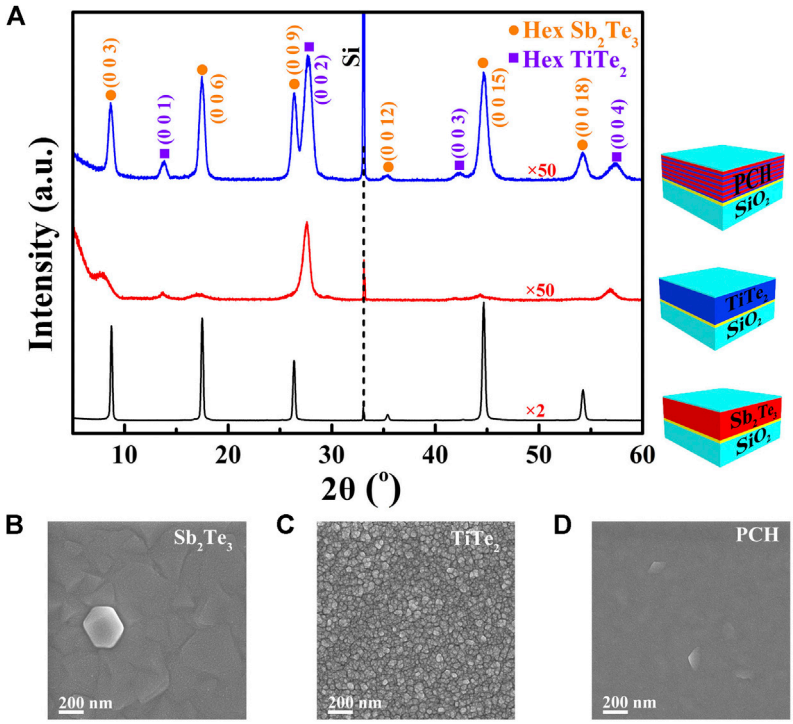


FIGURE 2 | (A) XRD curves of ~100 nm thick Sb_2Te_3, $TiTe_2$, and PCH films sandwiched between SiO_2 layers (*see* corresponding sketch). All the films were deposited at 300°C on the pregrown seed layer (yellow thin layer in each sketch), exhibiting hexagonal (Hex) lattice configuration with strong *c*-axis (0 0 *l*) orientation. For the sake of better observing the weak diffraction peaks, the diffraction intensity is multiplied by 2 for the Sb_2Te_3 curve and 50 for the $TiTe_2$ and PCH curves. **(B–D)** SEM images of the *in situ* heating deposited Sb_2Te_3, $TiTe_2$, and PCH films, respectively.

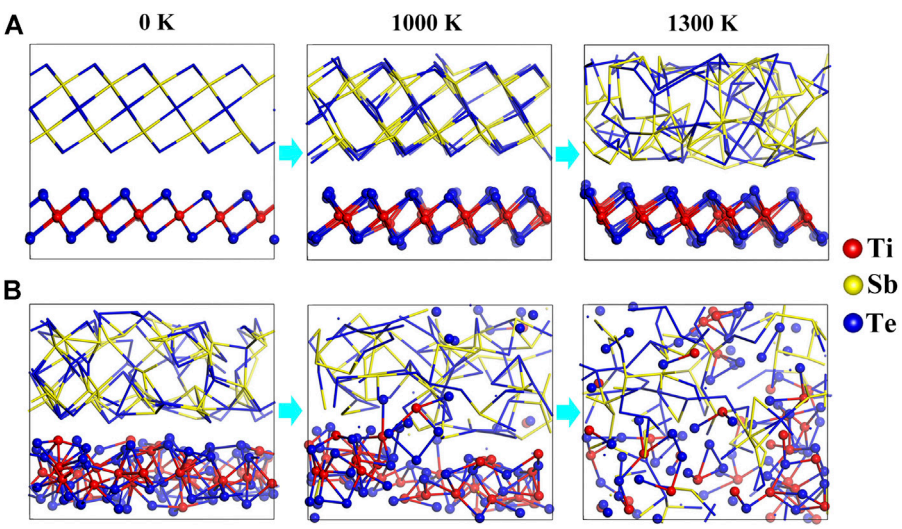


FIGURE 3 | (A) DFMD simulations of the melting process of the PCH model. The PCH model is firstly heated up to 1,000 K rapidly within 30 ps. The Sb_2Te_3 and $TiTe_2$ sublayers are ordered. The Sb_2Te_3 sublayer is fully melted at 1,300 K after 30 ps, while the $TiTe_2$ sublayer remains in the stable crystalline form. **(B)** DFMD simulations of the melting process of the SLL model. The SLL model contains partially ordered Sb_2Te_3 and $TiTe_2$ sublayers. After heating up to 1,000 K for 30 ps, the Sb_2Te_3 sublayer becomes fully disordered and mixes with the $TiTe_2$ sublayer. As the model is further heated at 1,300 K for 30 ps, the whole model becomes a fully disordered liquid phase.

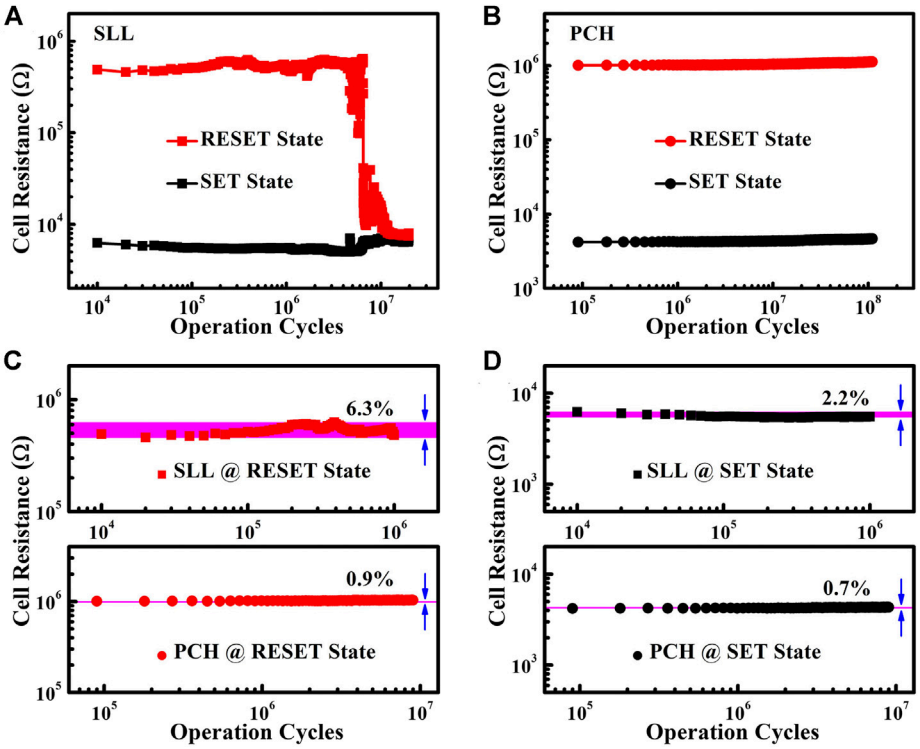


FIGURE 4 | (A) Approximately ~10^7 cycling endurance of the SLL-based PCRAM device that finally failed due to SET stuck: SET (under 2.1 V) and RESET (under 3.2 V) with 20 ns width voltage pulses. **(B)** Approximately ~10^8 cycles of the PCH device without failure under 10 ns width SET (1.6 V) and RESET (2.4 V) operations voltage pulses. **(C)** Comparison of the RESET resistance fluctuations between the SLL- and PCH-based device, with the RSD being 6.3 and 0.9%, respectively. **(D)** Comparison of the SET resistance fluctuations between the SLL- and PCH-based device, with the RSD being 2.2 and 0.7%, respectively.

Sb_2Te_3 and $TiTe_2$ nanolayers to construct the PCH architecture. Unsurprisingly, it also displays only the (0 0 *l*) diffraction peaks corresponding to the ones of its subunits (**Figure 2A**). Compared to the coarse granular surface of the SLL film, triangular crystals with rather bigger size (> ~100 nm) and pretty smooth texture were formed in the PCH film with (0 0 *l*) facets parallel to the substrate surface (**Figure 2D**).

Reliable Two-Dimensional Phase Transitions

In the well-textured $Sb_2Te_3/TiTe_2$ PCH, the $TiTe_2$ blocks with high chemical and thermal stabilities are capable of being the robust confinement layers to restrict the phase transitions of the Sb_2Te_3 blocks on 2D scale, as demonstrated by the DFMD simulations (**Figure 3A**). The initial PCH model was heated up to 1,000 K and maintained for 30 ps, where both Sb_2Te_3 and $TiTe_2$ blocks can still keep crystalline form. As the temperature further increases to 1,300 K, and after 30 ps, the Sb_2Te_3 block is fully melted, while the $TiTe_2$ block remains to be ordered.

Regarding the as-deposited and postannealed SLL films of poor crystallinity, there are most likely no clean and nonatomic (van der Waals-like) gaps between their sublayers; therefore the interlayer force cannot be weak enough. The closely bound sublayers may be merged into a compound (bulky) phase upon aggressive RESET operation or extensive programming, making the 3D switching dominate eventually. Our DFMD simulations also qualitatively illustrated such a phenomenon (**Figure 3B**), where initially the Sb_2Te_3 and $TiTe_2$ sublayers are chosen to be partially crystallized (or quite disordered) as according to the crystallographic results shown in **Figure 1** and **Supplementary Figure S2**. The SLL model experienced exactly the same heating process as the PCH model. The sublayers in the SLL model are totally disordered after heating at 1,000 K for 30 ps, and the subsequent heating at 1,300 K for another 30 ps finally melts them down into a mixed liquid.

Extended Cycling Endurance and Suppressed Programming Noise

We fabricated the mushroom-type SLL- and PCH-based PCRAM devices with the same geometry (*see Materials and Methods* section) to draw comparison of their electrical performances. The PCH-based device has lower SET and RESET voltages as compared to those of the SLL-based device (**Supplementary Figures S3, S4**), correlating to the reduced programming energy. As for the unconstrained 3D phase transitions, PCMs are subjected to nonisothermal and nonequilibrium shocks, giving rise to composition variation upon extensive programming caused by long-range element diffusions along the electrical current direction (Padilla et al., 2010; Xie et al.,

2018). Eventually, the device fails due to severe phase segregation and big void formation near the bottom electrode. The SLL-based device underwent repeated 3D phase transitions up to ~10^7 cycles until SET sticking failure took place (**Figure 4A**). We note similar SLL-based device also presented an approximate endurance (Shen et al., 2019), but showing considerably larger resistance fluctuations in both RESET and SET states than those of the SLL-based device studied in this work. The relative standard deviations (RSDs) of RESET and SET states of the SLL-based device are 6.3% (**Figure 4C**) and 2.2% (**Figure 4D**), respectively.

The reliable 2D phase transitions of the PCH architecture inhibit the large-scale atomic diffusion along the electrical pulse direction, which effectively prolongs the endurance of the PCH-based device to ~10^8 cycles without reaching failure (**Figure 4B**). Note that the fast speed (~10 ns) and long-life features of the PCH-based device offer a feasible route to develop DRAM-like phase-change working memory technology. Most importantly, the PCH-based device has pretty low resistance fluctuation in both RESET and SET states, with the RSDs being 0.9% (**Figure 4C**) and 0.7% (**Figure 4D**), respectively. The ultralow programming noise of the PCH-based device should also be ascribed to the reliable 2D switching manner of the confined Sb_2Te_3 sublayers, because the randomness of phase transitions (the stochastic crystallization in particular) (Rao et al., 2017) is markedly reduced, leading to more consistent resistance contrast and hence better-defined logic states.

CONCLUSION

In summary, we have demonstrated that, through fine-tuning the deposition techniques, the multilayer $Sb_2Te_3/TiTe_2$ stackings can be made into *c*-axis oriented heterostructure, which shall exhibit weakly coupled interactions among the Sb_2Te_3 and $TiTe_2$ building blocks. In contrast to the SLL structure that can only execute 3D phase transitions, the PCH architecture is able to perform reliable 2D switching of the confined Sb_2Te_3 sublayers. The long-range element migration during 3D phase transitions induces device failure after extensive cycling, which can be greatly inhibited in the 2D switching manner, leading to the remarkably extended cycling endurance of the PCH-based device as compared to the SLL-based one. This shall be conducive to the development of nonvolatile and long-life working memory to better renovate the classic von Neumann computing system. And above all, the PCH-based device presents rather smaller resistance fluctuations upon repeated programming than that of the SLL-based device. This low-noise feature is of necessity for the accomplishments of high-accuracy neuromorphic computing tasks.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

TL and KD fabricated the films and performed the XRD and SEM tests. KD and FR prepared the device samples and carried out electrical measurements. KD performed ab initio simulations. FR and KD wrote the paper with contributions from BC. All authors discussed the results and commented on the manuscript. The project was initiated and conceptualized by FR.

FUNDING

This work was supported by the National Natural Science Foundation of China (52032006, 62004130, 62004131), the Basic and Applied Basic Research Foundation of Guangdong (2020B1515120008), and the Science and Technology Foundation of Shenzhen (JCYJ20180507182248605, JCYJ20190808150605474).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnano.2021.649560/full#supplementary-material.

REFERENCES

Blöchl, P. E. (1994). Projector augmented-wave method. *Phys. Rev. B.* 50, 17953–17979. doi:10.1103/PhysRevB.50.17953

Boybat, I., Le Gallo, M., Nandakumar, S. R., Moraitis, T., Parnell, T., Tuma, T., et al. (2018). Neuromorphic computing with multi-memristive synapses. *Nat. Commun.* 9, 2514. doi:10.1038/s41467-018-04933-y

Burr, G. W., Shelby, R. M., Sebastian, A., Kim, S., Kim, S., Sidler, S., et al. (2017). Neuromorphic computing using non-volatile memory. *Adv. Phys.-X* 2, 89–124. doi:10.1080/23746149.2016.1259585

Chen, B., Chen, Y. M., Ding, K. Y., Li, K. L., Jiao, F. Y., Wang, L., et al. (2019). Kinetics features conducive to cache-type non-volatile phase-change memory. *Chem. Mater.* 31, 8794–8800. doi:10.1021/acs.chemmater.9b02598

Chia Tan, C., Shi, L. P., Zhao, R., Guo, Q., Li, Y., Yang, Y., et al. (2013). Compositionally matched nitrogen-doped $Ge_2Sb_2Te_5/Ge_2Sb_2Te_5$ superlattice-like structures for phase change random access memory. *Appl. Phys. Lett.* 103, 133507. doi:10.1063/1.4823551

Choe, J. (2017). Intel 3D XPoint memory die removed from Intel Optane™ PCM (Phase Change Memory). Available at: http://www.techinsights.com/about-techinsights/overview/blog/intel-3D-xpoint-memory-die-removed-from-intel-optane-pcm/ (Accessed May 18, 2017).

Chong, T. C., Shi, L. P., Zhao, R., Tan, P. K., Li, J. M., Lee, H. K., et al. (2006). Phase change random access memory cell with superlattice-like structure. *Appl. Phys. Lett.* 88, 122114. doi:10.1063/1.2181191

Ding, K. Y., Wang, J. J., Zhou, Y. X., Tian, H., Lu, L., Mazzarello, R., et al. (2019). Phase-change heterostructure enables ultralow noise and drift for memory operation. *Science* 366, 210–215. doi:10.1126/science.aay0291

Ielmini, D., and Wong, H.-S. P. (2018). In-memory computing with resistive switching devices. *Nat. Electron.* 1, 333–343. doi:10.1038/s41928-018-0092-2

Kestor, G., Gioiosa, R., Kerbyson, D. J., and Hoisie, A. (2013). "Quantifying the energy cost of data movement in scientific applications," in IEEE International

Symposium on Workload Characterization (IISWC), Portland, OR (Newyork, NY: IEEE), 55–65. doi:10.1109/IISWC.2013.6704670

Kresse, G., and Hafner, J. (1993). Ab initio molecular dynamics for liquid metals. *Phys. Rev. B.* 47, 558–561. doi:10.1103/PhysRevB.47.558

Kuzum, D., Jeyasingh, R. G. D., Lee, B., and Wong, H.-S. P. (2012). Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing. *Nano Lett.* 12, 2179–2186. doi:10.1021/nl201040y

Lam, C. H. (2010). "Storage class memory," in 10th IEEE International Conference Solid-State Integrated Circuit Technology, Shanghai, China, November 1–November 4, 2010 (Newyork, NY: IEEE), 1080–1083. doi:10.1109/ICSICT.2010.5667551

Le Gallo, M., Sebastian, A., Mathis, R., Manica, M., Giefers, H., Tuma, T., et al. (2018). Mixed-precision in-memory computing. *Nat. Electron.* 1, 246–253. doi:10.1038/s41928-018-0054-8

Li, X.-B., Chen, N.-K., Wang, X.-P., and Sun, H.-B. (2018). Phase-change superlattice materials toward low power consumption and high density data storage: microscopic picture, working principles, and optimization. *Adv. Funct. Mater.* 28, 1803380. doi:10.1002/adfm.201803380

Lu, Y. G., Song, S. N., Song, Z. T., Ren, W. C., Xiong, Y. L., Rao, F., et al. (2012). Superlattice-like $GaSb/Sb_2Te_3$ films for low-power phase change memory. *Scripta Mater.* 66, 702–705. doi:10.1016/j.scriptamat.2012.01.013

Padilla, A., Burr, G. W., Virwani, K., Debunne, A., Rettner, C. T., Topuria, T., et al. (2010). "Voltage polarity effects in GST-based phase change memory: physical origins and implications," in IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, December 6–December 8, 2010 (Newyork, NY: IEEE). 656–659. doi:10.1109/IEDM.2010.5703444

Rice, J. P., Burke, K., and Ernzerhof, M. (1996). Generalized gradient approximation made simple. *Phys. Rev. Lett.* 77, 3865–3868. doi:10.1103/PhysRevLett.77.3865

Rao, F., Song, Z., Cheng, Y., Liu, X., Xia, M., Li, W., et al. (2015). Direct observation of titanium-centered octahedra in titanium–antimony–tellurium phase-change material. *Nat. Commun.* 6, 10040. doi:10.1038/ncomms10040

Rao, F., Ding, K. Y., Zhou, Y. X., Zheng, Y. H., Xia, M. J., Lv, S. L., et al. (2017). Reducing the stochasticity of crystal nucleation to enable subnanosecond memory writing. *Science* 358, 1423–1427. doi:10.1126/science.aao3212

Rao, F., Zhang, W., and Ma, E. (2019). Catching structural transitions in liquids. *Science* 364, 1032–1033. doi:10.1126/science.aax6333

Raoux, S., Wełnic, W., and Ielmini, D. (2010). Phase change materials and their application to nonvolatile memories. *Chem. Rev.* 110, 240–267. doi:10.1021/cr900040x

Saito, Y., Fons, P., Bolotov, L., Miyata, N., Kolobov, A. V., and Tominaga, J. (2016). A two-step process for growth of highly oriented Sb_2Te_3 using sputtering. *AIP Adv.* 6, 045220. doi:10.1063/1.4948536

Sebastian, A., Le Gallo, M., Burr, G. W., Kim, S., BrightSky, M., and Eleftheriou, E. (2018). Tutorial: brain-inspired computing using phase-change memory devices. *J. Appl. Phys.* 124, 111101. doi:10.1063/1.5042413

Shen, J., Lv, S., Chen, X., Li, T., Zhang, S., Song, Z., et al. (2019). Thermal barrier phase change memory. *ACS Appl. Mater. Inter.* 11, 5336–5343. doi:10.1021/acsami.8b18473

Simpson, R. E., Fons, P., Kolobov, A. V., Fukaya, T., Krbal, M., Yagi, T., et al. (2011). Interfacial phase-change memory. *Nat. Nanotechnol.* 6, 501–505. doi:10.1038/nnano.2011.96

Soeya, S., Shintani, T., Odaka, T., Kondou, R., and Tominaga, J. (2013). Ultra-low switching power, crystallographic analysis, and switching mechanism for SnXTe100−X/Sb2Te3 diluted superlattice system. *Appl. Phys. Lett.* 103, 053103. doi:10.1063/1.4817068

Takaura, N., Ohyanagi, T., Tai, M., Kinoshita, M., Akita, K., Morikawa, T., et al. (2014). "55-μA Ge_xTe_{1-x}/Sb_2Te_3 superlattice topological-switching random access memory (TRAM) and study of atomic arrangement in Ge-Te and Sb-Te structures," in IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, December 15–December 17, 2014 (Newyork, NY: IEEE). 685–688. doi:10.1109/IEDM.2014.7047132

Tuma, T., Pantazi, A., Le Gallo, M., Sebastian, A., and Eleftheriou, E. (2016). Stochastic phase-change neurons. *Nat. Nanotechnol.* 11, 693–699. doi:10.1038/nnano.2016.70

Wong, H.-S. P., and Salahuddin, S. (2015). Memory leads the way to better computing. *Nat. Nanotechnol.* 10, 191–194. doi:10.1038/nnano.2015.29

Wuttig, M. (2005). Phase-change materials: towards a universal memory? *Nat. Mater.* 4, 265–266. doi:10.1038/nmat1359

Xie, Y., Kim, W., Kim, Y., Kim, S., Gonsalves, J., BrightSky, M., et al. (2018). Self-healing of a confined phase change memory device with a metallic surfactant layer. *Adv. Mater.* 30, 1705587. doi:10.1002/adma.201705587

Yu, S., and Chen, P. (2016). Emerging memory technologies: recent trends and prospects. *IEEE J. Solid-State Circuits Magazine* 8, 43–56. doi:10.1109/MSSC.2016.2546199

Zhou, X., Kalikka, J., Ji, X., Wu, L., Song, Z., and Simpson, R. E. (2016). Phase-change memory materials by design: a strain engineering approach. *Adv. Mater.* 28, 3007–3016. doi:10.1002/adma.201505865

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.

frontiers in Nanotechnology

ORIGINAL RESEARCH
published: 31 March 2021
doi: 10.3389/fnano.2021.660563

Memristors With Controllable Data Volatility by Loading Metal Ion-Added Ionic Liquids

Hiroshi Sato[1,2], *Hisashi Shima*[2]*, *Toshiki Nokami*[3], *Toshiyuki Itoh*[3], *Yusei Honma*[2], *Yasuhisa Naitoh*[2], *Hiroyuki Akinaga*[2] *and Kentaro Kinoshita*[1]

[1] Department of Applied Physics, Tokyo University of Science, Tokyo, Japan, [2] Device Technology Research Institute, National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan, [3] Center for Research on Green Sustainable Chemistry, Tottori University, Tottori, Japan

We demonstrate a new memristive device (IL-Memristor), in which an ionic liquid (IL) serve as a material to control the volatility of the resistance. ILs are ultra-low vapor pressure liquids consisting of cations and anions at room temperature, and their introduction into solid-state processes can provide new avenues in electronic device fabrication. Because the device resistance change in IL-Memristor is governed by a Cu filament formation/rupture in IL, we considered that the Cu filament stability affects the data retention characteristics. Therefore, we controlled the data retention time by clarifying the corrosion mechanism and performing the IL material design based on the results. It was found out that the corrosion of Cu filaments in the IL was ruled by the comproportionation reaction, and that the data retention characteristics of the devices varied depending on the valence of Cu ions added to the IL. Actually, IL-Memristors involving Cu(II) and Cu(I) show volatile and non-volatile nature with respect to the programmed resistance value, respectively. Our results showed that data volatility can be controlled through the metal ion species added to the IL. The present work indicates that IL-memristor is suitable for unique applications such as artificial neuron with tunable fading characteristics that is applicable to phenomena with a wide range of timescale.

Keywords: conductive bridge RAM, data retention characteristics, AI devices, fading memory, ionic liquids, reservoir devices

OPEN ACCESS

Edited by:
J. Joshua Yang,
University of Southern California, United States

Reviewed by:
Zhongrui Wang,
The University of Hong Kong, Hong Kong
Xiaobing Yan,
Hebei University, China

***Correspondence:**
Hisashi Shima
shima-hisashi@aist.go.jp

Specialty section:
This article was submitted to Nanodevices, a section of the journal Frontiers in Nanotechnology

Received: *29 January 2021*
Accepted: *05 March 2021*
Published: *31 March 2021*

Citation:
Sato H, Shima H, Nokami T, Itoh T, Honma Y, Naitoh Y, Akinaga H and Kinoshita K (2021) Memristors With Controllable Data Volatility by Loading Metal Ion-Added Ionic Liquids. Front. Nanotechnol. 3:660563. doi: 10.3389/fnano.2021.660563

INTRODUCTION

Memristors, which were proposed as the fourth fundamental elements of electric circuitry in 1971 (Chua, 1971), have been extensively investigated in memory and neuromorphic devices since the reports of the TiO_2 memristor in 2008 (Strukov et al., 2008; Yang et al., 2008). A memristor is a two-terminal passive device whose resistance changes with the amount of charge passing through it and is expected to advance electronics and electrochemical research (Sun B. et al., 2019). As the mechanism of resistance change in the memristor, not only the movement of oxygen vacancies (Sawa, 2008; Akinaga and Shima, 2010) but also the electrochemical metallization has been proposed (Gan et al., 2019). Here we propose an IL-Memristor in which ionic liquids (ILs) are introduced in a solid device as a new memristive material. ILs are ultra-low vapor pressure liquids comprised cations and anions at room temperature (Wasserscheid and Welton, 2002; Hallett and Welton, 2011) and can be used as chemical reaction field because of their wide potential window (Arimoto et al., 2008), making them suitable candidates for electronic devices

(Harada et al., 2015a,b, 2016; Saito and Iwasa, 2015; Kinoshita et al., 2017; Yamaoka et al., 2017). Because ILs are stable in vacuum, they can be incorporated into existing microfabrication processes. The stability of the metal in the IL impacts the data retention characteristics in electrochemical metallization type memristors. Expanding the range of device applications is possible by controlling the metal filament stability in IL-Memristor. The stable metal filament have been applied to a non-volatile memory, such as conducting-bridge RAM, which are expected to become next-generation memory devices because of their simple structure and low power consumption (Waser et al., 2009, 2016; Valov et al., 2011). Besides, the stable filament has been used to demonstrate electronics synapse devices, in which a long-term potentiation (LTP) and long-term depression (LTD) as well as spike-timing-dependent plasticity (STDP) in the biological synapse are successfully emulated (Jeong et al., 2016; Shi et al., 2018). Recently, the unstable filament also attracts considerable attention because the resultant temporal resistance change is applicable to the emerging devices for the human brain inspired computing (Hasegawa et al., 2011; Deng et al., 2015; Ascoli et al., 2016; Zhang et al., 2018; Midya et al., 2019; Wang et al., 2019; Zhu et al., 2020).

As an influencing factor on the Cu filament stability, we focused on the comproportionation reaction of Cu in IL reported by Murase et al. (2001). We directly detected the formation of Cu(I) from Cu(II) in IL dropped on a thin Cu film using X-ray photoelectron spectroscopy (XPS) by taking advantage of ultra-low vapor pressure of ILs in the vacuum. We used this reaction to control the data retention characteristics of the IL-Memristor. As expected from the XPS measurement results, the data retention time was more than 10 times longer in Cu(I)-doped IL-Memristor than that in Cu(II)-doped IL-Memristor although reproducible resistance change was observed in both devices. These results indicate that IL-Memristor with controllable data volatility can be produced through changing metal ion species in ILs.

MATERIALS AND METHODS

Device Fabrication

Figure 1A shows the SEM image of IL-Memristor. **Figure 1B** shows the cross-sectional view of IL-Memristor for lines A-B in **Figure 1A**. A Ta (1 nm)/Pt (20 nm)/Ta (1 nm) film was deposited on a SiO_2 substrate by sputtering, followed by chemical vapor deposition of SiO_2 (30 nm). Here, the Ta layer acts as the adhesion

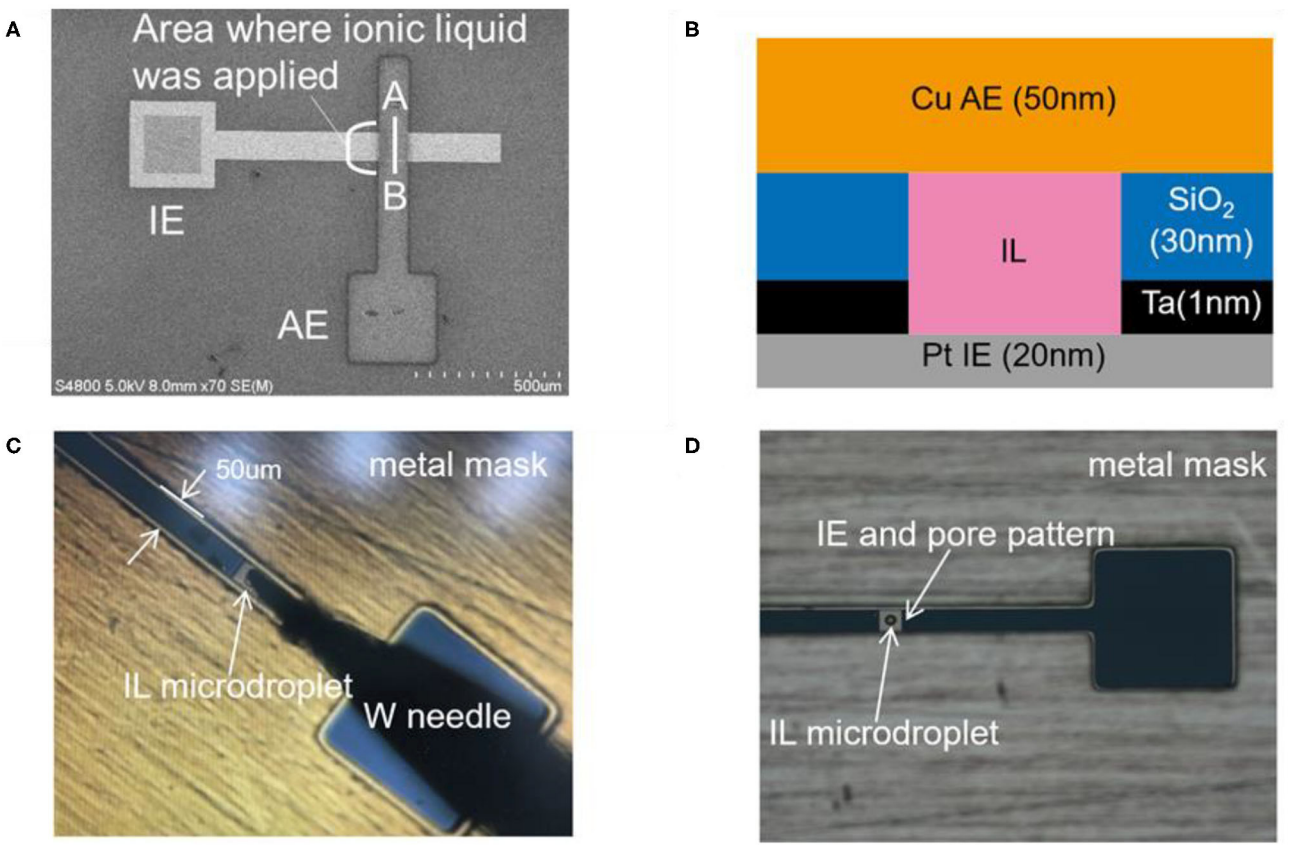


FIGURE 1 | (A) The SEM image of IL-Memristor and **(B)** the cross-sectional view of IL-Memristor for lines A-B in **(A)**. Photographs of **(C)** the W needle used to deliver IL microdroplet and **(D)** IL microdroplet on the IE and pore pattern. A metal mask for Cu AE pattering is also shown in **(C,D)**. IL is filled in a pore structure microfabricated in the 30 nm thick SiO_2 layer between Cu and Pt electrodes. In IL-Memristor, IL is used as a location for the Cu ion transport.

layer between the SiO_2 and Pt layers, and the Pt layer acts as the inert electrode (IE). A pore structure with an area of 1 × 1 μm, which determined the device size, was microfabricated in the SiO_2 layer by conventional photolithography and dry etching. As shown in **Figures 1C,D**, we used a W needle attached to a precision positioner to supply the IL microdroplet on the microfabricated IE and pore structure. The Cu active electrode (AE) patterns were prepared with the mask-through sputtering process using a metal mask. The thickness of Cu AE is 50 nm. The stacking structure of the present device is represented as Cu (50 nm)/IL (30 nm)/Pt (20 nm). We confirmed that the pore was successfully filled with IL from the results of the electrical measurement, which is explained in more detail in the **Supplementary Material**. 1-Butyl-3-methylimidazolium bis(trifluoromethyl sulfonyl)amide ([bmim][Tf_2N]) was used as the IL (Harada et al., 2015a). Cu(I) was introduced into this IL by electrolysis (Abedin et al., 2007; Qiu et al., 2010). The introduction of Cu(II) was conducted by dissolving $Cu(Tf_2N)_2$ metal salts. Hereafter, Cu(I)-doped IL and Cu(II)-doped IL are denoted as Cu(I)-IL and Cu(II)-IL, respectively. In addition to that, the device using Cu(I)-IL and Cu(II)-IL are represented as Cu(I)-IL-Memristor and Cu(II)-IL-Memristor, respectively.

Experimental Procedure

Optical microscopy of the Cu pattern in the IL was performed in the atmosphere at room temperature. In XPS measurements, ULVAC-PHI Quantera II applying a monochromatic Al Kα X-ray source (1486.6 eV) was used. The photoelectron take-off angle in XPS was 45°. Cu patterns deposited on SiO_2 substrates were introduced into the IL and observed by an optical microscope. After 1 h, the IL was sequentially washed away by acetone and ethanol, and the Cu pattern height was measured using a surface profiler. We measured current–voltage (I–V), data retention, and fading characteristics using a semiconductor parameter analyzer (Agilent B1500A). For the data retention characteristics, the reading voltage (−20 mV) was continuously applied during the measurement, and the current readings were taken at regular intervals.

RESULTS AND DISCUSSION

Optical Microscopy and XPS of the Cu Pattern in the IL

Shown in **Figure 2A** is the schematic illustration for the sample configuration to observe the Cu pattern dissolution process in IL. In this experiment, an IL droplet was dropped on the Cu patterns by using a micropipette. The corresponding photograph of the IL droplet and Cu patterns are depicted in **Figure 2B**. **Figures 2C–E** show the optical microscope images of the Cu pattern on SiO_2 immediately after, 15 min after, and 60 min after Cu(I)-IL dropping, respectively. **Figures 2F–H** show the optical microscope images of the Cu pattern on SiO_2 collected immediately after, 15 min after, and 60 min after Cu(II)-IL dropping, respectively. No change was observed in the shape of the Cu pattern in the case of Cu(I)-IL, whereas the Cu pattern was dissolved from the outside after 15 min in the case of Cu(II)-IL. Finally, the Cu pattern in Cu(II)-IL disappeared after 60 min. Although such electronics materials dissolution was an undesired negative phenomenon in terms of the device reliability and long-term use in the conventional electronics, its active utilization is proposed in the emerging field called transient electronics (Cheng and Vepachedu, 2016; Fu et al., 2016). The dissolution of Cu was also confirmed from the variation in the Cu pattern height measured by the surface profiler. **Figures 3A,B** show the surface profiler measurement results for the Cu pattern in Cu(I)- and Cu(II)-IL, respectively. The red and blue horizontal lines in **Figures 2E,H** correspond to the scanned location by the surface profiler shown in **Figures 3A,B**, respectively. When Cu(I)-IL was supplied, the Cu height pattern was almost the same as that of the as-prepared state, even after 60 min. This result indicates that the Cu pattern was not corroded in Cu(I)-IL. Moreover, when Cu(II)-IL was supplied, the Cu pattern height decreased over time and became <10 nm after 60 min, indicating that the Cu pattern was dissolved in Cu(II)-IL. As reported by Murase et al. (2001), Cu dissolution in Cu(II)-IL occurs because of the comproportionation reaction. To confirm this reaction, the chemical states of Cu in Cu(II)-IL on the Cu thin film were analyzed via XPS.

Because the ionization of the Cu metal [Cu(0)] is the possible origin for Cu dissolution, we conducted XPS measurement on IL/Cu (**Figure 4A**) and IL/SiO_2 (**Figure 4B**) to identify the change in Cu valence state during the dissolution. The former was prepared by dropping IL onto the Cu-sputtered SiO_2/Si substrate. The thickness of the Cu thin film was 50 nm. The latter was prepared as the control in which IL was dropped directly on the SiO_2/Si substrate. The area of the IL droplet in each case was ∼5 mm in diameter, which is much larger than the detection area of the XPS measurement (100 μm in diameter). In addition, the IL droplet on the substrate is thick enough to be visible for the human eye (roughly several 100 μm) and it is much thicker than the detection depth of the present XPS measurement (<10 nm). Therefore, from the viewpoint of the size and thickness of the IL droplet, the Cu signal comes only from IL. In the present study, XPS measurements were started 1 h after dropping IL to ensure adequate time for Cu dissolution.

Figure 5A shows the Cu $2p_{3/2}$ spectra for IL/Cu and IL/SiO_2. Because $Cu(Tf_2N)_2$ dissolved in [bmim][Tf_2N], the XPS signal of Cu species was detected even in IL/SiO_2. The signal intensity of the Cu $2p_{3/2}$ spectrum for IL/Cu is much larger than that for IL/SiO_2, which can be attributed to the increase in Cu content in IL as a result of the Cu dissolution. The chemical bonding state of Cu in IL/Cu can be estimated from the main peak position of the Cu $2p_{3/2}$ spectrum. The peak positions of the Cu $2p_{3/2}$ spectra in both IL/Cu and IL/SiO_2 are close to those in $Cu(NO_3)_2$ (935.51 eV) and $CuSO_4$ (936.00 eV) (Moretti and Beck, 2019). According to the previous reports, electrons delocalize within the S–N–S structure in Tf_2N anion (Forsyth et al., 2002; Hapiot and Lagrost, 2008; Smith et al., 2018). Additionally, because oxygen is more electronegative than nitrogen and sulfur, it is expected that the electrons of Cu in the IL are shared by S, N, and O in Tf_2N anion. Thus, the larger intensity of the Cu $2p_{3/2}$ spectrum for IL/Cu proves that the number of Cu cations interacting with Tf_2N anions increased during the Cu dissolution in the IL. As shown in **Figures 5B,C**, the waveform analysis for the Cu $2p_{3/2}$

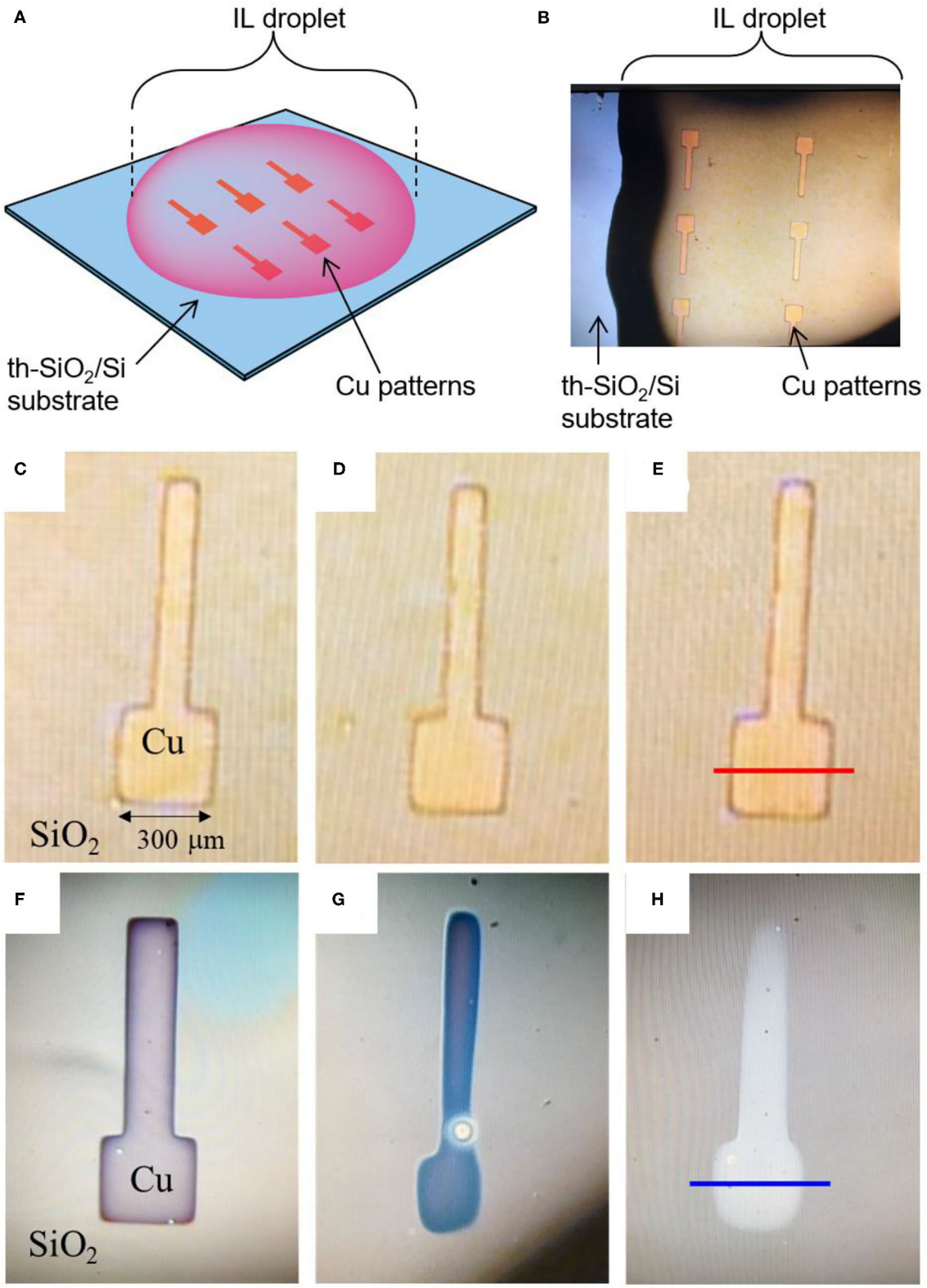


FIGURE 2 | (A) Schematic illustration and **(B)** photograph of Cu patterns immersed in the IL droplet. Magnified image of Cu thin film patterns **(C,F)** immediately after, **(D,G)** 15 min after, **(E,H)** 60 min after they were introduced into Cu(I)-IL and Cu(II)-IL, respectively. The thickness of Cu pattern is 50 nm. When Cu pattern is in Cu(I)-IL, the appearance of Cu pattern exhibited almost no change with time. On the other hands, Cu pattern gradually disappeared in Cu(II)-IL. The red and blue horizontal lines in **(E,H)** correspond to the scan position for the surface profiles in **Figures 3A,B**.

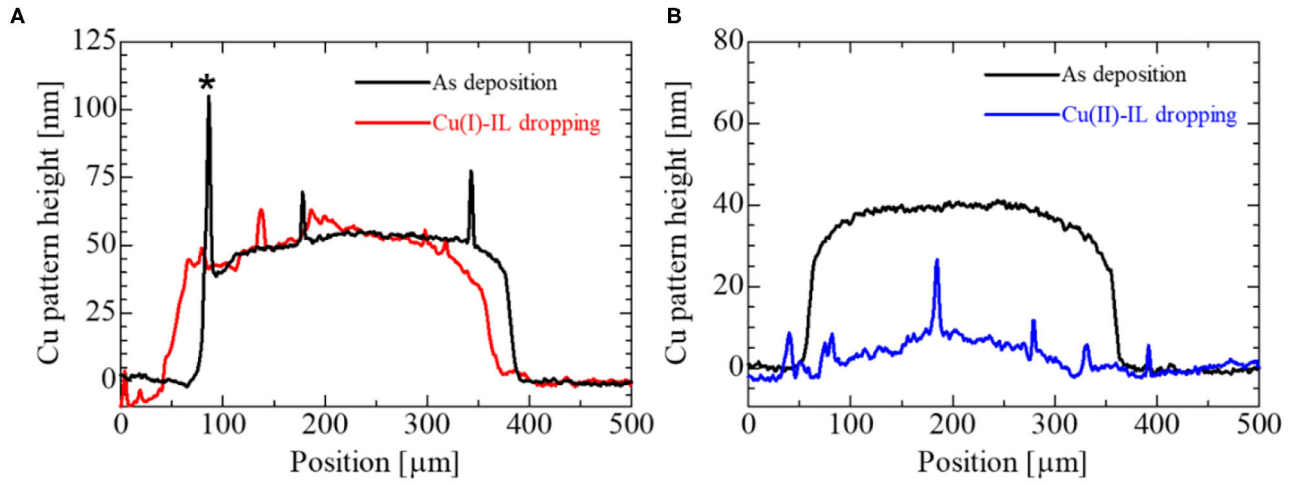


FIGURE 3 | The surface profiler results for **(A)** Cu(I)-IL and **(B)** Cu(II)-IL dropped onto the Cu pattern. The red curve in **(A)** is the surface profile measured along the red horizontal line in **Figure 2E**. The blue curve in **(B)** correspond to the surface profile along the blue horizontal line in **Figure 2H**. The black lines are the surface profiles before the Cu patterns are immersed in the IL. Note that the spiky peak shape at the edge of the Cu pattern plotted by the black line in **(A)** marked with an asterisk (*) is thought to be a side-wall fence structure formed during the mask-through sputtering process. The profiles for the Cu patterns dipped in IL were measured after removing IL by acetone and ethanol.

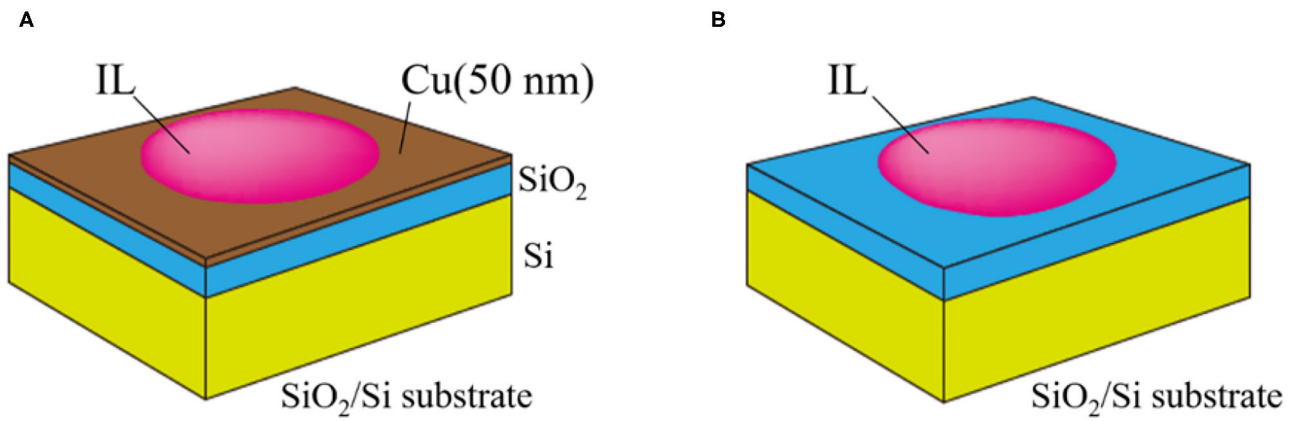


FIGURE 4 | Schematic illustrations of **(A)** IL/Cu and **(B)** IL/SiO_2 used for XPS measurements. The size of the substrate and IL drop (roughly 5 mm in diameter) is much larger than the detection area of the XPS instrument (100 μm in diameter).

XPS spectra was conducted in order to investigate the Cu valence state in the IL in more detail. Here, the main peak was labeled as Peak 1, and the satellite structure was split into two peaks labeled as Peak 2 and Peak 3. The peak position for Peak 1 in IL/Cu was 935.75 eV, whereas that in IL/SiO_2 was 936 eV. The lower binding energy value for Peak 1 in IL/Cu implied that not only Cu(II) but also Cu(I) was present in the IL on the Cu thin film. Although the constituent metal element is identical in the ionic material, the core-level binding energy becomes higher when the valence of the metal element (cation) increases because an increasing number of valence electrons is attracted by the neighboring anions. For instance, in the case of the Cu–O binary system (Cu_2O and CuO), the main peak position shifts to a higher binding energy when the valence of Cu is increased from Cu(I) in Cu_2O to Cu(II) in CuO (Moretti and Beck, 2019). By contrast, Peak 1 shifted to a lower binding energy in IL/Cu compared to that in IL/SiO_2, which can be related to the valence state decrease in Cu [i.e., the formation of Cu(I)]. Regarding the satellite structure in the Cu $2p_{3/2}$ spectra, a strong satellite structure was observed in both samples, indicating that Cu(II) was involved in the IL. This is reasonable because $Cu(Tf_2N)_2$ dissolved in [bmim][Tf_2N] has the divalent ions of Cu(II). However, the characteristics of the satellite structure of IL/Cu differ from that of IL/SiO_2. In the case of IL/SiO_2, the area ratio of Peak 3 to Peak 2, i.e., Peak 3/Peak 2, was ∼0.97, whereas the value of Peak 3/Peak 2 in IL/Cu was ∼1.1. The larger Peak 3/Peak 2 value in IL/Cu than that in IL/SiO_2 is

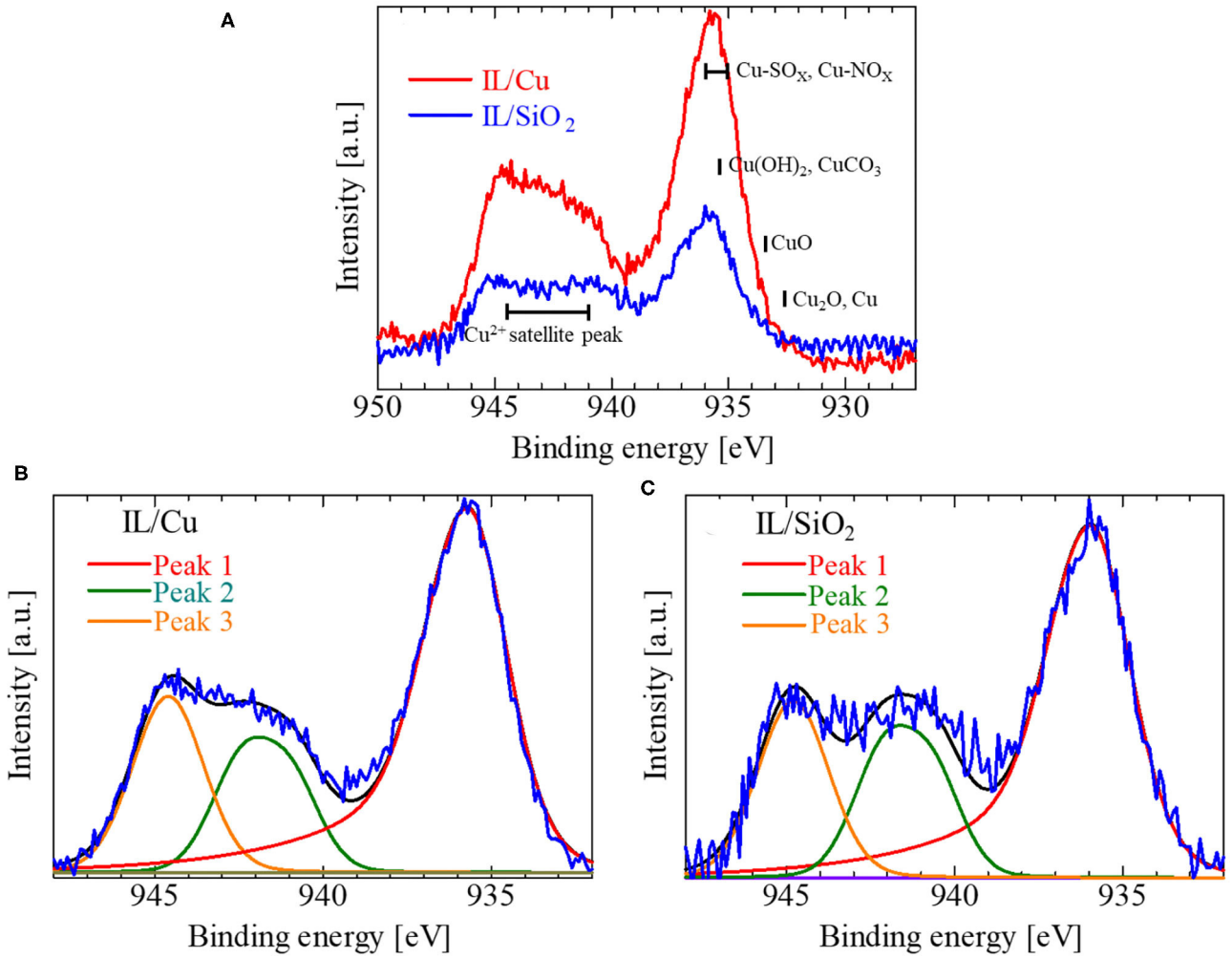


FIGURE 5 | (A) Cu $2p_{3/2}$ spectra for IL/Cu and IL/SiO_2 and the waveform analysis results for the Cu $2p_{3/2}$ spectrum in **(B)** IL/Cu and **(C)** IL/SiO_2. In **(A)**, the peak positions for some typical Cu compounds as well as Cu(II) satellite peaks are represented by the solid black bars. In **(B,C)**, red, green, and orange curves are the fitting results for the peak separation. Black curves are the summation of those curves.

due to the Cu(I) formation, which may be accompanied by the weak satellite structure in Cu $2p_{3/2}$ spectrum, as observed for Cu(I) in Cu_2O (Barreca et al., 2007; Wang et al., 2007).

Considering the above Cu $2p_{3/2}$ XPS spectra measurement results, the possible chemical reaction for Cu dissolution is

$$Cu(Tf_2N)_2 + Cu \rightarrow 2Cu(Tf_2N). \quad (1)$$

From the viewpoint of the Cu valence state, Cu dissolution is induced by the following comproportionation reaction:

$$Cu(II) + Cu \rightarrow 2Cu(I). \quad (2)$$

Two other signs of Cu(I) formation in IL/Cu were obtained. One was the Cu LMM Auger electron spectrum (**Figure 6**). The intensity of the Cu LMM Auger electron spectrum in IL/Cu is much larger than that in IL/SiO_2. A similar change in the Cu LMM Auger spectrum was observed when Cu(I) was introduced in the [MAP][Tf_2N] by the electrolysis of Cu (Qiu et al., 2010). Another was the shape of the N 1s XPS spectra in IL/Cu (**Figure 7A**) and IL/SiO_2 (**Figure 7B**). The N 1s XPS spectrum can be separated into two peaks (Peaks 4 and 6 in **Figure 7A**) for IL/Cu, whereas it can be separated into three peaks (Peaks 4–6 in **Figure 7B**) for IL/SiO_2. Peaks 4, 5, and 6 correspond to N in bmim cation, N in Tf_2N anion having the interaction with metal cations, and N in free Tf_2N anion (Caporali et al., 2016). Peak 5 in IL/Cu disappeared in Ag(I) containing [bmim][Tf_2N], whereas it was observed in [bmim][Tf_2N] with divalent metal cations, such as Cu(II), Ni(II), and Zn(II) (Caporali et al., 2016). As observed in the Cu $2p_{3/2}$ spectra, the binding energy between Cu cation and Tf_2N anion was weakened because of the formation of Cu(I), which may have caused the disappearance of Peak 5 in **Figure 7A**. Such feature in N 1s XPS spectrum, i.e., the disappearance of Peak 5, was also observed in Cu(I)-IL prepared by the electrolysis (see **Supplementary Material**). Regarding Cu(I) in the present IL, there is a possibility that Cu(I) partly forms the carbene complex with imidazolium cation according to the previous studies on

the N-heterocyclic carbenes coordinated to metals (Hapiot and Lagrost, 2008; Hopkinson et al., 2014).

Memory Operation

The operating mechanism of the IL-Memristor is as follows: when a voltage is applied to the AE, metal ions dissolve in the IL and deposit on the IE to form filaments, resulting in a low resistance state (LRS). Afterward, by applying a negative voltage to the AE in LRS, filaments are ruptured, resulting in a high resistance state (HRS). The switching from HRS to LRS is called SET, whereas that from LRS to HRS is called RESET. **Figure 8A** shows the *I*–*V* characteristics of the Cu(I)-IL-Memristor and Cu(II)-IL-Memristor, which were plotted using the median values calculated from 500 cycles. It should be noted that the number of DC sweep cycle in the present study for evaluating the statistical distribution in the operating voltages is comparable to or larger than those in the previous reports (Yan et al., 2017, 2019a). As indicated by the blue and red arrows in **Figure 8A**, the voltage values when the filament formation/rupture occurs are represented as V_{SET}/V_{RESET}, respectively. **Figure 8B** shows the cumulative probabilities of the operating voltage (V_{SET} and V_{RESET}) for the Cu(I)-IL-Memristor and Cu(II)-IL-Memristor. In **Figures 8A,B**, the blue circles show the Cu(II)-IL-Memristor, and the red triangles show the Cu(I)-IL-Memristor. The V_{SET} of the Cu(I)-IL-Memristor was lower than that of the Cu(II)-IL-Memristor. The comproportionation reaction affects each memristor differently: in the Cu(I)-IL-Memristor, Cu easily deposits on the IE, whereas in the Cu(II)-IL-Memristor, Cu easily dissolves from the AE. The reduction reaction (Cu deposition) on the IE is more dominant in affecting the SET process than the oxidation reaction (Cu dissolution) on the AE because the V_{SET} was lower in the Cu(I)-IL-Memristor than in the Cu(II)-IL-Memristor. This result was consistent with previous reports suggesting that the Helmholtz layer formed on the IE surface and the proton-accepting ability of ILs affect the SET process (Harada et al., 2016; Yamaoka et al., 2017). Additionally, the distribution of V_{RESET} is insusceptible to the Cu valence state in the IL. Assuming that the operating mechanism for the RESET process is mainly based on Joule heating (Tsuruoka et al., 2010; Sun et al., 2014), Cu valence insusceptibility of V_{RESET} may be because it masks the impact of corrosion by the comproportionation reaction. From **Figure 8B**, it is necessary to point out that the Cu(I)-IL-Memristors still have statistical variabilities in the values of V_{RESET} and V_{SET}. It is considered that such variabilities strongly affect the device reliabilities such

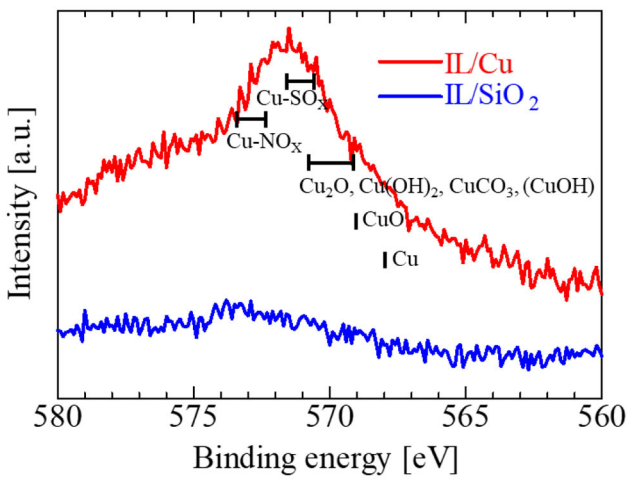


FIGURE 6 | Cu LMM Auger electron spectra for IL/Cu (red) and IL/SiO_2 (blue). The peak positions for some typical Cu compounds are depicted by black bars.

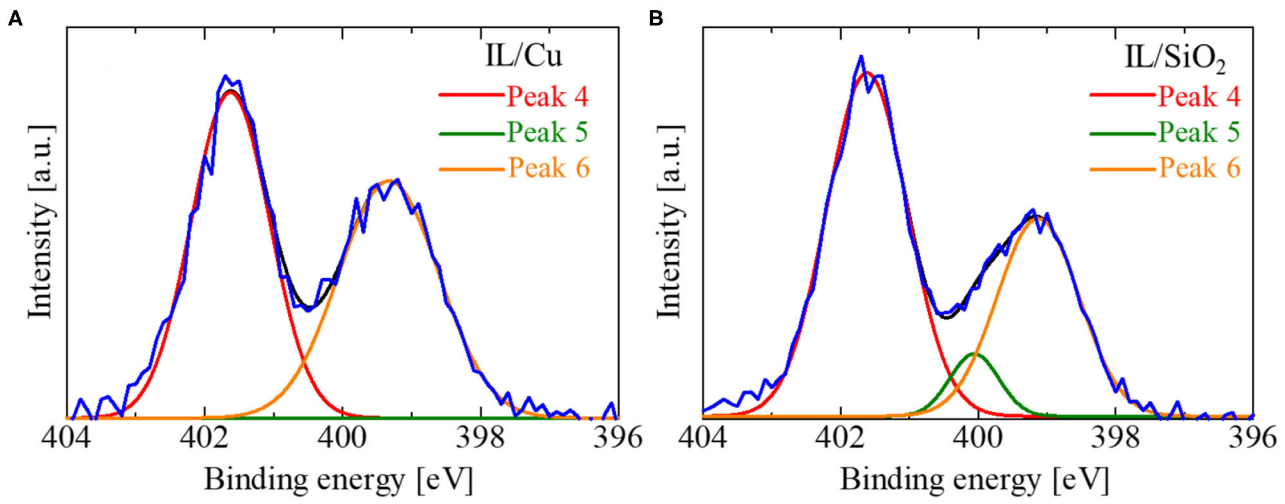


FIGURE 7 | Waveform analysis for N 1s XPS spectrum in **(A)** IL/Cu and **(B)** IL/SiO_2. Red, green, and orange curves are the fitting results for peak separation and black curves are the summation of those curves. For IL/Cu in **(A)**, the peak intensity for Peak 5 was zero.

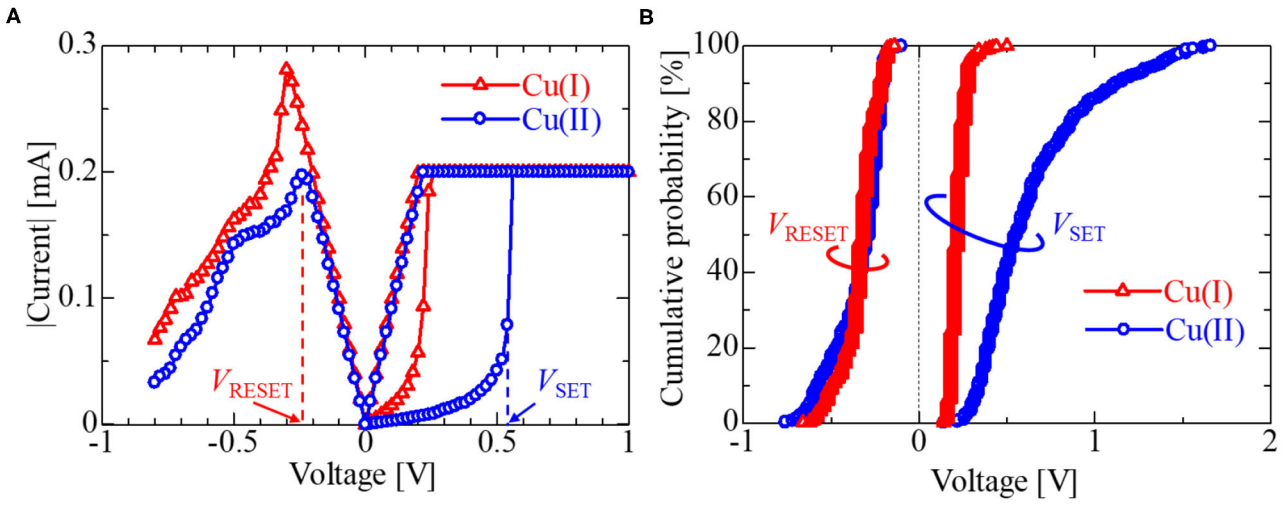


FIGURE 8 | (A) *I–V* characteristics and **(B)** cumulative probabilities of Cu(I)-IL-Memristor and Cu(II)-IL-Memristor. Red and blue plots, respectively corresponds to the data obtained from Cu(I)-IL-Memristor and Cu(II)-IL-Memristor. V_{SET} was defined as the voltage value just before the measured current reaches the current compliance value (I_{comp}). V_{RESET} was defined as the voltage at which the current value in the RESET process is maximized. The red and blue vertical dotted lines in **(A)** indicated the V_{RESET} and V_{SET} values for the case of Cu(II)-IL-Memristor shown as an example. Under the negative voltage, only the voltage scan from the origin to the negative direction was conducted in order to avoid the excessive voltage stress (For more detail, see **Supplementary Material**).

as cycle endurance characteristics because it results in the resistance switching failure during the cycle endurance test. Therefore, the suppression of the operating voltage variabilities through the development of materials, device structures, and fabrication processes for IL-Memristor is required. In terms of the device-to-device reproducibility, there is some device-to-device difference in the SET voltage distribution in **Figure 8B** (see **Supplementary Material**). However, the impact of the Cu valence on V_{SET} is qualitatively assured. One of the possible reasons for the device-to-device difference is that the volume of the IL microdroplet involved in each IL-Memristor is uncontrollable at present because it is transferred manually by using a W needle. The improvement of the device fabrication process such as adopting the ink-jet technology is required to confirm the device-to-device reproducibility.

Data Retention Characteristics

Figure 9 depicts the data retention characteristics of IL-Memristors. Data retention characteristics generally depend on the current level. To confirm the data retention characteristics at various current values, I_{comp} was set at 10, 20, and 200 μA. The Cu(II)-IL-Memristor showed a short data retention time ($<10^3$ s), whereas the Cu(I)-IL-Memristor showed a relatively long data retention time (more than 10^4 s). This indicates that the Cu filament was corroded by the comproportionation reaction in Cu(II)-IL, as expected from optical microscopy and XPS results of the Cu pattern in IL. Importantly, the volatility and non-volatility of the data retention characteristics in IL-Memristor can be controlled by changing metal ion species added to the IL. We found that the Cu(I)-IL-Memristor was more suitable for non-volatile memory applications compared with the Cu(II)-IL-Memristor. As shown in **Figure 9**, the resistance value

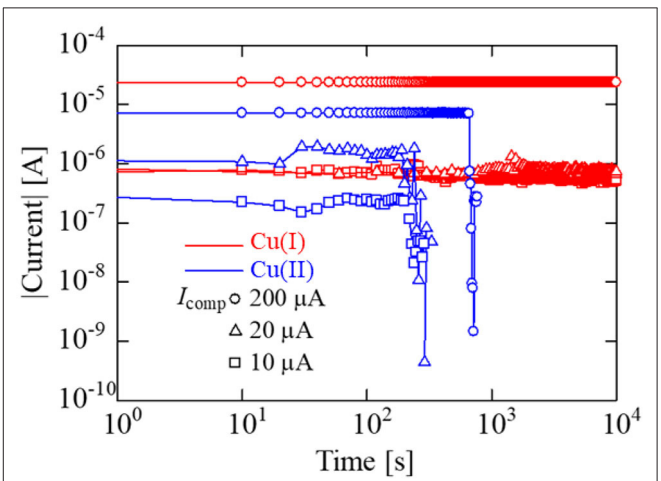


FIGURE 9 | Data retention characteristics of Cu(I)-IL-Memristor (red) and Cu(II)-IL-Memristor (blue). For both devices, I_{comp} was set at 10, 20, and 200 μA. The Cu(II)-IL-Memristor showed a short data retention time ($<10^3$ s), whereas the Cu(I)-IL-Memristor showed a relatively long data retention time (more than 10^4 s).

of LRS in Cu(I)-IL-Memristor can be increased by decreasing I_{comp}. Therefore, both the low current operation and good data retention are compatibly realized in Cu(I)-IL-Memristor. Although the minimum value of I_{comp} in the present study is 10 μA at present, the value of I_{comp} can be further decreased by decreasing the device area because the device resistance in HRS inversely scales with the device area. This is also favorable because the device size miniaturization is expected to promote the energy-conserving device operation.

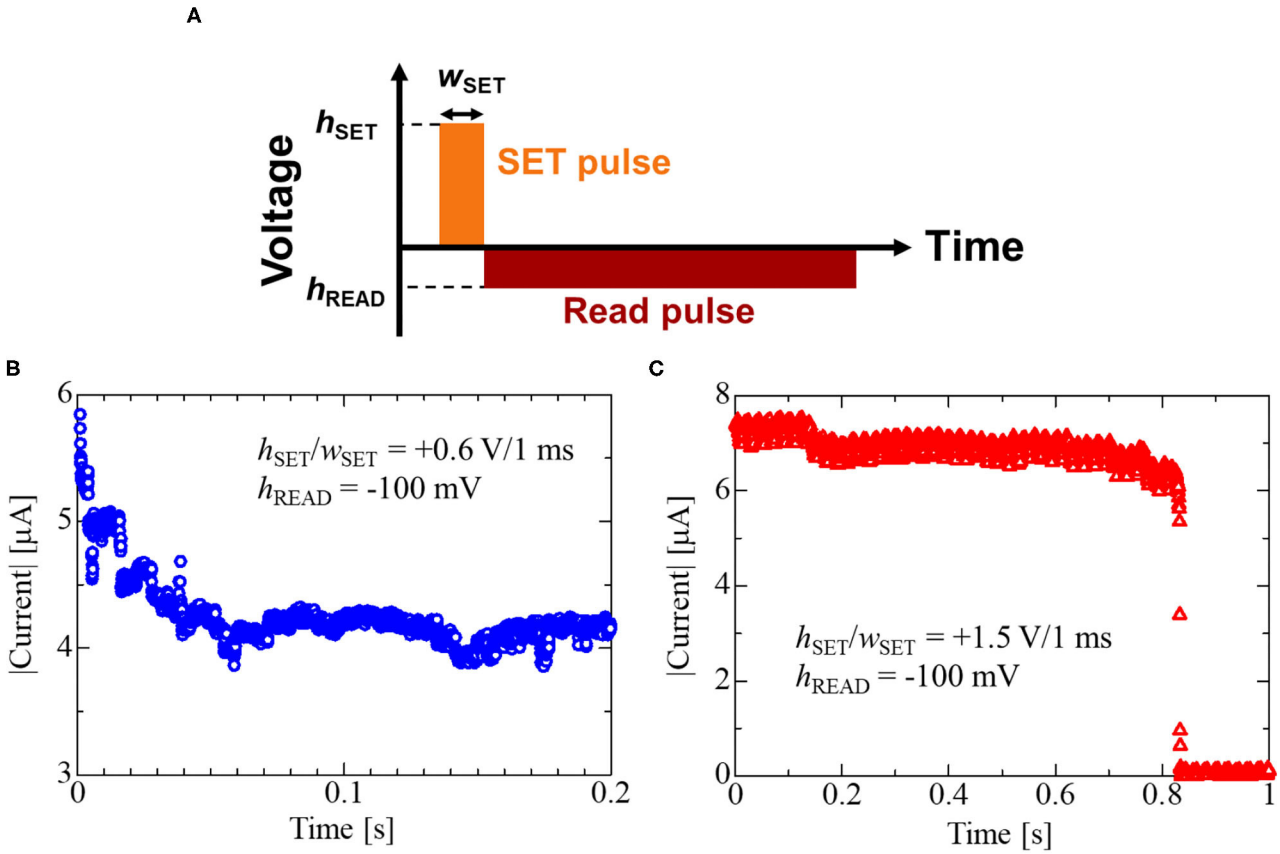


FIGURE 10 | (A) Measurement method for fading characteristics and the results of the fading characteristics of **(B)** the Cu(II)-IL-Memristor and **(C)** Cu(I)-IL-Memristor. In **(A)**, h_{SET} and w_{SET} denote the pulse voltage height and width for the SET pulse, while h_{READ} is the pulse voltage height for the read pulse. The values for h_{SET}, w_{SET}, and h_{READ} used in the present study are specified in **(B,C)**. The faster fading of the current value was observed in Cu(II)-IL-Memristor compared to Cu(I)-IL-Memristor.

Fading Characteristics

As shown in **Figure 9**, the Cu(II)-IL-Memristor demonstrated data volatility. Recently, such volatile memristors attract considerable attention because of the applicability to the human brain inspired computing. Various materials including solid oxide and sulfides exhibiting voltage resistance change are intensively investigated in order to realize artificial synaptic devices which mimic the information processing function of biological synapse (Hasegawa et al., 2012; Wang et al., 2016; Sun J. et al., 2019; Yan et al., 2019b). In addition, the memristors exhibiting such time-dependent resistance change are developed as the physical reservoir device for the reservoir computing (RC), in which the fading memory function is one of the essential requirements for the device (Tanaka et al., 2019). It has been pointed out that the timescale of the resistance change influences the time period between input signals to the reservoir and the information processing time (Midya et al., 2019). **Figure 10A** schematically illustrates the measurement method for the fading characteristics of the Cu(II)-IL-Memristor. At first, a SET pulse having a pulse height/width of h_{SET}/w_{SET} was applied to the IL-Memristor in HRS to slightly decrease the device resistance. Immediately after the application of the SET pulse, a read pulse with a height of h_{READ} was applied, and the current was monitored at regular intervals. For Cu(II)-IL-Memristor, h_{SET}/w_{SET} was +0.6 V/1 ms, while h_{READ} was −100 mV. Because the pulse voltage was used in this experiment, the SET process was incomplete although the pulse voltage height was larger than the median value of V_{SET} in **Figure 8**, which was evaluated by DC voltage sweep. **Figure 10B** shows the results of the fading characteristics of the Cu(II)-IL-Memristor, indicating a gradual current decrease. For comparison, fading characteristics of the Cu(I)-IL-Memristor was also evaluated (**Figure 10C**). For Cu(I)-IL-Memristor, h_{SET}/w_{SET} was +1.5 V/10 μs, while h_{READ} was −100 mV. During the reading process, about 45% increase of the resistance (from 15.6 to 22.8 kΩ was observed for 100 ms in Cu(II)-IL-Memristor. On the other hands, about 19% increase of the resistance (from 13.6 to 16.3 kΩ was observed for 800 ms in Cu(I)-IL-Memristor. The faster resistance change in Cu(II)-IL-Memristor suggests that the incomplete filament formation and subsequent dissolution by the comproportionation reaction in Cu(II)-IL lead to the time-dependent resistance change with a timescale of 100 ms. The timescale observed in Cu(II)-IL-Memristor is almost comparable to that observed in a solid diffusive memristor for RC (Midya et al., 2019). Therefore,

the present results indicate that the Cu(II)-IL-Memristor is suitable for AI applications such as RC because of their fading memory characteristics. Moreover, the present results also imply that controlling the time-dependent resistance change timescale depending on the task executed in the reservoir is expected by selecting the appropriate IL in IL-Memristor.

CONCLUSION

Cu corrosion in ILs was investigated by XPS. Based on this corrosion mechanism, we fabricated two types of devices, Cu(I)-IL-Memristor and Cu(II)-IL-Memristor, and evaluated their data retention characteristics. The Cu(II)-IL-Memristor showed data volatility because the comproportionation reaction promoted the corrosion of Cu filaments, whereas the Cu(I)-IL-Memristor was non-volatile. Our results show that data volatility and non-volatility can be easily controlled by changing metal ion species added to the IL. Additionally, the fading characteristics of the Cu(II)-IL-Memristor, where the current value gradually decreased over 100 ms, were confirmed. The Cu dissolution in IL and relevant time-dependent resistance change observed in IL-Memristor imply the adaptability of ILs and IL-Memristor to the emerging electronic and information processing technologies such as transient electronics and reservoir computing.

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

HSa, HSh, YN, HA, and KK conceived the experiments. HSh and YH prepared the devices. TN and TI synthesized the ionic liquids. HSa and HSh conducted the observations and measurements. HSa, HSh, HA, and KK analyzed the results. HSa, HSh, and KK wrote the manuscript. All authors contributed to the discussion about the results and commented on the manuscript.

FUNDING

A part of this work was financially supported by NAGASE & CO., LTD. New Value Creation Office.

ACKNOWLEDGMENTS

A part of this work was conducted at NIMS Nanofabrication Platform, supported by Nanotechnology Platform Program of the Ministry of Education, Culture, Sports, Science and Technology'(MEXT), Japan, Grant Number JPMXP09F20NM0072. We thank Enago (www.enago.jp) for the English language review.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnano.2021.660563/full#supplementary-material

REFERENCES

Abedin, S. Z. E., Saad, A. Y., Farag, H. K., Borisenko, N., Liu, Q. X., and Endres, F. (2007). Electrodeposition of selenium, indium and copper in an air- and water-stable ionic liquid at variable temperatures. *Electrochim. Acta* 52, 2746–2754. doi: 10.1016/j.electacta.2006.08.064

Akinaga, H., and Shima, H. (2010). Resistive Random Access Memory (ReRAM) based on metal oxides. *Proc. IEEE* 98, 2237–2251. doi: 10.1109/JPROC.2010.2070830

Arimoto, S., Kageyama, H., Torimoto, T., and Kuwabata, S. (2008). Development of *in situ* scanning electron microscope system for real time observation of metal deposition from ionic liquid. *Electrochem. Commun.* 10, 1901–1904. doi: 10.1016/j.elecom.2008.10.003

Ascoli, A., Tetzlaff, R., Chua, L. O., Strachan, J. P., and Williams, R. S. (2016). "Fading memory effects in a memristor for cellular nanoscale network applications," in *2016 Design, Automation, and Test in Europe Conference and Exhibition (DATE)* (Dresden), 421–425.

Barreca, D., Gasparotto, A., and Tondello, E. (2007). CVD Cu_2O and CuO nanosystems characterized by XPS. *Surf. Sci. Spectra* 14, 41–51. doi: 10.1116/11.20080701

Caporali, S., Pedio, M., Chiappe, C., Pomelli, C. S., Acres, R. G., and Bardi, U. (2016). Surface study of metal-containing ionic liquids by means of photoemission and absorption spectroscopies. *Surf. Sci.* 648, 360–365. doi: 10.1016/j.susc.2015.12.014

Cheng, H., and Vepachedu, V. (2016). Recent development of transient electronics. *Theor. Appl. Mech. Lett.* 6, 21–31. doi: 10.1016/j.taml.2015.11.012

Chua, L. (1971). Memristor-the missing circuit element. *IEEE Trans. Circuit Theory* 18, 507–519. doi: 10.1109/TCT.1971.1083337

Deng, L., Li, G., Deng, N., Wang, D., Zhang, Z., He, W., et al. (2015). Complex learning in bio-plausible memristive networks. *Sci. Rep.* 5:10684. doi: 10.1038/srep10684

Forsyth, C. M., MacFarlane, D. R., Golding, J. J., Huang, J., Sun, J., and Forsyth, M. (2002). Structural characterization of novel ionic materials incorporating the bis(trifluoromethanesulfonyl)amide anion. *Chem. Mater.* 14, 2103–2108. doi: 10.1021/cm0107777

Fu, K. K., Wang, Z., Dai, J., Carter, M., and Hu, L. (2016). Transient electronics: materials and devices. *Chem. Mater.* 28, 3527–3539. doi: 10.1021/acs.chemmater.5b04931

Gan, K. J., Chang, W. C., Liu, P. T., and Sze, S. M. (2019). Investigation of resistive switching in copper/InGaZnO/Al_2O_3-based memristor. *Appl. Phys. Lett.* 115:143501. doi: 10.1063/1.5116359

Hallett, J. P., and Welton, T. (2011). Room-temperature ionic liquids: solvents for synthesis and catalysis. 2. *Chem. Rev.* 111, 3508–3576. doi: 10.1021/cr1003248

Hapiot, P., and Lagrost, C. (2008). Electrochemical reactivity in room-temperature ionic liquids. *Chem. Rev.* 108, 2238–2264. doi: 10.1021/cr0680686

Harada, A., Yamaoka, H., Ogata, R., Watanabe, K., Kinoshita, K., Kishida, S., et al. (2015a). Enhanced stability of the HfO2 electrolyte and reduced working voltage of a CB-RAM by an ionic liquid. *J. Mater. Chem. C* 3, 6966–6969. doi: 10.1039/C5TC01127B

Harada, A., Yamaoka, H., Tojo, S., Watanabe, K., Sakaguchi, A., Kinoshita, K., et al. (2016). Improved performance of a conducting-bridge random access memory using ionic liquids. *J. Mater. Chem. C* 4, 7215–7222. doi: 10.1039/C6TC01486K

Harada, A., Yamaoka, H., Watanabe, K., Kinoshita, K., Kishida, S., Kishida, S., et al. (2015b). Copper ion-containing ionic liquids provide improved endurance and switching voltage distributions of conducting-bridge Random Access Memory. *Chem. Lett.* 44, 1578–1580. doi: 10.1246/cl.150773

Hasegawa, T., Itoh, Y., Tanaka, H., Hino, T., Tsuruoka, T., Terabe, K., et al. (2011). Volatile/nonvolatile dual-functional atom transistor. *Appl. Phys. Express* 4:015204. doi: 10.1143/APEX.4.015204

Hasegawa, T., Terabe, K., Tsuruoka, T., and Aono, M. (2012). Atomic switch: atom/ion movement controlled devices for beyond von-neumann computers. *Adv. Mater.* 24, 252–267. doi: 10.1002/adma.201102597

Hopkinson, M. N., Rishter, C., Schedler, M., and Glorius, F. (2014). An overview of N-heterocyclic carbenes. *Nature* 510, 485–496. doi: 10.1038/nature13384

Jeong, D. S., Kim, K. M., Kim, S., Choi, B. J., and Hwang, C. S. (2016). Memristors for energy-efficient new computing paradigms. *Adv. Electron. Mater.* 2:1600090. doi: 10.1002/aelm.201600090

Kinoshita, K., Sakaguchi, A., Harada, A., Yamaoka, H., Kishida, S., Fukaya, Y., et al. (2017). Improvement of switching endurance of conducting-bridge random access memory by addition of metal-ion-containing ionic liquid. *Jpn. J. Appl. Phys.* 56:04CE13. doi: 10.7567/JJAP.56.04CE13

Midya, R., Wang, Z., Asapu, S., Zhang, X., Rao, M., Song, W., et al. (2019). Reservoir computing using diffusive memristors. *Adv. Intell. Syst.* 1:1900084. doi: 10.1002/aisy.201900084

Moretti, G., and Beck, H. P. (2019). Relationship between the Auger parameter and the ground state valence charge of the core-ionized atom: the case of Cu(I) and Cu (II) compounds. *Surf. Interface Anal.* 51, 1359–1370. doi: 10.1002/sia.6704

Murase, K., Nitta, K., Hirato, T., and Awakura, Y. (2001). Electrochemical behaviour of copper in trimethyl-n-hexylammonium bis((trifluoromethyl)sulfonyl)amide, an ammonium imide-type room temperature molten salt. *J. Appl. Electrochem.* 31, 1089–1094. doi: 10.1023/A:1012255601793

Qiu, F., Taylor, A. W., Men, S., Villar-Garcia, I. J., and Licence, P. (2010). An ultra high vacuum-spectroelectrochemical study of the dissolution of copper in the ionic liquid (N-methilacetate)-4-picolinium bis(trifluoromethylsulfonyl)imide. *Phys. Chem. Chem. Phys.* 12, 1982–1990. doi: 10.1039/b924985k

Saito, Y., and Iwasa, Y. (2015). Ambipolar insulator-to-metal transition in black phosphorus by ionic-liquid gating. *ACS Nano* 9, 3192–3198. doi: 10.1021/acsnano.5b00497

Sawa, A. (2008). Resistive switching in transition metal oxides. *Mater. Today* 11, 28–36. doi: 10.1016/S1369-7021(08)70119-6

Shi, Y., Nguyen, L., Oh, S., Liu, X., Koushan, F., Jameson, J. R., et al. (2018). Neuroinspired unsupervised learning and pruning with subquantum CBRAM array. *Nat. Commun.* 9:5312. doi: 10.1038/s41467-018-07682-0

Smith, C. J., Gehrke, S., Hollóczk, O., Wagle, D. V., Heitz, M. P., and Baker, G. A. (2018). NMR relaxometric probing of ionic liquid dynamics and diffusion under mesoscopic confinement within bacterial cellulose ionogels. *J. Chem. Phys.* 148:193845. doi: 10.1063/1.5016337

Strukov, D. B., Snider, G. S., Stewart, D. R., and Williams, R. S. (2008). The missing memristor found. *Nature* 453, 80–83. doi: 10.1038/nature06932

Sun, B., Chen, Y., Xiao, M., Zhou,G., Ranjan, S., Hou, W., et al. (2019). A unified capacitive-coupled memristive model for the nonpinched current-voltage hysteresis loop. *Nano Lett.* 19, 6461–6465. doi: 10.1021/acs.nanolett.9b02683

Sun, J., Wand, H., Wang, Z., Song, F., Zhu, Q., Dang, B., et al. (2019). Physically transient memristive synapse with short-term plasticity based on magnesium oxide. *IEEE Electron Device Lett.* 40, 706–709. doi: 10.1109/LED.2019.2904752

Sun, P., Li, L., Lu, N., Li, Y., Wang, M., Xie, H., et al. (2014). Physical model of dynamic Joule heating effect for reset process in conductive-bridge random access memory. *J. Comput. Electron.* 13, 432–438. doi: 10.1007/s10825-013-0552-x

Tanaka, G., Yamane, T., Héroux, J. B., Nakane, R., Kanazawa, N., Takeda, S., et al. (2019). Recent advances in physical reservoir computing: a review. *Neural Netw.* 115, 100–123. doi: 10.1016/j.neunet.2019.03.005

Tsuruoka, T., Terabe, K., Hasegawa, T., and Aono, M. (2010). Forming and switching mechanisms of a cation-migration-based oxide resistive memory. *Nanotechnology* 21:425205. doi: 10.1088/0957-4484/21/42/425205

Valov, I., Waser, R., Jameson, J. R., and Kozicki, M. N. (2011). Electrochemical metallization memories-fundamentals, applications prospects. *Nanotechnology* 22:254003. doi: 10.1088/0957-4484/22/25/254003

Wang, Z., Jawhari, H. A., Al-, Nayak, P. K., Caraveo-Frescas, J. A., Wei, N., Hedhili, M. N., et al. (2007). Low temperature processed complementary metal oxide semiconductor (CMOS) device by oxidation effect from capping layer. *Sci. Rep.* 5:9617. doi: 10.1038/srep09617

Wang, Z., Joshi, S., Savel'ev, S. E., Jiang, H., Midya, R., Lin, P., et al. (2016). Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing. *Nat. Mater.* 16, 101–108. doi: 10.1038/nmat4756

Wang, Z., Li, C., Song, W., Rao, M., Belkin, D., Li, Y., et al. (2019). Reinforcement learning with analogue memristor arrays. *Nat. Electron.* 2, 115–124. doi: 10.1038/s41928-019-0221-6

Waser, R., Dittmann, R., Staikov, G., and Szot, K. (2009). Redox-based resistive switching memories nanoionic mechanisms, prospects, and challenges. *Adv. Mater.* 21, 2632–2663. doi: 10.1002/adma.200900375

Waser, R., Ielmini, D., Akinaga, H., Shima, H., Wong, H. S. P., Yang, J. J., et al. (2016). *Introduction to Nanoionic Elements for Information Technology, Resistive Switching: From Fundamentals of Nanoionic Redox Processes to Memristive Device Applications*. Hoboken, NJ: Wiley Online Library, 1–29.

Wasserscheid, P., and Welton, T. (2002). *Ionic Liquids in Synthesis*. Weinheim: Wiley-VCH, Verlag GmbH, 1.

Yamaoka, H., Yamashita, T., Harada, A., Sakaguchi, A., Kinoshita, K., Kishida, S., et al. (2017). Significantly improved performance of a conducting-bridge random access memory (CB-RAM) device using copper-containing Glyme Salt. *Chem. Lett.* 46, 1832–1835. doi: 10.1246/cl.170854

Yan, X., Pei, Y., Chen, H., Zhao, J., Zhou, Z., Wang, H., et al. (2019a). Self-assembled networked PbS distribution quantum dots for resistive switching and artificial synapse performance boost of memristors. *Adv. Mater.* 7:1805284. doi: 10.1002/adma.201805284

Yan, X., Zhang, L., Yang, Y., Zhou, Z., Zhao, J., Zhang, Y., et al. (2017). Highly improved performance in $Zr_{0.5}Hf_{0.5}O_2$ films inserted with graphene oxide quantum dots layer for resistive switching non-volatile memory. *J. Mater. Chem. C* 5, 11046–11052. doi: 10.1039/C7TC03037A

Yan, X., Zhao, Q., Chen, A. P., Zhao, J., Zhou, Z., Wang, J., et al. (2019b). Vacancy-induced synaptic behavior in 2d ws_2 nanosheet-based memristor for low-power neuromorphic computing. *Small* 15:1901423. doi: 10.1002/smll.201901423

Yang, J. J., Pickett, M. D., Li, X., Ohlberg, D. A. A., Stewart, D. R., and Williams, R. S. (2008). Memristive switching mechanism for metal/oxide/metal nanodevices. *Nat. Nanotech.* 3, 429–433. doi: 10.1038/nnano.2008.160

Zhang, Y., He, W., Wu, Y., Huang, K., Shen, Y., Su, J., et al. (2018). Highly compact artificial memristive neuron with low energy consumption. *Small* 14:1802188. doi: 10.1002/smll.201802188

Zhu, X., Wang, Q., and Lu, W. D. (2020). Memristor networks for real-time neural activity analysis. *Nat. Commun.* 11:2439. doi: 10.1038/s41467-020-16261-1

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.

frontiers in Nanotechnology

ORIGINAL RESEARCH
published: 15 April 2021
doi: 10.3389/fnano.2021.656026

Engineering Tunneling Selector to Achieve High Non-linearity for 1S1R Integration

Navnidhi K. Upadhyay[1,2], Thomas Blum[2], Petro Maksymovych[2], Nickolay V. Lavrik[2], Noraica Davila[3], Jordan A. Katine[3], A. V. Ievlev[2], Miaofang Chi[2], Qiangfei Xia[1] and J. Joshua Yang[1,4]**

[1] Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA, United States, [2] Oak Ridge National Laboratory, Center for Nanophase Materials Sciences, Oak Ridge, TN, United States, [3] Western Digital, San Jose, CA, United States, [4] Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, United States

Memristor devices have been extensively studied as one of the most promising technologies for next-generation non-volatile memory. However, for the memristor devices to have a real technological impact, they must be densely packed in a large crossbar array (CBA) exceeding Gigabytes in size. Devising a selector device that is CMOS compatible, 3D stackable, and has a high non-linearity (NL) and great endurance is a crucial enabling ingredient to reach this goal. Tunneling based selectors are very promising in these aspects, but the mediocre NL value limits their applications in large passive crossbar arrays. In this work, we demonstrated a trilayer tunneling selector based on the Ge/Pt/TaN_{1+x}/Ta_2O_5/TaN_{1+x}/Pd layers that could achieve a NL of 3×10^5, which is the highest NL achieved using a tunnel selector so far. The record-high tunneling NL is partially attributed to the bottom electrode's ultra-smoothness (BE) induced by a Ge/Pt layer. We further demonstrated the feasibility of 1S1R (1-selector 1-resistor) integration by vertically integrating a Pd/Ta_2O_5/Ru based memristor on top of the proposed selector.

Keywords: selectors, high non-linearity, vertically integrated 1S1R, crossbar arrays, memristor

OPEN ACCESS

Edited by:
Carlo Ricciardi, Politecnico di Torino, Italy

Reviewed by:
Vittorio Fra, Polytechnic University of Turin, Italy
Erika Covi, NaMLab gGmbH, Germany

****Correspondence:***
Navnidhi K. Upadhyay
nupadhyay@umass.edu
J. Joshua Yang
jjoshuay@usc.edu

Specialty section:
This article was submitted to Nanodevices, a section of the journal Frontiers in Nanotechnology

Received: *20 January 2021*
Accepted: *22 March 2021*
Published: *15 April 2021*

Citation:
Upadhyay NK, Blum T, Maksymovych P, Lavrik NV, Davila N, Katine JA, Ievlev AV, Chi M, Xia Q and Yang JJ (2021) Engineering Tunneling Selector to Achieve High Non-linearity for 1S1R Integration. Front. Nanotechnol. 3:656026. doi: 10.3389/fnano.2021.656026

INTRODUCTION

Originally CBA was proposed and adapted for telecommunication switching systems at the beginning of the twentieth century (Craft, 1925). A relay switch was placed at each crosspoint to automatically and efficiently route any permutation of its n input (e.g., rows) to its m output lines (e.g., column; Scudder and Reynolds, 1939). Given the CBA architecture's simplicity and extremely high-density capability, it has recently been adapted for memory applications (Kuekes et al., 2000; Kuekes and Williams, 2001; Chen et al., 2003). A memristor (also called ReRAM, resistive random access memory, when used for memory) has a simple two-terminal structure, which is highly desirable for CBA implementation (Xia and Yang, 2019). CBA makes it possible to achieve a device footprint of $4F^2$. An even higher memory density ($4F^2/n$; n: number of the stacked-layer) is achievable by 3D stacking the memristor devices (Lin et al., 2020). In light of these advantages, memristor-based CBA has emerged as one of the most promising technologies for high-density storage (Baek et al., 2005; Lee et al., 2009; Liu et al., 2014; Sills et al., 2014; Hudec et al., 2016) as well as memory-centric computing (Mouttet, 2008; Upadhyay et al., 2016, 2019; Rao et al., 2019; Lin et al., 2020; Wang et al., 2020). Memristor based CBAs can be used for solving

linear regression, logistic regression, linear equations, matrix eigenvectors, differential equations, neural networks, etc. in one computing cycle in principle, *in-situ* within the CBA by using physical law such as Ohm's law for multiplication and Kirchhoff's law for summation (Rao et al., 2019; Sun et al., 2019, 2020; Wang et al., 2020). CBAs enable a time-saving and energy-efficient approach to solving a wide range of practical problems in the era of big data nowadays.

On the other hand, a CBA suffers from the so-called sneak path current issue (Yang et al., 2013; Xia and Yang, 2019). Sneak path current could be suppressed if we can somehow make the current-voltage relation of the "ON" state of memristor non-linear (Joshua Yang et al., 2012). This could be achieved in two ways: (1) By engineering intrinsic NL into the memristor device (Xie et al., 2006; Choi et al., 2011; Joshua Yang et al., 2012), or (2) Introducing a non-linear device, so-called selector, in series with the memristor (Upadhyay et al., 2020) at each crosspoint. In the first approach combining the non-linear mechanism and switching characteristics in one structure makes it difficult to optimize both the memory and selector performance simultaneously and independently. So far, the maximum non-linearity demonstrated by such devices falls well short of the required NL for large array implementations. Connecting a selector in series with a memristor (the 2nd approach) gives freedom of optimizing memristor and selector independently. A transistor can also be used as a select element in the so-called 1T1R (1-transistor 1-ReRAM) array (Rao et al., 2019; Wang et al., 2020). However, the 3-terminal structure and large footprint of transistors are not ideal for CBAs. The high processing temperature of the transistors makes it almost impossible to be used in 3D stacked memories. So the best approach to solve the sneak path problem is to use two-terminal thin-film-based selector devices that can be scaled laterally and stacked vertically together with a memristor (Chen, 2015; Burr et al., 2016).

In **Supplementary Figure 1**, we presented a table (extension of the table from Xia and Yang, 2019) comparing the figure of merits of all types of selector devices proposed in the literature. The tunneling selector stands out as the most promising one because of the following reasons: (1) electroforming-free operation; (2) low cycle to cycle variation; (3) high endurance (theoretically infinite); (4) in-principle low-temperature dependence; and (5) high-speed operation. In one of our recent work, we have demonstrated that such a tunneling selector even withstands the memristor's electroforming operation in a vertically integrated 1S1R cell (Upadhyay et al., 2020). This paper proposes a trilayer tunneling selector with a stack structure of Ge/Pt/TaN_{1+x}/Ta_2O_5/TaN_{1+x}/Pd. Here TaN_{1+x}/Ta_2O_5/TaN_{1+x} layers form the trilayer tunneling barrier structure. Ge/Pt and Pt layers are BE and top electrode (TE), respectively. We engineer the Ge/Pt BE to provide an ultrasmooth surface on which a trilayer stack could be deposited. The root-mean-square (rms) roughness and peak-to-valley height distribution of the Ge/Pt layer were measured to be 185 and 700 pm, respectively.

The requirements of a smooth BE surface for the tunneling selector will be discussed in detail in the next section. Using the proposed trilayer tunneling device, we have shown a record NL of 3×10^5 and 10^7 for one-half and one-third biasing schemes, respectively. This is the highest NL among all tunneling based selectors reported so far. Furthermore, we integrated Pd/Ta_2O_5/Ru based memristor on top of the proposed selector device to realize a 1S1R cell. The 1S1R cell shows a maximum ON/OFF ratio of 100 and a NL of 10^4 and 10^6 for one-half and one-third biasing schemes, respectively. Again this is the highest NL demonstrated so far in any vertically integrated 1S1R cells to the best of our knowledge. The entire stack of the proposed 1S1R cell was deposited at room temperature, making it CMOS compatible and 3D stackable.

TUNNELING SELECTOR DESIGN

Even though tunneling selectors have many advantages over other types of selectors, their mediocre value of NL has been a shortcoming. The highest NL reported in tunneling selectors was 1.1×10^4, which we demonstrated in our previously reported work (Choi et al., 2016). We showed a trilayer tunneling barrier (TLTB) based selector could outperform the uniform barrier based selectors. As shown in **Figure 1A**, in the case of a uniform barrier device, the highest part of the barrier, closest to the electron source, is barely affected by the applied voltage. While for a "crested" barrier (**Figure 1B**) structure, the highest part of the barrier is in the middle and can be pulled down by the electric field quickly. Hence not only the barrier width but also the barrier height of such a crested barrier are much more sensitive to the applied electric field, resulting in a steeper current increase upon applied voltage and a higher non-linearity (Likharev, 1998; Jung and Cho, 2008). In other words, an applied voltage across the device only changes the barrier's width in the case of the regular rectangular barrier with a uniform barrier height. In contrast, both the barrier's width and height are reduced simultaneously upon applied voltages for the crested barrier. In practice, we can approximate a crested barrier structure with the staircase potential patterns formed in a trilayer structure, as shown in **Figure 1C**. We can achieve such a staircase energy barrier structure by sandwiching a dielectric layer with a small electron affinity between two other dielectric layers with larger electron affinities. This structure could be exploited for designing a high non-linearity selector. In this paper, we present a trilayer tunneling selector based on Ge/Pt/TaN_{1+x}/Ta_2O_5/TaN_{1+x}/Pd layers. The proposed TLTB (TaN_{1+x}/Ta_2O_5/TaN_{1+x}) layers form a staircase-like energy band structure, as shown in **Supplementary Figure 2**. This has been discussed in detail in our previous paper (Choi et al., 2016).

For designing a robust tunneling selector device, one needs to take care of two critical factors: (1) Depositing a high-quality dielectric layer that has minimal defects (e.g., Oxygen vacancies) and is stoichiometric and dense, which could be achieved by optimizing the deposition (sputtering in our case) recipe for the dielectric layers; (2) Having smooth surfaces and interfaces that can sustain a high electric field without breakdown. The roughness in the device stack could create hot spots due to the electric field concentration effect, resulting in an increase of the leakage current at relatively low voltage and low NL. With an increasing voltage, these hot spots can quickly

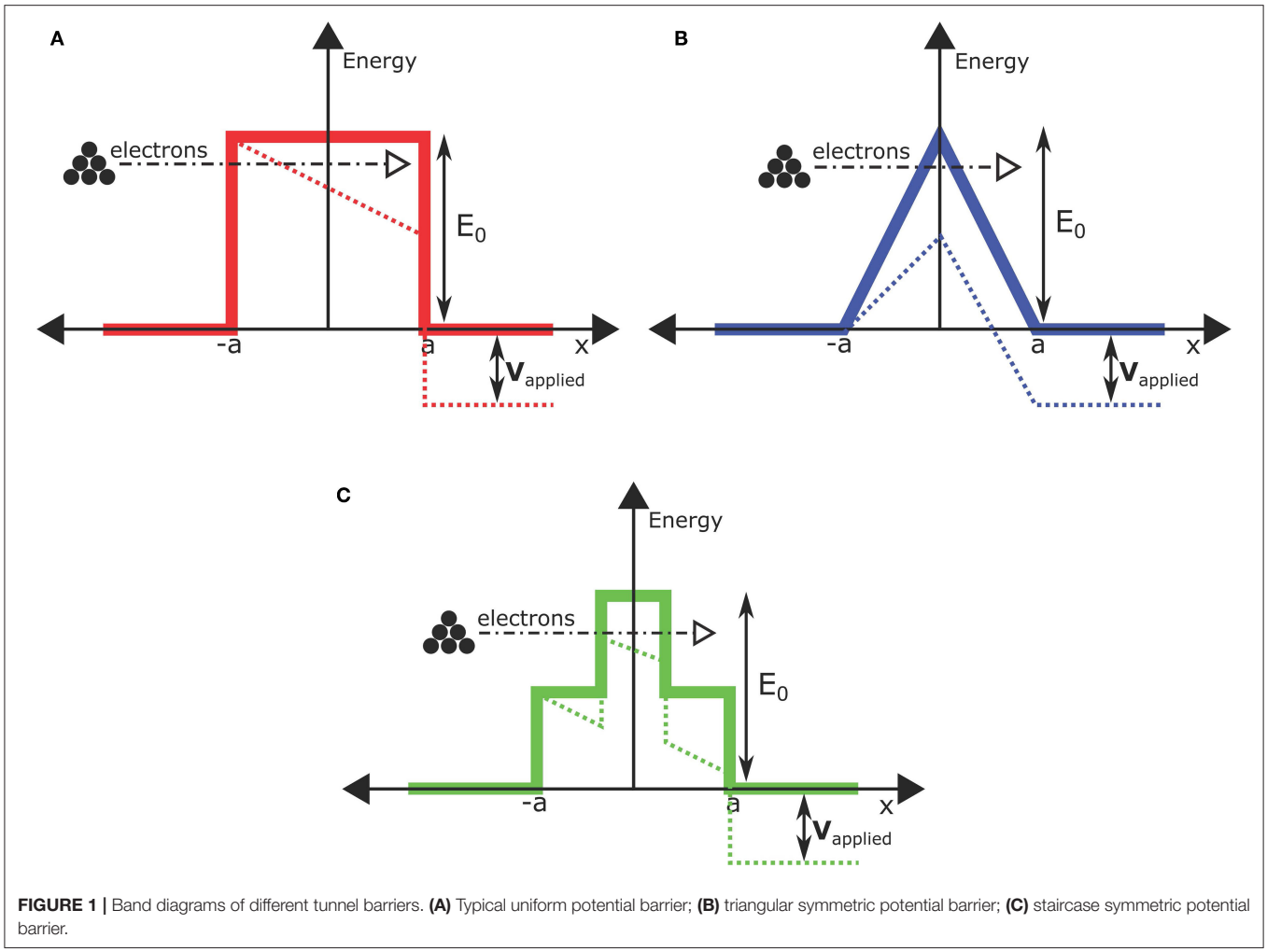


FIGURE 1 | Band diagrams of different tunnel barriers. **(A)** Typical uniform potential barrier; **(B)** triangular symmetric potential barrier; **(C)** staircase symmetric potential barrier.

cause a breakdown of the thin tunneling layers at relatively low voltages. To verify this hypothesis, we performed surface engineering for the device BE to obtain a much-smoothened surface. We found that depositing Pt on a thin Ge nucleation layer helps to achieve a smooth BE. For a comparative study, we deposited (evaporated) Ta (2 nm)/Pt (15 nm), Ti (2 nm)/Pt (15 nm), and Ge (2 nm)/Pt (15 nm) layers on a Si/SiO_2 substrate. **Figure 2** shows atomic force microscopy (AFM) topographic images of different surfaces. **Figure 2A** shows the rms (root mean square) roughness (R_q) of the substrate (Si/SiO_2) surface, which was measured to be 0.116 nm. **Figures 2B,C** present topographic images of the commonly used BE stacks Ta/Pt and Ti/Pt, with the R_q being 0.306 and 0.314 nm, respectively. **Figure 2D** presents the value of R_q for the Ge/Pt layer, which came out to be 0.185 nm. To ascertain uncertainties in the rms roughness measurement, we have done additional analysis presented in **Supplementary Figure 3**. The Ge/Pt layer not only has a significantly lower R_q but also a narrower peak-to-valley surface topological height distribution compared to those of Ta/Pt and Ti/Pt films, as shown in **Supplementary Figure 4**. The reason why a Ge nucleation layer provides smooth Pt film may be related to the activation energy of diffusion. If the activation energy of Pt diffusion on Ge is higher than those on the Ta and Ti surfaces, in that case, it could reduce the surface diffusion and mass transportation of Pt on the Ge nucleation layer, which could result in a smoother surface topology (Logeeswaran et al., 2009). Ge/Pt layer also provides good adhesion with the substrate but maybe not as great as Ti/Pt and Ta/Pt layers.

Next, we propose and demonstrate a highly non-linear TLTB selector device built upon the engineered smooth BE layers, as schematically shown in **Figure 3A**. Given that the BE smoothness is very critical to improving the performance of the TLTB selector device and the BE roughness will be affected by the surface roughness of the substrate itself, we took extensive measures to make sure the polished Si/SiO_2 (100 nm) substrate surface is clean and smooth. Starting with dipping the substrate in the Piranha solution [a mixture of H_2SO_4: H_2O_2 (3:1)] for 10 min, so residual organics were removed from the sample surface, followed by rinsing in the deionized water, followed by a blow-dry using compressed N_2 gun. Then to further smoothen out the substrate surface, $CHF_3 + O_2$ plasma cleaning (Turner and Chi, 2003; Ashraf et al., 2017) was performed in a Reactive

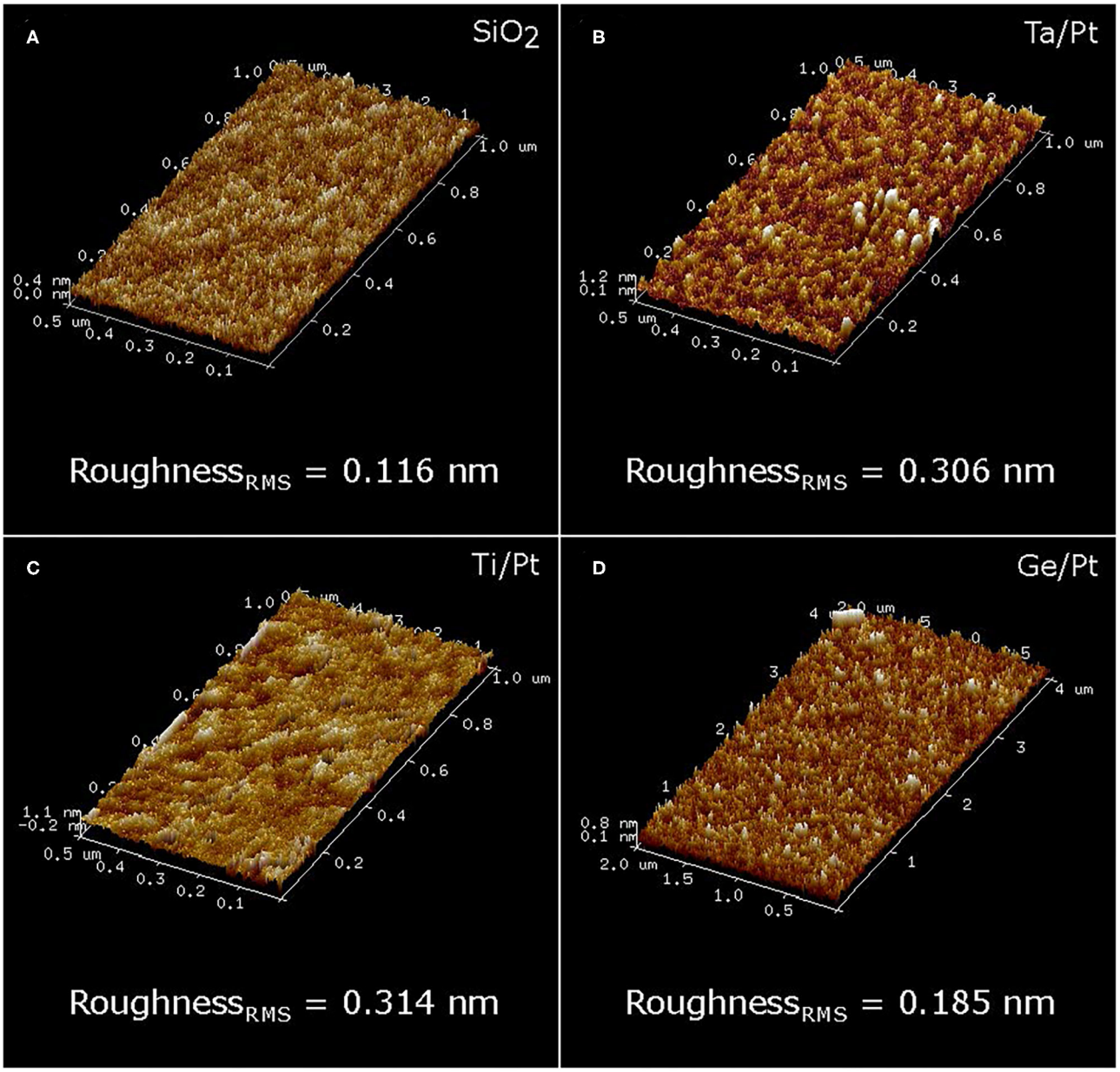


FIGURE 2 | Atomic force microscopy (AFM) topographic images of different surfaces. **(A)** SiO_2 (Substrate) surface. **(B)** Ta/Pt. **(C)** Ti/Pt. **(D)** Ge/Pt.

Ion Etching (RIE) chamber for 2 min. Finally, the substrate was cleaned with acetone in the ultrasonic bath for 10 min to remove any particle/contamination from the previous plasma cleaning step. We then rinsed it in IPA (Isopropanol) in the ultrasonic bath for 10 min to dissolve the acetone with the contaminant, followed by a blow-dry using compressed N_2. Immediately after finishing the substrate cleaning procedure, a photoresist (PR) was spin-coated on the substrate to maintain the surface cleanliness for the subsequent deposition of selector layers.

The PR coated substrate was then exposed by UV (ultraviolet) light through a mask for defining BE, then Ge (2 nm)/Pt (15 nm) layer was deposited using the e-beam evaporator followed by the standard lift-off process. Tri-layer tunneling stack [TaN_{1+x} (4 nm)/Ta_2O_5 (3 nm)/TaN_{1+x} (4 nm)] was deposited as a blanket layer without breaking the vacuum, using RF magnetron sputtering. The TaN_{1+x} layers were deposited using a ceramic TaN (99.99%) target in Ar + N_2 medium. The TaN target has a 1:1 ratio of Ta & N, which gives it metallic characteristics. Using Ar + N2 mixture as the deposition gas medium, the deposited TaN_{1+x} film's conductivity could be tuned based on the N_2 partial pressure (Yu et al., 2002). Similarly, an Ar + O_2 medium was used for Ta_2O_5 deposition using a ceramic Ta_2O_5 (99.99%) target. The

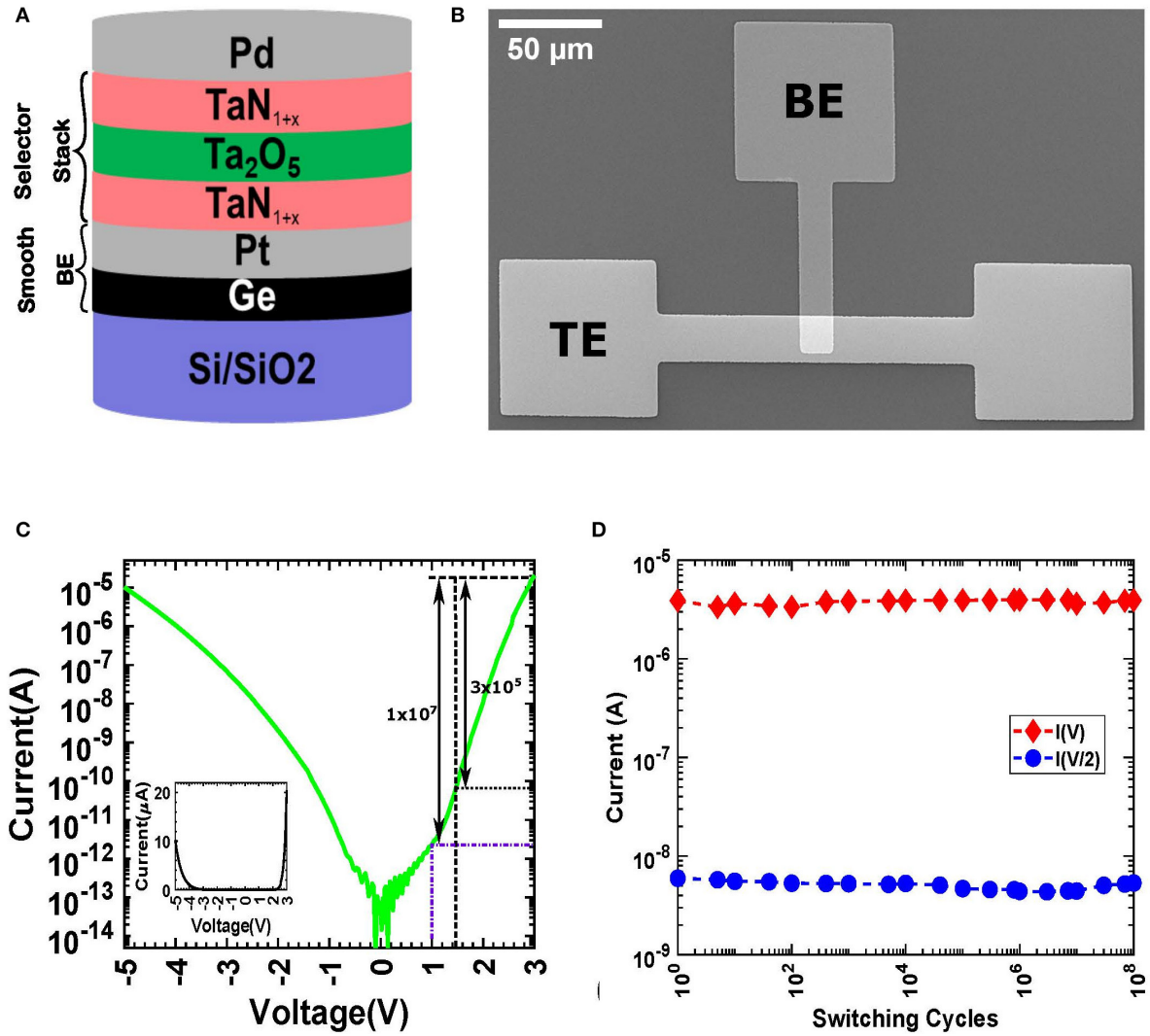


FIGURE 3 | Trilayer tunneling selector device performance. **(A)** Schematic of the proposed selector device. **(B)** Scanning electron microscopy image showing a top-view image of the fabricated device. **(C)** A typical non-linear I–V curve of the selector device. The inset shows a linear plot of the I–V curve. **(D)** Endurance data measured up to 100 million cycles.

presence of a low oxygen partial pressure during the sputtering improves the density and the stoichiometry of the sputtered Tantalum oxide film (Duggan et al., 1993). Finally, the standard photolithography process was used to define the TE, and Pd (40 nm) was deposited using DC sputtering, followed by the lift-off process. It is worth noting that the entire selector stack was deposited at room temperature, making this proposed device CMOS compatible. The SEM (scanning electron microscope) micrograph of the 20 μm × 15 μm crosspoint device is presented in **Figure 3B**.

Keysight B1500 device parameter analyzer was used to measure the I–V characteristics of the selector device. Bias was applied to the TE, and the BE was grounded. The I–V characteristics of the proposed TLTB selector device is plotted (semi-log) in **Figure 3C**, where the inset shows the linear plot of the same sweep cycles. The proposed TLTB selector device shows very insulating behavior under low bias (≈70 pA at +1.5 V). It becomes highly conductive at a high bias (≈20.4 μA at +3 V), which results in a highly non-linear I–V characteristic of the proposed selector device. For one-half-voltage scheme NL is defined as $NL_{1/2} = I(V_{read})/I(V_{read}/2)$ and for one-third biasing scheme NL could be given as $NL_{1/3} = I(V_{read})/I(V_{read}/3)$. The measured NL of the device is around 3×10^5 (10^7) for one-half (one-third biasing) schemes, as indicated in **Figure 3C**, which is the highest NL of any tunneling selector device reported so far. **Supplementary Figure 5** shows multiple I–V sweep of the selector device. A control sample with the same TLTB structure but on different BE layers, i.e., Ti/Pt, showed a NL of about 100, as shown in **Supplementary Figure 6**, which indicates the importance of using a smooth BE for tunneling

selector devices. Endurance measurement was conducted on the proposed selector device for 100 Million cycles without any noticeable degradation, as shown in **Figure 3D**, using 5 μs wide 3 V (V_{read}) and 1.5 V ($V_{read}/2$) pulses. It should be noted here that we used B1530A Waveform Generator/Fast Measurement Unit (WGFMU) for the endurance measurement. The noise floor of WGFMU unit is limited to a few nA of current, which resulted in a smaller NL (Compare to the actual NL of the selector device), observed during the endurance test.

To further demonstrate the effect of the proposed smooth BE (Ge/Pt) on tunneling selectors, we fabricated uniform barrier devices consisting of the TLTB stack's layers. **Supplementary Figure 7** shows the schematic of the devices and their electrical performance. **Supplementary Figures 7A,B** shows schematic of the Ge/Pt/Ta_2O_5/Pd and Ge/Pt/TaN_{1+x}/Pd devices and **Supplementary Figures 7C,D** shows their respective I–V characteristics. The NL of the single tunnel barrier selectors based on Ta_2O_5 and TaN_{1+x} layers was measured to be 2×10^3 and 1×10^3, respectively. It is worth noting that the NL demonstrated by these simple uniform barrier devices is an order of magnitude higher than the NL of previously reported similar single barrier devices and other tunneling selector devices (Kawahara et al., 2013; Woo et al., 2014). We attribute this improvement in NL to the use of a Ge/Pt based smooth BE layer.

VERTICALLY INTEGRATED 1S1R CELL

To demonstrate the feasibility of integrating the proposed selector device with a memristor, a vertically integrated 1S1R cell has been fabricated. We used a recently proposed Ru based memristor device (Yoon et al., 2020) for this demonstration. Ru based memristor exhibit forming free and low power switching operations, making it suitable for a 1S1R integration. **Figure 4A** presents a schematic of the vertically integrated 1S1R stack. A polished Si/SiO_2 (100 nm) substrate was cleaned and Ge (2 nm)/Pt (15 nm), TaN_{1+x} (4 nm)/Ta_2O_5 (3 nm)/TaN_{1+x} (4 nm) layers were deposited following the methods described in the previous section. Afterward, the Middle electrode (ME) was patterned using photolithography. A 40 nm thick Pd layer as the ME was sputter-deposited on top of the tri-layer tunneling stack, followed by the lift-off process. To isolate the selector layer from the memristor layer to be deposited on top of it, a 20 nm thick SiO_2 blanket isolation layer was deposited using sputtering. Then the SiO_2 isolation layer was patterned and etched away using the RIE ($CHF_3 + O_2$) to define the device region (see **Supplementary Figure 8** for details). A 10 nm thick blanket Ta_2O_5 switching layer was deposited using RF magnetron sputtering. Finally, the TE was patterned using the photolithography process, and a 40 nm thick Ru layer was sputter-deposited using the DC magnetron sputtering followed by the lift-off process (see the Experimental Section for more details). **Supplementary Figure 8** presents a wide-angle view of the cross-section of the 1S1R device. For the top memristor (1R) device, the Ru (TE) acts as an active electrode, while Pd (ME) serves as the inert electrode. It is worth noting that in this vertically integrated 1S1R cell, the selector was deliberately placed at the bottom to exploit the smooth BE (Ge/Pt) for achieving a high NL. We designed this testing structure in a way that the ME can be electrically accessed so that measurements can be made not only on the "1S1R" cell but also on the individual "S" and "R" to better understand the device stack.

For characterizing the Ru based memristor device, bias was applied to the TE of the integrated cell, the ME was grounded, and the BE was left floating. **Figure 4B** presents the I–V characteristics of the memristor device. To SET the device, a positive dual-sweep voltage (blue lines) was applied with current compliance (I_{cc}) set to 10 μA. Starting with a high resistance state (HRS), the device switched to a low resistance state (LRS) at 1.8 V and maintained its state during the reverse sweep. A negative dual-sweep voltage (red lines) was applied to RESET the device without any I_{cc}. Beginning with the LRS, device RESET started at about −0.4 V, and it switched to HRS at about −1.2 V and maintained its state afterward during the reverse sweep. The linear plot of the same I–V is presented in the inset. This switching mechanism is attributable to the Ru conductive path's formation/rupture in the switching layer. A metallic path in the

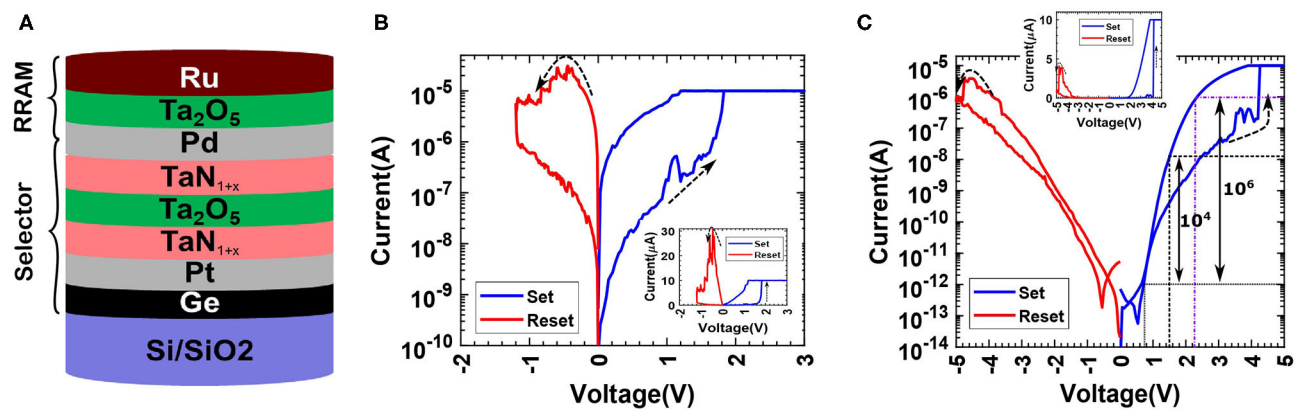


FIGURE 4 | Vertically integrated 1S1R cell. **(A)** Schematic of the 1S1R cell. **(B)** Typical I–V characteristics of the Pd/Ta_2O_5/Ru based memristor device. **(C)** Typical I–V characteristics of the 1S1R cell with highly non-linear ON state. The inset shows a linear plot of the same I–V.

LRS is corroborated also by the Ohmic behavior observed in the LRS (Yoon et al., 2020).

Next, the vertically integrated 1S1R cell was electrically tested to demonstrate successful 1S-1R operations. Bias was applied to the TE, and the BE was grounded with the ME left floating. The non-linear I–V characteristic of the device is presented in **Figure 4C**. The 1S1R cell exhibit a NL of 10^4 (10^6) for one-half (one-third) biasing scheme at $V_{read} = 1.5\,V$ with an ON/OFF ratio of around 50. The highest ON/OFF ratio of 100 could be achieved at $V_{read} = 3.4\,V$ but with a reduced NL value of 150. Interestingly there is a trade-off between high NL and maximum memory window size and can be leveraged depending on the use cases. It should be noted that adding a resistor (memristor in the 1S1R) reduces the NL of the selector by an amount depending on the relative resistance of the resistor and the selector. Nevertheless, the non-linearity exhibited in the proposed 1S1R cell is the highest among any vertically integrated 1S1R cell presented so far to the best of our knowledge. SEM micrograph showing the vertically integrated 1S1R device's top view is presented in **Figure 5a**.

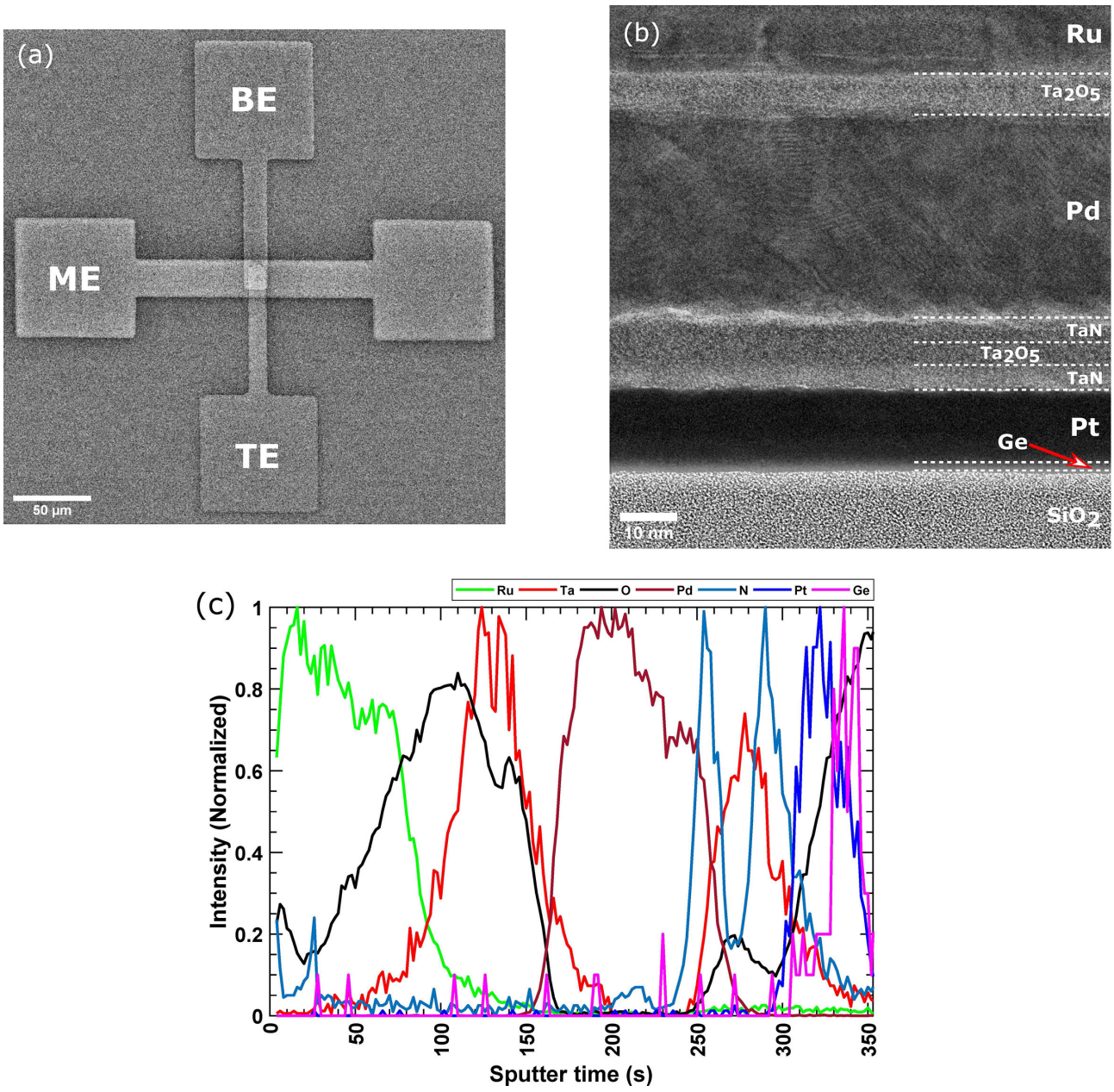


FIGURE 5 | Physical characterization of the vertically integrated 1S1R cell. **(a)** Scanning electron microscopy image of the 1S1R cell showing the top view of the device. **(b)** Cross-sectional STEM image of the 1S1R cell. **(c)** ToF-SIMS depth profile through the layers of the vertically integrated 1S1R cell.

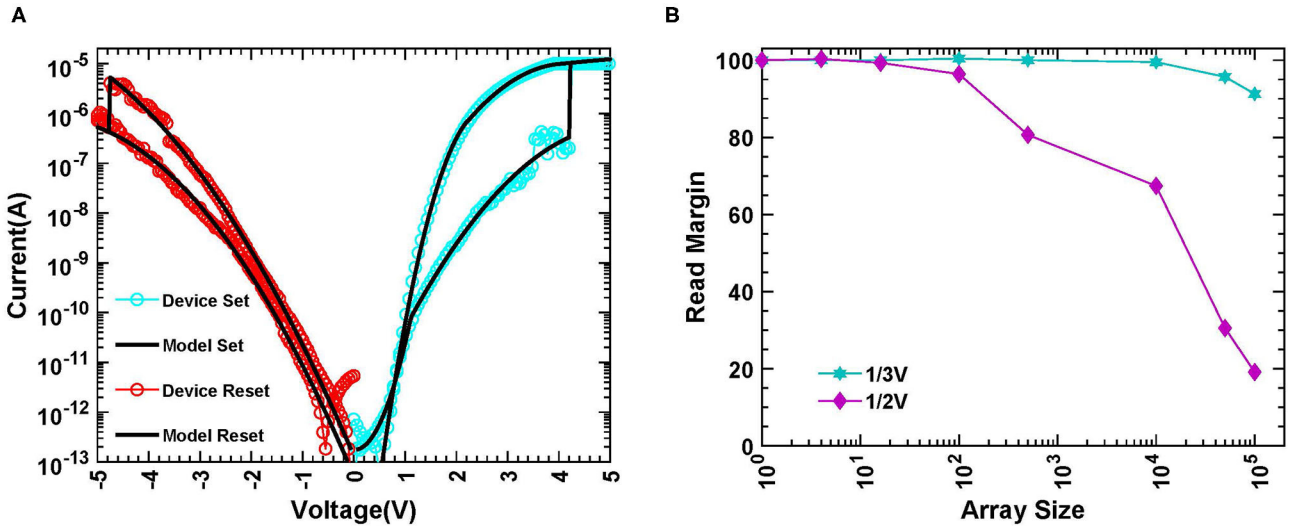


FIGURE 6 | Circuit level modeling of 1S1R based CBA. **(A)** SPICE simulation of single device IV sweep (solid line), **(B)** Normalize Readout Margin for different array sizes.

TE, ME, and BE are marked on the figure. Cross-section scanning transmission electron microscopy (STEM) image of the 1S1R cell ($Ge/Pt/TaN_{1+x}/Ta_2O_5/TaN_{1+x}/Pd/Ta_2O_5/Ru$) is presented in **Figure 5b**. STEM micrograph revealed that the TLTB layers were amorphous. Time-of-Flight Secondary Ion Mass Spectrometry (ToF-SIMS) was conducted at the active region of the 1S1R device. **Figure 5c** presents the depth profile through the vertically integrated 1S1R cell, identifying the stack's key elements.

To demonstrate the proposed selector capability, we did a circuit simulation of the 1S1R based CBA. The SPICE model was validated by comparing the experimental and simulated I–V characteristics of the single device, as shown in **Figure 6A**. The normalized readout margin for different sizes of the CBA is presented in **Figure 6B**. We considered the two most popular biasing schemes for our simulation: one-half and one-third voltage schemes. The readout margin for a 100 kbits CBA is around 20% for the one-half biasing scheme and more than 90% for the one-third biasing scheme. The one-third biasing scheme is more resilient to the sneak path current issue than the one-half biasing scheme. We can conclude from the results that the proposed selector helps mitigate the effect of the CBA's sneak path currents. It could potentially support a larger array size before the read margin hits the minimum criterion of 10% to differentiate the states (Lo et al., 2013). **Supplementary Figure 9** presents the readout resistance state of the selected device for various array sizes. HRS resistance decreases with array size because of an increase in sneak path current. Since readout current for LRS is larger than HRS, LRS current is only mildly affected by the sneak path current and so does the LRS resistance state.

CONCLUSION

In summary, we proposed and experimentally verified the critical role of layer smoothness and tunnel barrier shape in determining tunnel-based selectors' non-linearity. To prove the concepts, we developed a $Ge/Pt/TaN_{1+x}/Ta_2O_5/TaN_{1+x}/Pd$ based TLTB selector, which combined the benefit of a staircase potential barrier with the smooth BE. The proposed selector is CMOS compatible, 3D stackable, and exhibits a record NL value. We have engineered the BE layer (Ge/Pt) to make it an ultrasmooth surface. The measured rms roughness and peak-to-valley height distribution were 185 and 700 pm, respectively. This ultrasmooth BE surface and crested barrier lead to the demonstration of a record-high NL of 3×10^5. We further vertically integrated the proposed TLTB selector with a Ru based ($Pd/Ta_2O_5/Ru$) memristor device to demonstrate the feasibility of 1S1R integration and operation. The I–V characteristics recorded from this vertically integrated 1S1R cell show a maximum ON/OFF ratio of 100 and a NL of 10^4, also a record-high NL of any vertically integrated 1S1R cell ever reported. The excellent device NL performance suggests that our selector could be used to realize a large passive memristor array, which has remained elusive so far.

EXPERIMENTAL SECTION

Trilayer Selector Fabrication

A p-type (100) Si wafer with 100 nm thermal oxide was used as the substrate. The standard photolithography and lift-off process were used to define 15 μm wide Ge (2 nm)/Pt (15 nm) bottom electrode and 20 μm wide Pt (20 nm) top electrode using e-beam evaporation. A trilayer structure consists of TaN_{1+x} (4 nm)/Ta_2O_5 (3 nm)/TaN_{1+x} (4 nm) layers were sputter-deposited on top of the BE without breaking vacuum. TaN & Ta_2O_5 ceramic targets were used for RF magnetron sputtering in an Orion 8 (AJA international) sputtering system in the presence of an Ar-N_2 mixture (15:5) and Ar-O_2 mixture (20:1), respectively. Finally, 15 μm wide Pd (40 nm)

was sputter-deposited following standard photolithography and lift-off process.

Vertically Integrated 1S1R Device Fabrication

On a p-type (100) Silicon substrate with 100 nm thick thermal oxide, a 20 μm wide Ge (2 nm)/Pt (15 nm) line (BE) defined by photolithography and lift-off process, was fabricated. Then a trilayer structure consists of TaN_{1+x} (4 nm)/Ta_2O_5 (3 nm)/TaN_{1+x} (4 nm) layers were sputter-deposited on top of the BE without breaking vacuum. TaN & Ta_2O_5 ceramic targets were used for RF magnetron sputtering in an Orion 8 (AJA international) sputtering system in the presence of Ar-N_2 mixture and Ar-O_2 mixture, respectively. Afterward, a 15 μm wide middle electrode was patterned by photolithography, and a 40 nm thick Pd layer was deposited by RF magnetron sputtering followed by a lift-off process. A 20 nm thick SiO_2 isolation layer was sputter deposited on the top of the selector layer to isolate it (in the non-device-region) from the memristor layers to be deposited on top of it. The SiO_2 layer in the device region was patterned and etched away before a 10 nm thick blanket Ta_2O_5 switching layer was deposited using RF magnetron sputtering. Finally, 10 μm wide TE was patterned using photolithography, and then a Ru (40 nm) layer was deposited using sputtering followed by a lift-off process.

Device Characterization

The DC measurements were performed using a B1500A semiconductor parameter analyzer (Keysight), and B1530A (Keysight) was used for pulse measurement. All electrical measurements were performed by applying the bias to the TE and grounding the BE. The cross-sectional TEM study of the Ge/Pt/TaN_{1+x}/Ta_2O_5/TaN_{1+x}/Pd/Ta_2O_5/Ru device and the EDS element mapping was performed using JEOL NEOARM atomic-resolution STEM at an accelerating voltage of 200 kV. FEI Nova 200 Dual-Beam FIB was used to prepare the FIBed TEM lamella. ToF-SIMS measurements were done using a TOF.SIMS.5-NSC instrument, using a $Bi3^+$ ion gun (30 keV energy, 0.49 nA current) as the primary ion source and an $O2^-$ ion gun (1 keV energy, 120 nA current, 20 μm spot size) as the sputter source. ToF-SIMS measurements were performed in the non-interlaced mode, where every scan of chemical analysis with primary bismuth source was followed by sputtering using an $O2^-$ ion gun. A low energy electron flood gun was used for charge compensation between cycles. The vacuum level in the ToF-SIMS during the measurements ranged from 5 to 9×10^{-9} mbar.

SPICE Modeling

The device model was written using Verilog-A. Cadence Virtuoso was used for the circuit simulation. For all the simulations worst-case scenario has been considered. The selected device lies farthest to the row voltage source as well as farthest to the column current sense amplifier. All the unselected and half selected devices were switched to ON state before start of simulation.

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/s.

AUTHOR CONTRIBUTIONS

JY conceived the concept. JY, QX, and NU designed the experiments, fabricated the devices, and performed electrical measurements. MC, NU, and TB carried out the STEM experiments. AI and NU performed the TOF-SIMS experiments. PM, NL, ND, and JK helped with experiments and data analysis. JY and NU wrote the paper. All authors discussed the results and implications and commented on the manuscript.

FUNDING

Part of this research was conducted at the Center for Nanophase Materials Sciences, which is a DOE Office of Science User Facility. This work was also partially supported by Air Force Office of Scientific Research (AFOSR) for the support through the MURI program under Contract No. FA9550-19-1-0213 and the U.S. Air Force Research Laboratory (AFRL) (Grant No. FA8750-18-2-0122). This work was also partially supported by the National Science Foundation under contract No. 2023752. Any opinions, findings and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of AFRL.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnano.2021.656026/full#supplementary-material

REFERENCES

Ashraf, M., Sundararajan, S. V., and Grenci, G. (2017). Low-power, low-pressure reactive-ion etching process for silicon etching with vertical and smooth walls for mechanobiology application. *J. Micro Nanolithogr. MEMS MOEMS* 16:034501. doi: 10.1117/1.jmm.16.3.034501

Baek, I. G., Kim, D. C., Lee, M. J., Kim, H.-J., Yim, E. K., Lee, J. E., et al. (2005). "Multi-layer cross-point binary oxide resistive memory (OxRRAM) for post-NAND storage application," in *IEEE InternationalElectron Devices Meeting, 2005. IEDM Technical Digest* (Washington, DC), 750–753. doi: 10.1109/IEDM.2005.1609462

Burr, G. W., Shenoy, R. S., and Hwang, H. (2016). "Select device concepts for crossbar arrays," in *Resistive Switching: From Fundamentals of Nanoionic Redox Processes to Memristive Device Applications,* eds Ielmini, D., and Waser R. (Hoboken, NJ: John Wiley and Sons), 623–660. doi: 10.1002/9783527680870.ch22

Chen, A. (2015). "Memory select devices," in *Emerging Nanoelectronic Devices,* eds Chen, A., Hutchby, J., Zhirnov, V., and Bourianoff, G. (Hoboken, NJ: John Wiley and Sons), 227–245. doi: 10.1002/9781118958254.ch12

Chen, Y., Jung, G. Y., Ohlberg, D. A. A., Li, X., Stewart, D. R., Jeppesen, J. O., et al. (2003). Nanoscale molecular-switch crossbar circuits. *Nanotechnology* 14:462. doi: 10.1088/0957-4484/14/4/311

Choi, B. J., Zhang, J., Norris, K., Gibson, G., Kim, K. M., Jackson, W., et al. (2016). Trilayer tunnel selectors for memristor memory cells. *Adv. Mater.* 28, 356–362. doi: 10.1002/adma.201503604

Choi, H., Yi, J., Hwang, S., Lee, S., Song, S., Lee, S., et al. (2011). "The effect of tunnel barrier at resistive switching device for low power memory applications," in *2011 3rd IEEE International Memory Workshop* (Monterey, CA: IMW), 1–4. doi: 10.1109/IMW.2011.5873243

Craft, E. B. (1925). *Crossbar Line Switch*. U.S. Patent No. 1,551,190. Washington, DC: U.S. Patent and Trademark Office.

Duggan, M. J., Saito, T., and Niwa, T. (1993). Ionic conductivity of tantalum oxide by rf sputtering. *Solid State Ionics* 62, 15–20. doi: 10.1016/0167-2738(93)90247-Z

Hudec, B., Hsu, C. W., Wang, I. T., Lai, W. L., Chang, C. C., Wang, T., et al. (2016). 3D resistive RAM cell design for high-density storage class memory—a review. *Sci. China Inform. Sci.* 59, 1–21. doi: 10.1007/s11432-016-5566-0

Joshua Yang, J., Zhang, M. X., Pickett, M. D., Miao, F., Paul Strachan, J., Li, W., et al. (2012). Engineering nonlinearity into memristors for passive crossbar applications. *Appl. Phys. Lett.* 100, 98–102. doi: 10.1063/1.3693392

Jung, J.-W., and Cho, W.-J. (2008). Tunnel barrier engineering for non-volatile memory. *JSTS J. Semic. Technol. Sci.* 8, 32–39. doi: 10.5573/jsts.2008.8.1.032

Kawahara, A., Azuma, R., Ikeda, Y., Kawai, K., Katoh, Y., Hayakawa, Y., et al. (2013). An 8 mb multi-layered cross-point ReRAM macro with 443 MB/s write throughput. *IEEE J. Solid State Circ.* 48, 178–185. doi: 10.1109/JSSC.2012.2215121

Kuekes, P. J., and Williams, R. S. (2001). *Demultiplexer for a Molecular Wire Crossbar Network*. U.S. Patent No. 6,256,767. Washington, DC: U.S. Patent and Trademark Office.

Kuekes, P. J., Williams, R. S., and Heath, J. R. (2000). *Molecular Wire Crossbar Memory*. U.S. Patent No. 6,128,214. Washington, DC: U.S. Patent and Trademark Office.

Lee, M.-J., Kim, S. I., Lee, C. B., Yin, H., Ahn, S.-E., Kang, B. S., et al. (2009). Low-temperature-grown transition metal oxide based storage materials and oxide transistors for high-density non-volatile memory. *Adv. Funct. Mater.* 19, 1587–1593. doi: 10.1002/adfm.200801032

Likharev, K. K. (1998). Layered tunnel barriers for nonvolatile memory devices. *Appl. Phys. Lett.* 73, 2137–2139. doi: 10.1063/1.122402

Lin, P., Li, C., Wang, Z., Li, Y., Jiang, H., Song, W., et al. (2020). Three-dimensional memristor circuits as complex neural networks. *Nat. Electron.* 3, 225–232. doi: 10.1038/s41928-020-0397-9

Liu, T. Y., Yan, T. H., Scheuerlein, R., Chen, Y., Lee, J. K., Balakrishnan, G., et al. (2014). A 130.7-mm2 2-layer 32-gb reram memory device in 24-nm technology. *IEEE J. Solid State Circ.* 49, 140–153. doi: 10.1109/JSSC.2013.2280296

Lo, C. L., Hou, T. H., Chen, M. C., and Huang, J. J. (2013). Dependence of read margin on pull-up schemes in high-density one selector-one resistor crossbar array. *IEEE Trans. Electron Devic.* 60, 420–426. doi: 10.1109/TED.2012.2225147

Logeeswaran, V. J., Kobayashi, N. P., Islam, M. S., Wu, W., Chaturvedi, P., Fang, N. X., et al. (2009). Ultrasmooth silver thin films deposited with a germanium nucleation layer. *Nano Lett.* 9, 178–182. doi: 10.1021/nl8027476

Mouttet, B. (2008). "Logicless computational architectures with nanoscale crossbar arrays," in *Technical Proceedings of the 2008 NSTI Nanotechnology Conference and Trade Show, NSTI-Nanotech, Nanotechnology* (Boston, MA), 73–75.

Rao, M., Xia, Q., Yang, J. J., Wang, Z., Li, C., Jiang, H., et al. (2019). "Learning with resistive switching neural networks," in *Technical Digest - International Electron Devices Meeting, IEDM* (San Francisco, CA), 35–4. doi: 10.1109/IEDM19573.2019.8993465

Scudder, F. J., and Reynolds, J. N. (1939). Crossbar dial telephone switching system. *Bell Syst. Tech. J.* 18, 76–118. doi: 10.1002/j.1538-7305.1939.tb00808.x

Sills, S., Yasuda, S., Strand, J., Calderoni, A., Aratani, K., Johnson, A., et al. (2014). "A copper ReRAM cell for Storage Class Memory applications," in *Digest of Technical Papers - Symposium on VLSI Technology*, (Honolulu, HI). doi: 10.1109/VLSIT.2014.6894368

Sun, Z., Pedretti, G., Ambrosi, E., Bricalli, A., Wang, W., and Ielmini, D. (2019). Solving matrix equations in one step with cross-point resistive arrays. *Proc. Natl. Acad. Sci. U.S.A.* 116, 4123–4128. doi: 10.1073/pnas.1815682116

Sun, Z., Pedretti, G., Bricalli, A., and Ielmini, D. (2020). One-step regression and classification with cross-point resistive memory arrays. *Sci. Adv.* 6:eaay2378. doi: 10.1126/sciadv.aay2378

Turner, T., and Chi, W. (2003). *Process for Smoothing an Si Substrate after Etching of SiO_2*. NASA Tech Brief.

Upadhyay, N. K., Jiang, H., Wang, Z., Asapu, S., Xia, Q., and Joshua Yang, J. (2019). Emerging memory devices for neuromorphic computing. *Adv. Mater. Technol.* 4:1800589. doi: 10.1002/admt.201800589

Upadhyay, N. K., Joshi, S., and Yang, J. J. (2016). Synaptic electronics and neuromorphic computing. *Sci. China Inform. Sci.* 59, 1–26. doi: 10.1007/s11432-016-5565-1

Upadhyay, N. K., Sun, W., Lin, P., Joshi, S., Midya, R., Zhang, X., et al. (2020). A memristor with low switching current and voltage for 1S1R integration and array operation. *Adv. Electron. Mater.* 6:1901411. doi: 10.1002/aelm.201901411

Wang, Z., Wu, H., Burr, G. W., Hwang, C. S., Wang, K. L., Xia, Q., et al. (2020). Resistive switching materials for information processing. *Nat. Rev. Mater.* 5, 173–195. doi: 10.1038/s41578-019-0159-3

Woo, J., Song, J., Moon, K., Lee, J. H., Cha, E., Prakash, A., et al. (2014). "Electrical and reliability characteristics of a scaled (∼30nm) tunnel barrier selector ($W/Ta_2O_5/TaOx/TiO_2/TiN$) with excellent performance (JMAX > 107A/cm^2)," in *Digest of Technical Papers - Symposium on VLSI Technology* (Honolulu, HI). doi: 10.1109/VLSIT.2014.6894431

Xia, Q., and Yang, J. J. (2019). Memristive crossbar arrays for brain-inspired computing. *Nat. Mater.* 18, 309–323. doi: 10.1038/s41563-019-0291-x

Xie, Y. W., Sun, J. R., Wang, D. J., Liang, S., and Shen, B. G. (2006). Reversible electroresistance at the Ag/La0.67Sr 0.33MnO_3 interface. *J. Appl. Phys.* 100:033704. doi: 10.1063/1.2222069

Yang, J. J., Strukov, D. B., and Stewart, D. R. (2013). Memristive devices for computing. *Nat. Nanotechnol.* 8, 13–24. doi: 10.1038/nnano.2012.240

Yoon, J. H., Zhang, J., Lin, P., Upadhyay, N., Yan, P., Liu, Y., et al. (2020). A low-current and analog memristor with Ru as mobile species. *Adv. Mater.* 32:1904599. doi: 10.1002/adma.201904599

Yu, L., Stampfl, C., Marshall, D., Eshrich, T., Narayanan, V., Rowell, J. M., et al. (2002). Mechanism and control of the metal-to-insulator transition in rocksalt tantalum nitride. *Phys. Rev. B* 65, 2451101–2451105. doi: 10.1103/PhysRevB.65.245110

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.

frontiers in Nanotechnology

ORIGINAL RESEARCH
published: 27 April 2021
doi: 10.3389/fnano.2021.670762

Engineering Method for Tailoring Electrical Characteristics in TiN/TiO_x/HfO_x/Au Bi-Layer Oxide Memristive Devices

Seongae Park[1,2]*, *Stefan Klett*[1], *Tzvetan Ivanov*[1,2], *Andrea Knauer*[2], *Joachim Doell*[2] and *Martin Ziegler*[1,2]*

[1]Department of Electrical Engineering and Information Technology, Technische Universtität Ilmenau, Ilmenau, Germany, [2]Institute of Micro and Nanotechnologies MacroNano, Technische Universtität Ilmenau, Ilmenau, Germany

OPEN ACCESS

Edited by:
Huanglong Li,
Tsinghua University, China

Reviewed by:
Guangdong Zhou,
Southwest University, China
Dashan Shang,
Institute of Microelectronics, China

***Correspondence:**
Seongae Park
seongae.park@tu-ilmenau.de
Martin Ziegler
martin.ziegler@tu-ilmenau.de

Specialty section:
This article was submitted to Nanodevices, a section of the journal Frontiers in Nanotechnology

Received: 22 February 2021
Accepted: 26 March 2021
Published: 27 April 2021

Citation:
Park S, Klett S, Ivanov T, Knauer A, Doell J and Ziegler M (2021) Engineering Method for Tailoring Electrical Characteristics in TiN/TiO_x/HfO_x/Au Bi-Layer Oxide Memristive Devices. Front. Nanotechnol. 3:670762. doi: 10.3389/fnano.2021.670762

Memristive devices have led to an increased interest in neuromorphic systems. However, different device requirements are needed for the multitude of computation schemes used there. While linear and time-independent conductance modulation is required for machine learning, non-linear and time-dependent properties are necessary for neurobiologically realistic learning schemes. In this context, an adaptation of the resistance switching characteristic is necessary with regard to the desired application. Recently, bi-layer oxide memristive systems have proven to be a suitable device structure for this purpose, as they combine the possibility of a tailored memristive characteristic with low power consumption and uniformity of the device performance. However, this requires technological solutions that allow for precise adjustment of layer thicknesses, defect densities in the oxide layers, and suitable area sizes of the active part of the devices. For this purpose, we have investigated the bi-layer oxide system TiN/TiO_x/HfO_x/Au with respect to tailored I-V non-linearity, the number of resistance states, electroforming, and operating voltages. Therefore, a 4-inch full device wafer process was used. This process allows a systematic investigation, i.e., the variation of physical device parameters across the wafer as well as a statistical evaluation of the electrical properties with regard to the variability from device to device and from cycle to cycle. For the investigation, the thickness of the HfO_x layer was varied between 2 and 8 nm, and the size of the active area of devices was changed between 100 and 2,500 μm^2. Furthermore, the influence of the HfO_x deposition condition was investigated, which influences the conduction mechanisms from a volume-based, filamentary to an interface-based resistive switching mechanism. Our experimental results are supported by numerical simulations that show the contribution of the HfO_x film in the bi-layer memristive system and guide the development of a targeting device.

Keywords: memristive devices, neuromorphic systems, bi-layer oxide memristive devices, memristive layer engineering, numerical simulation

1 INTRODUCTION

Memristive devices have been under the spotlight as an ideal element for neuromorphic computing due to their outstanding characteristics to emulate bio realistic information processing (Versace and Chandler, 2010; Legenstein, 2015; Mohammad et al., 2016; Jeong and Shi, 2019; Krestinskaya et al., 2020). Their non-volatile memory property, which is induced by an adaptation of the resistance state by applying electrical signals, makes them ideal candidates for the emulation of synaptic functionalities in artificial neural networks (Sah et al., 2014). For this application, they enable the realization of extremely energy-efficient hardware (Massimiliano and Yuriy, 2013; Ignatov et al., 2017) and have the potential of a high integration capability due to their simple two-terminal device structure (Lin et al., 2020). In particular, the integration of memristive devices in crossbar structures is worthy of mentioning here, which makes it possible to implement efficient learning schemes (Prezioso et al., 2015; del Valle et al., 2018; Alibart et al., 2013).

When considering the wide range of different neuromorphic systems, two main fields of applications in neuromorphic systems can be distinguished (Ielmini and Ambrogio, 2020): (i) neurobiologically realistic learning schemes and (ii) machine learning based algorithms. In neurobiologically realistic learning schemes the synaptic connections of a network are tuned by time-encoded spike-like signals (Snider, 2008), which typically requires nonlinear memristive device characteristics in a time-dependent manner (Ziegler et al., 2015; Dittmann and Strachan, 2019). In contrast to that, machine learning based algorithms use vector-matrix multiplications in which an explicit time dependence is not required (Ziegler et al., 2018). For that application, it is more important to set very precisely different resistance values for the individual memristive cells in a crossbar array (Yakopcic et al., 2015). Therefore, a time-independent linear resistance modulation is desirable (Chandrasekaran et al., 2019) which requires a high symmetry between the setting and the resetting characteristic of the memristive device over a wide range of resistance states (Wang et al., 2016).

In last couple of years many memristive device structures have been presented that are adequate for the machine learning algorithms (Kim et al., 2017; Cüppers et al., 2019; Li et al., 2020; Yao et al., 2020). It has been shown that the use of the memristive devices can significantly simplify the training routine in massively interconnected networks (Wang et al., 2019). Among those devices, particularly, memristive devices with a metal oxide bi-layer structure gained considerable interest in that field. Those memristive devices showed a significant improvement in the resistance modulation linearity (Li et al., 2018a) and the number of resistance states (Stathopoulos et al., 2017) along with the reduced variability in the resistive switching characteristics (Wang et al., 2010). The bi-layer metal oxide devices typically consist of an oxide layer that serves as a reservoir of oxygen vacancies and a solid state electrolyte layer which builds a Schottky-like interface contact with the adjacent metallic electrode (Huang et al., 2012; Bousoulas et al., 2016; Kim et al., 2018; Xiong et al., 2019). The resistive switching mechanism can be described as follows (Cüppers et al., 2019): under an external bias voltage oxygen vacancies are injected from the reservoir layer into the solid state electrolyte layer in which the oxygen vacancies are forming a filamentary conduction path toward the metallic electrode. This reduces both the resistance of the electrolyte layer and the Schottky barrier height and leads to a lowering of the overall device resistance (Asanuma et al., 2009; Zhao et al., 2020). An alternative concept of a memristive bi-layer metal oxide device is the double barrier memristive devices (DBMD) (Hansen et al., 2015). In this device structure, an ultra-thin solid electrolyte layer is sandwiched between a metal oxide layer and a metal electrode forming a Schottky-like contact. Here, the metal oxide layer serves as a diffusion barrier for oxygen ions, but not as a reservoir (Hur et al., 2010; Yin et al., 2015; Clima et al., 2016; Dirkmann et al., 2016; Hansen et al., 2017). The resistive switching effect is based on a shift of the oxygen ions in the solid state electrolyte layer in the direction of the metal electrode, which also leads to a reduction of the Schottky barrier height (Dirkmann et al., 2016). The advantage of the non-filamentary type of devices is that they did not require an electroforming step (Yoon et al., 2014), and the switching effect is based on a defined interface effect (Govoreanu et al., 2013). However, a disadvantage compared to bi-layer metal oxide devices with oxygen vacancy filaments is the shorter retention time (Solan et al., 2017). Furthermore, DBMDs have a rectifying characteristic (Gao et al., 2015) and thus a high asymmetry in the voltage polarity. However, these devices allow the realization of selector-free crossbar structures (Ma et al., 2017; Hansen et al. 2018) and the realization of biologically realistic computational schemes (Wang et al., 2015; Diederich et al., 2018).

A common challenge in the development of memristive devices is a tailor-made design of memristive devices for a respective computational scheme (Pei et al., 2015). For this, a number of materials and technology parameters have to be considered, such as the concentration of oxygen vacancies (He et al., 2017) or active ions (Clima et al., 2016), materials for the active layers and interface (connecting) layers (Li et al., 2018b). But also geometrical parameters such as layer thicknesses (Park et al., 2015; Wang et al., 2016; Li et al., 2018a) and size of the active areas (Lee et al., 2010) have to be considered carefully. These parameters are often only slightly known or not known at all but must be related to the device performance for a reliable device functionality (Niu et al., 2010; Lee et al., 2011). This particularly requires systematic investigations of the individual parameters and suitable device technology combined with a profound understanding of the underlying physical processes (Sun et al., 2019).

The aim of this work is to bridge the gap between the material design and the electronic characteristics of memristive devices for a tailored development of bi-layer metal oxide devices for neuromorphic systems. For this purpose, the bi-layer system $TiN/TiO_x/HfO_x/Au$ is examined in more detail in this paper. In detail, a four-inch wafer technology is presented, which allows to vary different device parameters, such as layer thickness and area size of the devices over the wafer. Using automated electronic measurements, a statistic of important device characteristics is

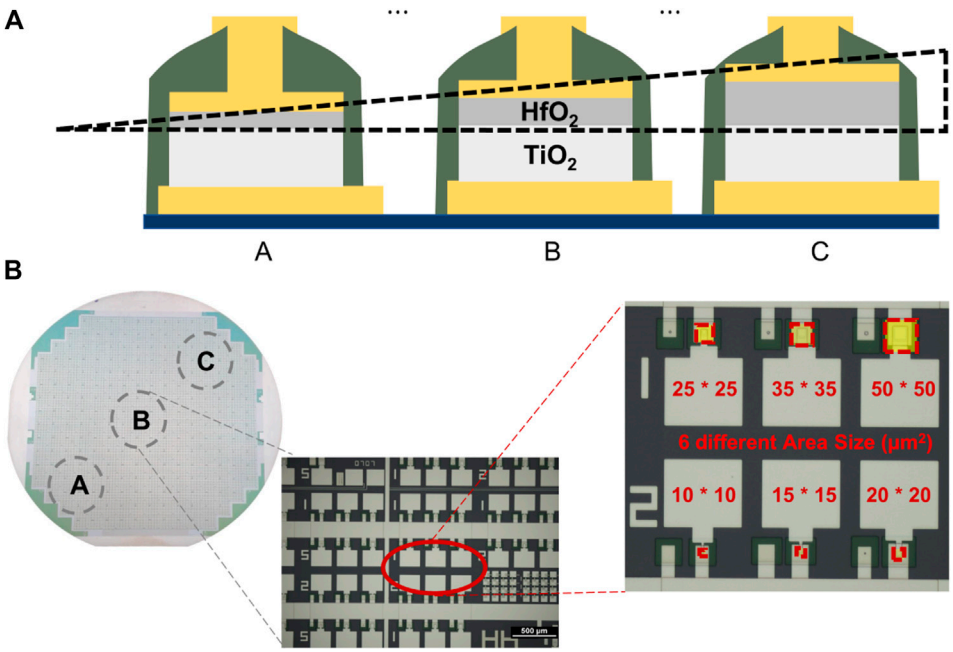


FIGURE 1 | Schematic representation of the technology for bi-layer oxide memristive devices: Using different sputtering geometries (referred to as methods 1 and 2 in the text), different types of devices have been produced, showing filamentary switching (referred to as type F) and area-based resistance switching (referred to as type R). **(A)** Cross-section of TiN/TiO_x/HfO_x/Au bi-layer oxide memristive devices over the wafer. The used deposition method (method 2) allows for a variation of the thickness of HfO_x layer from 2 to 8 nm. **(B)** Microscope images of a complete 4-inch wafer with 40,000 single devices and sections, showing the individual cells which consist of clusters with six devices, each with a different area size. Three regions on the wafer are indicated as A, B, and C. Each region has an area size of 0.5 × 0.5 cm^2. For type R devices a 2 nm HfO_x was deposited on all the three regions, while the thicknesses for type F have 2, 5, and 8 nm HfO_x films on region A, B, and C, respectively. Here, it was assumed that the HfO_x thickness is the same in each region due to the 1-dimensional thickness change over a 4-inch wafer and the relatively small area size. Approximately 60 memristive devices were measured in each area A, B, and C, respectively, for type F.

collected, and related to material properties and technology parameters. For a detailed understanding of the resistive switching mechanism, a physical device model is presented, which also allows a detailed examination of the individual device parameters. Essentially, we show that different sputtering conditions can influence oxygen ion and oxygen vacancy concentrations in the HfO_x layer. This causes different device characteristics. While an area-based switching mechanism leads to a rectifying current-voltage characteristic at high layer qualities with few oxygen vacancies, filamentary structures are formed in the HfO_x layer at higher concentrations of oxygen vacancies. This leads to a symmetrical current-voltage characteristic with multilevel resistant states and improved retention. In both cases, a change in the Schottky barrier between the HfO_x layer and the Au electrode can be identified as the reason for the observed switching effect. For a tailored design of memristive devices for their application, the different electronic characteristics are related to possible applications in neuromorphic systems.

The present work is structured as follows: In chapter 2, the implemented technology for manufacturing the memristive devices is presented first. Then the used methods for material and electrical characterization of the devices are discussed. Finally, chapter 2 presents a physical device model that serves to describe the underlying physical effects of the resistive switching mechanisms. In chapter 3, the obtained results are presented and discussed. For this purpose, first, the results of the electrical measurements and their statistics are shown in relation to an individual device and technology parameters. Then, important parameters of the devices are related to their electronic characteristics using the simulation model. Finally, the chapter discusses the application of the devices in neuromorphic computing architectures. The presented results are summarized in chapter 4.

2 MATERIALS AND METHODS

2.1 Device Technology

Figure 1 shows a developed device technology for bi-layer oxide memristive devices. In **Figure 1A** cross-sections of the fabricated TiN(50 nm)/TiO_x (30 nm)/HfO_x (2–8 nm)/Au(50 nm) bi-layer memristive devices with Al(300 nm) contact pads are sketched. They are fabricated on a 4-inch oxidized silicon wafer (1 µm of thermal SiO_x) in the full device technology. This technology is overviewed in **Figure 1B** and contains around 40,000 single devices, including test structures for the device development (see microscope images in **Figure 1B**). This allows the investigation of various device parameters, such as the active device area (six different area sizes are realized, as shown in **Figure 1B**), the thickness of the active HfO_x layer, and the material compositions over the wafer for a targeted

development of memristive devices. For a variation of the latter parameter, two different sputtering methods for the HfO_x layer were used. In particular, a variation in oxygen vacancies is required to achieve a desired resistive switching process in this class of memristive devices (He et al., 2017). Here, using a sputtering system equipped with three confocal source targets, two methods are employed for the deposition of the HfO_x which are referred to as method 1 and 2. During the deposition, the substrate is rotated to obtain a uniform film thickness, while a wedge film is formed without a rotation. The wedge is formed only along one direction. For method 1 the HfO_x layer was deposited on the wafer under optimal conditions, i.e., rotation of the substrate within a confocal sputtering arrangement. For method 2 the wafer was not rotated during the sputtering of the HfO_x layer. This leads to a reduced layer quality, but also a wedge over the wafer as shown in **Figure 1A** (further details are discussed below). As a result, we obtained two distinctive bi-layer oxide memristive device structures, which are referred to in the following as device R and device F.

In more detail, the TiO_x/HfO_x bi-layer was deposited on an inertial reactive sputtered TiN bottom-electrode via DC magnetron sputtering, where O_2/Ar of reactive gas was adjusted with the ratio of 10/40 and 10/29 for the TiO_x and HfO_x film, respectively. After the TiO_x was sputtered, the thickness of the HfO_x was controlled using the two discussed sputtering methods 1 (for device type R) and 2 (for device type F): as seen in **Figure 1A** a wedge layer with a variation of the HfO_x thickness from 2 to 8 nm was obtained for device F, where devices were fabricated along the axis (x-direction) perpendicular to the axis of the wedge (y-direction). Device R has a 2 nm uniform HfO_x layer. The layer deposition was finalized with an Au top-electrode layer. Thereafter, the material stack was patterned using photolithography and reactive ion etching for device R, while a lift-off in Dimethylsulfoxide (DMSO) was used for device F. The lift-off process was carried out due to the thickness variation of HfO_x in device F. Here, the investigation of the switching behaviors was preceded after we confirmed that the two patterning methods scarcely affected electrical characteristics. All devices were insulated with SiO_2 layers from the ambient air to avoid the influence of moisture in switching behaviors (Tsuruoka et al., 2012; Zhou et al., 2018; Zhou et al., 2020) (**Figure 1A**), and Al contact pads were deposited by e-beam evaporation.

2.2 Material Characterization

The development of the memristive devices was supported by a material characterization accompanying the manufacturing process. The thickness and the composition of the layers were characterized by ellipsometry measurements (SE500, Sentech) and surface profile measurements (Dektak 150, Veeco).

For a detailed material investigation, unstructured HfO_x films were deposited on silicon substrates. Therefore, the two described sputtering methods 1 and 2 were employed to deposit 37 nm thick HfO_x films. On those films ellipsometry measurements were performed at 632.8 nm at 70° of incidence. As the results, refractive indices of $n = 1.9889$ and $n = 2.0285$ were measured for, respectively, the uniform (method 1) and the wedge- (method 2) deposited HfO_x films. Thus, in agreement with previous investigations (Martínez et al., 2007) the film can be assumed to have amorphous crystallinities. However, the obtained n value from the uniform deposited film was higher than n of the wedge deposited HfO_2 film, which can be attributed to a reduced packing density (Gao et al., 2016).

Furthermore, X-ray photoelectron spectroscopy (XPS) measurement was utilized to study the quantitative atomic ratio O/Hf in sputtered HfO_x layers. The XPS analysis were carried out using monochromatic Al_K-alpha radiation (excitation energy hν = 1,486.68 eV) under charge neutralization using a SPECS SAGE HR 150 XPS system equipped with a 1D delayline detector and a Phoibos 150 analyzer. The calibration of the energy scale was ensured by reference measurements on a polycrystalline silver sample. Before the measurements, HfO_x was sputtered on Si/SiO_2 wafers for 900 s using the two different sputtering method 1 and 2. As a result, a ratio of O/Hf of 1.80/1 was observed for deposition method 2, while a ratio of 1.98/1 was recorded for samples sputtered via method 1 (see **Supplementary Data S1**). The sputtering method 1 provides a stoichiometry close to HfO_2, while the obtained stoichiometry via method 2 leads to optimal condition for the forming of oxygen vacancy filaments (McKenna, 2014). Thus, we can conclude that sputtering method 1 leads to a reduced number of oxygen vacancies than the sputtering method 2. Hence, a HfO_x layer with a higher density of oxygen vacancies can be assumed for device type F in respect to device type R.

2.3 Electrical Characterization

Current-voltage measurements (I-V curves) and voltage pulse measurements were carried out to characterize the electrical properties of TiO_x/HfO_x bi-layer memrsitive devices using a source measurement unit (Keysight b2901a). Therefore, a voltage is applied to the top-electrode of the device (bottom-electrode were grounded), while the current has been measured simultaneously. Furthermore, current compliance was imposed during the measurement to prevent the device from damage. The used current compliance was $I_{CC} = 10\,\mu A$, $I_{CC} = -5\,mA$ for R and F devices, respectively. For pulse measurements the device resistance was measured at, respectively, 1 and 0.1 V for R and F devices. The switching voltage to set and reset the device resistances was 3 V/−2 V for device type R and −1 V/1.5 V for F type devices. For both devices a pulse duration of ~10 ms was used. For a statistical evaluation of the electrical properties, median values were extracted taking into account the variability in cycle to cycle (C2C), and device to device (D2D). In the C2C investigation 10 times of DC voltage sweep cycles in one device were carried out. For reliable statistics, automated measurements of more than 10 memristive devices in each device parameter were performed, which means a total of 180 devices measured for three different thicknesses and six different area sizes. Both C2C and D2D statistics were investigated in DC conditions.

2.4 Physics Based Device Model

For a profound understanding of the resistive switching mechanisms and a targeted development of the devices a physics based device model was developed. In **Figure 2A** a sketch of this device model is shown: the model consists of

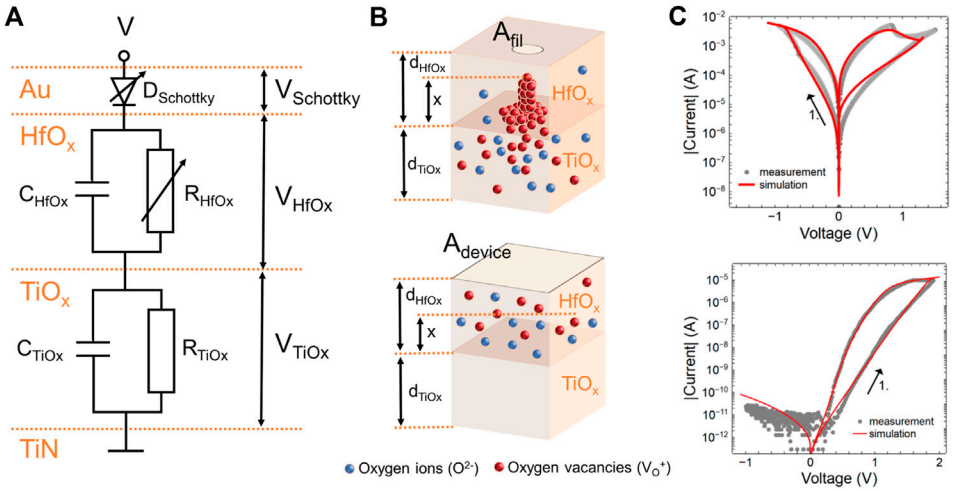


FIGURE 2 | (A) Sketch of the physics based device model for TiN/TiO_x/HfO_x/Au memristive devices. The model consists of two RC elements in the TiO_x and HfO_x layers and a Schottky diode at the interface of HfO_x and Au top-electrode. The external source voltage is divided into three local voltage drops at the HfO_x, TiO_x layers, and the Schottky diode. **(B)** Two different resistive switching mechanisms of type F devices **(top)** and type R devices **(bottom)** are considered. While in type F devices oxygen vacancies form a filamentary conductive path in the HfO_x film under an external voltage, in type R devices oxygen ions in the HfO_x film drift toward the top metal-electrode. As a result, in both device types the changed ion concentration leads to a modulation of the Schottky barrier height, which leads to the switching behaviors. Hereby originates the difference in the switching mechanisms from the density of the oxygen vacancies in the HfO_x film. **(C)** Measured I-V curves (gray) and the simulation results (red) for type F **(top)** and type R **(bottom)** devices. The arrows point to the SET direction.

two RC elements representing, respectively, the HfO_x and the TiO_x layer. The metal-semiconductor contact between the HfO_x layer and the Au electrode is considered by a Schottky diode ($D_{Schottky}$). Thus, an external applied voltage (V) is divided into the local voltage drops at the Schottky diode ($V_{Schottky}$), over the HfO_x layer (V_{HfOx}) and the TiO_x layer (V_{TiOx}) according to

$$V = V_{Schottky} + V_{HfOx} + V_{TiOx} \tag{1}$$

An important difference between here investigated two types of memristive devices is sketched in **Figure 2B**. While for the type F device a filament of oxygen vacancies is formed under the external voltage application, the type R device does not form any filaments. Essential for this is the concentration of oxygen ions and vacancies in the active HfO_x layer (Dirkmann et al., 2018). For the filamentary device F, we assumed that a number of oxygen vacancies are the mobile ions that vary between a minimum and a maximum concentration, denoted as N_{min} and N_{max}, respectively. In detail, for the filamentary device F we estimated $N_{min} = 4 \cdot 10^{24}\, m^{-3}$ and $N_{max} = 2 \cdot 10^{27}\, m^{-3}$ in accordance with the previous work (Menzel et al., 2011; Dirkmann et al., 2018). On the other hand, for the device type R we assumed a significantly lower concentration of oxygen vacancies due to a better layer quality. Here, the mobile species are oxygen ions where a concentration of $N = 10^{23}\, m^{-3}$ was used which is in qualitative agreement with (Dirkmann et al., 2016).

The concentration of the oxygen ions and vacancies has a particular effect on the active area used for resistance switching (cf. **Figure 2B**). Thus, for the filamentary device F only the filament area is relevant for the switching effect, i.e., $A = A_{fil}$ (see upper drawing in **Figure 2B**). For the type R device the whole device area is involved in the switching mechanism, i.e., $A = A_{device}$ (see lower drawing in **Figure 2B**). Both, the active area and the oxygen ion/vacancy concentration, are relevant for the resistance of the HfO_x layer:

$$R_{HfOx} = \frac{d_{HfOx}}{e \cdot zv_0 \cdot A \cdot \mu_n \cdot N} \tag{2}$$

where μ_n is the electron mobility, zv_0 is the ion charge number, and e is the elementary charge (Hardtdegen et al., 2018).

The layer thickness of TiO_x is significantly larger than that of HfO_x. Therefore, a much lower local electrical field strength is assumed ($E = V_{layer}/d_{layer}$). Thus, under an external bias voltage oxygen ion drift is suppressed within the TiO_x layer and the resistance R_{TiOx} of the TiO_x layer can be assumed to be constant. Nevertheless, the TiO_x layer plays a crucial role in the functionality of the bi-layer oxide structure: (i) it serves as a reservoir for oxygen vacancies in filament devices, and (ii) it stabilizes the switching process for both types of devices (Stathopoulos et al., 2017; Hardtdegen et al., 2018; Mikhaylov et al., 2020). For the latter point, the electronic contribution of the TiO_x layer is particularly important and has to be captured in the model. In general, the electronic charge transport through metal oxide layers can be determined by various transport mechanisms. It has been shown that a good approximation for the electron current is given by the following voltage realization (Jiang et al., 2016):

$$I_{TiOx} = j_0 \cdot A \cdot sinh(V_{TiOx}) \tag{3}$$

where j_0 is a fit parameter that has to be adapted to the real devices. The layer capacitances are given by

$$C_{layer} = \varepsilon \frac{A}{d_{layer}} \tag{4}$$

where $\varepsilon = \varepsilon_r \varepsilon_0$ is the permittivity of the respective layer.

The starting point of the switching model is the memristive behavior caused by a temporal and spatial change of the oxygen ions in the HfO_x layer. This effect is taken into account in the device model via an average ion velocity.

$$\frac{dx}{dt} = c_{drift} \cdot I_{Ion} \tag{5}$$

where x is the memristive state variable, i.e., the average position of oxygen ions or length of the filament in the HfO_x layer (cf. **Figure 2B**) and I_{Ion} is the ionic current of the oxygen ions. Furthermore, c_{drift} describes the resulting drift constant of the system, which is defined as

$$c_{drift} = \frac{\mu_n \cdot R_{mean}}{d_{HfOx} \cdot A} \tag{6}$$

Here, R_{mean} is the mean resistance of the HfO_x layer, which is given by $R_{mean} = \frac{1}{2} \cdot [R_{min} + R_{max}]$ for devices of type F and $R_{mean} = R_{HfOx}$ for the devices of type R. In particular, for devices of type F the resistance of the HfO_x layer can be specified as a function of the memristive state variable x:

$$R_{HfOx} = \frac{d_{HfOx}}{e \cdot z\nu_0 \cdot A \cdot \mu_n} \cdot \left[\frac{1}{N_{max}} \cdot x + \frac{1}{N_{min}} \cdot (1 - x) \right] \tag{7}$$

An essential important property of ionic based memristive devices is the back diffusion of the ions. The back diffusion determines the reliability and the storage time of the memristive device and is crucial parameter for a precise adjustment of multiple resistance states. In order to consider this behavior in the model, a further term was added to **Eq. 6**:

$$c_{drift} = \frac{\mu_n \cdot R_{mean}}{d_{HfOx} \cdot A} - c_{back} \cdot \left[1 - (2x - 1)^2\right] \tag{8}$$

Here c_{back} is a parameter that describes the strength of the back diffusion and must be adapted to the measured data.

The ion current can be written in the following form using the law of Mott and Gurney (Hardtdegen et al., 2018):

$$I_{Ion} = 4AeN_{mean}a\nu_0 \cdot \exp\left(-\frac{\Delta W}{V_T}\right) \cdot \sinh\left(\frac{a \cdot E_{HfOx}}{V_T}\right) \tag{9}$$

where ΔW is the diffusion barrier, which is reduced by the electric field E_{HfOx}. Furthermore, V_T is the thermal voltage, a the hopping distance, and ν_0 is the attempt frequency. N_{mean} determines the mean concentration of mobile ion species in the HfO_x layer, i.e., $N_{mean} = \frac{1}{2} \cdot (N_{max} + N_{min})$, while A is the active area of the device, which depends on the device type (cf. **Figure 2B**). Thus $A = A_{fil}$ for the filamentary device and $A = A_{device}$ for the interface based switching device (cf. **Figure 2B**).

The interface between the HfO_x/Au is assumed to be the relevant interface for the resistive switching process in both types of devices. In the simulation, this interface is modeled as a Schottky diode with variable Schottky barrier (ϕ_B). Using the thermionic emission theory, the charge transport over a Schottky barrier can be described in the following equation (Sze and Ng, 2006):

$$I_S = I_R\left(\exp^{\frac{eV}{nV_T}} - 1\right) \tag{10}$$

Where n is the ideality factor, which describes the deviation from an ideal diode characteristic, and I_R the reverse current, which is given by:

$$I_R = A^* A T^2 \cdot \exp^{\frac{-e\phi_B}{V_T}} \tag{11}$$

where A^* is the effective Richardson constant, which is $1.20173 \cdot 10^6 Am^{-2}K^{-2}$, T the local temperature, and A the active area. Under negative voltage polarities, however, the reverse current decreases gradually with the applied bias voltage. Therefore, on this polarity the reverse current is (Sze and Ng, 2006):

$$I_{R,v<0} = -A^* A T^2 \cdot \exp^{\frac{-e\phi_B}{V_T}} \exp^{\frac{-e\alpha_r\sqrt{|V|}}{V_T}} \tag{12}$$

Here α_r is a device dependent parameter. In our model we assumed that both quantities n and ϕ_B depend on the concentration of moved ions at the Au/HfO_x interface. A higher concentration of the negatively charged oxygen ions at that interface in R type devices increases the electron concentration locally. For devices of type F an increased concentration of oxygen vacancies increases the amount of acceptor states for electrons at the interface and thus there is also an accumulation of electrons at the interface. Thus, for both type of devices a reduction of the Schottky barrier is expected, which in turn has a significant effect on the charge transport through the complete device. In the model this was considered by a state variable dependency of those quantities:

$$\phi_B(x) = \phi_B^{LRS} \cdot \frac{x}{x_{max}} + \phi_B^{HRS} \cdot \left(1 - \frac{x}{x_{min}}\right) \tag{13}$$

$$n(x) = n^{LRS} \cdot \frac{x}{x_{max}} + n^{HRS} \cdot \left(1 - \frac{x}{x_{min}}\right) \tag{14}$$

the values for n^{LRS} and n^{HRS}, as well as ϕ_B^{HRS} and ϕ_B^{LRS} were obtained from the experimental I-V curves using **Eq. 10**. Another important parameter influencing the ion movement within the memristive device is the local temperature change. This includes mainly Joule heating and plays a crucial role particularly in filamentary-based device structures. This was taken into account in the simulation model as follows (Ielmini and Milo, 2017).

$$T = I \cdot V \cdot R_{therm} + T_0 \tag{15}$$

Here, R_{therm} is the effective thermal resistance and T_0 is the room temperature. The temperature along the filament is assumed to be relatively homogeneous and thus a uniform filament temperature can be assumed (Ielmini and Milo, 2017).

The device parameters have been carefully collected from measurements and literature and are summarized in **Table 1**. The I-V curves simulated with the model are shown in **Figure 2C** and compared with the measurement curves determined

TABLE 1 | Simulation Parameter.

Parameter	Value		Parameter	Value
	type F	*type R*		
$\Phi_B^{HRS}\,[eV]$	0.25	0.71	$\alpha_r\,[V/A\cdot s]$	$1.2\cdot10^6$
$\Phi_B^{LRS}\,[eV]$	0.06	0.61	$\mu_n\,[m^2/V\cdot s]$	10^{-5}
n^{LRS}	4.5	3.9	$\nu_0\,[Hz]$	$3\cdot10^{11}$
n^{LRS}	5	4.45	ε_r^{TiOx}	17
$N_{max}\,[m^{-3}]$	$2\cdot10^{27}$		ε_r^{HfOx}	5.5
$N_{min}\,[m^{-3}]$	$4\cdot10^{24}$		$\Delta W\,[eV]$	0.425
$N\,[m^{-3}]$		10^{23}	$a\,[nm]$	0.4
$d_{HfOx}\,[nm]$	2–8	2.5	$d_{TiOx}\,[nm]$	30
$j^0_{TiOx}\,[A/m^2]$	$5.8\cdot10^7$	$5.8\cdot10^4$	$R_{therm}\,[K/W]$	$1.1\cdot10^4$
$A_{device}\,[\mu m^2]$		100	$A_{fil}\,[nm^2]$	6,362
C_{back}/C_{drift}	0	$3.25\cdot10^{-11}$	$T_0\,[K]$	273

experimentally. As can be seen from this figure, the model presented here shows very good agreement with the experiment. A more detailed description of the results follows in the next chapter.

3 RESULTS AND DISCUSSION

3.1 Resistive Switching and Statistical Examinations

In **Figure 3A** typical obtained I-V curves of the two kinds of memristive devices (named as F and R) are shown. Common for both device types is that they show bipolar resistive switching with a gradual resistance change. A major difference between both types of devices is their voltage polarity. While type R devices require a positive voltage (applied to the top electrode) to set the device, type F devices require a negative voltage to be applied for the set process. The different polarity behaviors are originated from differences in concentration and species of mobile ions, which will be discussed in *Concentration of Mobile Icon*. Furthermore, while a highly rectifying memristive behavior is observed for device type R, a more symmetric memristive behavior is found for devices of type F together with a 3 times higher current level as compared to type R devices (cf. **Figure 3A**). In some more detail: the rectifying

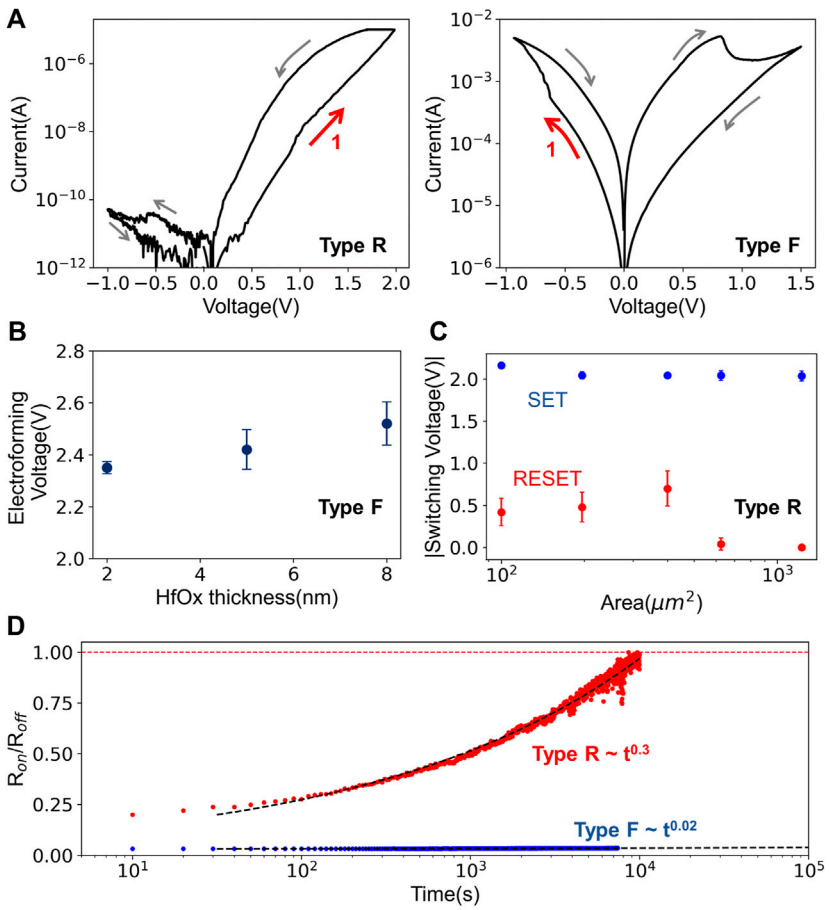


FIGURE 3 | (A) Representative I-V curves of TiN/TiO_x/HfO_x/Au bi-layer memristive devices for type R **(left)**, and type F **(right)**. The arrows point to the resistive switching direction. A clear rectifying behavior was observed in the type R, and a symmetric switching behavior in the type F. **(B)** Electroforming voltage (median values) as a function of HfO_x thickness. Electroforming voltages were tailored by the thickness of the HfO_x in type F. **(C)** SET (blue) and RESET (red) voltages (median values) as a function of device active area size. The smaller the area is, the lower SET/RESET voltage was observed. **(D)** Retention measurement and fitting curves. Type F devices (blue) showed an improved retention characteristic compared to type R (red). The fitted curves are shown with dashed lines, and the fitting constants were 0.02 and 0.3 for type F and type R, respectively.

behavior of devices of type R can be quantified by the ratio between the maximum and minimum current $r_{asym} = |I_{max}/I_{min}|$ at a voltage of ±0.5 V. Here we were able to determine $r_{asym} = 70$ for an active device area of 100 μm^2 which, however, has a strong area dependence. In particular, for an area of 625 μm^2 the asymmetry ratio r_{asym} is reduced from 70 to 4 (further information is provided in **Supplementary Figure 3**).

An important property of memristive devices and another difference between the here considered devices is the initial electroforming process. While no initial electroforming step was necessary for type R devices, type F devices had to be electroformed at the beginning. For a more precise discussion of the electroforming process of type F devices, the median of the required voltages as a function of the thickness of the HfO_x layer is depicted in **Figure 3B**. In detail, electroforming voltages of 2.35, 2.42, and 2.52 V have been observed for, respectively, a 2, 5, and 8 nm thick HfO_x layer. Thus, the electroforming voltage shows moderate thickness scalability. After the electroforming process type F device are operated typically at a maximum (minimum) voltage ±0.75 V. In terms of operating voltage, type F devices also differed from type R devices: type R devices require on average a 1.3 V higher operating voltage with a moderate area dependence (cf. **Figure 3C**). The operating voltages for type R devices were 2.2 V/−0.42 V (SET/RESET) for the smallest area size and 1.7V/0 V for the largest area size. However, the type R devices show a more gradual transition from the inertial high resistive state (HRS) to the low resistive state (LRS) (cf. **Figure 3A**).

A crucial property of memristive devices is their retention time. Furthermore, a detailed investigation of the retention characteristic already provides important conclusions about the underlying resistive switching mechanism (Hansen et al., 2015). The retention behavior for the here discussed two types of memristive bi-layer structures are shown in **Figure 3D**. For the measurement of the retention characteristics, the two types of devices were initially set to the low resistance state and then the resistance value of the devices was determined at regular intervals by means of voltage pulses. As can be seen in the figure, the two types of devices show quite different retention behaviors. For device type R, diffusive characteristics were observed (see red data points in **Figure 3D**), while much higher retention is observed for device type F. In order to analyze the retention characteristics in some more detail the retention curves were fitted using a power law according to the Curie-von Schweidler equation (Mikheev et al., 2014; Goossens et al., 2018):

$$R = R_{on}/R_{off} \propto t^{\alpha} \tag{16}$$

where α is a fit parameter, which is between 0 and 1 (Yang et al., 2010). As a result, $\alpha = 0.3$ is observed for devices of type R, whereas $\alpha = 0.02$ best reflects the experimental data for devices of type F. While $\alpha = 0.02$ describes a very good retention time for devices of type F, $\alpha = 0.3$ shows clearly lower retention for devices of type R. This difference can be explained by the different ion dynamics between the two devices. While in the type R device the filamentary structures are suppressed and mobile oxygen ions are shifted toward the electrode, in the type F devices oxygen vacancies are organized in filamentary structures. This leads to different activation energies of the ion dynamics. It has been shown that the activation energy of oxygen vacancies is in the range of 6–8 eV inside filaments (McKenna, 2014), while it is less than 1 eV outside filaments (Dirkmann et al., 2016; Dirkmann et al., 2018). Furthermore, it is worth mentioning that localized electronic states at the Au/HfOx interface can also contribute to the observed switching mechanism. The localized electronic states are filled or emptied depending on the applied bias voltage polarity (Hansen et al., 2015; Zhou et al., 2016). Even if the exact mechanism underlying the switching effect cannot be clearly explained by the presented measurements alone, the strong difference in the retention times and the different voltage polarity indicate that oxygen vacancies dominate the respective switching behavior in type F devices, while mobile oxygen ions lead to resistive switching in type R devices.

In order to be able to make suitable statements about possible applications of the memristive devices in neuromorphic systems and to tailor the device characteristics accordingly, a statistical investigation of relevant device parameters is required. As relevant device parameters we considered the thickness of the active HfO_x layer (d_{HfOx}), the active area size A, and the concentration of mobile ions N. The results of that investigation are shown in **Figure 4**. In the cumulative distribution function (CDF) of the resistance of type R devices (**Figure 4A**) and of type F devices (**Figure 4B**) the resistances were obtained from voltage sweep measurements by calculating the corresponding median values and the standard deviations. The resistance obeyed a lognormal distribution for all examined devices. For devices of type R (cf. **Figure 4A**) the resistance distributions for area sizes from 100 to 1,225 μm^2 are shown. For the devices of type F the different curves in **Figure 4B** originate from the different d_{HfOx}. As a result, we found that for devices of type R the resistance window decreased with increasing active area size A which can be attributed to the decreasing rectifying ratio (further details are in the supplement). Furthermore, the relatively small width of the CDF curve was observed for type R devices indicating a high device uniformity. For the devices of type F, the low resistant states show a steeper change in the CDF curve than the high resistant states. Even though the found variations in the resistances are small, the devices with a HfO_x thickness of 5 nm show here the best variability.

To be able to make detailed statements about the requirements to be met by the physical device parameters, the influence of the variability in relation to the physical parameters must be examined. Therefore, normalized standard deviations of the devices were determined and plotted as a function of the active volume, i.e., the layer thickness of the active HfO_x layer times the device area. The obtained results for both types of devices are shown in **Figure 4C**. The figure shows the different measured variabilities for the devices of type R (triangular data points) and the devices of type F (circular data points) as a function of the active volume of the device. For devices of type R it appears that the variability is only weakly affected by increased area size. Here the normalized standard deviation of 0.2 is quite constant over the investigated area sizes (see the dashed black line in **Figure 4C**). However, for devices of type F a parabolic curve best describes the found trend which indicates a clear optimum at

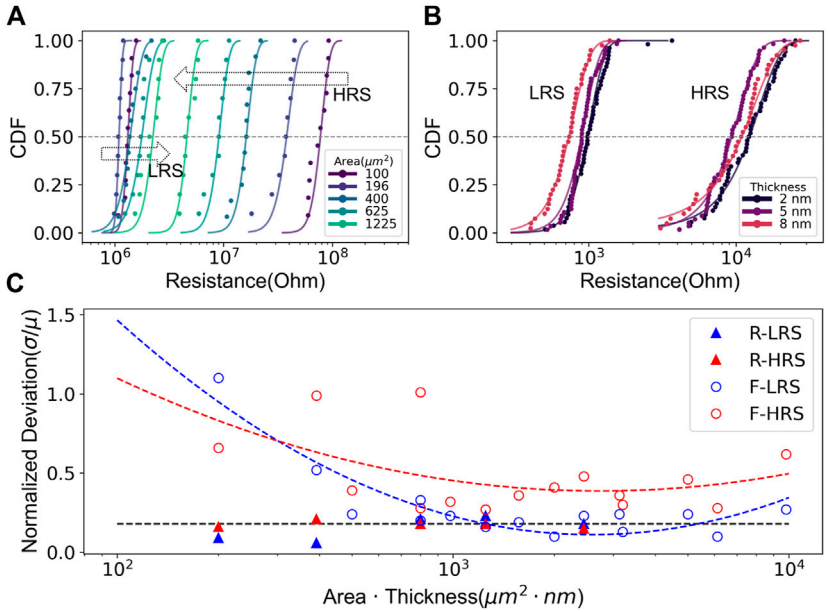


FIGURE 4 | Resistance distribution for different physical device parameters in type R devices **(A)**, and type F devices **(B)**. **(A)** CDF for different active device area sizes of device type R. The arrows point to the direction of the increasing area. **(B)** CDF for different HfO_x thicknesses in type F devices. **(C)** Device variability (σ/μ) in terms of physical device parameters. The variability of type R devices is marked with triangles, and type F devices with circles. Red color for high resistance states, and blue color for the low resistance were used. The fitted curves are shown with dashed lines. For type F devices a parabolic trend was observed, which shows a correlation between the area and the thickness of the HfO_x in the variability of devices.

approximately $2.6 \cdot 10^3\ \mu m^2 \cdot nm$ for the HRS and $2.73 \cdot 10^3\ \mu m^2 \cdot nm$ for the LRS. This means that a reduction from $d_{HfOx} = 8\ nm$ to $d_{HfOx} = 2\ nm$ increases the optimal device area from $A_{device} = (18 \times 18)\ \mu m^2$ to $A_{device} = (36 \times 36)\ \mu m^2$. Thus, the trend can be observed that with extremely small layer thicknesses, a larger area leads to a more stable device behavior.

3.2 Resistive Switching Mechanism and Device Requirements

A sound understanding of the resistive switching mechanism is important to enable a targeted design of the memristive devices for application in neuromorphic systems. For this reason, the device model described in *Physics Based Device Model* was used to interpret the experimental results described above. The obtained results are shown in **Figure 2C**. Therein the experimental I-V curves are compared with simulated curves. The used simulation parameters are summarized in **Table 1**. In both cases, i.e., in the case of the filamentary (type F device) and the interface-based device (type R device), one can see quite good agreement between simulation model and experiment. The main difference between the two I-V curves in the simulation model comes from (i) differences in concentration and species of mobile ions due to stoichiometric differences between Hf and O, (ii) different active areas that are responsible for the switching behavior (cf. **Figure 2C**), and (iii) the lowering of the Schottky barrier and the change of the ideality factor. In order to understand more exactly the underlying switching mechanisms that lead to the different device characteristics, the mentioned points (i–iii) will be discussed in the following in more detail.

3.2.1 Schottky Barrier Height Lowering

From the measured I-V curves the minimum and the maximum values of the variation of the Schottky barrier were determined. Therefore, **Eq. 11** was adapted to the experimental data at the voltage interval ranging from 0 to 100 mV for both device types. Furthermore, we made the assumption that the resistance of the device does not change in that interval. As a result, we found that the values for the Schottky barriers vary between 65 meV and 250 meV for the filamentary device F (upper graph in **Figure 2C**), whereby a barrier variation of 615 meV and 708 meV was obtained for the area-based device R (lower graph in **Figure 2C**). In addition, the fit procedure also considered the ideality factor as an adjustable parameter, whereby we obtained 5.0 and 4.54 for filamentary device F and 3.9 and 4.45 for the area-based device R. A key finding from that analysis is that area-based device has a much higher Schottky barrier, while for both devices a strong variation of the Schottky barrier is observed. In order to analyze that finding in more detail, simulations were carried out with a maximal barrier lowering of 200 meV. The results of the simulations for the two device types are shown in **Figure 5**. The obtained changes in the resistance value for the type R (**Figure 5A**) and type F devices (**Figure 5B**) are shown. For this purpose, the Schottky barrier of the high ohmic state (see indicated ϕ_B^{HRS} in the figures) was used as a starting value and the barrier height was successively reduced, i.e., $\Delta\phi_B = (\phi_B^{HRS} - \phi_B^{LRS})$. This confirms the experimentally observed finding of a strong

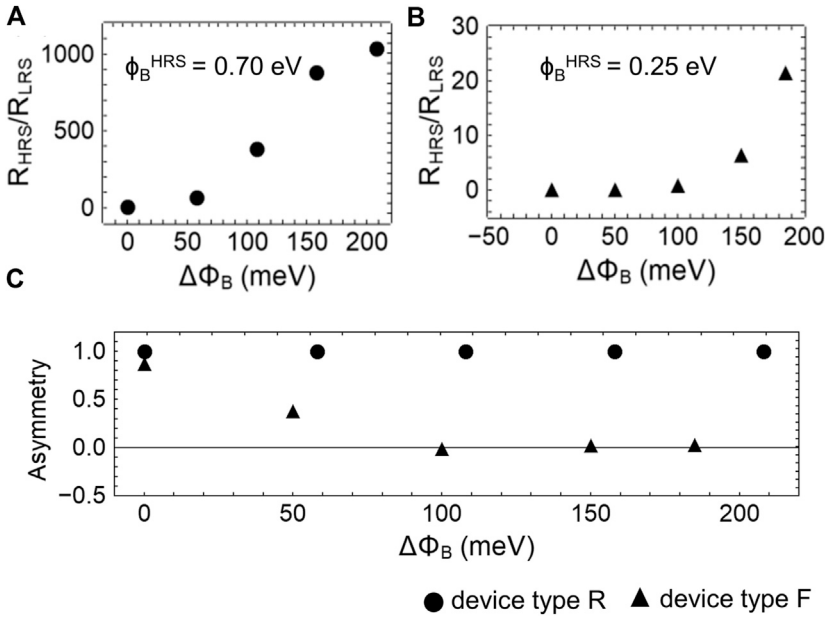


FIGURE 5 | Influence of the Schottky barrier height lowering $\Delta\Phi_B$ on the device resistance: the resistance value as a function of Schottky barrier lowering for type R devices **(A)** and type F devices **(B)**. Φ_B^{HRS} is the inertial Schottky barrier height for the respective R and F type device. **(C)** Asymmetry as a function of the barrier lowering. Strong asymmetric characteristics are observed in device R (circle) within 200 *meV* in the barrier lowering, while the asymmetry was destroyed in device F (triangles) at 100 *meV* barrier lowering.

dependence of the resistance change of the devices on the maximum barrier lowering for both types of devices. Two types showed different switching mechanisms: filamentary- and interface-type. However, Schottky contact adjustment was an essential factor in resistive switching behaviors for both typem F and type R devices. Thus, it can be stated that the Schottky barrier change is the main reason for the switching behavior of the two different memristive devices, which is in good agreement with previously published data (Hansen et al., 2015; Hardtdegen et al., 2018).

In fact, a significant influence on the Schottky barrier height and therewith an important technology parameter is the material used for the electrode and the oxide layer. Since the same electrode material (Au) was used for both types of devices, the observed difference can only come from the HfO_x layer. In this respect, the difference is mainly in the layer quality due to the different manufacturing processes that we used for the two devices. This has a particular effect on the number of oxygen ions and vacancies, which we will discuss in more detail below. However, a qualitative indicator of contact quality is the asymmetry between the minimum and the maximum current values in the I-V characteristics and the ideality factor n of the contacts. Here we observe that n is lower for the type R devices than for the type F devices and the type R devices show a clear asymmetry and therefore a stronger rectifying characteristic. To investigate this point, the asymmetry as a function of the barrier lowering $\Delta\phi_B$ is shown in **Figure 5C**. The asymmetry was determined by the following formula: $[(I_{max} - I_{min})/(I_{max} + I_{min})]$. As a result, we found that the barrier lowering of the area-based devices does not affect the asymmetry, whereas 100 mV of the type F devices is sufficient to completely destroy the asymmetry.

3.2.2 Concentration of Mobile Ions

One of the most important parameters for the resistance switching mechanism of memristive devices is the concentration of mobile ions. In the simulation model we have, therefore, investigated the concentration of mobile ions in the HfO_x layer as a further central device parameter. It turned out that for the rectifying device R a constant low concentration of negatively charged oxygen ions ($N_{min} = N_{max} = 10^{23}\ m^{-3}$) best describes the experimental I-V curve, where a variation of positively charged oxygen vacancies from $N_{min} = 4 \cdot 10^{24}\ m^{-3}$ to $N_{max} = 2 \cdot 10^{27}\ m^{-3}$ for the filamentary device F gives the best agreement with the experiment (cf. **Figure 2C**). These obtained results are in good agreement with previous investigations (Dirkmann et al., 2016; Hardtdegen et al., 2018) and support the model outlined in **Figure 2A**. In order to investigate these variations, the concentration of oxygen vacancies was varied in the range from $5.2 \cdot 10^{25}\ m^{-3}$ to $2.5 \cdot 10^{27}\ m^{-3}$ for F type device and a variation from $1 \cdot 10^{22}\ m^{-3}$ to $1 \cdot 10^{24}\ m^{-3}$ of oxygen ions were used for the rectifying device type R. The simulation results are summarized in **Figure 6**.

Figure 6A shows I-V curves for the filamentary device with different mean concentrations of oxygen vacancies. In particular, two major trends for the change in oxygen vacancies can be seen: (i) the hysteresis shows a clear variation with the change of the oxygen vacancies, and (ii) the values for set and reset voltages become smaller. To interpret these two properties in more detail, **Figure 6B** shows the ratio R_{HRS}/R_{LRS} at −0.1 V as a function of the

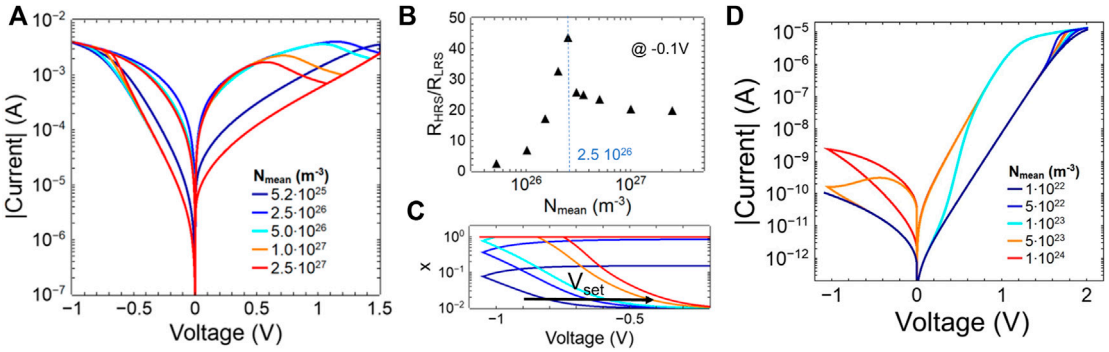


FIGURE 6 | Simulated I-V curves of TiN/TiO_x/HfO_x/Au bi-layer memristive devices with a variation of the mean concentration of mobile ions in the HfO_x layer. **(A)** The variation of the mean concentration of positively charged ions in the range from $5.2 \cdot 10^{25}\ m^{-3}$ to $2.5 \cdot 10^{27}\ m^{-3}$ is simulated for type F. **(B)** The ratio of R_{HRS}/R_{LRS} at −0.1 V as a function of the mobile ions. The optimum memristive hysteresis was observed at the mobile ion concentrations of $2.5 \cdot 10^{26}\ m^{-3}$. **(C)** The state variable (x) as a function of the applied voltages for different concentrations of mobile ions in the HfO_x layer. Operating voltages (SET/RESET voltage and threshold voltage) in the bi-layer memristive devices were affected by the concentration of mobile ions in the HfO_x layer. **(D)** The variation of the concentration of negatively charged ions in the range from $1 \cdot 10^{22}\ m^{-3}$ to $1 \cdot 10^{24}\ m^{-3}$ is simulated for type R. The resistance values and the hysteretic effects were influenced by the concentration of oxygen ions.

concentration of oxygen vacancies. What can be seen very clearly is that there is an optimum of the ratio at $2.5 \cdot 10^{26}\ m^{-3}$. The reason for this is the threshold value of the oxygen diffusion, which essentially determines R_{HRS} and R_{LRS}. For this purpose, **Figure 6C** shows the state variable x as a function of the applied voltage for the different concentrations at oxygen vacancies. According to **Eq. 7**, the concentration of the oxygen vacancies determines the change of the resistance of the active HfO_x layer, but also the ion drift (see **Eq. 6**), and thus the change of the state variable x. What can be observed from the simulation is that high oxygen vacancy concentrations cause a change in the state variable already at very low voltage values (cf. **Figure 6C**). This means that a threshold value for setting the device can no longer be set precisely, which already leads to a reduced resistance value for a voltage of −0.1 V. However, since threshold values are important for the application, a precise setting of the oxygen vacancies is an important device parameter that should be chosen carefully.

The results for the rectifying memristive device under varying concentrations of oxygen ions are shown in **Figure 6D**. Here it can be seen that the concentration of oxygen ions has an effect on the change of the resistance value as well as on the retention characteristics. Thus, at extremely low concentrations of oxygen ions, only small hysteric effects are observed, while a pronounced hysteresis is only observed at a concentration of $5 \cdot 10^{22}\ m^{-3}$. This concentration of oxygen ions, thus, defines a critical minimum for memristive switching behavior.

3.2.3 Area Dependence

An experimentally important indication of the type of resistive switching mechanism is given by the area dependence of the devices. For this purpose, the product of area times resistance (RA) as a function of the area of the devices is shown for both device types in **Figure 7**. While the upper graph of **Figure 7** is presenting the results from the rectifying device R, the graph on the bottom is showing the area dependency for the filamentary device F. Here, the data points are taken from the measurements

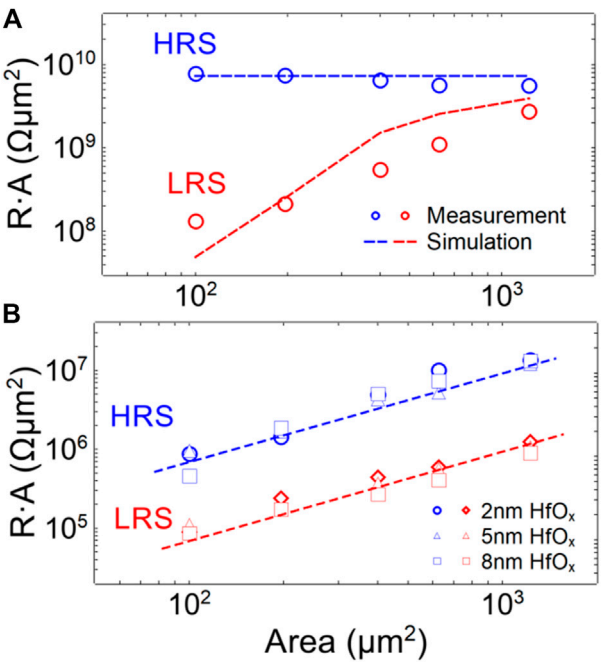


FIGURE 7 | The $R \cdot A$ product as a function of the device active area for **(A)** type R and **(B)** type F. The dashed lines present simulated device models. **(A)** The $R \cdot A$ product showed a constant behavior in HRS for type R, while an area dependent behavior for type F **(B)**. The area dependency of the $R \cdot A$ product in LRS for device R **(A)** was caused by the area dependency of the ion drift constant, which was in good agreement with the simulation results.

and the dashed lines are the results of the simulation model. The expected trend can be seen for the high resistant state of the devices: for the filamentary device F, an area independent behavior is seen, while a clear area dependence was found for the rectifying device R. It is noticeable that the low ohmic state of the rectifying device R shows a non-uniform area dependence.

One would actually expect a horizontal line in the $R \cdot A$ vs. A representation chosen here. This is relevant with the drift constant c_{drift} from **Eq. 8**, which depends on the layer thickness of the HfO_x layer (d_{Hfox}) and the active device area (A). While the changes in the layer thicknesses (d_{Hfox}) in the experimentally investigated interval cause only a small change in c_{drift}, the changes in the area for the rectiyfying device R have a considerable influence on c_{drift}. In this case, the drift constant (c_{drift}) is reduced, especially for large active areas, and thus a smaller change in the state variable x is induced during a voltage ramp. This in turn leads directly to a smaller change in the device resistance, which we can also observe experimentally. Thus, this shows that the choice of the device area has an influence on the dynamics of the oxygen ions and vacancies, especially for the rectifying memristive device. Furthermore, these results give good confirmation of the proposed switching mechanism, i.e., area-based switching for the R-type device and filamentary switching for the F-type device.

3.3 Applications for Neuromorphic Computing

The emulation of synaptic plasticity processes with memristive devices is one of the most important application fields of memristive devices in neuromorphic systems (Ziegler et al., 2018). In particular, this requires the design of suitable learning and training processes (Ielmini and Ambrogio, 2020), which needs a targeted adjustment of the resistance states of individual memristive devices in networks. In the following section, it is presented that type F devices fulfill requirements for machine learning based algorithms, whereas type R devices for neurobiologically inspired learning schemes.

The challenge in the machine learning based algorithm is to create suitable local learning rules that guarantee a local change of the device state so that a requested global network functionality is enabled. Therefore, a general framework is provided by the Hebbian learning rule (Ziegler et al., 2015), which can be systematized in the following equation:

$$\frac{d\omega_{ij}}{dt} = f\left(\omega_{ij}, A_j, A_i\right) \tag{17}$$

where ω_{ij} describes the coupling strength between the pre- and the post-synaptic neuron and $A_{j(i)}$ their activities, as sketched in **Figure 8A**. This formula translates Hebb's postulate, that synaptic connections change only when the respective pre- and post-synaptic neurons are active at the same time. The choice of the function f is thus decisive for the learning or training procedure of any artificial neural network. A common way to realize the weight update according to **Eq. 17** is provided by the delta rule (Kendall et al., 2020), which is at the heart of deep learning neural networks:

$$\Delta\omega_{ij} = \alpha \cdot \left(d_i - y_i\right) \cdot p_j \tag{18}$$

where the coefficient α is named learning rate and is usually positive. Furthermore, p_j is the activity of the pre-neuron (input value), y_i is the activity of the post-neuron (output value), and d_i the desired output value for a given input p_j used during learning. To convert that equation into hardware, the coupling strength ω_{ij} can be represented by the conductance G_{ij} of the memristive device, and y_j, p_j, and d_j by voltage- or current-dependent functions that either increase or decrease the conductance of the memristive device (Linares-Barranco et al., 2011). Thus, for the implementation of memristive devices in neuromorphic network structures via the delta rule a precise change of the conductance in dependence on applied voltage (or current) pulses is required (Payvand et al., 2018).

In order to investigate the resistance update behavior of the devices used here under voltage pulsing, AC pulses trains were used (see the sketch in **Figure 8B**). Therefore, a voltage train of 20 SET pulses followed by 20 RESET pulses was applied to the devices. Furthermore, the resistance states have been determined by a readout pulse that followed each switching pulse. The results obtained are shown in **Figure 8C** for type F devices and in (D) for type R devices. Read pulses of 1.0 and 0.1 V with a pulse width of 10 ms have been used for R and F devices, respectively. For the reset pulse, the width was 1 ms and the amplitudes were −2 and 1.5 V for R and F devices. As a result, a gradual transition change with multiple resistance states was observed in devices of type R, while a more binary behavior was recorded for devices of type F (cf. **Figures 8C,D**). In order to investigate the pulse behavior of the type F devices in more detail with respect to **Eq. 18**, the voltage amplitudes for SET and RESET pulses were successively changed in each pulse, as sketched in the inset of **Figure 8E**. The therewith obtained resistance change as a function of the voltage pulse amplitudes is shown in **Figure 8E**. Thus, a linear change in resistance with a successive incremental increase of the voltage pulse height was recorded for both set and reset. Furthermore, the resistance change was nearly symmetric in both resistance states, presenting 0.44 and 0.56 linearity for set and reset, respectively. Hence, this behavior fulfills nicely the requirement proposed by equation 18 and makes type F devices, together with their relatively good retention, perfect candidates for the hardware realization of deep learning neural networks. In this context, bi-layer oxide memristive devices of similar types have already proven their performance (Yao et al., 2020).

While the delta rule underlies a variety of machine learning systems and allows an effective implementation of Hebb's learning rule within artificial neural networks, there is no explicit time dependence. However, the time dependence of learning processes is an important parameter in biology and determines how the synaptic connection is strengthened or weakened (Panwar et al., 2017). Here, an important property is the memory effect of synapses which leads to a sustained strengthening of the synaptic connection after repeated (high frequency) excitation named long-term potentiation (LTP). Therefore, the respective time interval between the excitation is required. At this respect, the diffusive ionic processes of memristive devices and their memory behavior are unique properties for the emulation of bio-realistic time-dependent learning (Ziegler et al., 2018), such as spike-timing dependent plasticity (STDP) and paired-pulse facilitation (PPF), to only mention two important plasticity processes. Many ways to emulate such learning schemes have been presented in recent years with memristive devices (Wang et al., 2020). However, the

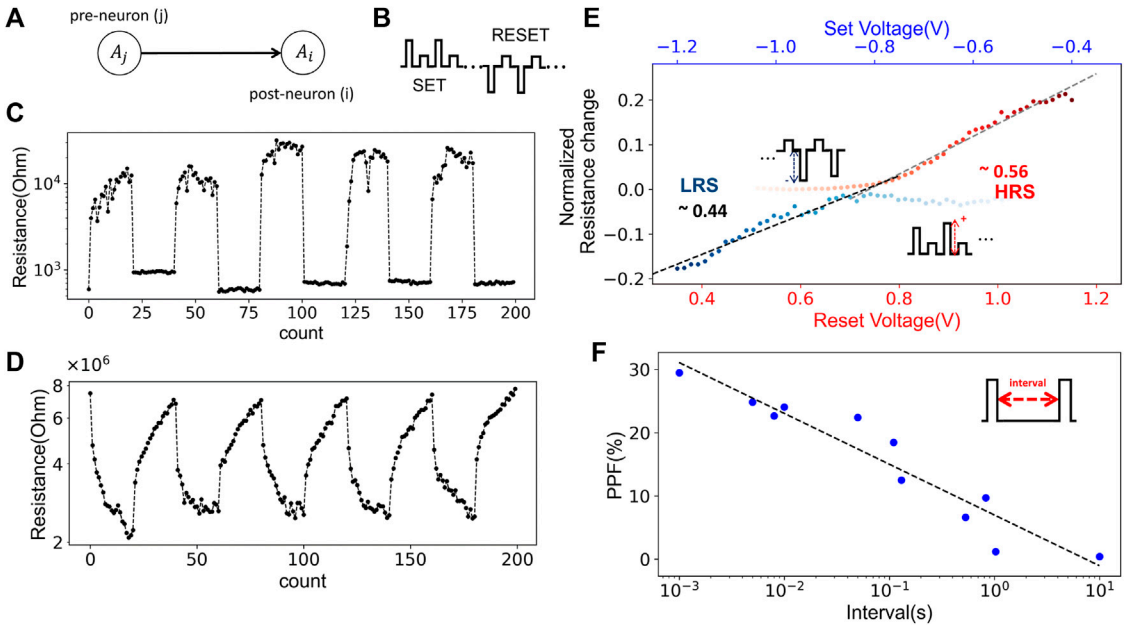


FIGURE 8 | (A) Sketch of synaptic plasticity process. **(B)** The waveform of voltage pulse trains; 20 SET pulses followed by 20 RESET pulses. A readout pulse followed each switching pulse. The resulting change in resistance states under the voltage train is shown in **(C)** for type F, **(D)** for type R. The linearity in resistance change was improved in the type R. **(E)** The resistance as a function of the amplitude of SET/RESET voltages in type F. The amplitude of the applied voltage for a SET (RESET) was decreasing (increasing) to lead a next level in the resistance, and a readout voltage followed each switching voltage pulse. The multistate resistance was observed along with the symmetricity between LRS and HRS. **(F)** PPF as a function of interval time between two sequent switching voltage pulses in type R. The dashed line presents a fitted curve in the experimental data (blue). Increasing the time interval results in the weaker resistance change.

challenge here is to select the correct voltage functions for the pre- and post-neurons, so that an appropriate voltage pulse is applied across the memristive device (Linares-Barranco et al., 2011; Ambrogio et al., 2013). To investigate this for the type R device, we took a closer look at the PPF scheme. The results obtained,therefore, are shown in **Figure 8F**. Two identical sequential SET pulses were applied using different time intervals. In this study, the PPF ratio was defined as the incremental percentage change in the resistance after the first and second pulses. As a result, we found, the longer the time interval is, the smaller the resistance changes with a linear trend. This, particularly, corresponds to the enhanced back diffusion of oxygen ions in R type devices, as discussed above, and resembles well with biology.

4 CONCLUSION

In summary, we have presented two bi-layer TiN/TiO_x/HfO_x/Au memristive devices. Depending on the respective sputtering method, we were able to realize different switching mechanisms. While mobile oxygen ions are responsible for resistance switching in type R devices, oxygen vacancies cause the switching mechanism in type F devices. Using a statistical analysis of the devices and a physical device model, we have investigated the relevant technology and device parameters, and related them to the electronic behavior of the devices. In **Figure 9** these parameters are graphically summarized and their relevance

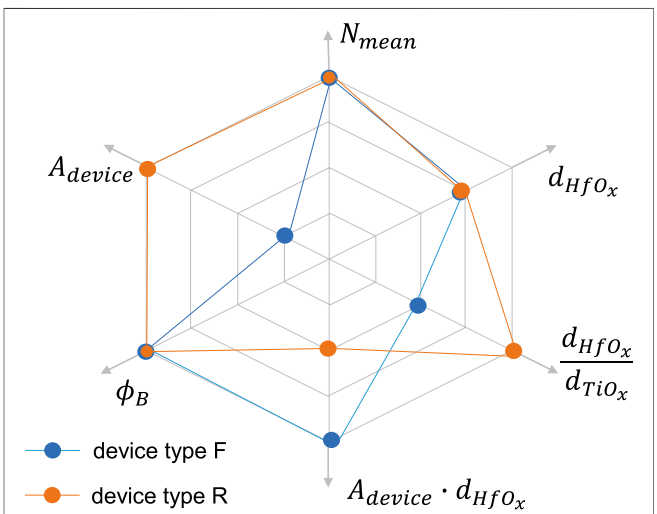


FIGURE 9 | Relevant technology and device parameters to realize a desired resistive switching behavior in the TiN/TiO_x/HfO_x/Au bi-layer memristive devices.

for the respective device type is shown. As you can see from that figure for devices of type R, whose resistive switching is induced by mobile oxygen ions, the device area (A_{device}), the Schottky barrier (ϕ_B), and the ratio d_{HfOx}/d_{TiOx} are important. In devices of type F, whose switching mechanism can be traced back to filamentary oxygen vacancies, also the Schottky barrier (ϕ_B) is

important. But, for that devices the product of A_{device} and d_{HfOx} is more in the focus for good device performance, than A_{device} or d_{HfOx} alone. For both types, however, it is important to adjust the concentration of the mobile charge carriers precisely to reach a reliable performance. In general, it can be concluded that the respective device properties must always be tailored to the specific application. Therefore, we hope that the framework described here helps to identify the relevant technology parameters for that purpose.

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 authors.

AUTHOR CONTRIBUTIONS

SP prepared the samples, JD developed the sputtering technology for the HfOx films, and the characteristics of HfOx films were analyzed using XPS by AK. SP performed the measurements, analyzed the experimental results, and co-wrote the manuscript. SK supervised the electronic measurement. MZ supported the measurements and data interpretation. MZ developed the simulation model. The simulation results were discussed and interpreted between SP, TI, SK, and MZ. TI and MZ conceived the idea, initiated, and supervised the experimental research. SP and MZ discussed the experimental results and contributed to the refinement of the manuscript.

FUNDING

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)–Project-ID 434434223–SFB 1461 and the Carl-Zeiss Foundation via the Project MemWerk.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnano.2021.670762/full#supplementary-material.

REFERENCES

Alibart, F., Zamanidoost, E., and Strukov, D. B. (2013). Pattern Classification by Memristive Crossbar Circuits Using *ex situ* and *in situ* Training. *Nat. Commun.* 4, 1–7. doi:10.1038/ncomms3072

Ambrogio, S., Balatti, S., Nardi, F., Facchinetti, S., and Ielmini, D. (2013). Spike-Timing Dependent Plasticity in a Transistor-Selected Resistive Switching Memory. *Nanotechnology* 24, 384012. doi:10.1088/0957-4484/24/38/384012

Asanuma, S., Akoh, H., Yamada, H., and Sawa, A. (2009). Relationship Between Resistive Switching Characteristics and Band Diagrams of Ti/Pr1−xCaxMnO3Junctions. *Phys. Rev. B.* 80, 235113. doi:10.1103/PhysRevB.80.235113

Bousoulas, P., Asenov, P., Karageorgiou, I., Sakellaropoulos, D., Stathopoulos, S., and Tsoukalas, D. (2016). Engineering Amorphous-Crystalline Interfaces in tio2-x/tio2-y-Based Bilayer Structures for Enhanced Resistive Switching and Synaptic Properties. *J. Appl. Phy.* 120, 154501. doi:10.1063/1.4964872

Chandrasekaran, S., Simanjuntak, F. M., Saminathan, R., Panda, D., and Tseng, T. Y. (2019). Improving Linearity by Introducing Al in hfo2 as a Memristor Synapse Device. *Nanotechnology* 30, 445205. doi:10.1088/1361-6528/ab3480

Clima, S., Chen, Y. Y., Chen, C. Y., Goux, L., Govoreanu, B., Degraeve, R., et al. (2016). First-Principles Thermodynamics and Defect Kinetics Guidelines for Engineering a Tailored RRAM Device. *J. Appl. Phy.* 119, 225107. doi:10.1063/1.4953673

Cüppers, F., Menzel, S., Bengel, C., Hardtdegen, A., von Witzleben, M., Böttger, U., et al. (2019). Exploiting the Switching Dynamics of hfo2-Based Reram Devices for Reliable Analog Memristive Behavior. *APL Mater.* 7, 091105. doi:10.1063/1.5108654

del Valle, J., Ramírez, J. G., Rozenberg, M. J., and Schuller, I. K. (2018). Challenges in Materials and Devices for Resistive-Switching-Based Neuromorphic Computing. *J. Appl. Phy.* 124, 211101. doi:10.1063/1.5047800

Diederich, N., Bartsch, T., Kohlstedt, H., and Ziegler, M. (2018). A Memristive Plasticity Model of Voltage-Based stdp Suitable for Recurrent Bidirectional Neural Networks in the Hippocampus. *Sci. Rep.* 8, 1–12. doi:10.1038/s41598-018-27616-6

Dirkmann, S., Hansen, M., Ziegler, M., Kohlstedt, H., and Mussenbrock, T. (2016). The Role of ion Transport Phenomena in Memristive Double Barrier Devices. *Sci. Rep.* 6. 35686. doi:10.1038/srep1375310.1038/srep35686

Dirkmann, S., Kaiser, J., Wenger, C., and Mussenbrock, T. (2018). Filament Growth and Resistive Switching in Hafnium Oxide Memristive Devices. *ACS Appl. Mater. Inter.* 10, 14857–14868. doi:10.1021/acsami.7b19836

Dittmann, R., and Strachan, J. P. (2019). Redox-Based Memristive Devices for new Computing Paradigm. *APL Mater.* 7, 110903. doi:10.1063/1.5129101

Gao, J., He, G., Deng, B., Xiao, D. Q., Liu, M., Jin, P., et al. (2016). Microstructure, Wettability, Optical and Electrical Properties of hfo2 Thin Films: Effect of Oxygen Partial Pressure. *J. Alloy. Compd.* 662, 339–347. doi:10.1016/j.jallcom.2015.12.080

Gao, S., Zeng, F., Li, F., Wang, M., Mao, H., Wang, G., et al. (2015). Forming-Free and Self-Rectifying Resistive Switching of the Simple pt/tao x/N-Si Structure for Access Device-Free High-Density Memory Application. *Nanoscale* 7, 6031–6038. doi:10.1039/c4nr06406b

Goossens, A. S., Das, A., and Banerjee, T. (2018). Electric Field Driven Memristive Behavior at the Schottky Interface of Nb-Doped srtio3. *J. Appl. Phy.* 124, 152102. doi:10.1063/1.5037965

Govoreanu, B., Redolfi, A., Zhang, L., Adelmann, C., Popovici, M., Clima, S., et al. (2013). "Vacancy-Modulated Conductive Oxide Resistive Ram: an Area-Scalable Switching Current, Self-Compliant, Highly Nonlinear and Wide on/off-Window Resistive Switching Cell," in IEEE International Electron Devices Meeting, Washington, DC, December 9–11, 2013 (New York, NY: IEEE), 10–12.

Hansen, M., Zahari, F., Kohlstedt, H., and Ziegler, M. (2018). Unsupervised Hebbian Learning Experimentally Realized With Analogue Memristive Crossbar Arrays. *Sci. Rep.* 8, 1–10. doi:10.1038/s41598-018-27033-9

Hansen, M., Zahari, F., Ziegler, M., and Kohlstedt, H. (2017). Double-Barrier Memristive Devices for Unsupervised Learning and Pattern Recognition. *Front. Neurosci.* 11, 91. doi:10.3389/fnins.2017.00091

Hansen, M., Ziegler, M., and Kolberg, L. (2015). A Double Barrier Memristive Device. *Sci. Rep.* 5, 13753. doi:10.1038/srep13753

Hardtdegen, A., Torre, C. L., Cüppers, F., Menzel, S., Waser, R., and Hoffmann-Eifert, S. (2018). Improved Switching Stability and the Effect of an Internal Series Resistor in hfo2/tiox Bilayer Reram Cells. *IEEE Trans. Electron Devices* 65, 3229–3236. doi:10.1109/TED.2018.2849872

He, W., Sun, H., Zhou, Y., Lu, K., Xue, K., and Miao, X. (2017). Customized Binary and Multi-Level HfO2−x-Based Memristors Tuned by Oxidation Conditions. *Sci. Rep.* 7, 10070. doi:10.1038/s41598-017-09413-9

Huang, C. H., Huang, J. S., Lin, S. M., Chang, W. Y., He, J. H., and Chueh, Y. L. (2012). Zno1−x Nanorod arrays/zno Thin Film Bilayer Structure: From Homojunction Diode and High-Performance Memristor to Complementary 1d1r Application. *Acs Nano* 6, 8407–8414. doi:10.1021/nn303233r

Hur, J. H., Lee, M.-J., Lee, C. B., Kim, Y.-B., and Kim, C.-J. (2010). Modeling for Bipolar Resistive Memory Switching in Transition-Metal Oxides. *Phys. Rev. B* 82, 155321. doi:10.1103/PhysRevB.82.155321

Ielmini, D., and Ambrogio, S. (2020). Emerging Neuromorphic Devices. *Nanotechnology* 31, 092001. doi:10.1088/1361-6528/ab554b

Ielmini, D., and Milo, V. (2017). Physics-Based Modeling Approaches of Resistive Switching Devices for Memory and in-Memory Computing Applications. *J. Comput. Electron* 16, 1121–1143. doi:10.1007/s10825-017-1101-9

Ignatov, M., Ziegler, M., Hansen, M., and Kohlstedt, H. (2017). Memristive Stochastic Plasticity Enables Mimicking of Neural Synchrony: Memristive Circuit Emulates an Optical Illusion. *Sci. Adv.* 3, e1700849. doi:10.1126/sciadv.1700849

Jeong, H., and Shi, L. (2019). Memristor Devices for Neural Networks. *J. Phys. D* 52, 023003. doi:10.1088/1361-6463/aae223

Jiang, Z., Wu, Y., Yu, S., Yang, L., Song, K., Karim, Z., et al. (2016). A Compact Model for Metal–Oxide Resistive Random Access Memory With Experiment Verification. *IEEE Trans. Electron Devices* 63, 1884–1892. doi:10.1109/ted.2016.2545412

Kendall, J. D., Pantone, R. D., and Nino, J. C. (2020). Data From: Deep Learning in Memristive Nanowire Networks. http://arXiv:2003.02642.

Kim, H. J., Zheng, H., Park, J. S., Kim, D. H., Kang, C. J., Jang, J. T., et al. (2017). Artificial Synaptic Characteristics With Strong Analog Memristive Switching in a pt-ceo2-pt Structure. *Nanotechnology* 28, 285203. doi:10.1088/1361-6528/aa712c

Kim, S. G., Han, J. S., Kim, H., Kim, S. Y., and Jang, H. W. (2018). Recent Advances in Memristive Materials for Artificial Synapses. *Adv. Mater. Technol.* 3, 1800457. doi:10.1002/admt.201800457

Krestinskaya, O., James, A. P., and Chua, L. O. (2020). Neuromemristive Circuits for Edge Computing: a Review. *IEEE Trans. Neural Netw. Learn. Syst.* 31, 4–23. doi:10.1109/TNNLS.2019.2899262

Lee, J., Shin, J., Lee, D., Lee, W., Jung, S., Jo, M., et al. (2010). "Diode-Less Nano-Scale zro x/hfo x rram Device With Excellent Switching Uniformity and Reliability for High-Density Cross-Point Memory Applications," in International Electron Devices Meeting. San Francisco, CA. December 6–8, 2010 (New York, NY: IEEE), 19–25.

Lee, J., Park, J., Jung, S., and Hwang, H. (2011). "Scaling Effect of Device Area and Film Thickness on Electrical and Reliability Characteristics of rram," in International Interconnect Technology Conference. Dresden, Germany. May 8–12 , 2011 (New York, NY: IEEE), 1–3.

Legenstein, R. (2015). Nanoscale Connections for Brain-Like Circuits. *Nature* 521, 37–38. doi:10.1038/521037a

Li, J., Zhang, T., Duan, Q., Li, L., Yang, Y., and Huang, R. (2018a). "Engineering Resistive Switching Behavior in Taox Based Memristive Devices for non-von Neuman Computing Applications," in China Semiconductor Technology International Conference. Shanghai, China. March 11–12, 2018 (New York, NY: IEEE), 1–3.

Li, Y., Wang, Z., Midya, R., Xia, Q., and Yang, J. J. (2018b). Review of Memristor Devices in Neuromorphic Computing: Materials Sciences and Device Challenges. *J. Phys. D* 51, 503002. doi:10.1088/1361-6463/aade3f

Li, Y., Fuller, E. J., Sugar, J. D., Yoo, S., Ashby, D. S., Bennett, C. H., et al. (2020). Filament-Free Bulk Resistive Memory Enables Deterministic Analogue Switching. *Adv. Mater.* 32, 2003984. doi:10.1002/adma.202003984

Lin, P., Li, C., Wang, Z., Li, Y., Jiang, H., Song, W., et al. (2020). Three-dimensional Memristor Circuits as Complex Neural Networks. *Nat. Electron.* 3, 225–232. doi:10.1038/s41928-020-0397-9

Linares-Barranco, B., Serrano-Gotarredona, T., Camuñas-Mesa, L. A., Perez-Carrasco, J. A., Zamarreño-Ramos, C., and Masquelier, T. (2011). On Spike-Timing-Dependent-Plasticity, Memristive Devices, and Building A Self-Learning Visual Cortex. *Front. Neurosci.* 5, 26. doi:10.3389/fnins.2011.00026

Ma, H., Feng, J., Lv, H., Gao, T., Xu, X., Luo, Q., et al. (2017). Self-Rectifying Resistive Switching Memory with Ultralow Switching Current in Pt/Ta2O5/HfO2-x/Hf Stack. *Nanoscale Res. Lett.* 12, 1–6. doi:10.1186/s11671-017-1905-3

Martínez, F. L., Toledano-Luque, M., Gandía, J. J., Cárabe, J., Bohne, W., Röhrich, J., et al. (2007). Optical Properties and Structure of hfo2thin Films Grown by High Pressure Reactive Sputtering. *J. Phys. D* 40, 5256. doi:10.1088/0022-3727/40/17/037

Massimiliano, B., and Yuriy, P. (2013). On the Physical Properties of Memristive, Memcapacitive and Meminductive Systems. *Nanotechnology* 24, 7. doi:10.1088/0957-4484/24/25/255201

McKenna, K. P. (2014). Optimal Stoichiometry for Nucleation and Growth of Conductive Filaments in hfox. *Model. Simul. Mater. Sci. Eng.* 22, 025001. doi:10.1088/0965-0393/22/2/025001

Menzel, S., Waters, M., Marchewka, A., Böttger, U., Dittmann, R., and Waser, R. (2011). Origin of the Ultra-Nonlinear Switching Kinetics in Oxide-Based Resistive Switches. *Adv. Funct. Mater.* 21, 4487–4492. doi:10.1002/adfm.201101117

Mikhaylov, A., Belov, A., Korolev, D., Antonov, I., Kotomina, V., Kotina, A., et al. (2020). Multilayer Metal-Oxide Memristive Device with Stabilized Resistive Switching. *Adv. Mater. Technol.* 5, 1900607. doi:10.1002/admt.201900607

Mikheev, E., Hoskins, B. D., and Strukov, D. B. (2014). Resistive Switching and its Suppression in pt/nb:srtio3 Junctions. *Nat. Commun.* 5, 3990. doi:10.1038/ncomms4990

Mohammad, B., Jaoude, M. A., Kumar, V., Al Homouz, D. M., Nahla, H. A., Al-Qutayri, M., et al. (2016). State of the art of Metal Oxide Memristor Devices. *Nanotechno. Rev.* 5, 311–329. doi:10.1515/ntrev-2015-0029

Niu, D., Chen, Y., Xu, C., and Xie, Y. (2010). "Impact of Process Variations on Emerging Memristor". in Proceedings of the 47th Design Automation Conference. San Francisco, CA. July 13–18, 2010 (DAS), (New York, NY: IEEE) 877–882.

Panwar, N., Rajendran, B., and Ganguly, U. (2017). Arbitrary Spike Time Dependent Plasticity in Memristor by Analog Waveform Engineering. *IEEE Electron Device Lett.* 38, 740–743. doi:10.1109/LED.2017.2696023

Park, T. H., Song, S. J., Kim, H. J., Kim, S. G., Chung, S., Kim, B. Y., et al. (2015). Thickness Effect of Ultra-Thin Ta 2 O 5 Resistance Switching Layer in 28 nm-Diameter Memory Cell. *Sci. Rep.* 5, 15965. doi:10.1038/srep15965

Payvand, M., Muller, L. K., and Indiveri, G. (2018). "Event-Based Circuits for Controlling Stochastic Learning with Memristive Devices in Neuromorphic Architectures," in International Symposium on Circuits and Systems. Florence, Italy. May 27–30, 2018 (New York, NY: IEEE), 1–5.

Pei, J.-S., Wright, J. P., Todd, M. D., Masri, S. F., and Gay-Balmaz, F. (2015). Understanding Memristors and Memcapacitors in Engineering Mechanics Applications. *Nonlinear Dyn.* 80, 457–489. doi:10.1007/s11071-014-1882-3

Prezioso, M., Merrikh-Bayat, F., Hoskins, B. D., Adam, G. C., Likharev, K. K., and Strukov, D. B. (2015). Training and Operation of an Integrated Neuromorphic Network Based on Metal-Oxide Memristors. *Nature* 521, 61–64. doi:10.1038/nature14441

Sah, M. P., Kim, H., and Chua, L. (2014). Brains are Made of Memristors. *IEEE Circuits Syst. Mag.* 14, 12–36. doi:10.1109/MCAS.2013.2296414

Snider, G. S. (2008). "Spike-Timing-Dependent Learning in Memristive Nanodevices," in International Symposium on Nanoscale Architectures. Anaheim, CA. June 12–13, 2008 (New York, NY: IEEE), 85–92.

Solan, E., Dirkmann, S., Hansen, M., Schroeder, D., Kohlstedt, H., Ziegler, M., et al. (2017). An Enhanced Lumped Element Electrical Model of a Double Barrier Memristive Device. *J. Phys. D* 50, 195102. doi:10.1088/1361-6463/aa69ae

Stathopoulos, S., Khiat, A., Trapatseli, M., Cortese, S., Serb, A., Valov, I., et al. (2017). Multibit Memory Operation of Metal-Oxide Bi-Layer Memristors. *Sci. Rep.* 7, 17532. doi:10.1038/s41598-017-17785-1

Sun, W., Gao, B., Chi, M., Xia, Q., Yang, J. J., Qian, H., et al. (2019). Understanding Memristive Switching via in Situ Characterization and Device Modeling. *Nat. Commun.* 10, 1–13. doi:10.1038/s41467-019-11411-6

Sze, S., and Ng, K. K. (2006). *Physics of Semiconductor Devices*. Hoboken, New Jersey: John Wiley & Sons, Ltd.

Tsuruoka, T., Terabe, K., Hasegawa, T., Valov, I., Waser, R., and Aono, M. (2012). Effects of Moisture on the Switching Characteristics of Oxide-Based, Gapless-Type Atomic Switches. *Adv. Funct. Mater.* 22 70–77. doi:10.1002/adfm.201101846

Versace, M., and Chandler, B. (2010). The Brain of a new Machine. *IEEE Spectr.* 47, 30–37. doi:10.1109/MSPEC.2010.5644776

Wang, C., Xiong, L., Sun, J., and Yao, W. (2019). Memristor-Based Neural Networks With Weight Simultaneous Perturbation Training. *Nonlinear Dyn.* 95, 2893–2906. doi:10.1007/s11071-018-4730-z

Wang, M., Luo, W. J., Wang, Y. L., Yang, L. M., Zhu, W., Zhou, P., et al. (2010). "A Novel Cuxsiyo Resistive Memory in Logic Technology with Excellent Data Retention and Resistance Distribution for Embedded Applications," in Symposium on VLSI Technology. Honolulu, HI. June 15–17, 2010 (New York, NY: IEEE), 89–90.

Wang, R., Yang, J.-Q., Mao, J.-Y., Wang, Z.-P., Wu, S., Zhou, M., et al. (2020). Recent Advances of Volatile Memristors: Devices, Mechanisms,

and Applications. *Adv. Intell. Sys.* 2, 2000055. doi:10.1002/aisy.202000055

Wang, Y.-F., Lin, Y.-C., Wang, I.-T., Lin, T.-P., and Hou, T.-H. (2015). Characterization and Modeling of Nonfilamentary Ta/TaOx/TiO2/Ti Analog Synaptic Device. *Sci. Rep.* 5, 10150. doi:10.1038/srep10150

Wang, Z., Yin, M., Zhang, T., Cai, Y., Wang, Y., Yang, Y., et al. (2016). Engineering Incremental Resistive Switching in Taoxbased Memristors for Brain-Inspired Computing. *Nanoscale* 8, 14015–14022. doi:10.1039/C6NR00476H

Xiong, W., Zhu, L. Q., Ye, C., Yu, F., Ren, Z. Y., and Ge, Z. Y. (2019). Bilayered Oxide-Based Cognitive Memristor with Brain-Inspired Learning Activities. *Adv. Electron. Mater.* 5, 1900439. doi:10.1002/aelm.20190043

Yakopcic, C., Hasan, R., and Taha, T. M. (2015). "Memristor Based Neuromorphic Circuit for *ex-situ* Training of Multi-Layer Neural Network Algorithms," in International Joint Conference on Neural Networks. July 12–17, 2015 (New York, NY: IEEE), 1–7.

Yang, C.-H., Kuo, Y., and Lin, C.-H. (2010). Charge Detrapping and Dielectric Breakdown of Nanocrystalline Zinc Oxide Embedded Zirconium-Doped Hafnium Oxide High-K Dielectrics for Nonvolatile Memories. *Appl. Phys. Lett.* 96, 192106. doi:10.1063/1.3429590

Yao, P., Huaqiang, W., Wu, H., Tang, J., Zhang, Q., Zhang, W., et al. (2020). Fully Hardware-Implemented Memristor Convolutional Neural Network. *Nature* 577, 641–646. doi:10.1038/s41586-020-1942-4

Yin, X.-B., Tan, Z.-H., and Guo, X. (2015). The Role of Schottky Barrier in the Resistive Switching of SrTiO3: Direct Experimental Evidence. *Phys. Chem. Chem. Phys.* 17, 134–137. doi:10.1039/c4cp04151h

Yoon, J. H., Song, S. J., Yoo, I.-H., Seok, J. Y., Yoon, K. J., Kwon, D. E., et al. (2014). Highly Uniform, Electroforming-Free, and Self-Rectifying Resistive Memory in the pt/ta2o5/hfo2-x/tin Structure. *Adv. Funct. Mater.* 24, 5086–5095. doi:10.1002/adfm.201400064

Zhao, M., Gao, B., Tang, J., Qian, H., and Wu, H. (2020). Reliability of Analog Resistive Switching Memory for Neuromorphic Computing. *Appl. Phy. Rev.* 7, 011301. doi:10.1063/1.5124915

Zhou, G., Duan, S., Li, P., Sun, B., Wu, B., Yao, Y., et al. (2018). Coexistence of Negative Differential Resistance and Resistive Switching Memory at Room Temperature in Tio x Modulated by Moisture. *Adv. Electron. Mater.* 4, 1700567. doi:10.1002/aelm.201700567

Zhou, G., Ren, Z., Sun, B., Wu, J., Zou, Z., Zheng, S., et al. (2020). Capacitive Effect: an Original of the Resistive Switching Memory. *Nano Energy* 68, 104386. doi:10.1016/j.nanoen.2019.104386

Zhou, G., Sun, B., Yao, Y., Zhang, H., Zhou, A., Alameh, K., et al. (2016). Investigation of the Behaviour of Electronic Resistive Switching Memory Based on Mose2-Doped Ultralong se Microwires. *Appl. Phys. Lett.* 109, 143904. doi:10.1063/1.4962655

Ziegler, M., Riggert, C., Hansen, M., Bartsch, T., and Kohlstedt, H. (2015). Memristive Hebbian Plasticity Model: Device Requirements for the Emulation of Hebbian Plasticity Based on Memristive Devices. *IEEE Trans. Biomed. Circuits Syst.* 9, 197–206. doi:10.1109/tbcas.2015.2410811

Ziegler, M., Wenger, C., Chicca, E., and Kohlstedt, H. (2018). Tutorial: Concepts for Closely Mimicking Biological Learning With Memristive Devices: Principles to Emulate Cellular Forms of Learning. *J. Appl. Phy.* 124, 152003. doi:10.1063/1.5042040

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.

frontiers in Nanotechnology

REVIEW
published: 28 April 2021
doi: 10.3389/fnano.2021.654418

Modeling-Based Design of Memristive Devices for Brain-Inspired Computing

Yudi Zhao[1,2]*, *Ruiqi Chen*[2], *Peng Huang*[2] and *Jinfeng Kang*[2]*

[1] Key Laboratory of the Ministry of Education for Optoelectronic Measurement Technology and Instrument, Beijing Information Science and Technology University, Beijing, China, [2] Institute of Microelectronics, Peking University, Beijing, China

Resistive switching random access memory (RRAM) has emerged for non-volatile memory application with the features of simple structure, low cost, high density, high speed, low power, and CMOS compatibility. In recent years, RRAM technology has made significant progress in brain-inspired computing paradigms by exploiting its unique physical characteristics, which attempts to eliminate the energy-intensive and time-consuming data transfer between the processing unit and the memory unit. The design of RRAM-based computing paradigms, however, requires a detailed description of the dominant physical effects correlated with the resistive switching processes to realize the interaction and optimization between devices and algorithms or architectures. This work provides an overview of the current progress on device-level resistive switching behaviors with detailed insights into the physical effects in the resistive switching layer and the multifunctional assistant layer. Then the circuit-level physics-based compact models will be reviewed in terms of typical binary RRAM and the emerging analog synaptic RRAM, which act as an interface between the device and circuit design. After that, the interaction between device and system performances will finally be addressed by reviewing the specific applications of brain-inspired computing systems including neuromorphic computing, in-memory logic, and stochastic computing.

Keywords: memristive devices, RRAM, physics-based models, brain-inspired computing, neuromorphic computing, computing in-memory, stochastic computing

OPEN ACCESS

Edited by:
Huanglong Li,
Tsinghua University, China

Reviewed by:
Peng Yao,
Tsinghua University, China
Vishal Saxena,
University of Delaware, United States

***Correspondence:**
Yudi Zhao
zhaoyd@pku.edu.cn
Jinfeng Kang
kangjf@pku.edu.cn

Specialty section:
This article was submitted to Nanodevices, a section of the journal Frontiers in Nanotechnology

Received: *16 January 2021*
Accepted: *24 February 2021*
Published: *28 April 2021*

Citation:
Zhao Y, Chen R, Huang P and Kang J (2021) Modeling-Based Design of Memristive Devices for Brain-Inspired Computing. Front. Nanotechnol. 3:654418. doi: 10.3389/fnano.2021.654418

INTRODUCTION

In the 1960s, the resistive switching phenomenon in metal–insulator–metal structure was first reported by Hickmott in binary oxides (Hickmott, 1962). As the development of material processing and device integration technologies, the research into the resistive switching in memristive devices was revived in the late 1990s (Asamitsu et al., 1997; Sawa, 2008; Waser et al., 2009; Wong et al., 2012; Yang et al., 2013; Pan et al., 2014; Jeong et al., 2016; Wu H. et al., 2017). The resistive switching random access memory (RRAM) are widely investigated in recent years for their potential to be used as a promising candidate for non-volatile memories (Asamitsu et al., 1997; Sawa, 2008; Waser et al., 2009; Wong et al., 2012). A typical RRAM device consists of a metal oxide-resistive switching layer sandwiched between two electrodes. The resistance of the device can be switched reversibly between the high-resistance state (HRS) and the low resistance state (LRS). Up to now, significant technical advances have been achieved in the device performance of RRAM, including great scalability (<10 nm), fast speed (<1 ns), low operation voltage (<1.5 V) and current

(<1 μA), high endurance (>10^{12} cycles), an long retention (>10 years at room temperature for binary state RRAM) (Lee et al., 2008, 2010, 2012; Chen et al., 2009; Chien et al., 2010; Govoreanu et al., 2011; Wang et al., 2012; Li K. S. et al., 2014).

So far, to reveal the origins of resistive switching in RRAM, a large variety of physical mechanisms have been proposed leading to the resistive switching effects such as oxygen vacancy (Vo) generation and recombination, ion migration, charge trapping and de-trapping, thermal reaction, insulator-to-metal transition, charge transfer, and so on (Russo et al., 2007; Wei et al., 2008; Degraeve et al., 2010; Kwon et al., 2010; Goux et al., 2011; Kang et al., 2015). Multiple experimental techniques have been utilized, so far, in order to identify the resistive switching mechanism such as high-resolution X-ray photoelectron spectroscopy (XPS), scanning electron microscopy (SEM), conductive atomic force microscopy (C-AFM), and transmission electron microscopy (TEM) (Baek et al., 2004; Janousch et al., 2007; Yun et al., 2007; Yang et al., 2012). These techniques are widely used in the conductive-bridge random access memory (CBRAM) with fruitful findings. However, for the metal oxide-based RRAM, it is difficult to directly observe the Vo defects. It is now commonly accepted that the switching behavior in metal oxide-based RRAM is due to the formation and rupture of the conductive filament (CF) composed of Vo in the resistive switching layer (Sawa, 2008; Waser et al., 2009; Wong et al., 2012; Pan et al., 2014; Wu H. et al., 2017).

In the early work, the RRAM devices with single resistive switching layer are widely studied. The typical binary oxides that exhibit resistive switching characteristics includes HfOx, Al_2O_3, TaOx, TiOx, and NiO (Sawa, 2008; Waser et al., 2009; Wong et al., 2012; Yang et al., 2013; Pan et al., 2014; Jeong et al., 2016; Wu H. et al., 2017). To specifically optimize the device performance, RRAM devices with multi-layer electrolyte stack are also proposed and investigated such as HfOx/Al_2O_3, Ta_2O_5/TaOx, and HfOx/TaOx, where one electrolyte layer acts as the resistive switching layer, and the other acts as an assistant layer to enhance the performance. After inserting an assistant layer, the device uniformity and reliability can be improved, and other additional function such as self-compliance, self-rectifying, and even analog switching can be realized (Lee et al., 2011; Hsu et al., 2014; Azzaz et al., 2015; Chou et al., 2015; Zhao et al., 2015, 2016; Woo et al., 2016a; Wu W. et al., 2017; Wu et al., 2018). Compared with the typical binary switching with two stable resistance states, analog switching is an attractive device property to mimic the function of biological synapse.

Due to the unique characteristics, RRAM has been suggested for use as building blocks for brain-inspired computing systems (Yang et al., 2013; Philip Wong and Salahuddin, 2015; Chi et al., 2016; Jeong et al., 2016; Yu, 2018). The brain-inspired computing paradigms are highly desired to overcome the bottleneck of the so-called "memory wall" from the traditional von Neumann architecture. The brain-inspired computing aims to carry out calculations where the data are located, which is similar to the information processing in the human brain. The RRAM electrical characteristics can mimic the signal processing of biological synapse, making it feasible to be applied into neuromorphic applications to perform energy-efficient, fault-tolerant, and highly parallel computing tasks (Yu et al., 2012; Gao et al., 2014, 2016; Prezioso et al., 2015; Wang et al., 2017). RRAM was also proposed and demonstrated to implement the stateful logic, in which Boolean logic states were operated and stored in the resistance of RRAM (Borghetti et al., 2010; Li et al., 2015a; Huang P. et al., 2016). With the feature of inherent variability, RRAM shows great potential to be used as low-cost and energy-efficient stochastic number generator enabling stochastic computing, which emulates the generation of neural spikes processed by the human brain in the form of long sequences of noisy voltage spikes (Gaba et al., 2013; Suri et al., 2013; Knag et al., 2014; Moons and Verhelst, 2014; Ielmini and Wong, 2018; Wang et al., 2018; Carboni and Ielmini, 2019; Zhao et al., 2019). For the design and optimization of these brain-inspired computing systems, related physics-based models and simulation platforms have been developed to bridge the link between device, circuit, and system, which aims to meet the requirement for the device–circuit–system co-design (Gao et al., 2011; Guan et al., 2012; Huang et al., 2013, 2017, 2018; Chen et al., 2017; Larcher et al., 2017; Pedretti et al., 2017; Zhao et al., 2019; Cai et al., 2020; Liao et al., 2020).

In this work, we will review the latest advances in the design and optimization of metal oxide-based RRAM in the applications of brain-inspired computing systems based on physics-based models. First, the physical effects in both the resistive switching layer and the multifunctional assistant layer of RRAM are discussed in the *Physical Effects of Resistive Switching Behaviors in Resistive Switching Random Access Memory* section. Then, the physics-based compact models of typical binary RRAM and the analog synaptic RRAM are presented in the *Physics-Based Compact Models of Resistive Switching Random Access Memory* section. In the *Applications in Brain-inspired Computing* section, the design and optimization of system applications of RRAM in novel brain-inspired computing paradigms are explored. The review will be concluded with a short summary and future prospect.

PHYSICAL EFFECTS OF RESISTIVE SWITCHING BEHAVIORS IN RESISTIVE SWITCHING RANDOM ACCESS MEMORY

Understanding the dominant physical effects in the resistive switching behaviors in metal oxide RRAM is crucial for designing and optimizing the device performance. In this section, we will first address the physical effects correlated with the resistive switching layer in detail, and then discuss the various functions of assistant layers in the bilayer device.

Physical Effects in the Resistive Switching Layer

The resistive switching of the metal oxide RRAM has been attributed to the filamentary modification of conduction properties since the early 2000s (Waser et al., 2009; Wong et al., 2012; Pan et al., 2014). To reveal the physical effects and the resistive switching mechanism of Ox-RRAM, multiple experimental techniques have been utilized. For the metal oxide-based RRAM, although it is difficult to directly observe the

Vo defects, the resistive switching behaviors can be detected by the change in electrostatic potential distribution through *in situ* electron holography, which is based on the change of transmitted electron wave phase triggered by the accumulated charges in the sample (Li et al., 2017). This is because the electrons traveling along the CF would change the potential of the HfOx layer. The *in situ* low-energy-filtered images can then be used to describe the change in oxygen concentrations in HfOx layer. Based on this technique, the bias-induced phases featuring $\Delta\varphi^{bias}(x,y)$ of the TiN/HfOx/AlOy/Pt structure in the forming process are shown in **Figure 1A**. During the forming process, positive bias is applied to the TiN top electrode (TE), and the increasing bias would enhance the positive potential with the most positive charges aggregated near the interface between the HfOx and AlOy layers. With the bias increasing over 3 V, the potential of the AlOy layer changes to nearly zero and then becomes negative. At the same time, in the lower half of the HfOx layer, a negative potential emerges and then diffuses vertically toward TE. The positive charges originated from Vo, while the negative potential can be attributed to the transport electrons residual in the migration path, which can be used to track the CF formation process in the HfOx layer. The RESET process can also be monitored by the hologram images similarly, which demonstrates that the CF starts to rupture from the interface of TE and the HfOx layer. Based on the above experimental results, the CFs in the resistive switching layer are formed due to the fact that Vo are generated and ruptured at the top interface of the HfOx layer.

To explain the physical origin of generation and rupture of the CF, multiple switching mechanisms have been proposed in recent years (Russo et al., 2007; Wei et al., 2008; Degraeve et al., 2010; Kwon et al., 2010; Goux et al., 2011; Kang et al., 2015). Combining with the experimental evidence, one widely accepted physical mechanism is the generation and combination of Vo with O^{2-} (Gao et al., 2011; Guan et al., 2012; Huang et al., 2013; Kang et al., 2015). Based on the mechanism, the microscopic physical processes of switching of the typical TiN/HfOx/Pt device are shown in **Figure 1B**. In the SET process, O^{2-} are ionized from the HfOx lattice accompanied by the generation of Vo. The O^{2-} will be driven toward TE under the electric field and restored at the oxygen reservoir, which is the TiN electrode in the TiN/HfOx/Pt structure. The probability of above microscopic processes can be described as Guan et al. (2012):

$$P_g = f \cdot \exp\left(-\frac{E_0 - \Delta\varphi}{k_B T}\right) \quad (1)$$

where f is the vibration frequency of the oxygen atom, E_0 denotes the average active energy of V_O generation or O^{2-} hopping, $\Delta\phi$ is the barrier height reduction induced by the electric field, and T is the local temperature. In the RESET process, the electrons in the vicinity of Vo are depleted under the electric field, and then the positively charged Vo would recombine with the dissociated O^{2-} released by the oxygen reservoir. The recombination of Vo and O^{2-} finally results in the rupture of CF.

For other resistive switching materials, such as TiO_2 and Ta_2O_5, the phase transition also takes place during the resistive switching (Wei et al., 2008; Kang et al., 2015). The phase transitions in TiO_2 and Ta_2O_5 were calculated by *ab initio* calculations as shown in **Figure 1C** (Kang et al., 2015). In the Ta_2O_5-based RRAM, the phase transitions take place between Ta_2O_5 and TaO_2, and Ta_2O_5 is semiconductive, while TaO_2 is metallic. During the resistive switching, the CF is composed of both Vo and TaO_2. Although the effects of Vo generation/recombination and phase transition coexist during switching, the Vo generation/recombination is the dominant effect based on the device simulation results (Zhao et al., 2016).

Based on the basic principle of Vo generation and recombination, the bipolar and unipolar switching characteristics can be explained by a unified model (Gao et al., 2011). Their physical origins of CF formation and rupture between the bipolar switching and unipolar switching are roughly similar. The difference is the location that stores and releases O^{2-}. In the unipolar RRAM, the dissociated O^{2-} would be absorbed or released by the easily reduced oxide clusters near the CF, and several different phases of oxide clusters coexist in the electrolyte material. The O^{2-} will be thermally activated and recombine with the neighbor Vo in the RESET process. For both bipolar and unipolar RRAM, the electron transport in the CF is metallic, and the conductivity decreases with increasing temperature following the Arrhenius law (Ielmini et al., 2010). In the region with low Vo concentration, the electrons hop among the dispersive Vo, and the hopping rate can be calculated by the Mott hopping model (Mott and Davis, 1972). Therefore, the I–V characteristics are nonlinear for the HRS device as shown in **Figure 1D**. Based on the physical effects of resistive switching, the kinetic Monte Carlo simulations can be performed to investigate the switching dynamics in atomic scale. **Figure 1D** shows the CF evolution processes during RESET and SET processes (Huang et al., 2013). In the RESET process, the CF first ruptures at the interface between the TiN and HfOx layer, and then the gap region enlarges gradually. In the SET process, a thin CF first connects the electrode and residual CF, and then the thin CF would grow along the radius direction.

Even the filament effect and the correlated physical effects have been widely accepted for resistive switching, the direct experiment evidences of the physical effects in microscopic characterizations are still lacking. Future breakthroughs in atomic level characterization technologies may finally help people to clarify the underlying physical origins.

Device Optimization With Multifunctional Assistant Layer

The RRAM characteristics can be improved or modified by inserting an assistant layer adjacent with the resistive switching layer, which composes a multifunctional electrolyte stack. A typical example is the Ta_2O_5/TaO_X bilayer stack, which aims to improve the endurance characteristics (Wei et al., 2008; Lee et al., 2011). In the Ta_2O_5/TaO_X stack, the oxygen-deficient TaO_X layer, instead of TiN electrode in the HfO_X-RRAM, acts as the oxygen reservoir. The generated O^{2-} in the SET process would be absorbed by the TaOx layer, in which part of O^{2-} will continue hopping in the TaOx layer under the electric field,

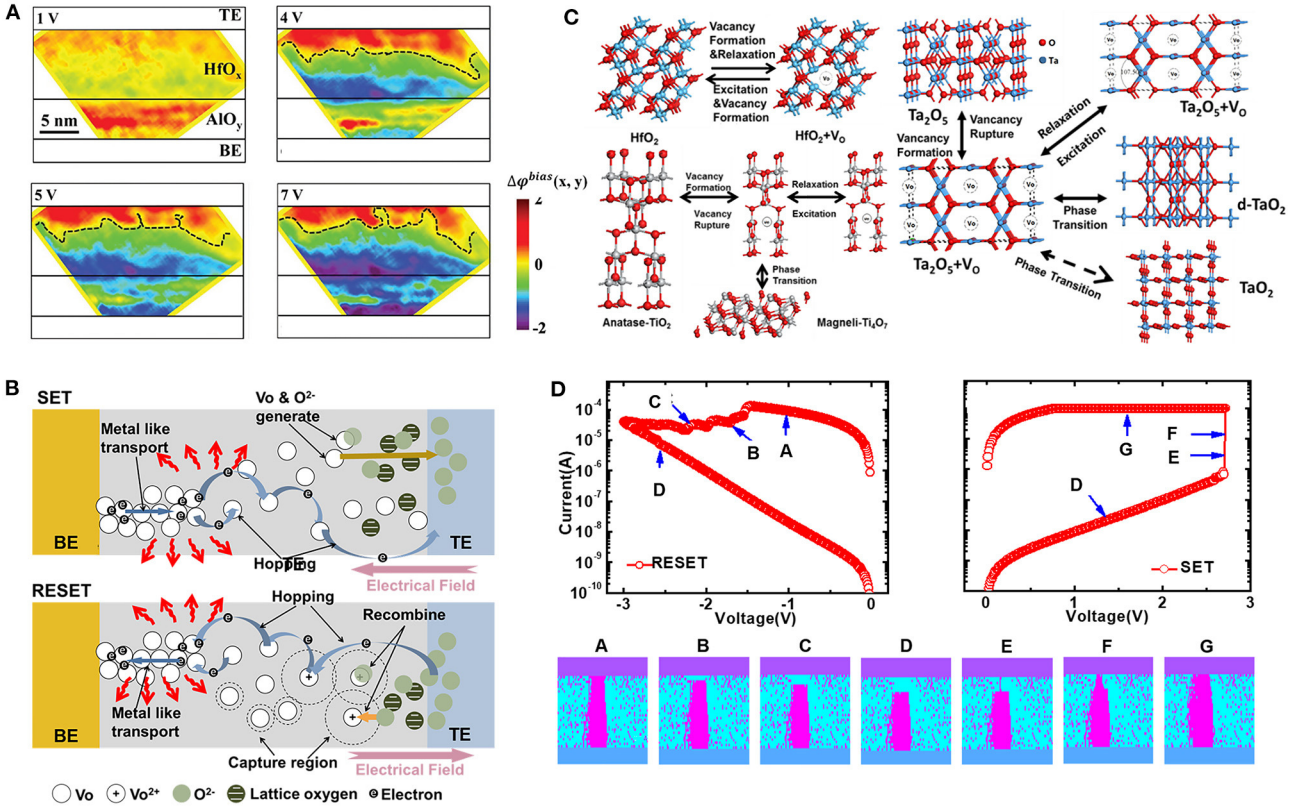


FIGURE 1 | (A) The bias-induced phases featuring $\Delta\varphi^{bias}$(x,y) in the forming process of HfOx- resistive switching random access memory (RRAM). The intrinsic inner potential is removed by using the phase image of the pristine sample without bias. The dash curve is the boundary between the positive and negative phases. TE, top electrode; BE, bottom electrode. **(B)** Physical effects during SET and RESET processes. **(C)** *Ab initio* calculation results for various metal oxide materials such as HfO_2, TiO_2, and Ta_2O_5. **(D)** I–V curves and the corresponding CF evolutions simulated by the kinetic Monte Carlo method during RESET and SET processes. Reprinted from Kang et al. (2015), Huang et al. (2013), and Li et al. (2017).

while the rest will take the redox reaction with the oxygen-deficient TaOx and be stored as lattice oxygen. The oxygen concentration in the TaOx layer increases as O^{2-} gradually oxidizes TaOx, leading to the resistance increase in the TaOx assistant layer. In this way, the current during the SET process can be adjusted dynamically and prevented from being too large. This can explain the self-compliance behavior observed in measured I–V characteristics in Ta_2O_5/TaO_X-based RRAM as shown in **Figure 2A** (Zhao et al., 2016). Besides that, one remarkable characteristic of Ta_2O_5/TaO_X-based RRAM is the superior endurance performance. The endurance can reach up to 10^{12} as shown in **Figure 2A** (Lee et al., 2011). Moreover, the endurance can be enhanced when choosing lower oxygen partial pressure during the deposition of TaOx. The enhanced endurance can be attributed to the capability of TaOx to take redox reactions with O^{2-}, which can then be stored concentrated near CF in the TaOx layer. **Figure 2A** schematically shows the endurance model in the bi-layered TaOx-based RRAM (Zhao et al., 2015). During the resistive switching process, the concentrated distribution of absorbed oxygen guarantees the sufficient supply of O^{2-} in each RESET cycle, otherwise, the O^{2-} would distribute more dispersively in the oxygen reservoir. If the TaOx material is easy to take redox reactions with O^{2-}, the endurance can be highly enhanced, otherwise the endurance behavior would be degraded.

For HfOx-based RRAM, recent studies demonstrated that by introducing a thin Al_2O_3 layer into the HfO_2-based RRAM devices, the switching uniformity, memory window, as well as the operating current can be improved compared with the single-layer HfOx RRAM (Yu et al., 2011; Goux et al., 2012; Azzaz et al., 2015). **Figure 2B** shows the LRS and HRS retention behaviors for the HfO_2/Al_2O_3 device at 200°C. The comparison between the retention of HfO_2 and HfO_2/Al_2O_3 are also shown in **Figure 2B**. The insertion of the Al_2O_3 assistant layer greatly improves the device thermal stability. This can be explained by the increase in Vo diffusion barrier due to the incorporation of Al into the HfO_2 matrix.

The assistant layer can also help the device realize self-rectifying property. Due to the sneak current issue in the RRAM crossbar array, the maximum array size is limited, which requires an additional selector to suppress the current crosstalk. One solution to reduce the cell area and fabrication complexity is to construct a RRAM device with highly non-linear I–V characteristics, which is also known as selector-less or self-rectifying. A Ta/TaOx/TiO_2/Ti RRAM cell is constructed with a

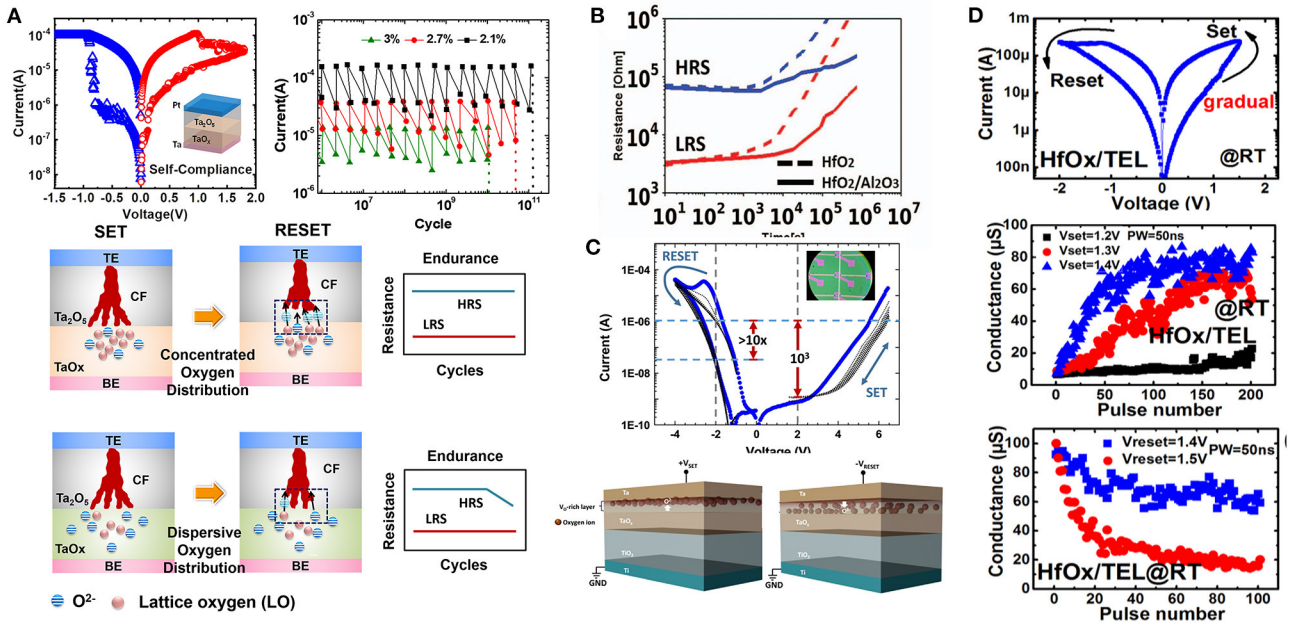


FIGURE 2 | (A) The self-compliance and enhanced endurance characteristics of Ta_2O_5/TaO_X-based RRAM and its schematic endurance model. **(B)** The retention characteristics at 200°C of the HfO_2/Al_2O_3 device and its comparison with the single-layer HfO_2 device. **(C)** Self-rectifying characteristics in the $TaOx/TiO_2$-based RRAM with 10^3 rectifying ratio and the schematic mechanism. **(D)** Analog switching behavior under both SET and RESET pulses in the HfOx/TaOx RRAM. Reprinted from Lee et al. (2011), Azzaz et al. (2015), Chou et al. (2015), Zhao et al. (2015, 2016), Wu W. et al. (2017).

high self-rectifying ratio up to 10^3 for sneak current suppression (Chou et al., 2015). **Figure 2C** shows the I–V characteristics of the proposed device. No obvious SET transition is observed during the switching from HRS to LRS. Compared with a positive-bias current at 2 V, the device shows a three-order rectifying ratio at +2 V and −2 V. Different from the filamentary switching in a single-layer device, the switching mechanism in the $TaOx/TiO_2$-based device can be attributed to the O^{2-} migration under the electric field and the Schottky barrier modulation at the Ta/TaOx interface as shown in **Figure 2C**.

Compared with the abovementioned binary RRAM with two stable resistance states, the analog RRAM with hundreds of resistance levels is an attractive device to mimic the function of biological synapse for neuromorphic computing. A gradual resistance change requires analog modulation of CF evolutions, while it contrasts with the presence of the gap, as the current depends exponentially on the band offset and thickness of the gap. Another issue that contrasts the analog switching is the exponential dependence of physical effects on the field (Larcher et al., 2017). Mitigating the strong field dependence is the key to achieve analog switching, which can be achieved by introducing an assistant layer in the device. Several methods have been used to form the assistant layer such as introducing an AlOx layer in the HfOx-based RRAM (Woo et al., 2016a; Chuang et al., 2019), introducing a SiO_2 layer at the TiN/TaOx interface (Wang et al., 2016), insertion of a TiO_2 layer in the TaOx/Ti interface (Gao et al., 2015), and the Ar plasma treatment at the Ti/HfO_2 interface (Ku et al., 2019). **Figure 2D** shows the analog switching behavior by introducing an oxygen-deficient TaOx layer in the HfOx/Ti RRAM cell at room temperature (Wu W. et al., 2017; Wu et al., 2018). For the HfOx/Ti RRAM cell, the experimental measurements indicate that when increasing the temperature in the HfOx layer, the abrupt switching changes to analog switching due to the thermal effect. Based on this principle, a thermal enhanced layer is designed with less thermal conductivity than metal, therefore it will confine the heat in the HfOx switching layer. In the HfOx/TaOx RRAM, the DC I–V characteristics exhibits gradual current change in both SET and RESET processes. For the operation scheme of identical pulses, the gradual conductance modulations are achieved in both SET and RESET processes as shown in **Figure 2D**. Besides the thermal effect, simulations also show that the slower diffusion of O^{2-} in the bi-layer device would benefit the gradual resistance change (Larcher et al., 2017). The slower diffusion is due to the lower electric field within the oxygen reservoir layer, originated by the voltage distribution and the lower dielectric constant of the assistant layer compared with the resistive switching layer. Therefore, a careful thermal and electric design is required to achieve analog switching behavior.

For the analog RRAM, the distribution of multi-level resistance states is widely spread. The wide conductance distribution causes the overlap of neighboring conductance states, resulting in retention degradation (Huang et al., 2018). In addition, after programming the device to the target conductance state, the conductance of the device may experience a notable change in a short time scale, forming tail bits (Xu et al., 2020). This is called conductance relaxation effect, which is different from retention degradation. The relaxation effect and retention

degradation are mainly due to the stochastic diffusion of O^{2-} and Vo, thus can be suppressed by the restriction of O^{2-} and Vo diffusion. For instance, Al doping in HfOx-based RRAM and HfO_2/Al_2O_3 multilayer stack are used to suppress the Vo diffusion (Chen et al., 2013; Fantini et al., 2014). However, doping may introduce dopant variations with the device scaling down to a small size. A post annealing process after Hf/HfO_2 RRAM formation was used to form an HfOx interface layer to enhance retention by slowing down the oxygen diffusion (Huang X. et al., 2016). Devices with worse state instability and retention need a short refresh interval to ensure accuracy of neural network, which brings extra power consumption.

PHYSICS-BASED COMPACT MODELS OF RESISTIVE SWITCHING RANDOM ACCESS MEMORY

The compact model is very important for the development of emerging devices. It can provide fast calculations of the device electrical properties and be implemented into standard IC design software to evaluate the performance of the target system. Moreover, a compact model involving the device physics can act as an interface between the device and the circuit. For RRAM device, based on the understanding on the microscopic properties of CF evolution and the correlated device characteristics, the physics-based compact models are investigated to capture the essential characteristics, which can be used to design and optimize the brain-inspired systems.

Binary Resistive Switching Random Access Memory

The first model of RRAM is the memristor model proposed by Chua (1971). Then a physical model for the device that behaves like a perfect memristor is proposed with a simplified explanation of current–voltage anomalies (Strukov et al., 2008). With the development of understanding of physical effects in RRAM, a compact model by considering the generation and recombination of Vo is proposed and implemented in Ngspice (Guan et al., 2012). Numerical compact models have also been developed based on the temperature and field-driven ion migrations (Larentis et al., 2012; Kim et al., 2013). By invloving the electro-themal effect, a physics-based compact model is proposed by bridging the switching behaviors with the evolution of CF configuration (Huang et al., 2013). The model is implemented into HSPICE and used for simulation of large-scale circuit by Verilog-A. In this section, this physics-based electro-thermal model will be discussed in detail.

Based on the kinetic Monte Carlo simulations in **Figure 1D**, the model with 3-D CF evolution process is developed as shown in **Figure 3A** (Huang et al., 2013). For the initial state of RESET, a cylindrical CF with the diameter w_0 bridges two electrodes. The RESET process is modeled by the increase in gap distance x between the CF tip and the top electrode when the bias increases. The increase rate of x is expressed as dx/dt. The x determines the HRS resistance, and dx/dt determines the RESET speed. The dx/dt can be calculated by the slowest process among: (1) electrode releasing O^{2-}, (2) O^{2-} hopping in the switching layer, and (3) recombination between O^{2-} and V_O.

As an example, to illustrate the modeling process, we consider the O^{2-} hopping process as the slowest process, which is also called the dominant process. During RESET, the amount of O^{2-} flowing through the unit area of cross-section per unit time can be written as:

$$J_{O^{2-}} = 1/2(P_h(E,T,dt) - P_h(-E,T,dt))/(a^2 dt) \quad (2)$$

where $J_{O^{2-}}$ is the O^{2-} flow rate, a is the distance between two V_O. The coefficient 1/2 is due to the two hopping directions of O^{2-}. In dt, the amount of O^{2-} hopping to V_O is:

$$N_{O^{2-}} = J_{O^{2-}}\pi(w_0/2)^2 dt \quad (3)$$

and the amount of V_O that take recombination reaction with O^{2-} is:

$$N_{V_O} = \pi(w_0/2)^2 dx/a^3 \quad (4)$$

Combining Equations (3) and (4), we can get:

$$\frac{dx}{dt} = af\exp(-\frac{E_h}{k_B T})\sinh(\frac{\alpha_h ZeE}{k_B T}) \quad (5)$$

where E_h is the hopping barrier of O^{2-}, E is the electric field, α_h is the enhancement factor of the electric field for the lowering of E_h, and Z is the charge number of oxygen ion. If the O^{2-} releasing or Vo recombination is the dominant process, the dx/dt can be calculated similarly.

For the SET process, the CF evolution is divided into two steps as shown in **Figure 3A**. First, a thin CF would grow from the residual CF and then connect to the electrode. Then, the thin CF would expand laterally along the radius direction. The reduction speed of gap distance dx/dt and the increase in speed of CF radius dw/dt can be calculated similarly, which are the two factors that influence the SET operation. The equivalent circuit of RRAM is shown in **Figure 3B**. It consists of a parallel capacitance (C_p), a parallel resistance (R_p), contact resistance (R_c), and the resistive switching elements (R_s). The conduction of the switching element can be modeled with metallic conduction in the CF region and hopping conduction in the gap region as shown in **Figure 3C** (Huang et al., 2013). The temperature also plays a very important role in resistive switching. In the model, we assume uniform temperature in the electrolyte layer, and the temperature at LRS can be written as Russo et al. (2009):

$$T = T_0 + IVR_{th} \quad (6)$$

where T_0 is the environment temperature, R_{th} is the effective thermal resistance of the electrolyte. Involving the model of conduction and temperature, the I–V characteristics can be calculated.

The calculated DC and AC electrical characteristics are shown in **Figures 3D,E**. The compact model can accurately reproduce the gradual RESET and the abrupt SET in the DC I–V characteristics. The transient response current waveforms

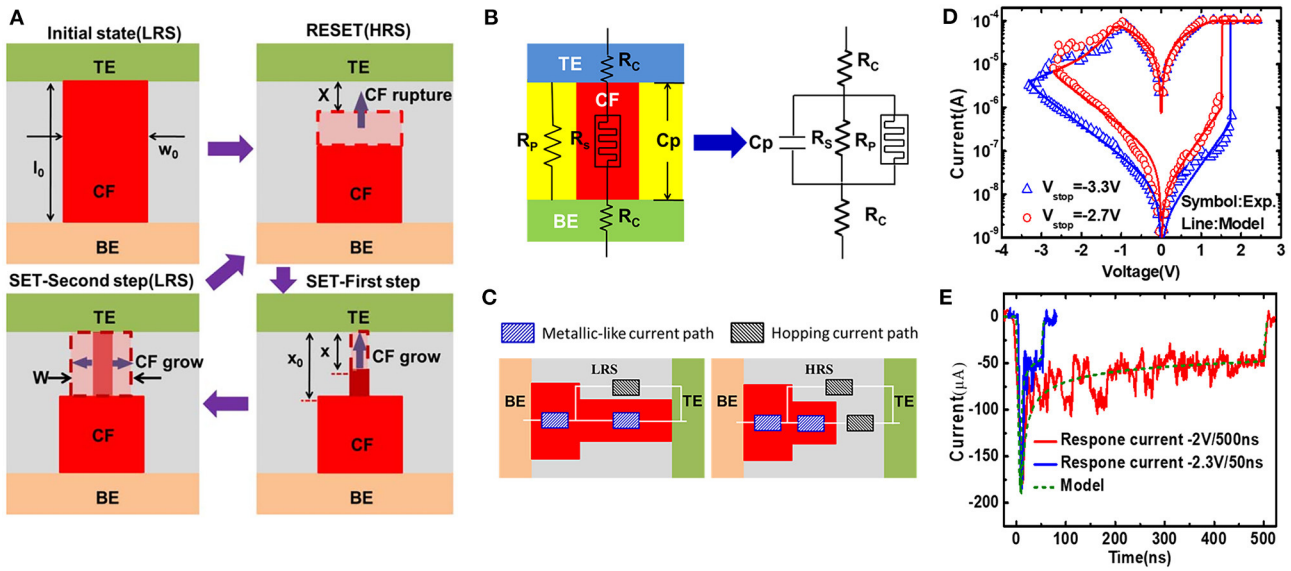


FIGURE 3 | (A) Schematic of the conductive filament (CF) evolution model in the RESET and SET process. **(B)** Equivalent circuit with parasitic effects. **(C)** The low resistance state (LRS) and high resistance state (HRS) conduction model. **(D)** Measured and calculated DC and **(E)** AC characteristics for different operation schemes. Reprinted from Huang et al. (2013).

for different RESET programming schemes of −2 V/500 ns and −2.3 V/50 ns can also be successfully reproduced. The excellent agreement between the modeling and measured results shows the validity and universality of this compact model to capture the main features of the RRAM devices. Using the model, the critical parameters during switching can be extracted from the physical view, thus providing design space for device optimization and device–circuit co-design.

Besides the basic resistive switching characteristics, the compact model for synaptic features of HfOx-based RRAM is developed to satisfy the co-design requirements of RRAM synapses and the CMOS neurons in the neuromorphic computing systems (Huang et al., 2017). The conductance change in HfOx-based RRAM can emulate the activating or deactivating ion channels of biological synapse, and the gradual RESET and stochastic SET can emulate the biological depression and potentiation processes. During RESET process, multiple intermediate states can be achieved under proper spike pulses, and they can be divided into three stages as shown in **Figure 4A** (Huang et al., 2017). **Figure 4B** schematically shows the model of gradual RESET with three stages. In the first stage, with O^{2-} released by the electrode, the Vo density near the electrode would decrease, resulting in the slimming of CF. The conductance in this stage is linear with CF width, so the conductance decrease is relatively low. In the second stage, CF is ruptured from the tip, and the O^{2-} released by the electrode would continue recombining with V_O in the CF. In this stage, the resistance is approximately exponentially dependent on the gap distance, and thus, the conductance decreases fast. In the third stage, due to the decrease in electric field in the gap, the reaction rates of O^{2-} hopping and Vo recombination decreases; hence, the resistance would tend to saturate.

The SET process in the single-layer HfOx-based RRAM is typically abrupt; thus, only binary states can be achieved. The SET also demonstrates the stochastic transition behavior as shown in **Figure 4D**. **Figure 4E** shows the model of stochastic SET. The device will be switched to LRS with the probability P after a positive pulse, which is related with the pulse amplitude V and pulse width T_w. The probability P can be written as:

$$P = \int_0^{T_w} \nu \exp\left(-\frac{E_a - \alpha_a ZeV/x}{k_B T}\right) dt \tag{7}$$

P follows a distribution even for the same device.

The proposed model is verified with measurement data as shown in **Figure 4C**. The gradual resistance modulation under consecutive identical pulses can be well-reproduced. The figure indicates that more intermediate states can be achieved with lower initial LRS, which is beneficial for synapse application. **Figure 4F** shows the measured and calculated SET voltage distributions in 1,000 cycles for the same device. They both roughly follow a normal distribution with similar mean value and standard variation. Good agreements between measurements and calculations demonstrate the validity of the model to capture the RRAM synaptic features. In addition, the SET stochasticity can be employed to generate stochastic numbers, which demonstrates great potential in the application of stochastic computing. This will be further discussed in the *Stochastic Computing* section.

Analog Resistive Switching Random Access Memory

As discussed in the *Device Optimization With Multifunctional Assistant Layer* section, analog RRAM devices have been realized

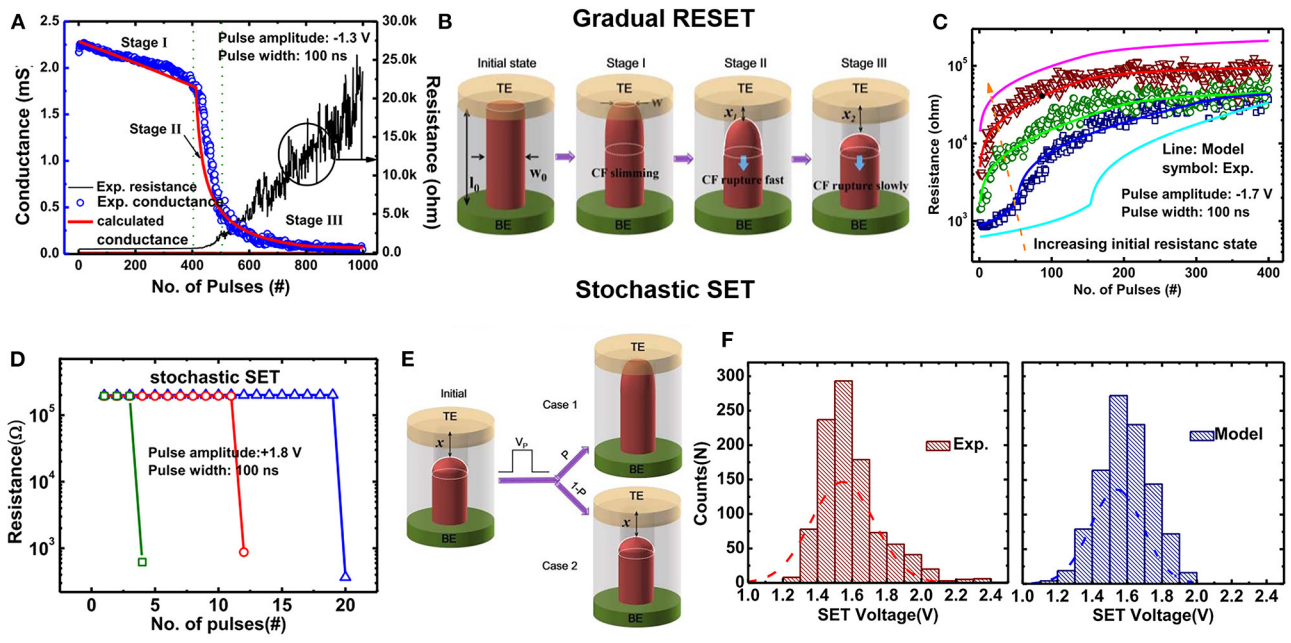


FIGURE 4 | (A) Measured gradual RESET under consecutive identical pulses. **(B)** CF evolution model of gradual RESET. **(C)** Measured and calculated gradual RESET with different initial states. **(D)** Measured binary stochastic SET under consecutive identical pulses. **(E)** Model for stochastic SET. **(F)** Measured and calculated SET voltage distributions in 1,000 cycles. Reprinted from Huang et al. (2017).

by introducing an assistant layer. Many efforts have been made to mitigate the non-ideal effects of analog RRAM including the programming non-linearity and asymmetry, variability, and tuning voltage sensitivity (Woo et al., 2016a; Wu W. et al., 2017; Wu et al., 2018). Compact models for analog RRAM have been developed to provide insights into the influence of electrical and thermal effects of assistant layer on the device characteristics and provide guidance for the optimization of non-ideal effects. In addition, the compact models can provide fast and accurate evaluation of the training accuracy. Multiple theories have been used to explain the analog switching behavior. One is the multiple-weak-filament theory, in which the local Vo concentration in the CF region is lower than the binary RRAM; thus, multiple weak CFs are assumed to be formed due to the percolation effect (Liao et al., 2020). The number of weak CFs and their conductivity are strongly dependent on the Vo concentration. Another theory describes CF with one resistive switching (RS) region and one Vo-rich (VR) region (Cai et al., 2020). The Vo concentration varies in the RS region during resistive switching processes, thus, leading to the gradual resistance modulation. Based on above theories, the key factor for analog properties is to control the Vo concentration and distribution in the CF, and the Vo modulation in multiple weak CFs can be treated as the Vo density redistribution in the RS region. The compact model with Vo modulation in the RS region will be introduced in detail in the following part.

In the model, the CF is modeled with the RS region and one VR region as shown in **Figure 5A**. In the SET process, due to the generation of Vo, the percentage of Vo in the RS region ($\Delta C_V{}^+$) increases, which can be described as:

$$\Delta C_V^+ = \Delta t \cdot f \cdot \exp(-\frac{E_a - \lambda ZeE}{k_B T})(1 - C_V) \tag{8}$$

where C_V is the Vo concentration. For RESET process, the Vo recombination leads to the decrease in C_V. Besides the kinetic barrier E_o, the releasing of O^{2-} also relies on C_V at the interface of CF and the intermediate modulation layer (IML). The O^{2-} percentage in the RS region C_O is changed by the released O^{2-}, which can be described as:

$$\Delta C_O = \Delta t \cdot f \cdot \exp(-\frac{E_o - \lambda ZeE}{k_B T}) \cdot \frac{a}{l}(1 - C_O) \tag{9}$$

The reduced percentage of Vo in the RS region is expressed as:

$$\Delta C_V^- = f \cdot \exp(-\frac{E_r}{k_B T}) \cdot C_V \cdot (\Delta C_O + C_O) \tag{10}$$

where E_r is the recombination barrier. The conduction of the analog RRAM is modeled in **Figure 5B**. In the RS region, the effective conductivity can be calculated based on the effective medium theory, while the conductivity of IML can be calculated by evolving the O^{2-} concentration in IML. Based on above model, the I–V characteristics of the analog RRAM can be calculated as shown in **Figure 5C**. Gradual SET and RESET behavior can be well-reproduced by the model, which is in good accordance with measurement data obtained from the

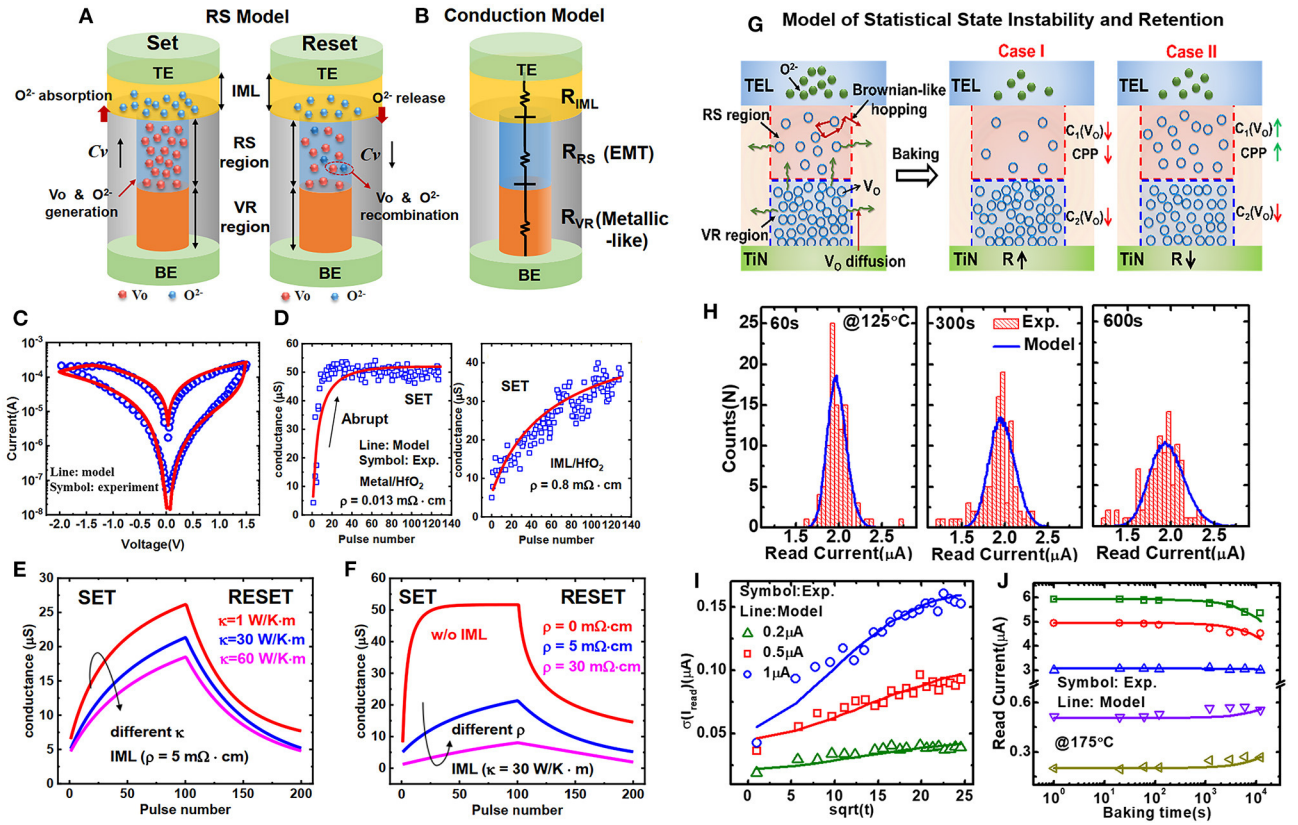


FIGURE 5 | (A) The physical switching model of the analog RRAM. **(B)** The conduction model. **(C)** The measured and calculated I–V characteristics. **(D)** The comparison between abrupt and gradual SET by introducing an intermediate modulation layer (IML). **(E)** The influence of thermal conductivity and **(F)** resistivity on the analog switching behavior. **(G)** Physical model of state instability and retention degradation. **(H)** Measured and calculated evolution of read current distributions. **(I)** Measured and calculated standard deviation of read current *vs* baking time. **(J)** Measured and calculated retention degradation for a long baking time at 175°C. Reprinted from Cai et al. (2020) and Huang et al. (2018).

TiN/TaOx/HfOx/TiN device in Wu et al. (2018). Based on the model, the continuous conductance accumulation can be reproduced under identical pulses as shown in **Figure 5D**. By adjusting the resistivity ρ of IML, the compact model shows good agreement with experiments about the linearity improvement. The non-linearity of conductance is influenced by both the electrical and thermal effects of IML. The impacts of electrical and thermal effects of IML on potentiation and depression are investigated as shown in **Figures 5E,F**. The results indicate that reduced thermal conductivity κ enlarges the tuning window due to the acceleration of Vo generation under high temperature, and the switching window is reduced with increased resistivity ρ. Increasing ρ and κ of IML both improve the linearity of conductance tuning, but the impact of resistivity is more obvious. Therefore, IML material with high resistivity would be more recommended to improve the linearity for learning accuracy in the application of neuromorphic computing.

Although analog RRAM shows great potential in weight storage and weight updating, it suffers from serious state instability and retention degradation issues, which greatly affect the performance of neural network. A physics-based analytic model is developed to describe the statistical state instability and retention behaviors of analog RRAM (Huang et al., 2018). In the model, the diffusion of Vo, the Brownian-like hopping of Vo during diffusion, and the recombination of Vo are considered. **Figure 5G** shows the physical model of the state instability and retention degradation. In a relatively short time, the Vo hopping is similar to the random Brownian movement. The Brownian-like hopping of Vo at the critical site of the current percolation path (CPP) results in the fluctuation of conductance, which is also called as the state instability. In a relatively long time, Vo diffuses along the radius direction and recombines with the O^{2-} released by the IML, thus the Vo concentration C(Vo) in the RS/VR region and the corresponding conductance decrease (case I). The diffusion of Vo from the VR region to the RS region will increase the conductance because the cell resistance mainly depends on the C(Vo) in the RS region (case II). To sum up, the diffusion and recombination of Vo will result in the retention degradation. C(Vo) in the RS and VR regions are the key parameters to characterize the state instability and retention degradation. The mean C(Vo) can be obtained as a function of time by calculating the diffusion and recombination of Vo. **Figure 5H** shows the measured and calculated read current distribution at different baking times. The distribution becomes

wide with time. The mean and standard deviation of the read current are in good accordance with the measured data. The measured and calculated standard deviations of the read current at different states are shown in **Figure 5I**, which indicates that the model can reproduce the statistical state instability. To further verify the model, the 1-kb analog RRAM array is measured under higher temperature and longer time. **Figure 5J** shows the retention behavior under 175°C of 1.2 × 10^4s, which agrees well with the model prediction. The results indicate that the mean read current of high current states decrease with time, while the mean read current of low current states increase with time. The model can be used to evaluate and optimize the performance neural network. Optimized synapse structures and refresh operation schemes can be proposed under the guidance of the model to mitigate the performance degradation, which can significantly enhance the reliability of the RRAM-based neural network.

APPLICATIONS IN BRAIN-INSPIRED COMPUTING

In the era of big data, the amount of data is explosively growing every day especially the non-structured data such as pattern, voice, and video. However, due to the von Neumann bottleneck, the traditional computing paradigm has a hard time in handling the task of a large amount of non-structured data. Fortunately, in recent years, brain-inspired computing has developed rapidly and has demonstrated great advantages in the fields of recognition and information processing, which could supplement the shortcoming of the traditional computing. In this section, the specific applications of RRAM-based brain-inspired computing including neuromorphic computing, computing in memory, and stochastic computing will be introduced.

Neuromorphic Computing

Neuromorphic computing is a kind of computing paradigm for accelerating neural networks used in data-centric computing, which paves the way for artificial intelligence with low power consumptions, mimicking the synapse- and neuron-interconnected biosystems in the human brain. RRAM is widely regarded as one of the promising candidates of artificial synaptic device, and its crossbar structure can be utilized for the hardware acceleration of the neural networks (Hochreiter and Schmidhuber, 1997; Hinton et al., 2006; Russo et al., 2009; Krizhevsky et al., 2012; Graves et al., 2013; Silver et al., 2016). The Vo/ion-based mechanism of RRAM controlling the device conductance can emulate the synaptic plasticity, acting as the base for learning and memory operations of the brain. RRAM enables high-precision synaptic weight over 6 bits, bidirectional conductance modulation, and tiny weight accumulation, so that a high-performance deep neural network algorithm could be realized; besides, RRAM could also implement the basic functions of biological synaptic, such as spike time-/rate-dependent plasticity (STDP/SRDP) and paired-pulse facilitation (PPF), which provides an approach to establish spike neural networks (SNN) (Yu et al., 2012; Gao et al., 2014, 2016; Prezioso et al., 2015; Wang et al., 2017).

In a neural network composed of neurons and synapses, neurons are connected by synapses with different weights. A two-layer neural network can be directly mapped to a RRAM crossbar array, where WLs are connected to the pre-neurons, and BLs are connected to the post-neurons as shown in **Figure 6A**. Through the RRAM-based synapse, the signals sent by the pre-neurons can be transmitted to the post-neurons. The synapse weights are mapped to the RRAM conductance. The output current I_j at the j_{th} column can be written as:

$$I_j = \sum_{i=1}^{m} V_i G_{i,j} \tag{11}$$

where V_i is the voltage applied to the i_{th} row, and $G_{i,j}$ is the conductance of RRAM at row i and column j. Therefore, the weighted sum, which is a time- and energy-consuming step for neuromorphic computing based on conventional computing system, can be performed by the RRAM crossbar array in one step. Generally, the integrated current at each column will be converted to voltage pulse by the neuron circuit and sent to the post-neuron.

The weights of the RRAM-based synapses can be updated in two ways. The first way is based on the working mechanism of the biological neural networks, in which the weight can be updated based on certain modification rules, such as the STDP (Jo et al., 2010; He et al., 2014; Du et al., 2015; Eryilmaz et al., 2015; Prezioso et al., 2016). For an STDP synapse, the weight update direction depends on the time difference Δt of the spikes from the pre-neuron and post-neuron as shown in **Figure 6B** (Jo et al., 2010). When spikes from the pre-neuron are before (or after) the post-neuron, the synaptic weight increases (or decreases). It can be found that the relation between the change in the synaptic weight and Δt can be well-fitted with exponential decay functions, which is similar to the STDP characteristics of biological synaptic systems as shown in **Figure 6B**. Arbitrary STDP behaviors, such as anti-STDP, symmetric STDP, and STDP with sin decay function can be achieved with this feature. In addition to STDP, several other synaptic functions have been realized by RRAMs, such as SRDP, short-term plasticity (STP), and long-term plasticity (LTP) (Yu et al., 2012; Gao et al., 2014, 2016; Prezioso et al., 2015; Wang et al., 2017). All these achievements are helpful to the researcher of biological neural network and will significantly enhance the intelligence of neuromorphic hardware. Although various functions of biological synapse have been realized by the RRAM, a large neural network based on such synapse update rule is still lacking due to the fact that the working mechanism of the brain is not clear. Moreover, for the SNN, the training is mainly achieved using the biology-like unsupervised learning rules, which makes it difficult to support complex practical cognitive applications.

Another principle to update the weight is the backpropagation (BP) learning rule, which has shown its advance in pattern and speech recognitions. The HfOx-based RRAM synaptic device has been demonstrated with sub-pJ energy per spike

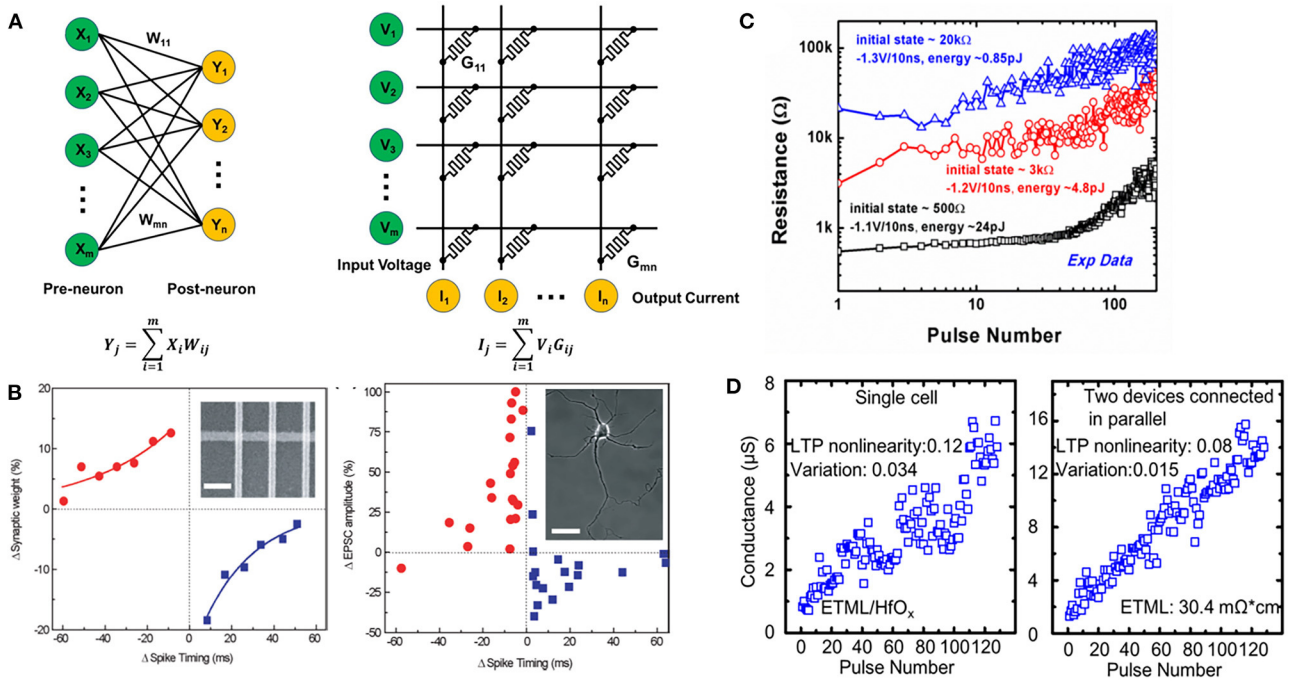


FIGURE 6 | (A) Schematic of the RRAM-based neural network. **(B)** Demonstration of the spike time-/rate-dependent plasticity (STDP) in the RRAM-based synapse and the excitatory postsynaptic current of rat hippocampal neurons. **(C)** Measured gradual training process of RRAM under consecutive identical pulses. **(D)** Improved linearity of conductance tuning in RRAM synaptic devices by introducing an electro-thermal modulation layer. Reprinted from Wu et al. (2018), Yu et al. (2012), and Jo et al. (2010).

to build a neuromorphic visual system. The measured gradual training process of RRAM under consecutive identical pulses are shown in **Figure 6C** (Yu et al., 2012). According to the BP algorithm, the desirable characteristic of the RRAM synapse is multilevel (states > 64) and low power (<0.1 pJ/spiking) switching, and the linear and symmetric responses of synapses to electric pulses are required for the training process. However, that is a quite difficult task for RRAM-based synapse. To modulate the characteristics of the RRAM-based synaptic device, the optimization of linearity and symmetry of conductance modulation is essential to realize efficient training tasks. The programming schemes can be optimized by varying the operation voltage, pulse width, gate voltage in 1T1R structure, and compliance current (Wu et al., 2012; Park et al., 2013; Woo et al., 2016b; Ku et al., 2019). However, this method brings additional circuit overhead and power consumption. Then a more favorable solution was proposed to optimize an identical programming scheme independent of device conductance states, and the abrupt resistance change can be avoided (Woo et al., 2016b). Besides the operation scheme, the non-linearity can be mitigated by the device engineering. As has been discussed in the *Device Optimization With Multifunctional Assistant Layer* section and the *Analog Resistive Switching Random Access Memory* section, an electro-thermal modulation layer has been inserted between the top electrode and resistive layer to control the distribution of electric field and temperature in the filament region; the linearity of conductance tuning is improved as shown in **Figure 6D** (Wu et al., 2018). However, the dynamic range decreases by this method.

The multilevel conductance capability of the RRAM-based synapse can impact the inference accuracy. **Figure 7A** shows the impact of weight precision on the accuracy of a two-layer fully-connected neural network for MNIST dataset (Chen et al., 2017). At least six bits are required for online training, and one or two bits are sufficient for offline classification. Higher weight precision is required for complicated convolutional neural network as shown in **Figure 7B** (Yang and Sze, 2019). To meet this requirement, a large ON/OFF ratio with multiple intermediate resistance states is essential. Regarding the issue of non-ideal device characteristics, other possible solutions may be from the interaction and optimization between devices and algorithms or architectures. For example, in the incorporation with recently proposed binarized neural networks (BNNs) based on modified BP algorithm, the impact of non-linearity in RRAM-based synapses on system performance can be effectively eliminated. A new BNN-based hardware implementation approach to utilize the non-linear synaptic cells to achieve highly efficient online training is shown in **Figure 7C** (Zhou et al., 2018). Based on the presented implementation approach, the conductance tuning non-linearity has little impact on the recognition accuracy of neural network. However, the binarization of weight would lead to the information loss, and the discontinuity of its quantization function increases the difficulty of the optimization of neural networks (Qin et al., 2020).

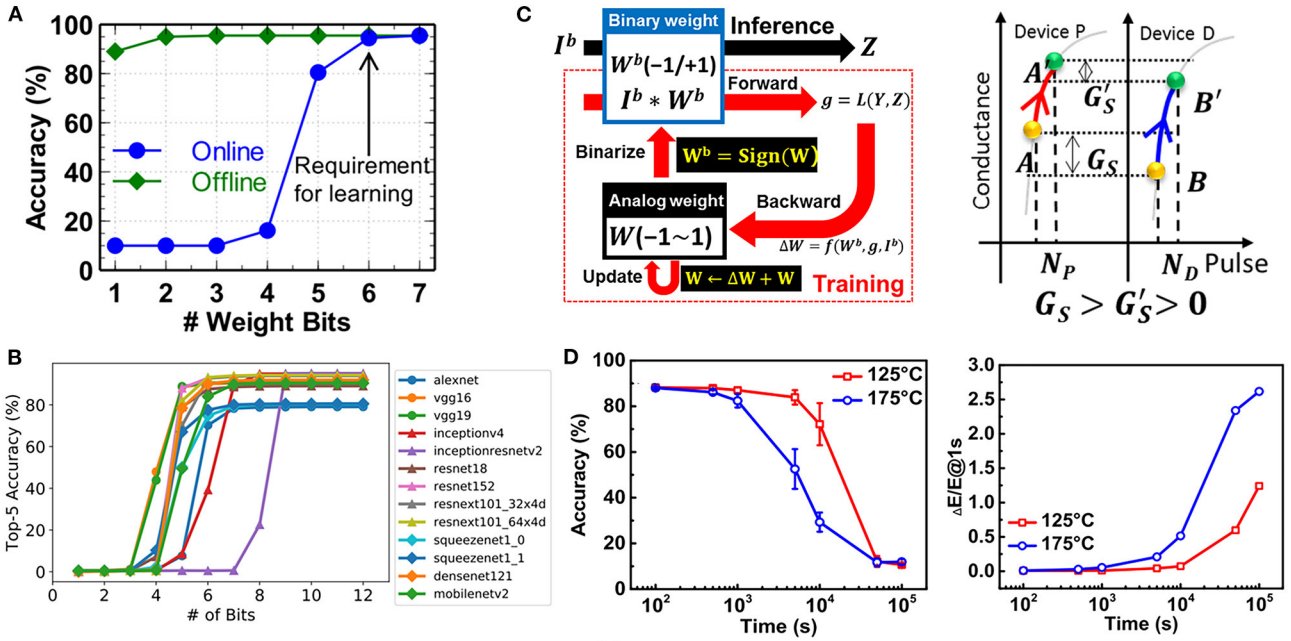


FIGURE 7 | (A) Impact of weight precision on accuracy in a two-layer fully neural network for MNIST dataset. **(B)** Impact of weight precision on accuracy of representative DNNs. **(C)** Schematic of binarized neural network (BNN) algorithm and the typical example of weight accumulation process. **(D)** Recognition accuracy and normalized energy consumption change as a function of baking time. Reprinted from Chen et al. (2017), Yang and Sze (2019), Zhou et al. (2018), and Xiang et al. (2019).

The robustness of RRAM-based neural network is related with the reliability of the RRAM-based synapse such as retention, endurance, and immunity to noise. The impacts of device state instability and retention on the performance of DNN was investigated (Xiang et al., 2019). Using the analytic model for RRAM state instability and retention degradation in the *Analog Resistive Switching Random Access Memory* section, the performance of the 11-layer RRAM-based DNN for CIFAR-10 recognition can be evaluated. **Figure 7D** shows the dependence of the recognition accuracy on the baking time at 125 and 175°C. The accuracy decreases remarkably with time due to the overlap among neighboring resistance levels. Meanwhile, the energy consumption during the inference increases with time as shown in **Figure 7D**. This is because, for the proposed neural network, more than 90% of the weight is located near 0, which means most of the RRAMs are in the low conductance states. More importantly, the differential pairs are used to store weight, and one device is in the conductance state at least. Therefore, the conductance of a large proportion of RRAMs increases with the baking time, which dominates the energy consumption. To enhance the reliability of DNN, both the device characteristics and the operation scheme should be optimized.

To design and optimize the RRAM-based neuromorphic system, modeling platforms have been developed to design the neuromorphic computing circuits and find the algorithmic constraints with device properties (Chen et al., 2017; Larcher et al., 2017; Haensch, 2018). A comprehensive model for SNN based on STDP is developed to predict the learning efficiency and time for unsupervised learning from detailed spice-like models to high-level analytical compact models (Pedretti et al., 2017). The analytic model includes all possible pattern/noise and noise/pattern sequences of input spikes as driving forces for potentiation and depression, and can predict the time evolution of pattern weight and noise weight for any set of input variables. Using the model, the impacts of noise density, pattern density, and pattern/noise probabilities on learning efficiency can be investigated, and a learning efficiency improvement up to 92% can be realized by using optimized noise in unsupervised learning of handwritten digits from the MNIST database. In terms of system-level learning accuracy and hardware performance metrics, an integrated device-to-algorithm framework NeuroSim+ for benchmarking synaptic devices and array architectures was developed (Chen et al., 2017). The framework includes the technology and memory models in the device level, the synaptic array architectures and neuron periphery in the circuit level, and the neural network topologies in the algorithm level. The impact of device non-ideal properties on learning accuracy, the area, latency and energy estimation in the circuit level can then be investigated by this framework. A two-layer multilayer perceptron (MLP) neural network with MNIST handwritten digits is adopted as the training and testing dataset to implement online learning and offline classification. In the MLP neural network, the MNIST input images are converted to black and white data to reduce the encoding complexity. The weights are mapped to the synaptic cores, which are the computation units for performing weighted

sum and weight update. The synaptic core can be categorized into the binary RRAM and analog RRAM, where binary type is more mature. When a weighted sum or weight update instruction is given during feed forward and BP, the instruction will be sent to the RRAM array and device behavior model for calculating the computation error and sent to NeuroSim to evaluate the circuit performance. The framework facilitates the design space exploration from device to algorithm, which is helpful to benchmark different synaptic device candidates and array architectures for neuromorphic applications.

For RRAM-based neuromorphic computing, although some small-scale neural networks have been demonstrated, it is still far from being applied. The challenges come from the design and fabrication of RRAM arrays with high performances, device characteristic engineering, neuron circuit design, and algorithm modification. Possible solutions should consider the interaction and optimization between devices and algorithm or architectures.

In-memory Logic

The conventional computation systems process information and store information separately, which brings huge energy cost and time wasting in data transfer between the computing units and memories. In order to break the von Neumann bottleneck in both the device and architecture level and meet the requirement for energy-efficient information system, the RRAM-based logic is proposed as a promising solution, which can perform logic operation and store the output in the same physical location (Borghetti et al., 2010; Li et al., 2015a; Huang P. et al., 2016).

In 2010, the RRAM-based stateful logic operation was first proposed and experimentally demonstrated (Borghetti et al., 2010). The basic logic operation is the implication (IMP), and the operation is based on two RRAM devices (P and Q) and one resistor as shown in **Figure 8A**. The resistance state stored in P and Q represents the logical value. IMP is performed by two simultaneous pulses applied on P and Q to execute conditional toggling on Q depending on the state of P and Q. The output of the operation is then stored in Q. If we define HRS as "1" and LRS as "0," the IMP result is summarized in **Figure 8A**. Based on this principle, other logic computations can also be performed. However, the initial state of Q is covered during the operation, which hinders the logic cascading, and the Q needs a copy operation if the value is used more than once (Li et al., 2015b).

To prevent the input value from being covered, a method to execute NAND and logic operations in one step was proposed (Huang P. et al., 2016). The subcircuit to realize a NAND operation is shown in **Figure 8B**. In the circuit, the device top electrodes are connected to a common WL. A strong pulse is applied to the WL via a reference resistor, and a small pulse is applied to devices A and B through BL. For device Y, the BL is grounded. The input for the operation is the resistance states of A and B, and the output will be stored in Y, whose initial state has been switched to HRS. If A and B are both "1," the potential of common WL is close to V_{DD}, then Y will be programmed to "0" after the operation. If any input device is "0," the potential of common WL is close to V_R; thus, the output Y will still be "1." By this way, the NAND logic operation is performed. The value of V_R, V_{DD}, and R_G should be carefully designed to guarantee the NAND operation. V_{DD} should be larger than the SET voltage in order to compensate the voltage drop across R_G. As for V_R, on one side, it should be large enough to avoid the switching of A and B; on the other side, it should be small enough to avoid the switching of Y. The experimental demonstration of the NAND logic is shown in **Figure 8B**. The logic function of the subcircuit can be reconfigured by changing the applied voltage. For example, the AND logic can also be realized using the same subcircuit by exchanging the V_{DD} and V_R.

Besides the basic logic operation, compound logic operation can be executed with latching the NAND logic operation. **Figure 8C** shows an example of a full adder. The subcircuit is composed of nine RRAM devices including three input devices (addend A, summand B, and carry-in C_i), two output devices (summary S and carry-out C_o), and four assisted devices (AS_1-AS_4). The computation procedure is shown in **Figure 8C**, which needs 10 sequential steps. The corresponding logical states after each procedure are read out and demonstrated as gray-scale maps. The measured data indicate that the function of a full adder can be realized correctly. In order to realize the logic operation in arbitrary positions in the RRAM array, the structure of devices with the same BL was also proposed and verified (Huang P. et al., 2016). The same computing task can be performed parallelly by cells in different rows or columns in the RRAM array by simultaneously applying the pulses to the corresponding ports of BL and WL.

One challenge for the RRAM-based stateful logic is the device variations, which may cause errors to the logic operation. Therefore, the logic operation should be robust to these device variations, which include the SET voltage variation and resistance variation. To quantitatively describe the robustness of the logic operation, the dependence of maximum tolerance to SET voltage variation on the resistance window (R_H/R_L) was investigated by HSPICE simulation (Shen et al., 2019). The results indicate that compared with the conventional scheme based on 1R structure, the dual gate voltage scheme in the 1T1R array shows higher robustness to the SET voltage variations as R_H/R_L changes from 25 to 10,000. The variation of resistance in HRS and LRS will reduce the effective resistance window. For each given SET voltage variation, there exists a tolerable resistance window to ensure the successful logic operation.

The Boolean logic computing is closer to the off-the-shelf system compared with the neuromorphic computing paradigm, which does not require new algorithm or software. However, the development of the RRAM-based in-memory logic is very slow due to the lack of application scenarios, and the demonstration of complete computing and memory unit is still missing.

Stochastic Computing

Stochastic computing (SC) is a highly fault-tolerant and energy-efficient computing paradigm, which can realize complex functions with simple logic units (Gaines, 1969; Lv and Wang, 2017; Hu et al., 2019). Different from the traditional binary computing, SC operates on stochastic bit streams (SBSs), which emulate the neural spikes processed by the brain in the form of long sequences of noisy voltage spikes as shown in **Figure 9A**.

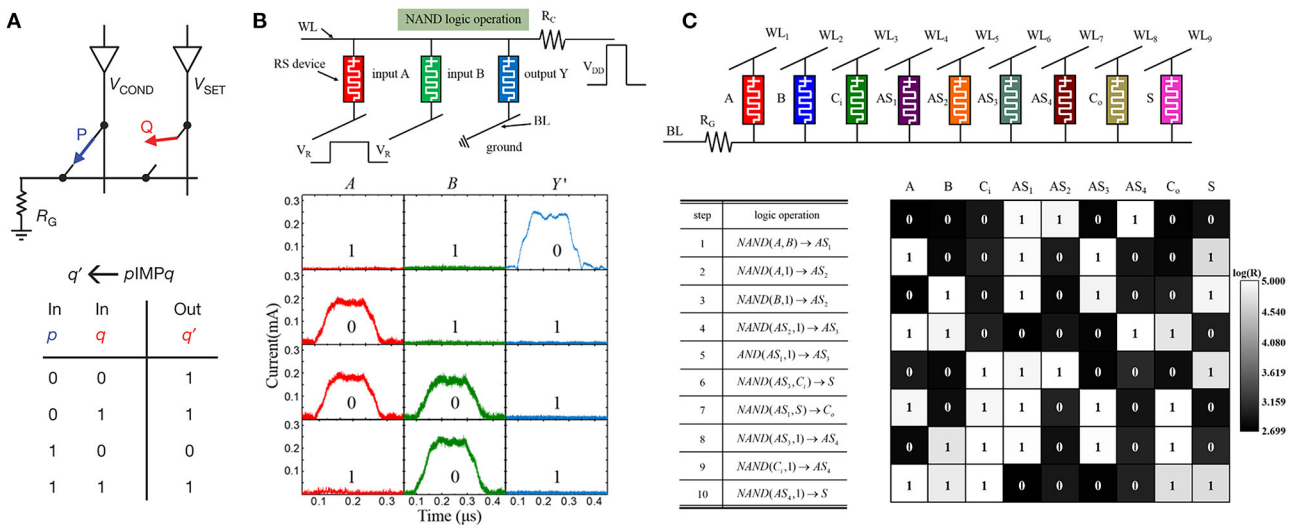


FIGURE 8 | (A) RRAM-based IMP logic operation and the truth table. **(B)** The subcircuit of the NAND operation by using three RRAM devices with common WL and the corresponding measured results. The logic inputs are stored in devices A and B, while Y stores the output. **(C)** The subcircuit of one-bit full adder and the corresponding computation procedure. Here "NAND(A, B) → Y" represents that states of devices A and B are executed; the NAND operation and the output are stored into device Y. Reprinted from Borghetti et al. (2010) and Huang P. et al. (2016).

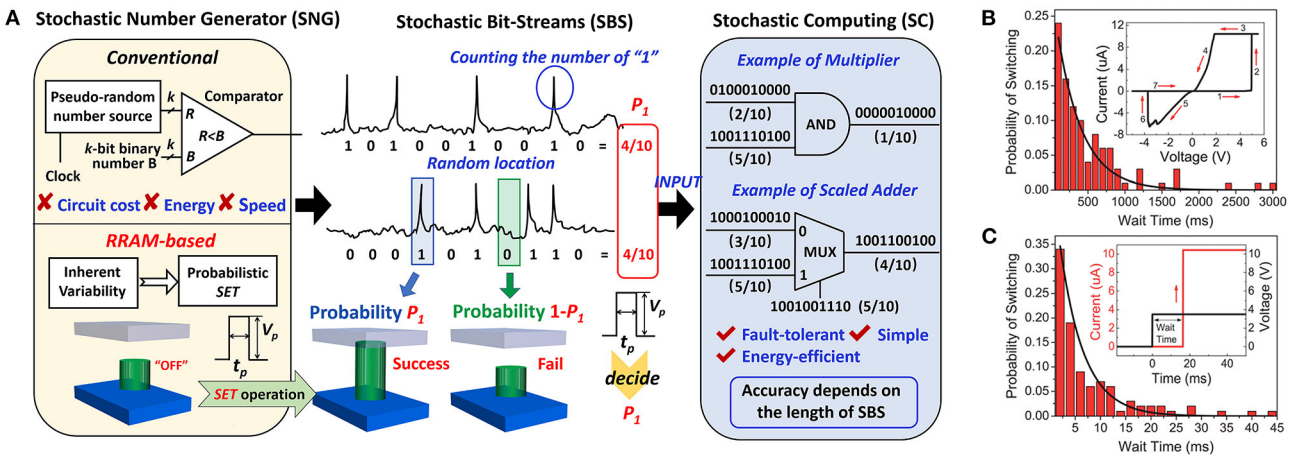


FIGURE 9 | (A) The schematic of the RRAM-based stochastic number generator (SNG) in the stochastic computing (SC) system. The RRAM-based SNG utilizes the probabilistic SET to randomly generate "1" or "0." **(B)** The SET waiting time distribution when the pulse amplitude is 2.5 V and **(C)** 3.5 V. Reprinted from Knag et al. (2014) and Zhao et al. (2019).

The information contained in the SBS is the frequency at which the spikes appear randomly within a period of time. For example, the value 0.4 can be represented by a 10-bit SBS {1,0,0,1,0,1,0,0,1,0}, where the probability of "1" is 0.4. The position of "1"s in the SBS is random, so different SBSs can represent the same value. Moreover, SC can be implemented with simple arithmetic units. For example, A multiplied by B can be operated with an AND gate, while A plus B can be operated with a MUX (Lv and Wang, 2017; Yang et al., 2017; Hu et al., 2019). Compared with the binary system, the SBS is more fault tolerant because one-bit flip is almost negligible. Therefore, SC can be used in highly fault-tolerant applications such as parity-check decoding, image processing, filter design, and neural networks (Gaudet and Rapley, 2003; Ma et al., 2012; Alaghi et al., 2013; Li P. et al., 2014; Canals et al., 2016; Li B. et al., 2016; Li Z. et al., 2016).

The biggest challenge to realize SC is to generate SBS efficiently. The traditional stochastic number generator (SNG) is composed of a pseudo stochastic number-generating unit such as the linear feedback shift register and a comparator. Compared with the simple computation unit of SC, the CMOS-based SNG occupies up to 80% of the system circuit area, which brings huge hardware overhead. RRAM devices, with the feature of

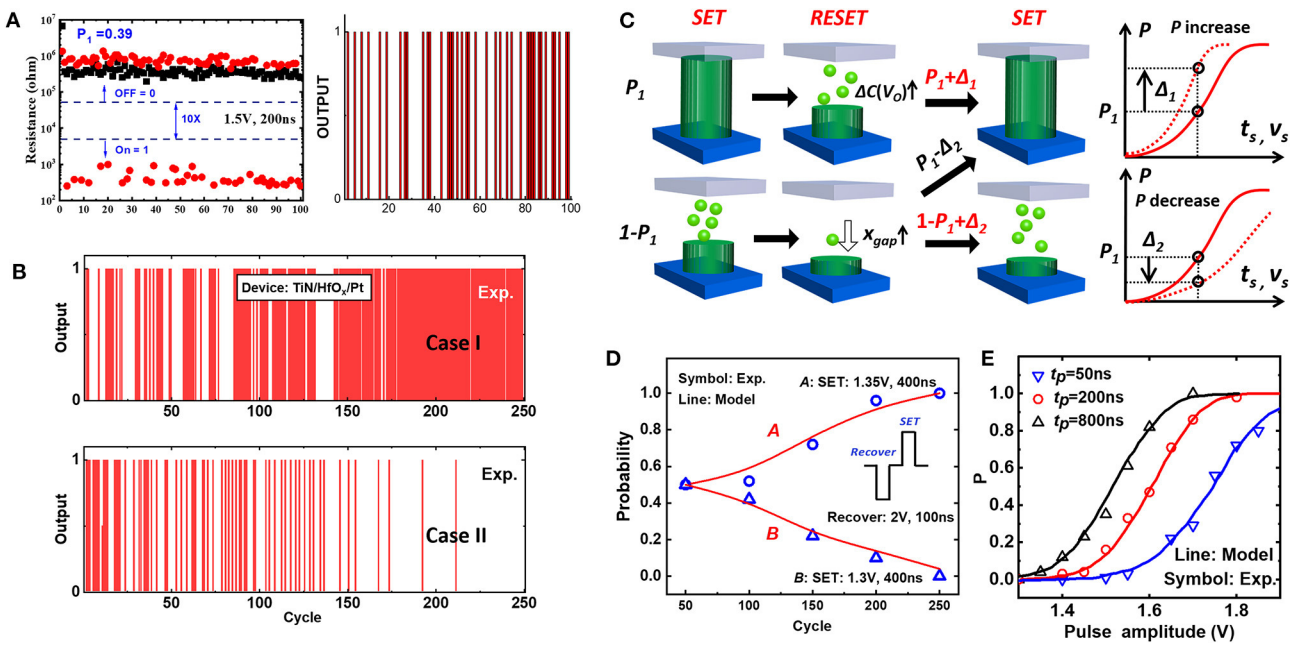


FIGURE 10 | (A) The measured resistance state after 100 consecutive SET and RESET pulses and the corresponding SBS. **(B)** The measured SBS with probability shift behavior between continuous cycles with the same initial probability. **(C)** The model of probability shift behavior. **(D)** The measured and calculated SET probability shift behavior. **(E)** The measured and calculated SET probability curve with different pulse intensity under optimized operation scheme. Reprinted from Zhao et al. (2019).

inherent variability, shows great potential to be used as low-cost and energy-efficient SNG (Gaba et al., 2013; Suri et al., 2013; Knag et al., 2014; Moons and Verhelst, 2014; Ielmini and Wong, 2018; Wang et al., 2018; Carboni and Ielmini, 2019; Zhao et al., 2019). The inherent variability of RRAM originates from the probabilistic SET process as have been discussed in the *Binary Resistive Switching Random Access Memory* section (**Figure 4**). **Figures 9B,C** are the measurement results of the waiting time distribution during the SET process (Knag et al., 2014). The SET waiting time can be obtained by performing continuous RESET and SET operations on the device and then recording the time before the transition from HRS to LRS during each SET operation. Based on the measurements, the SET waiting time roughly follows the Poisson distribution, and the distribution curve will shift left or right when changing the pulse amplitude. Therefore, when consecutively applying SET and RESET pulses on the device, whether the CF would be generated inside the device is random, so a sequence of different current levels can be obtained, as shown in **Figure 10A**. Using "1" representing the LRS and "0" representing HRS, an SBS of n bits can be achieved. The SET probability is determined by the intensity of SET pulse; thus, by adjusting the pulse amplitude and pulse width, the numerical value represented by the SBS can be adjusted.

To accurately control and predict the SET probability, the probability should be quantitatively modeled considering the device physics, as a small deviation of the input signal could affect the probability significantly. By considering multiple variation sources including the atom thermal vibration, manufacturing parameter variation, and cycle–cycle gap distance fluctuation, the behavior of the RRAM-based SNG can be modeled (Zhao et al., 2019). However, the RRAM SET probability may shift upward or downward between continuous cycles. **Figure 10B** shows the measured SBS with probability shift behavior of $TiN/HfO_2/Pt$ device. The unstable SET probability will influence the accuracy of the SBS, which must be mitigated for the application of RRAM-based SC. The probability shift behavior is modeled as shown in **Figure 10C**. Due to the different SET results in the $N{-}1_{\text{th}}$ cycle, the SET probability between the $N{-}1_{th}$ and the N_{th} cycles will increase or decrease. For example, the upper figure in **Figure 10C** corresponds to the situation where the CF successfully connected the electrodes during the $N{-}1_{th}$ SET process, and the device represents "1" after this operation. At this time, the concentration of the remaining Vo increases after RESET. The probability of generating "1" in the next SET operation increases, and the corresponding SET probability distribution curve would shift left. The model can well-reproduce the probability shift behavior observed in experiments as shown in **Figure 10D**. The increase or decrease of SET probability with cycles is due to the mismatch between SET and RESET pulses; thus, an optimized operation scheme is proposed by the model to suppress the probability shift behavior by applying an additional deterministic SET before each RESET operation. After suppressing the probability shift behavior, the SET probability dependence on pulse amplitude and pulse width can be investigated. **Figure 10E** shows the calculated and

measured SET probability curve with different pulse strengths. The SET probability changes with pulse strength; thus, one can use this curve to obtain the device operation scheme depending on the desired probability, which is the value represented by SBS in the SC application.

In addition to the SET operation, the RESET operation also has a great influence on the SET probability. When increasing the amplitude of RESET pulse, the probability distribution curve shifts to the right. This is because a stronger RESET pulse will increase the gap length before each SET, which will reduce the probability of a successful SET. Therefore, to obtain the expected SET probability in a RRAM-based SNG, the SET and RESET operations should be both carefully designed. Moreover, due to the randomness of resistive switching and the noise in the pulse signal, the length of SBS should be properly selected to avoid a large error. The accuracy of SBS can be improved by using longer SBS, but the energy consumption and calculation time will also increase exponentially (Gaines, 1969). Therefore, according to the requirements of the SC application scenarios, the accuracy, energy consumption, and calculation time should be collaboratively designed.

The challenge facing the RRAM-based SC is the uncontrollable device stochasticity, so the distribution and probability of switching cannot be accurately predicted, which would seriously affect the accuracy of SC. Although the improvement of accuracy can be realized by using a longer bit stream length, the energy consumption will be greatly increased, resulting in the design trade-off between accuracy and energy consumption. The cost-effective design techniques that minimize the disadvantages such as low precision and long bit-streams are highly required.

SUMMARY

The RRAM-based brain-inspired computing systems has achieved remarkable progresses in the past decades. Various computing paradigms have been proposed to exploit the device physics to perform neuromorphic computing, in-memory logic, and stochastic computing. However, some key issues still need to be addressed such as the device variability, forming voltage, selector device, and non-linearity/symmetry of RRAM-based synapses; thus, the design and optimization of structures, materials, and operation schemes in the device level, by means of the deeply physical understanding and innovative device-engineering methods, are still required. Moreover, the corresponding architectures and algorithms that can be utilized to construct power-efficient brain-inspired computing systems are still being developed, and it highly desires the persistent and creative research to the interaction and optimization between devices and algorithms or architectures.

AUTHOR CONTRIBUTIONS

YZ and RC contributed to the writing of the manuscript. RC and JK revised the manuscript. PH and JK helped with the supervision of the study. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported, in part, by the National Key Research and Development (2018YFE0100800), National Natural Science Foundation of China (61841404 and 62004005), and the 111 project (B18001).

REFERENCES

Alaghi, A., Cheng, L., and Hayes, J. P. (2013). "Stochastic circuits for real-time image-processing applications," in *Proceeding Design Automation Conference* (Austin: ACM/EDAC/IEEE), 1–6. doi: 10.1145/2463209.2488901

Asamitsu, A., Tomioka, Y., Kuwahara, H., and Tokura, Y. (1997). Current switching of resistive states in magnetoresistive manganites. *Nature* 388, 50–52. doi: 10.1038/40363

Azzaz, M., Benoist, A., Vianello, E., Garbin, D., Jalaguier, E., Cagli, C., et al. (2015). "Benefit of Al_2O_3/HfO_2 bilayer for BEOL RRAM integration through 16kb memory cut characterization," in *Proceeding European Solid State Device Research Conference* (Graz: IEEE), 266–269. doi: 10.1109/ESSDERC.2015.7324765

Baek, G., Lee, M. S., Seo, S., Lee, M. J., Seo, D. H., Suh, D.-S., et al. (2004). "Highly scalable nonvolatile resistive memory using simple binary oxide driven by asymmetric unipolar voltage pulses," in *Proceeding International Electron Devices Meeting* (San Francisco, CA: IEEE), 587–590 .

Borghetti, J., Snider, G. S., Kuekes, P. J., Yang, J. J., Stewart, D. R., and Williams, R. S. (2010). 'Memristive' switches enable 'stateful' logic operations via material implication. *Nature* 464, 873–876. doi: 10.1038/nature08940

Cai, L., Chen, W., Zhao, Y., Liu, X., Kang, J., Zhang, X., et al. (2020). A physics-based analytic model of analog switching resistive random access memory. *IEEE Electron Device Lett.* 41, 236–239. doi: 10.1109/LED.2019.2961697

Canals, V., Morro, A., Oliver, A., Alomar, M. L., and Rosselló, J. L. (2016). A new stochastic computing methodology for efficient neural network implementation. *IEEE Trans. Neural Netw. Learn. Syst.* 27, 551–564. doi: 10.1109/TNNLS.2015.2413754

Carboni, R., and Ielmini, D. (2019). Stochastic memory devices for security and computing. *Adv. Electron. Mater.* 5:1900198. doi: 10.1002/aelm.201900198

Chen, P.-Y., Peng, X., and Yu, S. (2017). "NeuroSim+: an integrated device-to-algorithm framework for benchmarking synaptic devices and array architectures," in *Proceeding International Electron Devices Meeting* (San Francisco: IEEE), 135–138. doi: 10.1109/IEDM.2017.8268337

Chen, Y. S., Lee, H. Y., Chen, P. S., Gu, P. Y., Chen, C. W., Lin, W. P., et al. (2009). "Highly scalable hafnium oxide memory with improvements of resistive distribution and read disturb immunity," in *Proceeding International Electron Devices Meeting* (Washington, DC: IEEE), 105–108. doi: 10.1109/IEDM.2009.5424411

Chen, Y. Y., Komura, M., Degraeve, R., Govoreanu, B., Goux, L., A., et al. (2013). "Improvement of data retention in HfO_2/Hf 1T1R RRAM cell under low operating current," *Proceeding International Electron Devices Meeting* (Washington, DC: IEEE), 252–255. doi: 10.1109/IEDM.2013.6724598

Chi, P., Li, S., Xu, C., Zhang, T., Zhao, J., Liu, Y., et al. (2016). "PRIME: a novel processing-in-memory architecture for neural network computation in ReRAM-based main memory," in *Proceeding Annual International Symposium on Computer Architecture* (Seoul: IEEE), 27–39. doi: 10.1145/3007787.3001140

Chien, W. C., Chen, Y. R., Chen, Y. C., Chuang, A. T. H., Lee, F. M., Lin, Y. Y., et al. (2010). "A Forming-free WO_X resistive memory using a novel self-aligned field enhancement feature with excellent reliability and scalability," in *Proceeding International Electron Devices Meeting* (San Francisco, CA: IEEE), 440–443. doi: 10.1109/IEDM.2010.5703390

Chou, C.-T., Hudec, B., Hsu, C.-W., Lai, W.-L., Chang, C.-C., and Hou, T.-H. (2015). Crossbar array of selector-less TaOx/TiO_2 bilayer RRAM. *Microelectron. Reliab*. 55, 2220–2223. doi: 10.1016/j.microrel.2015.04.002

Chua, L. (1971). Memristor—missing circuit element. *IEEE Trans. Circuit Theory* 18, 507–519. doi: 10.1109/TCT.1971.1083337

Chuang, K. C., Chu, C. Y., Zhang, H. X., Luo, J. D., Li, W. S., Li, Y. S., et al. (2019). Impact of the stacking order of HfOx and AlOx dielectric films on RRAM switching mechanisms to behave digital resistive switching and synaptic characteristics. *IEEE J. Electron Device Soc.* 7:589. doi: 10.1109/JEDS.2019.2915975

Degraeve, R., Roussel, P., Goux, L., Wouters, D., Kittl, J., Altimime, L., et al. (2010). "Generic learning of TDDB applied to RRAM for improved understanding of conduction and switching mechanism through multiple filaments," in *Proceeding International Electron Devices Meeting* (San Francisco, CA: IEEE), 632–635. doi: 10.1109/IEDM.2010.5703438

Du, C., Ma, W., Chang, T., Sheridan, P., and Lu, W. D. (2015). Biorealistic implementation of synaptic functions with oxide memristors through internal ionic dynamics. *Adv. Funct. Mater*. 25, 4290–4299. doi: 10.1002/adfm.201501427

Eryilmaz, S. B., Kuzum, D., Yu, S., and Wong, H. S. P. (2015). "Device and system level design considerations for analog-non-volatile-memory based neuromorphic architectures," in *Proceeding International Electron Devices Meeting* (Washington, DC: IEEE), 64–67. doi: 10.1109/IEDM.2015.7409622

Fantini, A., Goux, L., Clima, S., Degraeve, R., Redolfi, A., Adelmann, C., et al. (2014). "Engineering of $Hf_{1-x}Al_xO_y$ amorphous dielectrics for high-performance RRAM applications," in *Proceeding International Memory Workshop* (Taipei: IEEE), 1–4.

Gaba, S., Sheridan, P., Zhou, J., Choi, S., and Lu, W. (2013). Stochastic memristive devices for computing and neuromorphic applications. *Nanoscale* 5, 5872–5878. doi: 10.1039/c3nr01176c

Gaines, B. R. (1969). "Stochastic computing systems," in *Advances in Information Systems Science*, ed J. T. Tou (Boston, MA: Springer), 37–172. doi: 10.1007/978-1-4899-5841-9_2

Gao, B., Bi, Y., Chen, H.-Y., Liu, R., Huang, P., Chen, B., et al. (2014). Ultra-low-energy three-dimensional oxide-based electronic synapses for implementation of robust high-accuracy neuromorphic computation systems. *ACS Nano* 8, 6998–7004. doi: 10.1021/nn501824r

Gao, B., Kang, J. F., Chen, Y. S., Zhang, F. F., Chen, B., Huang, P., et al. (2011). "Oxide-based RRAM: unified microscopic principle for both unipolar and bipolar switching," in *Proceeding International Electron Devices Meeting* (Washington, DC: IEEE), 417–420. doi: 10.1109/IEDM.2011.6131573

Gao, L., Chen, P.-Y., and Yu, S. (2016). Demonstration of convolution kernel operation on resistive crossbar array. *IEEE Electron Device Lett*. 37, 870–873. doi: 10.1109/LED.2016.2573140

Gao, L., Wang, I.-T., Chen, P.-Y., Vrudhula, S., Seo, J., Cao, Y., et al. (2015). Fully parallel write/read in resistive synaptic array for accelerating on-chip learning. *Nanotechnology* 26:455204. doi: 10.1088/0957-4484/26/45/455204

Gaudet, V. C., and Rapley, A. C. (2003). Iterative decoding using stochastic computation. *Electron. Lett*. 39, 299–301. doi: 10.1049/el:20030217

Goux, L., Degraeve, R., Govoreanu, B., Chou, H.-Y., Afanas'ev, V. V., Meersschaut, J., et al. (2011). "Evidences of anodic-oxidation reset mechanism in TiN/NiO/Ni RRAM cells," in *Proceeding Symposium on VLSI Technology* (Kyoto: IEEE), 24–25.

Goux, L., Fantini, A., Kar, G., Chen, Y.-Y., Jossart, N., Degraeve, R., et al. (2012). "Ultralow sub-500nA operating current high-performance TiN/Al_2O_3/HfO_2/Hf/TiN bipolar RRAM achieved through understanding based stack-engineering," in *Proceeding Symposium on VLSI Technology* (Honolulu: IEEE), 159–160. doi: 10.1109/VLSIT.2012.6242510

Govoreanu, B., Kar, G. S., Chen, Y.-Y., Paraschiv, V., Kubicek, S., Fantini, A., et al. (2011). "10x10nm^2 Hf/HfOx crossbar resistive RAM with excellent performance, reliability and low-energy operation," in *Proceeding International Electron Devices Meeting* (Washington: IEEE), 729–732. doi: 10.1109/IEDM.2011.6131652

Graves, A., Mohamed, A., and Hinton, G. (2013). "Speech recognition with deep recurrent neural networks," in *Proceeding International Conference on Acoustics, Speech and Signal Processing,* Vancouver, BC: IEEE), 6645–6649. doi: 10.1109/ICASSP.2013.6638947

Guan, X., Yu, S., and Wong, H. S. P. (2012). On the switching parameter variation of metal-oxide RRAM—Part I: physical modeling and simulation methodology. *IEEE Trans. Electron Devices* 59, 1172–1182. doi: 10.1109/TED.2012.2184545

Haensch, W. (2018). "Analog computing for deep learning: algorithms, materials & architectures," in *Proceeding International Electron Devices Meeting* (San Francisco, CA: IEEE), 59–62. doi: 10.1109/IEDM.2018.8614681

He, W., Huang, K., Ning, N., Ramanathan, K., Li, G., Jiang, Y., et al. (2014). Enabling an integrated rate-temporal learning scheme on memristor. *Sci. Rep*. 4:4755. doi: 10.1038/srep04755

Hickmott, T. W. (1962). Low-frequency negative resistance in thin anodic oxide films. *J. Appl. Phys*. 33:2669. doi: 10.1063/1.1702530

Hinton, E., Osindero, S., and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Comput.* 18, 1527–1554. doi: 10.1162/neco.2006.18.7.1527

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput*. 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Hsu, C.-W., Wang, Y.-F., Wan, C.-C., Wang, I.-T., Chou, C.-T., Lai, W.-L., et al. (2014). Homogeneous barrier modulation of TaOx/TiO_2 bilayers for ultra-high endurance three-dimensional storage-class memory. *Nanotechnology* 25:165202. doi: 10.1088/0957-4484/25/16/165202

Hu, J., Li, B., Ma, C., Lilja, D., and Koester, S. J. (2019). Spin-hall-effect-based stochastic number generator for parallel stochastic computing. *IEEE Trans. Electron Devices* 66, 3620–3627. doi: 10.1109/TED.2019.2920401

Huang, P., Kang, J., Zhao, Y., Chen, S., Han, R., Zhou, Z., et al. (2016). Reconfigurable nonvolatile logic operations in resistance switching crossbar array for large-scale circuits. *Adv. Mater*. 28, 9758–9764. doi: 10.1002/adma.201602418

Huang, P., Liu, X. Y., Chen, B., Li, H. T., Wang, Y. J., Deng, Y. X., et al. (2013). A physics-based compact model of metal-oxide-based RRAM DC and AC operations. *IEEE Trans. Electron Devices* 60, 4090–4097. doi: 10.1109/TED.2013.2287755

Huang, P., Xiang, Y. C., Zhao, Y. D., Liu, C., Gao, B., Wu, H. Q., et al. (2018). "Analytic model for statistical state instability and retention behaviors of filamentary analog RRAM array and its applications in design of neural network," in *Proceeding International Electron Devices Meeting,* San Francisco, CA: IEEE), 937–940. doi: 10.1109/IEDM.2018.8614567

Huang, P., Zhu, D., Chen, S., Zhou, Z., Chen, Z., Gao, B., et al. (2017). Compact model of HfOx-based electronic synaptic devices for neuromorphic computing. *IEEE Trans. Electron Devices* 62, 614–621. doi: 10.1109/TED.2016.2643162

Huang, X., Wu, H., Gao, B., Sekar, D. C., Dai, L., Kellam, M., et al. (2016). HfO_2/Al_2O_3 multilayer for RRAM arrays: a technique to improve tail-bit retention. *Nanotechnology* 27, 395201. doi: 10.1088/0957-4484/27/39/395201

Ielmini, D., Nardi, F., Cagli, C., and Lacaita, A. L. (2010). Size-dependent retention time in NiO-based resistive-switching memories. *IEEE Electron Device Lett*. 31, 353–355. doi: 10.1109/LED.2010.2040799

Ielmini, D., and Wong, H. S. P. (2018). In-memory computing with resistive switching devices. *Nat. Electron*. 1, 333–343. doi: 10.1038/s41928-018-0092-2

Janousch, M., Meijer, G. I., Staub, U., Delley, B., Karg, S. F., and Andreasson, B. P. (2007). Role of oxygen vacancies in Cr-doped $SrTiO_3$ for resistance-change memory. *Adv. Mater*. 19, 2232–2235. doi: 10.1002/adma.200602915

Jeong, D. S., Kim, K. M., Kim, S., Choi, B. J., and Hwang, C. S. (2016). Memristors for energy-efficient new computing paradigms. *Adv. Electron. Mater*. 2:1600090. doi: 10.1002/aelm.201600090

Jo, S. H., Chang, T., Ebong, I., Bhadviya, B. B., Mazumder, P., and Lu, W. (2010). Nanoscale memristor device as synapse in neuromorphic systems. *Nano Lett*. 10, 1297–1301. doi: 10.1021/nl904092h

Kang, J. F., Gao, B., Huang, P., Li, H. T., Zhao, Y. D., Chen, Z., et al. (2015). "Oxide-based RRAM: Requirements and challenges of modeling and simulation," in *Proceeding International Electron Devices Meeting* (Washington, DC: IEEE), 113–116. doi: 10.1109/IEDM.2015.7409634

Kim, S., Kim, S.-J., Kim, K. M., Lee, S. R., Chang, M., Cho, E., et al. (2013). Physical electro-thermal model of resistive switching in bi-layered resistance-change memory. *Sci. Rep*. 3:1680. doi: 10.1038/srep01680

Knag, P., Lu, W., and Zhang, Z. (2014). A native stochastic computing architecture enabled by memristors. *IEEE Trans. Nanotechnol*. 13, 283–293. doi: 10.1109/TNANO.2014.2300342

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "ImageNet classification with deep convolutional neural networks," in *Proceeding International*

Conference on Neural Information Processing Systems (Lake Tahoe: ACM), 1097–1105.

Ku, B., Abbas, Y., Kim, S., Sokolov, A. S., Jeon, Y. R., and Choi, C. (2019). Improved resistive switching and synaptic characteristics using Ar plasma irradiation on the Ti/HfO_2 interface. *J. Alloy Compd*. 797, 277–283. doi: 10.1016/j.jallcom.2019.05.114

Kwon, D.-H., Kim, K. M., Jang, J. H., Jeon, J. M., Lee, M. H., Kim, G. H., et al. (2010). Atomic structure of conducting nanofilaments in TiO_2 resistive switching memory. *Nat. Nanotechnol*. 5, 148–153. doi: 10.1038/nnano.2009.456

Larcher, L., Padovani, A., and Lecce, V. D. (2017). "Multiscale modeling of neuromorphic computing: from materials to device operations," in *Proceeding International Electron Devices Meeting* (San Francisco, CA: IEEE), 282–285. doi: 10.1109/IEDM.2017.8268374

Larentis, S., Nardi, F., Balatti, S., Ielmini, D., and Gilmer, D. C. (2012). "Bipolar-switching model of RRAM by field- and temperature-activated ion migration," in *Proceeding International Memory Workshop* (Milan: IEEE), 1–4. doi: 10.1109/IMW.2012.6213648

Lee, H. Y., Chen, P. S., Wu, T. Y., Chen, Y. S., Wang, C. C., Tzeng, P. J., et al. (2008). "Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust HfO_2 based RRAM," in *Proceeding International Electron Devices Meeting* (San Francisco, CA: IEEE), 297–300. doi: 10.1109/IEDM.2008.4796677

Lee, J., Shin, J., Lee, D., Lee, W., Jung, S., Jo, M., et al. (2010). "Diode-less nano-scale ZrOx/HfOx RRAM device with excellent switching uniformity and reliability for high-density crossbar memory applications," in *Proceeding International Electron Devices Meeting* (San Francisco, CA: IEEE), 452–453. doi: 10.1109/IEDM.2010.5703393

Lee, M.-J., Lee, C. B., Lee, D., Lee, S. R., Chang, M., Hur, J. H., et al. (2011). A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta_2O_{5-x}/TaO_{2-x} bilayer structures. *Nat. Mater.* 10, 625–630. doi: 10.1038/nmat3070

Lee, M.-J., Lee, D., Kim, H., Choi, H.-S., Park, J.-B., Kim, H. G., et al. (2012). "Highly-scalable threshold switching select device based on chaclogenide glasses for 3D nanoscaled memory arrays," in *Proceeding International Electron Devices Meeting* (San Francisco, CA: IEEE), 33–35. doi: 10.1109/IEDM.2012.6478966

Li, B., Majafi, M. H., and Lilja, D. J. (2016). "Using stochastic computing to reduce the hardware requirements for a restricted Boltzmann machine classifier," in *Proceeding International Symposium on Field-Programmable Gate Arrays* (Monterey: ACM/SIGDA), 36–41. doi: 10.1145/2847263.2847340

Li, C., Gao, B., Yao, Y., Guan, X., Shen, X., Wang, Y., et al. (2017). Direct observations of nanofilament evolution on switching processes in HfO_2-based resistive random access memory by in situ TEM studies. *Adv. Mater.* 29:1602976. doi: 10.1002/adma.201602976

Li, H., Chen, Z., Ma, W., Gao, B., Huang, P., Liu, L., et al. (2015b). Nonvolatile logic and in situ data transfer demonstrated in crossbar resistive RAM array. *IEEE Electron Device Lett.* 36, 1142–1145. doi: 10.1109/LED.2015.2481439

Li, H., Gao, B., Chen, Z., Zhao, Y., Huang, P., Ye, H., et al. (2015a). A learnable parallel processing architecture towards unity of memory and computing. *Sci. Rep*. 5:13330. doi: 10.1038/srep13330

Li, K. S., Ho, C., Lee, M.-T., Chen, M.-C., Hsu, C.-L., Lu, J. M., et al. (2014). "Utilizing sub-5 nm sidewall electrode technology for atomic-scale resistive memory fabrication," in *Proceeding Symposium on VLSI Technology* (Honolulu: IEEE), 164–165.

Li, P., Lilja, D. J., Qian, W., Bazargan, K., and Riedel, M. D. (2014). Computation on stochastic bit streams digital image processing case studies. *IEEE Trans. Very Large Scale Integrat. Syst.* 22, 449–462. doi: 10.1109/TVLSI.2013.2247429

Li, Z., Ren, A., Li, J., Qiu, Q., Wang, Y., and Yuan, B. (2016). "DSCNN: Hardware-oriented optimization for stochastic computing based deep convolutional neural networks," in *Proceeding International Conference on Computer Design* (Scottsdale: IEEE), 678–681. doi: 10.1109/ICCD.2016.7753357

Liao, Y., Gao, B., Xu, F., Yao, P., Chen, J., Zhang, W., et al. (2020). A compact model of analog RRAM with device and array nonideal effects for neuromorphic systems. 67, 1593–1599. doi: 10.1109/TED.2020.2975314

Lv, Y., and Wang, J.-P. (2017). "A single magnetic-tunnel-junction stochastic computing unit," in *Proceeding International Electron Devices Meeting* (San Francisco, CA: IEEE), 800–803. doi: 10.1109/IEDM.2017.8268504

Ma, C., Zhong, S., and Dang, H. (2012). "High fault tolerant image processing system based on stochastic computing," in *Proceeding International Conference on Computer Science and Service System* (Nanjing: IEEE), 1587–1590. doi: 10.1109/CSSS.2012.397

Moons, B., and Verhelst, M. (2014). Energy-efficiency and accuracy of stochastic computing circuits in emerging technologies. *IEEE J. Emerg. Select. Topics Circu. Syst.* 4, 475–486. doi: 10.1109/JETCAS.2014.2361070

Mott, F., and Davis, E. A. (1972). Electronic processes in non-crystalline materials. *Phys. Today* 25:55. doi: 10.1063/1.3071145

Pan, F., Gao, S., Chen, C., Song, C., and Zeng, F. (2014). Recent progress in resistive random access memories: materials, switching mechanisms, and performance. *Mater. Sci. Eng. R.* 83, 1–59. doi: 10.1016/j.mser.2014.06.002

Park, S., Sheri, A., Kim, J., Noh, J., Jang, J., Jeon, M., et al. (2013). "Neuromorphic speech systems using advanced ReRAM-based synapse," in *Proceeding International Electron Devices Meeting* (Washington, DC: IEEE), 625–628. doi: 10.1109/IEDM.2013.6724692

Pedretti, G., Bianchi, S., Milo, V., Calderoni, A., Ramaswamy, N., and Ielmini, D. (2017). "Modeling-based design of brain-inspired spiking neural networks with RRAM learning synapses," in *Proceeding International Electron Devices Meeting* (San FranciscoCA: IEEE), 653–656. doi: 10.1109/IEDM.2017.8268467

Philip Wong, H. S., and Salahuddin, S. (2015). Memory leads the way to better computing. *Nat. Nanotechnol*. 10, 191–194. doi: 10.1038/nnano.2015.29

Prezioso, M., Bayat, F. M., Hoskins, B., Likharev, K., and Strukov, D. (2016). Self-adaptive spike-time-dependent plasticity of metaloxide memristors. *Sci. Rep*. 6:21331. doi: 10.1038/srep21331

Prezioso, M., Merrikh-Bayat, F., Hoskins, B. D., Adam, G. C., Likharev, K. K., and Strukov, D. B. (2015). Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* 521, 61–64. doi: 10.1038/nature14441

Qin, H., Gong, R., Liu, X., Bai, X., Song, J., and Sebe, N. (2020). Binary neural networks: a survey. *Pattern Recognition* 107281. doi: 10.1016/j.patcog.2020.107281

Russo, U., Ielmini, D., Cagli, C., and Lacaita, A. L. (2009). Self-accelerated thermal dissolution model for reset programming in unipolar resistive switching memory (RRAM) devices. *IEEE Trans. Electron Devices* 56, 193–200. doi: 10.1109/TED.2008.2010584

Russo, U., Ielmini, D., Cagli, C., Lacaita, A. L., Spiga, S., Wiemer, C., et al. (2007). "Conductive-filament switching analysis and self-accelerated thermal dissolution model for reset in NiO-based RRAM," in *Proceeding International Electron Devices Meeting* (Washington, DC: IEEE), 775–778. doi: 10.1109/IEDM.2007.4419062

Sawa, A. (2008). Resistive switching in transition metal oxides. *Mater. Today* 11, 28–36. doi: 10.1016/S1369-7021(08)70119-6

Shen, W., Huang, P., Fan, M., Han, R., Zhou, Z., Gao, B., et al. (2019). Stateful logic operations in one-transistor-one-resistor resistive random access memory array. *IEEE Electron Device Lett.* 40, 1538–1541. doi: 10.1109/LED.2019.2931947

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Driessche, G., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 484–489. doi: 10.1038/nature16961

Strukov, D. B., Snider, G. S., Stewart, D. R., and Williams, R. S. (2008). The missing memristor found. *Nature* 453, 80–83. doi: 10.1038/nature06932

Suri, M., Querlioz, D., Bichler, O., Palma, G., Vianello, E., Vuillaume, D., et al. (2013). Bio-inspired stochastic computing using binary CBRAM synapses. *IEEE Trans. Electron Devices* 60, 2402–2409. doi: 10.1109/TED.2013.2263000

Wang, X. P., Fang, Z., Li, X., Chen, B., Gao, B., Kang, J. F., et al. (2012). "Highly compact 1T-1R architecture ($4F^2$ footprint) involving fully CMOS compatible vertical GAA nano-pillar transistors and oxide-based RRAM cells exhibiting excellent NVM properties and ultra-low power operation," in *Proceeding International Electron Devices Meeting* (San Francisco, CA: IEEE), 493–496. doi: 10.1109/IEDM.2012.6479082

Wang, Z., Joshi, S., Savel've, S. E., Jiang, H., Midya, R., Lin, P., et al. (2017). Memristors with diffusive dynamics as synaptic emulators for brain-inspired computing. *Nat. Mater.* 16, 101–108. doi: 10.1038/nmat4756

Wang, Z., Midya, R., Joshsi, S., Jiang, H., Li, C., Lin, P., et al. (2018). "Unconventional computing with diffusive memristors," in *Proceeding International Symposium on Circuits and Systems* (Florence: IEEE), 1–5. doi: 10.1109/ISCAS.2018.8351882

Wang, Z., Yin, M., Zhang, T., Cai, Y., Wang, Y., Yang, Y., et al. (2016). Engineering incremental resistive switching in TaOx based memristors for brain-inspired computing. *Nanoscale* 8:14015. doi: 10.1039/C6NR00476H

Waser, R., Dittmann, R., Staikov, G., and Szot, K. (2009). Redox-based resistive switching memories - nanoionic mechanisms, prospects, and challenges. *Adv. Mater*. 21, 2632–2663. doi: 10.1002/adma.200900375

Wei, Z., Kanzawa, Y., Arita, K., Katoh, Y., Kawai, K., Muraoka, S., et al. (2008). "Highly reliable TaOx ReRAM and direct evidence of redox reaction mechanism," in *Proceeding International Electron Devices Meeting* (San Francisco, CA: IEEE), 293–296. doi: 10.1109/IEDM.2008.4796676

Wong, H. S. P., Lee, H.-Y., Yu, S., Chen, Y.-S., Wu, Y., Chen, P.-S., et al. (2012). Metal-oxide RRAM. *Proc. IEEE* 100, 1951–1970. doi: 10.1109/JPROC.2012.2190369

Woo, J., Moon, K., Song, J., Kwak, M., Park, J., and Hwang, H. (2016b). Optimized programming scheme enabling linear potentiation in filamentary HfO2 RRAM synapse for neuromorphic systems. *IEEE Trans. Electron Devices* 63, 5064–5067. doi: 10.1109/TED.2016.2615648

Woo, J., Moon, K., Song, J., Lee, S., Kwak, M., Park, J., et al. (2016a). Improved synaptic behavior under identical pulses using AlOx/HfO_2 bilayer RRAM array for neuromorphic systems. *IEEE Electron Device Lett*. 37, 994–997. doi: 10.1109/LED.2016.2582859

Wu, H., Wang, X. H., Gao, B., Deng, N., Lu, Z., Haukness, B., et al. (2017). Resistive random access memory for future information processing system. *Proc. IEEE* 105, 1770–1789. doi: 10.1109/JPROC.2017.2684830

Wu, W., Wu, H., Gao, B., Deng, N., Yu, S., and Qian, H. (2017). Improving analog switching in HfOx-based Resistive memory with a thermal enhanced layer. *IEEE Electron Device Lett*. 38, 1019–1022. doi: 10.1109/LED.2017.2719161

Wu, W., Wu, H., Gao, B., Yao, P., Zhang, X., Peng, X., et al. (2018). "A methodology to improve linearity of analog RRAM for brain-inspired computing," in *Proceeding Symposium on VLSI Technology* (Hawaii: IEEE), 103–104. doi: 10.1109/VLSIT.2018.8510690

Wu, Y., Yu, S., Wong, H. S. P., Chen, Y.-S., Lee, H.-Y., Wang, S.-M., et al. (2012). "AlOx-based resistive switching device with gradual resistance modulation for neuromorphic device application," in *Proceeding International Memory Workshop* (Milan: IEEE), 1–4. doi: 10.1109/IMW.2012.6213663

Xiang, Y., Huang, P., Zhao, Y., Zhao, M., Gao, B., Wu, H., et al. (2019). Impacts of state instability and retention failure of filamentary analog RRAM on the performance of deep neural network. *IEEE Trans. Electron Devices* 66, 4517–4522. doi: 10.1109/TED.2019.2931135

Xu, F., Gao, B., Xi, Y., Tang, J., Wu, H., and Qian, H. (2020). "Atomic-device hybrid modeling of relaxation effect in analog RRAM for neuromorphic computing," in *Proceeding International Electron Devices Meeting* (Online: IEEE), 263–266. doi: 10.1109/IEDM13553.2020.9372114

Yang, J. J., Strukov, D. B., and Stewart, D. R. (2013). Memristive devices for computing. *Nat. Nanotechnol*. 8, 13–24. doi: 10.1038/nnano.2012.240

Yang, M., Hayes, J. P., Fan, D., and Qian, W. (2017). "Design of accurate stochastic number generators with noisy emerging devices for stochastic computing," in *Proceeding International Conference on Computer-Aided Design* (Irvine: IEEE/ACM), 638–644. doi: 10.1109/ICCAD.2017.8203837

Yang, T.-J., and Sze, V. (2019). "Design considerations for efficient deep neural networks on processing-in-memory accelerators," in *Proceeding International Electron Devices Meeting* (San Francisco, CA: IEEE), 514–517. doi: 10.1109/IEDM19573.2019.8993662

Yang, Y., Gao, P., Gaba, S., Chang, T., Pan, X., and Lu, W. (2012). Observation of conductive filament growth in nanoscale resistive memories. *Nat. Commun*. 3:732. doi: 10.1038/ncomms1737

Yu, S. (2018). Neuro-inspired computing with emerging nonvolatile memory. *Proc. IEEE* 106, 260–285. doi: 10.1109/JPROC.2018.2790840

Yu, S., Gao, B., Fang, Z., Yu, Y., Kang, J., and Wong, H. S. P. (2012). "A neuromorphic visual system using RRAM synaptic devices with sub-pJ energy and tolerance to variability: experimental characterization and large-scale modeling," in *Proceeding International Electron Devices Meeting,* San Francisco, CA: IEEE), 239–242. doi: 10.1109/IEDM.2012.6479018

Yu, S., Wu, Y., Chai, Y., Provine, J., and Philip Wong, H. S. (2011). "Characterization of switching parameters and multilevel capability in HfOx/AlOx bi-layer RRAM devices," in *Proceeding VLSI Technology, Systems and Applications* (Hsinchu: IEEE), 1–2. doi: 10.1109/VTSA.2011.5872251

Yun, J.-B., Kim, S., Seo, S., Lee, M.-J., Kim, D.-C., Ahn, S.-E., et al. (2007). Random and localized resistive switching observation in Pt/NiO/Pt. *Phys. Stat. Sol.* 1, 280–282. doi: 10.1002/pssr.200701205

Zhao, Y., Huang, P., Chen, Z., Liu, C., Li, H., Chen, B., et al. (2016). Modeling and optimization of bilayered TaOx-RRAM based on defect evolution and phase transition effects. *IEEE Trans. Electron Devices* 63, 1524–1532. doi: 10.1109/TED.2016.2532470

Zhao, Y., Shen, W., Huang, P., Xu, W., Fan, M., Liu, X., et al. (2019). "A Physics-based model of RRAM probabilistic switching for generating stable and accurate stochastic bit-streams," in *Proceeding International Electron Devices Meeting* (San Francisco, CA: IEEE), 767–770. doi: 10.1109/IEDM19573.2019.8993559

Zhao, Y. D., Huang, P., Chen, Z., Liu, C., Li, H. T., Ma, W. J., et al. (2015). "Understanding the underlying physics of superior endurance in bi-layered TaOx-RRAM," in *Proceeding Silicon Nanoelectronics Workshop* (Kyoto: IEEE), 1–2.

Zhou, Z., Huang, P., Xiang, Y. C., Shen, W. S., Zhao, Y. D., Feng, Y. L., et al. (2018). "A new hardware implementation approach of BNNs based on nonlinear 2T2R synaptic cell," in *Proceeding International Electron Devices Meeting* (San Francisco, CA: IEEE), 488–491. doi: 10.1109/IEDM.2018.8614642

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.

frontiers in Nanotechnology

ORIGINAL RESEARCH
published: 12 May 2021
doi: 10.3389/fnano.2021.633026

System-Theoretic Methods for Designing Bio-Inspired Mem-Computing Memristor Cellular Nonlinear Networks

Alon Ascoli[1,2]*, *Ronald Tetzlaff*[1,2], *Sung-Mo Steve Kang*[3] and *Leon Chua*[4]

[1]Fundamentals of Electrical Engineering, Institute of Circuits and Systems, Faculty of Electrical and Computer Engineering, Technische Universität, Dresden, Germany, [2]Department of Microelectronics, Brno University of Technology, Brno, Czech Republic, [3]Department of Electrical and Computer Engineering, University of California Santa Cruz, Santa Cruz, CA, United States, [4]Department of Electrical Engineering and Computer Sciences, University of California Berkeley, Berkeley, CA, United States

The introduction of nano-memristors in electronics may allow to boost the performance of integrated circuits beyond the Moore era, especially in view of their extraordinary capability to process and store data in the very same physical volume. However, recurring to nonlinear system theory is absolutely necessary for the development of a systematic approach to memristive circuit design. In fact, the application of linear system-theoretic techniques is not suitable to explore thoroughly the rich dynamics of resistance switching memories, and designing circuits without a comprehensive picture of the nonlinear behaviour of these devices may lead to the realization of technical systems failing to operate as desired. Converting traditional circuits to memristive equivalents may require the adaptation of classical methods from nonlinear system theory. This paper extends the theory of time- and space-invariant standard cellular nonlinear networks with first-order processing elements for the case where a single non-volatile memristor is inserted in parallel to the capacitor in each cell. A novel nonlinear system-theoretic method allows to draw a comprehensive picture of the dynamical phenomena emerging in the memristive mem-computing array, beautifully illustrated in the so-called Primary Mosaic for the class of uncoupled memristor cellular nonlinear networks. Employing this new analysis tool it is possible to elucidate, with the support of illustrative examples, how to design variability-tolerant bio-inspired cellular nonlinear networks with second-order memristive cells for the execution of computing tasks or of memory operations. The capability of the class of memristor cellular nonlinear networks under focus to store and process information locally, without the need to insert additional memory units in each cell, may allow to increase considerably the spatial resolution of state-of-the-art purely CMOS sensor-processor arrays. This is of great appeal for edge computing applications, especially since the Internet-of-Things industry is currently calling for the realization of miniaturized, lightweight, low-power, and high-speed mem-computers with sensing capability on board.

Keywords: memristor, bio-inspired mem-computing machines, cellular nonlinear networks, nonlinear circuit theory, nonlinear system theory

OPEN ACCESS

Edited by:
Huanglong Li,
Tsinghua University, China

Reviewed by:
Carlos Sánchez-López,
Autonomous University of Tlaxcala, Mexico
Xiaoping Wang,
Huazhong University of Science and Technology, China

***Correspondence:**
Alon Ascoli
alon.ascoli@tu-dresden.de

Specialty section:
This article was submitted to Nanodevices, a section of the journal Frontiers in Nanotechnology

Received: *24 November 2020*
Accepted: *25 January 2021*
Published: *12 May 2021*

Citation:
Ascoli A, Tetzlaff R, Kang S-MS and Chua L (2021) System-Theoretic Methods for Designing Bio-Inspired Mem-Computing Memristor Cellular Nonlinear Networks. Front. Nanotechnol. 3:633026. doi: 10.3389/fnano.2021.633026

1 INTRODUCTION

On August 28th, 2018 GlobalFoundries announced to halt the 7 nm chip development. After installing at least one Extreme-Ultraviolet Lithography (EUV) machine at one of its fabs, the foundry reckoned there would not be enough customers, interested in the cutting-edge 7 nm node technology process, to make chip development profitable (GlobalFoundries Ltd, 2018). Despite Intel, Samsung, and TSMC are still making efforts to reduce the size of *integrated circuits* (ICs) further, transistor scaling is approaching atomic boundaries, with an inevitable concurrent rise in manufacturing costs. This issue is known as *Moore wall*. With the Moore era (Moore, 1965) coming to a natural end (Williams, 2017), a great deal of resources have been deployed over the past decade toward the development of innovative disruptive nanotechnologies, which may enable the development of versatile multi-functional devices, that, opening the door toward the implementation of peculiar signal processing paradigms, would allow to boost the performance of conventional circuits and systems without the need to shrink the size of transistors any further. Two are the factors for the inefficiency of machines based upon the von Neumann architecture: 1) There is a large mismatch between processing time and shuttling time. This issue is known as *memory wall* or *von Neumann bottleneck*. 2) The energy dissipated by digital switching units is no longer following the exponentially decreasing trend, predicted by Landauer (Landauer, 1988), with the reduction in IC dimensions. This issue, known as *heat wall*, poses serious risks for the lifetime of transistors. Memristors (Chua, 1971; Chua and Kang, 1976) represent one of the most promising nanotechnologies to address the problems affecting state-of-the-art electronics. A current (voltage)-controlled non-volatile memristor (Chua, 2014; Chua, 2015) is a two-terminal device, whose resistance (conductance) can be tuned to some desired value by applying a current (voltage) signal through (across) it, and which remembers its resistance (conductance) after the current (voltage) source, in parallel to it, is disconnected (Chua, 2018a). It remembers its past! (Chua, 2018b). The most impressive and peculiar virtue of non-volatile memristors is the combined capability to store data, thanks to excellent data retention levels, achievable without the need for external batteries, and to process signals, through the rich nonlinear dynamics of the memory state, within a single nano-scale volume, which enables the implementation of *in-memory computing* and *mem-computing* paradigms[1], mimicking the distributed nature of memory and processing operations in the human brain, in future computing machines. Other distinctive qualities of memristors are low-power and high-speed operation, superior endurance, and, very importantly, good compatibility with CMOS technology. While in conventional memories data are stored as voltage levels, in memristors the physical quantity, which holds the information content, is the resistance. Given that all nonvolatile memories based upon resistance switching phenomena, irrespective of their constitutive materials and operating principles, are memristors (Chua, 2011), a wide range of nanotechnologies, including Resistive Random Access Memories (ReRAMs), Phase Change Memories (PCMs), Magnetic Tunnel Junctions (MTJs), Spin-Transfer-Torque Magneto-Resistive Random Access Memories (STTM-RRAMs), and Ferroelectric Tunnel Junctions (FTJs), are competing one with the other to produce the best performing data storage device for future brain-like computers. While many people believe that non-volatile nano-memristors will eventually replace conventional memories, including Flash Memories, Dynamic RAMs (DRAMs), and Hard Disk Drives (HDDs), the aforementioned nanotechnologies are not yet mature enough to draw conclusions on the portion of the nonvolatile memory market, which memristors will be able to cover in the next five years from now. However, *edge computing* technical systems already make use of memristive memories. Panasonic (Panasonic Ltd., 2013) have been launching mass production of micro-computers with 64 kB ReRAM-based data storage for battery-powered equipment, including portable devices for medical, healthcare, and security applications already in 2013, while Fujitsu (Fujitsu Ltd, 2019) has recently taken a step forward by offering 1024 kB ReRAMs for wearable units–e.g., smart watches and glasses–and hearing aids. Even when used simply as tunable resistors, memristors offer unique opportunities to enhance the performance of conventional data processing systems. Most computing tasks in artificial intelligence (AI) applications consist of machine-learning operations, such as object, image, and speech recognition, which require the calculation of a massive number of vector-matrix multiplications (VMMs). Nowadays these calculations are executed through expensive and bulky supercomputers. But with the advent of the memristor, which, leveraging Ohm's law, naturally carries out a multiplication operation between the conductance it holds and the voltage falling between its terminals, outputting the result into the current flowing through it, it is possible to use a crossbar array (Li et al., 2018) to compute at unprecedented rates the product between a vector of voltages, distributed along the rows, and a matrix of memristor crosspoint conductances, with the computation result available in current form along the columns (refer to the Dot Product Engine (DPE) lab prototype developed at Hewlett Packard Enterprise (Hu et al., 2016)). Last but not least, given that the two constitutive elements of the human brain, namely the synapse and the neuron, are made of nonvolatile and volatile memristors, respectively, resistance switching memories allow to develop innovative neuromorphic circuits, which promise to outperform conventional purely CMOS counterparts in mimicking the functionalities of the human brain. Non-volatile memristor devices, in which the resistance may be

[1]*In-memory computing* refers to the partial/temporary use of data storage units for information processing purposes, while *mem-computing* is associated to the adoption of computing systems to perform memory read/write operations on demand, as is the case for the processing elements of the proposed memristive cellular array.

finely tuned under excitation, may reproduce most closely the plastic response of biological synapses to external stimuli. Furthermore, there exist a large class of memristor physical nano-scale realisations, which, despite being unable to store data–for this reason they are classified as volatile memories –, feature the extraordinary capability to amplify infinitesimal fluctuations in energy (Bohaichuk et al., 2019; Kumar et al., 2020), a property which is referred to as *local activity* (Chua, 2005), and which enables the development of realistic electronic realisations of spiking neurons[2] (Chua et al., 2012), the so-called neuristors.

Besides constitutive the ideal framework for modeling biological systems (Chua, 1998; Chua and Roska, 2002), *Cellular Nonlinear Networks* (CNNs) (Chua and Yang, 1988a; Chua and Yang, 1988b) represent a powerful multi-variate signal processing paradigm, which, featuring a bio-inspired architecture, operates in a massively parallel fashion, allowing to process data at very high rates, as necessary in time-critical *Internet-of-Things* (IoT) applications, nowadays. Purely CMOS analogue hardware implementations of the CNN signal processing paradigm are typically co-integrated with highly selective equal-sized sensor arrays to allow the solution of complex computing tasks directly where the acquisition of specific data takes place (Vázquez et al., 2018). A technological issue, which limits the applicability scope of these sensor-processor arrays, is related to the huge difference between the typically small minimum size of an element of the sensor matrix, and the relatively large minimum *integrated circuit* (IC) area, which a processing element of the CNN hardware realization usually occupies, due to the fact that it needs to accommodate memory units, which endow the resulting computing machine with local stored programmability on board, allowing to harness thoroughly the advantages associated with the massive parallelism of the CNN signal processing paradigm. The adoption of non-volatile memristors (Chua, 2015), capable to combine data processing and storage functionalities within a common nanoscale physical volume, in CNN circuit design may allow to increase significantly the spatial resolution of the cellular computing machine. Moreover, leveraging the rich nonlinear dynamics of resistance switching memories, the computing capabilities of the processing elements of a memristive CNN hardware implementation (Duan et al., 2015; Di Marco et al., 2017a; Di Marco et al., 2017b; Di Marco et al., 2018) may be extended beyond the operational boundaries of the cells of a traditional purely CMOS implementation.

CNNs process information through the analogue dynamics of the cells' states, which converge toward distinct attractors depending upon inputs and/or initial conditions. While *wave-based computing*, where the cellular array carries out data processing tasks through the generation of specific dynamic patterns, is an active field of research (Weiher et al., 2019), there exists a huge library (Karacs et al., 2018) of image processing operations, which the nonlinear dynamic array may execute as the cells' states approach predefined equilibria. This paper focuses on the performance of CNNs (M-CNNs) as *equilibria-based computing (mem-computing)* engine. Now, for a full exploration of the potential of memristors in electronics, recurring to concepts from *nonlinear system theory* is necessary. In fact, linear system-theoretic methods are not suitable for the analysis and design of memristor-based circuits. However, as is the case here, converting traditional nonlinear circuits to memristive equivalents may require the extension of classical nonlinear system-theoretic techniques. The *Memristor Cellular Nonlinear Network* (M-CNN), proposed in Tetzlaff et al. (2020), differs from a standard time- and space-invariant two-dimensional CNN (Chua and Yang, 1988a; Chua and Yang, 1988b), characterized by first-order cells, and typically implemented in hardware (Vázquez et al., 2018), for the inclusion of a single non-volatile memristor in parallel to the capacitor in the circuit implementation of each processing element. One of the most powerful tools for the analysis of nonlinear dynamical systems with one degree of freedom is the *Dynamic Route Map* (DRM) (Chua, 2018a), which represents the system-theoretic technique of reference for the investigation of CNNs with first-order processing elements. Since the memristive cell in the proposed M-CNN features two degrees of freedom, the investigation of the cellular array calls for the generalization of the DRM graphical tool, applicable to first-order systems only. The modified DRM graphical tool, applicable to second-order dynamical systems, is known as *Second-Order Dynamic Route Map* (DRM_2) (Tetzlaff et al., 2020). The application of this novel system-theoretic technique to the model of the proposed M-CNN allows to gain a deep insight into the rich nonlinear behaviour of its second-order processing elements, unveiling dynamical phenomena, which may not emerge in the original cellular array (Ascoli et al., 2020b). The DRM_2 graphical tool lies at the basis of a systematic methodology to design variability-tolerant mem-computing arrays with second-order memristive cells (Ascoli et al., 2020a).

The structure of the paper is organized as follows. **Section 2** revisits the theory of CNNs, explaining the invaluable role of the classical DRM graphical technique to analyze standard arrays of locally coupled processing elements, and elucidating through an illustrative example the traditional method, based upon this system-theoretic tool, to program the cellular computing engine for the execution of a predefined image processing task. **Section 3** first defines the class of M-CNNs under study, including the model of the non-volatile memristor hosted in each cell, secondly extends the DRM graphical tool to second-order dynamical systems, elucidates how this allows to draw a comprehensive picture of the nonlinear dynamics of each memristive processing element, and finally presents a rigorous procedure, based upon the DRM_2 system-theoretic technique, to design cellular mem-computing structures with second-order memristive cells. Sections 4 and 5 are devoted to the application of the M-CNN design methodology for operating the multifunctional memristive cellular computing engine as image processing system and as memory bank, respectively. A brief discussion, summarizing the significance of the research work, is provided in **section 6**. Conclusions are finally drafted in **section 7**.

[2]Interestingly, it has been recently shown (Zhang et al., 2020) that the Cardiac Purkinje Fiber (CPF), which is the last branch of the heart conduction system, may be described via a modified variant of the memristive Hodgkin-Huxley equations, revealing the ubiquituous presence of memristors in living cells.

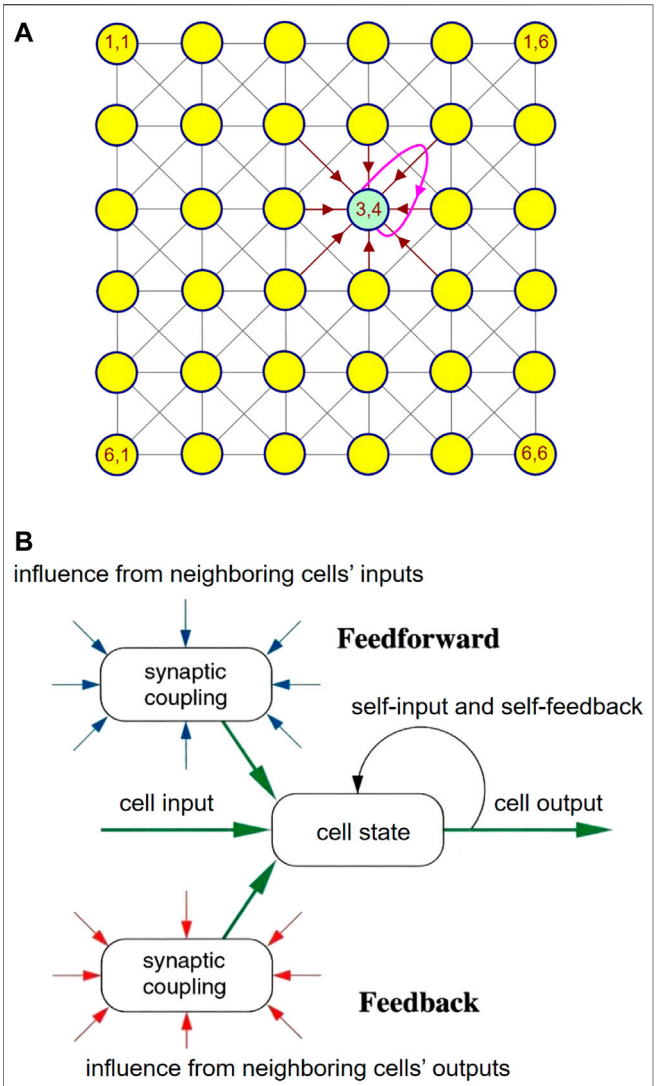


FIGURE 1 | (A) Physical connectivity among the locally coupled cells of a two-dimensional CNN with six rows and six columns. **(B)** Schematic illustration of the main features of a CNN cell, revealing some of its analogies with a biological neuron, which explains the nomenclature originally introduced to characterize the locally coupled nonlinear dynamic array, namely *Cellular Neural Network*.

2 ANALYSIS AND DESIGN OF MEMRISTOR CELLULAR NONLINEAR NETWORKS

Cellular Nonlinear Networks (CNNs) constitute a bio-inspired multivariate signal processing paradigm, which, based upon a massively parallel information flow, enables computations at very high rates, and is amenable to a Very Large Scale Integration (VLSI) circuit realization, which, centered around a non-von-Neumann machine architecture, enables computational universality. Introducing memristive devices in CNN VLSI design may provide two main benefits. Firstly, the rich spectrum of nonlinear dynamic phenomena, appearing in resistance switching memories, may simplify or extend the functionalities of traditional CNNs. Secondly, the unique combined capability of nonvolatile memristors to compute and store data within the same nanoscale physical medium may render unnecessary the need to include spacious data storage units within the circuit implementation of each cell, allowing to improve considerably the number of processing elements fitting into the IC design area allocated to the non-von-Neumann computing machine.

2.1 Theory of Cellular Nonlinear Networks

The theoretical foundations of CNNs were laid in 1988 by L. Chua (Chua and Yang, 1988a; Chua and Yang, 1988b). In the most general case, a CNN consists of a spatially discrete collection of locally coupled k^{th}-order continuous-time processing elements, called cells, arranged at regular positions within a l-dimensional lattice. The architecture of a small two-dimensional CNN with $M = 6$ rows and $N = 6$ columns is presented in **Figure 1A**, under the assumption that each cell $C(i,j)$ – $i \in \{1,\ldots,M\}, j \in \{1,\ldots,N\}$ – is physically coupled to its 8 adjacent neighbors only[3]. Each cell is assigned a state, an input, an output, as well as a threshold. The rate of change of the state of a cell is influenced by the inputs and outputs of its 8 adjacent neighbors, as well as by its own input and output, as respectively sketched through eight brown directed segments and through one magenta directed loop in **Figure 1A** for the processing element located where the 3rd row crosses the 4th column. The block diagram in plot (b) of **Figure 1** illustrates once more the key factors affecting the dynamical behaviour of a CNN cell. The neighbors' inputs (outputs) are modulated by feedforward (feedback) synaptic weights before accessing the cell to affect the time evolution of its state, similarly as it occurs in biological neural networks. The cell $C(i,j)$ of a standard time- and space-invariant two-dimensional CNN (Chua and Roska, 2002) is implemented by the circuit of **Figure 2**, where the computing core is mathematically described by[4] ($i \in \{1,\ldots,M\}, j \in \{1,\ldots,N\}$)

$$\frac{dv_{x_{i,j}}}{dt} = -\frac{v_{x_{i,j}}}{C_{x_{i,j}} \cdot R_{x_{i,j}}} + \frac{z \cdot I}{C_{x_{i,j}}} + \frac{1}{C_{x_{i,j}}} \cdot \sum_{k=-r}^{k=r} \sum_{l=-r}^{l=r} \left(i_{a_{k,l}} + i_{b_{k,l}}\right), \qquad (1)$$

in case it is physically coupled to its 8 adjacent neighbors only[5] i.e., $r = 1$. With reference to the circuit of **Figure 2**, the main physical quantity within the input stage, the computing core, and the output stage respectively are the input voltage $v_{u_{i,j}}$, the voltage

[3]A CNN cell, exhibiting physical couplings to the eight closest processing elements, is said to have a sphere of influence of unitary radius, or, alternatively, a 3×3 local neighborhood.

[4]The determination of a numerical solution for the $M{\times}N$ ODEs, governing the time evolution of the states of all the processing elements, requires the preliminary assignment of an *initial condition* $v_{x_{i,j}}(0)$ to each cell $C(i,j)$, as well as the preparatory specification of the *boundary conditions* (Chua and Roska, 2002), fixing the input voltage $v_{u_{m,n}}$ and the output voltage $v_{y_{m,n}}$ of each *virtual cell*. With reference to a two-dimensional CNN, in which each cell exhibits a sphere of influence of unitary radius, a processing element $C(m,n)$ is said to be virtual if it does not belong to the nonlinear dynamic array i.e., $m \notin \{1,\ldots,M\}$ and/or $n \notin \{1,\ldots,N\}$, but is part of the 3×3 neighborhood of a cell, which belongs to the cellular network, being listed in the set $\{C(1,1:N), C(M,1:N), C(2:M-1,1), C(2:M-1,N)\}$.

[5]It is important to note that only CNN hardware realizations with such a basic coupling configuration have been developed so far (Vázquez et al., 2018).

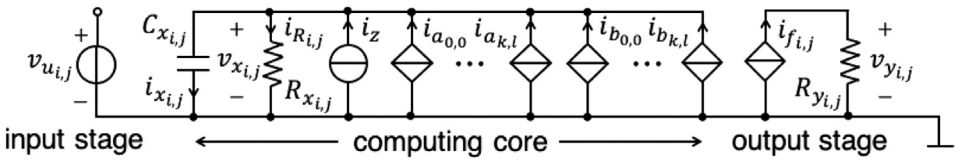


FIGURE 2 | Input stage, computing core, implementing the state **Eq. 1**, and output stage of the circuit realization of a standard space-invariant CNN cell $C(i,j)$.

$v_{x_{i,j}}$ across a capacitor with capacitance $C_{x_{i,j}}\mathbb{R}_{>0}$, expressing the state, and the output voltage $v_{y_{i,j}}$. Focusing on the output stage, the latter physical quantity is defined as

$$v_{y_{i,j}} = R_{y_{i,j}} \cdot i_{f_{i,j}}, \tag{2}$$

where $R_{y_{i,j}}\mathbb{R}_{>0}$ is the resistance of a passive linear resistor, whereas $i_{f_{i,j}}$ is the current of a voltage-controlled source, featuring the piecewise linear expression

$$i_{f_{i,j}} \triangleq i_f\left(v_{x_{i,j}}\right) = g_{lin} \cdot \frac{\left|v_{x_{i,j}} + v_{sat}\right| - \left|v_{x_{i,j}} - v_{sat}\right|}{2}, \tag{3}$$

generally known as *standard nonlinearity*, where g_{lin} and v_{sat} are positive parameters with units Ω^{-1} and V, respectively. Importantly, the piecewise-linear standard nonlinearity identifies three domains, specifically the *negative saturation region* $v_{x_{i,j}} < -v_{sat}$, the *linear region* $\left|v_{x_{i,j}}\right| \leq v_{sat}$, and the *positive saturation region* $v_{x_{i,j}} > +v_{sat}$, where the cell output voltage $v_{y_{i,j}}$ attains the negative saturation voltage $-v_{sat}$, is a linear function of the state $v_{x_{i,j}}$, and attains the positive saturation voltage v_{sat}, respectively. Inspecting now the computing core in the cell circuit of **Figure 2**, i_z, defined as

$$i_z \triangleq z \cdot I, \tag{4}$$

where z is a dimensionless parameter referred to as *threshold* in CNN theory (Chua and Roska, 2002), while I is a coefficient with positive 1 A value, symbolizes the threshold current. Further, $R_{x_{i,j}}\mathbb{R}_{>0}$ stands for the resistance of a passive linear resistor[6], while, importantly, $i_{a_{k,l}}$ ($i_{b_{k,l}}$), defined as

$$i_{a_{k,l}} \triangleq a_{k,l} \cdot v_{y_{i+k,j+l}} \tag{5}$$

$$i_{b_{k,l}} \triangleq b_{k,l} \cdot v_{u_{i+k,j+l}}, \tag{6}$$

where $a_{k,l}$ ($b_{k,l}$), with $k,l \in \{-1,0,1\}$, is known as *feedback (feedforward) synaptic weight* in CNN theory (Chua and Roska, 2002), represents the feedback (feedforward) synaptic current due to the neighboring cell $C(i+k,j+l)$, and constituting one of the 18 contributions to the cell capacitor current $i_{x_{i,j}}$ enclosed within the round brackets to the right of the double sum in **Eq. 1**. It is worth mentioning that, the CNN mathematical description reported in Eq. 1 is known as Chua-Yang model (Chua and Yang, 1988a; Chua and Yang, 1988b). Despite an alternative mathematical description, known as Full-Range model (Vázquez et al., 1993), facilitates the hardware implementation of the CNN paradigm by restricting the set of allowable values for the cells' states, this paper adopts the original Chua-Yang mathematical description for pedagogical purposes.

Typically, the right hand side of the CNN cell state **Eq. 1** may be recast as

$$\frac{dv_{x_{i,j}}}{dt} = \frac{i_{g_{i,j}} + i_{w_{i,j}}}{C_x}, \tag{7}$$

where $i_{g_{i,j}}$, the so-called *Internal Driving Point (DP) Component* is a function of the cell state, being defined as

$$i_{g_{i,j}} \triangleq i_g\left(v_{x_{i,j}}\right) = \begin{cases} -\dfrac{v_{x_{i,j}}}{R_x} - a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat} & \text{if } v_{x_{i,j}} < -v_{sat}, \quad (8) \\ \left(a_{0,0} \cdot R_y \cdot g_{lin} - \dfrac{1}{R_x}\right) \cdot v_{x_{i,j}} & \text{if } \left|v_{x_{i,j}}\right| \leq v_{sat}, \quad (9) \\ -\dfrac{v_{x_{i,j}}}{R_x} + a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat} & \text{if } v_{x_{i,j}} > +v_{sat}. \quad (10) \end{cases}$$

while $i_{w_{i,j}}$, known as *offset current*, mostly accounts for the coupling effects, being expressed by

$$\begin{aligned} i_{w_{i,j}} = i_w\left(\left\{v_{u_{i+k,j+l}}\right\}, \left\{v_{y_{i+k,j+l}}\right\}\right) \triangleq z \cdot I + b_{0,0} \cdot v_{u_{i,j}} \\ + \sum_{\substack{k,l=-1 \\ (k,l)\neq(0,0)}} \left(a_{k,l} \cdot v_{y_{i+k,j+l}} + b_{k,l} \cdot v_{u_{i+k,j+l}}\right), \end{aligned} \tag{11}$$

where $\{v_{u_{i+k,j+l}}\}$ ($\{v_{y_{i+k,j+l}}\}$) denotes the set of input (output) voltages of all the processing elements in the 3×3 neighborhood of the cell $C(i,j)$. Nineteen are the key parameters, which define the image processing operation, which a CNN may accomplish, for a predefined input/initial condition combination, and under suitable boundary conditions, specifically the threshold z, the feedback synaptic weights $\{a_{-1,-1}, a_{-1,0}, a_{-1,+1}, a_{0,-1}, a_{0,0}, a_{0,+1}, a_{+1,-1}, a_{+1,0}, a_{+1,+1}\}$, and the+feedforward synaptic weights $\{b_{-1,-1}, b_{-1,0}, b_{-1,+1}, b_{0,-1}, b_{0,0}, b_{0,+1}, b_{+1,-1}, b_{+1,0}, b_{+1,+1}\}$. Given the crucial role that this 19-number set plays on the dynamical evolution of the cells' states, it is generally known as *gene* in CNN theory. A gene defines the set of rules, which apply concurrently in the neighborhood of each cell, allowing the standard space-invariant CNN to carry out a given computation.

Remark 1. A CNN may be used to carry out any computing task. The calculations are based upon the analogue dynamics of the cell states. As transients vanish, depending on cell parameter settings, inputs and initial conditions, each capacitor voltage may tend toward an equilibrium, or converge to an oscillatory waveform, which may be periodic, quasi-periodic, or even

[6]Since the CNN is space-invariant, from now onwards the following assumptions are made: $C_{x_{i,j}} = C_x$, $R_{x_{i,j}} = R_x$, and $R_{y_{i,j}} = R_y$.

chaotic. A CNN may then process the information, inserted as cell inputs and/or initial conditions, in two different forms i.e., producing computation results in the form of equilibria or waves, hence the names CNN equilibria-based computing or CNN wave computing attributed to the respective operating principle. In this manuscript the attention is focused on CNNs computing via equilibria[7].

Let us elucidate the classical method to design a CNN, so that it may execute a fundamental image processing task[8], carrying out the result of the computation in the cell equilibria.

A rigorous technique to synthesize the gene of a standard CNN, so as to allow the successful execution of a given equilibria-based computing task, even in the presence of deviations of some cell circuit parameters from their nominal values, was introduced in (Zarándy, 2003), and comprehensively presented in (Itoh and Chua, 2003). Before summarizing the line-of-thought at the basis of this classical methodology, let us provide a brief overview of the nonlinear dynamics of a CNN processing element.

2.2 Key Features of the CNN Cell Dynamics

The first aspect to consider for the synthesis of a suitable CNN gene is the choice of a proper value for the self-feedback synaptic weight $a_{0,0}$. As will be clarified through qualitative sketches below, this parameter crucially influences the directed *Internal DP Characteristic*, consisting of the arrowed locus of the rate of change of the state $\dot{v}_{x_{i,j}}$ vs. the state $v_{x_{i,j}}$ itself under $i_{w_{i,j}} = 0$ A. As anticipated in **section 2.2** in the context of memristors, this type of graphical representation is typically referred to as State Dynamic Route (SDR). In the upper (lower) half of the $v_{x_{i,j}}$–$\dot{v}_{x_{i,j}}$ plane, where $\dot{v}_{x_{i,j}} > (<) 0\,\mathrm{V} \cdot \mathrm{s}^{-1}$, arrows, supeimposed over the Internal DP Characteristic, point toward the east (west) to indicate an increase (a decrease) in the state $v_{x_{i,j}}$ over time. The intersections between this $\dot{v}_{x_{i,j}}$–$v_{x_{i,j}}$ locus and the horizontal $v_{x_{i,j}}$-axis, marked as filled (hollow) circles, denote the stable (unstable) equilibria of the cell state equation under null offset current. All in all, fixing the value for $a_{0,0}$ unequivocally determines the dynamical behaviour of the cell state $v_{x_{i,j}}$ from any initial condition $v_{x_{i,j},0}$ under zero offset current. With reference to[9] **Figure 3**, plot (a) shows how $a_{0,0}$ affects the shape of the locus of the state rate of change $\dot{v}_{x_{i,j}}$ vs. the state $v_{x_{i,j}}$ itself under $i_{w_{i,j}} = 0$ A. As anticipated in **section 2.2** in the context of memristors, a family of directed loci of this kind, one for each value of a control parameter (in this case $a_{0,0}$), is called DRM, here more specifically referred to as cell DRM. Varying $a_{0,0}$ from $-\infty$ to $+\infty$, the CNN processing element may exhibit three distinct stability characters:

- If $a_{0,0} < R_x^{-1} \cdot g_{lin}^{-1} \cdot R_y^{-1}$ (see the green and brown curves, associated to non-null negative and null self-feedback values, respectively) the cell is monostable with one and only one GAS equilibrium at $\overline{v}_{x_{i,j}} = 0$ V.
- If $a_{0,0} = R_x^{-1} \cdot g_{lin}^{-1} \cdot R_y^{-1}$ (see the pink curve) each state value within the set $[-1, 1]$V is a stable but not GAS equilibrium $\overline{v}_{x_{i,j}}$ for the processing element, which is said to admit a line of equilibria.
- If $a_{0,0} > R_x^{-1} \cdot g_{lin}^{-1} \cdot R_y^{-1}$ (see the black curve) the cell is bistable, featuring two locally stable equilibria, one lying at $\overline{v}_{x_{i,j}} = -a_{0,0}$ in the negative saturation region, and the other at $\overline{v}_{x_{i,j}} = a_{0,0}$ in the positive saturation region, besides the separatrix between their basins of attractions, namely the unstable equilibrium in the origin.

The ordinates of the two breakpoints of the three-segment[10] piecewise-linear Internal DP Characteristic, located at $(-v_{sat}, -(a_{0,0} \cdot R_y \cdot g_{lin} - R_x^{-1}) \cdot v_{sat})$, and at $(+v_{sat}, +(a_{0,0} \cdot R_y \cdot g_{lin} - R_x^{-1}) \cdot v_{sat})$, respectively, are of significant importance in the analysis of the *Shifted*[11] *DP Characteristic* (Chua and Roska, 2002), i.e. the locus of the state rate of change $\dot{v}_{x_{i,j}}$ vs. the state $v_{x_{i,j}}$ itself under non-null offset current. First of all, it is important to point out that, under specific hypotheses, including fixed values for all input voltages and thresholds, a standard space-invariant[12] CNN is *completely stable* (Chua and Roska, 2002), in the sense that the state $v_{x_{i,j}}$ of each cell $C(i, j)$ converges asymptotically toward an equilibrium point $\overline{v}_{x_{i,j}}$, which, in general, depends upon the initial conditon $v_{x_{i,j},0}$. Moreover, for a completely stable standard space-invariant CNN, according to the *bistability criterion* (Chua and Roska, 2002), in case the condition

$$a_{0,0} > R_x^{-1} \cdot g_{lin}^{-1} \cdot R_y^{-1} \tag{12}$$

holds true, the slope of the Internal DP Characteristic in the linear region–refer to **Eq. 9** – is positive, and, consequently, the stable equilibria of each cell under $i_{w_{i,j}} \neq 0$ A are found to lie in either of the two saturation regions of the standard nonlinearity of **Eq. 3**, as will be elucidated, shortly, through the analysis of the resulting Family of Shifted DP Characteristics, implying that, given the form of the standard nonlinearity **Eq. 3**, the final value for the output voltage of any processing element is one of the two saturation levels in the set $\{-v_{sat}, v_{sat}\}$. This is highly desirable for image processing, where, as will be shown through an illustrative example shortly, a CNN equilibria-based

[7]As reported in the template library (Karacs et al., 2018), a very large class of computing tasks may be executed on the basis of the stable equilibria, the CNN states dynamically evolve to, or, more precisely, of the respective cell output voltages.

[8]Images to be processed, consisting of $M \times N$ pixels, are encoded into the cells' input voltages and/or into the cells' initial states of a CNN, which features M rows and N columns. The output voltages, that the cells exhibit at equilibrium, and that constitute the result of a given computing task, are mapped onto an output image for visualization purposes. Regarding the correspondence between the color of the image pixel at row i and column j and the real value of the respective CNN cell input voltage $v_{u_{i,j}}$ or initial state $v_{x_{i,j},0} \triangleq v_{x_{i,j}}(0\,s)$ or output voltage $v_{y_{i,j}}$, the following convention is adopted in the EDGE CNN design. A black (white) image pixel is associated with a positive (negative) 1 V value, while a gray image pixel is associated with a real number, properly extracted from the range $(-1, +1)$ V, so as to reveal the intensity of the color tone. The lighter (darker) is the gray color, the closer to $-(+)1$ V would be the corresponding real value.

[9]The viewgraphs in **Figure 3** have been derived under the following parameter setting: $C_x = 1$ F, $R_x = 1\,\Omega$, $g_{lin} \cdot R_y = 1$, and $v_{sat} = 1$ V.

[10]In the case $a_{0,0} = 0\,\Omega^{-1}$ the Internal DP Characteristic features a single segment only, as illustrated graphically through the pink $i_{g_{i,j}}$–$v_{x_{i,j}}$ locus in **Figure 3A**.

[11]The effect of the offset current is to shift the Internal DP Characteristic, explaining the origin for the name of the $\dot{v}_{x_{i,j}}$–$v_{x_{i,j}}$ locus for $i_{w_{i,j}} \neq 0$ A.

[12]This theorem holds also for standard space-variant CNNs.

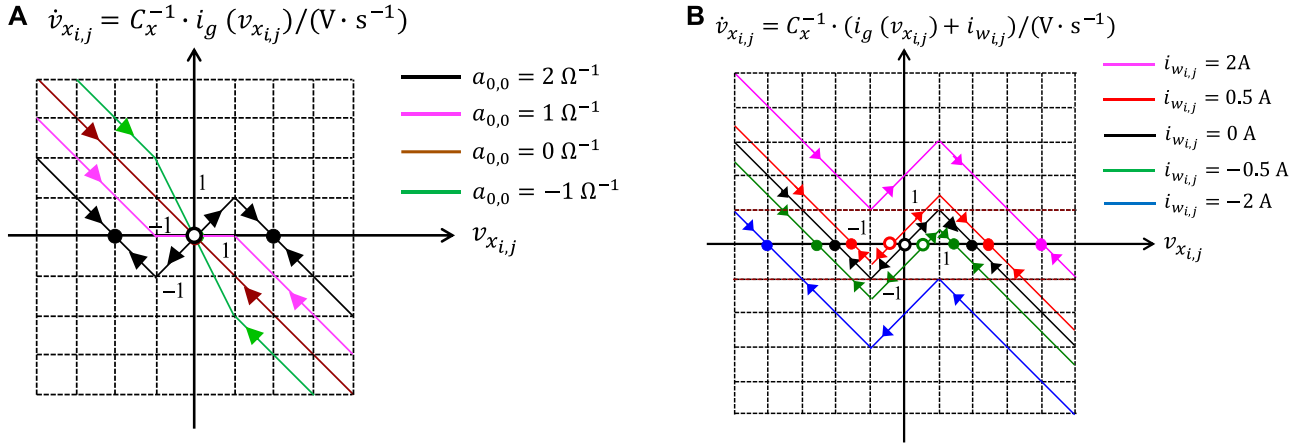


FIGURE 3 | (A) Family of Internal DP Characteristics, which a CNN cell admits for each value of the self-feedback synaptic weight $a_{0,0}$ in the set $\{-1, 0, 1, 2\}\ \Omega^{-1}$. **(B)** Family of Shifted DP Characteristics, which a CNN cell admits for $a_{0,0} = 2\ \Omega^{-1}$ for each value of the offset current $i_{w_{i,j}}$ in the set $\{-2, -0.5, 0, 0.5, 2\}$. The set of cell circuit parameter values for the derivation of these viewgraphs is provided here: $C_x = 1$ F, $R_x = 1\ \Omega$, $g_{lin} \cdot R_y = 1$, and $v_{sat} = 1$ V.

computation is typically visualized in the form of a binary output image, coding the final values of the cell output voltages. Importantly, the output voltage of each processing element will attain its final value i.e., one of the two possible saturation levels, already at a finite time instant, let us denote it as $t_{i,j}^{(s)}$, at which its state $v_{x_{i,j}}$ enters the saturation region, which hosts the equilibrium point $\overline{v}_{x_{i,j}}$. For all $t \geq t^{(s)} \triangleq \max_{1 \leq i \leq M, 1 \leq j \leq N} \{t_{i,j}^{(s)}\}$ the CNN may be considered at *steady state* as far as its cell output voltages are concerned[13]. It follows that, irrespective of the location of the CNN cell $C(i,j)$ under analysis, the offset current $i_{w_{i,j}}$, which, in general, changes over time during transients, keeps a fixed value for all $t \geq t^{(s)}$. This is of significant importance, since from this time instant onward, the Shifted DP Characteristic will no longer bounce, as, on the other hand, may be the case during transients, facilitating the study of the dynamic behaviour of the state $v_{x_{i,j}}$ from any initial condition $v_{x_{i,j},0}$.

With reference to the viewgraphs of **Figure 3B**, neglecting the marginal cases, three are the possible equilibria configurations, which a cell $C(i,j)$ may feature under the bistable criterion hypothesis for $i_{w_{i,j}} \neq 0$ A.

- If $i_{w_{i,j}} < -(a_{0,0} \cdot R_y \cdot g_{lin} - R_x^{-1}) \cdot v_{sat}$ (see the blue curve) the cell turns into a monostable dynamical system, featuring one and only one globally asymptotically stable (GAS) equilibrium in the negative saturation region at $\overline{v}_{x_{i,j}} = -R_x \cdot (a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat} - i_{w_{i,j}})$.
- If $-(a_{0,0} \cdot R_y \cdot g_{lin} - R_x^{-1}) \cdot v_{sat} < \left|i_{w_{i,j}}\right| < +(a_{0,0} \cdot R_y \cdot g_{lin} - R_x^{-1}) \cdot v_{sat}$ (see the green, and red curves, associated to negative and positive offset current values, respectively) the processing element keeps its bistable character, featuring an unstable equilibrium in the linear region at $\overline{v}_{x_{i,j}} = -i_{w_{i,j}} \cdot (a_{0,0} \cdot R_y \cdot g_{lin} - R_x^{-1})^{-1}$, and two-locally stable equilibria, one in the negative saturation region at $\overline{v}_{x_{i,j}} = -R_x \cdot (a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat} - i_{w_{i,j}})$, and one in the positive saturation region at $\overline{v}_{x_{i,j}} = +R_x \cdot (a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat} + i_{w_{i,j}})$.
- If $i_{w_{i,j}} > +(a_{0,0} \cdot R_y \cdot g_{lin} - R_x^{-1}) \cdot v_{sat}$ (see the magenta curve) the cell turns into yet another monostable dynamical system, admitting one and only one GAS equilibrium in the positive saturation region at $\overline{v}_{x_{i,j}} = +R_x \cdot (a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat} + i_{w_{i,j}})$.

2.3 A Systematic DRM-Based Methodology to Design Robust CNNs

The standard method (Zarándy, 2003; Itoh and Chua, 2003) to synthesize a suitable CNN gene for the execution of a given computing task is based upon the set up and later solution of an ad-hoc set of inequalities, expressed in terms of unknown gene parameters, ensuring a robust accomplishment of the computing task of interest. The functionalities of a given uncoupled[14] standard space-invariant completely stable CNN, satisfying the hypotheses of the bistability criterion,

[13]For $t > t^{(s)}$ the cell states continue their evolution toward the respective equilibria, but the output stage resistor voltages keep unchanged. This makes the CNN calculations insensitive to small variations in the nominal locations of the cell equilibria, and allows to read the result of a certain image processing operation at some finite time.

[14]A standard space-invariant CNN, in which the cell $C(i,j)$ features a sphere of influence of unitary radius and is described by the ODE (1), is said to be *uncoupled* (Chua and Roska, 2002) if each feedback synaptic weight $a_{k,l}$ – $k, l \in \{-1, 0, 1\}$ – except for $a_{0,0}$, is null. Further, if at least one feedforward synaptic weight $b_{k,l}$ – $k, l \in \{-1, 0, 1\}$ – is non-null, then the CNN is said to be *non-autonomous*. Each rule in a suitable set, which an uncoupled nonlinear dynamic array is obliged to obey, so as to execute a preliminarily specified data processing task, dictates the equilibrium $(\overline{x}_{m_{i,j}} = x_{m_{i,j}}(\infty), \overline{v}_{x_{i,j}} = v_{x_{i,j}}(\infty))$ of the cell $C(i,j)$, or the corresponding output voltage $v_{y_{i,j}}(\infty)$ on the basis of conditions involving input $v_{u_{i+k,j+l}}$ and/or initial condition $v_{x_{i+k,j+l}}(0)$ of each processing element $C(i+k, j+l)$ in the 3×3 neighborhood of the cell $C(i,j)$ itself, only. Under these circumstances, the rules are said to be *local*. In case the aforementioned CNN is coupled, on the other hand, some of the rules – referred to as global – for the cell $C(i,j)$ may also depend upon the input voltage $v_{u_{i+m,j+n}}$ and/or the initial condition $v_{x_{i+m,j+n}}(0)$ of a remote processing element ($m \notin \{-1, 0, 1\}$ and/or $n \notin \{-1, 0, 1\}$).

are dictated by a set of local rules, establishing the asymptotic value for the state $v_{x_{i,j}}(\infty) \equiv \overline{v}_{x_{i,j}}$, and, correspondingly, the steady-state output voltage $v_{y_{i,j}}(t_{i,j}^{(s)})$ of each cell $C(i,j)$, depending upon inputs, and initial conditions of all the processing elements within its 3×3 neighbourhood.

For the reasons motivated above, in CNN designs for image processing applications, it is common to set the numerical value for the self-feedback synaptic weight $a_{0,0}$ so as to guarantee the satisfaction of inequality **Eq. 12**. Typically, to facilitate the CNN design process, the structure of the gene under synthesis is simplified as much as possible, given the degree of complexity of the computing task, which the cellular array is expected to execute, and/or the values of some of the elements from the 19-number parameter set are assumed to be known. The family of all the possible invariable arrowed Shifted DP Characteristics, which a cell may admit for all $t > t^{(s)}$ for any value, which the offset current may assume, then, under the hypothesis of each of the rules, is then examined, so as to identify the worst-case scenario, where deviations of cell circuit parameters from their nominal values are most likely to induce a change in the cell equilibria configuration[15], and, consequently, the emergence of CNN computation errors. The next step is to write down an inequality, establishing a constraint for the offset current, and ensuring that, in such critical scenario, $\dot{v}_{x_{i,j}}$ is negative (positive) at the specified initial condition $v_{x_{i,j}}0$, so that the state $v_{x_{i,j}}$ would approach a desired equilibrium $\overline{v}_{x_{i,j}}$ in the negative (positive) saturation region. Repeating this procedure for each rule allows to derive an inequality set (IS), whose solutions may be determined through numerical techniques, or, in case the number of unknowns is small, via a geometry-based approach. The particular values assigned to the unknowns, allowing to program the CNN with a suitable gene, are then selected to endow the computation with the highest degree of tolerance against parameter variation.

2.4 Edge CNN

The aim of this section is to synthesize the gene of a standard space-invariant non-autonomous uncoupled CNN so that the resulting nonlinear dynamic array is able to extract the edges from an input binary image. Next, the classical CNN design methodology, briefly described earlier, will be applied to the cell ODE (1) in order to achieve this purpose. The three local rules, which each cell should obey, are reported in **Table 1**, where n_B defines how many of the 8 adjacent neighbors feature a positive one V-valued input voltage. The first rule establishes that, if the cell $C(i,j)$ features an input voltage $v_{u_{i,j}}$ equal to $-1\,V$, its output voltage $v_{y_{i,j}}(\infty)$ at equilibrium is found to attain the negative saturation voltage $-v_{sat}$ irrespective of the value of n_B. Plot (a) in **Figure 4** depicts a possible 3×3 pattern around a white pixel at row i and column j in the input binary image

TABLE 1 | Local rule triplet, which, irrespective of its location within the cellular array, a processing element $C(i,j)$ is requested to obey, for the extraction of edges from an input binary image, preliminarily discretized into a $M \times N$ matrix of pixels ($i \in \{1,\ldots,M\}$, $j \in \{1,\ldots,N\}$).

Local rule	$\mathbf{v_{u_{i,j}}/V}$	$\mathbf{v_{y_{i,j}}(\infty)}$	Conditions on $\mathbf{n_B}$
1	−1	$-v_{sat}$	Irrespective of n_B
2	+1	$-v_{sat}$	If $n_B = 8$
3	+1	v_{sat}	If $n_B \neq 8$

under rule 1. Here 3 of the 8 neighboring pixels are black. The pixel in the position (i,j) of the output binary image is white at equilibrium, as depicted on the bottom of the input pattern. The second (third) rule imposes that, in case the cell $C(i,j)$ features an input voltage $v_{u_{i,j}}$ equal to +1 V, its output voltage $v_{y_{i,j}}(\infty)$ at equilibrium is found to attain the negative (positive) saturation voltage $-v_{sat}$ in case n_B is exactly equal to 8 (is less or equal to 7). Plot (b) ((c)) in **Figure 4** depicts the only (a) possible 3×3 pattern around a black pixel at row i and column j in the input binary image under rule 2 (3). Here all (4) of the 8 neighboring pixels are black. The pixel in the position (i,j) of the output binary image is white (black) at equilibrium, as depicted on the bottom of the input pattern.

As clarified by **Figure 5A**, the CNN under design is expected to extract the edges from an input binary image, visualizing them in the output binary image at steady state. Since the CNN is meant to be uncoupled, the offset current from **Eq. 11** reduces to[16]

$$i_{w_{i,j}} = z \cdot I + b_{0,0} \cdot v_{u_{i,j}} + \sum_{\substack{k,l=-1 \\ (k,l)\neq(0,0)}}^{1} b_{k,l} \cdot v_{u_{i+k,j+l}}. \tag{13}$$

Assuming that all the feedforward synaptic weights, with the exclusion of $b_{0,0}$, are identical one to the other, namely $b_{k,l} = b$ for all $k,l \in \{-1,0,+1\}$ such that $(k,l) \neq (0,0)$, indicating how many, among the 8 neighbours of the cell $C(i,j)$, feature a negative one V-valued input voltage through the variable n_W, and noting that $n_B + n_W = 8$, the formula **Eq. 13** for $i_{w_{i,j}}$ reduces to

$$i_{w_{i,j}}\left(v_{u_{i,j}}, n_B\right) = z \cdot I + b_{0,0} \cdot v_{u_{i,j}} + b \cdot (2 \cdot n_B - 8)V, \tag{14}$$

where the argument reveals the dependence of the offset current upon $v_{u_{i,j}}$ and n_B. Assuming that $a_{0,0} \in \mathbb{R}_{>0}$ and $b \in \mathbb{R}_{<0}$ are given parameters, the only two unknowns for the specification of

[15]In the theory of nonlinear dynamics, a quantitative change in the behaviour of a system, occurring during the sweep of a control parameter, is referred to as a *bifurcation* phenomenon (Strogatz, 2000).

[16]It is instructive to observe that a very large number of fundamental image processing operations are possible adopting the class of standard space-invariant uncoupled CNNs, as may be inferred by inspecting the template library (Karacs et al., 2018). For each CNN, belonging to this class and processing still images, the cell offset current from **Eq. 13** is always a constant, and the respective Shifted DP Characteristic is invariant over time. As anticipated earlier, with reference to a coupled CNN, which process still images, a cell, which satisfies the bistability condition **Eq. 12**, typically features a Shifted DP Characteristic, which continually moves vertically up or down until the time instant $t_{i,j}^{(s)}$, at which the state of each cell has entered the particular saturation region hosting the equilibrium it asymptotically converges to, keeping unchanged thereafter.

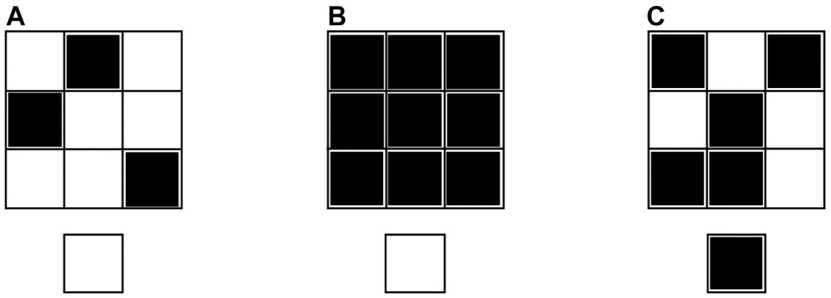


FIGURE 4 | Graphical illustration of the application of the EDGE CNN local rules 1 for $n_B = 3$ **(A)**, 2 **(B)**, and 3 for $n_B = 4$ **(C)**. Each of the three plots visualizes, on top, a 3 × 3 pattern around the pixel located at row *i* and column *j* in the input binary image, and, below, the pixel at position (i,j) in the output binary image at equilibrium.

a suitable gene are then $b_{0,0}$ and z. **Figure 5B** shows the directed Internal DP Characteristic[17]. The state **Eq. 1** under $i_{w_{i,j}} = 0$ A admits two locally stable equilibria, located one in the negative saturation region, namely $\overline{v}_{x_{i,j}} = -a_{0,0} \cdot R_y \cdot g_{lin} \cdot R_x \cdot v_{sat}$, and one in the positive saturation region, specifically $\overline{v}_{x_{i,j}} = a_{0,0} \cdot R_y \cdot g_{lin} \cdot R_x \cdot v_{sat}$, and separated by an unstable equilibrium, i.e. $\overline{v}_{x_{i,j}} = 0$ V, positioned in the linear region.

Let us set the initial condition $v_{x_{i,j}}(0)$ of the cell ODE (1) to +1 V. Following the line-of-thought inspiring the classical CNN design methodologies, discussed in the seminal papers (Zarándy, 2003) and (Itoh and Chua, 2003), and briefly reviewed above, let us now examine the Family of arrowed Shifted DP Characteristics, which a processing element may admit in all scenarios, which may possibly emerge under the hypothesis of each of the three rules from **Table 1**.

In order to fulfill rule 1, where $v_{u_{i,j}} = -1$ V, a condition should be enforced to ensure that the maximum value, which the offset current may ever attain i.e., $\max_{0 \le n_B \le 8} \{i_{w_{i,j}}(-1\,\mathrm{V}, n_B)\} = i_{w_{i,j}}(-1\,\mathrm{V}, 0)$, is smaller than the ordinate $-(a_{0,0} \cdot R_y \cdot g_{lin} - G_x) \cdot v_{sat}$ of the left breakpoint of the $\dot{v}_{x_{i,j}}$–$v_{x_{i,j}}$ piecewise linear characteristic of **Figure 5B**. This would guarantee a negative sign for the ordinate of the right breakpoint of the resulting $\dot{v}_{x_{i,j}}$–$v_{x_{i,j}}$ piecewise-linear characteristic, as is the case for the arrowed blue locus in **Figure 6A**, illustrating the dynamic route followed by the cell state for $i_{w_{i,j}}(v_{u_{i,j}}, n_B) = -2 \cdot (a_{0,0} \cdot R_y \cdot g_{lin} - G_x) \cdot v_{sat}$, where $v_{u_{i,j}} = -1$ V, and $n_B = 0$, under the parameter setting, reported in the caption of **Figure 5B**. As a result, for all possible n_B values in $\{0, 1, 2, 3, 4, 5, 6, 7, 8\}$, the CNN cell would be monostable, and $v_{x_{i,j}}$ would decrease monotonically over time from the initial condition $v_{x_{i,j}}(0)$ toward an equilibrium, i.e. $\overline{v}_{x_{i,j}} = (-a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat} + i_{w_{i,j}}(-1\,\mathrm{V}, n_B)) \cdot R_x$, located in the negative saturation region, as established by rule 1. Therefore, the first EDGE CNN design constraint sets an upper bound for the maximum offset current, according to

$$i_{w_{i,j}}(-1\,\mathrm{V}, 0) < -\left(a_{0,0} \cdot R_y \cdot g_{lin} - G_x\right) \cdot v_{sat}. \tag{15}$$

It is worth pointing out that the farther away from the horizontal axis, within the plane lower half, would be positioned the right breakpoint of the $\dot{v}_{x_{i,j}}$–$v_{x_{i,j}}$ piecewise-linear characteristic in the worst-case scenario from rule 1, and the more robust would be the EDGE CNN design[18].

Let us now derive the condition allowing the CNN to apply rule 2 from **Table 1** in the sphere of influence of any processing cell $C(i,j)$, which features, as each of its eight neigbours, a positive one V-valued input voltage. Since the expected cell steady-state output voltage $v_{y_{i,j}}(t_{i,j}^{(s)})$ is once again −1 V, as in rule 1, **Figure 6A** can be reused to work out a suitable inequality for rule 2 under $v_{u_{i,j}} = +1$ V and $n_B = 8$. The second EDGE CNN design condition, ensuring that the state $v_{x_{i,j}}$ of a cell $C(i,j)$ with $v_{u_{i,j}} = +1$ V and $n_B = 8$ would asymptotically approach an equilibrium, specifically $\overline{v}_{x_{i,j}} = (-a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat} + i_{w_{i,j}}(+1\,\mathrm{V}, 8)) \cdot R_x$, located in the negative saturation region, is then similar to the inequality **Eq. 5**, reading as

$$i_{w_{i,j}}(+1\,\mathrm{V}, 8) < -\left(a_{0,0} \cdot R_y \cdot g_{lin} - G_x\right) \cdot v_{sat}. \tag{16}$$

In order for the CNN under design to apply rule 3 from **Table 1** in the 3 × 3 neighbourhood of each processing element, which features a positive one V-valued input voltage, and is physically coupled to at least one neighbour with a negative one V-valued input voltage, the minimum value, which the offset current may ever attain, namely $\min_{1 \le n_B \le 7} \{i_{w_{i,j}}(+1\,\mathrm{V}, n_B)\} = i_{w_{i,j}}(+1\,\mathrm{V}, 7)$ should be larger than the ordinate $-(a_{0,0} \cdot R_y \cdot g_{lin} - G_x) \cdot v_{sat}$ of the left breakpoint of the $\dot{v}_{x_{i,j}}$–$v_{x_{i,j}}$ piecewise-linear characteristic of **Figure 5B**. This would ensure a positive sign for the ordinate of the right breakpoint of the resulting $\dot{v}_{x_{i,j}}$ vs. $v_{x_{i,j}}$ piece-wize linear characteristic, as is the case for the arrowed blue locus in

[17]The values of some of the parameters in the cell circuit of **Figure 2** are fixed as reported in the caption of **Figure 5B**. Since the bistability condition **Eq. 12** holds true, the slope of the piecewise-linear locus in the linear region–refer to **Eq. 9** is strictly positive.

[18]In case $i_{w_{i,j}}(-1\,\mathrm{V}, 0)$ were found to be equal to $-(a_{0,0} \cdot R_y \cdot g_{lin} - G_x) \cdot v_{sat}$, the right breakpoint of the resulting $\dot{v}_{x_{i,j}}$ versus $v_{x_{i,j}}$ locus, depicted in red in the example of **Figure 6A**, would lie on the horizontal axis. Under this hypothesis, in the worst-case scenario from rule 1, the state $v_{x_{i,j}}$ would keep equal to the initial condition $v_{x_{i,j}}(0) = +1$ V at all times, and the cell $C(i,j)$ would fail to operate as requested. As a result, in the worst-case scenario from rule 1, the right breakpoint of the Shifted DP Characteristic should lie within the plane lower half at some safety distance from the horizontal axis. Note that a half-filled black circle denotes a *semistable equilibrium*, which attracts only trajectories, which are initiated from one of its two sides.

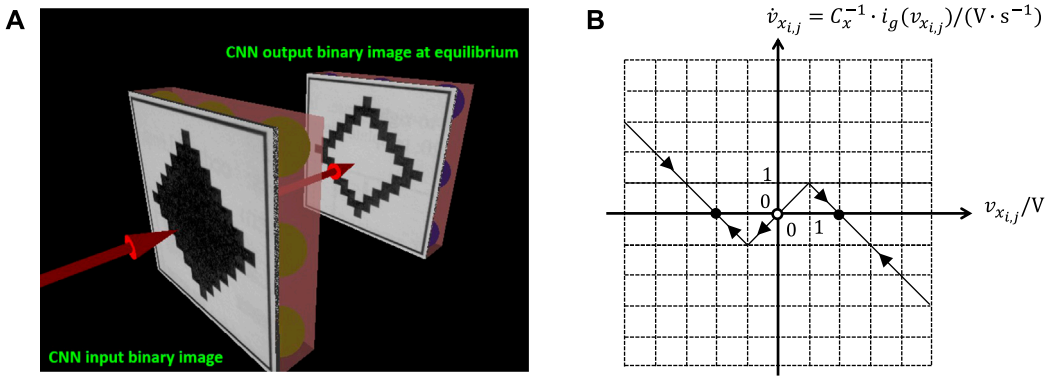


FIGURE 5 | (A) Graphical illustration of the operating principles of the CNN under design. **(B)** EDGE CNN SDR for zero offset current. Here $C_x = 1\,\mathrm{F}$, $R_x = 1\,\Omega$, $R_y \cdot g_{lin} = 1$, and $v_{sat} = 1\,\mathrm{V}$. The self-feedback synaptic weight $a_{0,0}$ (the b value for each of the feedforward synaptic weights, except for $b_{0,0}$) is set to $2\,\Omega^{-1}$ ($-1\,\Omega^{-1}$) ahead of the application of the classical CNN design methodology from Itoh and Chua (2003). The cell equilibria lie at $\overline{v}_{x_{i,j}} = -2\,\mathrm{V}$, at $\overline{v}_{x_{i,j}} = 0\,\mathrm{V}$, and at $\overline{v}_{x_{i,j}} = 2\,\mathrm{V}$.

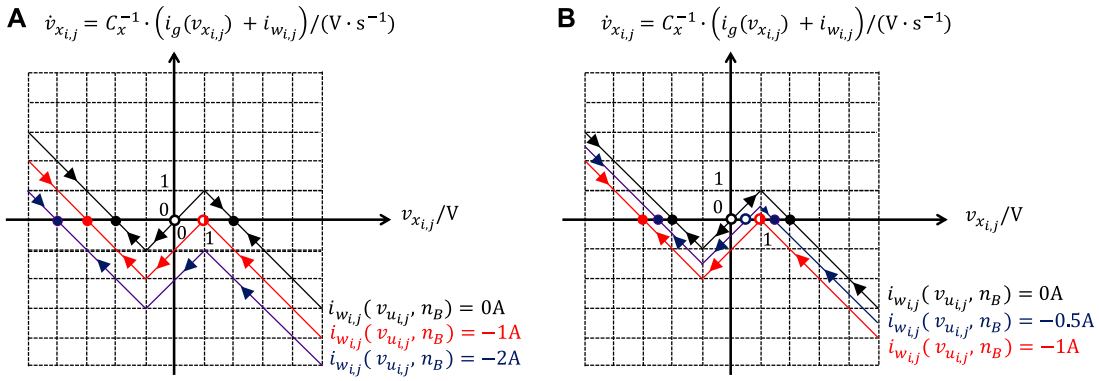


FIGURE 6 | Graphs clarifying the line of reasoning behind the DRM synthesis strategy adopted in Itoh and Chua (2003) to select a suitable gene allowing the resulting CNN to apply the local rule triplet of the binary image edge extraction operation in the 9-cell neighborhood of each processing element. The worst-case scenario in rule 1 is analyzed in **(A)**, where $v_{u_{i,j}} = -1\,\mathrm{V}$ and $n_B = 0$. Setting $v_{u_{i,j}} = +1\,\mathrm{V}$ and $n_B = 8$, plot **(A)** allows to investigate rule 2 as well. The worst-case scenario in rule 3 is illustrated in plot **(B)**, where $v_{u_{i,j}} = +1\,\mathrm{V}$ and $n_B = 7$. The setting of the known parameters of the cell circuit of **Figure 2** is reported in the caption of **Figure 5B**. With reference to plot **(A)**, in the worst-case scenario from rule 1 (in rule 2) the cell state $v_{x_{i,j}}$ would evolve from the initial condition $v_{x_{i,j}}(0) = +1\,\mathrm{V}$ toward the equilibrium $\overline{v}_{x_{i,j}} = -4\,\mathrm{V}$, as dictated by the arrowed blue locus, in case $i_{w_{i,j}}(-(+)1\,\mathrm{V}, 0\,(8))$ were found to be equal to $-2\,\mathrm{A}$, while it would keep its initial value $v_{x_{i,j}}(0) = +1\,\mathrm{V}$ at all times, as governed by the arrowed red locus, if, as a result of the CNN design, the value $-1\mathrm{A}$ would be assigned to $i_{w_{i,j}}(-(+)1\,\mathrm{V}, 0\,(8))$. In case a cell would feature the blue (red) SDR, either in the worst-case scenario from rule 1 or in rule 2, the CNN would operate (would fail to function) as required. With reference to plot **(B)**, in the worst-case scenario from rule 3, $v_{x_{i,j}}$ would evolve along the arrowed blue dynamic route from the initial condition toward the equilibrium $\overline{v}_{x_{i,j}} = 1.5\,\mathrm{V}$ provided $i_{w_{i,j}}(+1\,\mathrm{V}, 7)$ were found to be equal to $-0.5\mathrm{A}$, while it would keep its initial value $v_{x_{i,j}}(0) = +1\,\mathrm{V}$, as established by the arrowed red locus, if, as a result of the CNN design, the value $-1\mathrm{A}$ would be assigned to $i_{w_{i,j}}(+1\,\mathrm{V}, 7)$. Theoretically a CNN would properly function if a cell would exhibit the red SDR in the worst-case scenario from rule 3. However, if the cell featured the blue SDR, instead, it would additionally exhibit a little tolerance to deviations of parameters from their nominal values. The directed Internal DP Characteristic, shown in **Figure 5B**, is depicted once again in black in both plots as a reference. This SDR would induce the cell state $v_{x_{i,j}}$ to converge toward the equilibrium $\overline{v}_{x_{i,j}} = 2\mathrm{V}$. It follows that a cell with such a SDR under $v_{u_{i,j}} = -1\,\mathrm{V}$ and $n_B = 0$ or under $v_{u_{i,j}} = +1\,\mathrm{V}$ and $n_B = 8$ (under $v_{u_{i,j}} = +1\,\mathrm{V}$ and $n_B = 7$) would seriously fail to operate as desired (would function properly, exhibiting a good robustness against parameter variability).

Figure 6B, illustrating the dynamic route of the cell state for $i_{w_{i,j}}(v_{u_{i,j}}, n_B) = -0.5 \cdot (a_{0,0} \cdot R_y \cdot g_{lin} - G_x) \cdot v_{sat}$, where $v_{u_{i,j}} = +1\,\mathrm{V}$, and $n_B = 7$, under the parameter setting, reported in the caption of **Figure 5B**. Consequently, for all admissible n_B values in $\{0, 1, 2, 3, 4, 5, 6, 7\}$, the cell state $v_{x_{i,j}}$ would monotonically increase over time toward an equilibrium, i.e. $\overline{v}_{x_{i,j}} = (a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat} + i_{w_{i,j}}(v_{u_{i,j}}, n_B)) \cdot R_x$, located in the positive saturation region, as required in rule 3. The third EDGE CNN design inequality is then establishing a lower bound for the minimum offset current, i.e.[19]

$$i_{w_{i,j}}(+1\,\mathrm{V}, 7) > -\left(a_{0,0} \cdot R_y \cdot g_{lin} - G_x\right) \cdot v_{sat}. \tag{17}$$

[19]In the worst-case scenario from rule 3 the cell is found to be bistable provided $\left|i_{w_{i,j}}(+1\,\mathrm{V}, 7)\right| < (a_{0,0} \cdot R_y \cdot g_{lin} - G_x) \cdot v_{sat}$, and monostable otherwise.

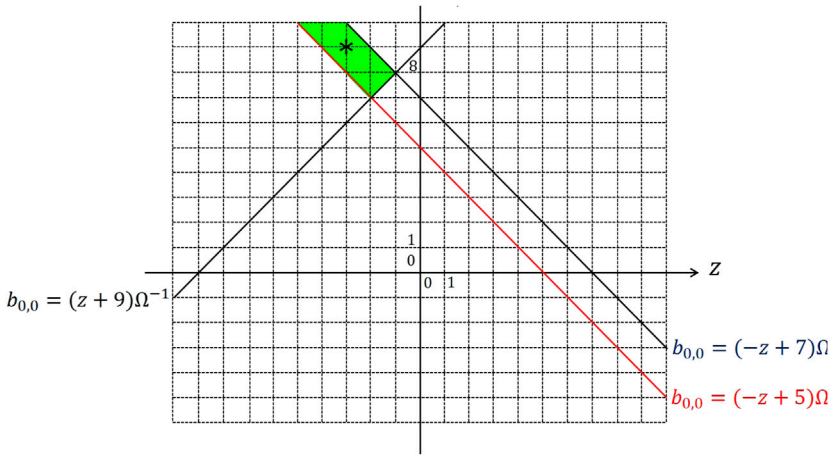


FIGURE 7 | Illustration of the geometrical analysis adopted to solve the three inequalities **Eqs. 15–17**, derived through the classical CNN design method (Itoh and Chua, 2003) to synthesize a suitable gene for a cellular array, intended to extract edges from a given input binary image, and reducing to $b_{0,0} > (z+9)\,\Omega^{-1}$, $b_{0,0} < (-z+7)\,\Omega^{-1}$, and $b_{0,0} > (-z+5)\,\Omega^{-1}$, respectively, under the parameter setting reported in the caption of **Figure 5B**. The set of admissible solutions are enclosed within the green area. The asterisk symbol, located at $(z^*, b^*_{0,0}) = (-3, 9\,\Omega^{-1})$, indicates a reasoned parameter pair choice for the specification of a robust EDGE CNN gene.

For a robust CNN design the right breakpoint of the $\dot{v}_{x_{i,j}}$–$v_{x_{i,j}}$ piecewise-linear characteristic of a cell $C(i,j)$ in the worst-case scenario from rule 3 should be positioned as farther away as possible from the horizontal axis within the plane upper half[20].

For the parameter setting reported in the caption of **Figure 5**, the three inequalities **Eqs. 15–17** are solved through a geometric approach on the z–$b_{0,0}$ parameter plane, as shown in **Figure 7**, where the green region visualizes the set of admissible solutions. For the specification of a suitable gene, guaranteeing the expected EDGE CNN functionality even in the presence of some small deviation of either of the two parameters z and $b_{0,0}$ from their nominal values, it is adviceable to choose a particular solution $(z^*, b^*_{0,0})$, whose graphical point-based representation on the parameter plane features an adequate distance from the boundaries of the green region, as indicated by means of an asterisk marker in **Figure 7**. The gene, synthesized in this section, allows the CNN to extract edges from an input binary image, as displayed in plot (a) of **Figure 5**.

2.5 Limitations of the CNN Paradigm and of Its Hardware Implementation

Since each of their processing elements interacts simultaneously with the respective neighbors, CNNs may process multi-variate signals in a massively parallel fashion, as crucially necessary in time-critical application fields, such as industry process control, electronic surveillance, medical augmented reality, and IoT smart sensing. In order to harness more efficiently the bio-inspired operating principles of these nonlinear dynamic arrays, which make them a suitable mathematical framework for modeling neural systems, Chua and Roska proposed an innovative computer, called CNN Universal Machine (CNN-UM) (Roska, 1993), to implement their signal processing paradigm. The CNN-UM, fabricated in various forms over the years through the well-established CMOS technology[21] (Vázquez et al., 2018), consists of an array of locally coupled computing units, each of which is endowed with data storage blocks, which allow to distribute the memory across the cellular array, endowing the computing machine with a truly non-von Neumann architecture, and to reconfigure the array so as to solve any computation problem. Thanks to their massively parallel computing power, CNN-UM hardware realizations (Vázquez et al., 2018) may process images at rates as high as 30,000 frames per second. Considering that, furthermore, a universal cellular array may be physically realized within the IC area of a single chip (Vázquez et al., 2018), CNNs are particularly suitable for the development of miniaturized IoT technical systems, in which the integration between a matrix of sensing elements and a network of locally coupled computing units with local stored programmability on board enables information processing at the same location, where data detection takes place. A major problem, which prevents to widen the applicability scope of this class of sensor-processor arrays, is the limited degree of complexity of the dynamical phenomena, which may possibly emerge within their physical media, due to the simplicity of the input-output behaviours of the electrical components employed

[20]In case $i_{w_{i,j}}(+1\,\mathrm{V}, 7)$ were found to be equal to $-(a_{0,0} \cdot R_y \cdot g_{lin} - G_x) \cdot v_{sat}$, as shown in red in the example of 6(b), the cell state would keep its initial value at all times. Here, at least theoretically, rule 3 would hold true. However, in the presence of any infinitesimally small negative-signed additive constant perturbation of the offset current, the cell would become monostable with a globally asymptotically stable equilibrium, specifically $\overline{v}_{x_{i,j}} = (-a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat} + i_{w_{i,j}}(+1\,\mathrm{V}, 7)) \cdot R_x$, in the negative saturation region, and the CNN would fail to impose rule 3 in the neighborhood of each cell with $v_{u_{i,j}} = +1\,\mathrm{V}$ and $n_B = 7$. Thus, in the worst-case scenario from rule 3, the right breakpoint of the Shifted DP Characteristic should lie at some safety distance from the horizontal axis within the plane upper half.

[21]Typically, the Full-Range model (Vázquez et al., 1993) is used in place for the Chua-Yang mathematical description of **Eq. (1)** to limit the range of admissible values for the cells' states, thus simplifying the hardware realisation of the CNN paradigm.

in the CNN-UM constitutive blocks. Thanks to their extremely rich dynamics, memristors may be adopted in novel designs of cellular computing arrays so as to extend significantly the spectrum of asymptotic spatio-temporal behaviours, which purely CMOS CNNs may currently exhibit. Another critical issue, which affects the performance of technical systems, combining sensing and processing functionalities on the same physical platform, is due to the rather low spatial resolution of state-of-the-art CNN-UM hardware realizations, originating from the presence of spacious data storage units within their computing units, as discussed earlier. This limits the maximum number of sensing and processing elements, which may be paired[22] one-to-one within the available IC area of these IoT commercial products (Toshiba Ltd., 2012). The adoption of memristive devices, endowed with memprocessing capabilities, may allow to obviate the inclusion of additional memory banks within the IC design of each CNN-UM computing unit, allowing to shrink considerably its size, and enabling the future realization of sensor-processor arrays with unprecedented spatial resolution, of great appeal to the IoT industry, nowadays. In this respect, it is timely to commence investigations aimed to explore the functionalities of Memristor CNNs (M-CNNs). In general, introducing memristors in the circuit implementation of a CNN processing element[23] increases the order of its ODE model, calling for the development of a new theory to investigate the operating principles of the resulting nonlinear dynamic array, and to program its gene to allow the accomplishment of a pre-defined memcomputing task. The theoretical foundations of M-CNNs shall be discussed in the section to follow.

3 THEORY OF MEMRISTOR CELLULAR NONLINEAR NETWORKS

Memristors are the key technology enabler for the hardware implementation of innovative memcomputing paradigms. This section provides some evidence for this claim, establishing the theoretical foundations of a class of cellular memprocessing structures, which we call M-CNNs, as anticipated in **section 2.5**. In order to realize one of the proposed M-CNNs a first-order non-volatile memristor[24] $\mathcal{M}_{x_{i,j}}$ is placed in parallel with the capacitor in the circuit implementation of each cell of the two-dimensional standard time- and space-invariant CNN (Chua, 1998), which was discussed in **section 2.1**. The memristive cell of the novel nonlinear dynamic array is shown in **Figure 8**.

The next section reports the mathematical description of the proposed memristive cellular array.

3.1 M-CNN Model

The M-CNN cell $C(i,j)$ of **Figure 8** may be described by the following pair of first-order coupled ODEs[25] ($i \in \{1,\ldots,M\}$, $j \in \{1,\ldots,N\}$):

$$\frac{dx_{m_{i,j}}}{dt} = g\left(x_{m_{i,j}}, v_{x_{i,j}}\right), \text{ and} \tag{18}$$

$$\frac{dv_{x_{i,j}}}{dt} = \frac{\tilde{i}_{g_{i,j}} + i_{w_{i,j}}}{C_x}. \tag{19}$$

The first ODE **Eq. 18** governs the time evolution of the state $x_{m_{i,j}}$ of the first-order nonvolatile resistance switching memory $\mathcal{M}_x$, which the M-CNN cell $C(i,j)$ acccommodates, according to an enhanced variant of a voltage-controlled memristor model, originally formulated by Pershin and Di Ventra (Pershin et al., 2009), and capable to capture the switching kinetics of real memristor devices (Jo et al., 2009), as discussed in Pershin and Di Ventra (2011). The model of the resistance switching memory in the cell $C(i,j)$ is a first-order element from the class of *generic memristors*, defined by the DAE set

$$\frac{dx_{m_{i,j}}}{dt} = g\left(x_{m_{i,j}}, v_{m_{i,j}}\right), \tag{20}$$

$$i_{m_{i,j}} = G\left(x_{m_{i,j}}\right) \cdot v_{m_{i,j}}. \tag{21}$$

Note that, within the processing element $C(i,j)$, the memristor voltage $v_{m_{i,j}}$ coincides with the capacitor voltage $v_{x_{i,j}}$, thus the expression for the memristor current $i_{m_{i,j}}$ in **Eq. 21** reduces to

$$i_{m_{i,j}} = G\left(x_{m_{i,j}}\right) \cdot v_{x_{i,j}}. \tag{22}$$

The state evolution function $g(x_{m_{i,j}}, v_{m_{i,j}})$ and the memductance function $G(x_{m_{i,j}})$ in the Pershin and Di Ventra model of the memristor in the M-CNN cell $C(i,j)$ are respectively expressed by

$$g\left(x_{m_{i,j}}, v_{m_{i,j}}\right) = \kappa\left(v_{x_{i,j}}\right) \cdot \left(\text{step}\left(v_{m_{i,j}}\right) \cdot f_+^{(p)}\left(x_{m_{i,j}}\right), +\text{step}\left(-v_{m_{i,j}}\right) \cdot f_-^{(p)}\left(x_{m_{i,j}}\right)\right), \text{ and} \tag{23}$$

$$G\left(x_{m_{i,j}}\right) = \frac{1}{x_{m_{i,j}}}. \tag{24}$$

The memristor state $x_{m_{i,j}}$, representing the device memristance, is constrained to lie at all times within the closed set $\mathcal{D} \triangleq [x_{on}, x_{off}]$, where x_{on} and x_{off} denote the lowest and highest possible device

[22]In state-of-the-art sensor-processor arrays the input to the processing element, lying in correspondence to the i^{th} row and j^{th} column of a CNN-UM hardware realization, is derived from the output of a sensing unit, located in the same position within a matrix of data detectors with same cell number count as the analog-and-logic computer ($i \in \{1,\ldots,M\}$, $j \in \{1,\ldots,N\}$). Since, for a given IC area, a sensing matrix may feature a much higher density as compared to a cellular computing machine, the low cell number count, which purely CMOS sensor processor arrays typically feature (Vázquez et al., 2018), could be significantly increased by leveraging the memcomputing capability of memristors to execute the memory functions, currently accomplished by additional data storage elements, within the CNN-UM computing units.

[23]In the class of M-CNNs, investigated in this thesis, memristors are employed only for the design of the cell circuit. Their use for the circuit implementation of the synaptic couplings shall be the focus of future studies.

[24]Throughout this chapter the state, voltage, and current of the memristor $\mathcal{M}_{x_{i,j}}$ in the cell $C(i,j)$ are denoted as $x_{m_{i,j}}$, $v_{m_{i,j}}$, and $i_{m_{i,j}}$, respectively.

[25]The numerical integration of the $2 \cdot (M \times N)$ ODEs, dictating the time evolution of the state vectors of all the memprocessing elements, calls for the preliminary assignment of an *initial condition* $\{(x_{m_{i,j}}(0), v_{x_{i,j}}(0))\}$ to each cell $C(i,j)$, and for the preparatory specification of the boundary conditions (Chua and Roska, 2002), fixing the input voltage and the output voltage of each virtual cell.

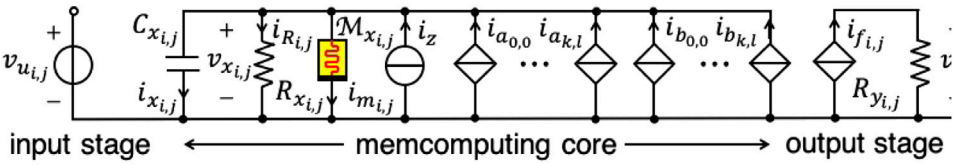


FIGURE 8 | Circuit implementation of the M-CNN cell $C(i,j)$ ($i \in \{1,\ldots,M\}, j \in \{1,\ldots,N\}$). In this study the cell circuit parameters are assumed to be invariant across the $M \times N$ bio-inspired memristive array. As a result, the following assumptions are made: $C_{x_{i,j}} = C_x$, $\mathcal{M}_{x_{i,j}} = \mathcal{M}_x$, $R_{x_{i,j}} = R_x$, and $R_{x_{i,j}} = R_y$. Two are the main contributions to the capacitor current $i_{x_{i,j}}$: one, given by the addition between $i_{a_{0,0}}$, $i_{R_{i,j}}$, and $i_{m_{i,j}}$, is a function of the two cell states, while the other, expressed by the sum of the memcomputing core currents, which flow through the 18 branches appearing to the right of the memristor, except for the self-feedback synaptic current, capture mostly the impact of input and output voltages of the 8 neighbors on the dynamics of the cell states themselves.

resistances, respectively. With reference to **Eq. 23**, step$(\cdot)$ stands for the Heaviside function, while $\kappa(v_{m_{i,j}})$ is a piecewise-linear nonlinearity of the form

$$\kappa\left(v_{m_{i,j}}\right) = -\beta \cdot v_{m_{i,j}} + \frac{\beta - \alpha}{2} \cdot \left(\left|v_{m_{i,j}} + V_t\right| - \left|v_{m_{i,j}} - V_t\right|\right), \tag{25}$$

where $\alpha \in \mathbb{R}_{>0}$ and $\beta \in \mathbb{R}_{>0}$ are coefficients, measured in units $\Omega \cdot V^{-1} \cdot s^{-1}$, denoting the smaller and larger slopes of the characteristic for $\left|v_{m_{i,j}}\right| \leq V_t$ and $\left|v_{m_{i,j}}\right| > V_t$, respectively, where $V_t \in \mathbb{R}_{>0}$ represents the memristor switching threshold voltage. **Figure 9A** depicts the $k(v_{m_{i,j}})$–$v_{m_{i,j}}$ chapacteristic for the parameter setting reported in its caption.

Since the memristor state existence domain $\mathcal{D}$ is finite, the state evolution function in **Eq. 23** is endowed with boundary conditions, which ensure that $x_{m_{i,j}}$ never decreases below (increases above) its lowest (largest) possible value under $v_{m_{i,j}} > (<) 0\,\text{V}$. In order to facilitate the numerical simulation of the memristor DAE set, we reformulate the boundary conditions as compared to their original definition[26] in the Pershin and Di Ventra model (Pershin et al., 2009), adopting continuous and differentiable functions, inspired to Biolek's window (Biolek et al., 2009), and reading as

$$f_{+}^{(p)}\left(x_{m_{i,j}}\right) = 1 - \left(\frac{x_{m_{i,j}} - x_{on}}{x_{off} - x_{on}} - 1\right)^{2 \cdot p}, \text{ and} \tag{26}$$

$$f_{-}^{(p)}\left(x_{m_{i,j}}\right) = 1 - \left(\frac{x_{m_{i,j}} - x_{on}}{x_{off} - x_{on}}\right)^{2 \cdot p}, \tag{27}$$

where $p \in \mathbb{N}_{>0}$ controls the decay rate of the window function **Eqs. 26**, **27** as $x_{m_{i,j}}$ approaches x_{on} (x_{off}). As graphically illustrated in plot (b) ((c)) of **Figure 9** for the parameter configuration provided in its caption, the window function $f_{+\,(-)}^{(p)}(x_{m_{i,j}})$ in **Eqs. 26**, **27** enforces the memory state evolution function **Eq. 23** to feature a zero, and, consequently, the memristor ODE **Eq. 18** to admit an equilibrium at $\overline{x}_{m_{i,j}} = x_{on\,(off)}$ under positive (negative) values of the capacitor voltage. Since the memory state ODE **Eq. 18**, with evolution function expressed by **Eq. 23**, is of first-order, the classical DRM graphical tool (Chua, 2018a) may be applied to investigate the memristor nonlinear dynamics. The DRM of the modified Pershin and Di Ventra memristor model is illustrated in **Figure 10A** for the parameter arrangement defined in its caption. The DC value $V_{m_{i,j}}$, assigned to the voltage falling across the resistance switching memory, parametrizes the family of memristor SDRs. Within the family of $\dot{x}_{m_{i,j}}$ vs. $x_{m_{i,j}}$ loci, the characteristic obtained for $V_{m_{i,j}} = 0\,\text{V}$, known as POP, provides hints on the nonvolatile memory capability of the circuit element. On the basis of the memristor model under focus, the POP is a segment of the $x_{m_{i,j}}$ axis lying between x_{on} and x_{off}. Each of the points on this segment–shown in black in **Figure 10A**–represents a stable but not asymptotically stable equilibrium (Strogatz, 2000) for the ODE **Eq. 18** with state evolution function **Eq. 23**. Particularly, the existence of a continuum of equilibria, namely $\overline{x}_{m_{i,j}} \in \mathcal{D}$, for the memristor state equation under zero input clearly reveals that the resistance switching device is an analogue non-volatile memory. Any value for $x_{m_{i,j}}$ within its existence domain $\mathcal{D}$ is a possible state, which the memristor may store, from the time at which the power is turned off, till the time at which a new voltage stimulus is applied between its terminals. With regard to the $\dot{x}_{m_{i,j}}$–$x_{m_{i,j}}$ loci, associated to nonzero values for $V_{m_{i,j}}$, in **Figure 10A**, the device asymptotically approaches the fully off (fully on) resistive equilibrium state $\overline{x}_{m_{i,j}} = x_{off\,(on)}$ in case any negative (positive) DC voltage is applied continually between its terminals, as indicated by the arrow superimposed on each blue (red) characteristic, which dictates a memory state rate of change increasing monotonically with $\left|V_{m_{i,j}}\right|$. Irrespective of the negative (positive) DC value assigned to the memristor voltage, the upper (lower) bound in the memristor state existence domain $\mathcal{D}$ is found to be a globally asymptotically stable equilibrium for the ODE (18) with state evolution function **Eq. 23**. For the very same parameter setting, **Figure 10B** demonstrates now the smooth periodic change, which the state $x_{m_{i,j}}$ of the memristor in the cell $C(i,j)$ undergoes over each cycle of a sinusoidal voltage appearing between its terminals, and mathematically expressed by $v_{m_{i,j}} = \widehat{v}_{m_{i,j}} \cdot \sin(2 \cdot \pi \cdot f_{m_{i,j}} \cdot t)$, where $\widehat{v}_{m_{i,j}} = 2\,\text{V}$ and $f_{m_{i,j}} = 100\,\text{Hz}$. Clearly, at any given time instant, the cell is effectively a second-order dynamical system with degrees of freedom provided one by the memristor state and one by the capacitor voltage, which is also illustrated in plot (b) of **Figure 10**. Visualizing the memristor current flowing through the memristor as a result of the capacitor

[26]The Pershin and Di Ventra model from (Pershin et al., 2009) adopts the Heaviside functions step$(x_{m_{i,j}} - x_{on})$ and step$(x_{off} - x_{m_{i,j}})$ in place for the proposed continuous and differentiable variants, expressed by **Eqs. 26**, **27**, respectively. The use of these discontinuous functions does not always prevent the memristor state $x_{m_{i,j}}$ from exiting its existence domain $\mathcal{D}$ in numerical simulation of the original model. The proposed Biolek window-based boundary condition reformulation resolves this issue.

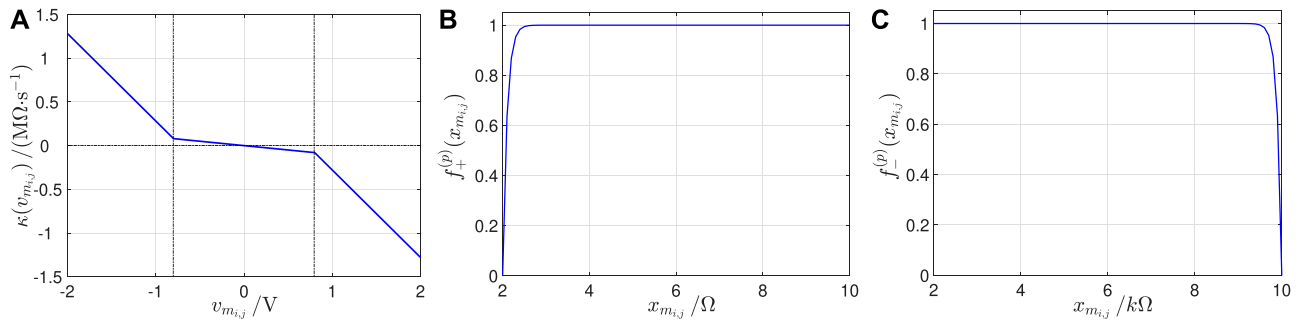


FIGURE 9 | (A) Course of $\kappa(v_{m_{ij}})$ as a function of $v_{m_{ij}}$ for $\alpha = 10^5\ \Omega{\cdot}\mathrm{V}^{-1}\ \mathrm{s}^{-1}$, $\beta = 10^6\ \Omega{\cdot}\mathrm{V}^{-1}\ \mathrm{s}^{-1}$, and $V_t = 0.8\,\mathrm{V}$. **(B,C)** Window functions, appearing in the state evolution function **Eq. 23**, and preventing $x_{m_{ij}}$ from decreasing below (increasing above) the lower **(upper)** bound x_{on} (x_{off}) under positive (negative) memristor voltages. Here $p = 40$, $x_{on} = 2\,\mathrm{k}\Omega$, and $x_{off} = 10\,\mathrm{k}\Omega$.

voltage from plot (b) vs. the capacitor voltage itself, the resulting pinched hysteresis loop, shown in **Figure 10C**, gives further evidence for the analogue dynamic behaviour of the cell memristor. The second M-CNN cell ODE **Eq. 19** governs the time evolution of the cell capacitor voltage $v_{x_{i,j}}$ within the memcomputing core of the circuit of **Figure 8**. Its right hand side is identical as in the ODE **Eq. 1** dictating the rate of change of the capacitor voltage within the computing core of the cell of the standard time- and space-invariant two-dimensional CNN discussed in **section 2.1**, except for the presence of an additional addend, resulting from the current through the memristor. It follows that the expression for the offset current $i_{w_{i,j}}$ of the memristive processing element of **Figure 8** is still given by **Eq. 11**, while, using **Eq. 22** to express the current through the memristor, the formula for the cell Internal DP Component $\tilde{g}_{i,j}$ features the new form

$$\tilde{i}_{g_{i,j}} \triangleq \tilde{i}_g\left(x_{m_{i,j}}, v_{x_{i,j}}\right) = -\frac{v_{x_{i,j}}}{x_{m_{i,j}}} - \frac{v_{x_{i,j}}}{R_x} - a_{0,0}\cdot R_y\cdot g_{lin}\cdot v_{sat} \text{ if } v_{x_{i,j}} < -v_{sat}, \tag{28}$$

$$\left(a_{0,0}\cdot R_y\cdot g_{lin} - \frac{1}{R_x} - \frac{1}{x_{m_{i,j}}}\right)\cdot v_{x_{i,j}} \text{ if } \left|v_{x_{i,j}}\right| \le v_{sat}, \tag{29}$$

$$-\frac{v_{x_{i,j}}}{x_{m_{i,j}}} - \frac{v_{x_{i,j}}}{R_x} + a_{0,0}\cdot R_y\cdot g_{lin}\cdot v_{sat} \quad \text{if } v_{x_{i,j}} > +v_{sat}. \tag{30}$$

in which **Eqs. 22**, **24** were employed to model the cell memristor current $i_{m_{i,j}}$ and the memductance function $G(x_{m_{i,j}})$, respectively. It is worth to note that the number of variables in the argument of $i_{g_{i,j}}$ is a signature for the order of the cell, as can be inferred by comparings **Eqs. 8**–**10** and **Eqs. 28**–**30**. The classical cell DRM technique (Chua, 2018a), reviewed in **section 2.2**, and adopted for the analysis and synthesis of standard CNNs with first-order processing elements, is applicable to dynamical systems with one degree of freedom only. As a result, the development of a systematic procedure to investigate and design M-CNNs with second-order memristive processing elements calls for a preliminary generalization of the DRM graphic tool. Drawing inspiration from the *phase portrait* concept from the theory of nonlinear dynamics (Strogatz, 2000), the next section introduces a new system-theoretic notion, which we name *Second-Order DRM* (DRM_2), enabling the investigation of the memcomputing capabilities of cellular nonlinear arrays with second-order memristive cells.

3.2 A Generalized DRM Technique for the Analysis of M-CNNs With Second-Order Processing Elements

In this section we extend the classical DRM methodology (Chua, 2018a) for the analysis of a nonlinear dynamic system with two degrees of freedom. Focusing, in particular, on the second-order M-CNN cell under study, the $x_{m_{i,j}}$–$v_{x_{i,j}}$ phase plane is the most natural domain, where the dynamical evolution of the two states of the system, described by **Eqs. 18**, **19**, may be studied. Let us first introduce the concept of *State Dynamic Portrait* (SDP).

Remark 2. With reference to the qualitative drawing in **Figure 11**, a SDP is a two-dimensional graph associated to a prescribed choice for the offset current value. It may be obtained as follows. First, the phase plane $x_{m_{i,j}}$–$v_{x_{i,j}}$ is partitioned into at most 4 distinct regions, differing in the *sign* $(\dot{x}_{m_{i,j}})$ and/or in the *sign* $(\dot{v}_{x_{i,j}})$, and distinguished according to the following coding map:

- Green region I: $\dot{v}_{x_{i,j}} < 0\ \mathrm{V/s}$ and $\dot{x}_{m_{i,j}} < 0\ \Omega/s$.
- Yellow region II: $\dot{v}_{x_{i,j}} > 0\ \mathrm{V/s}$ and $\dot{x}_{m_{i,j}} > 0\ \Omega/s$.
- Cyan region III: $\dot{v}_{x_{i,j}} > 0\ \mathrm{V/s}$ and $\dot{x}_{m_{i,j}} < 0\ \Omega/s$
- Gray region IV: $\dot{v}_{x_{i,j}} < 0\ \mathrm{V/s}$ and $\dot{x}_{m_{i,j}} > 0\ \Omega/s$

Then the loci $\dot{x}_{m_{i,j}} = 0\ \Omega/\mathrm{s}$ and $\dot{v}_{x_{i,j}} = 0\ \mathrm{V/s}$ – respectively known as $x_{m_{i,j}}$ and $v_{x_{i,j}}$ nullclines (Strogatz, 2000) – as well as their intersections – i.e., the equilibria of the ODE set **Eqs. 18**, **19** – are marked on the phase plane using the following symbolism:

- Red crosses: $\dot{x}_{m_{i,j}} = 0\ \Omega/\mathrm{s}$.
- Magenta diamonds: $\dot{v}_{x_{i,j}} = 0\ \mathrm{V/s}$.
- Black circles: $\dot{v}_{x_{i,j}} = 0\ \mathrm{V/s}$ and $\dot{x}_{m_{i,j}} = 0\ \Omega/\mathrm{s}$.

Particularly, the local instability (stability) of an equilibrium, studied by linearizing the state equations and studying the properties of the Jacobian, is graphically illustrated in a given SDP by means of a hollow (filled) black circle. The dynamical

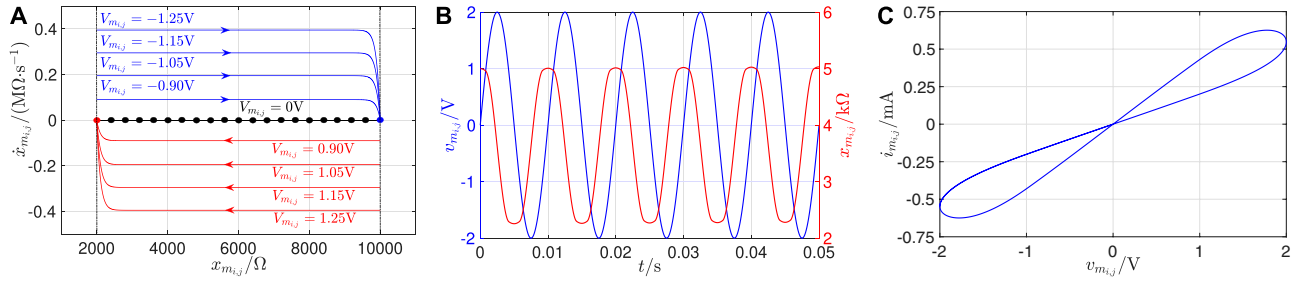


FIGURE 10 | (A) SDRs foliating from the memristor DRM for $V_{m_{ij}} \in \{-1.25, -1.15, -1.05, -0.90, 0, 0.90, 1.05, 1.15, 1.25\}$V. The blue (red) arrowed loci are associated to negative (positive) values for the memristor voltage. In the first (latter) case, the larger is $|V_{m_{ij}}|$, and the higher is the speed of the memristor state in its motion toward the equilibrium $\bar{x}_{m_{ij}} = x_{off\ (on)}$. The black locus represents the memristor POP. **(B)** Proof of evidence for the analogue dynamic response of the memristor state, hosted by the cell $C(i,j)$, to a sinusoidal voltage of the form $v_{m_{ij}} = \hat{v}_{m_{ij}} \cdot \sin(2 \cdot \pi \cdot f_{m_{ij}} \cdot t)$, with amplitude $\hat{v}_{m_{ij}} = 2$ V and frequency $f_{m_{ij}} = 100$ Hz, appearing between its terminals. **(C)** Pinched hysteresis loop emerging on the $v_{m_{ij}}$–$i_{m_{ij}}$ plane as a result of the device periodic excitation illustrated in **(B)**. The memristor model parameters are set as follows: $\alpha = 10^5\ \Omega\cdot\mathrm{V}^{-1}\ \mathrm{s}^{-1}$, $\beta = 10^6\ \Omega\cdot\mathrm{V}^{-1}\ \mathrm{s}^{-1}$, $V_t = 0.95$ V, $p = 40$, $x_{on} = 2\,\mathrm{k\Omega}$, $x_{off} = 10\,\mathrm{k\Omega}$.

behaviour of the state variables from any initial condition of interest may be qualitatively inferred by inspecting the direction of the vector field $(\dot{x}_{m_{ij}}, \dot{v}_{x_{ij}})$. In fact, phase plane trajectories[27], moving through regions I, II, III, and IV, proceed in the south-west, north-east, north-west, and south-east directions, as time goes by, respectively. The numerical integration of the pair of first-order coupled ODEs **Eqs. 18**, **19**, for initial conditions in the set of interest, allows to confirm this qualitative investigation on a quantitative basis, allowing to endow the partitioned plane, already accomodating nullclines and equilibria, with a number of phase plane trajectories, extracted by plotting the two solutions $v_{x_{ij}}(t)$ and $x_{m_{ij}}(t)$ of the model equations one against the other, and indicating, through the guide of arrows, placed on top of them, how the second-order M-CNN cell state evolves with time from prescribed starting points. An arrowed phase plane trajectory, marked in blue on a given SDP, is called a Second-Order SDR (SDR_2). Finally, the family of SDPs, obtained for each offset current value within a certain set of interest, takes the name of Second-Order DRM (DRM_2).

The proposed generalized DRM methodology may be used to analyze the operating principles of a given M-CNN with second-order memristive cells. Most importantly, the DRM_2 graphical tool allows to develop a systematic procedure to program one of the memristive cellular arrays under focus for the execution of a predefined memcomputing task, as outlined in the next section.

Remark 3. The DRM_2 graphic tool features a much more general applicability scope than this paper demonstrates. In fact, it allows to investigate any second-order dynamical system, including memristive circuit elements with two degrees of freedom.

3.3 A Rigorous DRM_2-Based Methodology for Robust M-CNN Design

The proposed DRM_2-based M-CNN design methodology (Ascoli et al., 2020a) allows to program the memristive nonlinear dynamic array i.e., to choose numerical values for the 19 cell core parameters[28] $\{\{a_{k,l}\}, \{b_{k,l}\}, z\}$ $(k, l \in \{-1, 0, 1\})$, in such a way that the processing element $C(i,j)$ may implement a predefined set of rules[29] (Chua, 1998), which, depending upon the specific data storage or processing operation to be executed, dictate the steady-state value[30] of its output voltage $v_{y_{ij}}(t_{i,j}^{(s)})$ for any combination of input voltage $v_{u_{ij}}$ and initial conditions $x_{m_{ij}}(0)$ and $v_{x_{ij}}(0)$ of its two

[27]A phase plane trajectory is the locus of points $(x_{m_{ij}}(t), v_{x_{ij}}(t))$, with first (second) coordinate at a given time extracted from the temporal succession of values of the memristor state (capacitor voltage) in the solution of the second-order ODE (18)–(19) for a given initial condition $(x_{m_{ij}}(0), v_{x_{ij}}(0))$.

[28]This 19-parameter set is often referred to as *gene* in CNN theory (Chua and Roska, 2002), since its 19 elements crucially affect the spatio-temporal phenomena, which may emerge in the space-invariant array of locally coupled cells, similarly as the DNA genetic content has a significant impact on the dynamical evolution of living beings. The 9 feedback (feedforward) synaptic weights in the set $\{a_{k,l}\}$ ($\{b_{k,l}\}$), with $k, l \in \{-1, 0, 1\}$, are typically arranged in a 3×3 matrix **A** (**B**), referred to as *feedback (feedforward) template* in CNN theory (Chua and Roska, 2002).

[29]The conditions, under which a space-invariant M-CNN is uncoupled, are the same as defined in **section 2.3** for a space-invariant CNN: $a_{k,l} = 0$ for all $k, l \in \{-1, 0, 1\}$ such that $(k, l) \neq (0, 0)$. For a memristive cellular array from this class the rules are said to be *local* (Chua and Roska, 2002), i.e., in general, they depend upon the input voltage $v_{u_{i+k,j+l}}$ and/or the initial conditions $x_{m_{i+k,j+l}}(0)$ and $v_{x_{i+k,j+l}}(0)$ of the two dynamical states of each cell $C(i+k, j+l)$ within the 3×3 sphere of influence of $C(i,j)$, only. In coupled M-CNNs the applicability of the some of the rules–referred to as *global* (Chua and Roska, 2002)–for the cell $C(i,j)$ may be conditioned by the input voltage $v_{u_{i+m,j+n}}$ and/or the initial conditions $x_{m_{i+m,j+n}}(0)$ and $v_{x_{i+m,j+n}}(0)$ of the two dynamical states of some remote cell $C(i+m, j+n)$, with $m \notin \{-1, 0, 1\}$ and/or $n \notin \{-1, 0, 1\}$.

[30]The M-CNN computing paradigm, this section is focused upon, revolves around the asymptotic convergence of the state vector $(x_{m_{ij}}, v_{x_{ij}})$ of the cell $C(i,j)$ to a relevant stable equilibrium $(\bar{x}_{m_{ij}}, \bar{v}_{x_{ij}})$, with ordinate located in either of the two saturation regions of the standard nonlinearity of **Eq. 3**, on the basis of a set of predefined task-dependent rules. However, as is the case for standard space-invariant CNNs satisfying the bistability condition, the output voltages of all the processing elements attain their final positive or negative saturation levels within a finite time frame. Similarly as in **section 2.2**, denoting with $t_{i,j}^{(s)}$ the time instant, at which the capacitor voltage $v_{x_{ij}}$ of the cell $C(i,j)$ enters the saturation region, which accommodates the equilibrium $(\bar{x}_{m_{ij}}, \bar{v}_{x_{ij}})$ the cell state vector $(x_{m_{ij}}, v_{x_{ij}})$ is expected to approach as time goes to infinity, the M-CNN may be considered at steady state, with respect to the output voltages of all its processing elements, from the time instant $t^{(s)} = \max_{1 \le i \le M, 1 \le j \le N}(t_{ij}^{(s)})$. This makes the memcomputing task outcome insensitive to a potential change in the location of some of the cell equilibria, which may occur due to non-idealities, including the intrinsic variability of nanodevices employed in the nonlinear dynamic array, especially the memristors.

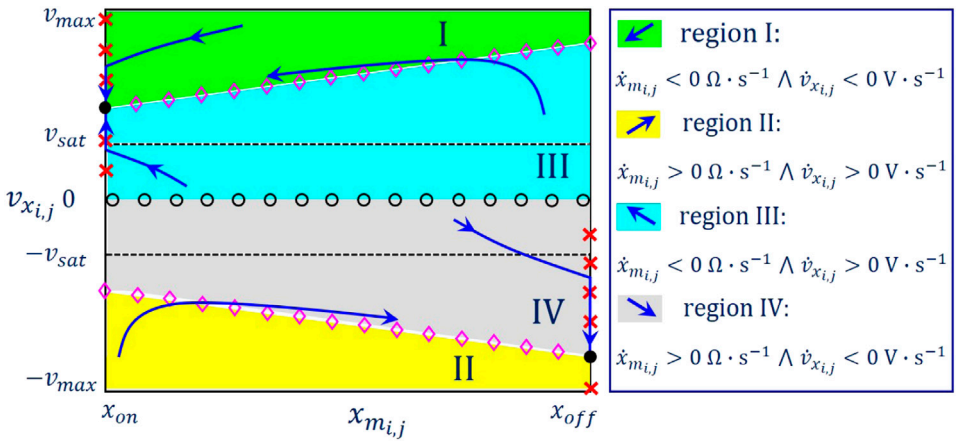


FIGURE 11 | Exemplifying SDP, which the M-CNN cell $C(i,j)$ would typically feature under $i_{w_{i,j}} = 0$ A, if $a_{0,0} > G_x + x_{on}^{-1}$. The capacitor voltage range under display is $[-v_{max}, v_{max}]$, with $v_{max} > v_{sat}$. The linear region $|v_{x_{i,j}}| \le v_{sat}$ is the rectangular domain enclosed within the two black dashed horizontal lines. The direction of motion of the state vector $(x_{m_{i,j}}, v_{x_{i,j}})$ in regions I, II, III, and IV is graphically illustrated in the legend.

dynamical states, and under specific conditions involving neighboring or remote processing elements. The proposed M-CNN design methodology complementing similar works – discussed in **section 2.3** – on the synthesis of CNN genes (Zarándy, 2003; Itoh and Chua, 2003), is based upon the following steps:

1. On the basis of the memcomputing task assigned to a given M-CNN, and with reference to the processing element $C(i,j)$, the designer should first roughly identify, under any possible combination of input voltage $v_{u_{i,j}}$, and of initial capacitor voltage $v_{x_{i,j}}(0)$ and memory resistance $x_{m_{i,j}}(0)$, and for any condition involving neighboring and/or remote cells, envisaged by the rule set, the most suitable partition of the two-dimensional state space $x_{m_{i,j}}$–$v_{x_{i,j}}$, which would guide the respective phase-plane trajectory toward an appropriate equilibrium. In other words, this step allows to specify the Family of SDPs i.e., the cell DRM_2, under target.
2. In order to derive numerical values for the parameter set $\{\{a_{k,l}\}, \{b_{k,l}\}, z\}$, where $k, l \in \{-1, 0, 1\}$, so as to endow the cell with the specified DRM_2, a number of inequalities, constraining, for each scenario of any rule, the behaviour of the sign $(\dot{x}_{m_{i,j}})$ and of the sign $(\dot{v}_{x_{i,j}})$ across the phase plane $x_{m_{i,j}}$–$v_{x_{i,j}}$ so as to control the number and stability properties of the equilibria, which it accommodates, are written down[31] through the use of the second-order ODE system **Eqs. 18**, **19**, with the expression for the offset current $i_{w_{i,j}}$, appearing in the latter state equation, preliminarily simplified as much as possible as compared to its general formula from **Eq. 11**, so as to implement the given data storage or processing task as efficiently as feasible.
3. A set of cell parameter values, satisfying concurrently all the aforementioned inequalities, shall be determined by means of a graphical approach, or through a numerical algorithm, depending upon the number of unknowns. Integrating numerically the state **Eqs. 18**, **19** of the M-CNN cell $C(i,j)$ for prescribed input voltage $v_{u_{i,j}}$ and vector state initial condition $(x_{m_{i,j}}(0), v_{x_{i,j}}(0))$ in each scenario encompassed in any rule, the resulting phase plane trajectories on the relevant SDP shall be found to evolve progressively toward the desired equilibria, allowing the cellular array to accomplish a predefined memcomputing task.

3.4 Application of the M-CNN Design Methodology to Execute Fundamental Memcomputing Tasks

In this section the proposed cell DRM_2 synthesis technique is applied to the cell model **Eqs. 18**, **19** to program the M-CNN to execute an image processing operation and a couple of memory functions, namely the data storage and retrieval tasks. Before presenting the M-CNN design examples, it is instructive to identify the most important properties of the second-order system **Eqs. 18**, **19** through the application of fundamental concepts from the theory of nonlinear dynamics (Strogatz, 2000).

From the first M-CNN cell ODE **Eq. 18**, the formulas for the $x_{m_{i,j}}$ nullclines (Strogatz, 2000) are

$$x_{m_{i,j}} = x_{off}, \text{ for } v_{x_{i,j}} < 0\,V, \tag{31}$$

[31]Despite template optimization (Chua and Roska, 2002) does not constitute the focus of this research study, in some cases, in order to improve the robustness of the M-CNN design, it may be adviceable to include additional inequalities, which may endow the design with a good tolerance to parameter variability. As examples, one could enforce a minimal distance between certain $x_{m_{i,j}}$ and $v_{x_{i,j}}$ nullclines to prevent the emergence of unwanted equilibria, or between the location of a desired equilibrium and the frontier between the saturation region, where it is due to reside, and the linear region. Moreover, in certain M-CNN designs, the use of the resistor of strictly positive conductance G_x in parallel to the cell capacitor allows to keep the power dissipation in the memristor within reasonable limits at equilibrium. Finally, within the domain of admissible solutions of a given IS, one should select a particular one holding some safety distance from the boundary with the remainder of the space of the cell unknown core parameters.

$$x_{m_{i,j}} \in \mathcal{D}, \text{for } v_{x_{i,j}} = 0\,\text{V}, \quad \text{and} \tag{32}$$

$$x_{m_{i,j}} = x_{on}, \text{for } v_{x_{i,j}} > 0\,\text{V}. \tag{33}$$

Employing now the second M-CNN cell ODE **Eq. 19**, the $v_{x_{i,j}}$ nullclines are found to be expressed by

$$v_{x_{i,j}} = \frac{i_{w_{i,j}} - a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat}}{G_x + \frac{1}{x_{m_{i,j}}}}, \tag{34}$$

for $v_{x_{i,j}} < -v_{sat}$, by

$$v_{x_{i,j}} = \frac{-i_{w_{i,j}}}{a_{0,0} \cdot R_y \cdot g_{lin} - G_x - \frac{1}{x_{m_{i,j}}}}, \tag{35}$$

for $\left|v_{x_{i,j}}\right| \leq v_{sat}$, and by

$$v_{x_{i,j}} = \frac{i_{w_{i,j}} + a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat}}{G_x + \frac{1}{x_{m_{i,j}}}}, \tag{36}$$

for $v_{x_{i,j}} > v_{sat}$

Remark 4. As it follows from **Eq. 29**, under $i_{w_{i,j}} = 0\,\text{A}$, the $v_{x_{i,j}}$ nullclines in the linear region consist of the segment, lying along the $x_{m_{i,j}}$ axis, and comprised between x_{on} and x_{off}, and, in case $a_{0,0}^{(-)} < a_{0,0} < a_{0,0}^{(+)}$ (see **Figure 12B** for an example, where $G_x = 0\,\Omega^{-1}$), where

$$a_{0,0}^{(-)} \triangleq \frac{G_x + x_{off}^{-1}}{R_y \cdot g_{lin}}, \text{ and} \tag{37}$$

$$a_{0,0}^{(+)} \triangleq \frac{G_x + x_{on}^{-1}}{R_y \cdot g_{lin}} \tag{38}$$

also of the two disjoint sets $v_{x_{i,j}} \in [-v_{sat}, 0\,\text{V})$, and $(0\,\text{V}, v_{sat}]$ for $x_{m_{i,j}} = \frac{1}{a_{0,0} \cdot R_y \cdot g_{lin} - G_x}$.

Remark 5. The application of the proposed design method to the specific M-CNN cell model **Eqs. 18**, **19** is unable to control existence and/or massage the shape of the $x_{m_{i,j}}$ nullclines, which are invariably set by equations **Eqs. 31–33**. However, the number and graphical look of the $v_{x_{i,j}}$ versus $x_{m_{i,j}}$ loci from equations **Eqs. 34–36** may be altered by tuning the cell model parameters, they are function of, so as to allow the synthesis of a suitable cell DRM_2 for the accomplishment of a predefined memcomputing task.

The equilibria, lying at the intersections between the $x_{m_{i,j}}$ and $v_{x_{i,j}}$ nullclines, are located at

$$Q^{(-)} \triangleq \left(\overline{x}_{m_{i,j}}^{(-)}, \overline{v}_{x_{i,j}}^{(-)}\right) = \left(x_{off}, \frac{i_{w_{i,j}} - a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat}}{G_x + x_{off}^{-1}}\right), \tag{39}$$

if

$$\frac{i_{w_{i,j}} - a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat}}{G_x + x_{off}^{-1}} < -v_{sat}, \tag{40}$$

in the negative saturation region, at

$$Q^{(0,-)} \triangleq \left(\overline{x}_{m_{i,j}}^{(0,-)}, \overline{v}_{x_{i,j}}^{(0,-)}\right) = \left(x_{off}, \frac{i_{w_{i,j}}}{-a_{0,0} \cdot R_y \cdot g_{lin} + G_x + x_{off}^{-1}}\right), \tag{41}$$

if

$$-v_{sat} \leq \frac{i_{w_{i,j}}}{-a_{0,0} \cdot R_y \cdot g_{lin} + G_x + x_{off}^{-1}} < 0\,\text{V}, \tag{42}$$

as well as at

$$Q^{(0,+)} \triangleq \left(\overline{x}_{m_{i,j}}^{(0,+)}, \overline{v}_{x_{i,j}}^{(0,+)}\right) = \left(x_{on}, \frac{i_{w_{i,j}}}{-a_{0,0} \cdot R_y \cdot g_{lin} + G_x + x_{on}^{-1}}\right), \tag{43}$$

if

$$0\,\text{V} < \frac{i_{w_{i,j}}}{-a_{0,0} \cdot R_y \cdot g_{lin} + G_x + x_{on}^{-1}} \leq v_{sat}, \tag{44}$$

in the linear region, and at

$$Q^{(+)} \triangleq \left(\overline{x}_{m_{i,j}}^{(+)}, \overline{v}_{x_{i,j}}^{(+)}\right) = \left(x_{on}, \frac{i_{w_{i,j}} + a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat}}{G_x + x_{on}^{-1}}\right), \tag{45}$$

if

$$\frac{i_{w_{i,j}} + a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat}}{G_x + x_{on}^{-1}} > v_{sat}, \tag{46}$$

in the positive saturation region.

Remark 6. Under $i_{w_{i,j}} = 0\,\text{A}$, each point defined as

$$Q^{(0)} = \left(x_{m_{i,j}}^{(0)}, v_{x_{i,j}}^{(0)}\right), \text{ with } x_{m_{i,j}}^{(0)} \in \mathcal{D}, \text{ and } v_{m_{i,j}}^{(0)} = 0\,\text{V}, \tag{47}$$

represents a possible equilibrium for the M-CNN cell in the linear region. Moreover, in case $a_{0,0} = a_{0,0}^{(-)}$ ($a_{0,0} = a_{0,0}^{(+)}$), with $a_{0,0}^{(-)}$ ($a_{0,0}^{(+)}$) defined in equation **Eqs. 37**, **38**, also each point along the vertical line of the $x_{m_{i,j}}$–$v_{x_{i,j}}$ phase plane, passing through the memristor state upper (lower) bound x_{off} (x_{on}), and stretching over the capacitor voltage range $v_{x_{i,j}} \in [-v_{sat}, 0\,\text{V})$ ($v_{x_{i,j}} \in (0\,\text{V}, v_{sat}]$) denotes an additional M-CNN cell equilibrium in the linear region (Ascoli et al., 2020b). From the first M-CNN cell ODE **Eq. 18**, it follows that the memristor state $x_{m_{i,j}}$ increases if

$$v_{x_{i,j}} < 0 \quad \text{and} \quad x_{m_{i,j}} \in \left[x_{on}, x_{off}\right) \tag{48}$$

and decreases if

$$v_{x_{i,j}} > 0 \quad \text{and} \quad x_{m_{i,j}} \in \left(x_{on}, x_{off}\right] \tag{49}$$

Thus, as revealed by the illustrative cell SDP example of **Figure 11**, the motion of a trajectory point $(x_{m_{i,j}}, v_{x_{i,j}})$ on a given SDP points toward the east (west) in the phase plane lower (upper) half. Inspecting now the second M-CNN cell ODE (19), the capacitor voltage $v_{x_{i,j}}$ is found to increase provided

$$v_{x_{i,j}} < \frac{i_{w_{i,j}} - a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat}}{G_x + \frac{1}{x_{m_{i,j}}}}, \tag{50}$$

for $v_{x_{i,j}} < -v_{sat}$, provided

$$\left(v_{x_{i,j}} + \frac{i_{w_{i,j}}}{a_{0,0} \cdot R_y \cdot g_{lin} - G_x - \frac{1}{x_{m_{i,j}}}}\right) \cdot \left(x_{m_{i,j}} - \frac{1}{a_{0,0} \cdot R_y \cdot g_{lin} - G_x}\right) > 0, \tag{51}$$

for $\left|v_{x_{i,j}}\right| \leq v_{sat}$, and provided

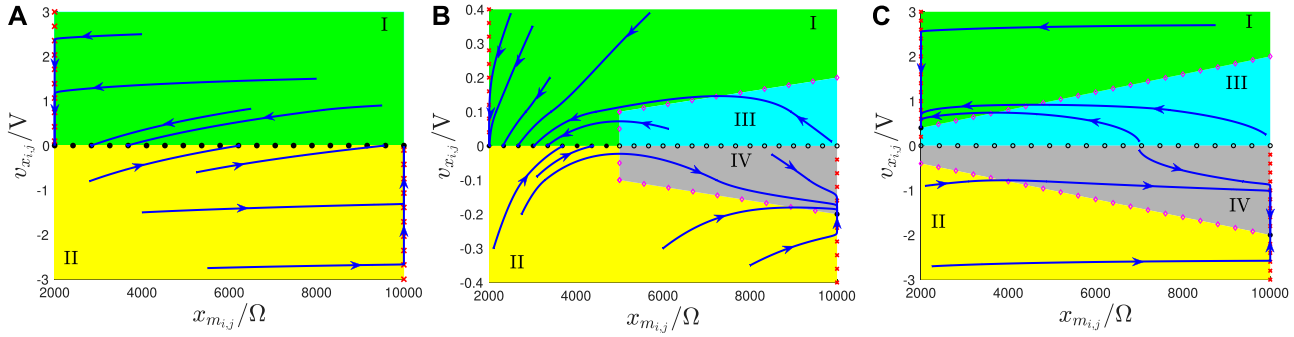


FIGURE 12 | Cell SDP, emerging for the fixed circuit parameter setting from **Table 1**, under $i_{w_{ij}} = 0\,\mathrm{A}$, and $G_x = 0\,\Omega^{-1}$, and featuring a continuum of stable equilibria for $a_{0,0} = 0\Omega^{-1}$ **(A)**, a stable isolated equilibrium, as well as a line of equilibria with stable (unstable) character to the left **(right)** of a bifurcation point for $a_{0,0} = 2 \cdot 10^{-4}\,\Omega^{-1}$ **(B)**, and, finally, two stable isolated equilibria, as well as a continuum of unstable equilibria for $a_{0,0} = 2 \cdot 10^{-3}\Omega^{-1}$ **(C)**.

$$v_{x_{i,j}} < \frac{i_{w_{i,j}} + a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat}}{G_x + \frac{1}{x_{m_{i,j}}}}, \tag{52}$$

for $v_{x_{i,j}} > v_{sat}$, and it decreases provided the inequality sign in each of the $i_{w_{i,j}}$-dependent conditions **Eqs. 50–52** is inverted. On the basis of inequalities **Eqs. 50**, **52**, dictating the conditions under which $\mathrm{sign}\dot{v}_{x_{i,j}} > 0\mathrm{V} \cdot \mathrm{s}^{-1}$ in the negative, linear, and saturation region, respectively, it is now possible to understand the reason why the trajectory point $(x_{m_{i,j}}, v_{x_{i,j}})$ moves northward or southward in the illustrative cell SDP example of **Figure 11**.

3.4.1 Zero Offset Current Scenario

It may be proved that, unlike a standard CNN processing element, the M-CNN cell $C(i,j)$ may never exhibit *monostability* for $i_{w_{i,j}} = 0\,\mathrm{A}$. In other words, under no circumstances may the respective SDP host one and only one globally asymptotically stable equilibrium (Strogatz, 2000). **Table 2** sums up (Ascoli et al., 2020b) the location and local stability property of each equilibrium, which the M-CNN cell $C(i,j)$ may admit under zero offset current depending upon the self-feedback synaptic weight $a_{0,0}$.

Specifying the values[32], reported in **Table 3**, for all the fixed parameters of the cell circuit of **Figure 8**, the viewgraphs in plots (a), (b), and (c) of **Figure 12** illustrate the SDP of a M-CNN processing element, accommodating no linear resistor in the memcomputing core, under zero offset current, and for a specific value of the self-feedback synaptic weight $a_{0,0}$ in the first, second, and third of the three sets reported in **Table 2** and Ascoli et al. (2020b).

3.4.2 Non-Zero Offset Current Scenario

Allowing a non-null offset current, accounting mostly for the coupling effects, to flow through the capacitor in the circuit of **Figure 8** may endow the processing element with monostability, which is useful for the accomplishment of certain mem-computing tasks, as will be clear from the discussion of some M-CNN designs in the sections to follow. **Table 4**, in which $i_1(a_{0,0})$ and $i_2(a_{0,0})$ are defined as

$$i_1(a_{0,0}) \triangleq \left(a_{0,0} \cdot R_y \cdot g_{lin} - G_x - x_{off}^{-1}\right) \cdot v_{sat}, \text{ and} \tag{53}$$

$$i_2(a_{0,0}) \triangleq \left(-a_{0,0} \cdot R_y \cdot g_{lin} + G_x + x_{on}^{-1}\right) \cdot v_{sat}, \tag{54}$$

classifies the number, location, and local stability property of all the equilibria which a M-CNN cell may possibly admit for all the possible combinations of self-feedback synaptic weight $a_{0,0}$ and offset current $i_{w_{ij}}$.

Remark 7. Interestingly, this table allows to draw the codimension-2 bifurcation diagram of **Figure 13**, in which, without loss of generality, G_x *was set to* $0\,\Omega^{-1}$. This graph, which, taking inspiration from CNN theory (Chua, 1998; Chua and Roska, 2002) is called M-CNN Primary Mosaic, visualizes the partitioning of the $a_{0,0}$–$i_{w_{ij}}$ plane in domains differing one from the other in at least one of the stable equilibria, which the solutions of the ODE of a cell from the class of uncoupled M-CNNs may possibly approach depending upon the initial conditions. For each of such domains in **Figure 13**, a distinct color is chosen to fill the space within its boundaries, and the indication of the stable and unstable equilibria, which the M-CNN cell admits for any pair $(a_{0,0}, i_{w_{ij}})$ residing therein, is given.In this manuscript the proposed DRM_2-based M-CNN design methodology shall be applied to the model **Eqs. 18**, **19** of a cell belonging to the class of uncoupled M-CNNs, and featuring an offset current, which, in comparison to its most general formula, namely **Eq. 11**, reduces to **Eq. 13**.

4 M-CNN AS A BIO-INSPIRED IMAGE PROCESSING ENGINE

A M-CNN may be programmed to carry out any image processing operation, which a classical CNN is able to execute. To provide some evidence for this claim, the next section discusses the system-theoretic design of a memristive cellular array for the extraction of edges from a binary image.

[32]These very same values will be assigned to all the invariable cell circuit parameters in each of the M-CNN designs discussed below.

TABLE 2 | Location and local stability property of each of the equilibrium points, which a M-CNN cell may possibly admit, depending upon $a_{0,0}$, under $i_{w_{i,j}} = 0$ A (Ascoli et al., 2020b). The coordinates of $Q^{(-)}$, $Q^{(0)}$, and $Q^{(+)}$ are indicated in **Eqs. 39**, **47**, and **45**, respectively. With reference to the table content, we define $a_{0,0}^{(-)} \triangleq \breve{x}_{off}^{-1} \cdot R_y^{-1} g_{lin}^{-1}$, and $a_{0,0}^{(+)} \triangleq \breve{x}_{on}^{-1} \cdot R_y^{-1} g_{lin}^{-1}$. The marginal case $a_{0,0} = a_{0,0}^{(-)}$ ($a_{0,0} = a_{0,0}^{(+)}$), in which, as mentioned in Remark 6, an additional line of equilibria, namely each point along the vertical line passing through the memristor state upper (lower) bound and stretching across the capacitor voltage range $v_{x_{i,j}} \in [-v_{sat}, 0\,\mathrm{V})$ ($v_{x_{i,j}} \in (0\,\mathrm{V}, v_{sat}]$), appears in the linear region, is not tabulated here, but the interested reader is invited to consult (Ascoli et al., 2020b).

Self-feedback synaptic weight range	Cell equilibrium location	Local stability property
$a_{0,0} < a_{0,0}^{(-)}$	$Q^{(0)} = (\bar{x}_{m_{i,j}}^{(0)}, \bar{v}_{x_{i,j}}^{(0)})$ with $x_{m_{i,j}}^{(0)} \in \mathcal{D}$, and $v_{m_{i,j}}^{(0)} = 0\,\mathrm{V}$	Stable for all $x_{m_{i,j}}^{(0)}$ values
$a_{0,0} \in (a_{0,0}^{(-)}, a_{0,0}^{(+)})$	$Q^{(-)} = \left(x_{off}, \frac{-a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat}}{G_x + x_{off}^{-1}}\right)$	Stable
—	$Q^{(0)} = (\bar{x}_{m_{i,j}}^{(0)}, \bar{v}_{x_{i,j}}^{(0)})$ with $x_{m_{i,j}}^{(0)} \in \mathcal{D}$, and $v_{m_{i,j}}^{(0)} = 0\,\mathrm{V}$	Stable if $x_{m_{i,j}}^{(0)} \in \left[x_{on}, \frac{1}{a_{0,0} \cdot R_y \cdot g_{lin} - G_x}\right)$
$a_{0,0} > a_{0,0}^{(+)}$	$Q^{(-)} = \left(x_{off}, \frac{-a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat}}{G_x + x_{off}^{-1}}\right)$	Stable
—	$Q^{(0)} = (\bar{x}_{m_{i,j}}^{(0)}, \bar{v}_{x_{i,j}}^{(0)})$ with $x_{m_{i,j}}^{(0)} \in \mathcal{D}$, and $v_{m_{i,j}}^{(0)} = 0\,\mathrm{V}$	Unstable for all $x_{m_{i,j}}^{(0)}$ values
—	$Q^{(+)} = \left(x_{on}, \frac{a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat}}{G_x + x_{on}^{-1}}\right)$	Stable

4.1 Edge M-CNN

This section is devoted to the design of a $M \times N$ memristive array capable to extract the edges from an input binary image featuring as many rows and columns as the M-CNN. The local rule triplet, each M-CNN cell, featuring the circuitry shown in **Figure 8**, is requested to comply with, so as to execute this image processing task[33], are reported in **Table 1** from **section 2.4**. In order to ensure that the memprocessing elements obey this local rule set, it is wise to synthesize the cell SDP pertaining to each scenario from rules 1 and 2 (rule 3) in such a way that it accommodates one and only one equilibrium located in the negative (positive) saturation region i.e., $Q^{(-)}$ ($Q^{(+)}$), as specified by **Eqs. 39**, **45**. In order to ease the understanding of the steps of the proposed M-CNN design methodology, it is adviceable to provide its result in advance. Plots (a), (b), and (c) of **Figure 14** respectively illustrate the SDP of a M-CNN processing element, which obeys[34] rule 1 for $n_B = 0$, rule 2, and rule 3 for $n_B = 7$. As may be inferred by inspecting plots (a) and (b) (plot (c)), here the cell monostability in both rules 1 and 2 (in rule 3) is enforced by making sure that only the negative (positive) saturation region hosts a $v_{x_{i,j}}$ nullcline, as expressed by **Eqs. 34**, **36**, and highlighted by means of magenta diamonds, and imposing that such $v_{x_{i,j}} = v_{x_{i,j}}(x_{m_{i,j}})$ characteristic form a point of intersection, as defined by **Eqs. 39**, **45**, and marked via a black circle, with the vertical $x_{m_{i,j}}$ nullcline **Eqs. 31**, **33** indicated through red crosses in the phase plane lower (upper) half. Adopting such a cell DRM_2 synthesis strategy, in any scenario of rule 1 and for rule 2 (under all circumstances in rule 3), a state vector $(x_{m_{i,j}}, v_{x_{i,j}})$ positioned below/above the $v_{x_{i,j}}$ nullcline **Eqs. 34**, **36** is constrained to move in the north/south direction, bending eastward or westward in the phase plane lower or upper half, respectively, toward the point **Eqs. 39**, **45**, denoting, as a result, a globally asymptotically stable equilibrium for the second-order ODE system **Eqs. 18**, **19**, as the filling of the respective black circle marker in plots (a) and (b) (plot (c)) of **Figure 14** clearly indicates. Plots (a.1) (a.2), and (a.3) ((b.1), (b.2), and (b.3)) of **Figure 15** graphically visualize the steps, envisaged by the proposed cell SDP synthesis approach, and discussed shortly, to shape the phase portrait of the second-order ODE **Eqs. 18**, **19** in the linear, negative (positive) saturation, and positive (negative) saturation regions, respectively, so as to enforce local rules 1 and 2 (rule 3) from **Table 1**. Through a rigorous mathematical analysis of **Eqs. 18**, **19** in each region of the standard output nonlinearity **Eq. 3** we shall next derive an ad-hoc IS set, allowing to massage the cell DRM_2, as illustrated in **Figure 15**. Previous to initiate this investigation, a couple of aspects should be pinpointed. Firstly, the cell ODE initial condition $(x_{m_{i,j}}(0), v_{x_{i,j}}(0))$ may be chosen arbitrarily, since, as mentioned earlier, irrespective of the rule, the phase plane will be allowed to host one and only one GAS equilibrium in any possible scenario. Secondly, we assume the same expression for the offset current as in the EDGE CNN design, namely **Eq. 14**. Let us suppose that the parameter values for $b \in \mathbb{R}_{<0}$ and $z \in \mathbb{R}_{<0}$ are known. As a result,

TABLE 3 | Setting of specific M-CNN cell circuit parameters, specifically α, β, and V_t from **Eq. 25**, x_{on}, x_{off}, and p from **Eqs. 26**, **27**, C_x from **Eq. 19**, I from **Eq. 4**, R_y from **Eq. 2**, as well as g_{lin} and v_{sat} from **Eq. 3**, which are kept unchanged in the design examples to follow.

$\alpha/\ (\Omega \cdot \mathbf{s}^{-1} \cdot \mathbf{V}^{-1})$	$\beta/\ (\Omega \cdot \mathbf{s}^{-1} \cdot \mathbf{V}^{-1})$	$\mathbf{V_t}/\mathbf{V}$	$\mathbf{x_{on\ (off)}}/\ \mathbf{k\Omega}$	p
10^5	10^6	0.8	2 (10)	40
$C_x/\,\mu$F	I/A	$R_y/\,$kΩ	$g_{lin}/\,\mathrm{m}\Omega^{-1}$	$v_{sat}/$V
10	1	1	1	0.1

[33]Indicating the memristor resistance $x_{m_{i,j}}$ (capacitor voltage $v_{x_{i,j}}$) of the cell $C(i,j)$ as state 1 (2), the convention adopted in the EDGE M-CNN design for mapping a given input image with $M \times N$ pixels ($M \cdot N$ real-valued initial cell capacitor voltages) onto $M \cdot N$ real-valued cell input voltages (onto an initial state 2 image with $M \times N$ pixels) is identical to the approach followed in **section 2.1**, while discussing the operating principles of the standard array implementation. However, since v_{sat} is set to 0.1 V here, the colour coding map for the visualization of the steady-state cell output voltages differs from the strategy used in **section 2.1**. A negative (positive) saturation voltage level for the steady-state output voltage of a cell is converted into a white (black) pixel for the steady-state output image.

[34]As will be clarified shortly, the scenario $n_B = 0$ (7) represents the most critical setting for the cell SDP synthesis in rule 1 (3), and, for this reason, is referred to as worst-case scenario.

TABLE 4 | Location and local stability property of each of the equilibrium points, which a M-CNN cell may possibly admit, depending upon $a_{0,0}$ and $i_{w_{i,j}}$. The coordinates of equilibria $Q^{(-)}$, $Q^{(0)}$, $Q^{(0)}$, and $Q^{(+)}$ are respectively specified in **Eqs. 39**, **41**, **43**, and **45**. The formulas for $a_{0,0}^{(-)}$, $a_{0,0}^{(+)}$, $i_1\,(a_{0,0})$, and $i_2\,(a_{0,0})$ are respectively expressed by **Eqs. 37**, **38**, **53**, and **54**. The local stability nature of each of the possible cell equilibria is also revealed. The analysis of the marginal cases is omitted from this table.

Offset current range	Self-feedback synaptic weight range	Cell equilibrium location	Local stability property
$i_{w_{i,j}} < i_1\,(a_{0,0})$	For all $a_{0,0}$ values	$Q^{(-)} = \left(x_{off}, \frac{i_{w_{i,j}} - a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat}}{G_x + x_{off}^{-1}}\right)$	Stable
$i_{w_{i,j}} \in (i_1\,(a_{0,0}), 0\,\mathrm{A})$	$a_{0,0} < a_{0,0}^{(-)}$	$Q^{(0)} = \left(x_{off}, \frac{i_{w_{i,j}}}{-a_{0,0} \cdot R_y \cdot g_{lin} + G_x + x_{off}^{-1}}\right)$	Stable
$i_{w_{i,j}} \in (0\,\mathrm{A}, i_1\,(a_{0,0}))$	$a_{0,0} > a_{0,0}^{(-)}$	—	Unstable
$i_{w_{i,j}} \in (0\,\mathrm{A}, i_2\,(a_{0,0}))$	$a_{0,0} < a_{0,0}^{(+)}$	$Q^{(0)} = \left(x_{on}, \frac{i_{w_{i,j}}}{-a_{0,0} \cdot R_y \cdot g_{lin} + G_x + x_{on}^{-1}}\right)$	Stable
$i_{w_{i,j}} \in (i_2\,(a_{0,0}), 0\,\mathrm{A})$	$a_{0,0} > a_{0,0}^{(+)}$	—	Unstable
$i_{w_{i,j}} > i_2\,(a_{0,0})$	For all $a_{0,0}$ values	$Q^{(+)} = \left(x_{on}, \frac{i_{w_{i,j}} + a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat}}{G_x + x_{on}^{-1}}\right)$	Stable

the M-CNN gene synthesis technique will target the derivation of suitable values for $a_{0,0}$ and $b_{0,0}$. An appropriate IS in these two unknowns is derived next. The analysis of **Eqs. 18**, **19** focuses first on the linear region of the phase plane.

4.1.1 Edge M-CNN Cell DRM_2 Synthesis in the Linear Region

With reference to plot (a.1) (b.1) of **Figure 15**, the aim of this section is to make sure that, under all circumstances in rule 1 and for rule 2 (in all scenarios of rule 3), the characteristic $v_{x_{i,j}} = v_{x_{i,j}}(x_{m_{i,j}})$ of **Eq. 35** lies entirely within the domain $v_{x_{i,j}} > (<) + (-)v_{sat}$, as indicated by means of a dashed brown curve without magenta diamonds. The inequality

$$a_{0,0} \cdot R_y \cdot g_{lin} - G_x > x_{on}^{-1}, \tag{55}$$

ensures a positive sign for the denominator of the rational function on the right hand side of **Eq. 35** irrespective of the value assumed by the memristor state $x_{m_{i,j}}$ throughout its existence domain $\mathcal{D}$. Provided the constraint **Eq. 55** holds true, enforcing a negative (positive) polarity for the offset current[35] in each scenario of rule 1 and for rule 2 (under all circumstances in rule 3) via

$$i_{w_{i,j}}\left(v_{u_{i,j}}, n_B\right) < (>)\,0\,\mathrm{A}, \tag{56}$$

ensures that the $v_{x_{i,j}}$–$x_{m_{i,j}}$ locus, expressed by **Eq. 35**, falls entirely within the phase plane positive (negative) half in these cases. It is simple to show that, under the satisfaction of constraint **Eq. 56** with the first (second) inequality sign in rules 1 and 2 (rule 3), the function $v_{x_{i,j}} = v_{x_{i,j}}(x_{m_{i,j}})$ features an upward (downward) concavity, decreasing (increasing) monotonically with the memristor state, as shown in plot (a.1) ((b.1)) of **Figure 15**. As a result, under all possible circumstances in rule 1 and for rule 2 (in all scenarios of rule 3) the characteristic $v_{x_{i,j}} = v_{x_{i,j}}(x_{m_{i,j}})$ lies completely within the positive (negative) saturation region, as depicted in **Figure 15** (a.1) ((b.1)), provided

$$\frac{-i_{w_{i,j}}\left(v_{u_{i,j}}, n_B\right)}{a_{0,0} \cdot R_y \cdot g_{lin} - G_x - x_{off}^{-1}} > (<) + (-)v_{sat} \tag{57}$$

Let us now study the direction of motion of the state vector $(x_{m_{i,j}}, v_{x_{i,j}})$ throughout the linear region. The enforcement of inequality **Eq. 55** endows the second factor on the left hand side of constraint **Eq. 51** with a strictly positive sign. It follows that, within the domain $\left|v_{x_{i,j}}\right| \leq v_{sat}$, the capacitor voltage of the cell circuit of **Figure 8** increases over time if

$$v_{x_{i,j}} > -\frac{i_{w_{i,j}}}{a_{0,0} \cdot R_y \cdot g_{lin} - G_x - \frac{1}{x_{m_{i,j}}}} \tag{58}$$

Since, as established by constraint **Eq. 57**, in rules 1 and 2 (rule 3) the right hand side of inequality **Eq. 58** assumes values larger (lower) than $+(-)v_{sat}$ throughout the memristor state existence domain, phase plane trajectories of the linear region move toward the south (north), bending eastward or westward in the phase plane lower or upper half, as established by condition **Eq. 48** or **Eq. 49**, visiting[36] the gray (yellow) region IV (II) or the green (cyan) region I (III), respectively, as indicated in plot (a.1) ((b.1)) of **Figure 15**.

4.1.2 Edge M-CNN Cell DRM_2 Synthesis in the Saturation Regions

The goal of this section is twofold. On one hand, in each scenario of rule 1 and for rule 2 (under all circumstances from rule 3) the cell SDP is expected to accommodate one and only one GAS equilibrium point, specifically **Eqs. 39**, **45**, over the domain $v_{x_{i,j}} < (>) - (+)v_{sat}$, as indicated by the black-filled circle in plot (a.2) (b.2) of **Figure 15**. On the other hand, in order to avoid the existence of a $\dot{v}_{x_{i,j}} = 0\,\mathrm{V} \cdot \mathrm{s}^{-1}$ locus in the positive (negative) saturation region under any circumstance in rule 1 and for rule 2 (in all scenarios of rule 3), the whole $v_{x_{i,j}} = v_{x_{i,j}}(x_{m_{i,j}})$ characteristic, expressed by **Eqs. 34**, **36** should fall below (above) the horizontal line $v_{x_{i,j}} = +(-)v_{sat}$, as sketched in **Figure 15** (a.3) ((b.3)), where the three dashed brown curves show its three possible shape variants. It is straightforward to verify that, in view of inequality **Eq. 56**, in any of the possible scenarios of rule 1 and for rule 2 (under all circumstances in rule 3), the $v_{x_{i,j}}$ nullcline **Eqs. 34**, **36** features upward (downward) concavity as it decreases (increases) monotonically with the memristor state. As a result,

[35]Given that, as discussed in **section 3.4**, under $i_{w_{i,j}} = 0$ A, the M-CNN cell is unable to exhibit monostable behaviour, its capacitor current necessarily includes a nonzero offset current in this design.

[36]The direction of motion of the state vector $(x_{m_{i,j}}, v_{x_{i,j}})$ within each of the four possible regions, which may partition a cell SDP, is qualitatively indicated in the legend of **Figure 11**.

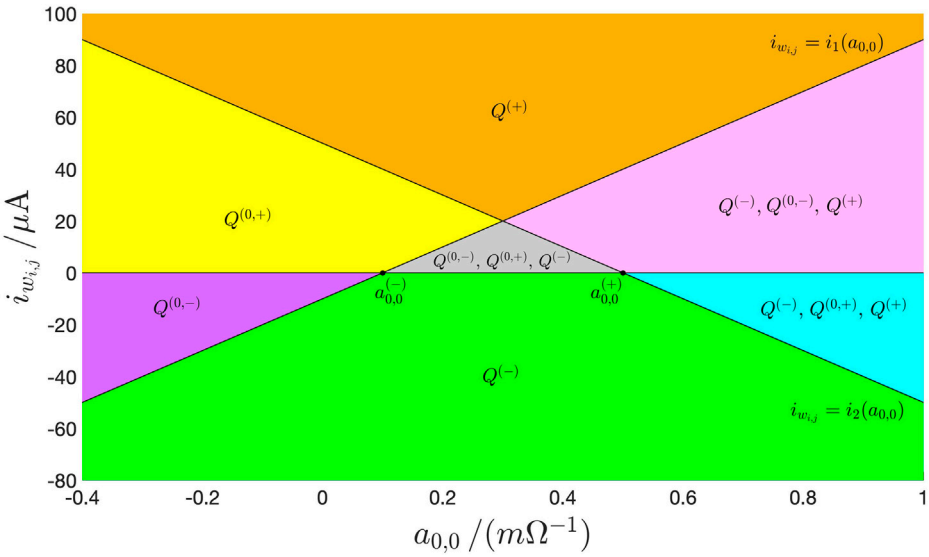


FIGURE 13 | The *M-CNN Primary Mosaic*: codimension-2 bifurcation diagram illustrating all the admissible equilibria the two states of the second-order cell from the class of uncoupled M-CNNs may approach asymptotically depending upon the specific region of the $a_{0,0}$–$i_{w_{ij}}$ parameter plane, in which the values assigned to the self-feedback synaptic weight and to the offset current reside. A red (black) color is adopted for the symbol of each unstable (stable) M-CNN cell equilibrium, as specified in **Table 2**. Without loss of generality, here G_x was set to 0 Ω^{-1}. The coordinates of $Q^{(-)}$, $Q^{(0)}$, $Q^{(0)}$, and $Q^{(+)}$ are indicated in **Eqs. 39**, **41**, **43**, and **45**. The formulas for $a_{0,0}^{(-)}$, $a_{0,0}^{(+)}$, $i_1(a_{0,0})$, and $i_2(a_{0,0})$ are respectively expressed by **Eqs. 37**, **38**, **53**, and **54**. Details on the possible steady-state behaviours of the M-CNN cell in the marginal cases $i_{w_{ij}} = 0$ A, $i_{w_{ij}} = i_1(a_{0,0})$, and $i_{w_{ij}} = i_2(a_{0,0})$ have not been reported in the bifurcation diagram to keep the illustration as clear as possible. Importantly, as studied earlier, under zero offset current, irrespective of the value specified for $a_{0,0}$, in the linear region the cell admits each equilibrium $Q^{(0)}$, which, according to **Eq. 47**, lies along the segment of the horizontal axis of the $x_{m_{ij}}$–$v_{x_{ij}}$ phase plane comprised between x_{on} and x_{off}.

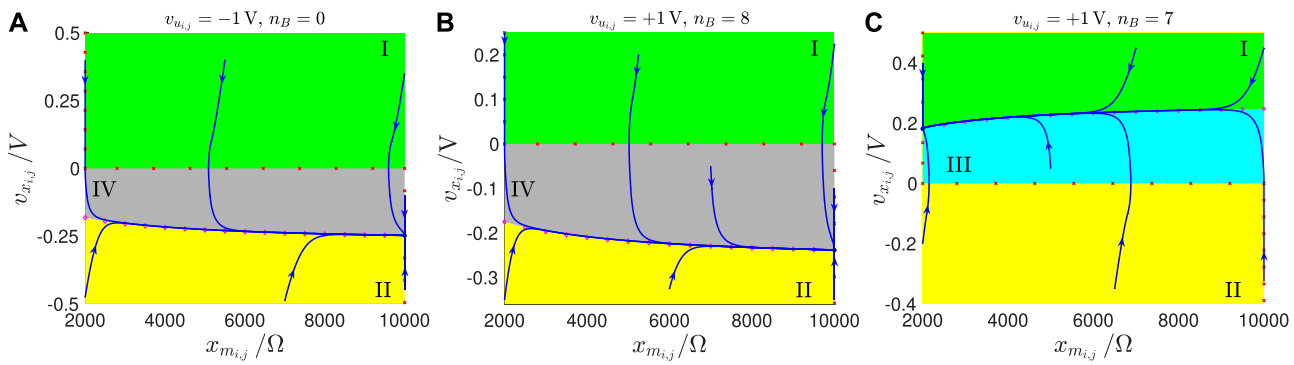


FIGURE 14 | (A) SDP of a cell $C(i,j)$ featuring the input voltage $v_{u_{ij}} = -1$ V in the worst-case scenario $n_B = 0$ of rule 1. **(B,C)** SDP of a cell $C(i,j)$ featuring the input voltage $v_{u_{ij}} = +1$ V in the only scenario of rule 2 (in the worst-case scenario $n_B = 7$ of rule 3).

imposing that it further assumes a value smaller (larger) than the negative (positive) saturation voltage, when the memristor state sits at its lowest bound i.e.,

$$\frac{i_{w_{ij}}\left(v_{u_{ij}}, n_B\right) - (+)a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat}}{G_x + x_{on}^{-1}} < (>) - (+)\, v_{sat}, \qquad (59)$$

in each of the first (latter) set of scenarios, ensures that this unique $\dot{v}_{x_{ij}} = 0\,\mathrm{V} \cdot \mathrm{s}^{-1}$ locus lies within the negative (positive) saturation region for all $x_{m_{ij}} \in \mathcal{D}$, as visualized through a dashed brown curve with magenta diamonds in plot (a.2) ((b.2)) of **Figure 15**. This in turn allows the formation of a cell equilibrium, as expressed by **Eqs. 39**, **45**, and visualized through a black circle in plot (a.2) ((b.2)) of **Figure 15**, for each value of $n_B \in \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$ under $v_{u_{ij}} = -1$ V, and for $n_B = 8$ under $v_{u_{ij}} = +1$ V (for each value of $n_B \in \{0, 1, 2, 3, 4, 5, 6, 7\}$ under $v_{u_{ij}} = +1$ V). Thus, with regard to rules 1 and 2 (rule 3), recalling that $\dot{x}_{m_{ij}}$ features a positive sign throughout the phase plane lower half, as established by **Eq. 48**, and recalling the condition **Eqs. 50**, **52**, which guarantees an increase of the cell capacitor voltage over time in the negative (positive) saturation region, phase plane trajectories, visiting the domain

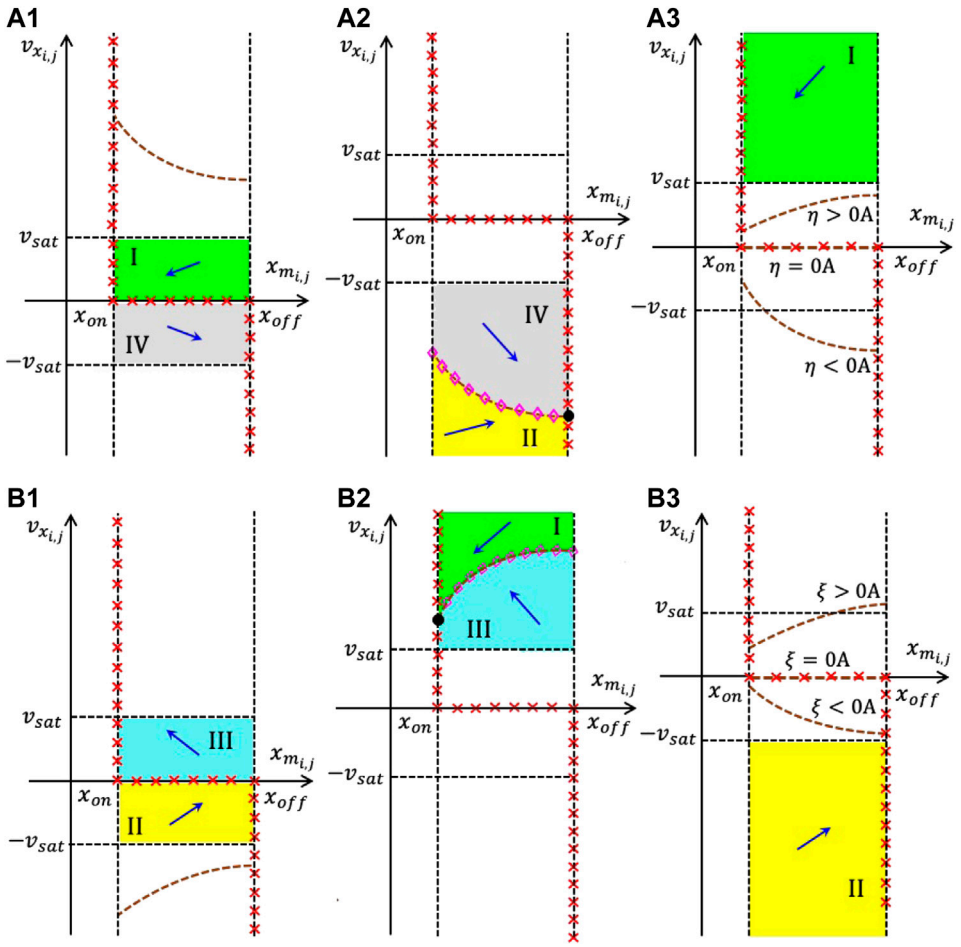


FIGURE 15 | Qualitative visualization of the strategy adopted to massage the shape of the cell DRM_2 in the EDGE as well as in the STORE M-CNN designs. Here the stepwise application of the proposed systematic gene synthesis methodology in the linear region (a.1) ((b.1)), in the negative (positive) saturation region (a.2) ((b.2)), and in the positive (negative) saturation region (a.3) ((b.3)) of the phase plane $x_{m_{i,j}}$–$v_{x_{i,j}}$ enables to enforce the appearance of a single $v_{x_{i,j}}$ nullcline, namely **Eqs. 34**, **36**, and the existence of one and only one equilibrium, specifically **Eqs. 39**, **45**, in the EDGE cell SDP, which emerges in each scenario of rule 1 and for rule 2 (under all circumstances in rule 3) from **Table 4**, as well as in the STORE cell SDP, which forms under the hypothesis of rule 1 (2) from **Table 5**. With reference to the first (latter) set of scenarios, combining plots (a.1), (a.2), and (a.3) ((b.1), (b.2), and (b.3)) provides an ad-hoc cell SDP, given that the phase-plane partition guides all trajectories toward the unique equilibrium in the negative (positive) saturation region, as desired in the EDGE as well as in the STORE M-CNN designs. The dashed brown curve without magenta diamonds in (a.1) ((b.1)) is the $v_{x_{i,j}} = v_{x_{i,j}}(x_{m_{i,j}})$ characteristic, expressed by **Eq. 35**, and constrained to lie in the region $v_{x_{i,j}} > (<) + (-)v_{sat}$, so as to keep the linear region free of $v_{x_{i,j}}$ nullclines. The dashed brown curve with magenta diamonds in (a.2) ((b.2)) represents the only locus of points of the phase plane, where $\dot{v}_{x_{i,j}} = 0\,\mathrm{V}\cdot\mathrm{s}^{-1}$, as expressed by **Eqs. 34**, **36**. Finally, the set of three dashed brown curves without magenta diamonds in (a.3) (b.3) constitute the possible courses of the $v_{x_{i,j}} = v_{x_{i,j}}(x_{m_{i,j}})$ characteristic, expressed by **Eqs. 34**, **36**, constrained to lie in the region $v_{x_{i,j}} < (>) + (-)v_{sat}$, so as to keep the positive (negative) region free of $\dot{v}_{x_{i,j}} = 0\,\mathrm{V}\cdot\mathrm{s}^{-1}$ loci, depending upon the sign of $\eta \triangleq i_{w_{i,j}} + a_{0,0}\cdot v_{sat}$ ($\xi \triangleq i_{w_{i,j}} - a_{0,0}\cdot v_{sat}$). Interestingly, the function $v_{x_{i,j}}(x_{m_{i,j}})$ of **Eqs. 34**, **36** is either concave down and monotone increasing if $\eta > 0\,\mathrm{A}$ ($\xi > 0\,\mathrm{A}$) or coinciding with the horizontal axis if $\eta = 0\,\mathrm{A}$ ($\xi = 0\,\mathrm{A}$), or even concave up and monotone decreasing if $\eta < 0\,\mathrm{A}$ ($\xi < 0\,\mathrm{A}$).

$v_{x_{i,j}} < (>) - (+)v_{sat}$, bend toward the east (west), evolving over time in the north or south direction from initial conditions lying below or above the $v_{x_{i,j}}$ nullcline **Eqs. 34**, **36**, respectively, going through the yellow (cyan) region II (III) or gray (green) region IV (I), as clearly indicated in **Figure 15** (a.2) ((b.2)). Manipulating the constraint, obtained by choosing the first (second) inequality sign in **Eq. 57**, it is simple to demonstrate that, depending upon its polarity, namely the sign of $\eta(v_{u_{i,j}}, n_B) \triangleq i_{w_{i,j}}(v_{u_{i,j}}, n_B) + a_{0,0}\cdot R_y\cdot g_{lin}\cdot v_{sat}$ ($\xi(v_{u_{i,j}}, n_B) \triangleq i_{w_{i,j}}(v_{u_{i,j}}, n_B) - a_{0,0}\cdot R_y\cdot g_{lin}\cdot v_{sat}$), the $v_{x_{i,j}}$–$x_{m_{i,j}}$ locus **Eqs. 34**, **36** may exhibit one of three possible graphs, as depicted in plot (a.3) ((b.3)) of **Figure 15**, lying, nevertheless, entirely below (above) the horizontal line $v_{x_{i,j}} = +(-)v_{sat}$, under all circumstances from rule 1 and in rule 2 (for each scenario of rule 3). Given that, with reference to the positive (negative) saturation region, the memristor state experiences a strictly monotonic decrease (increase) over time according to constraint **Eqs. 48**, **49**, while $\dot{v}_{x_{i,j}} > 0\,\mathrm{V}\cdot\mathrm{s}^{-1}$ provided condition **Eqs. 50**, **52** is satisfied, phase plane trajectories, visiting the domain $v_{x_{i,j}} > (<) + (-)v_{sat}$ in the SDP of any cell obeying rules 1 and 2 (rule 3), are expected to evolve in the south-west (north-east) direction, passing through the green (yellow) region I (II), as visualized in plot (a.3) ((b.3)) of **Figure 15**. Looking at the direction of motion of phase-plane trajectories across the phase plane (refer to plots (a.1)-(a.3) ((b.1)-(b.3)) of **Figure 15**), it is

evident that the unique equilibrium **Eqs. 39**, **45**, which the cell admits under the hypotheses of rules 1 and 2 (rule 3), is GAS, as indicated through the filling of its black circle marker in **Figure 15** (a.2) ((b.2)). In regard to rule 1 (3), taking into account the negative sign assumed for the common value b of the off-center synaptic weigths in the feedforward template, it may be easily realized that, under $n_B = 0$ (7), the $\dot{v}_{x_{i,j}} = 0\,\mathrm{V}\cdot\mathrm{s}^{-1}$ loci (34), (35), and (36) are closest to the horizontal lines $v_{x_{i,j}} = -v_{sat}$ (refer to plot (a.2) ((b.3))), $v_{x_{i,j}} = +\,(-)v_{sat}$ (refer to plot (a.1) ((b.1))), and $v_{x_{i,j}} = +v_{sat}$ (refer to plot (a.3) ((b.2))), respectively. It follows that, with regard to rule 1 (3), the constraint triplet, obtained from **Eqs. 56**, **57**, and **59** by choosing the first (second) inequality sign, should be evaluated only in the worst-case scenario, in which none (seven) of the eight neighbours of the M-CNN cell $C(i, j)$ features (feature) a positive one V-valued input voltage. Combining the resulting six conditions with the rule 2-based constraint triplet **Eqs. 56**, **57**, and **59** under the first inequality sign option, and with **Eq. 55**, provides a total number of 10 inequalities in the unknowns $a_{0,0}$ and $b_{0,0}$. Fixing the values for two cell core circuit parameters, namely z, and b, respectively set to $-1\cdot10^{-4}$ and to $-1\cdot10^{-4}\,\Omega^{-1}$, and assigning the value $1\cdot10^{-3}\,\Omega^{-1}$ to the conductance G_x of the linear resistor[37] in parallel to the capacitor in the memcomputing core of **Figure 8**, the two conditions, respectively descending from constraint **Eqs. 56**, **57** under the first inequality sign option for the worst-case scenario $n_B = 0$ from rule 1, and under the second inequality sign option for the worst-case scenario $n_B = 7$ from rule 3, are found to be identical one to the other, allowing to discard a couple of inequalities from the 10 aforementioned constraints. Manipulating the remaining 8 inequalities, it may be shown that only 3 of them are non-redundant, specifically constraint **Eq. 55**, and the pair of conditions, which respectively originate from the variant of **Eq. 57**, which is associated to the choice of the first inequality sign in the worst-case scenario $n_B = 0$ from rule 1 and in the only scenario $n_B = 8$ from rule 2. These 3 inequalities may be solved numerically, but, given their low number, a geometric approach is adopted here to derive suitable values for the self-feedforward and self-feedback synaptic weights. The magenta region in the $a_{0,0}$–$b_{0,0}$ parameter plane of **Figure 16** depicts the domain of admissible solutions of the non-redundant inequality triplet. Choosing the particular solution, which is indicated through an asterisk symbol, namely $(a^*_{0,0}, b^*_{0,0}) = (1.675\cdot10^{-3}\,\Omega^{-1}, 80.5\cdot10^{-5}\,\Omega^{-1})$, plots (a), (b), and (c) of **Figure 14** illustrate the SDPs, which foliate from the cell DRM_2 in the worst-case scenario $n_B = 0$ from rule 1, where $i_{w_{i,j}} = -0.105\,\mathrm{mA}$, in the sole scenario $n_B = 8$ admissible in rule 2, where $i_{w_{i,j}} = -0.095\,\mathrm{mA}$, and in the worst-case scenario $n_B = 7$ from rule 3, where $i_{w_{i,j}} = +0.105\,\mathrm{mA}$, respectively. Inspecting the cell SDP from plot (a), (b), and (c), the GAS equilibrium is found to be located at $(10\,\mathrm{k}\Omega, -0.2477\,\mathrm{V})$, at $(10\,\mathrm{k}\Omega, -0.2386\,\mathrm{V})$, and at $(2\,\mathrm{k}\Omega, +0.1817\,\mathrm{V})$, respectively. With reference to **Figure 16**, programming the cell core circuit parameters as indicated above, a M-CNN with $M = 64$ rows and $N = 60$ columns is capable to extract the edges of an input binary image, such as the one depicted in plot (b), providing them in the output binary image at steady-state, as in the example of plot (d), under null initial conditions for all the capacitor voltages, as shown in plot (c) [38], upon setting the resistance of each memristor to $5\,\mathrm{k}\Omega$ at the beginning of the simulation, and for fixed or Dirichlet boundary conditions (Chua and Roska, 2002), with each virtual cell input voltage value fixed to negative 1 V.

Remark 8. The insertion of a single memristor within the circuit implementation of the cell of a standard time- and space-invariant CNN allows to endow the resulting memristive array with novel functionalities, including the capability to read and write data locally within each processing element without the need to accommodate additional memory units, which are currently responsible for the poor spatial resolution of state-of-the-art CNN-UM hardware realizations.

5 M-CNN AS A MEMORY BANK: WRITE/READ FUNCTIONALITIES

The operating principles of a M-CNN programmed to write or read input binary data into or from the resistances of its memristors are elucidated below.

5.1 Store M-CNN

The aim of this section is to synthesize the gene for programming the cell $C(i, j)$ of a $M \times N$ M-CNN to store the negative (positive) one value, which is assigned to its input voltage $v_{u_{i,j}}$ on the basis of the white (black) color of the pixel in the corresponding location in a given input binary image with same spatial resolution as the cellular array, as off (on) resistive state x_{off} (x_{on}) in its memristor $\mathcal{M}_{x_{i,j}}$ at equilibrium[39], for all $i \in \{1, \ldots, M\}$, and for all $j \in \{1, \ldots, N\}$. Thus, as reported in **Table 5**, two are the local rules, which each M-CNN processing element is requested to comply with, so as to accomplish the data storage task.

Taking inspiration from the strategy adopted earlier on in the synthesis of a suitable gene for programming the bio-inspired memristive array to extract edges from an input binary image, a possible approach to design the STORE M-CNN is to make sure that the cell SDP accommodates one and only one globally asymptotically stable equilibrium i.e., $Q^{(-)}$ ($Q^{(+)}$), located in the position specified in **Eqs. 39**, **45** under $v_{u_{i,j}} = -(+)1\,\mathrm{V}$, as expected from rule 1 (2).

[37]The function of the linear resistor is to decrease the absolute value of the $v_{x_{i,j}}$ coordinate of the cell GAS equilibrium in each scenario from any of the three rules from **Table 1**. In turn this expedient would reduce the power, which the memristor device dissipates at equilibrium, allowing to extend its lifetime expectancy.

[38]As anticipated earlier, for each combination of indices $i \in \{1, \ldots, M\}$ and $j \in \{1, \ldots, N\}$, a real value within the set $(-1\mathrm{V}, 1\mathrm{V})$ $((-v_{sat}, +v_{sat}))$, associated to either the cell input voltage $v_{u_{i,j}}$ or the cell state 2 initial condition $v_{x_{i,j}}(0)$ (to the cell output voltage $v_{y_{i,j}}$), is mapped into a suitable tone on the grayscale for visualization purposes.

[39]Despite, theoretically, under the hypothesis of either rule, the two-dimensional state vector $(x_{m_{i,j}}, v_{x_{i,j}})$ converges toward the respective equilibrium $(\overline{x}_{m_{i,j}}, \overline{v}_{x_{i,j}})$ as t tends to ∞, in practice it is infinitesimally close to its final destination after a finite amount of time.

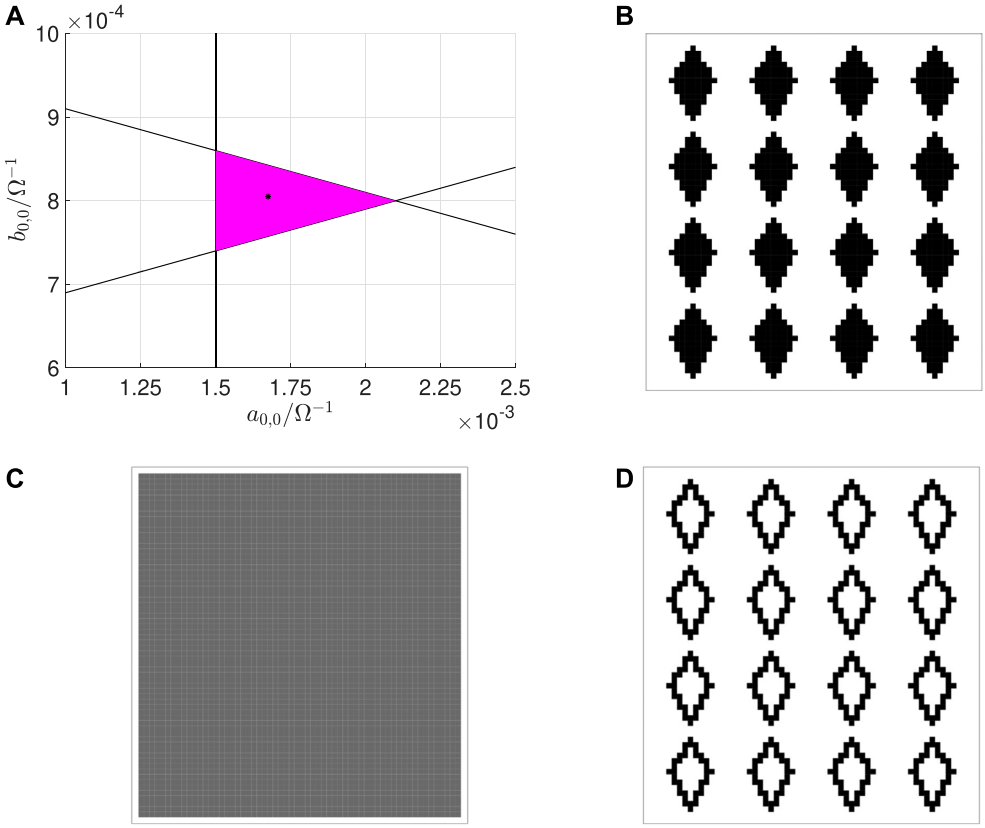


FIGURE 16 | (A) Geometry-based approach to the solution of the system of three non-redundant inequalities, resulting from the application of the systematic M-CNN design methodology from **section 3.3** for the accomplishment of the binary image edge extraction task. The magenta region contains the set of admissible solutions of the constraint triplet, which consist of inequality **Eq. 55**, and of the couple of conditions descending from the constraint pair **Eq. 57** under the first inequality sign option in the worst-case scenario $n_B = 0$ from rule 1, and in the only possible scenario $n_B = 8$ from rule 2, respectively. Recalling the invariable cell circuit parameter setting, reported in **Table 3**, and setting z, b, and G_x to $-1 \cdot 10^{-4}$, $-1 \cdot 10^{-4}\ \Omega^{-1}$, and $1 \cdot 10^{-3}\ \Omega^{-1}$, respectively, these three inequalities are in turn found to assume the analytical expressions $a_{0,0} > 1.5 \cdot 10^{-3}\ \Omega^{-1}$, $b_{0,0} > 1 \cdot 10^{-1} \cdot a_{0,0} + 5.9 \cdot 10^{-4}\ \Omega^{-1}$, and $b_{0,0} < -1 \cdot 10^{-1} \cdot a_{0,0} + 10.1 \cdot 10^{-4}\ \Omega^{-1}$. For the particular solution, highlighted by means of an asterisk marker, specifically $(a^*_{0,0}, b^*_{0,0}) = (1.675 \cdot 10^{-3}\ \Omega^{-1}, 80.5 \cdot 10^{-5}\ \Omega^{-1})$, and residing inside the magenta triangle at some safety distance from its sides, plots **(A–C)** of **Figure 13** show the SDPs foliating from the resulting EDGE M-CNN cell DRM_2 in the worst-case scenario $n_B = 0$ from rule 1, in the only possible scenario $n_B = 8$ from rule 2, and in the worst-case scenario $n_B = 7$ from rule 3, respectively. For the sake of completeness, referring to **Figure 14**, the specification of the aforementioned solution for the non-redundant IS defines the course of the graph of the function $v_{x_{i,j}}(x_{m_{i,j}})$, expressed by **Eqs. 34**, **36**, in plot (a.3) ((b.3)) for rules 1 and 2 (for rule 3), setting, to name but the most important cases, $\eta(v_{u_{i,j}}, n_B)$ ($\xi(v_{u_{i,j}}, n_B)$) to 62.5 μA for $n_B = 0$ and to −137.5 μA for $n_B = 1$ under rule 1, as well as to 72.5 μA under rule 2 (to −62.5 μA for $n_B = 7$ and to 137.5 μA for $n_B = 6$ under rule 3). **(B–D)** Proof of evidence for the proper functionality of a $M \times N$ M-CNN programmed through the gene synthesized in this section ($M = 64$, $N = 60$). Plot **(D)** depicts the steady-state output binary image of the EDGE M-CNN, once the black-and-white image, illustrated in plot **(B)**, is loaded to its input, a zero is assigned to the voltage falling across each capacitor at the beginning of the simulation, as visualized through the gray image in plot **(C)**, the initial condition on the resistance of each memristor is set to 5 kΩ , and a negative 1 V value is attributed to the input voltage of any virtual cell (Chua and Roska, 2002).

In order to facilitate the comprehension of the STORE M-CNN design, let us anticipate its outcome. Plot (a) ((b)) in **Figure 17** depicts the cell SDP synthesized through the proposed DRM_2-based system-theoretic technique to store the input voltage $v_{u_{i,j}} = -(+)1\,\mathrm{V}$ as off (on) resistance x_{off} (x_{on}) in the memristor $\mathcal{M}_{x_{i,j}}$ according to rule 1 (2). As graphically illustrated in the first (latter) plot, an ad-hoc IS shall be set in place to ensure that the cell SDP under negative (positive) one V-valued input voltage accommodates only the $\dot{v}_{x_{i,j}} = 0\,\mathrm{V} \cdot \mathrm{s}^{-1}$ locus **Eqs. 34**, **36**, indicated via magenta diamonds and located in the negative (positive) saturation region, and that the $v_{m_{i,j}}$ nullcline intersects the $\dot{x}_{m_{i,j}} = 0\,\Omega \cdot \mathrm{s}^{-1}$ locus **Eqs. 31**, **33**, identified via red crosses in the phase-plane lower (upper) half, in the equilibrium **Eqs. 39**, **45**, where, as marked by means of a black circle, the memristor stores the off (on) resistive state x_{off} (x_{on}) expected from rule 1 (2). Shaping the cell DRM_2 this way, any trajectory of either SDP, visiting the region below (above) the single $v_{x_{i,j}}$ nullcline, would move upward (downward), bending toward the east/west in the phase-plane lower/upper half, toward the unique equilibrium, which, as a result, would feature global asymptotic stability, as highlighted through the filling of the respective black circle marker in each of plots (a) and (b) of **Figure 17**. Given that the gene synthesis strategy, adopted here for programming the bio-inspired memristive array to write binary data into its memristors, is analogous to the one considered previously in the EDGE M-CNN design, **Figure 15** (a.1), (a.2),

and (a.3) ((b.1), (b.2), and (b.3)) are used once more to illustrate graphically the way we wish to massage the cell SDP in the linear, negative (positive), and positive (negative) saturation regions, respectively, for the satisfaction of rule 1 (2) from **Table 5**. The stepwise mathematical analysis of the second-order ODE **Eqs. 18**, **19**, following shortly, shall enable to combine the graphs in plots (a.1)-(a.3) ((b.1)-(b.3)) for the synthesis of the desired cell SDP under $v_{u_{i,j}} = -(+)1$ V. Before commencing the investigations, it is worth stressing a couple of points. Firstly, as a result of our strategy to enforce the existence of one and only one globally asymptotically stable equilibrium in each of the two possible cell SDPs, the choice for the initial condition $(x_{m_{i,j}}(0), v_{x_{i,j}}(0))$ of the ODEs **Eqs. 18**, **19** is arbitrary. Secondly, since an isolated[40] non-autonomous array is expected to suffice for the accomplishment of the binary data writing task, the expression for the offset current in **Eq. 13** reduces to

$$i_{w_{i,j}}\left(v_{u_{i,j}}\right) = I \cdot z + b_{0,0} \cdot v_{u_{i,j}} \tag{60}$$

Under the hypothesis that a value is preliminarily assigned to the self-feedback synaptic weight $a_{0,0}$ the proposed DRM_2 system-theoretic method will aim to the determination and later solution of an ad-hoc IS in the pair of unknown parameters $b_{0,0}$ and z. The proposed gene synthesis technique is first applied to derive the necessary constraints for well-behaved phase-plane trajectories in the linear region of each cell SDP.

5.1.1 Store M-CNN Cell DRM_2 Synthesis in the Linear Region

The purpose of the following mathematical derivations is to make sure that, under negative (positive) one V-valued cell input voltage, the locus of points, lying on the characteristic $v_{x_{i,j}} = v_{x_{i,j}}(x_{m_{i,j}})$, expressed by **Eq. 35**, assumes values in the positive (negative) saturation region, as shown through a dashed brown curve without magenta diamonds in **Figure 15** (a.1) (b.1). With reference to the right hand side of **Eq. 35**, a strictly positive sign is imposed on the denominator of the rational function for all $x_{m_{i,j}} \in \mathcal{D}$ under the constraint established by inequality **Eq. 55**. Under this hypothesis, enforcing a negative (positive) polarity for the offset current under $v_{u_{i,j}} = -(+)1$ V through the inequality

$$i_{w_{i,j}}\left(v_{u_{i,j}}\right) < (>)\, 0 \text{ A}, \tag{61}$$

the graph of the function $v_{x_{i,j}}(x_{m_{i,j}})$, described by **Eq. 35**, is found to lie on the phase plane upper (lower) half, and to feature a monotonic decrease (increase) with $x_{m_{i,j}}$ with upward (downward) concavity. In the first (latter) case this $v_{x_{i,j}}$–$x_{m_{i,j}}$ locus may thus be forced to fall completely over the domain $v_{x_{i,j}} > (<) + (-)v_{sat}$, as visualized in plot (a.1) ((b.1)) of **Figure 15**, via the additional constraint

$$\frac{-i_{w_{i,j}}\left(v_{u_{i,j}}\right)}{a_{0,0} \cdot R_y \cdot g_{lin} - G_x - x_{off}^{-1}} > (<) + (-)\, v_{sat} \tag{62}$$

[40]A standard space-invariant uncoupled CNN, in which each cell features a 3×3 local neighbourhood, is said to be *isolated* (Chua and Roska, 2002) if each feedforward synaptic weight $b_{k,l} - k, l \in \{-1, 0, 1\}$ – except for $b_{0,0}$, is null.

TABLE 5 | Pair of local rules, which the M-CNN cell $C(i,j)$ is requested to obey, so as to map the white (black) pixel in the corresponding position of a given $M \times N$ input binary image into the off (on) resistive state of its memristor $\mathcal{M}_{x_{i,j}}$ at equilibrium ($i \in \{1, \ldots, M\}$, $j \in \{1, \ldots, N\}$).

Local rule	$\mathbf{v_{u_{i,j}}/V}$	$\bar{\mathbf{x}}_{\mathbf{m_{i,j}}}$
1	−1	x_{off}
2	+1	x_{on}

Under rule 1 (2), the whole phase-plane region $\left|v_{x_{i,j}}\right| \leq v_{sat}$ lies below (above) the characteristic $v_{x_{i,j}} = v_{x_{i,j}}(x_{m_{i,j}})$ of **Eq. 35**. As a result, on the basis of condition **Eq. 58**, descending from inequality **Eq. 51** by taking into account that the second factor on the left hand side is strictly positive in view of constraint **Eq. 55**, the trajectory, which each point $(x_{m_{i,j}}, v_{x_{i,j}})$ traces in the linear region, evolves in the south (north) direction, bending eastward or westward in the phase-plane lower or upper half over time, as dictated by constraint **Eq. 48** or **Eq. 49**, visiting the gray IV (yellow II) or green I (cyan III) regions, as shown in plot (a.1) ((b.1)) of **Figure 15**. Next, our systematic M-CNN design methodology is applied to massage the STORE cell DRM_2 in the phase-plane saturation regions.

5.1.2 Store M-CNN Cell DRM_2 Synthesis in the Saturation Regions

Two are the aims of the mathematical treatment to follow. Firstly, referring to **Figure 15** (a.2) ((b.2)), we shall make sure that, under $v_{u_{i,j}} = -(+)1$ V, the characteristic $v_{x_{i,j}} = v_{x_{i,j}}(x_{m_{i,j}})$ of **Eqs. 34**, **36** lie in the negative (positive) saturation region, as indicated through a dashed brown curve with magenta diamonds, and intersect the vertical $x_{m_{i,j}}$ nullcline **Eqs. 31**, **33**, which crosses the horizontal axis at the memristor state upper (lower) bound, in a GAS equilibrium point, namely **Eqs. 39**, **45**, as marked through a black-filled circle. Secondly, looking now at **Figure 15** (a.3) ((b.3)), in order to enforce that the processing element is monostable under the hypothesis of either rule i.e., that the function $v_{x_{i,j}}(x_{m_{i,j}})$ of **Eqs. 34**, **36** denote the only possible $\dot{v}_{x_{i,j}} = 0\,\text{V} \cdot \text{s}^{-1}$ locus, which the cell SDP may ever accommodate under rule 1 (2), we shall impose that the characteristic $v_{x_{i,j}} = v_{x_{i,j}}(x_{m_{i,j}})$ of **Eqs. 34**, **36** does not go through any phase-plane point belonging to the positive (negative) saturation region, featuring, in particular, one of three possible graphs, as visualized by means of dashed brown curves without magenta diamonds. It may be shown that, due to inequality **Eq. 61**, under the hypothesis of rule 1 (rule 2), the $\dot{v}_{x_{i,j}} = 0\,\text{V} \cdot \text{s}^{-1}$ locus **Eqs. 34**, **36** exhibits upward (downward) concavity in its monotonic decrease (increase) with the memristor state. Thus, making sure it lies below (above) the horizontal line $v_{x_{i,j}} = -(+)v_{sat}$ at the memristor state lower bound, via the additional inequality

$$\frac{i_{w_{i,j}}\left(v_{u_{i,j}}\right) - (+)a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat}}{G_x + x_{on}^{-1}} < (>) - (+)\, v_{sat}, \tag{63}$$

under $v_{u_{i,j}} = -(+)1$ V, this unique $v_{x_{i,j}}$ nullcline is found assume values within the negative (positive) saturation region over the entire memristor state existence domain $\mathcal{D}$, as depicted by means of a dashed brown curve with magenta diamonds in **Figure 15** (a.2) ((b.2)). The existence of a cell equilibrium, as specified in **Eqs. 39**, **45**,

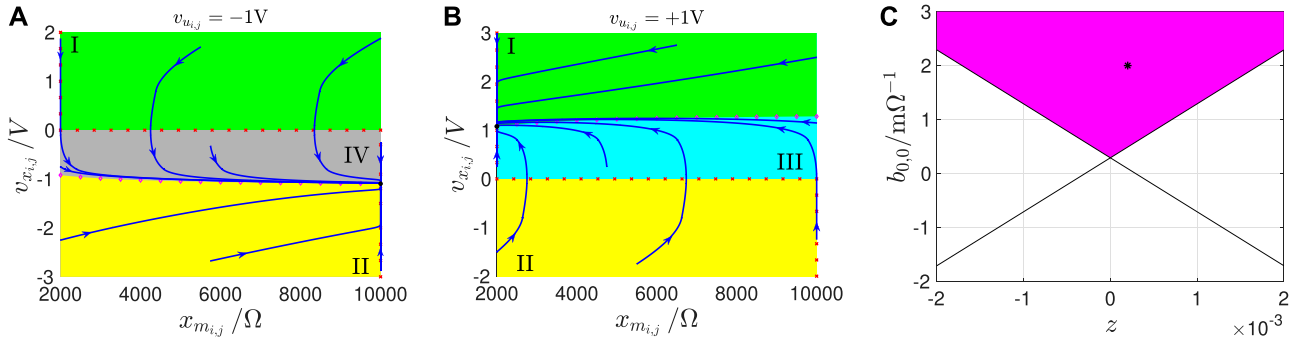


FIGURE 17 | (A,B) SDP of the processing element $C(i,j)$ for the storage of the negative (positive) one value, assigned to its input voltage $v_{u_{ij}}$ on the basis of the white (black) pixel at row i and column j of a given input binary image, in the form of off (on) resistance x_{off} (x_{on}) into its memristor $\mathcal{M}_{x_{ij}}$. In the first (latter) case the cell is monostable, featuring a single equilibrium, attracting all phase-plane trajectories, and lying in the negative (positive) saturation region, along the vertical line crossing the horizontal $x_{m_{ij}}$–axis at the upper (lower) bound of the memristor state existence domain D, as expected from rule 1 (2) of **Table 5**. **(C)** Graphical determination of the solutions of the non-redundant inequality pair **Eq. 62**, descending from the application of the proposed DRM$_2$-centered system-theoretic bio-inspired array design method to the model **Eqs. 18**, **19** of the processing element $C(i,j)$, in each region of the standard nonlinearity (3), as discussed in detail in the text, and illustrated graphically in **Figure 14**, for the synthesis of an ad-hoc family of STORE M-CNN cell SDPs. The magenta region constitutes the domain of the two-dimensional z–$b_{0,0}$ parameter space, which hosts all the admissible solutions of the system of inequalities **62**, which, taking into account **Table 3**, and assigning the values of $5 \cdot 10^{-3}\,\Omega^{-1}$, and of $20 \cdot 10^{-4}\,\Omega^{-1}$, to $a_{0,0}$, and G_x, respectively, feature the analytical expressions $b_{0,0} > (2.9 \cdot 10^{-4} + (-)z)\,\Omega^{-1}$, with the positive (negative) polarity option defining a rule 1 (2)-based constraint. The particular cell DRM$_2$, foliating in the SDPs of plots **(A)** and **(B)**, associated to a nonzero offset current $i_{w_{ij}}$ of value −1.8 and +2.2 mA, and admitting a GAS equilibrium $(\bar{x}_{m_{ij}}, \bar{v}_{x_{ij}})$, sitting at (10 kΩ, −1.10 V), and at (2 kΩ, 1.08 V), respectively, was derived for the specific solution pair $(z^*, b^*_{0,0}) = (2 \cdot 10^{-4}, 2 \cdot 10^{-3}\,\Omega^{-1})$, indicated through an asterisk in plot **(C)**, and located well within the magenta region, conveniently away from its boundaries. For the sake of completeness, with reference to **Figure 14** (a.3) ((b.3)), the final solution of the non-redundant inequality system defines the shape of the characteristic $v_{x_{ij}} = v_{x_{ij}}(x_{m_{ij}})$ of **Eqs. 34**, **36**, since it sets $\eta(v_{u_{ij}})$ ($\xi(v_{u_{ij}})$) to −2.3 mA (+2.7 mA) for rule 1 (2).

and indicated via a black circle in plot (a.2) ((b.2)) of **Figure 15**, is then guaranteed under the hypothesis of rule 1 (2). With regard to the flow of the vector field $(\dot{x}_{m_{ij}}, \dot{v}_{x_{ij}})$ over time, all trajectories, visiting points lying below or above the $v_{x_{ij}}$ nullcline **Eqs. 34**, **36** in the phase-plane negative (positive) saturation region under $v_{u_{ij}} = -(+)1$ V, feature a northeastward (northwestward) or southeastward (southwestward) direction of motion, in view of inequalities **Eqs. 50**, **52** and **Eqs. 48**, **49**, crossing the yellow II (cyan III) or gray IV (green I) regions, defined in the legend of **Figure 11**, as illustrated in **Figure 15** (a.2) ((b.2)).

Basic mathematical analysis reveals that, under the hypothesis of rule 1 (2), the function $v_{x_{ij}}(x_{m_{ij}})$, expressed by **Eqs. 34**, **36**, may exhibit three possible distinct courses, depending upon the polarity of its numerator, i.e. upon the sign of $\eta(v_{u_{ij}}) \triangleq i_{w_{ij}}(v_{u_{ij}}) + a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat}$ $(\xi(v_{u_{ij}}) \triangleq i_{w_{ij}}(v_{u_{ij}}) - a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat})$, but, irrespectively, keeps always below (above) the horizontal line $v_{x_{ij}} = +(-)v_{sat}$, as sketched by means of dashed brown curves without magenta diamonds in **Figure 15** (a.3) ((b.3)). It descends that all the points, residing in the positive (negative) saturation region of the cell SDP for $v_{u_{ij}} = -1\,(+1)$ V, lie above (below) the characteristic $v_{x_{ij}} = v_{x_{ij}}(x_{m_{ij}})$ of **Eqs. 34**, **36**. Therefore, on the basis of conditions **Eqs. 48**, **49** and **Eqs. 50**, **52**, a trajectory point $(x_{m_{ij}}, v_{x_{ij}})$ evolves over time in the south-west (north-east) direction, as it passes across the phase-plane positive (negative) saturation region, exploring the green (yellow) region I (II) of the two-dimensional state-space, as graphically shown in plot (a.3) ((b.3)) of **Figure 15**. The global asymptotic stability of the only equilibrium **Eqs. 39**, **45**, which the cell SDP hosts under a negative (positive) one V-valued input voltage, explaining the filling of the respective black circle in plot (a.2) ((b.2)) of **Figure 15**, may be inferred by inspecting the flow of the vector field $(\dot{x}_{m_{ij}}, \dot{v}_{x_{ij}})$ throughout the phase-plane (refer to **Figure 15** (a.1)-(a.3) ((b.1)-(b.3))). Overall, our rigorous system-theoretic M-CNN design methodology has identified a set of 7 constraints, including condition **Eq. 55**, and a trio of inequality couples, namely **Eqs. 61**–**63**, where the first (second) sign option applies under the hypothesis of rule 1 (2). Setting the values for $a_{0,0}$ and for G_x to[41] $5 \cdot 10^{-3}\,\Omega^{-1}$, and $20 \cdot 10^{-4}\,\Omega^{-1}$, respectively, the first condition **Eq. 55** holds automatically true. Further, the two **Eq. 62** already account for all the remaining four conditions, expressed by the inequality pairs **Eqs. 61**, **63**. Adopting a geometric approach to solve the two non-redundant inequalities **Eq. 62**, one for each of the two possible sign choices, in the z–$b_{0,0}$ parameter plane, the coordinates of all the points, residing within the magenta region of **Figure 17C**, enable to program the M-CNN so as to accomplish the data writing task. With reference to this same figure, as anticipated earlier, plot (a) ((b)) visualizes the SDP, which the M-CNN cell features under $v_{u_{ij}} = -(+)1$ V, upon the assignment of $z^* = 2 \cdot 10^{-4}$, and $b^*_{0,0} = 2 \cdot 10^{-3}\Omega^{-1}$ to z, and $b_{0,0}$, respectively, as it descends from the selection of the non-redundant inequality pair solution, sitting at a precautionary distance from the boundaries of the magenta region, and indicated through an asterisk marker in plot (c).

[41]Similarly as in the EDGE M-CNN design, the addition of a linear resistor in parallel to the capacitor within the memcomputing core of the processing element circuit of **Figure 8** allows to keep within reasonable limits the modulus of the voltage, falling across the resistance switching memory at equilibrium, for each of the two possible cell input voltage values. This preventive measure is of particular importance in view of a future hardware realization of the bio-inspired memristive array under study.

Setting the values for the core circuit parameters of each cell of a M-CNN, featuring $M = 145$ rows and $N = 147$ columns, as established through the gene synthesis procedure, numerical simulations reveal the capability of the resulting bio-inspired memristive array to store binary data into its locally distributed memristive memory bank. A white (black) pixel at row i and column j of a binary image, featuring the same spatial resolution as the M-CNN, and shown in **Figure 18A**, is first mapped onto a negative (positive) one V-valued input voltage $v_{u_{i,j}}$ for the M-CNN cell $C(i,j)$ $(i \in \{1,\ldots,M\},\ j \in \{1,\ldots,N\})$. Letting a white (black) pixel code the lowest (highest) possible resistive level for the initial condition on state 1, as well as a negative (positive) 1 V voltage for the initial condition on state 2, $x_{m_{i,j}}$ and $v_{x_{i,j}}$ are randomly initialized to one of two possible values from the sets $\{x_{on}, x_{off}\}$, and $\{-1,+1\}$ V, respectively, as graphically illustrated through binary images in plots (b) and (c) of **Figure 18**, respectively. Plot (d) from the same figure visualizes the data written into the memristors at equilibrium through the use of a white (black)-coloured pixel in each location corresponding to a M-CNN cell, which stores the on (off) memristance level x_{on} (x_{off}). Given the scheme adopted for choosing the input voltage of each cell, and for visualizing the resistive state of each memristor at equilibrium, it follows that a white (black) pixel in correspondence of the i^{th} row and j^{th} column of the input image from plot (a) is mapped onto a black (white) pixel in the corresponding location of the illustrative picture from plot (d). The complementary operation to information storage is data retrieval. The next section elucidates the principles behind the choice of a suitable gene for the execution of this task.

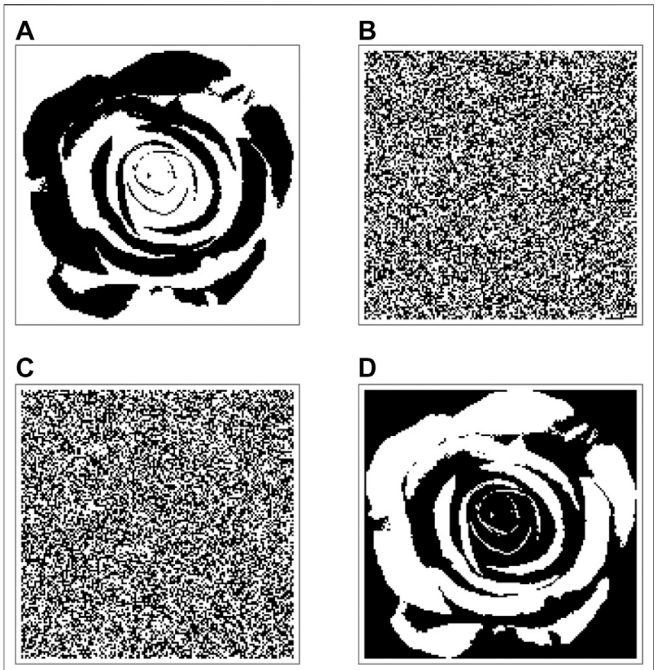


FIGURE 18 | (A) Input binary image with $M \times N$ pixels ($M = 145$, $N = 147$). **(B,C)** Black-and-white picture coding the random initial condition assigned to the capacitor voltage (memristor state) in each cell of a M-CNN with same rows and columns as the input image. **(D)** Graphical illustration of the data stored in the memristive memory at the end of the data writing operation. The binary picture in **(D)** appears to be the logically inverted version of the black-and-white image in **(A)**, due to the convention, intentionally adopted here, to map each white (black) pixel of the input binary image to a negative (positive) one V-valued input voltage for the corresponding M-CNN cell, and to code the memory content of each off (on) memristor at equilibrium through a black (white) pixel.

5.2 Recall M-CNN

The purpose of this section is to shape the DRM_2 of the processing element $C(i,j)$ of a $M \times N$ M-CNN, so as to allow the cell itself to retrieve the initial resistive state of the memristor $\mathcal{M}_{x_{i,j}}$, mapping the memory content into the steady-state[42] output voltage $v_{y_{i,j}}(t_{i,j}^{(s)})$ as negative (positive) saturation level $-(+)v_{sat}$, in case $x_{m_{i,j},0} \triangleq x_{m_{i,j}}(0\,\text{s})$ is found to be the upper (lower) bound x_{off} (x_{on}) of the closed set $\mathcal{D}$. The local rule pair, dictating the operating principles of each RECALL M-CNN cell, is reported in **Table 6**.

Given that the RECALL M-CNN is autonomous, differently from the strategy adopted in the STORE M-CNN design, the gene synthesis approach, followed in this section, aims to massage one and only one SDP, hosting the solutions of the second-order ODE **Eqs. 18**, **19** for both local rules from **Table 6**, and constituting, as a result, the DRM_2 itself. In order to achieve this purpose, the cell should be programmed so as to operate as a bistable dynamical system: in case the memristor, it accommodates, sits in the off (on) resistive state, the vector field flow should guide the trajectory toward the equilibrium $Q^{(-)}$ $(Q^{(+)})$, which features coordinates specified in **Eqs. 39**, **45**, and is indicated via a black circle in the phase-plane negative (positive) saturation region from **Figure 19A**, anticipating the bistable cell SDP, which will be synthesized shortly by means of the proposed approach. The existence of the first (latter) equilibrium is ensured by enforcing the existence of a $v_{x_{i,j}}$ nullcline, which is defined in **Eqs. 34**, **36**, in the negative (positive) saturation region, as indicated via magenta diamonds, and ensuring it would admit a point of intersection with the $\dot{x}_{m_{i,j}} = 0\,\Omega \cdot \text{s}^{-1}$ locus, which is marked with red crosses, and expressed by **Eqs. 31**, **33**. A fundamental step in the RECALL M-CNN design regards the selection of a suitable initial condition $v_{x_{i,j},0}$ for the capacitor voltage. It should be based upon the necessity to ensure that, with the memristor $\mathcal{M}_{x_{i,j}}$ storing the off (on) resistance x_{off} (x_{on}), the initial condition $(x_{m_{i,j}}(0), v_{x_{i,j}}(0))$ of the cell ODE **Eqs. 18**, **19** should belong to the basin of attraction of the equilibrium **Eqs. 39**, **45**.

With reference to the cell SDP in **Figure 19A**, in our strategy we first imposed the existence of a $\dot{v}_{x_{i,j}} = 0\,\text{V} \cdot \text{s}^{-1}$ locus, which is expressed by **Eq. 35**, also within the phase-plane domain

[42]The task of the RECALL M-CNN may be considered accomplished as soon as the outputs of all its processing elements attain their final values. This occurs at the steady-state time instant $t^{(s)} \triangleq \max_{1 \le i \le M, 1 \le j \le N}\{t_{i,j}^{(s)}\}$, where $v_{y_{i,j}}(t) = -(+)v_{sat}$ for all $t \ge t_{i,j}^{(s)}$, with $t_{i,j}^{(s)}$ denoting the time instant, at which the phase-plane trajectory point $(x_{m_{i,j}}, v_{x_{i,j}})$, evolving in time according to the second-order ODE (18)–(19), which models the dynamics of the cell $C(i,j)$, enters the SDP negative (positive) saturation region, hosting the equilibrium, it is asymptotically converging to, in case the memristor $\mathcal{M}_{x_{i,j}}$ initially sits in the highest (lowest) possible resistive state x_{off} (x_{on}).

$\left|v_{x_{i,j}}\right| \leq v_{sat}$, as indicated via magenta diamonds, and then ensured it would cross the $x_{m_{i,j}}$ nullcline, marked through red crosses, and defined in **Eq. 31**, forming, as a result, the additional equilibrium $Q^{(0)}$, with coordinates reported in **Eq. 41**, symbolized through the black circle, and located on the linear region negative side. Furthermore, we enforced that the $v_{x_{i,j}} = v_{x_{i,j}}(x_{m_{i,j}})$ characteristics **Eqs. 34**, **35** would assume values only over a limited range of the close set $\mathcal{D}$, namely $x_{m_{i,j}} \in [\tilde{x}_{m_{i,j}}, x_{off}]$ within the regions $v_{x_{i,j}} < -v_{sat}$ and $\left|v_{x_{i,j}}\right| \leq v_{sat}$, respectively, with $\tilde{x}_{m_{i,j}}$ representing the abscissa of their point of intersection, residing on the frontier between negative saturation and linear regions, away, to some extent, from the vertical line $x_{m_{i,j}} = x_{on}$. This design plan was instrumental for the creation of a special domain, lying within the region $v_{x_{i,j}} < 0\,\mathrm{V}$, and accommodating trajectories moving in the south-east direction, whereas the vector field flows toward the north-east across the remainder of the phase plane lower half. Besides revealing the unstable nature of the equilibrium **Eq. 41**, as indicated by the hollow structure of its black circle marker, the formation of this special domain ensures that, setting the initial condition $v_{x_{i,j}}(0)$ on the capacitor voltage to an intermediate value between the ordinates of the two equilibria, lying along the vertical line $x_{m_{i,j}} = x_{off}$, the phase-plane trajectory, which would emerge on the cell SDP, in case the memristor initially sits in the highest (lowest) possible resistive state, would asymptotically approach the equilibrium located in the negative (positive) saturation region, revealing its locally stable nature, as highlighted through the filling of the relative black circle symbol. The steps, to be mathematically formulated below, which our system-theoretic methodology entails, for shaping the cell DRM_2 in the linear, negative saturation, and positive saturation region, as desired (refer, once again to **Figure 19A**), are visualized through illustrative viewgraphs in **Figures 20A–C**, respectively.

A rigorous mathematical analysis of the cell ODE **Eqs. 18**, **19** allows the derivation of a suitable IS for the creation of an ad-hoc cell SDP, combining the coloured phase-plane regions in plots (a), (b), and (c) of **Figure 20**. A preliminary requirement for initiating the investigations is to fix the expression for the offset current. Conjecturing that the use of a simple isolated autonomous M-CNN would be sufficient for the accomplishment of the data recall operation, simplifying **Eq. 13**, the following formula may be assigned to $i_{w_{i,j}}$:

$$i_{w_{i,j}} = I \cdot z. \tag{64}$$

Assuming that the conductance G_x of the linear resistor, appearing in parallel with the capacitor in the cell circuit memcomputing core of **Figure 8**, is a given design parameter, the IS under determination will be expressed in terms of two unknowns only, specifically $a_{0,0}$, and z, which will enable the determination of the domain of admissible solutions on the basis of a geometrical analysis. Let us commence the systematic mathematical treatment from the linear region of the standard nonlinearity of **Eq. 3**.

5.2.1 Recall M-CNN Cell DRM_2 Synthesis in the Linear Region

Looking at **Figure 20**, the purpose of this section is to make sure that $\dot{v}_{x_{i,j}} = 0\,\mathrm{V} \cdot \mathrm{s}^{-1}$ on the locus of points, expressed by **Eq. 35**, and indicated via a dashed brown curve with magenta diamonds in plot (a), that such $v_{x_{i,j}} = v_{x_{i,j}}(x_{m_{i,j}})$ characteristic lies over the phase plane region $v_{x_{i,j}} \in [-v_{sat}, 0\,\mathrm{V})$, forming together with the $x_{m_{i,j}}$ nullcline **Eq. 31** an unstable equilibrium in the point, defined in **Eq. 31**, and marked through a black hollow circle in plot (a), intersecting the frontier between negative saturation and linear regions in a point, specifically $(\tilde{x}_{m_{i,j}}, -v_{sat})$, which, as may be easily verified through maths (refer to plot (b) as well), belongs also to the graph of the function $v_{x_{i,j}}(x_{m_{i,j}})$ of **Eq. 34**, residing at some distance from the vertical line $x_{m_{i,j}} = x_{on}$.

TABLE 6 | Set of local rules, which are imposed on the processing element $C(i,j)$ of a $M \times N$ M-CNN, so as to allow the reading of the memory content, initially stored in the memristor $\mathcal{M}_{x_{i,j}}$, and its transfer to the steady-state output voltage $v_{y_{i,j}}(t_{i,j}^{(S)})$ ($i \in /1, \ldots, M/$, $j \in /1, \ldots, N/$). The initial condition of the memristor state $x_{m_{i,j}}(0\,\mathrm{s})$ is requested to have a crucial impact on the dynamic behaviour of the capacitor voltage $v_{x_{i,j}}$: if $\mathcal{M}_{x_{i,j}}$ initially sits in the off (on) resistive state x_{off} (x_{on}), $v_{x_{i,j}}$ is expected to converge asymptotically toward an equilibrium value lower (higher) than the negative (positive) saturation level, fixing, consequently, $v_{y_{i,j}}(t_{i,j}^{(S)})$ to $-(+)v_{sat}$.

Local rule	$\mathbf{x_{m_{i,j}}(0\,s)/\,k\Omega}$	$\mathbf{v_{y_{i,j}}(t_{i,j}^{(s)})}$
1	x_{off}	$-v_{sat}$
2	x_{on}	$+v_{sat}$

Enforcing the inequality **Eq. 55**, and assuming a positive polarity for the offset current according to

$$i_{w_{i,j}} > 0\,\mathrm{A}, \tag{65}$$

the function $v_{x_{i,j}}(x_{m_{i,j}})$, expressed by **Eq. 35**, is found to be strictly negative, and to feature downward concavity as it increases monotonically with the memristor state. It follows that, through the additional condition

$$\frac{-i_{w_{i,j}}}{a_{0,0} \cdot R_y \cdot g_{lin} - G_x - x_{off}^{-1}} > -v_{sat}, \tag{66}$$

the characteristic $v_{x_{i,j}} = v_{x_{i,j}}(x_{m_{i,j}})$ of **Eq. 35** falls within the domain $v_{x_{i,j}} \in [-v_{sat}, 0\,\mathrm{V})$, as illustrated via a dashed brown curve with magenta diamonds in **Figure 20A**, and crosses the $x_{m_{i,j}}$ nullcline **Eq. 31** in the equilibrium point $Q^{(0)}$, defined in **Eq. 41**, and depicted as a black circle in the same figure. Furthermore, the constraint

$$\tilde{x}_{m_{i,j}} \triangleq \frac{v_{sat}}{\left(a_{0,0} \cdot R_y \cdot g_{lin} - G_x\right) \cdot v_{sat} - i_{w_{i,j}}} > x_{on}, \tag{67}$$

establishes the requirement for the point of intersection between the $v_{x_{i,j}}$ nullcline of **Eq. 35** and the horizontal line $v_{x_{i,j}} = -v_{sat}$ to lie, at least to some extent, away from the $x_{m_{i,j}} = x_{on}$ locus.

Given that, with inequality **Eq. 55** holding true, **Eq. 58** expresses the condition under which $\dot{v}_{x_{i,j}} > 0\,\mathrm{V \cdot s^{-1}}$ in the linear region, the phase plane trajectories, lying, therein, below (above) the $v_{x_{i,j}}$ nullcline of **Eq. 35**, evolve over time in the south (north) direction, bending to the east or to the west, as dictated by constraint **Eq. 48** or **Eq. 49**, across the domain $v_{x_{i,j}} = [-v_{sat}, 0)$ or $v_{x_{i,j}} = (0\,v_{sat}]$, exploring the gray IV (yellow II or cyan III) region(s), as illustrated in plot (a) of **Figure 20**, unveiling the unstable nature of the equilibrium point (41), which is then visualized as a black hollow circle. The analytical treatment of the

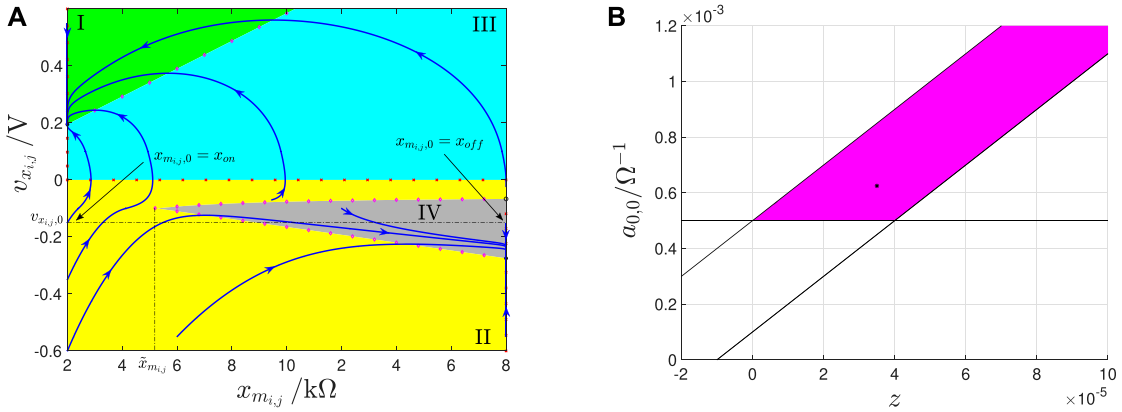


FIGURE 19 | (A) SDP of the bistable cell $C(i,j)$ of the isolated and autonomous RECALL M-CNN. Choosing a suitable value for the initial condition $v_{x_{ij},0}$ on the capacitor voltage, here –0.15 V, if the memristor sits in the highest (lowest) possible resistive state x_{off} (x_{on}) at the onset of the data recall procedure, the state vector $(x_{m_{ij}}, v_{x_{ij}})$ evolves in time toward the equilibrium (39) ((45)) located in the phase-plane region $v_{x_{ij}} < (>) -(+) v_{sat}$, as expected from rule 1 (2) from **Table 6**. Importantly, since the memristor state $\overline{x}_{m_{ij}}$ at equilibrium is found to be identical as its initial condition $x_{m_{ij},0}$, the data stored in the locally distributed memristive memory bank are unaltered by the RECALL operation. Importantly, the values of the self-feedback synaptic weight $a_{0,0}$ and of the offset current $i_{w_{ij}}$ in the SDP identify a point located in the pink domain of the M-CNN Primary Mosaic of **Figure 13**. **(B)** Graphical illustration of the geometric analysis, carried out in the parameter plane z–$a_{0,0}$, for the determination of valid solutions of the non-redundant inequality trio, composed of conditions **Eqs. 55**, **66**, and **67**. These three inequalities, obtained through the system-theoretic methodology, proposed in **section 3.3**, allow to massage the DRM_2 of each M-CNN processing element in such a way to retrieve the information stored in the memristor $\mathcal{M}_{x_{ij}}$. On the basis of **Table 3**, and setting G_x to $0\,\Omega^{-1}$, they are found to feature formulas $a_{0,0} > 5 \cdot 10^{-4}\,\Omega^{-1}$, $a_{0,0} > (10 \cdot z + 1 \cdot 10^{-4})\,\Omega^{-1}$, and $a_{0,0} < (10 \cdot z + 5 \cdot 10^{-4})\,\Omega^{-1}$, respectively. The coordinates of each point $(z, a_{0,0})$ within the magenta domain satisfy them concurrently. The cell SDP, depicted in plot **(A)**, was derived for the particular solution $(z^*, a_{0,0}^*) = (3.5 \cdot 10^{-5}, 6.25 \cdot 10^{-4}\,\Omega^{-1})$, which, as indicated via an asterisk marker in plot **(B)**, resides, to some extent, away from the white parameter plane region, where at least one of the three conditions does not hold true.

second-order cell ODE (18)–(19) is now focused on the saturation regions of the standard nonlinearity of **Eq. 3**.

5.2.2 Recall M-CNN Cell DRM_2 Synthesis in the Saturation Regions

Referring to Figure 20, the intention of this section is to establish the existence of a $\dot{v}_{x_{ij}} = 0\,\mathrm{V{\cdot}s^{-1}}$ locus in the negative (positive) saturation region, as expressed by **Eqs. 34**, **36**, and highlighted via a dashed brown curve with magenta diamonds in plot (b) ((c)), and to ensure that it forms, together with the $\dot{x}_{m_{ij}} = 0\,\Omega{\cdot}\mathrm{s^{-1}}$ locus, defined in **Eqs. 31**, **33**, a stable equilibrium point, as given in **Eqs. 39**, **45**, and shown as a black filled circle in plot (b) ((c)).

As anticipated earlier, mathematical calculations reveal that the $v_{x_{ij}} = v_{x_{ij}}(x_{m_{ij}})$ characteristic of **Eq. 34** intersects the frontier between the negative saturation and linear regions in the point of abscissa $\tilde{x}_{m_{ij}}$, defined on the left hand side of inequality **Eq. 67**. Therefore, with inequality **Eq. 65** holding true, taking into account that, on the basis of condition **Eq. 67**, $i_{w_{ij}} < a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat}$, the $v_{x_{ij}} = v_{x_{ij}}(x_{m_{ij}})$ characteristic **Eq. 34** is found to fall in the phase plane negative region over the memristor state range $[\tilde{x}_{m_{ij}}, x_{off}]$), featuring an upward concavity, while it decreases monotonically with $x_{m_{ij}}$, as depicted through a brown curve with magenta diamonds in **Figure 20B**, and forming, together with the $x_{m_{ij}}$ nullcline **Eq. 31**, the equilibrium **Eq. 39**, indicated via a black circle in the same plot.

Focusing now on the domain $v_{x_{ij}} > v_{sat}$, in view of condition **Eq. 65**, the $v_{x_{ij}} = v_{x_{ij}}(x_{m_{ij}})$ characteristic **Eq. 36** is found to be strictly positive, and to exhibit downward concavity as it monotonically increases with the memristor state. As a result, imposing the new condition

$$\frac{i_{w_{ij}} + a_{0,0} \cdot R_y \cdot g_{lin} \cdot v_{sat}}{G_x + x_{on}^{-1}} > v_{sat} \tag{68}$$

ensures that the graph of the function $v_{x_{ij}}(x_{m_{ij}})$ of **Eq. 36** assumes values within the domain $v_{x_{ij}} > v_{sat}$ for all $x_{m_{ij}} \in \mathcal{D}$, as indicated via the dashed brown curve with magenta diamonds in **Figure 20C**, creating, in conjunction with the $\dot{x}_{m_{ij}} = 0\,\Omega{\cdot}\mathrm{s^{-1}}$ locus, the equilibrium **Eq. 45**, shown as a black circle on the same plot. On the basis of the behaviour of the vector field in the negative (positive) saturation region, as established by conditions **Eqs. 48**, **49** and **Eqs. 50**, **52**, phase-plane trajectories below or above the $\dot{v}_{x_{ij}} = 0\,\mathrm{V{\cdot}s^{-1}}$ locus **Eqs. 34**, **36** are bound to move north-eastward (north-westward) or south-eastward (south-westward), visiting the yellow II (cyan III) or gray IV (green I) regions, as qualitatively sketched in the viewgraph of **Figures 20B,C**, unveiling the local stability of the equilibrium point **Eqs. 39**, **45**, as indicated through the filling structure of its black circle symbol.

All in all, the application of our stepwise system-theoretic M-CNN design method to the cell model **Eqs. 18**, **19**, identifies five inequalities, specifically **Eqs. 55**, **65**–**68**. Replacing the linear resistor in parallel to the capacitor in the memcomputing core of the cell circuit of **Figure 8** with an open circuit[43], the system of five inequalities may be reduced to the triplet of non-redundant conditions **Eqs. 55**, **66**, and **67**. Solving

[43]The modulus of the voltage, falling across the cell memristor at equilibrium, was found to be reasonably small, even for $G_x = 0\,\Omega^{-1}$, irrespective of the initial condition.

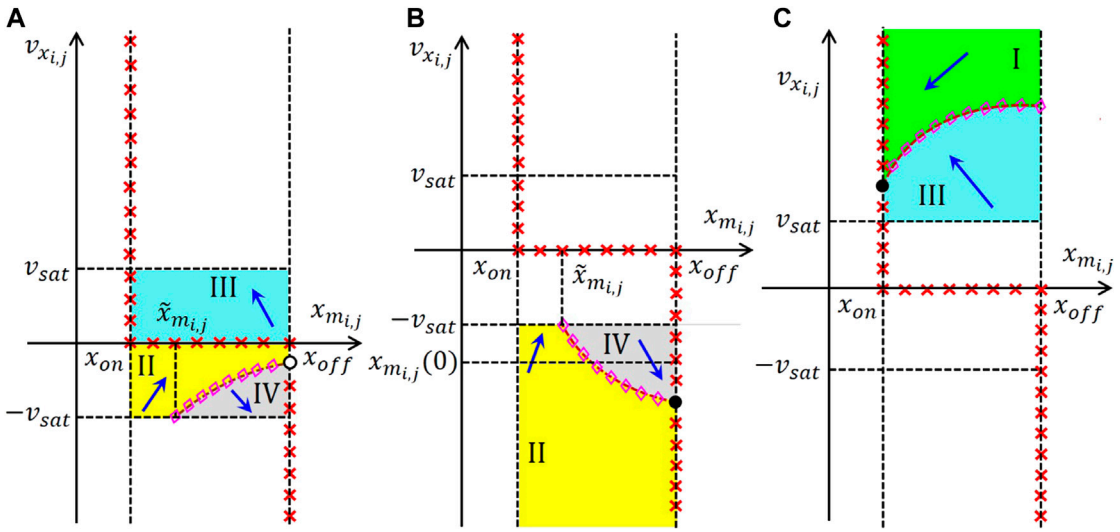


FIGURE 20 | Qualitative sketches illustrating graphically the steps of the system-theoretic method to massage the cell SDP in the linear **(A)**, negative saturation **(B)**, and positive saturation **(C)** regions for the derivation of a suitable gene to program the M-CNN for the data retrieval task.

them through the geometric analysis method, all the points lying in the magenta region of the z–$a_{0,0}$ parameter plane of **Figure 19B** allow to program the M-CNN to retrieve the memory content stored in the locally distributed memristive bank. With reference to **Figure 19**, the bistable cell SDP, shown in plot (a), was derived for the particular solution $(z^*, a^*_{0,0}) = (3.5 \cdot 10^{-5}, 6.25 \cdot 10^{-4}\,\Omega^{-1})$, which, as revealed via an asterisk marker in plot (b), is safely distanced from the white region of the parameter plane, where the three non-redundant inequalities would not be simultaneously satisfied. With the stable equilibrium of the positive saturation region, positioned at $(2\,k\Omega, +0.1950\,\text{V})$, choosing, for the initial condition $v_{x_{i,j},0}$ on the capacitor voltage, an intermediate value, specifically $-0.15\,\text{V}$, between the ordinate of the unstable equilibrium of the linear region, lying at $(10\,\text{k}\Omega, -0.0667\,\text{V})$, and the ordinate of the stable equilibrium of the negative saturation region, residing at $(10\,\text{k}\Omega, -0.275\,\text{V})$, the phase-plane trajectory is found to converge asymptotically toward the equilibrium over the domain $v_{x_{i,j}} < -(+)v_{sat}$ if the memristor initially sits in the off (on) resistive state x_{off} (x_{on}), as expected from rule 1 (2) from **Table 6**. Remarkably, the memristor state approaches asymptotically the same value it stored before its memory content interrogation commenced, revealing that no unwanted secondary effect accompanies the data reading operation. **Figure 21** demonstrates that a $M \times N$ M-CNN, accommodating cells regularly positioned along $M = 177$ rows and $N = 240$ columns, operates as desired, after its gene is programmed as established by the DRM$_2$-centered M-CNN design methodology. The binary image, shown in plot (a), illustrates graphically the resistive states of all the memristors. A white (black) pixel in a given position of this image reveals that the memristor in the equivalent location of the RECALL M-CNN stores the lowest (highest) possible resistance before the memory reading task is initiated. The image in plot (b) visualizes through a uniform gray color of appropriate tone[44] the common initial condition assigned to each capacitor voltage, set to $-0.15\,\text{V}$. Plot (c) codes the steady-state output voltages of all the cells. If the pixel, lying at the crossing between i^{th} row and j^{th} column of this image, is black (white), the steady-state output voltage of the cell in the equivalent location of the RECALL M-CNN is the positive (negative) saturation level.

6 DISCUSSION

The theory presented in this work is independent of the memristor model adopted in M-CNN circuit design. In fact, it is a general theory, which may be applied to a much wider range of nonlinear dynamical circuits other than the cellular arrays analyzed in the manuscript. It is worth to pinpoint that the Second-Order Dynamic Route Map (DRM$_2$), around which the design methodology proposed in this paper is centered, extends the classical Dynamic Route Map (DRM) (Chua, 1998; Chua and Roska, 2002), applicable to first-order systems only, allowing to draw a complete picture of the local and global behaviour of any second-order dynamical system. For example, it shall constitute the system-theoretic tool of reference for a thorough study of the nonlinear dynamics of memristive devices with two state variables. In this work the DRM$_2$ is adopted to investigate the spatio-temporal phenomena emerging in each of the second-order memristive cells of a two-dimensional cellular

[44]The graphical illustration convention, adopted here for visualizing each initial cell capacitor voltage is analogous as the one established in the discussion of the EDGE M-CNN design: the closer is its real value, lying in the set $(-1, 1)$ V, to the lower (upper) bound $-(+)1$ V, and the lighter (darker) is the tone of the gray colour attributed to the respective pixel.

FIGURE 21 | (A) Black-and-white checkerboard visualizing the binary data stored in the locally distributed memristive memory bank previous to their retrieval. A white (black) pixel at row i and column j in this image reveals the on (off) resistance of the memristor in the M-CNN cell $C(i,j)$ ($i \in \{1,\ldots,M\}$, $j \in \{1,\ldots,N\}$, $M = 177$ $N = 240$). **(B)** Grey-scale image indicating the common −0.15 V-valued initial condition on each capacitor voltage. **(C)** Output binary image coding the steady-state output voltages of all the RECALL M-CNN processing elements. A black (white) pixel in correspondence to the i^{th} row and j^{th} column of this image denotes a positive (negative) saturation level $+(-)v_{sat}$ for the steady-state output voltage of the cell located in the corresponding position of the memristive cellular array.

array. The two degrees of freedom of each cell are the voltage across a linear capacitor and the state of a first-order non-volatile memristor. The model (Pershin et al., 2009) adopted for the resistance switching memory in this work is a simple yet accurate mathematical description of a physical memristor realization (Jo et al., 2009). The reason behind the choice of this particular piecewise-linear memristor model for our study lies behind the pedagogical nature of this paper. In other words, since the aim of this manuscript is to provide researchers with powerful system-theoretic methods to analyze memristive cellular arrays, we found it useful to adopt an analytically tractable memristor model, in order to allow the determination of closed-form analytical expressions for the nullclines as well as for the equilibria of each State Dynamic Portrait (SDP) of a given DRM_2. The systematic M-CNN design methodology, presented in this paper, allows to derive optimal values for the cell parameters of each second-order cell $C(i,j)$ of a two-dimensional $M \times N$ array – $i \in \{1,\ldots,M\}$, $j \in \{1,\ldots,N\}$ – on the basis of the solution of an IS, which is preliminarily set up to obtain a desired task-dependent partition of the $x_{m_{i,j}}$–$v_{x_{i,j}}$ phase plane for each value of the input $v_{u_{i,j}}$ of the cell $C(i,j)$ itself and of the input $v_{u_{i+k,j+l}}$ of any of the 8 cells $\{C(i+k,j+l)\}$ – $k,l \in \{-1,0,1\}$, $k,l \neq (0,0)$ – in its neighboorhod. This rigorous system-theory-based design strategy represents one the first examples of a systematic memristor circuit design approach. As outlined in the paper, choosing the particular IS solution, which holds the largest distance from the parameter space regions, where the system would fail to operate as desired, allows to obtain a variability-aware design, which is of great interest, given the intrinsic cycle-to-cycle and device-to-device variability affecting memristive nanodevices. All in all, the system-theoretic analysis and design strategies, presented in this paper, are applicable to any M-CNN with second-order cells, irrespective of the particular models adopted for their constitutive first-order dynamical components, particularly the capacitor and the non-volatile memristor. Of course, in case one wished to use a first-order capacitor (memristor) model with a more complicated mathematical description, and pertaining to some other real-world electrical energy storage device (resistance switching memory), the appearance of the SDPs of a given cell DRM_2 would undergo inevitable changes, since the shape of the nullclines, the number, position, and stability of the second-order cell equilibria in the memristor state-capacitor voltage phase plane, the rules dictating how the sign of the time derivatives of the two states change across the $x_{m_{i,j}}$–$v_{x_{i,j}}$ phase plane, and, consequently, the final IS, leading to the selection of an optimal cell circuit parameter set for the implementation of a predefined memcomputing task crucially depend upon the particular cell model, but what matters is that the proposed theory, the highlight of this work, would keep its validity. The only downside associated with the adoption of more involved second-order cell models lies in the need, which would resultingly emerge, to recur to numerical methods for the investigation of the $x_{m_{i,j}}$–$v_{x_{i,j}}$ phase plane partitioning. Importantly, our future research efforts will be devoted to validate the system theory-centered memcomputing M-CNN designs by experimental verification on memristive hardware prototypes.

7 CONCLUSION

The motto "linearize-then-analyze", which electrical engineers have been advocating for generations, should not drive the investigation of highly nonlinear memristive devices, circuits and systems, which are being developed in our times through disruptive nanotechnologies with the intention to foster progress in integrated circuit (IC) design beyond the Moore era. In fact, given that linear analysis techniques are unable to gain a deep insight into the behaviour of a nonlinear system, the availability of a partial picture of the dynamics of a novel nano-device prevents its conscious use in IC design. Recurring to nonlinear system theory is thus absolutely necessary to unfold the full potential of memristors in electronics. However, the conversion of classical circuits to memristive equivalents might require the adaptation of classical nonlinear system-theoretic analysis and design techniques, as is the case in this study. Cellular Nonlinear Networks (CNNs) (Chua and Yang, 1988a; Chua and Yang, 1988b) constitute one of the earliest examples of a non-von Neumann computing architecture, where data processing and storage tasks are locally distributed across a multi-dimensional array of locally coupled dynamical systems. In analogue hardware implementations of these bio-inspired computing structures, the cells typically feature one degree of freedom. As a result, the Dynamic Route Map (DRM) graphical

tool, a powerful system-theoretic technique for the analysis of first-order systems, is applicable to gain a full understanding of the dynamics of these cells. Further, a rigorous procedure, employing the DRM analysis method, and leading to the derivation of an optimal solution for an inequality set (IS), which constrain number, and stability of cell equilibria for each of the possible combinations of inputs and/or initial conditions, allows to program the cellular network for the robust execution of a predefined computing task. The adoption of memristors in new designs of cell and coupling circuitry may allow to extend the processing functionalities and/or the computing efficiency of traditional dynamic arrays, thanks to the enrichment of the spectrum of dynamical phenomena, which may emerge within the cellular medium, while allowing to improve the spatial resolution of CNN analogue hardware realisations, concurrently. It is thus timely to investigate the impact of the introduction of memristors in new CNN designs. This work consider a first class of Memristor CNNs (M-CNNs), in which a first-order non-volatile resistance switching memory is inserted in parallel to the capacitor in each cell of a two-dimensional time- and space-invariant standard CNN. Given that the cells of each M-CNN from the proposed class feature two degrees of freedom, the DRM analysis methodology is no longer pertinent to gain insight into their data processing capabilities. A novel graphical tool, inspired to the Phase Portrait concept (Strogatz, 2000) from the theory of nonlinear dynamics, constituting the natural extension of the classical DRM system-theoretic technique to dynamical systems with two degrees of freedom, and called Second-Order Dynamic Route Map (DRM_2) (Tetzlaff et al., 2020), may allow to gain a deep insight into the dynamical phenomena emerging in cellular arrays with second-order memristive cells (Ascoli et al., 2020b), enabling to draw, finally, a codimension-2 bifurcation diagram, referred to as M-CNN Primary Mosaic, which specifies all the possible stable and unstable equilibria, which a cell may admit for each combination of self-feedback synaptic weight $a_{0,0}$ and offset current $i_{w_{i,j}}$. Finally, a rigorous procedure (Ascoli et al., 2020a), employing the DRM_2 graphical tool, and leading to the derivation of an optimal solution of an IS, which shape the phase portrait of each cell in such a way that solutions of the CNN model equations may approach predefined equilibria for each of the possible combinations of inputs and initial conditions, allows to tune the parameters of the cellular array for a variability-tolerant accomplishment of a prescribed signal processing task or of a predefined memory operation. This work, contributing to the establishment of solid foundations of M-CNN theory, highlights the huge potential of memristive mem-processing structures for edge computing applications, and is expected to serve as a source of inspiration for future studies intended to verify the theoretical predictions on the beneficial impact of resistance switching memories on the performance of cellular nonlinear arrays.

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

AA and RT conceived the main idea of the paper. AA developed the systematic M-CNN design methodology, made all the analytical calculations, run the complete set of simulations to confirm the theoretical derivations, and wrote the whole manuscript. RT, SK, and LC supported the research with inspiring suggestions, precious guidelines, and insightful advices.

FUNDING

This work has been partially supported by the Czech Science Foundation under grant No. 18–21608S, Czech Republic. LC is supported in part by AFOSR Grant No. FA 9550-18-1-0016.

REFERENCES

Ascoli, A., Tetzlaff, R., Kang, S. M, and Chua, L. O. (2020a). Theoretical foundations of memristor cellular nonlinear networks: a DRM-based method to design memcomputers with dynamic memristors. *IEEE Trans. Circuits Systems–I: Regular Pap.* 67 (8), 2753–2766. doi:10.1109/tcsi.2020.2978460

Ascoli, A., Messaris, L., Tetzlaff, R., and Chua, L. O. (2020b). Theoretical foundations of memristor cellular nonlinear networks: stability analysis with dynamic memristors. *IEEE Trans. Circuits Systems–I: Regular Pap.* 67 (4), 1389–1401. doi:10.1109/tcsi.2019.2957813

Biolek, Z., Biolek, D., and Biolkova, V. (2009). Spice model of memristor with nonlinear dopant drift. *Radioengineering* 18 (2), 210–214.

Bohaichuk, S. M., Kumar, S., Pitner, G., McClellan, C. J., Jeong, J., Samant, M. G., et al. (2019). Fast spiking of a mott VO – carbon nanotube composite device. *Nano Lett.* 19, 6751–6755. doi:10.1021/acs.nanolett.9b01554

L. O. Chua (Editor) (1998). *CNN: a paradigm for complexity. World scientific series on nonlinear science, series A, vol. 31.* Singapore: World Scientific.

Chua, L. O. (2015). Everything you wish to know about memristors but are afraid to ask. *Radioengineering* 24 (2), 319–368. doi:10.13164/re.2015.0319

Chua, L. O. (2018a). Five non-volatile memristor enigmas solved. *Appl. Phys. A.* 124 (8), 563. doi:10.1007/s00339-018-1971-0

Chua, L. O., Sbitnev, V., and Kim, K. (2012). Hodgkin-Huxley axon is made of memristors. *Int. J. Bifurcation Chaos* 22 (3), 1230011. doi:10.1142/s021812741230011X

Chua, L. (2014). If It's Pinched, It's a Memristor. *Semicond. Sci. Technol. Spec. Issue. Memristive Devices* 29 (10), 42. doi:10.1088/0268-1242/29/10/104001

Chua, L. O., and Kang, S. M. (1976). Memristive devices and systems. *Proc. IEEE* 64 (2), 209–223. doi:10.1109/proc.1976.10092

Chua, L. O. (2005). Local activity is the origin of complexity. *Int. J. Bifurcation Chaos* 15 (11), 3435–3456. doi:10.1142/s0218127405014337

Chua, L. O. (1971). Memristor: the missing circuit element. *IEEE Trans. Circuit Theor.* 18 (5), 507–519. doi:10.1109/tct.1971.1083337

Chua, L. O. (2018b). Memristors: remembrance of things past. *IEEE Micro* 38 (5), 7–12. doi:10.1109/mm.2018.053631136

Chua, L. O. (2011). Resistance switching memories are memristors. *Appl. Phys. A.* 102, 765–783. doi:10.1007/s00339-011-6264-9

Chua, L. O., and Roska, T. (2002). *Cellular neural networks and visual computing: foundations and applications.* 1 edition. Cambridge, England: Cambridge University Press.

Chua, L. O., and Yang, L. (1988b). Cellular neural networks: applications. *IEEE Trans. Circuits Systems–i* 35 (10), 1273–1290. doi:10.1109/31.7601

Chua, L. O., and Yang, L. (1988a). Cellular neural networks: theory. *IEEE Trans. Circuits Systems–I* 35 (10), 1257–1272. doi:10.1109/31.7600

Di Marco, M., Forti, M., and Pancioni, L. (2017a). Memristor standard cellular neural networks computing in the flux-charge domain. *Neural Netw.* 93, 152–164. doi:10.1016/j.neunet.2017.05.009

Di Marco, M., Forti, M., and Pancioni, L. (2017b). Convergence and Multistability of Nonsymmetric Cellular Neural Networks With Memristors. *IEEE Trans. Cybern.* 47 (10), 2970–2983. doi:10.1109/TCYB.2016.2586115

Di Marco, M., Forti, M., and Pancioni, L. (2018). New Conditions for Global Asymptotic Stability of Memristor Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* 29, 1822–1834. doi:10.1109/TNNLS.2017.2688404

Duan, S., Hu, X., Dong, Z., Wang, L., and Mazumder, P. (2015). Memristor-Based Cellular Nonlinear/Neural Network: Design, Analysis, and Applications. *IEEE Trans. Neural Netw. Learn. Syst.* 26 (6), 1202–1213. doi:10.1109/TNNLS.2014.2334701

Fujitsu Ltd (2019). Fujitsu announcement. Available at: https://www.fujitsu.com/global/products/devices/semiconductor/memory/reram/ (Accessed July 30, 2019).

GlobalFoundries Ltd (2018). GlobalFoundries announcement. Available at: https://spectrum.ieee.org/nanoclast/semiconductors/devices/globalfoundries-halts-7nm-chip-development (Accessed August 28th, 2018).

Hu, M., Strachen, J. P., Li, Z., Grafals, E. M., Davila, N., Lam, S., et al. (2016). "Dot-product engine for neuromorphic computing: programming 1T1M crossbar to accelerate vector-matrix multiplication," in Proceedings of the ACM/EDAC/IEEE 53rd Annual Design Automation Conference (DAC), Austin, TX, USA, June 5–9, 2016.

Itoh, M., and Chua, L. O. (2003). Designing CNN genes. *Int. J. Bifurcations Chaos* 13 (10), 2739–2824. doi:10.1142/s0218127403008375

Jo, S. H., Kim, K. H., and Lu, W. (2009). High-density crossbar arrays based on a Si memristive system. *Nano Lett.* 9 (2), 870–874. doi:10.1021/nl8037689

Karacs, K., Cserey, Gy., Zarándy, Á., Szolgay, P., Rekeczky, Cs., Kék, L., et al. (2018). CNN template library. Available at: http://cnn-technology.itk.ppke.hu/Template_libraryv4.0beta2.pdf (Accessed April 19, 2021).

Kumar, S., Williams, R. S., and Wang, Z. (2020). Third-order nano-circuit elements for neuromorphic engineering. *Nature* 585, 518–523. doi:10.1038/s41586-020-2735-5

Landauer, R. (1988). Dissipation and noise immunity in computation and communication. *Nature* 335 (6193), 779–784. doi:10.1038/335779a0

Li, C., Hu, M., Li, Y., Jiang, N., Ge, N., Montgomery, E., et al. (2018). Analogue signal and image processing with large memristor crossbars. *Nat. Electron.* 1, 52–59. doi:10.1038/s41928-017-0002-z

Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electronics* 38 (8), 114–117. doi:10.2307/3756714

Panasonic Ltd. (2013). Panasonic announcement. Available at: https://news.panasonic.com/global/press/data/2013/07/en130730-2/en130730-2.html (Accessed July 30th, 2013).

Pershin, Y. V., and Di Ventra, M. (2011). Memory effects in complex materials and nanoscale systems. *Adv. Phys.* 60 (2), 145–227. doi:10.1080/00018732.2010.544961

Pershin, Y. V., Fontaine, S. L., and Di Ventra, M. (2009). Memristive model of amoeba learning. *Phys. Rev. E.* 89, 021926. doi:10.1103/PhysRevE.80.021926

Roska, T. (1993). The CNN universal machine: an analogic array computer. *IEEE Trans. Circuits Syst.–II: Analog Digital Signal Process.* 40 (3), 163–173. doi:10.1109/82.222815

Strogatz, S. H. (2000). *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering (studies in nonlinearity).* 1 edition. Florida, US: CRC Press.

Tetzlaff, R., Ascoli, A., Messaris, I., and Chua, L. O. (2020). Theoretical foundations of memristor cellular nonlinear networks: memcomputing with bistable-like memristors. *IEEE Trans. Circuits Syst.–I: Regular Pap.* 67 (2), 502–515. doi:10.1109/tcsi.2019.2940909

Toshiba Ltd. (2012). Smart photo sensor SPS02. Available at: https://www.toshiba-teli.co.jp/pdf/industrial-camera-en/sps02e.pdf (Accessed February 1st, 2012).

Vázquez, A. R., Espejo, S., Domínguez-Castro, R., Huertas, J. L., and Sánchez-Sinencio, E. (1993). Current-mode techniques for the implementation of continuous- and discrete-time cellular neural networks. *IEEE Trans. Circuits Syst. II: Analog Digit. Signal Process.* 40 (3), 132–146. doi:10.1109/82.222812

Vázquez, A. R., Fernández-Berni, J., Leñero-Bardallo, J. A., Vornicu, I., and Carmona-Galán, R. (2018). CMOS vision sensors: embedding computer vision at imaging front-ends. *IEEE Circuits Syst. Mag.* 18 (2), 90–107. doi:10.1109/MCAS.2018.2821772

Weiher, M., Herzig, M., Tetzlaff, R., Ascoli, A., Mikolajick, T., and Slesazeck, S. (2019). Pattern formation with local active S-type NbOx memristors. *IEEE Trans. Circuits Syst.–I: Regular Pap.* 66 (7), 2627–2638. doi:10.1109/tcsi.2019.2894218

Williams, R. S. (2017). What's next? [The end of Moore's law]. *IEEE Comput. Sci. Eng.* 19 (2), 7–13. doi:10.1109/mcse.2017.31

Zarándy, Á. (2003). The art of CNN template design. *Int. J. Circuit Theor. Appl.* 27 (1), 5–23. doi:10.1002/(sici)1097-007x(199901/02)27:1<5::aid-cta38>3.0.co;2-c

Zhang, X., Wu, Z., and Chua, L. (2020). Hearts are poised near the Edge of Chaos. *Int. J. Bifurcation Chaos* 30 (9), 2030023. doi:10.1142/s0218127420300232

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.

frontiers in Nanotechnology

ORIGINAL RESEARCH
published: 26 May 2021
doi: 10.3389/fnano.2021.675792

Spoken Digit Classification by In-Materio Reservoir Computing With Neuromorphic Atomic Switch Networks

Sam Lilak[1], Walt Woods[2], Kelsey Scharnhorst[1], Christopher Dunham[1], Christof Teuscher[2], Adam Z. Stieg[3,4] and James K. Gimzewski[1,3,4,5]**

[1]Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA, United States, [2]Department of Electrical and Computer Engineering, Portland State University, Portland, OR, United States, [3]California NanoSystems Institute, University of California, Los Angeles, Los Angeles, CA, United States, [4]WPI Center for Materials Nanoarchitectonics (MANA), National Institute for Materials Science (NIMS), Tsukuba, Japan, [5]Research Center for Neuromorphic AI Hardware, Kyutech, Kitakyushu, Japan

Atomic Switch Networks comprising silver iodide (AgI) junctions, a material previously unexplored as functional memristive elements within highly interconnected nanowire networks, were employed as a neuromorphic substrate for physical Reservoir Computing This new class of ASN-based devices has been physically characterized and utilized to classify spoken digit audio data, demonstrating the utility of substrate-based device architectures where intrinsic material properties can be exploited to perform computation in-materio. This work demonstrates high accuracy in the classification of temporally analyzed Free-Spoken Digit Data These results expand upon the class of viable memristive materials available for the production of functional nanowire networks and bolster the utility of ASN-based devices as unique hardware platforms for neuromorphic computing applications involving memory, adaptation and learning.

Keywords: atomic switch networks, memristive, neuromorphic, reservoir computing, in-materio

OPEN ACCESS

Edited by:
Huanglong Li,
Tsinghua University, China

Reviewed by:
Qijun Sun,
Beijing Institute of Nanoenergy and Nanosystems (CAS), China
Jianshi Tang,
Tsinghua University, China

****Correspondence:***
Adam Z. Stieg
stieg@cnsi.ucla.edu
James K. Gimzewski
gim@chem.ucla.edu

Specialty section:
This article was submitted to Nanodevices, a section of the journal Frontiers in Nanotechnology

Received: *03 March 2021*
Accepted: *30 April 2021*
Published: *26 May 2021*

Citation:
Lilak S, Woods W, Scharnhorst K, Dunham C, Teuscher C, Stieg AZ and Gimzewski JK (2021) Spoken Digit Classification by In-Materio Reservoir Computing With Neuromorphic Atomic Switch Networks. Front. Nanotechnol. 3:675792. doi: 10.3389/fnano.2021.675792

INTRODUCTION

Speech recognition is a seminal task in the field of artificial intelligence and natural language processing. Typical algorithmic approaches to speech recognition break apart sections of raw speech data and bin them into hidden Markov models manipulating Markov chains. While effective, these approaches are more computationally intensive than some recently developed neural network models, which may prove a more suitable compute framework for handling increasingly larger data sets (Schatz and Feldman, 2018; Mustafa et al., 2019; Deshmukh, 2020). Artificial Neural Networks (ANNs) have also been a promising avenue for more efficient speech recognition tasks which offer the benefit of being trained for natural language processing and are believed to be a more suitable candidate for handling the varied complexity of each person's unique voice and accent. Implementation of ANNs in modern computing hardware remains computationally burdensome and often requires access to and utilization of high-performance computing clusters. A suitable hardware architecture for local execution of complex tasks such as natural language processing must be able to process dynamic, temporal data in real-time while remaining energy efficient. Memristive materials have been identified as strong candidate for such applications as they offer an opportunity to alleviate the bus latency between memory and processing elements in traditional von Neumann

architectures while also performing in-memory computation with reduced power consumption (Ielmini and Wong, 2018). The nonlinear character of memristors, resulting from the underlying physics of the material itself, is essential for enabling simultaneous storage of data (memory) and performance of complex tasks with it (processing) through a relatively new technique known as evolution in-materio (Miller and Downing, 2002; Harding and Miller, 2009; Miller et al., 2014; Dale et al., 2017).

The growing field of evolution in-materio computing has sought to optimize computational architectures *via* evolutionary (search) algorithms (Harding and Miller, 2009; Dale et al., 2017). The materials and architectures employed vary with the desired facet of computation, but ideally these materials are computationally and energetically efficient at employing a litany of machine learning based algorithms. Utilizing a single hardware element capable of exhibiting both memory and processing alleviates the burden of busing information between two separate hardware components, reducing latency in computation (Mustafa et al., 2019). The most robust currently known architecture that combines the aforementioned elements is the mammalian brain, which has been both a foundation and inspiration toward the development of architectures which can efficiently process multi-input, chaotic, and/or time-varying (temporal) datasets.

This work focuses on the class of neuromorphic computing devices known as Atomic Switch Networks (ASN), comprising a highly interconnected network of memristive nanowire junctions as shown schematically in **Figure 1**. Ongoing efforts to develop memristive hardware for neuromorphic computing include not only ASNs, but also patterned crossbar arrays, and nanoparticle clusters (Moon et al., 2019; Du et al., 2017; Alibart et al., 2013; Sattar et al., 2013; Tappertzhofen et al., 2012). ASN-based devices provide a physical system with structure and functional dynamics reminiscent of the mammalian brain (Srinivasa and Cruz-Albrecht, 2012; Avizienis et al., 2012; Türel et al., 2004; Calimera et al., 2013) that has previously been employed as a computational material for applications in Reservoir Computing (RC) (Lukoševičius and Jaeger, 2009; Schrauwen et al., 2007; Snyder et al., 2012; Du et al., 2017; Goudarzi et al., 2014; Sillin et al., 2013; Fu et al., 2020). The atomic switch is a nanoscale electroionic element consisting of a Metal-Insulator-Metal (MIM) junction whose properties can be manipulated via a time-dependent input signal (Zhu et al., 2020; Kuncic et al., 2020; Manning and et al., 2018; Manning et al., 2017). Individual atomic switches have been shown to produce memristive, nonlinear responses, exhibiting both short and long-term memory as well as quantized conductance (Sattar et al., 2013; Tappertzhofen et al., 2012; Terabe et al., 2005; Hasegawa et al., 2010). For electrochemical metallization memristors filament growth is dominated by cation transport through the insulating medium as shown in **Figure 1** and has been experimentally observed *in-situ* (Guo et al., 2007; Yang et al., 2012; Sun et al., 2019). These properties render atomic switches and other memristive systems as ideal circuit elements for use within a network architecture that can serve as a dynamic physical reservoir used to solve complex computational tasks, including speech recognition and natural language processing (Kan et al., 2021; Zhong et al., 2021).

RC provides a framework for computing complex functions using a dynamical system as a "reservoir" (Lukoševičius and Jaeger, 2009; Hashmi et al., 2011; Lukoševičius et al., 2012; Sillin et al., 2013). The RC framework is ideal for the processing of dynamic, temporal real-time signals and can be used in many of the same situations as recurrent feed-forward neural networks. RC also offers advantages such as fault-tolerance and the capacity for learning (Hashmi et al., 2011; Stieg et al., 2014). Passing a time varying input through a dynamic reservoir produces a higher dimensional representation of the signal through nonlinear transformation, where different points on the reservoir are measured and linearly combined to reproduce an arbitrary output signal as shown in **Figure 2**. Training is only performed on the linear readout coefficients (voltage readouts are shown in **Figure 3** and in **Supplementary Figure S2** demonstrating a reproducible response over time); the reservoir dynamics themselves are generally considered fixed. Limiting training to the weights between the reservoir and output layer alleviates the need to use gradient-descent based methods, greatly minimizing the associated computational burden.

As an alternative to simulation-driven RC, in-materio RC leverages material complexity for computational purposes (Teuscher, 2017; Konkoli et al., 2018; Tanaka et al., 2019; Nakajima, 2020). Whereas early implementations of RC simply utilized a body of a liquid acting as the dynamic reservoir, more recent works harnessed the intrinsic properties of complex physical systems, including ASNs, as the basis for a computation (Lukosevicius, 2011; Lukoševičius et al., 2012; Snyder et al., 2012; Goudarzi et al., 2014; Fu et al., 2020). Software RC has historically been demonstrated as a suitable method for a litany of complex tasks including pattern classification, signal generation and temporal based logic tasks (Tanaka et al., 2019). Hardware based approaches to RC commonly leverage photonic interactions or memristor dynamics, though photonic systems aren't performing computations in-materio in contrast to memristors (Vandoorne et al., 2010; Tanaka et al., 2019). In-materio approaches to traditional RC have recently garnered attention as potential candidates to accelerate compute times while achieving higher power efficiency. Recent in-materio studies have demonstrated high accuracy in time-series analysis (Moon et al., 2019; Zhong et al., 2021), handwritten digit identification (Midya et al., 2019) and biosignal processing (Kudithipudi et al., 2016).

Computational neural models such as the perceptron and support vector machine can also be used as reservoirs; however, long convergence times can be a drawback depending on the task. Material-based reservoirs have the benefit of efficiently performing these tasks *in-situ*, enabling low-power, on-chip computing (Loppacher et al., 2003; Kuzum et al., 2012; Bürger et al., 2015). This alternative approach offers the opportunity to employ neural networks and machine learning algorithms offline, without the need to access servers, clusters and other high-performance computing infrastructures.

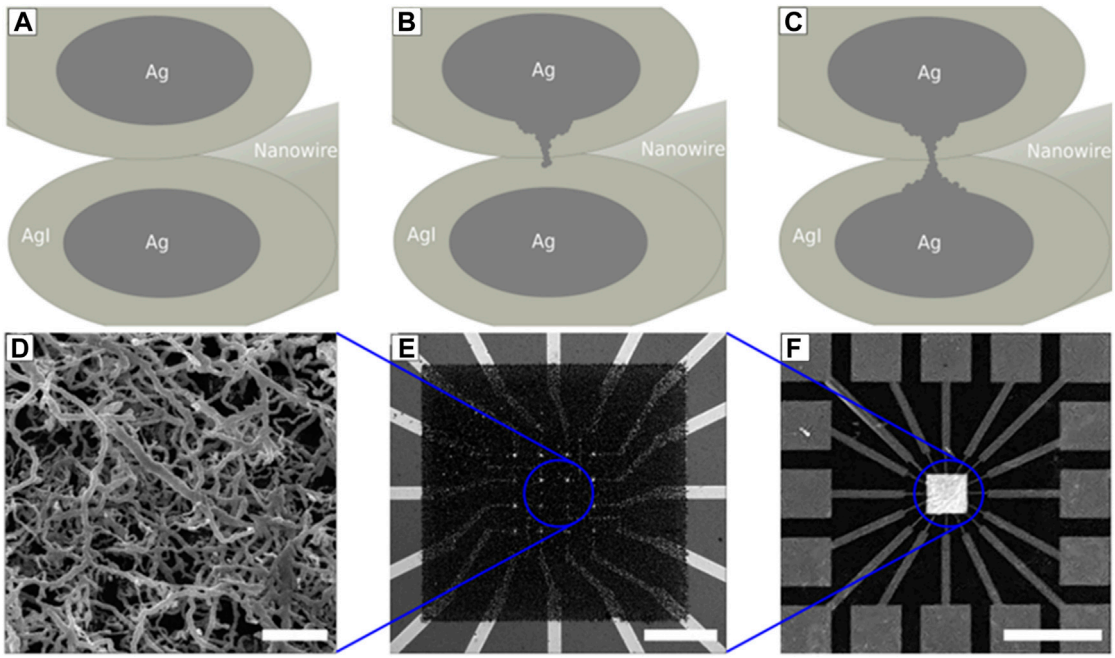


FIGURE 1 | Schematic diagram of an AgI-based ASN device, from nanowire junction to chip. **(A)** initial high resistance state of the system. **(B)** filament formation process under an applied bias **(C)** completed silver filament short circuits between overlapping nanowires (low resistance state). Yellow-gray represents AgI. Dark-gray represents Ag. Filament formation occurs as a gapless junction between Ag nanowires. **(D)** SEM image of the interconnected nanowire (scale bar = 20 um). **(E)** Optical image of microelectrode array at center of the ASN device (scale bar = 360 um). **(F)** Optical image of a complete 16-electrode ASN device (scale bar = 5 mm).

ASNs have been shown to represent uniquely suitable class of materials for implementation of hardware-based RC, namely complex network architectures with the requisite material complexity (Avizienis et al., 2012; Stieg et al., 2014; Nayak et al., 2010). These self-organized systems offer a unique opportunity to produce highly interconnected memristive networks, where a density of atomic switch junctions of up to $10^8/cm^{30}$ has been previously reported. The fabrication scheme, based on electroless deposition, produces a diverse ensemble of silver nanowires with varying lengths, widths and thereby junction dimensions. This structural diversity in the material substrate imparts a distribution of operational characteristics that improves the capacity to perform non-linear transformations of input signals.

Herein, we report the use of a new memristive material, silver iodide (AgI), as the functional element in the ASN framework (Liang et al., 2007; Tappertzhofen et al., 2012; Cai et al., 2013). Silver iodide can be robustly prepared in a brief vapor phase reaction of iodine vapor with silver nanowires at room temperature in contrast to the lengthy formation times at elevated temperatures of previously reported silver sulfides. This promising material provides voltage-controlled resistance in both the bulk and when integrated into crossbar architectures, rendering it suitable as a memristive material for RC applications which require non-linear transformations and quantized conductance states (Stieg et al., 2014). This work expands the catalog of investigated ASN materials by fabricating and testing AgI for non-linear, temporal computation through the classification of spoken digits.

METHODS

Device Fabrication

The substrate for ASN devices, a multielectrode array enabling spatiotemporal stimulation and monitoring, was fabricated using standard thermally oxidized (500 nm) silicon wafers as the base substrate. A 16-electrode grid of Pt (150 nm) was patterned by photolithography and deposited using a negative photoresist (AZ NLOF 2020) onto a Cr or Ti wetting layer (5 nm). Liftoff was induced overnight in N-methyl-2-pyrrolidone (NMP) at 60°C. Point contact electrodes were prepared using a patterned insulating layer of SU-8 (400 nm) which was soft baked (90°C), exposed to UV, post exposure baked (90°C), developed for 3 min, and hard baked at 180°C for 30 min. An array of copper (300 nm) seed sites with 5 × 5 μm spacing in a grid were patterned onto inner point contact electrodes and deposited onto AZ NLOF 2020 *via* metal evaporation at 3 nm/s followed by lift-off overnight in NMP (60°C). The resultant device platforms consist of a stack of Si/SiO2/Cr/Pt-electrodes/SU-8/Cu-posts (**Supplementary Figure S1**) and were stored in inert atmosphere until bottom-up silver nanoarchitecture construction (Sillin et al., 2013; Demis et al., 2016).

This substrate was placed into a 50 mm solution of silver nitrate ($AgNO_3$) for 30–60 min. Silver nanowires formed through

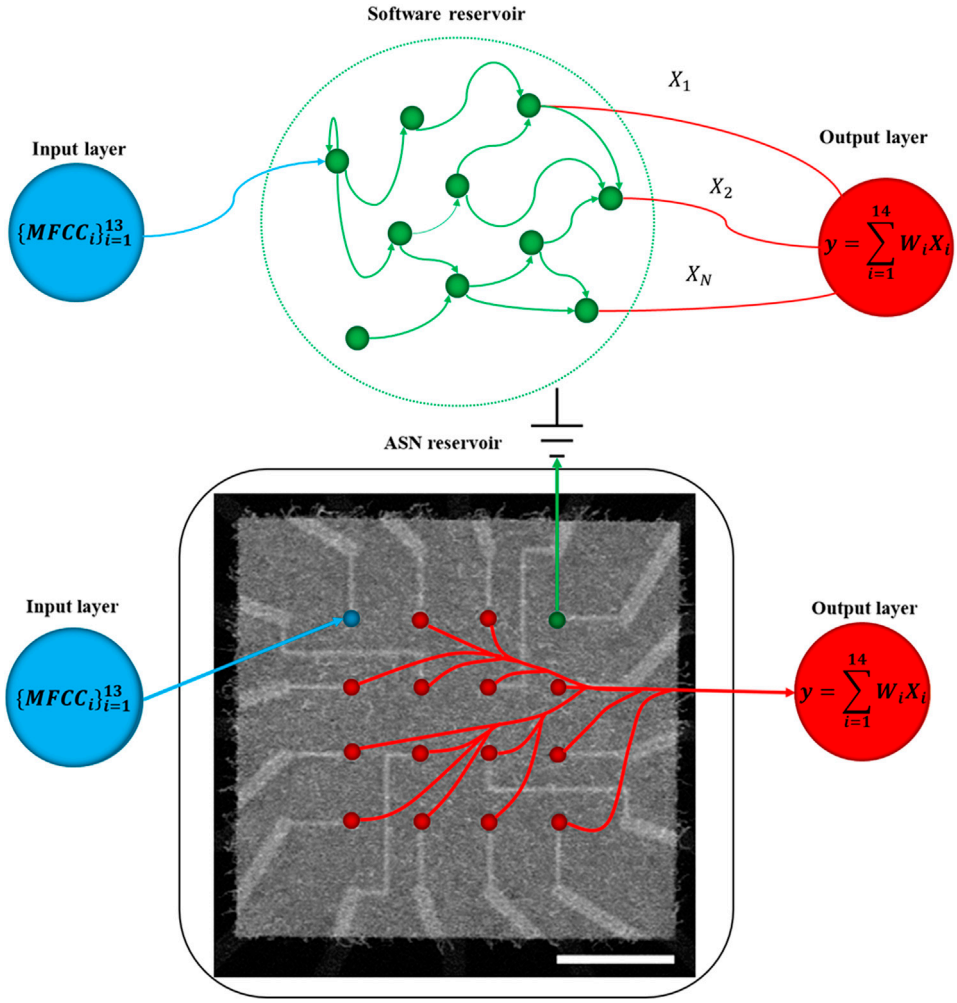


FIGURE 2 | Overview of a traditional software-based reservoir **(top)** in contrast to the ASN acting as an in-materio reservoir **(bottom)**, in which 13 MFCC's are sequentially delivered to the ASN in the form of a time-varying voltage to a single electrode. Utilizing physical nodes enables hardware acceleration at a lower power cost. Simultaneous, real-time voltage measurements are carried out at each of the remaining 14 electrodes and provided to an output layer for regression analysis.

an electroless deposition reaction involving the reduction of silver and the oxidation of copper through the following reaction:

$$Cu^{0}_{(s)} + 2Ag^{+}_{(aq)} \rightarrow Cu^{+2}_{(aq)} + 2Ag^{0}_{(s)}$$

The ordered copper posts (5 × 5 μm) directed a density-controlled formation of interconnected silver nanowires, whereby each ASN exhibited a unique structure determined by the bottom-up fabrication of metal cations. Subsequent silver iodide was formed in a nitrogen purged and sealed glass chamber with the ASN chip suspended over a small iodine pellet. Two different experimental techniques, one under ambient conditions (5 min exposure time) and the other with added heat (30°C, 2–3 min exposure time) were employed with both techniques successfully iodizing the silver nanowires.

$$2Ag_{(s)} + I_{2\,(g)} \rightarrow 2AgI_{(s)}$$

UV-Vis and XPS samples were prepared using transparent silver thin films (20 nm). These films were deposited on glass cover slides *via* a silver target in a Hummer 6.2 sputter system at 15 mA from Anatech Ltd. (Hayward, CA, United States) under an argon vacuum environment (80 mtorr).

Material Characterization

Optical and scanning electron microscopy (SEM) were used to characterize the as-fabricated structure of the nanowire network. SEM images were acquired using the JEOL JSM-7500F. X-Ray photoelectron (XPS) and UV-VIS spectroscopy were employed using transparent Ag thin film substrates with Ag as a control. Absorbance spectra of thin films were collected using the HP 8453 spectrophotometer. XPS spectra were obtained on an AXIS Ultra DLD XPS instrument from Kratos Analytical. The X-ray source was Al Kα at 1,486.6 eV. Survey (1,200 eV) and high-resolution scans were integrated over 4 and 16 sweeps, respectively.

Electrical Characterization

Characterization of ASN devices involves the spatially defined stimulation and monitoring of electrical activity throughout the network in the form of current and voltage traces. All input-output signals were generated/acquired using a purpose-built software package developed in Labview in conjunction with dedicated hardware manufactured by National Instruments. A data acquisition card (DAQ) (model PXIe-6368) was used to deliver input signals routed through a shielded connector box (model SCB-68A) to the ASN device. A source measurement unit (model PXIe-4141) was used to measure current flow through the ASN at user-selected electrodes, where acquired and applied signals were routed using a 16 × 32 switch matrix terminal block (model TB-2642B). Voltage traces were simultaneously monitored at all 16 electrodes using the DAQ card. All components were housed in a National Instruments chassis (model PXIe-1078) with an embedded controller.

Prior to any FSDD output signals, each ASN was driven through an initialization (activation) process in which the electrodes were sequentially stimulated with 7 Hz triangle waves. This process was repeated with increasing voltages (0.01–1 V) to realize switching patterns within the network. The switch matrix was employed in conjunction with the DAQ to calculate the resistance of every electrode combination prior to and after initialization, where successful activation was characterized by a sharp reduction in the network-wide parallel resistance as compared to the virgin metal system. Current-voltage and voltage-voltage measurements utilized triangle wave outputs from the DAQ card. The FSDD signal outputs were also produced by the DAQ card at selected electrode locations via the switch matrix.

Reservoir Computing

The AgI ASN's were evaluated for their potential RC applications through three different tests: non-temporal logic operations, temporal logic operations and recall of previous inputs and spoken digit classification. The non-linear XOR task was chosen for all logic operations and the assessment of the networks temporal properties as described in the Supplementary Information.

Spoken digit classification was implemented in AgI ASN devices via RC using the FSDD. The task was not performed using raw audio data, but rather using Mel-Frequency Cepstrum Coefficients (MFCCs) of the data, similar to previously reported techniques. Each 8 kHz wave-format sound file from the FSDD was zero-padded up to 1 s of recording length and then converted into MFCCs using the "python_speech_features" *Python* package. Mel-frequency cepstrum is a short-term power spectrum of the sound waves, using a linear cosine transform of a log power spectrum and is a nonlinear mel scale of frequency that approximates the human auditory response better than standard linear spacing of frequency components.

Default settings were used, resulting in an array of MFCCs where each 25 ms window of signal was parameterized by 13 MFCCs. Windows were offset by 10 ms, resulting in 1,287 total coefficients. To reduce device thrashing, the resulting MFCC array was flattened and fed to the network one at a time. The entire temporal sequence of the lowest-frequency coefficient was passed first, then the next-lowest-frequency coefficient's values, and so on. The resulting 1,287 Hz signal (shown in **Figure 2**) was sent to an input electrode, 14 electrodes were measured, and another electrode was grounded. Both the input and 14 read electrodes were recorded at 1 kHz. For RC, the resulting voltage streams were sampled at the end of sub-windows of computation, and the entire collection of sampled recordings was linearly regressed to indicate which digit was spoken (see **Figure 4**). Twelve unique spoken digit recordings were used, characterized by two speakers, saying three digits, two unique times. The FSDD speakers were "Jackson" and "Theo", the digits spoken were zero, one, or two, and the first two instances of each digit were used. As a baseline, regressions were performed on only the input electrode's voltage reading ("input only" mode) as well as on the full electrode suite of the input electrode and the 14 readout electrodes ("reservoir" mode).

RESULTS AND DISCUSSION

Material and Device Characterization

Silver nanowire networks like those shown in **Figure 1B** were reliably produced based on previously developed protocols. The network functionalization process requires conversion of silver nanowire junctions to silver iodide. The protocol for the formation of silver iodide was validated using UV-Vis and X-ray Photoelectron Spectroscopies (XPS). **Figure 4** provides representative visible absorption spectra of as-prepared Ag and AgI thin films. Ag thin films prepared by desktop sputtering exhibited a Surface Plasmon Resonance (SPR), suggesting the presence of silver islands within the film (Bharathi Mohan et al., 2007). These results are in line with previous reports which have demonstrated that silver exposed to iodine decreases SPR intensity coupled with a buildup of excitons. An absorbance peak around 420 nm has been previously reported and longer exposure to iodine at ambient temperature yielded a red-shifted maximum, which has been associated with the formation of larger AgI particles (Bharathi Mohan et al., 2007; Gnanavel and Sunandana, 2008). XPS results shown in **Figure 5** confirmed the presence of characteristic peaks for iodide $3d_{5/2}$ and $3d_{3/2}$ core level energies previously reported in metal iodides at binding energies of 620 and 631 electron volts (eV) which are absent in silver control samples (Kato and et al., 2015). While both functionalization protocols successfully produced AgI, the heated method was used for all ASN devices due to quicker sublimation of solid iodine.

To confirm the viability of AgI networks as a physical substrate for in-materio RC, the spatially distributed nonlinear characteristics of the ASN were examined. Voltage traces acquired at each of the 14 measurement electrodes enabled the analysis of Lissajous plots (V-V) as shown in **Figure 6**. AgI devices demonstrated distributed

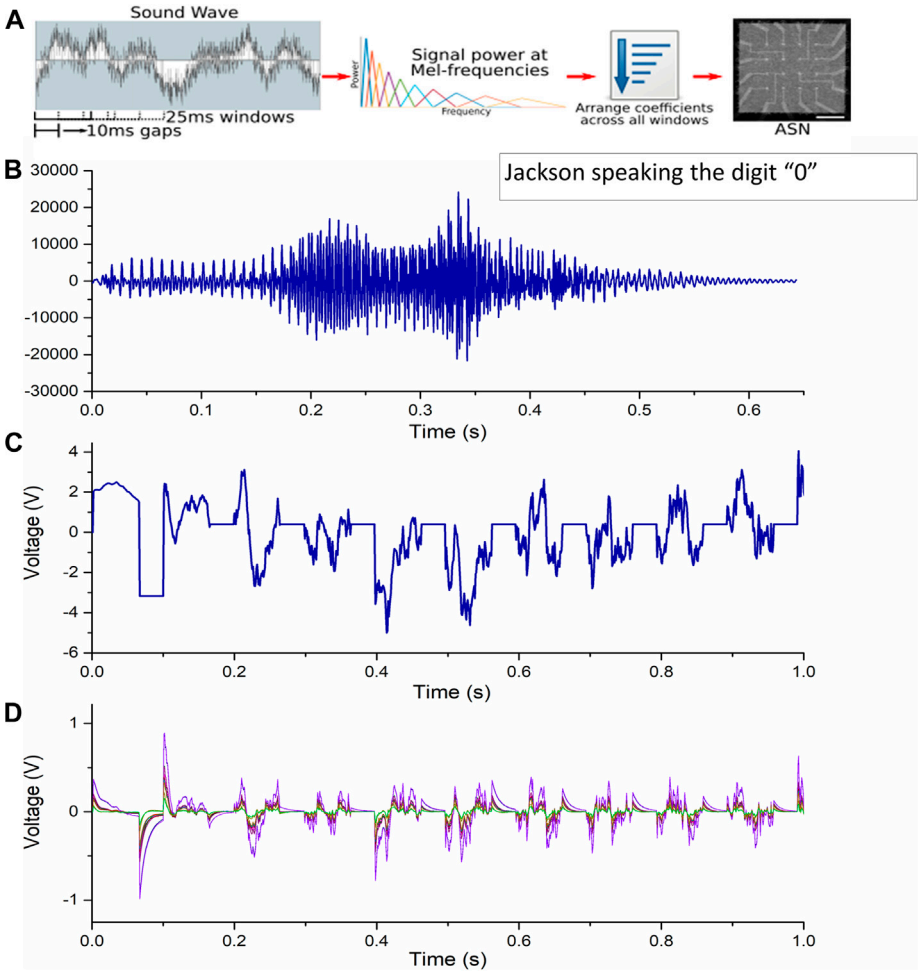


FIGURE 3 | (A) The workflow for RC-based speech recognition using ASN-based devices involved encoding and separation of raw audio data - spoken digits - data into overlapping windows, each of which was converted into 13 MFCCs. Individual MFCCs were arranged to minimize input thrashing and then delivered as input voltage to a single electrode of the ASN device. Output data, in the form of voltage traces, was collected at all remaining electrodes. **(B)** The raw FSDD audio signal of "Jackson" speaking the digit zero and its subsequent conversion to a voltage signal **(C)**. The resultant 14 voltage recordings and their unique responses are overlaid in **(D)** with additional detail provided in **Supplementary Material**.

nonlinear dynamics throughout the entirety of the nanowire network as a consequence of their highly interconnected nature, where a stable and reproducible nonlinear transformation of the input signal was observed. Different switching regimes emerge throughout the network (**Figure 6**) demonstrating different dynamics dominating spatial regions, suggesting there is a combination of switching dominated (blue, green plots) and capacitance dominated (red, pink) regions distributed throughout the network under an applied bias at any given electrode combination. The switching for mechanism for AgI junctions is accepted to arise from the formation metallic filaments between the insulating material classifying them as electrochemical metallization cells (Guo et al., 2007; Yang et al., 2012; Sun et al., 2019; Yang et al., 2013). The memristive properties of individual AgI junctions have been well characterized by Tappertzhofen et al. (2012), Sánta et al. (2020) and clearly demonstrate pinched hysteresis in their I-V curves. The unique dynamics observed in ASNs are the result of coupled memristive switching events among many interconnected junctions, where measurements at a point electrode capture the dynamics of an ensemble of memristive elements rather than a single one. Consequently, Lissajous plots of ASN device operation do not commonly produce the characteristic pinched hysteresis loops associated with individual memristive junctions. This capacity for the non-linear transformation of time-varying signals and temporal datasets renders the AgI nanowire network ideal for the performance RC-based speech recognition tasks.

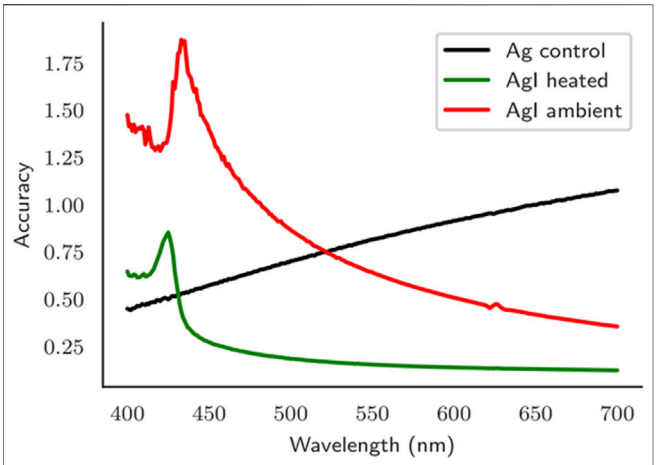


FIGURE 4 | UV-Vis spectra of silver thin films before and after iodization under ambient (λ max = 433 nm) and heated to 30°C (λ max = 424 nm). The presence of surface plasmon resonance in the blank silver samples suggests the thin films are discontinuous small islands of metal formed during the sputtering process.

AgI Atomic Switch Networks-Based Reservoir Computing

AgI nanowire networks were evaluated for their RC potential in spoken digit recognition as shown schematically in **Figure 4**. To effectively benchmark the value of the nanowire network in the performance of a spoken digit classification task, linear regression was performed in two ways. First, linear regression of the input voltages only - defined as "Input Only"-was carried out in the absence of the physical reservoir. Second, the full reservoir system–defined as "Reservoir" - employed regression of both the input signal and all device outputs. Inclusion of the input signal allows the regression to more accurately discern correlations between the transformed output signals and the input itself. FSDD digits encoded as MFCCs and passed to the network as a temporal sequence at 1,287 Hz were successfully classified as shown in **Figure 7**. A sufficient number of training examples were found to stabilize the reservoir's behavior, and evaluating testing data on only a single array of readout coefficients was found to be valid (Scharnhorst et al., 2017).

The target function was regressed by dividing the measured electrode data into N segments and using the last data point from each segment, this post-processing of voltage traces is done offline. For the "input only" mode, this means that $N = 80$ used 80 values in the regression. For the "reservoir" mode, this means that $N = 80$ used $80 \times 15 = 1{,}200$ values in the regression. To determine the accuracy at each N value, 12-fold cross-validation was employed using 11 of the audio files as training data and the 12th audio file as testing data. Each file was delivered to the device multiple times on a loop, aggregating far more than twelve tests to compute the accuracy. Nonetheless, there were only 12 unique data streams used. As a result, this problem suffered from significant overfitting, indicated by the "input only" results decreasing in accuracy as more points were used for the regression. This overfitting manifested as significant noise in the accuracy; $N = 100$ might give an accuracy as high as 100%, while $N = 101$ would give an accuracy of 54%. To account for this, the space of points N tested was divided into windows of size 25, and the average and standard deviation of accuracy within this window is shown in **Figure 7**. For instance, the mean and standard deviation shown at $N = 100$ indicate the statistics for $N \in \{88, ..., 112\}$. The ASN reservoir also demonstrated highly accurate results across a wide range of input voltages (0.5–10 V), suggesting potential utility of these devices for low-power applications.

These results clearly demonstrate the added stability provided by the ASN reservoir, evidenced by consistent accuracy at higher points of regression in the reservoir. The ASN's robustness and versatility was demonstrated by its capability to discern spoken digits when stimulated by both high and low voltage signals without a significant loss in accuracy. The ASN also provided a moderate benefit in accuracy, even before the input-only lines began overfitting. The lack of overfitting on the reservoir lines could be interpreted as a side-effect of the temporal, non-linear properties of the reservoir. This is corroborated by the fact that the reservoir lines achieved higher accuracy than the input only lines, a phenomenon that could not be achieved without non-linear or temporal behavior. Rather than relying on a stream of individual values, each of which has some noise associated, the reservoir readout mode could rely on 15 such streams. Assuming the noise on each electrode is somewhat independent, averaging these channels could have significantly reduced noise.

CONCLUSION

Neuromorphic nanowire networks such as the ASN represent a burgeoning class of material architectures whose dynamical

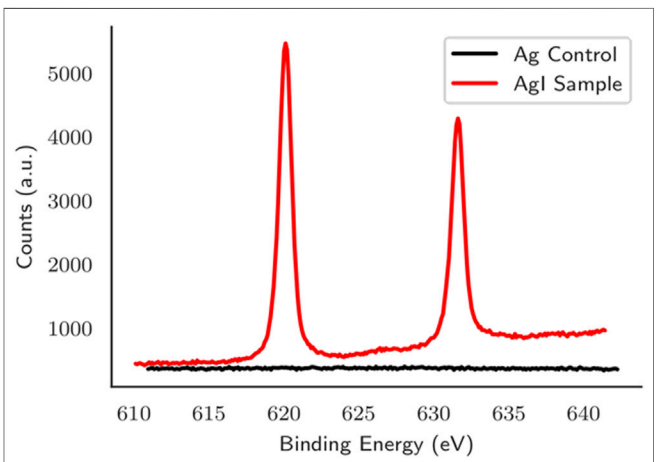


FIGURE 5 | XPS spectra of the iodine $3d_{5/2}$ and $3d_{3/2}$ core levels in silver-based ASN devices exposed to (sample, red) and not exposed to (control, black) iodization procedures. The two peaks at 620 and 631 eV correspond to the expected I^- bands for $I_{3/d7}$.

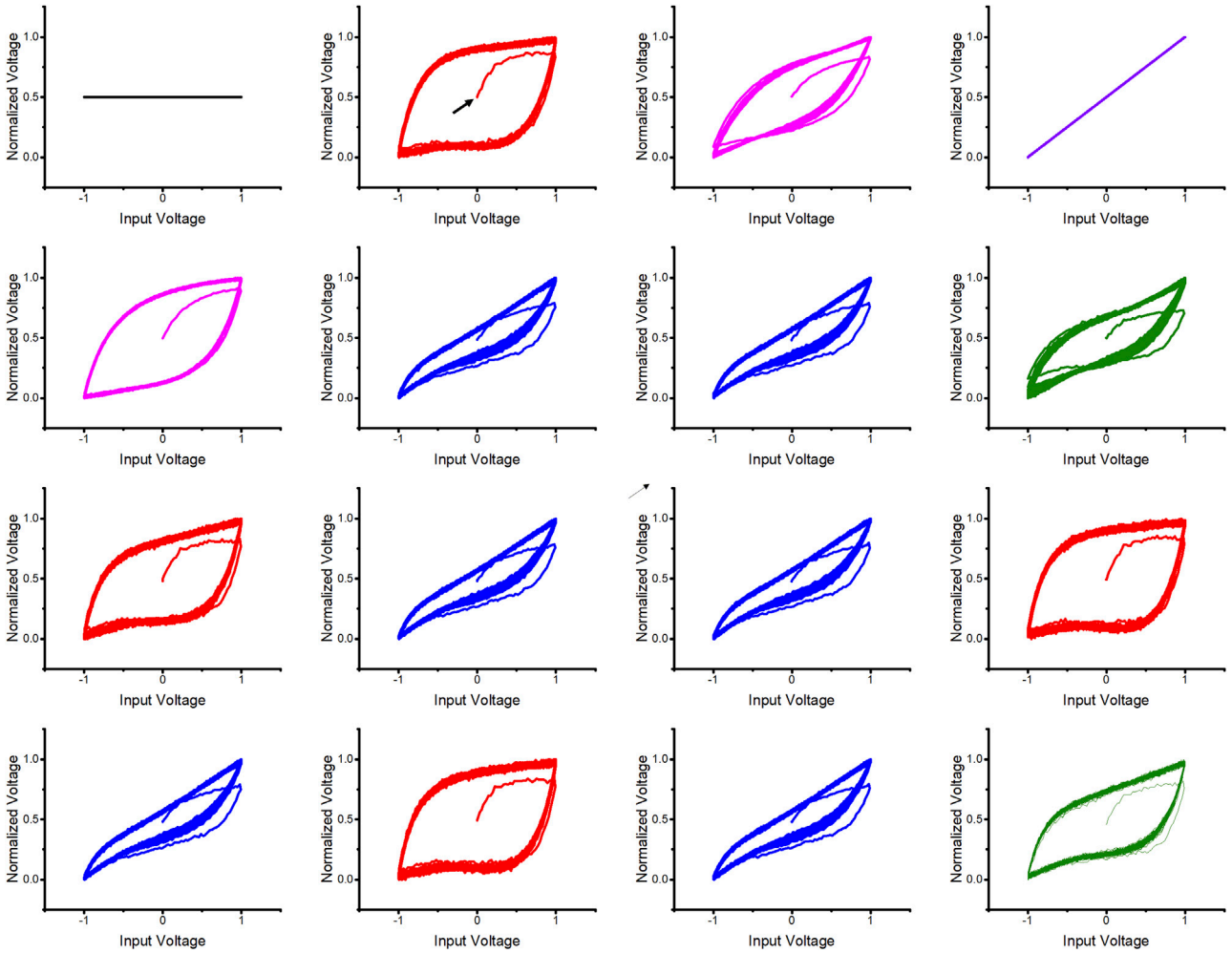


FIGURE 6 | Representative normalized Lissajous plots of all 16 electrodes measured simultaneously using a 7 Hz triangle waveform swept from −1 to +1 V over the course of 23 s with the grounded electrode **(top left)** and input signal **(top right)** recorded. Different colors correspond to different emergent dynamics spatially distributed throughout the network. The first sweep can be seen in all plots as indicated by the black arrow. The network demonstrated a spatially diverse system with reproducible, non-linear behavior distributed throughout the networks.

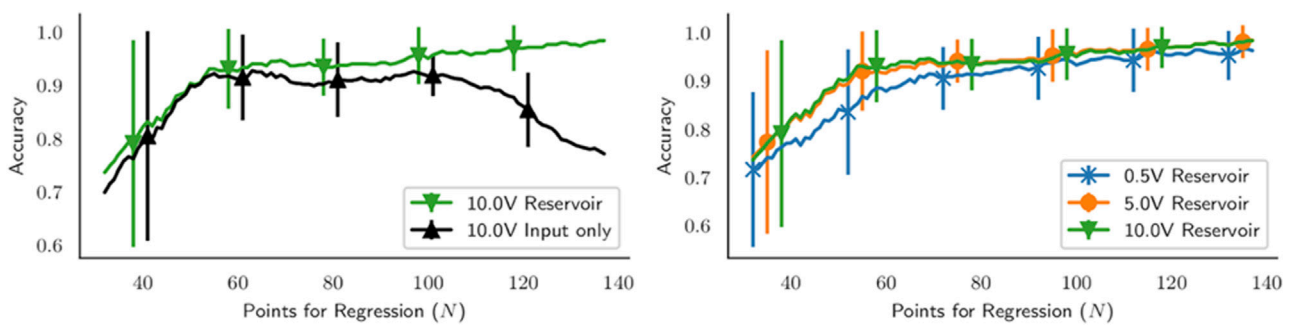


FIGURE 7 | (Left) Performance of the spoken digit classification task using AgI nanowire networks for in-materio RC to tap the temporal sequence of spoken digit MFCCs at N different points and regressing to identify the digit spoken. Mean accuracy and standard deviation clearly shows that the "Reservoir" readout method avoided overfitting and improved task performance as compared to using the "Input Only" mode **(Right)**. The input signal amplitude (voltage) was observed to have minimal. impact on accuracy, indicating the potential for maintaining task performance under low-power operation of AgI ASNs.

nature makes them uniquely suited to serve as physical substrates for hardware-based, in-materio computing. While the ever-increasing demands for computational capacity and complexity continue to challenge even the most advanced computing architectures, dynamical in-memory compute platforms such as the ASN may provide an alternative solution that is scalable, energy-efficient, adaptive, and capable of processing complex, time-varying data without the need for pre-programming or remote intervention. Expanding the catalog of memristive materials amenable to production of ASN-based devices, and thereby the diversity of network dynamics available for task performance, further increases their potential utility as a platform technology for next-generation computing applications. The new AgI-based ASN devices reported here served as a dynamic, memristive reservoir for the nonlinear transformation of temporal data and demonstrated the capacity to reliably classify spoken digits with high accuracy across a wide range of input voltages. Combined with the relative ease and low cost of the fabrication process, these AgI nanowire networks represent both a new material system that is ripe for future study and an opportunity to further develop the concept of in-materio computing toward real-world applications.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

SL and WW contributed equally to this work. AS, WW, and KS conceived and designed the experiments. SL, WW, KS, and CD performed experiments and analyzed data. All authors discussed the results and contributed to preparation of the manuscript. S.L, WW, KS, and AS wrote the manuscript. CT and JG reviewed and edited the manuscript.

FUNDING

This work was partially supported by the World Premier International Center for Materials Nanoarchitectonics (MANA) at the National Institute for Materials Science (Tsukuba, Japan). This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No (NSF grant number), and Semiconductor Research Corp. under Grant No (2015,209,024). The views expressed are those of the author(s) and do not reflect the official policy or position of the Department of Defense or the United States Government. Approved for public release, distribution is unlimited.

ACKNOWLEDGMENTS

The authors gratefully acknowledge past group members: Cristina Martin-Olmos, Henry Sillin and Audrius V. Avizienis for their foundation work in chip fabrication, LabView programming and crystal growth, respectively. Physical ASN chips were fabricated in the UCLA Nanofabrication Laboratory at the California Nanosystems Institute (CNSI). Instruments used in this study were operated and maintained by the Molecular Instrumentation Center at the Department of Chemistry and Biochemistry at UCLA.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnano.2021.675792/full#supplementary-material

REFERENCES

Alibart, F., Zamanidoost, E., and Strukov, D. B. (2013). Pattern Classification by Memristive Crossbar Circuits Using Ex Situ and In Situ Training. *Nat. Commun.* 4, 2072. doi:10.1038/ncomms3072

Avizienis, A. V., Sillin, H. O., Martin-Olmos, C., Shieh, H. H., Aono, M., Stieg, A. Z., et al. (2012). Neuromorphic Atomic Switch Networks. *PLoS One* 7–e42772. doi:10.1371/journal.pone.0042772

Bharathi Mohan, D., Sreejith, K., and Sunandana, C. S. (2007). Surface Plasmon-Exciton Transition in Ultra-thin Silver and Silver Iodide Films. *Appl. Phys. B* 89, 59–63. doi:10.1007/s00340-007-2768-6

Bürger, J., Goudarzi, A., Stefanovic, D., Stefanovic, D., and Teuscher, C. (2015). Computational Capacity and Energy Consumption of Complex Resistive Switch Networks. *AIMS Mater. Sci.* 2, 530–545. doi:10.3934/matersci.2015.4.530

Cai, K., Sun, J., Li, B., and Zhou, J. (2013). Hysteretic Current-Voltage Characteristics and Memristive Behaviors in AgI Nano-Particles Assembly. *ECS J. Solid State. Sci. Technol.* 2, N6–N10. doi:10.1149/2.017301jss

Calimera, A., Macii, E., and Poncino, M. (2013). The Human Brain Project and Neuromorphic Computing. *Funct. Neurol.* 28, 191–196. doi:10.11138/FNeur/2013.28.3.191

Dale, M., Miller, J. F., Stepney, S., and Adamatzky, A. (2017). Reservoir Computing as a Model for In-Materio Computing. *Adv. Unconv. Comput.* 1, 533–571. doi:10.1007/978-3-319-33924-5_22

Demis, E. C., Aguilera, R., Scharnhorst, K., Aono, M., Stieg, A. Z., and Gimzewski, J. K. (2016). Nanoarchitectonic Atomic Switch Networks for Unconventional Computing. *Jpn. J. Appl. Phys.* 55, 1102B2. doi:10.7567/jjap.55.1102b2

Deshmukh, A. M. (2020). Comparison of Hidden Markov Model and Recurrent Neural Network in Automatic Speech Recognition. *Ejers* 5, 958–965. doi:10.24018/ejers.2020.5.8.2077

Du, C., Cai, F., Zidan, M. A., Ma, W., Lee, S. H., and Lu, W. D. (2017). Reservoir Computing Using Dynamic Memristors for Temporal Information Processing. *Nat. Commun.* 8, 1–10. doi:10.1038/s41467-017-02337-y

Fu, K., Zhu, R., Loeffler, A., Hochstetter, J., Diaz-Alvarez, A., Stieg, A., Gimzewski, J., Nakayama, T., and Kuncic, Z. (2020). "Reservoir Computing with Neuromemristive Nanowire Networks," in Proceedings of the International Joint Conference on Neural Networks (. Institute of Electrical and Electronics Engineers Inc). doi:10.1109/IJCNN48605.2020.9207727

Gnanavel, M., and Sunandana, C. S. (2008). "Optical Absorption and Photoluminescence in Ultra Thin Silver and Silver Iodide Films," in 2008 IEEE PhotonicsGlobal at Singapore, doi:10.1109/IPGC.2008.4781352

Goudarzi, A., Lakin, M. R., and Stefanovic, D. (2014). *Reservoir Computing Approach to Robust Computation Using Unreliable Nanoscale Networks*. Berlin, Germany: Springer. arXiv.org. doi:10.1145/2770287.2770315

Guo, X., Schindler, C., Menzel, S., and Waser, R. (2007). Understanding the Switching-Off Mechanism in Ag+ Migration Based Resistively Switching Model Systems. *Appl. Phys. Lett.* 91, 133513. doi:10.1063/1.2793686

Harding, S., and Miller, J. F. (2009). "Evolution in Materio," in Encyclopedia Of Complexity And Systems Science *3220–3233*. New York: Springer, 3220–3233. doi:10.1007/978-0-387-30440-3_190

Hasegawa, T., Ohno, T., Terabe, K., Tsuruoka, T., Nakayama, T., Gimzewski, J. K., et al. (2010). Learning Abilities Achieved by a Single Solid-State Atomic Switch. *Adv. Mater.* 22, 1831–1834. doi:10.1002/adma.200903680

Hashmi, A., Berry, H., Temam, O., and Lipasti, M. (2011). "*Automatic Abstraction and Fault Tolerance in Cortical Microachitectures*," in Proceedings of the 38th Annual International Symposium on Computer Architecture, San Jose, CA, June 4–8, 2011. doi:10.1145/2000064.2000066

Ielmini, D., and Wong, H. S. P. (2018). In-memory Computing with Resistive Switching Devices. *Nat. Electron.* 1, 333–343. doi:10.1038/s41928-018-0092-2

Kan, S., Nakajima, K., Takeshima, Y., Asai, T., Kuwahara, Y., and Akai-Kasaya, M. (2021). Simple Reservoir Computing Capitalizing on the Nonlinear Response of Materials: Theory and Physical Implementations. *Phys. Rev. Appl.* 15, 024030. doi:10.1103/physrevapplied.15.024030

Kato, Y., Ono, L. K., Lee, M. V., Wang, S., Raga, S. R., and Qi, Y. (2015). Silver Iodide Formation in Methyl Ammonium Lead Iodide Perovskite Solar Cells with Silver Top Electrodes. *Adv. Mater. Inter.* 2, 1500195. doi:10.1002/admi.201570065

Konkoli, Z., Nichele, S., Dale, M., and Stepney, S. (2018). "Reservoir Computing with Computational Matter," in *Natural Computing Series*, Springer-Verlag, 269–293. doi:10.1007/978-3-319-65826-1_14

Kudithipudi, D., Saleh, Q., Merkel, C., Thesing, J., and Wysocki, B. (2016). Design and Analysis of a Neuromemristive Reservoir Computing Architecture for Biosignal Processing. *Front. Neurosci.* 9, 61907. doi:10.3389/fnins.2015.00502

Kuncic, Z., Kavehei, O., Zhu, R., Loeffler, A., Fu, K., Hochstetter, J., et al. (2020). Neuromorphic Information Processing with Nanowire Networks, in Institute of Electrical and Electronics Engineers (IEEE), 1–5. doi:10.1109/iscas45731.2020.9181034

Kuzum, D., Jeyasingh, R. G. D., Yu, S., and Wong, H.-S. P. (2012). Low-Energy Robust Neuromorphic Computation Using Synaptic Devices. *IEEE Trans. Electron. Devices* 59, 3489–3494. doi:10.1109/TED.2012.2217146

Liang, X. F., Chen, Y., Shi, L., Lin, J., Yin, J., and Liu, Z. G. (2007). Resistive Switching and Memory Effects of AgI Thin Film. *J. Phys. D: Appl. Phys.* 40, 4767–4770. doi:10.1088/0022-3727/40/16/004

Loppacher, C. H., Guggisberg, M., Pfeiffer, O., Meyer, E., Bammerlin, M., Lüthi, R., et al. (2003). Direct Determination of the Energy Required to Operate a Single Molecule Switch. *Phys. Rev. Lett.* 90, 4. doi:10.1103/physrevlett.90.066107

Lukoševičius, M., and Jaeger, H. (2009). Reservoir Computing Approaches to Recurrent Neural Network Training. *Comput. Sci. Rev.* 3, 127–149.

Lukoševičius, M., Jaeger, H., and Schrauwen, B. (2012). Reservoir Computing Trends. *KI-Künstliche Intelligenz* 26, 365–371. doi:10.1007/s13218-012-0204-5

Lukosevicius, M. (2011). *Reservoir Computing and Self-Organized Neural Hierarchies*. Bremen, Germany: IRC-Library, Information Resource Center der Jacobs University Breme.

Manning, H. G., Biswas, S., Holmes, J. D., and Boland, J. J. (2017). Nonpolar Resistive Switching in Ag@TiO2 Core-Shell Nanowires. *ACS Appl. Mater. Inter.* 9, 38959–38966. doi:10.1021/acsami.7b10666

Manning, H. G., et al. (2018). Emergence of Winner-Takes-All Connectivity Paths in Random Nanowire Networks. *Nat. Commun.* 9, 1–9. doi:10.1038/s41467-018-05517-6

Midya, R., Wang, Z., Asapu, S., Zhang, X., Rao, M., Song, W., et al. (2019). Reservoir Computing Using Diffusive Memristors. *Adv. Intell. Syst.* 1, 1900084. doi:10.1002/aisy.201900084

Miller, J. F., and Downing, K. (2002). "Evolution in Materio: Looking beyond the Silicon Box," in Proceedings - NASA/DoD Conference On Evolvable Hardware, EH Vols 2002-January. Piscataway, NJ: Institute of Electrical and Electronics Engineers Inc., 167–176.

Miller, J. F., Harding, S. L., and Tufte, G. (2014). Evolution-in-materio: Evolving Computation in Materials. *Evol. Intelligence*, 7 49–67.

Moon, J., Ma, W., Shin, J. H., Cai, F., Du, C., Lee, S. H., et al. (2019). Temporal Data Classification and Forecasting Using a Memristor-Based Reservoir Computing System. *Nat. Electron.* 2, 480–487. doi:10.1038/s41928-019-0313-3

Mustafa, M. K., Allen, T., and Appiah, K. (2019). A Comparative Review of Dynamic Neural Networks and Hidden Markov Model Methods for Mobile On-Device Speech Recognition. *Neural Comput. Applic* 31, 891–899. doi:10.1007/s00521-017-3028-2

Nakajima, K. (2020). Physical Reservoir Computing-An Introductory Perspective. *Jpn. J. Appl. Phys.* 59, 060501. doi:10.35848/1347-4065/ab8d4f

Nayak, A., Tamura, T., Tsuruoka, T., Terabe, K., Hosaka, S., Hasegawa, T., et al. (2010). Rate-Limiting Processes Determining the Switching Time in a Ag2S Atomic Switch. *J. Phys. Chem. Lett.* 1, 604–608. doi:10.1021/jz900375a

Sánta, B., Molnár, D., Haiber, P., Gubicza, A., Szilágyi, E., Zolnai, Z., et al. (2020). Nanosecond Resistive Switching in Ag/AgI/PtIr Nanojunctions. *Beilstein J. Nanotechnol.* 11, 92–100. doi:10.3762/bjnano.11.9

Sattar, A., Fostner, S., and Brown, S. A. (2013). Quantized Conductance and Switching in Percolating Nanoparticle Films. *Phys. Rev. Lett.* 111, 136808. doi:10.1103/physrevlett.111.136808

Scharnhorst, K., Woods, W., Teuscher, C., Stieg, A., and Gimzewski, J. (2017). "Non-Temporal Logic Performance of an Atomic Switch Network," in 2017 Proceedings of the IEEE/ACM International Symposium on Nanoscale Architectures (. Institute of Electrical and Electronics Engineers Inc.). NANOARCH 2017 133–138doi:10.1109/NANOARCH.2017.8053728

Schatz, T., and Feldman, N. H. (2018). "Neural Network vs. HMM Speech Recognition Systems as Models of Human Cross-Linguistic Phonetic Perception," in Proceedings of the Conference on Cognitive Computational Neuroscience http://kaldi-asr.org/. doi:10.32470/ccn.2018.1240-0

Schrauwen, B., Defour, J., Verstraeten, D., and Van Campenhout, J. (2007). The Introduction of Time-Scales in Reservoir Computing, Applied to Isolated Digits Recognition. *Proc. 15th Eur. Symp. Artif. Neural Networks.*, 471–479. doi:10.1007/978-3-540-74690-4_48

Sillin, H. O., Aguilera, R., Shieh, H. H., Avizienis, A. V., Aono, M., Stieg, A. Z., et al. (2013). A Theoretical and Experimental Study of Neuromorphic Atomic Switch Networks for Reservoir Computing. *Nanotechnology* 24. doi:10.1088/0957-4484/24/38/384004

Snyder, D. R., Goudarzi, A., and Teuscher, C. (2012). *Computational Capabilities of Random Automata Networks for Reservoir Computing*. College Park, MD: Phys Rev E Stat Nonlin Soft Matter Phys. arXiv.org. doi:10.7551/978-0-262-31050-5-ch035

Srinivasa, N., and Cruz-Albrecht, J. M. (2012). Neuromorphic Adaptive Plastic Scalable Electronics: Analog Learning Systems. *IEEE Pulse* 3, 51–56. doi:10.1109/mpul.2011.2175639

Stieg, A. Z., Avizienis, A. V., Sillin, H. O., Martin-Olmos, C.., Lam, M-L., Aono, M., et al. (2014). Self-organized Atomic Switch Networks. *Jpn. J. Appl. Phys.* 53. doi:10.7567/jjap.53.01aa02

Sun, W., Gao, B., Chi, M., Xia, Q., Yang, J. J., Qian, H., et al. (2019). Understanding Memristive Switching via In Situ Characterization and Device Modeling. *Nat. Commun.* 10, 3453. doi:10.1038/s41467-019-11411-6

Tanaka, G., Yamane, T., Héroux, J. B., Nakane, R., Kanazawa, N., Takeda, S., et al. (2019). Recent Advances in Physical Reservoir Computing: A Review. *Neural Netw.* 115, 100–123. doi:10.1016/j.neunet.2019.03.005

Tappertzhofen, S., Valov, I., and Waser, R. (2012). Quantum Conductance and Switching Kinetics of AgI-Based Microcrossbar Cells. *Nanotechnology* 23, 145703. doi:10.1088/0957-4484/23/14/145703

Terabe, K., Hasegawa, T., Nakayama, T., and Aono, M. (2005). Quantized Conductance Atomic Switch. *Nature* 433, 47–50. doi:10.1038/nature03190

Teuscher, C. (2017). The Weird, the Small, and the Uncontrollable: Redefining the Frontiers of Computing. *Computer* 50, 52–58. doi:10.1109/mc.2017.3001242

Türel, Ö., Lee, J. H., Ma, X., and Likharev, K. K. (2004). Neuromorphic Architectures for Nanoelectronic Circuits. *Int. J. Circ. Theor. Appl.* 32, 277–302. doi:10.1002/cta.282

Vandoorne, K., Fiers, M., Verstraeten, D., Schrauwen, B., Dambre, J., and Bienstman, P. (2010). "Photonic Reservoir Computing: A New Approach to Optical Information Processing," in 2010 12th International Conference on Transparent Optical Networks. doi:10.1109/ICTON.2010.5548990

Yang, J. J., Strukov, D. B., and Stewart, D. R. (2013). Memristive Devices for Computing. *Nat. Nanotech* 8, 13–24. doi:10.1038/nnano.2012.240

Yang, Y., Guo, P., Gaba, S., Chang, T., Pan, X., and Lu, W. (2012). Observation of Conducting Filament Growth in Nanoscale Resistive Memories. *Nat. Commun.* 3, 1–8. doi:10.1038/ncomms1737

Zhong, Y., Tang, J., Li, X., Gao, B., Qian, H., and Wu, H. (2021). Dynamic Memristor-Based Reservoir Computing for High-Efficiency Temporal Signal Processing. *Nat. Commun.* 12, 1–9. doi:10.1038/s41467-020-20692-1

Zhu, R., Hochstetter, J., Loeffler, A., Diaz-Alvarez, A., Stieg, A., Gimzewski, J., et al. (2020). "Harnessing Adaptive Dynamics in Neuro-Memristive Nanowire Networks for Transfer Learning," in 2020 International Conference on Rebooting Computing (ICRC). doi:10.1109/icrc2020.2020.00007

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.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 19 August 2021
doi: 10.3389/fnins.2021.709053

A Brain-Inspired Homeostatic Neuron Based on Phase-Change Memories for Efficient Neuromorphic Computing

*Irene Muñoz-Martin†, Stefano Bianchi†, Shahin Hashemkhani, Giacomo Pedretti, Octavian Melnic and Daniele Ielmini**

Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Politecnico di Milano, Milan, Italy

One of the main goals of neuromorphic computing is the implementation and design of systems capable of dynamic evolution with respect to their own experience. In biology, synaptic scaling is the homeostatic mechanism which controls the frequency of neural spikes within stable boundaries for improved learning activity. To introduce such control mechanism in a hardware spiking neural network (SNN), we present here a novel artificial neuron based on phase change memory (PCM) devices capable of internal regulation via homeostatic and plastic phenomena. We experimentally show that this mechanism increases the robustness of the system thus optimizing the multi-pattern learning under spike-timing-dependent plasticity (STDP). It also improves the continual learning capability of hybrid supervised-unsupervised convolutional neural networks (CNNs), in terms of both resilience and accuracy. Furthermore, the use of neurons capable of self-regulating their fire responsivity as a function of the PCM internal state enables the design of dynamic networks. In this scenario, we propose to use the PCM-based neurons to design bio-inspired recurrent networks for autonomous decision making in navigation tasks. The agent relies on neuronal spike-frequency adaptation (SFA) to explore the environment via penalties and rewards. Finally, we show that the conductance drift of the PCM devices, contrarily to the applications in neural network accelerators, can improve the overall energy efficiency of neuromorphic computing by implementing bio-plausible active forgetting.

Keywords: brain-inspired computing, unsupervised learning, reinforcement learning, spike-timing-dependent plasticity, hardware resilience, homeostatic scaling, synaptic scaling, phase change memory

OPEN ACCESS

Edited by:
Huanglong Li,
Tsinghua University, China

Reviewed by:
Ming Xu,
Huazhong University of Science and Technology, China
SangBum Kim,
Seoul National University, South Korea

***Correspondence:**
Daniele Ielmini
daniele.ielmini@polimi.it

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Neuromorphic Engineering, a section of the journal Frontiers in Neuroscience

Received: *13 May 2021*
Accepted: *27 July 2021*
Published: *19 August 2021*

Citation:
Muñoz-Martin I, Bianchi S, Hashemkhani S, Pedretti G, Melnic O and Ielmini D (2021) A Brain-Inspired Homeostatic Neuron Based on Phase-Change Memories for Efficient Neuromorphic Computing. Front. Neurosci. 15:709053. doi: 10.3389/fnins.2021.709053

1. INTRODUCTION

The field of artificial intelligence (AI) has recently seen significant breakthroughs in the research, showing high performance in several tasks such as image recognition, natural language processing and playing games (Collobert et al., 2011; Krizhevsky et al., 2012; Mikolov et al., 2012; Silver et al., 2016). The most widespread approach to AI has focused on deep learning, where the intelligent systems are trained via specific algorithms such as backpropagation (LeCun et al., 2015). However, the pre-tuning of the training parameters, which requires time and power intensive procedures, deprives the systems of the plastic adaptation to the environment which, on the other hand, is one

of the fundamental properties of the biological organisms. This lack of resilience with respect to a constantly changing environment is what actually hinders the current AI to achieve human-like accuracy in daily-life tasks (Parisi et al., 2019).

Biological organisms collect, settle and modulate the information relying on specific mechanisms of synaptic plasticity and neural activity (Turrigiano, 1999). In particular, the learning procedure is usually explained in terms of Hebbian-type plasticity, where the time correlation between the pre-synaptic and post-synaptic spikes induces variations of the synaptic weights (Fox and Stryker, 2017; Lisman, 2017), as in spike-timing-dependent plasticity (STDP) (Masquelier and Thorpe, 2007). On the other hand, Hebbian learning cannot completely describe the learning procedure of the brain, since the only STDP theory foresees a continual synaptic potentiation and depression as a consequence of the correlation between the neuronal responses and the corresponding inputs (Miller and MacKay, 2008). In fact, biological systems adopt homeostatic regulation to keep the overall neuronal and synaptic activities within safe boundaries, which also helps to counteract unwanted changes of the firing rate due to external perturbations (Turrigiano, 1999). In this framework, the synaptic scaling, or homeostatic scaling (Turrigiano, 2008), refers to the biological mechanism able to counteract a chronically high firing rate of a population of neurons. Thus, Hebbian learning and homeostatic regulation sustain each other for the optimization of experience-based knowledge toward continual adaptation of real-life information (Abraham and Robins, 2005; Zenke et al., 2017).

Experience-based knowledge, where agents learn a behavioral policy by interacting with the world and consequently receiving penalties and rewards, is a scientific field shared between neuroscience and computer science known as "reinforcement learning" (Kaelbling et al., 1996). One of the leading reinforcement mechanism is associated with dopamine, a pleasure-related neurotransmitter, which is released in the brain when a person succeeds in solving a problem (Schultz et al., 1997). In the literature, several approaches have been proposed to facilitate reinforcement learning. For instance, reinforcement techniques have been shown to enable the learning of optimized behavioral policy for a given model of the space, where the agent continually looks for the maximization of the reward thus acquiring an accurate mapping of the environment (Sutton, 1988). However, in real life, an agent must build its own model by incremental experience of positive and negative events, as studied by model-free methods such as (i) Q-learning (Watkins and Dayan, 1992) and (ii) temporal difference learning, TD(λ) (Doya, 2000). In particular, in the last few years, such cognitive functions have been widely discussed in the framework of attractor neural networks for the key role of cognitive functions, such as context dependent decision making (Doya, 2000; Kuzum et al., 2012), thus gaining momentum as viable networks to replicate human-like behaviors (Chicca et al., 2014).

The combination of the benefits introduced by homeostatic mechanism and reinforcement learning would thus improve the artificial intelligence systems toward the ability to autonomously interact with the environment in real life situations.

In this framework, several neuromorphic spiking neural networks (SNNs) based on CMOS technology have been proposed, demonstrating VLSI synaptic circuits with homeostatic neurons (Bartolozzi and Indiveri, 2006; Chicca et al., 2014; Qiao et al., 2017) and reward-based decision-making circuits (Wunderlich et al., 2019; Yan et al., 2019). At the same time, non-volatile memory devices, such as phase change memory (PCM), have raised considerable interest as promising synaptic connections for neuromorphic computation, thanks to the 3D stacking capability, the low-voltage operation and the ability to serve as embedded non-volatile memory in computing systems (Suri et al., 2012; Xu et al., 2020; Ren et al., 2021). In particular, PCMs have recently demonstrated outstanding multi-level capability (Kuzum et al., 2013; Ren et al., 2021), which enables continual learning in neural networks (Bianchi et al., 2019; Muñoz-Martín et al., 2019) and decision making in brain-inspired cognitive systems (Eryilmaz et al., 2014).

In this work, we present a novel artificial integrate-and-fire (I&F) neuron based on PCM devices implementing homeostatic mechanisms. In particular, the gradual crystallization of a PCM device enables the continual tuning of the internal threshold of the neuron as a function of the level of firing excitation. This adaptation process improves the learning capability and directly translates in hardware the homeostatic control mechanism that manages the synaptic weight update during STDP. We show that the homeostatic neuron can optimize the pattern specialization of large images, e.g., those taken from the Fashion-MNIST dataset, while enabling high robustness against errors and external perturbations (Muñoz-Martín et al., 2020). In this framework, we propose the use of PCM-based homeostatic neurons for achieving continual learning in standard convolutional neural network. We also analyze the impact of device programming failure in relation to the multilevel capability of the PCM devices. The impact of PCM conductance drift is also studied (Suri et al., 2012; Xu et al., 2020; Ren et al., 2021), demonstrating that this device non-ideality could implement bio-inspired features, such as active forgetting. Finally, we propose a novel bio-inspired recurrent neural network (RNN) capable of solving reinforcement learning tasks. The internal state of each neuron of the RNN is mapped by the self-adaptive threshold using a PCM device, which modulates, as before, the firing excitability. The more the neuron fires, the more the control PCM conductance increases, thus mapping the dynamic behavior of the network in real time (Bianchi et al., 2020b). In this work, the recurrent PCM device enables the study of several reinforcement learning tasks such as decision making during autonomous navigation, with particular attention in terms of power-efficiency. This work highlights the importance of PCM devices as key elements to achieve adaptation, learning and autonomous navigation exploiting the benefits of local edge computing.

2. BIO-INSPIRED LEARNING IN ARTIFICIAL NEURAL NETWORKS

Figure 1A shows a schematic illustration of spike-frequency adaptation (SFA) in a neuronal cell. When a signal excites a

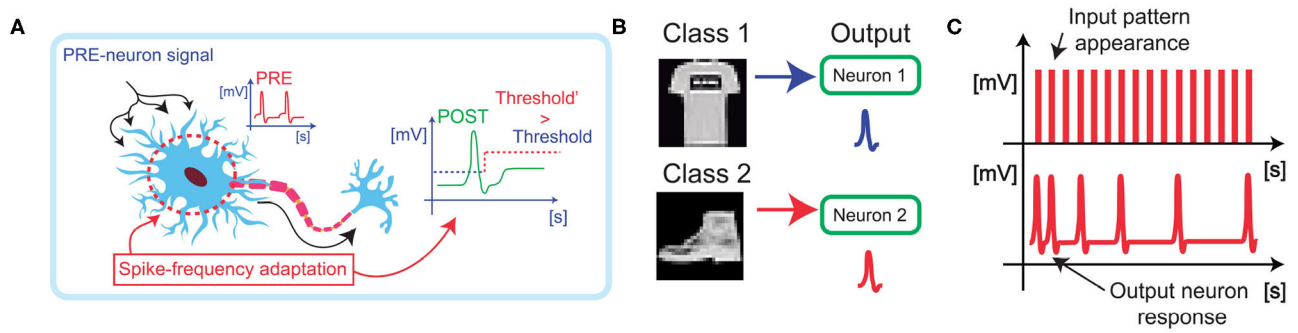


FIGURE 1 | (A) Biological neurons are stimulated by spikes coming from the synaptic connections and modulate their response in frequency as a function of the spiking activity. **(B)** By implementing the spike frequency adaptation in hardware, it is possible to introduce a boundary for the learning activity of large images, e.g., from the Fashion-MNIST dataset, thus boosting the overall specialization accuracy. **(C)** Furthermore, the specialization of the output neuron results in a decrease the firing activity of the neuron, thus optimizing the energy consumption.

neuron, the output firing rate is balanced between an increase due to the synaptic potentiation and a decrease due to the homeostatic mechanism (Indiveri et al., 2011). In synaptic learning processes, this threshold regulation aims at stabilizing the learning activity and limiting the growth of the synaptic weights, thus enabling low energy consumption and better accuracy of classification.

The homeostatic adaptation has been studied in the case of a winner-take-all (WTA) network for the classification of large images. The output homeostatic neurons (POSTs) must specialize on different classes of images presented at the input of the WTA, **Figure 1B**, thus enabling the spike-frequency adaptive mechanism that limits the power consumption and enables efficient classification (**Figure 1C**; Pedretti et al., 2018). Classification is achieved by using both excitatory synapses, which evolve by increasing or decreasing the conductance accordingly to STDP, and inhibitory synapses, which prevent the same specialization on different patterns by discharging the integration at each POST firing activity (Bianchi et al., 2020a). Synaptic excitatory dynamics are reproduced by using PCM devices switching from low resistive state (LRS) to high resistive state (HRS), and vice versa. Potentiation is achieved when the POST fires after the pre-neurons (PREs), while depression is achieved when the POST fires before the PRE (Bianchi et al., 2020c).

2.1. Hardware Realization of the Homeostatic Neuron

Figure 2 illustrates the artificial neuron circuit, where the threshold is managed by a control PCM directly connected to the comparator which compares the membrane potential with the threshold. PCM devices typically show multilevel storage with a large number of analog conductance states (Kim et al., 2019). In **Figure 2**, the multilevel behavior is obtained by the applications of repeated set pulses to the top electrode for gradual crystallization or amorphization, thus causing a modulation of the neuronal threshold (Suri et al., 2011; Wright et al., 2013; Tuma et al., 2016).

The incoming PRE spikes are weighted by PCM synapses which induce a synaptic current collected by the "integration" block in **Figure 2**. The synaptic current spikes are integrated until the internal potential hits the threshold of the neuron. This event causes the generation of two spikes, namely (i) a POST spike which is applied to the next layer of neurons, and (ii) a second spike which is applied to the top electrode of the internal PCM device to induce partial crystallization, which is responsible for a self-threshold regulation. Each crystallization pulse leads to an incremental set transition of the PCM device to higher conductive values G_{PCM}. The PCM conductance is the leading element setting the responsivity of the neurons since it maps the fire threshold V_{TH} of the neuron. In particular, V_{TH} is obtained as the read current of the PCM biased at negative values ($V_{read} < 0$) after conversion by the trans-impedance amplifier of **Figure 2**, namely $V_{TH} = -R_L G_{PCM} V_{read}$, where R_L is the feedback resistance and G_{PCM} also includes the conductance of the series transistor M_1. Initially, the PCM device is prepared in the HRS, thus resulting in low current I_C and low threshold voltage V_{TH}. As the POST fires, the incremental crystallization of the PCM causes the increase of the threshold with respect to the first reference firing value. The gradual crystallization procedure is thus iterated at every POST fire, causing a continuous increase of V_{TH}. As a result, more input spikes are needed to induce the fire of the neuron or, equivalently, the spiking frequency of the POST decreases at increasing crystallization of the control PCM.

2.2. Characteristics of the PCM Devices

The PCM is programmed by set (with current I_{SET}) and reset signal pulses as shown in **Figure 3A**. **Figure 3B** shows the cumulative distribution of the LRS and HRS resistances after the application of the programming signals, with two orders of magnitude of resistive window. On the other hand, note that the PCM shows a gradual increase of conductance which suitably reproduces the adaptive threshold regulation of V_{TH}. In particular, the variation of LRS distributions can be modulated by proper choice of I_{PULSE}, thus enabling multilevel states. The multilevel behavior of the PCMs can be obtained by both starting from a full LRS and applying incremental amorphizing pulses,

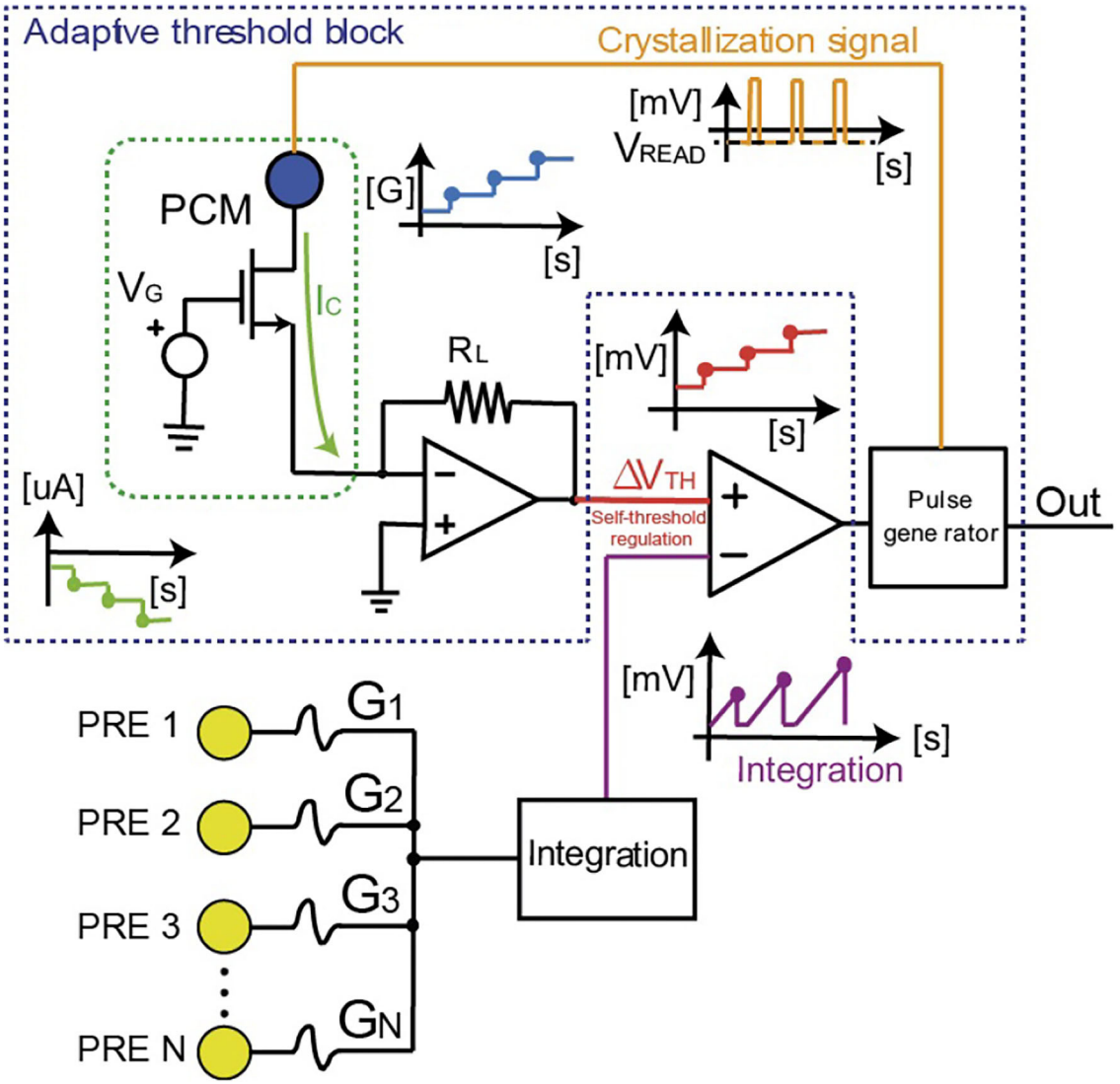


FIGURE 2 | Scheme of the homeostatic neuron with the control PCM device which regulates the internal threshold. The spike signals coming from other neurons (e.g., pre-synaptic neurons) are integrated ("Integration" block) using an Arduino microcontroller (2 or Mega2560 in the measurements we performed). Arduino also manages the fire activity when the threshold of the neuron is overcome. When this happens, two signals are generated: (i) the "Out" response of the neuron and (ii) the crystallization pulse for the gradual increase of the PCM conductance. In this way the internal threshold V_{TH} of the neuron increases.

as indicated in **Figure 3A**, or from a partial HRS and applying crystallizing pulses. Note that the crystallization depends on both the amplitude and duration of the pulses. In general, G_{PCM} is more easily modulated by using shorter pulses and intermediate set voltages. In this way, the conductive multilevel states can be spread over one order of magnitude, thus enabling the possibility of effective modulation of the threshold (Wong et al., 2010).

Note that the PCM resistance suffers from the conductance drift in time, which is due to the structural relaxation of the device (Kim et al., 2019). **Figures 3C–E** illustrate the time evolution of three different resistance distributions. Experimental data show that the conductance drift is higher for higher initial resistances, thus obtaining a non-linear increase in time of the initial programmed conductive value if the device is not continuously re-programmed. Such variation in time of the synaptic weights implemented with PCM devices is a key limitation for the design of neural accelerators (Kim et al., 2019; Joshi et al., 2020). The progressive decrease of the conductance also affects the homeostatic mechanism. However, the drift can also have a beneficial effect in our bio-inspired neuron, since it gives the possibility of spontaneous forgetting. In fact, the threshold of the neuron naturally decreases during drift, thus increasing the neuronal firing excitability and enabling an active forgetting mechanism.

Note that the PCM devices can be also programmed in multilevel states by applying repetitive voltage rectangular pulses,

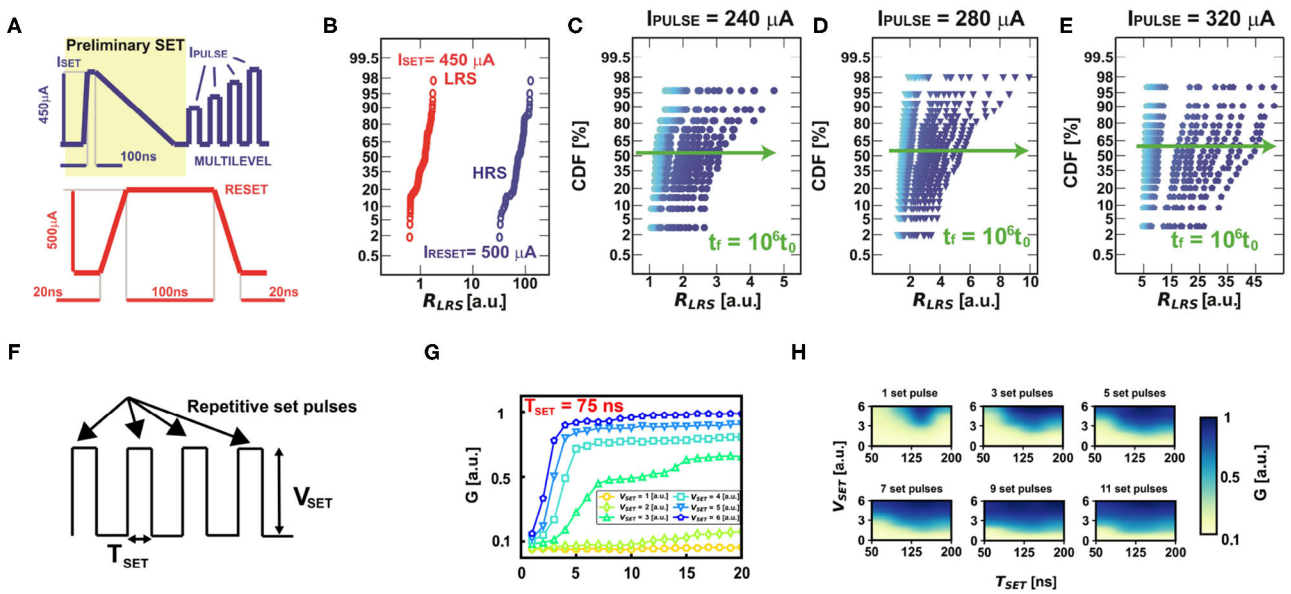


FIGURE 3 | (A) Typical current programming signals used to set and reset the PCM device to low resistive state (LRS) and high resistive state (HRS), respectively. **(B)** LRS and HRS experimental distributions of the PCM devices. **(C–E)** Distributions of three different programmed multilevel resistive states to highlight the effect of the conductance drift in time. Note that the conductance drift is more evident at higher initial values of resistances. **(F)** Voltage-based rectangular programming pulses for achieving multilevel resistive states starting from a partial HRS. **(G)** Multilevel characteristics at different set voltages as a function of the number of rectangular pulses for a pulse duration T_{SET} = 75 ns. **(H)** Experimental color maps varying the number of pulses for achieving multilevel states as a function of the rectangular pulse amplitude and duration.

as highlighted in **Figure 3F** starting from a partial HRS. In particular, it is possible to modulate the number of multilevel states by proper choice of the voltage amplitude V_{SET} at fixed pulse duration T_{SET}, as highlighted in **Figure 3G** for T_{SET} = 75 ns. Note also that it is possible to have a modulation of the resistive states at various combination of duration and amplitude of the repetitive programming pulses, as depicted in **Figure 3H**, thus giving rise to an extensive resistive modulation as a function of the target programming condition. This is very important for the development of neuromorphic and neural networks with PCM-based homeostatic neurons, as it is going to be analyzed in the following.

3. UNSUPERVISED STDP WITH HOMEOSTATIC MECHANISM

To study the properties of the homeostatic neuron with respect to the classification accuracy of input images, we designed a spiking neural network capable of unsupervised learning by STDP. The input patterns are submitted asynchronously, which means that not all the patterns are presented with fixed density and shape to the network. Note also that the input signal consists of an alternation of the asynchronous pattern and random noise spikes, where noise, used for background depression, has lower density and input appearance probability in order to assure circuital and learning stability during operation (Bianchi et al., 2020c). **Figure 4A** illustrates the SNN, where PCM synapses have 1-transistor/1-resistor (1T1R) structure with the gates of the transistors connected by wordlines (WLs) and the PCM top electrodes connected by bitlines (BLs). The bitlines are directly linked to the neurons, since the feedback neuronal signal is used to adjust the synaptic weights involved in the STDP protocol (Ambrogio et al., 2016a). Thus, with respect to **Figure 2**, which represents the main structure of the homeostatic neuron, a further signal line is needed for the unsupervised learning with STDP. Input spikes are applied to the WLs to induce synaptic currents that are summed at each column to feed the I&F POSTs with self-adaptive threshold, according to the scheme of **Figure 2**. The feedback spike consists of a set pulse of voltage V_{TE}, followed by a pulse of reset voltage. The overlap between the PRE spike and the POST spike induces potentiation (set transition) or depression (reset transition) for positive or negative delay between the two spikes (Bianchi et al., 2020c). During potentiation the synaptic element switches to LRS, while during depression the synaptic element switches to HRS. Thus, the STDP is mapped in a binary framework, which enables simpler hardware computation with respect to bio-inspired analog STDP (Bianchi et al., 2020c). Note that an extra column of PCM synapses programmed in the HRS is used to discriminate pattern and noise, i.e., in particular, spike integration is enabled only for the presentation of an input pattern, to prevent a decay of the overall accuracy due to noise (Ambrogio et al., 2016b).

Figure 4B shows the measured weights of the 16 PCM synapses, divided in pattern synapses and background synapses which were not stimulated by input pattern spikes. Once the internal potential overcomes the threshold V_{TH}, the POST generates a spike, thus enabling the synaptic

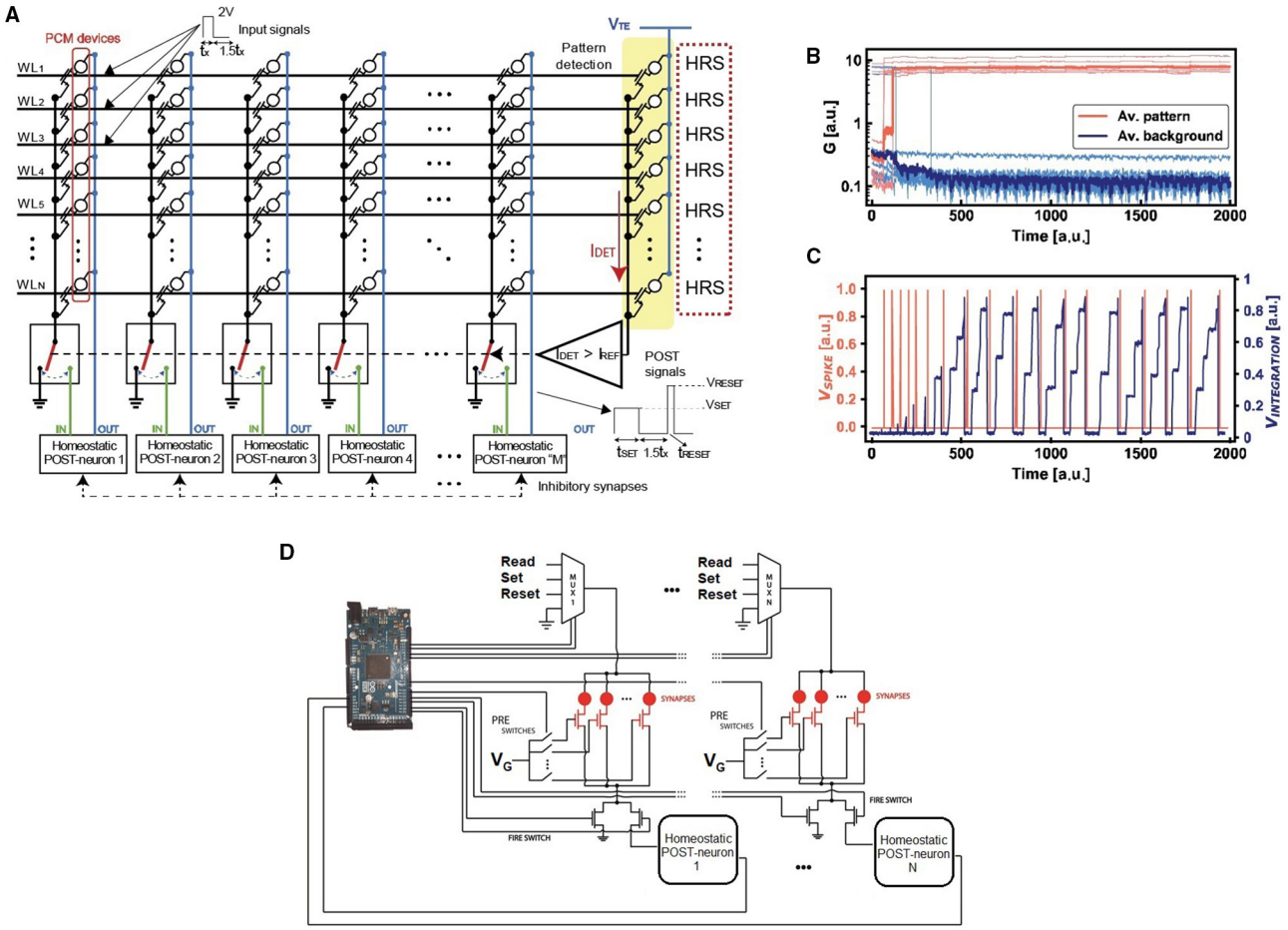


FIGURE 4 | (A) Experimental setup for the asynchronous STDP, where the inputs are submitted at the wordlines (WLs). Every column connects the WLs to a specific POST by using 1T1R PCM excitatory synapses, while the inhibitory synapses are implemented via discharge signals of the integrators. A further column of HRS devices is used for pattern/noise detection. The integration activity of each neuron is enabled only for $I_{DET} > I_{REF}$. **(B)** Experimental evolution of the pattern and background synapses under STDP. Note that the inputs are presented to the network asynchronously, since the potentiation and depression are gradual. **(C)** At every firing activity, the internal PCM device of the neuron is incrementally set thus obtaining an overall reduction of the spiking frequency due to the increase of the internal threshold of the neuron. This spike frequency adaptation enables optimized pattern specialization and reduced energy consumption. **(D)** Schematic representation of the experimental setup with several POST-synaptic neurons in order to implement a WTA network. Note that the microcontroller (we used both Arduino 2 and Arduino Mega2560) acts as master of the system.

potentiation/depression (depending on the PRE/POST spike delay) and the increase of the homeostatic PCM conductance. In turn, the PCM conductance increase causes the increase of V_{TH}, hence the homeostatic control mechanism. This is evidenced by the decreased POST spiking frequency in **Figure 4C**, which ensures an improved energy efficiency of the SNN. The integration is disabled when the POST fires in order to avoid the integration of set/reset pulses to prevent excessive charge storage in the integrator block of **Figure 2**.

Figure 4D shows a simplified schematic to explain the management of the homeostatic neuron for the STDP measurements in a WTA network. An Arduino 2 (or Mega2560) microcontroller acts as master of the whole setup, managing both the gate voltages and the proper top electrode biases of the synaptic elements implemented with PCMs. The microcontroller also manages the results of the integration signal with respect to the adaptive internal thresholds of the homeostatic neurons. Note also that, at fire, the multiplexers enable the passage of the top electrode voltage of the synapses in order to implement the STDP learning paradigm.

3.1. Fashion-MNIST Accuracy and Robustness

To study the effect of homeostatic scaling on multi-pattern unsupervised learning, we simulated our SNN for the average classification of images from the Fashion-MNIST dataset, characterized by 10 different classes of clothes. **Figure 5** shows the confusion matrices from Monte Carlo simulations for the learning accuracies without homeostasis (**Figure 5A**) and with homeostasis (**Figure 5B**). The study is carried out by

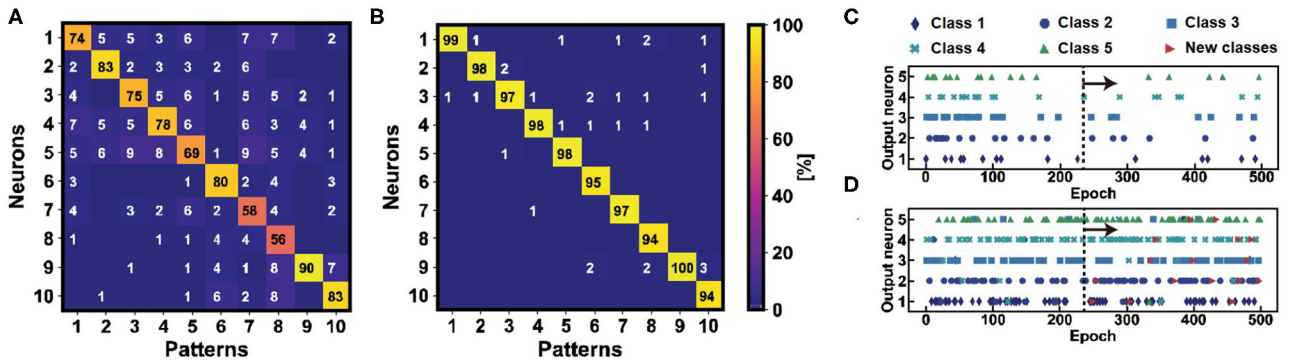


FIGURE 5 | Confusion matrices for the study of the average accuracy of the learning activities for the 10 classes of the Fashion-MNIST training dataset without **(A)** and with **(B)** homeostasis. The learning accuracy highlights a high and stable tendency only when homeostatic neurons are used as post-synaptic neurons in the WTA network, reaching a value of 97%. **(C)** Fire activities of 5 homeostatic and non-homeostatic **(D)** neurons in 500 epochs of pattern and noise presentations. The homeostatic neurons are robust against "false" patterns presentations from another dataset, e.g., MNIST (here submitted after the 250*th* epoch).

considering one image for each of the 10 classes of the training dataset, replicating the study for the available 60,000 images and implementing the WTA protocol with a single-layer perceptron of 784 input neurons and 10 output neurons for each case (Ambrogio et al., 2016a). The learning accuracies are then averaged for each class to assess the overall efficiency. Homeostatic scaling allows for an accuracy increase by about 20% on average for the pattern specialization during learning of ten different images from the Fashion-MNIST dataset, which highlights the importance for unsupervised learning of PCM-based adaptive threshold. Such adaptive mechanism is also fundamental for achieving better accuracy in deep neural networks, where the homeostatic scaling improves the neuronal specialization for a pattern of a specific class of the dataset (Martin et al., 2020). The improvement of the accuracy can be directly referred to the better specialization achieved by the control PCM device which assures an optimized threshold level for each specific neuronal spiking activity. In fact, the homeostatic mechanism allows to exceed the threshold only when the learnt pattern appears at the input. Note that, thanks to the additional bitline of **Figure 4A** used for pattern/noise detection, the low-density inputs are neglected, thus avoiding spurious firing activity.

Homeostatic scaling also improves the robustness of the network for the classification when external perturbations, such as disturbs, errors or false patterns from other datasets, are presented at the input. To test the classification robustness of the network, **Figure 5** show the output neuronal spikes during the classification of five images from Fashion-MNIST with homeostasis (**Figure 5C**) and without homeostasis (**Figure 5D**). In the first phase of the experiment, five images from Fashion-MNIST are presented and classified. In this phase, the non-homeostatic neurons show some errors due to the lack of a dedicated "specialization," while no significant errors are evident among the homeostatic neurons. In the second phase of the experiment, handwritten digit patterns from the MNIST dataset are presented along with the Fashion-MNIST patterns. The homeostatic neurons do not show erroneous spikes since they have been specialized on the Fashion-MNIST patterns during the previous learning procedure. On the other hand, the non-homeostatic neurons show spurious spikes in correspondence of the presentations of the false patterns, due to the fact that the similarity between the patterns of the two datasets is sufficient to induce a false fire. Such behavior is avoided using the threshold modulation mechanism which allows to set a specific threshold for a specific learnt pattern, thus highlighting the higher classification robustness thanks to the homeostatic scaling procedure.

3.2. Active Forgetting by Conductance Drift

The PCM device is programmed by set pulses (with current I_{SET}) and reset transitions. The variation of the resistive distributions can be modulated by incremental application of pulsed signals at the top electrode of the device, thus enabling multilevel states. These states are affected by conductance drift if the device is not constantly re-programmed in time. During standard STDP procedures, the conductance drift does not affect the overall behavior of the network, since the devices are continually set and reset in the pattern and background positions. Similarly, the internal state used to calibrate the threshold does not suffer too, since the drift effect is not appreciable in the reference timescale, as already seen in **Figure 4C**.

STDP has been recently used in the final classification layer of deep convolutional networks for achieving continual learning (Muñoz-Martín et al., 2019). In this kind of neural networks, the convolutional filters generate responses which constitute artificial patterns that are learnt and classified afterwards via unsupervised WTA STDP. This procedure enables the incremental learning of new patterns during inference, since the convolutional filters give (for the new classes) a combination of responses which is original with respect to the others. However, since the variability among the new artificial patterns is high there is the possibility of having neurons which commit errors, specializing on input patterns that are unlikely to appear again at the input of the WTA STDP. In this situation, the internal PCM device is not activated for a long time, thus causing a decrease of

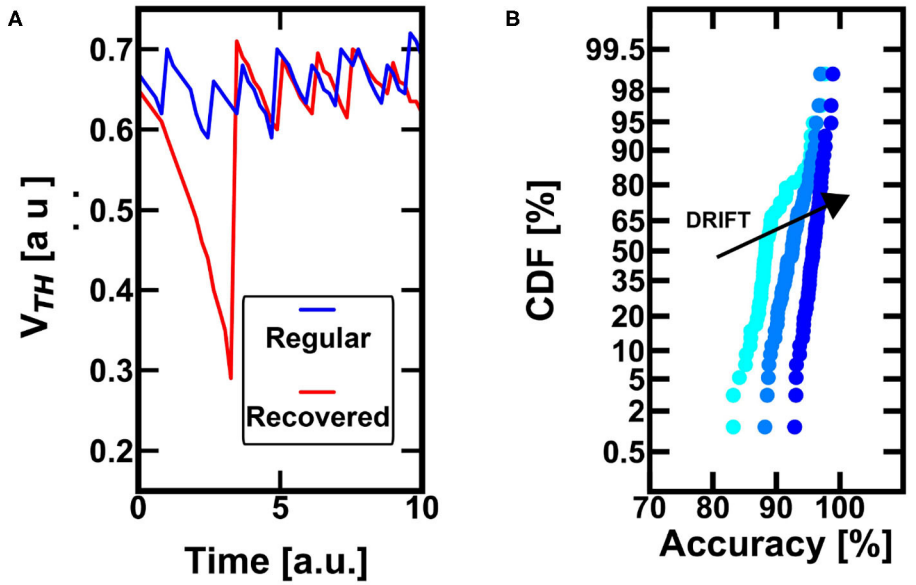


FIGURE 6 | (A) The conductive drift leads to a substantial decrease of the threshold whenever the neuron is not excited (and the device is not reprogrammed), red line. This behavior well fits the bio-inspired forgetting and can lead to the recovery of a wrong spiking specialization toward improved classification. On the other hand, blue line, if the neuron is regularly excited (even if not often in time), the drift effect does not lead to active forgetting. **(B)** The conductance drift of the PCM devices has a positive effect for the recovery of neurons which committed error during the classification, such as neurons that have specialized on "wrong" patterns. At increasing drift of the control PCM device, the internal threshold gets progressively smaller, and the neuron is induced to fire again to the presentation of another pattern (eventually the "good" one). This favorable scenario is due to the fact the pattern information is correlated in time, while the errors are not. Thus, the drift effect can recover the error and increase the probability of accurate spiking activity in time.

the threshold, as shown by the Monte Carlo simulations in **Figure 6A**. Here, in particular, you can see that a regular spiking activity continually adjusts the threshold of the device, thus avoiding the lowering of the threshold. On the other hand, once a spurious spike activity is taken into consideration (red line), the internal threshold decreases considerably in time, since the spurious firing activity is not correlated. Note that such behavior can induce a neuron to change specialization, since the reduction of the threshold is proportional to an increase of neuronal excitability.

Furthermore, the conductance drift in time could be directly referred to the bio-plausible active forgetting, which erases previously stored information as a complementary procedure with respect to the homeostatic scaling consolidation (Davis and Zhong, 2017). Such active scaling forgetting gets rid of the unwanted pattern specialization and allows for a further specialization neuron able to be dedicated to more likely patterns at the input. **Figure 6B** shows the Monte Carlo simulations of the probability of recovering a past incorrect spiking event toward a fair accurate specialization at decreasing threshold conditions. In particular, it is evident that, increasing the conductance drift in time, it is possible to increase the firing excitability too. This is very relevant, since an incorrect specialization due to an uncorrelated error can be recovered by the correct excitation of a time-correlated input (i.e., a pattern), which is far more probable to contribute to the firing activity. Note that the presented figure is referred only to previously misunderstood firing activities, that are the only cases for which the drift plays a positive role.

4. HOMEOSTATIC NEURON IN RECURRENT NEURAL NETWORKS

The bio-inspired spike-frequency adaptation modulates the fire excitability of a neuron inside a neural network. In other words, the fire responsivity directly depends on the past specialization history of the network. Such behavior along a temporal sequence is the key element for the recurrent neural networks (RNN) which can be thus re-designed taking advantage of the SFA mechanism (Amit, 1989).

To support the spike-frequency adaptation of the neurons for reinforcement learning tasks, we considered a free-model decision-making test where an agent has to move in an environment until it finds a global reward. In particular, we considered the navigation problem of **Figure 7A**, where an agent explores the maze via penalties and rewards until it is successful in finding the escape path. In this case study, each point of the environment is configured as a homeostatic neuron which modulates its internal state as a function of the firing history of that particular position inside the environment. In particular, the reward is given when the agent reaches the prize causing the decrease of the internal threshold of the rewarded positions, while the punishment arises when the agent touches a barrier causing the increase of the internal threshold (Frémaux et al., 2013). Once the agent finds the escape path, it starts to remember the successful way by progressive rewards, i.e., the internal thresholds of successive positions decrease. Thus, the network evolves relying only

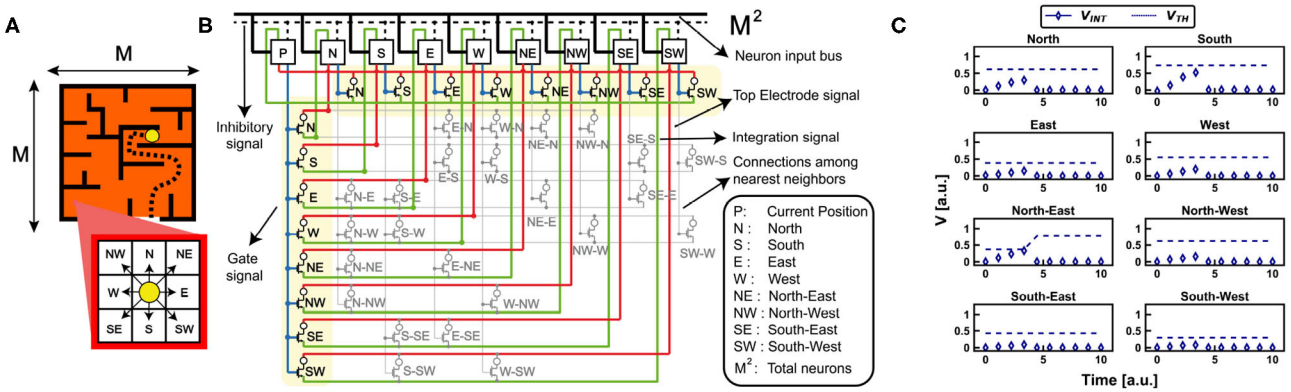


FIGURE 7 | (A) Case study maze for the investigation of the reinforcement learning capabilities of the bio-inspired RNN. **(B)** The synapses of the RNN link symmetrically each neuron to and from the nearest neighbors. At every position P, the neuron sends a signal to the synaptic gates of its neighbors. When one neuron integrates enough current to overcome its internal threshold, it fires and inhibits all the network. Every firing activity maps the movement of the agent. The schematic of the circuit also shows the connections among the nearest neighbors. **(C)** Experimental measurement for a single movement of the agent inside the case study maze. The nearest neurons start to integrate current until one (North-East neuron) overcomes the threshold and fires. Note that the fire activity of the neuron causes an increase of the PCM internal threshold.

on the self-adaptive threshold mechanism of reward and penalty and on the synaptic plasticity, without any further external aid.

We addressed the problem of a maze of size MxM (M = 30) by a brain-inspired RNN with M^2 self-adaptive neurons, where each neuron represents a position within the maze. **Figure 7B** shows a section of the RNN limited to the current position P and the eight nearest neighbors, which map the eight fundamental cardinal directions. Note that the RNN is completely symmetrical, since each connection between the current position and one of the adjacent is configured by two symmetric synapses to and from P. Each synapse has a 1T1R structure where the PCM device is randomly initialized in HRS or LRS. Note that the further synapses connecting the nearest neurons also contribute to the definition of a symmetric matrix with respect to the diagonal of the RNN. Synaptic weights along the diagonal are all zero because a neuron, i.e., a position, is not self-connected. Note that an inhibitory signal enables a WTA algorithm, as already described in the first section of this manuscript.

4.1. The Movement of the Agent

The environmental boundaries are initially defined by programming the thresholds of each position. The goal of the network is to find the escape route across the maze via reinforcement learning, thus supporting the relevance of the PCM plastic properties for typical neuromorphic abilities (Frémaux et al., 2013).

At any time, only the occupied neuron P is activated by external spike stimulation. The firing activity of the neuron P induces two types of event: first, the threshold V_{TH} of neuron P increases, due to the homeostatic mechanism; second, nearest neighbor neurons are stimulated by the spiking activity of neuron P. This dynamics was experimentally validated by the RNN with PCM neurons and synapses of **Figure 7B**, where each neuron is connected to the nearest neighbor positions, e.g., E is connected to P, NE and SE. **Figure 7C** shows the measured internal potential V_{INT} for the eight nearest neurons during stimulation of neuron P with an external spiking signal of limited duration. Since all synapses are initially programmed in random state (i.e., 50% in LRS, 50% in HRS), only those neurons which are connected by synapses with relatively high conductance show substantial current integration. Once the first neuron reaches the threshold, namely neuron North-East in the example of **Figure 7C**, the agent moves to the corresponding position and a new cycle can be started by zeroing the internal potential V_{INT} of all the neurons (i.e., the typical inhibitory signal already discussed for the WTA network). Note that, as the agent position changes, the synaptic weights must be reinitialized to enable trial-to-trial variations of the random walk, thus boosting the effect of penalties and rewards. Note also that the self-adaptive threshold mechanism induces partial crystallization of the control PCM of the firing neuron, thus preventing the agent to come back to previously occupied positions. In fact, as visible in **Figure 7C**, once a neuron fires it increases its internal threshold, thus making less probable the coming back to that position from the surrounding ones during the next movements of the agent.

4.2. Penalty/Reward Mechanisms and Optimization of the Solution

Figure 8A shows the random walks of the agent during successive trials. Each experiment is limited in time, since the agent has to find the reward by elaborating a strategy, rather than testing each single position (Frémaux et al., 2013). If the agent cannot escape within 400 spikes, (i.e., steps of the agent), a new trial starts by reinitializing the agent position and the synaptic weights. The reinforcement learning is instead retained from trial to trial and only relies on (i) penalties, when the agent touches a wall, or (ii) rewards, when the escape paths is found. Both penalties and rewards are mapped by acting on the internal V_{TH} of the neuron, thus increasing or decreasing the neuronal responsivity. When

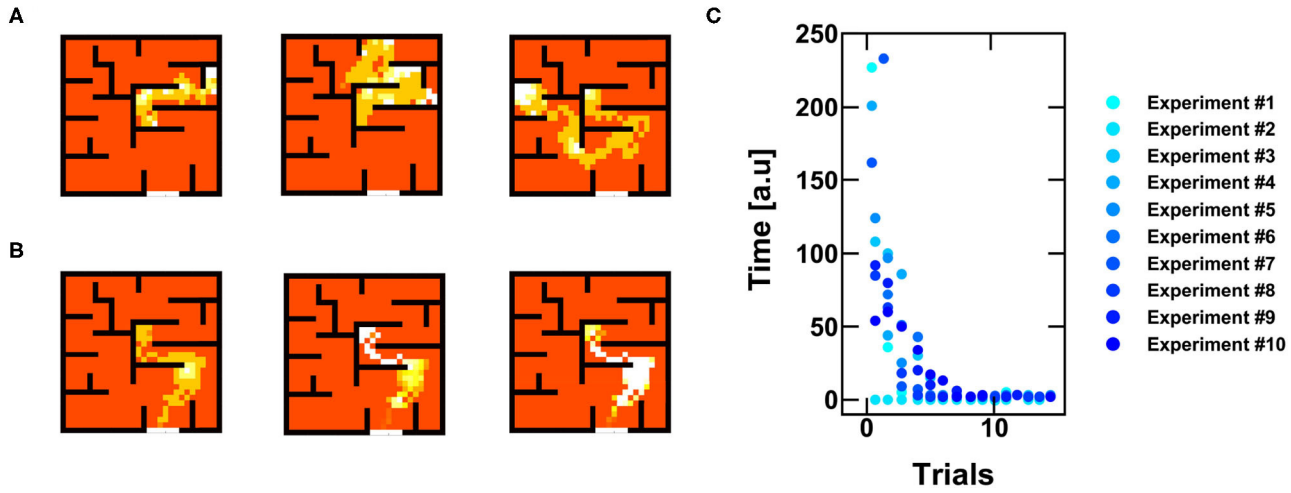


FIGURE 8 | (A) Example of three random walks for successive trials of exploration of the agent. **(B)** Example of three successive trials after finding the escape path: the agent progressively improves its policy for finding the reward, eventually not committing errors anymore. **(C)** Time to get to the escape path as a function of the number of trials for 10 different experiments of 15 trials each.

the agent touches a wall, a penalty is assigned to that position by increasing the corresponding V_{TH}. On the other hand, when the agent finds the escape path, a reward is given by lowering the V_{TH} of the last positions occupied by the agent.

As shown in **Figure 8B**, once the escape path has been disclosed, the system tends to follow the preferential path toward the objective. This happens because the reward policy introduces a positive feedback, which reduces the V_{TH} of the path thus improving the preference of the agent to follow the escape path. **Figure 8C** shows the time to find the reward as a function of successive trials. Note that the reward has two main effects, namely (i), the system self-optimizes its policy map by increasing the time efficiency, and (ii) the spiking activities concentrate in the positions close to the target, thus reducing any unwanted energy consumption along ordinary positions which do not give any reward. As a result, the experience-based evolution of our RNN relies on PCM-based neurons and synaptic plasticity and enables the optimization of reinforcement learning for autonomous decision-making navigation.

4.3. Impact of Drift on Reinforcement Learning

To study the impact of the drift, we studied the effect of the drift-induced decrease of the internal neuronal threshold in **Figure 9A**. The decrease of the internal threshold causes a decrease of the necessary time to get to the final reward for each trial. On the other hand, the drift also affects the threshold of the punished neurons, but the drift does not drive such positions to a condition comparable with the ordinary ones.

The difference between the reinforcement learning with and without PCM drift decreases at increasing trial of specialization, since the reward facilitate the identification of the successful path by acting on the threshold of the corresponding positions (less integration time per single step is needed to follow the rewarded path). **Figure 9B** shows the accuracy (i.e., the ability of finding the escape path considering a fixed number of trials per experiment) over 1,000 Monte Carlo simulations. The study indicates that the drift of the PCMs increases the error probability, lowering the overall efficiency of the network. As a result, drift does not introduce significant benefit in the case of reinforcement learning, with respect to the STDP learning. In more complex situations, where the surrounding boundaries change continuously thus requiring a constant reconfiguration by the agent, the drift-induced forgetting mechanism could become favorable, since it would boost the quest toward other points of the environment.

4.4. Energetic Efficiency

The energy efficiency of reinforcement learning can be improved by operating the devices in burst-mode (Bianchi et al., 2019), which consists of the application of fast pulsed signals at the electrodes of the PCM devices, thus enabling a consistent reduction of the required energy per single operation. In our simulations, we stimulated the devices with pulsed signal with duration of 100 ns separated by silent periods of 10 μs as shown in **Figure 10A**.

Figure 10B shows the average energy per single exploration trial of the agent, indicating that the energy consumption decreases as the agent refines its strategy. During the initial trials, the energy consumption due to integration needed to explore the environment is larger than the other contributions, since the agent requires many steps to explore the surroundings. Once the final reward is achieved, the integration procedure requires less energy, thanks to the threshold decrease in the path positions close to the objective. Note also that the simulation without drift indicates a higher integration energy, which is due to the fact that the internal states undergo a decrease of the respective threshold due to conductance drift, thus requiring less power per

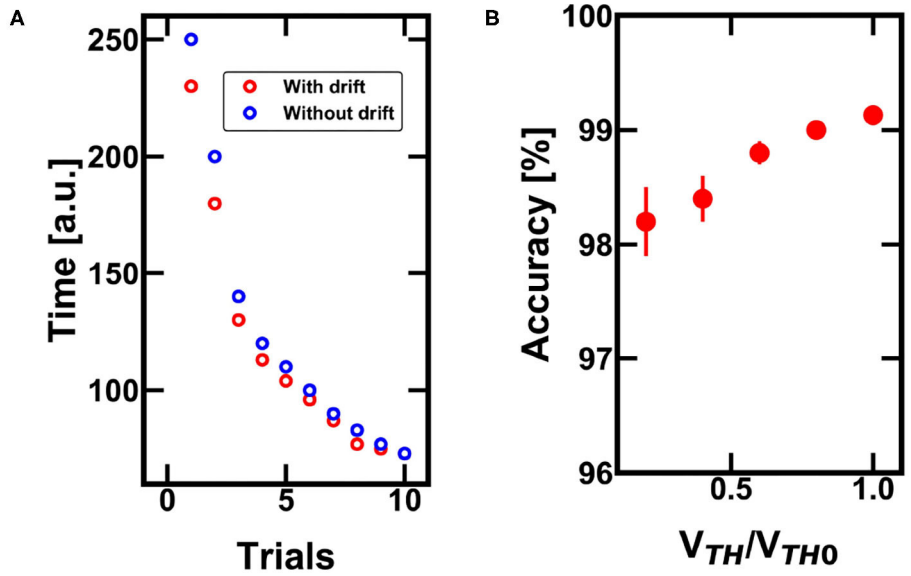


FIGURE 9 | (A) Monte Carlo simulations of the minimum time needed to successfully find the escape path with and without the drift effect of the PCM devices. The larger the drift, the lower the time to get to the final reward. **(B)** Impact of the drift on the accuracy for finding the escape path over 1,000 trials of the same experiment. Note that the drift is not a benefit since the decreasing V_{TH} (with respect to the nominal V_{TH0}) can lead to misunderstanding in the policy map definition.

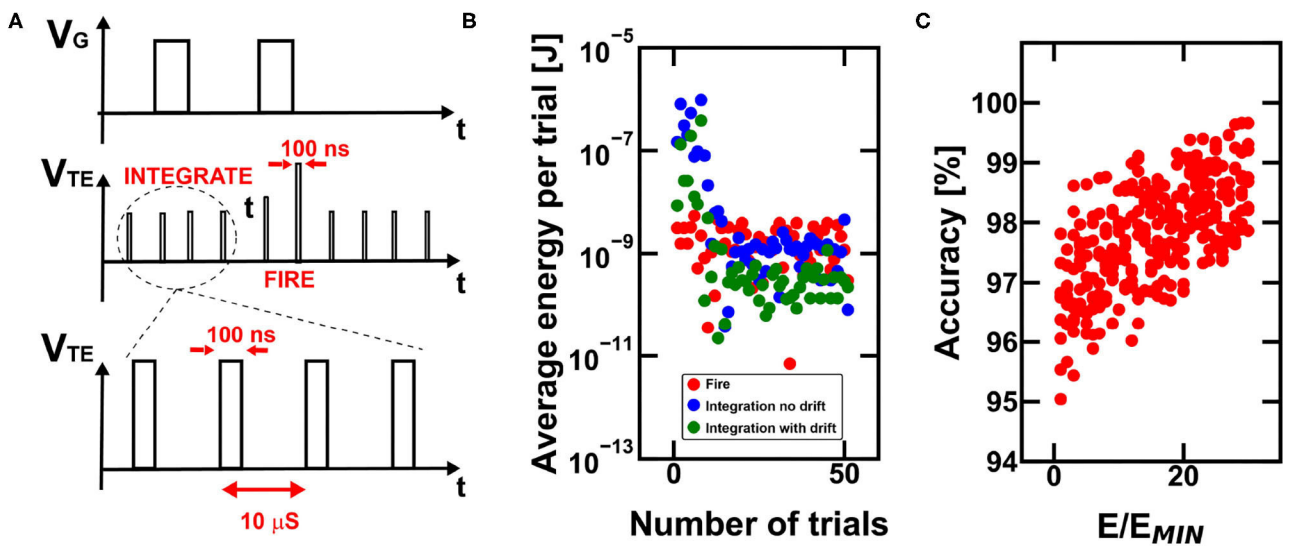


FIGURE 10 | (A) Burst-mode operation for power-saving during PCM-based working procedure of the RNN. **(B)** Note that the required energy for the operations carried out by the RNN is dependent on the grade of specialization of the network and on the final achievement with respect to the disclosure of the escape path. In fact, once the final reward is found, the network progressively decreases the total need of integration energy. Note that the simulated energy consumption reduction also comes with a decrease in the overall accuracy for finding the escape path when conductance drift is considered. **(C)** Monte Carlo simulations of the global accuracy for the case study maze considering increasing trial and error procedures for the programming of the internal state and of the inter-neuronal synaptic devices.

single trial. The energy consumption decrease, as well as the time decrease to get to the solution, depends on the timescale of the reinforcement learning execution in hardware, since longer times means larger conductance drift.

Figure 10C shows the accuracy for finding the reward as a function of the number of memory access per single device (e.g., the PCM internal state of the neurons) in order to assure the theoretical conductance value assessed during the simulations. However, a 30 times higher energy consumption for best programming condition only improves the accuracy by 1.5%, on average. This result indicates the substantial robustness and efficiency of

bio-inspired neuromorphic computing for reinforcement learning tasks.

5. CONTINUAL LEARNING IN ARTIFICIAL NEURAL NETWORKS

STDP-based unsupervised learning with homeostatic neurons is a robust approach for achieving continual learning in artificial neural networks. In particular, STDP has been already introduced in the last layer of convolutional neural networks (CNNs) in order to get resilience in neural systems trained with the backpropagation algorithm (Muñoz-Martín et al., 2019). These kinds of hybrid supervised/unsupervised neural networks rely on custom training algorithms to extract, after convolution, single-bit responses per each filter relative to a found/not found trained feature, as illustrated in **Figure 11**. After convolution, a novel feature map arises, which is then classified by means of post-synaptic neurons under the STDP learning paradigm. In order to study the effect of the introduction of PCM-based SFA neurons in this neural system, we built a WTA network with ten POSTs capable of spike frequency adaptation, as in **Figure 2**, and inhibitory signals. The inhibition, in particular, enables the drop of the internal potential of all the neurons when a fire event occurs (Pedretti et al., 2017; Bianchi et al., 2020c).

The use of neurons with SFA control mechanism in the last layer of the network of **Figure 11** introduces robustness

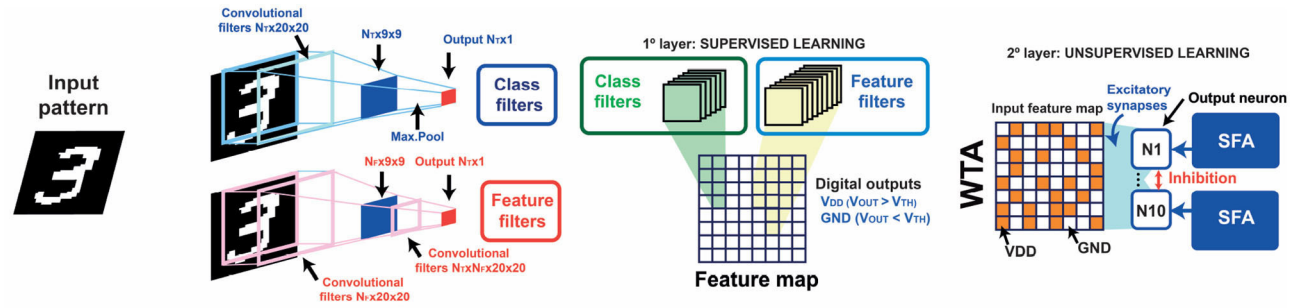


FIGURE 11 | Schematic architecture of the hybrid supervised-unsupervised neural network. The input patterns coming from the dataset are convolved with pre-trained convolutional filters. Each filter, which can recognize a generic feature, "feature-filter," or a specific class, "class filter," gives a single-bit response (found/not found response). The responses of the convolutional filters give thus rise to a binary feature map, which is then classified by homeostatic neurons using the STDP paradigm in the WTA architecture.

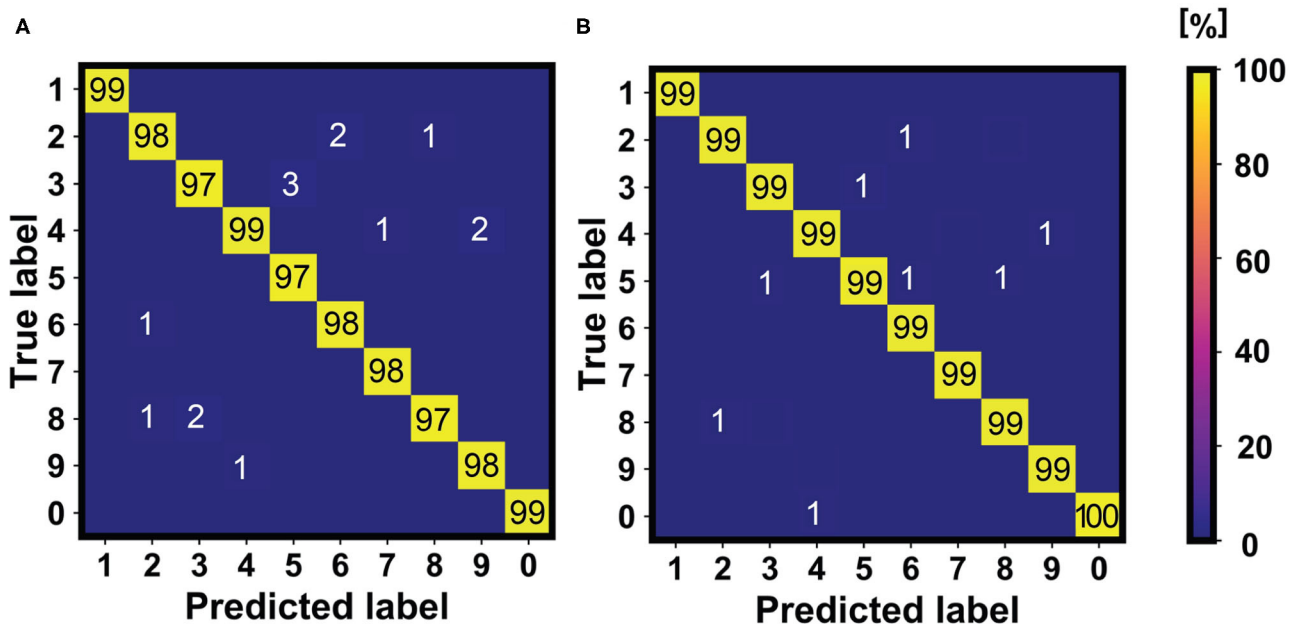


FIGURE 12 | Comparison of the accuracies of previous works (Muñoz-Martín et al., 2019) **(A)** with the accuracies obtainable after using homeostatic neurons in the last layer of the network **(B)**. Note that the accuracy results increase in the second case, which is mainly due to the improved specialization capability and to the active forgetting mechanism introduced by the SFA homeostatic neurons.

and improved accuracy with respect to previous works, as reported in **Figure 12** for the inference of the MNIST dataset (10,000 patterns of handwritten digits). This is due to two main contributions, namely: (i) the improved specialization capability of the neurons to get specialized on specific input patterns (each neuron modulates its internal threshold on a specific feature map arising from the patterns joining the same class, as also studied in **Figure 5**); (ii) errors in the WTA classification are prone to be corrected thanks to the spontaneous forgetting mechanism studied in **Figure 6**. This latter point, in particular, is due to the fact the classification errors are not correlated in time, thus driving a wrong fire event to be forgotten in time.

Thus, the homeostatic neurons appear as key elements to introduce both resilience and accuracy in artificial neural networks, paving the way for the next technological steps of artificial intelligent computation.

6. CONCLUSIONS

In this work we introduced a novel artificial neuron based on phase change memory (PCM) devices capable of internal regulation via homeostatic and plastic procedures. The neuron relies on the definition of the internal threshold by multilevel programming of the control PCM devices, thus enabling the specialization of large patterns and the continual learning capability of CNNs by introducing the STDP procedure in a supervised framework. The novel neuron is also used to introduce a bio-inspired recurrent neural network which directly creates a directed experienced-graph in time by keeping trace of the fire history of each neuron of the network. Such recurrent connections based on neurons capable of spike frequency adaptation demonstrate decision-making capabilities for navigation tasks. Furthermore, we show that conductance drift of the PCM devices can be used to emulate active forgetting in neural networks. This work supports the suitability of PCM devices for the optimization of synaptic dynamics and the implementation of brain-inspired computing in artificial intelligence.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

AUTHOR CONTRIBUTIONS

IM-M and SB have contributed equally in the planning, design and implementation of the system, the extraction and the interpretation of the results, the figures realization, and the text writing. SH, GP, and OM have contributed to the experimental setup. DI has supervised the planning and the design of this project. All authors contributed to the article and approved the submitted version.

FUNDING

This article has received fundings from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 648635) and from the Italian Minister for University and Research (grant agreement No. R164TYLBZP).

REFERENCES

Abraham, W. C., and Robins, A. (2005). Memory retention-the synaptic stability versus plasticity dilemma. *Trends Neurosci*. 28, 73–78. doi: 10.1016/j.tins.2004.12.003

Ambrogio, S., Balatti, S., Milo, V., Carboni, R., Wang, Z.-Q., Calderoni, A., et al. (2016a). Neuromorphic learning and recognition with one-transistor-one-resistor synapses and bistable metal oxide RRAM. *IEEE Trans. Electron Devices* 63, 1508–1515. doi: 10.1109/TED.2016.2526647

Ambrogio, S., Ciocchini, N., Laudato, M., Milo, V., Pirovano, A., Fantini, P., et al. (2016b). Unsupervised learning by spike timing dependent plasticity in phase change memory (PCM) synapses. *Front. Neurosci*. 10:56. doi: 10.3389/fnins.2016.00056

Amit, D. J. (1989). *Modeling Brain Function: The World of Attractor Neural Networks*. Cambridge University Press. doi: 10.1017/CBO9780511623257

Bartolozzi, C., and Indiveri, G. (2006). "Silicon synaptic homeostasis," in *Brain Inspired Cognitive Systems, BICS 2006* (Island of Lesvos).

Bianchi, S., Martín, I. M., and Ielmini, D. (2020a). Bio-inspired techniques in a fully digital approach for lifelong learning. *Front. Neurosci*. 14, 379–393. doi: 10.3389/fnins.2020.00379

Bianchi, S., Muñoz-Martín, I., Hashemkhani, S., Pedretti, G., and Ielmini, D. (2020b). "A bio-inspired recurrent neural network with self-adaptive neurons and PCM synapses for solving reinforcement learning tasks," in *2020 IEEE International Symposium on Circuits and Systems (ISCAS)* (Seville). doi: 10.1109/ISCAS45731.2020.9181103

Bianchi, S., Muñoz-Martin, I., Pedretti, G., Melnic, O., Ambrogio, S., and Ielmini, D. (2019). "Energy-efficient continual learning in hybrid supervised-unsupervised neural networks with PCM synapses," in *2019 Symposium on VLSI Technology* (Kyoto), T172–T173. doi: 10.23919/VLSIT.2019.8776559

Bianchi, S., Pedretti, G., Muñoz-Martín, I., Calderoni, A., Ramaswamy, N., Ambrogio, S., et al. (2020c). A compact model for stochastic spike-timing-dependent plasticity (STDP) based on resistive switching memory (RRAM) synapses. *IEEE Trans. Electron Devices* 67, 2800–2806. doi: 10.1109/TED.2020.2992386

Chicca, E., Stefanini, F., Bartolozzi, C., and Indiveri, G. (2014). Neuromorphic electronic circuits for building autonomous cognitive systems. *Proc. IEEE* 102, 1367–1388. doi: 10.1109/JPROC.2014.2313954

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. (2011). Natural language processing (almost) from scratch. *J. Mach. Learn. Res*. 12, 2493–2537. doi: 10.5555/1953048.2078186

Davis, R., and Zhong, Y. (2017). The biology of forgetting—a perspective. *Neuron* 95, 490–503. doi: 10.1016/j.neuron.2017.05.039

Doya, K. (2000). Reinforcement learning in continuous time and space. *Neural Comput*. 12, 219–245. doi: 10.1162/089976600300015961

Eryilmaz, S. B., Kuzum, D., Jeyasingh, R., Kim, S., BrightSky, M., Lam, C., et al. (2014). Brain-like associative learning using a nanoscale non-volatile phase change synaptic device array. *Front. Neurosci*. 8:205. doi: 10.3389/fnins.2014.00205

Fox, K. D., and Stryker, M. P. (2017). Integrating hebbian and homeostatic plasticity: introduction. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20160413. doi: 10.1098/rstb.2016.0413

Frémaux, N., Sprekeler, H., and Gerstner, W. (2013). Reinforcement learning using a continuous time actor-critic framework with spiking neurons. *PLoS Comput. Biol.* 9:e1003024. doi: 10.1371/journal.pcbi.1003024

Indiveri, G., Linares-Barranco, B., Hamilton, T. J., van Schaik, A., Etienne-Cummings, R., Delbruck, T., et al. (2011). Neuromorphic silicon neuron circuits. *Front. Neurosci.* 5:73. doi: 10.3389/fnins.2011.00073

Joshi, V., Gallo, M. L., Haefeli, S., Boybt, I., Nandakumar, S. R., Piveteau, C., et al. (2020). Accurate deep neural network inference using computational phase-change memory. *Nat. Commun.* 11:2473. doi: 10.1038/s41467-020-16108-9

Kaelbling, L. P., Littman, M. L., and Moore, A. W. (1996). Reinforcement learning: a survey. *J. Artif. Intell. Res.* 4, 237–285. doi: 10.1613/jair.301

Kim, W., Bruce, R. L., Masuda, T., Fraczak, G. W., Gong, N., Adusumilli, P., et al. (2019). "Confined PCM-based analog synaptic devices offering low resistance-drift and 1000 programmable states for deep learning," in *2019 Symposium on VLSI Technology* (Kyoto), T66–T67. doi: 10.23919/VLSIT.2019.8776551

Krizhevsky, A., Sutskever, I., and Hinton, G. (2012). "Imagenet classification with deep convolutional neural networks," in *Neural Information Processing Systems* (Lake Tahoe), 25.

Kuzum, D., Jeyasingh, R. G. D., Yu, S., and Wong, H. P. (2012). Low-energy robust neuromorphic computation using synaptic devices. *IEEE Trans. Electron Dev.* 59, 3489–3494. doi: 10.1109/TED.2012.2217146

Kuzum, D., Yu, S., and Wong, H.-S. P. (2013). Synaptic electronics: materials, devices and applications. *Nanotechnology* 24:382001. doi: 10.1088/0957-4484/24/38/382001

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Lisman, J. (2017). Glutamatergic synapses are structurally and biochemically complex because of multiple plasticity processes: long-term potentiation, long-term depression, short-term potentiation and scaling. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20160260. doi: 10.1098/rstb.2016.0260

Martin, I. M., Bianchi, S., Covi, E., Piccolboni, G., Bricalli, A., Regev, A., et al. (2020). "A SiOx RRAM-based hardware with spike frequency adaptation for power-saving continual learning in convolutional neural networks," in *2020 Symposia on VLSI Technology and Circuits* (Honolulu, HI). doi: 10.1109/VLSITechnology18217.2020.9265072

Masquelier, T., and Thorpe, S. J. (2007). Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Comput. Biol.* 372:e30031. doi: 10.1371/journal.pcbi.0030031

Mikolov, T., Deoras, A., Povey, D., Burget, L., and Cernocky, J. (2012). "Strategies for training large scale neural network language models," in *2011 IEEE Workshop on Automatic Speech Recognition and Understanding* (Waikoloa, HI). doi: 10.1109/ASRU.2011.6163930

Miller, K. D., and MacKay, D. J. C. (2008). The role of constraints in Hebbian learning. *Neural Comput.* 6, 100–126. doi: 10.1162/neco.1994.6.1.100

Muñoz-Martín, I., Bianchi, S., Hashemkhani, S., Pedretti, G., and Ielmini, D. (2020). "Hardware implementation of PCM-based neurons with self-regulating threshold for homeostatic scaling in unsupervised learning," in *2020 IEEE International Symposium on Circuits and Systems (ISCAS)* (Seville). doi: 10.1109/ISCAS45731.2020.9181033

Muñoz-Martín, I., Bianchi, S., Pedretti, G., Melnic, O., Ambrogio, S., and Ielmini, D. (2019). Unsupervised learning to overcome catastrophic forgetting in neural networks. *IEEE J. Explor. Solid-State Comput. Dev. Circ.* 5, 58–66. doi: 10.1109/JXCDC.2019.2911135

Parisi, G., Kemker, R., Part, J., Kanan, C., and Wermter, S. (2019). Continual lifelong learning with neural networks: a review. *Neural Netw.* 113, 54–71. doi: 10.1016/j.neunet.2019.01.012

Pedretti, G., Milo, V., Ambrogio, S., Carboni, R., Bianchi, S., Calderoni, A., et al. (2017). Memristive neural network for on-line learning and tracking with brain-inspired spike timing dependent plasticity. *Sci. Rep.* 7:5288. doi: 10.1038/s41598-017-05480-0

Pedretti, G., Milo, V., Ambrogio, S., Carboni, R., Bianchi, S., Calderoni, A., et al. (2018). Stochastic learning in neuromorphic hardware via spike timing dependent plasticity with RRAM synapses. *IEEE J. Emerg. Select. Top. Circ. Syst.* 8, 77–85. doi: 10.1109/JETCAS.2017.2773124

Qiao, N., Bartolozzi, C., and Indiveri, G. (2017). An ultralow leakage synaptic scaling homeostatic plasticity circuit with configurable time scales up to 100 ks. *IEEE Trans. Biomed. Circ. Syst.* 11, 1271–1277. doi: 10.1109/TBCAS.2017.2754383

Ren, Y., Sun, R., Chen, S. H. Y., Du, C., Han, S.-T., and Zhou, Y. (2021). Exploring phase-change memory: From material systems to device physics. *Phys. Status Solidi* 15:2000394. doi: 10.1002/pssr.202000394

Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science* 275, 1593–1599. doi: 10.1126/science.275.5306.1593

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 484–489. doi: 10.1038/nature16961

Suri, M., Bichler, O., Querlioz, D., Cueto, O., Perniola, L., Sousa, V., et al. (2011). "Phase change memory as synapse for ultra-dense neuromorphic systems: application to complex visual pattern extraction," in *2011 International Electron Devices Meeting* (Washington, DC), 4.4.1–4.4.4. doi: 10.1109/IEDM.2011.6131488

Suri, M., Bichler, O., Querlioz, D., Traoré, B., Cueto, O., Perniola, L., et al. (2012). Physical aspects of low power synapses based on phase change memory devices. *J. Appl. Phys.* 112:054904. doi: 10.1063/1.4749411

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Mach. Learn.* 3, 9–44. doi: 10.1007/BF00115009

Tuma, T., Pantazi, A., Gallo, M. L., Sebastian, A., and Eleftheriou, E. (2016). Stochastic phase-change neurons. *Nat. Nanotechnol.* 11, 693–699. doi: 10.1038/nnano.2016.70

Turrigiano, G. G. (1999). Homeostatic plasticity in neuronal networks: the more things change, the more they stay the same. *Trends Neurosci.* 5, 221–227. doi: 10.1016/S0166-2236(98)01341-1

Turrigiano, G. G. (2008). The self-tuning neuron: synaptic scaling of excitatory synapses. *Cell* 135, 422–435. doi: 10.1016/j.cell.2008.10.008

Watkins, C. J. C. H., and Dayan, P. (1992). Q-learning. *Mach. Learn.* 8, 279–292. doi: 10.1023/A:1022676722315

Wong, H. S. P., Raoux, S., Kim, S., Liang, J., Reifenberg, J. P., Rajendran, B., et al. (2010). Phase change memory. *Proc. IEEE* 98, 2201–2227. doi: 10.1109/JPROC.2010.2070050

Wright, D., Hosseini, P., and Vazquez Diosdado, J. (2013). Beyond von-neumann computing with nanoscale phase-change memory devices. *Adv. Funct. Mater.* 23, 2248–2254. doi: 10.1002/adfm.201202383

Wunderlich, T., Kungl, A., Müller, E., Schemmel, J., and Petrovici, M. (2019). Brain-inspired hardware for artificial intelligence: accelerated learning in a physical-model spiking neural network. *arXiv[Preprint]. arXiv:1909.11145*. doi: 10.1007/978-3-030-30487-4_10

Xu, M., Mai, X., Lin, J., Zhang, W., Li, Y., He, Y., et al. (2020). Recent advances on neuromorphic devices based on chalcogenide phase-change materials. *Adv. Funct. Mater.* 30:2003419. doi: 10.1002/adfm.202003419

Yan, Y., Kappel, D., Neumaerker, F., Partzsch, J., Vogginger, B., Höppner, S., et al. (2019). "Efficient reward-based structural plasticity on a spinnaker 2 prototype," in *IEEE Transactions on Biomedical Circuits and Systems*. doi: 10.1109/TBCAS.2019.2906401

Zenke, F., Poole, B., and Ganguli, S. (2017). "Continual learning through synaptic intelligence," in *ICML'17: Proceedings of the 34th International Conference on Machine Learning, Vol. 70* (Sidney, NSW), 3987–3995.

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.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 23 September 2021
doi: 10.3389/fnins.2021.730566

Memristive Hodgkin-Huxley Spiking Neuron Model for Reproducing Neuron Behaviors

Xiaoyan Fang[1], *Shukai Duan*[2,3,4,5] *and Lidan Wang*[1,3,4,5]*

[1] School of Electronic and Information Engineering, Southwest University, Chongqing, China, [2] College of Artificial Intelligence, Southwest University, Chongqing, China, [3] Brain-Inspired Computing and Intelligent Control of Chongqing Key Lab, Chongqing, China, [4] National and Local Joint Engineering Laboratory of Intelligent Transmission and Control Technology, Chongqing, China, [5] Chongqing Brain Science Collaborative Innovation Center, Chongqing, China

The Hodgkin-Huxley (HH) spiking neuron model reproduces the dynamic characteristics of the neuron by mimicking the action potential, ionic channels, and spiking behaviors. The memristor is a nonlinear device with variable resistance. In this paper, the memristor is introduced to the HH spiking model, and the memristive Hodgkin-Huxley spiking neuron model (MHH) is presented. We experimentally compare the HH spiking model and the MHH spiking model by applying different stimuli. First, the individual current pulse is injected into the HH and MHH spiking models. The comparison between action potentials, current densities, and conductances is carried out. Second, the reverse single pulse stimulus and a series of pulse stimuli are applied to the two models. The effects of current density and action time on the production of the action potential are analyzed. Finally, the sinusoidal current stimulus acts on the two models. The various spiking behaviors are realized by adjusting the frequency of the sinusoidal stimulus. We experimentally demonstrate that the MHH spiking model generates more action potential than the HH spiking model and takes a short time to change the memductance. The reverse stimulus cannot activate the action potential in both models. The MHH spiking model performs smoother waveforms and a faster speed to return to the resting potential. The larger the external stimulus, the faster action potential generated, and the more noticeable change in conductances. Meanwhile, the MHH spiking model shows the various spiking patterns of neurons.

Keywords: HH, MHH, memristor, neuron, spiking

OPEN ACCESS

Edited by:
Huanglong Li,
Tsinghua University, China

Reviewed by:
Jie-Ning Wu,
Fudan University, China
Mauro Forti,
University of Siena, Italy

***Correspondence:**
Lidan Wang
ldwang@swu.edu.cn

Specialty section:
This article was submitted to Neuromorphic Engineering, a section of the journal Frontiers in Neuroscience

Received: *25 June 2021*
Accepted: *16 August 2021*
Published: *23 September 2021*

Citation:
Fang X, Duan S and Wang L (2021) Memristive Hodgkin-Huxley Spiking Neuron Model for Reproducing Neuron Behaviors. Front. Neurosci. 15:730566. doi: 10.3389/fnins.2021.730566

1. INTRODUCTION

Neurons with highly nonlinear characteristics act as the basic functional unit of receiving and propagating signals. The whole procedure of processing signals in the nerve system needs the cooperation of neurons. Some theoretical knowledge and research methods are beneficial to unveil the mechanism of information propagation in neurons. Italian scientist Camillo Golgi worked on the nervous system structure and earned the Nobel Prize for physiology and medicine in 1906 (Dröscher, 1998). In 1998, Ramon y Cajal pointed out that the neurons without directly connecting each other in the nerve system (Raviola and Mazzarello, 2011). To replicate the functions and mechanisms of neurons, we urgently need to construct the biophysical model. A variety of neuron models are emerging, and the Hodgkin-Huxley (HH) spiking neuron model is the original

(Hodgkin and Huxley, 1989). Stochastic Hodgkin-Huxley Neuron Systems with the NEF is helpful to study neuron sensitivity (Chen and Li, 2010). The Hodgkin-Huxley Model with automatic parameter estimation is applied to the neuromimetic chips (Buhry et al., 2011). The space-clamped Hodgkin-Huxley model effectively inhibits the production of spikes under the injection of the noisy synaptic input (Tuckwell and Ditlevsen, 2016). The Langevin is combined with the Hodgkin-Huxley system performs accurate interspike interval (ISI) and realizes the accuracy minimal loss (Pu and Thomas, 2020). The Berger-Levy theory is introduced to the Hodgkin-Huxley model, demonstrate that the information communication between neurons is related to the presynaptic firing rate and the synchronization (Ghavami et al., 2018).

The memristor with the non-volatility and variable resistance characteristics is regarded as the fourth passive circuit element. Therefore, it becomes a hot topic in neural computing (Le et al., 2015), learning and memorizing (Sayyaparaju et al., 2018), micro-circuitry design (Berdan et al., 2014), biological synapse (Mandal and Saha, 2016), and neuron modeling (Maheshwar et al., 2014), and so on. The synaptic plasticity of biological neuronal systems can be realized by memristors and memristive crossbar in 3-D architecture to mimic the human brain (Truong et al., 2016). The memristor with hysteresis and memory characteristics is the most promising candidate for establishing the brain-like neuromorphic system (Mokhtar et al., 2017). The key features of biological neurons and synapses can be mimicked by memristors (Berdan et al., 2016; Mandal and Saha, 2016). The ion motion in neurons is represented by the electrical conductance change of a memristor (Xia and Yang, 2019). A memristor is used as a two-terminal resistor with memory (Chua, 1971; Strukov et al., 2008) performs well in storing information according to the physical laws (Yang et al., 2013). The memristor entirely avoids the data transformation bottleneck between the memory and computation (Li and Wang, 2019). The memristor crossbar array can be used to integrate the co-processor chip, which will realize machine learning algorithms and neuromorphic computing (James, 2019).

This work elaborates on the construction of the memristive Hodgkin-Huxley spiking neuron model. The mathematical expressions and the circuit of the HH spiking model are presented and analyzed in sections 2, 3. Section 4 describes the MHH spiking model and discusses the memristors used to mimic the ion channels. The comparison between two models under the different stimuli is conducted in section 5. Section 6 is the conclusion of the paper.

2. THE HODGKIN-HUXLEY (HH) SPIKING NEURON MODEL

The neuron cell membrane is a voltage-gated ion channel, which has high selectivity for the permeability of external and internal ions in body fluid. Only one type of ion can pass through specific channels. There involves four ionic components, sodium, potassium, calcium, and chloride. The transmembrane current depends on the rapid inward current caused by sodium and the slow outward current caused by potassium (Häusser, 2000). The ion concentration difference inside and outside of the cell is the primary driving force of neural activities. When the sodium channels are opened, the high concentration sodium flows from extracellular to intracellular, the depolarization is produced, the action potential is generated. And then, the sodium channels are closed, and the potassium channels are opened, the potassium permeates from intracellular to extracellular, the repolarization is performed. Finally, the membrane potential undergoes a hyperpolarization phase, the membrane potential shifts back to the resting potential. The above process is the generation mechanism of the action potential in a neuron.

The inside of the axon membrane is full of ionic fluids (cytoplasm), the outside of the axon membrane is filled with body fluids. The fluids (conductor) of intracellular and extracellular are separated by the axon membrane (insulator). When an insulator separates two conductors, the capacitor emerges to model the charge storage capacity. The part of the axon membrane without ion channels is equivalent to a capacitor (C_m). The axon membrane of the neuron consists of the lipid bilayer, the membrane protein, and ion channels (the upper image in **Figure 1**). The sodium ion channel is represented by a nonlinear conductance (g_{Na}), the potassium ion channel is denoted by a nonlinear conductance (g_K), and other ion channels are described as a linear conductance (g_L) (Beck et al., 2020). When the neuron is in the resting state, a potential difference is caused by the ionic concentration between the intracellular and extracellular fluids. The potential difference is called the equilibrium potential of each ion (E_{Na}, E_K, and E_L), which is equivalent to a driving power supply (the lower image in **Figure 1**).

When the neuron is in the resting state, there is a resting potential. Here, we choose $v_{rest} = -65$ mV as the resting potential in experiments (Hodgkin and Huxley, 1952). The V_m denotes the membrane potential, E_{Na} (50 mV), E_K (−70 mV), and E_L (−50 mV) represent the Nernst equilibrium potentials. When the potassium current passes through the potassium channel, the potassium current is proportional to the difference between the membrane potential and E_K (Hodgkin and Huxley, 1989; Börgers, 2017):

$$I_K = g_K(V_m - E_K) \tag{1}$$

Here, g_K is the potassium conductance, (V_m-E_K) is the potassium driving force. The sodium current and the leaky current are described as:

$$I_{Na} = g_{Na}(V_m - E_N a) \tag{2}$$

$$I_L = g_L(V_m - E_L) \tag{3}$$

The ion channels are sensitive to membrane potential, which control the open and close states of channels.

In the Hodgkin-Huxley spiking model, the conductance value of each ion channel is decided by the gate-controlled variables m, n, h, and $0 \leq m \leq 1$, $0 \leq n \leq 1$, $0 \leq h \leq 1$. The potassium channel depends on four active gate variables (n). The sodium channel is controlled by three active gate variables (m) and

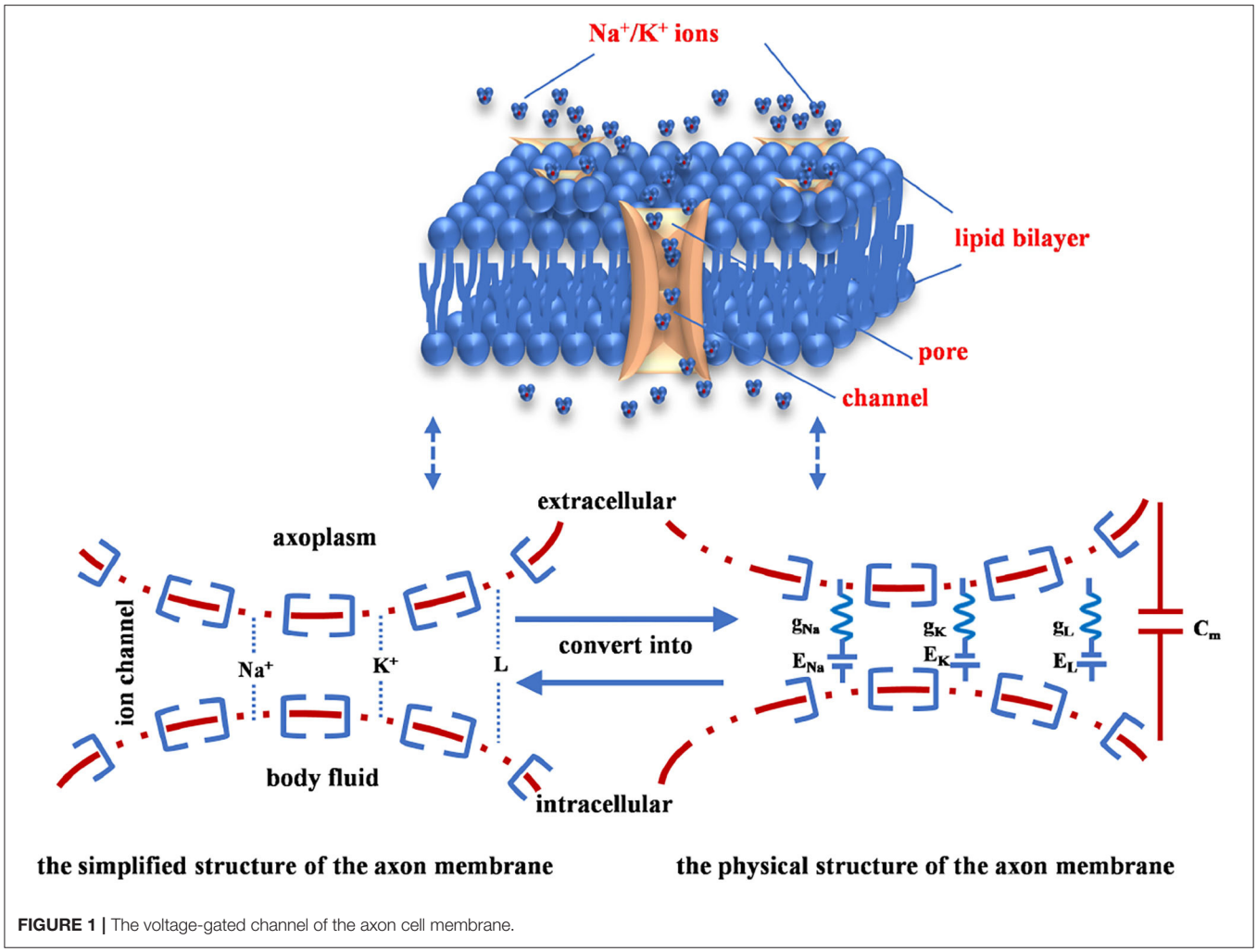


FIGURE 1 | The voltage-gated channel of the axon cell membrane.

one inactive gate variable (h). The potassium conductance, the sodium conductance, and the leaky conductance are described as:

$$g_K = g_{Kmax} n^4 \tag{4}$$

$$g_{Na} = g_{Namax} m^3 h \tag{5}$$

$$g_L = g_{Lmax} \tag{6}$$

Here, g_{Kmax}, g_{Namax}, and g_{Lmax} denote the maximum values of potassium, sodium, and leaky conductances, accordingly. Their values are 36, 120, 0.3 $Ohm^{-1}cm^{-2}$ (Hodgkin and Huxley, 1952, 1989). The expressions of gate-controlled variables of ion channels are written as follows:

$$dm/dt = 1/\tau_m(m_\infty - m) \tag{7}$$

$$dn/dt = 1/\tau_n(n_\infty - n) \tag{8}$$

$$dh/dt = 1/\tau_h(h_\infty - h) \tag{9}$$

The time constants τ_m, τ_n, and τ_h change with m, n, and h, accordingly. The transition rate α characterizes the ion channels change from the close state to the open state. The transition rate β indicates the ion channels vary from the open state to the close state. m_∞, n_∞, and h_∞ are the steady-state values of the gate variables m, n, and h, accordingly (Saïgai et al., 2011). They are all the functions of the membrane potential. Their expressions are:

$$m_\infty = \alpha_m/(\alpha_m + \beta_m) \tag{10}$$

$$n_\infty = \alpha_n/(\alpha_n + \beta_n) \tag{11}$$

$$h_\infty = \alpha_h/(\alpha_h + \beta_h) \tag{12}$$

$$\tau_m = 1/(\alpha_m + \beta_m) \tag{13}$$

$$\tau_n = 1/(\alpha_n + \beta_n) \tag{14}$$

$$\tau_h = 1/(\alpha_h + \beta_h) \tag{15}$$

$$\alpha_m = \varphi(2.5 - 0.1(V_m - V_{rest}))/(e^{(2.5-0.1(V_m-V_{rest}))} - 1) \quad (16)$$

$$\alpha_n = \varphi(0.1 - 0.01(V_m - V_{rest}))/(e^{(1-0.1(V_m-V_{rest}))} - 1) \quad (17)$$

$$\alpha_h = 0.07\varphi e^{(-(V_m-V_{rest}))/20} \quad (18)$$

$$\beta_m = 4\varphi e^{(-(V_m-V_{rest}))/20} \quad (19)$$

$$\beta_n = 0.125\varphi e^{(-(V_m-V_{rest}))/80} \quad (20)$$

$$\beta_h = \varphi/(e^{(3.0-0.1(V_m-V_{rest}))} + 1) \quad (21)$$

Here, φ=$3^{(T-6.3)/10}$. The relationship between the transition state and the membrane potential is shown in **Figure 2** (Hodgkin and Huxley, 1952, 1989; Börgers, 2017).

The HH spiking neuron model is strongly dependent on the temperature, and the early experiments were carried out under the temperatures $T = 6.3°C$ and $T = 18.5°C$. When the temperature is $6.3°C$, the transition rates of the active gates α_n and α_m (**Figure 2A**), the inactive rate β_h (**Figure 2B**) increase with the rise of the membrane potential. The inactive transition rate α_h (**Figure 2A**), the active transition rates β_n and β_m (**Figure 2B**) decrease with the increase of the membrane potential. When the temperature is increased to $18.5°C$, the transition rates α and β show the same experimental phenomena (**Figures 2C,D**) as above. We compare the transition rates at different temperatures, and the difference is performed in the light blue ellipse. When the temperature is $6.3°C$, α_n varies from 0 to 10, α_m alters from 0 to 1, α_h changes from 0.5 to 0 (the enlarged plot in **Figure 2A**). When the temperature is $18.5°C$, α_n varies from 0 to 36, α_m adjusts from 0 to 3.5, α_h changes from 2 to 0 (the enlarged plot in **Figure 2C**). When the temperature is set to $6.3°C$, β_n varies from 37 to 0, β_m adjusts from 0.2 to 0, β_h changes from 0 to 1 (the enlarged plot in **Figure 2B**). When the temperature is increased to $18.5°C$, β_n varies from 140 to 0, β_m adjusts from 0.8 to 0, β_h changes from 0 to 4 (the enlarged plot in **Figure 2D**). The higher the temperature, the greater the range of conversion rates, the longer time needed to return to the critical value of the transition rate.

When the temperatures are $T = 6.3°C$ and $T = 18.5°C$, the simulation plots between the steady values of gate variables (m_∞, n_∞, and h_∞) and the membrane potential, the relationship between the time constant (τ_m, τ_n, and τ_h) and the membrane potential, as shown in **Figure 3**.

The steady-state values (m_∞ and n_∞) of activation gate variables (m and n) change from 0 to 1 with the increase of the membrane potential. The steady-state value (h_∞) of the inactivation gate variable (h) decreases with the increase of the membrane potential (**Figures 3A,C**). The steady-state values are not affected by the change of temperature. When the temperature is 6.3°C, τ_n varies from 5.8 to 1, τ_m adjusts from 0.8 to 0, τ_h changes from 9 to 1. When the temperature is increased to 18.5°C, τ_n varies from 1.5 to 0.25, τ_m adjusts from 0.2 to 0, τ_h changes from 2.25 to 0.25 (**Figures 3B,D**). The higher temperature, the smaller the range of τ.

3. THE ELECTRICAL CIRCUIT OF THE HODGKIN-HUXLEY SPIKING NEURON

The significant electrical properties of a neuron can be precisely replicated by the HH circuit model, as shown in **Figure 4A** (Hodgkin and Huxley, 1989).

Here, C is the membrane capacitor. g_{Na} is the sodium conductance, g_K is the potassium conductance, and g_L is the leaky conductance. V_m is the membrane potential. I_C is the capacitor current, I_{Na} is the sodium current, I_K is the potassium current, and I_L is the leaky current. I_{ext} is the external stimulus. E_{Na}, E_K, and E_L are ion concentration differences of sodium, potassium, and leakage [namely, the equilibrium potentials (Emili et al., 2003) are calculated by the Nernst equation (Hill, 1992)]. The arrow directions of currents are pointing from inside to outside of the membrane. The value of the extracellular potential is set to zero (V_{out}= 0, namely, the extracellular is grounded) (Hodgkin and Huxley, 1989).

According to Kirchhoff's voltage-current law, the circuit equations are described as:

$$V_m = V_{in} - V_{out} \quad (22)$$

$$I_C = dQ/dt \quad (23)$$

$$Q = CV_m \quad (24)$$

$$I_m = I_{Na} + I_K + I_L \quad (25)$$

$$I_{ext} = I_C + I_{Na} + I_K + I_L = I_C + I_m \quad (26)$$

In the giant squid axon experiment, the current through the axon membrane is expressed as the current density J(t, x). It represents the amount of the electric current per square centimeter, and its unit is $mAcm^{-2}$. Based on the mathematical analysis of the RC equivalent circuit (**Figure 4A**), the following voltage-current equations are obtained.

$$C\partial V_m(t,x)/\partial t = -J_m(t,x) + J_{ext}(t,x) + 1/(2r_{in})\partial^2 V_m(t)/\partial x^2 \quad (27)$$

$$J_m = J_{Na} + J_K + J_L \quad (28)$$

$$J_{Na} = g_{Na}(V_m - E_{Na}) \quad (29)$$

$$J_K = g_K(V_m - E_K) \quad (30)$$

$$J_L = g_L(V_m - E_L) \quad (31)$$

The left side of (27) is the charging or discharging rate per unit area for the capacitor. $J_m(t,x)$ is the total current density that flows through the membrane. J_{Na} is the current density passing through sodium conductance. J_K is the current density of potassium. V_m is the membrane potential. J_{ext}(t, x) is the external stimulus. The last term is the charge rate of longitudinal current

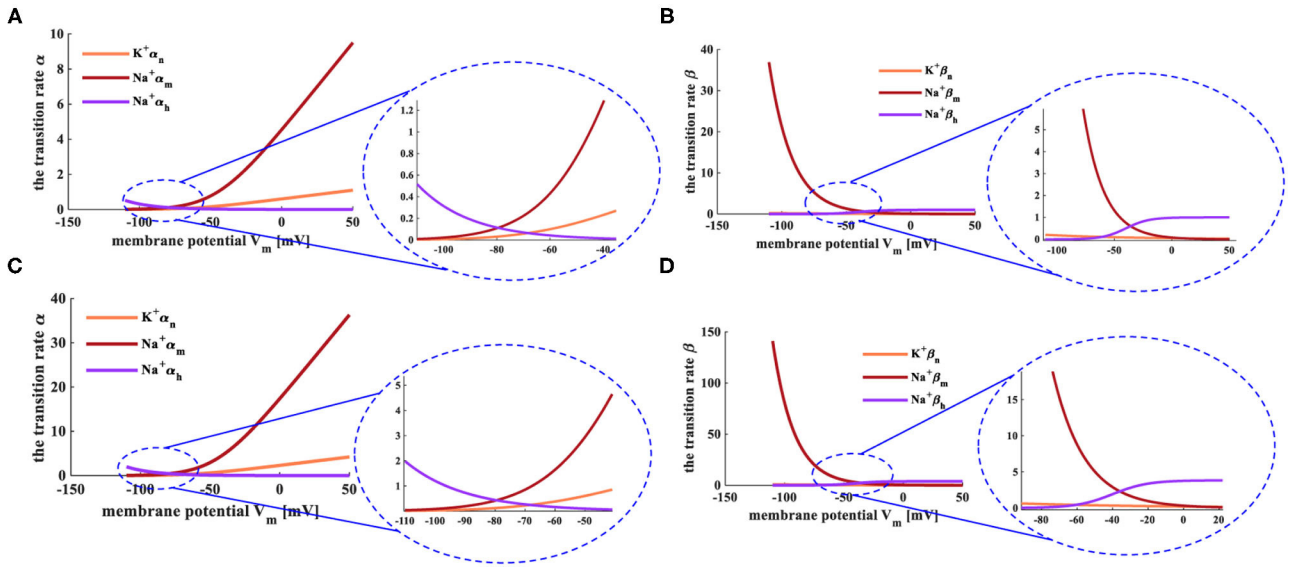


FIGURE 2 | The relationship between transition state and membrane potential. **(A)** The evolution of the transition rate α at the temperature of 6.3°C. **(B)** The variation of the transition rate β at the temperature of 6.3°C. **(C)** The evolution of the transition rate α at the temperature of 18.5°C. **(D)** The change of the transition rate β at the temperature of 18.5°C.

along the inside membrane surface. It depends only on the time t rather than the location x, so the quadratic partial differential term equals zero, (27) can be rewritten as:

$$C\partial V_m(t,x)/\partial t = -J_m(t,x) + J_{ext}(t,x) \tag{32}$$

The propagated action potential is performed by (32). The action potential is sensitive to the temperature. The action potential of the cell membrane shows distinct firing behaviors under various temperatures.

When the temperature is $6.3°C$, the HH spiking model generates three action potentials in 20 ms, the duration of a spike is 7.65 ms (**Figure 5A**). When the temperature becomes $15°C$, the HH spiking model generates six action potentials in 20 ms, the duration of a spike decreases to 3.35 ms (**Figure 5B**). When the temperature is increased to $20°C$, the HH spiking model generates nine action potentials in 20 ms, the duration of a spike reduces to 1.95 ms (**Figure 5C**). We increase the temperature to $35°C$, and there is no action potential produced after one action potential is generated (**Figure 5D**). We decrease the temperature to $-20°C$, and the action potential cannot be obtained (**Figure 5E**). The temperature affects the time duration of the spike, the generation of action potentials, and the firing frequency of a neuron. It is hard to achieve the action potential when the temperature is too high or low. The increase of temperature has significantly decreased the time duration of the spike and remarkably produced a higher firing frequency.

The external stimuli with various intensities act on the HH spiking model, which performs different action potentials. When the current density is $0.001 mA cm^{-2}$, the HH spiking model cannot produce the action potential (**Figure 6A**). When the current densities are increased to 0.01 and $0.09 mA cm^{-2}$, the action potentials are obtained (**Figures 6B,C**). However, when the current density becomes $0.2 mA cm^{-2}$, the HH spiking model generates one action potential. After that, it cannot produce the action potentials (**Figure 6D**). The external stimulus is related to the generation of the action potential. The larger the external stimulus, the higher the firing frequency. If the external stimulus is too larger or small, the HH spiking model cannot reproduce the action potential.

When the action time of the external stimulus is 1 ms, there is not enough time to show the complete firing process (**Figure 7A**). Therefore, the action time is increased to 10 ms, and the action potential is generated (**Figure 7B**). When the action time becomes 20 or 50 ms, the HH spiking model produces more action potentials (**Figures 7C,D**). Thus, the action time of the external stimulus has a strong influence on the generation of the action potential. The longer the action time, the more action potentials generated. But when the action time is too long or short, the HH spiking model cannot perform the firing process.

4. THE MEMRISTIVE HODGKIN-HUXLEY (MHH) SPIKING NEURON MODEL

In the HH circuit model, the potassium conductance and the sodium conductance are voltage-gated channels, which can be described by time and membrane potential. The flux-controlled memristor with the nonvolatile property is the function of time and voltage, which can be used in a nonlinear circuit system (Petras, 2010; Corinto and Forti, 2017; Corinto et al., 2018). Based on the HH spiking model, we replace the sodium and potassium conductances with the flux-controlled memristors

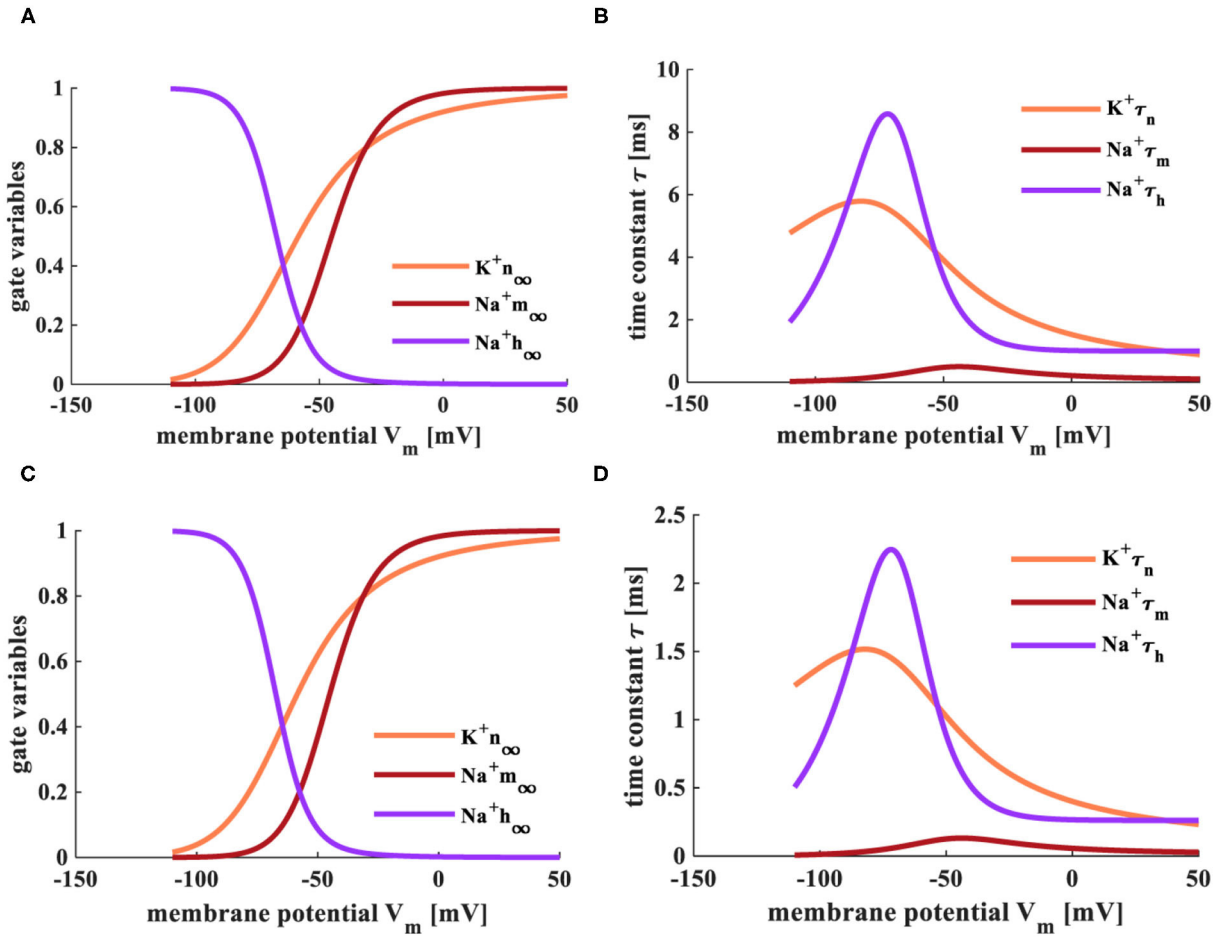


FIGURE 3 | The relationship between gate-controlled variables, the time constant, and the membrane potential. **(A)** The evolution of the gate-controlled variables (m, n, and h) at the temperature of 6.3°*C*. **(B)** The change of the time constant (τ) at the temperature of 6.3°*C*. **(C)** The change of the gate-controlled variables (m, n, and h) at the temperature of 18.5°*C*. **(D)** The evolution of the time constant (τ) at the temperature of 18.5°*C*.

(Wang et al., 2012), and the memristive Hodgkin-Huxley spiking neuron model is constructed (**Figure 4B**).

Some of the mathematical expressions in the HH spiking model need to be modified. g_{Na} and g_K in (4) and (5) are replaced by the memristance and rewritten as:

$$g_{MK} = 1/M_K n^4 \tag{33}$$

$$g_{MNa} = 1/M_{Na} m^3 h \tag{34}$$

The conductance values of the sodium and potassium ion channels become the function of time, and the membrane potential will change with the evolution of the memristance.

The flux-controlled memristor is described as (Wang et al., 2012):

$$M(\phi(t)) = \begin{cases} 20000 & \phi(t) < -0.75 \\ \sqrt{-3.98 \times 10^8 \phi(t) + 10^8} & \phi(t) \geq -0.75 \text{ and } \phi(t) < 0.25 \\ 100 & \phi(t)) \geq 0.25 \end{cases} \tag{35}$$

Where M_K=M_{Na}=M is the function of time. The potassium memristance (g_{MK}) and the sodium memristance (g_{MNa}) are functions involved with time and membrane potential. When the various external stimuli act on the MHH spiking neuron model, changes in g_{MK} and g_{MNa} are performed in **Figure 8**.

The initial values of memductances and reductance g_{MK}= 0.5×10^{-4} $Ohm^{-1}cm^{-2}$, g_{MNa}= 0.5×10^{-4} $Ohm^{-1}cm^{-2}$, and g_L= 0.3×10^{-3} $Ohm^{-1}cm^{-2}$ [0.5×10^{-4} is the reciprocal of the maximum value (20,000 $Ohmcm^{-2}$) of a memristor]. The temperature is 6.3°C, C is 1 μF m^{-2}. E_{Na} is 50 mV, E_K is −70 mV, and E_L is −50 mV.

When the external stimulus [0.008 mA cm^{-2} (g_{MNa})] is applied to the MHH spiking model, the sodium memductance (the coral color curve) does not change in the time range from 0 to 1.025 ms (the enlarged plot in **Figure 8A**). Then, the sodium memductance increases to 0.029 $Ohm^{-1}cm^{-2}$ and then decreases to zero. When the MHH spiking model receives the external stimulus [0.08 $mAcm^{-2}$ (g_{MNa1})], the sodium memductance (the dark red curve) remains the same in the time range from 0ms to

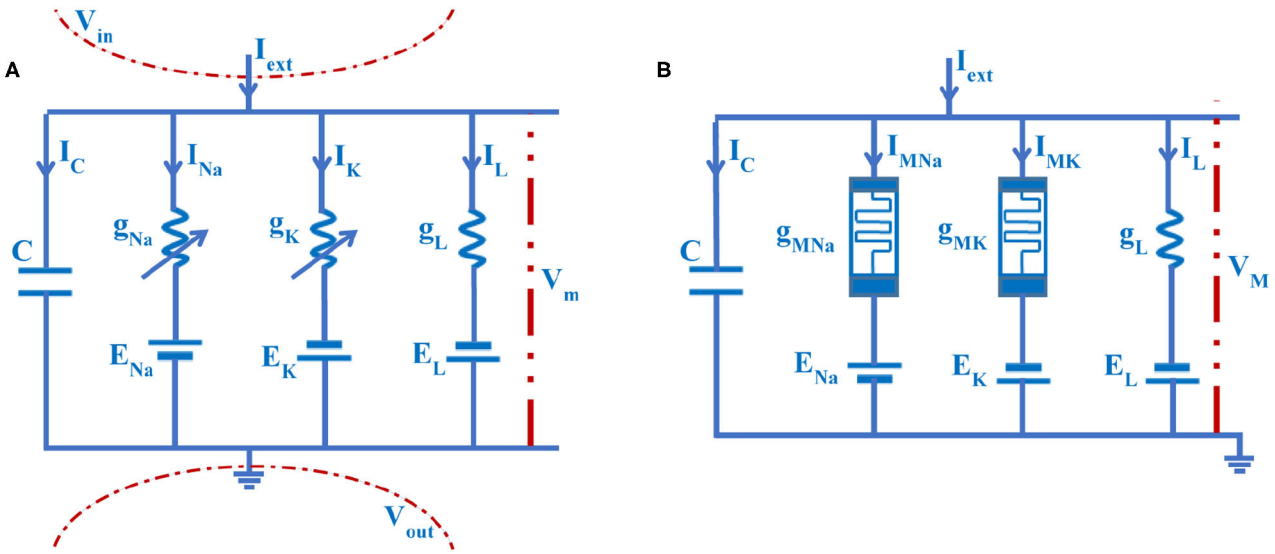


FIGURE 4 | The electrical circuit of the axon cell membrane. **(A)** The HH circuit model. **(B)** The MHH circuit model.

1.38ms (the enlarged plot in **Figure 8A**). And the maximum value of the sodium memductance is 0.031 $Ohm^{-1}cm^{-2}$. Likewise, when the external stimulus [0.8 $mAcm^{-2}$ (g_{MNa2})] acts on the MHH spiking model, the sodium memductance (the purple curve) does not change in the time range from 0 to 0.97 ms (the enlarged plot in **Figure 8A**). And the maximum value of the sodium memductance is 0.038 $Ohm^{-1}cm^{-2}$.

When the external stimulus [0.04 $mAcm^{-2}$ (g_{MK})] is injected into the MHH spiking model, the potassium memductance (the coral color curve) does not change from 0 to 1.5 ms (the enlarged plot in **Figure 8B**). Then, the potassium memductance increases and attains 0.0324 $Ohm^{-1}cm^{-2}$. Likewise, the MHH spiking model receives the external stimuli [(0.08 $mAcm^{-2}$ (g_{MK1}) and 0.16 $mAcm^{-2}$ (g_{MK2})], the potassium memductance (the dark red curve reaches 0.0348 $Ohm^{-1}cm^{-2}$ and the purple curve attains 0.0359 $Ohm^{-1}cm^{-2}$ (the enlarged plot in **Figure 8B**) are stable at constant values (**Figure 8B**).

The sodium memductance and the potassium memductance are associated with the external stimulus. The stronger the external input, the faster the memductance changes, the larger the memductance value. The change curves of sodium and potassium memductance are similar to the theoretical curves (refer to Hodgkin and Huxley, 1989). Therefore, the memristors can mimic the sodium ion channel and the potassium ion channel.

The temperature is selected as 6.3°C, and the external current is 0.08 $mAcm^{-2}$. The transition rate parameters (α and β), gate variables (m_∞, n_∞, and h_∞), and the time constant (τ) in the MHH spiking model are shown in **Figure 9**.

The transition rates of the active gates (α_n and α_m, **Figure 9A**), the inactive transiton rate (β_h, **Figure 9B**) enhance with the increase of the membrane potential. The inactive transition rate (α_h, **Figure 9A**), the active transition rates (β_n and β_m, **Figure 9B**) decrease with the rise in the membrane potential. The steady-state values (m_∞ and n_∞) of activation gate variables (m and n) change from 0 to 1 with the increase of the membrane potential. The steady-state value (h_∞) of the inactivation gate variable (h) decreases with the increase of the membrane potential (**Figure 9C**). The time constant τ_n changes from 4.52 to 0, τ_m adjusts from 0.5 to 0, and τ_h varies from 8.57 to 0 (**Figure 9D**). The changing processes of the transition rate, gate variables, and the time constant in the MHH spiking model have high similarities with those of the HH spiking model in **Figures 2**, **3**. Therefore, the memristors can be utilized as the sodium ion channel and the potassium ion channel.

When the current density J_m in (28) is replaced by J_M, conductances g_{Na} and g_K in (29) and (30) are replaced by g_{MNa} and g_{MK}, and the current equations are rewritten as:

$$J_M = J_{MNa} + J_{MK} + J_L \tag{36}$$

$$J_{MNa} = g_{MNa}(V - E_{Na}) \tag{37}$$

$$J_{MK} = g_{MK}(V - E_K) \tag{38}$$

The membrane potential V_m in (32) is replaced by V_M, and the membrane potential of the MHH spiking neuron model is described as:

$$C\partial V_M(t,x)/\partial t = -J_M(t,x) + J_{ext}(t,x) \tag{39}$$

The electrical equivalent circuit of the HH spiking model is based on the voltage-clamp experimental method. When the voltage-clamp values are distinct, the variables perform various variations in the HH and MHH spiking models. Here, the temperature T = 6.3°C. The clamp voltage is denoted by V_{clamp}, and its value is selected as +20 or +80 mV. The resting potential V_{rest} = −65 mV. The membrane potential V_m=V_{clamp} +V_{rest}.

When the clamp-voltage value is 20 mV, the membrane potential becomes −45 mV. Changes of Na^+ and K^+ gate

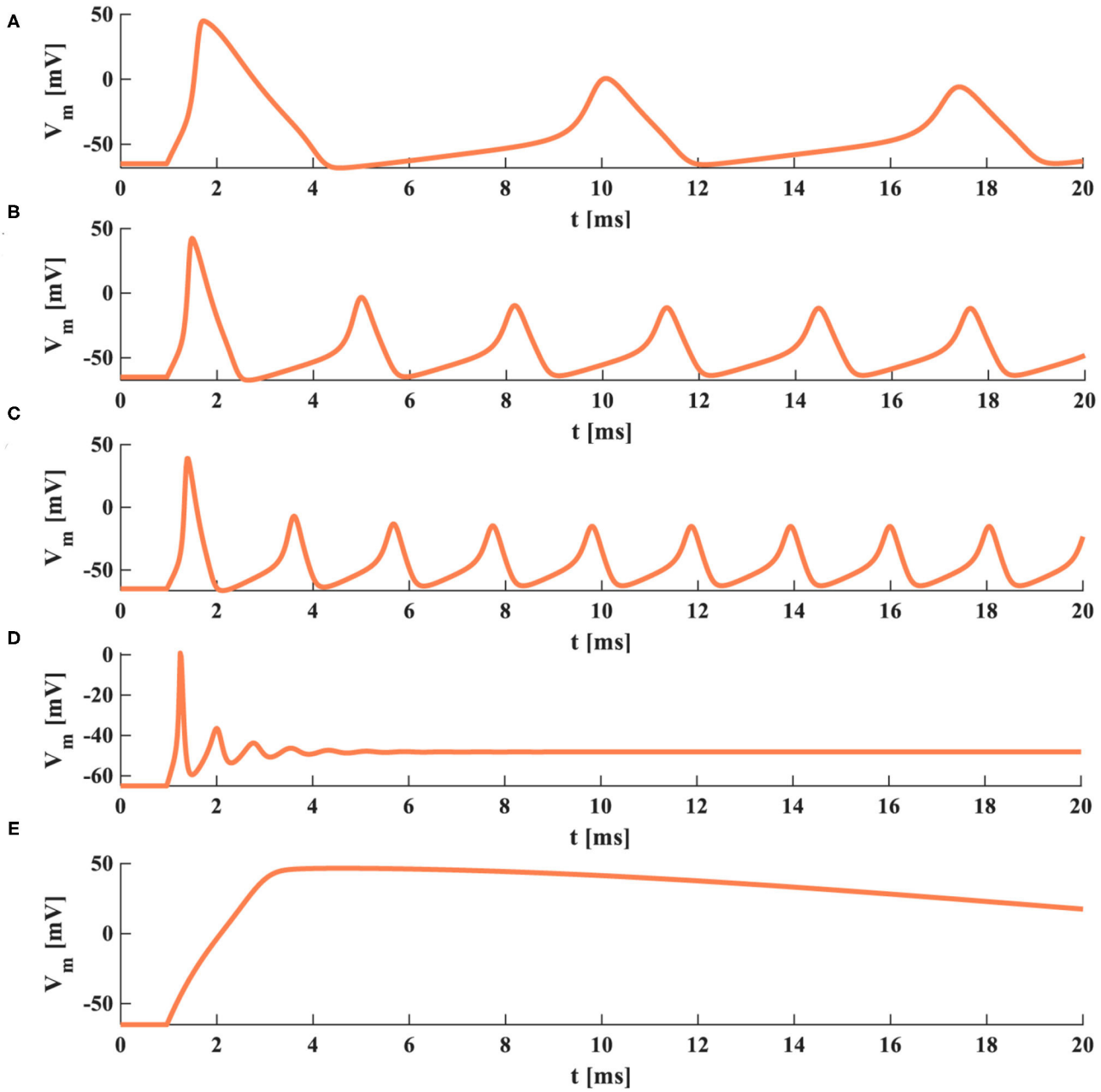


FIGURE 5 | The action potentials under the different temperatures (the external stimulus is 0.08 $mAcm^{-2}$). **(A)** The temperature is 6.3°*C*. **(B)** The temperature is 15°*C*. **(C)** The temperature is 20°*C*. **(D)** The temperature is 35°*C*. **(E)** The temperature is −20°*C*.

variables in the MHH spiking model (the plots on the left in **Figure 10B**) are the same as those in the HH spiking model (the plots on the left in **Figure 10A**). The HH spiking model generates the reverse curves of J_{Na} and J_m, and their maxima are −0.17 and −0.21 $mAcm^{-2}$. The maximum of the forward curve J_K is 0.14 $mAcm^{-2}$, and the forward curve J_L reaches 0.009 $mAcm^{-2}$. The peak values of g_K and g_{Na} are 4.54 and 2.25 $mOhm^{-1}cm^{-2}$ (the plots on the right in **Figure 10A**). The MHH spiking model produces the reverse curves of J_{MNa} and J_M, and their maxima are −0.18 and −0.16 $mAcm^{-2}$. The forward curves of J_{MK} and J_L attain their maxima 0.04 and 0.009 $mAcm^{-2}$. The maxima of g_{MK} and g_{MNa} are 1.26 and 1.88 $mOhm^{-1}cm^{-2}$ (the plots on the right in **Figure 10B**).

The variable values of the HH spiking model are more significant than those of the MHH spiking model (because the memristance is large, its initial value is 10,000 $Ohmcm^{-2}$). When the clamp-voltage value is 20 mV, both spiking models cannot generate the action potential.

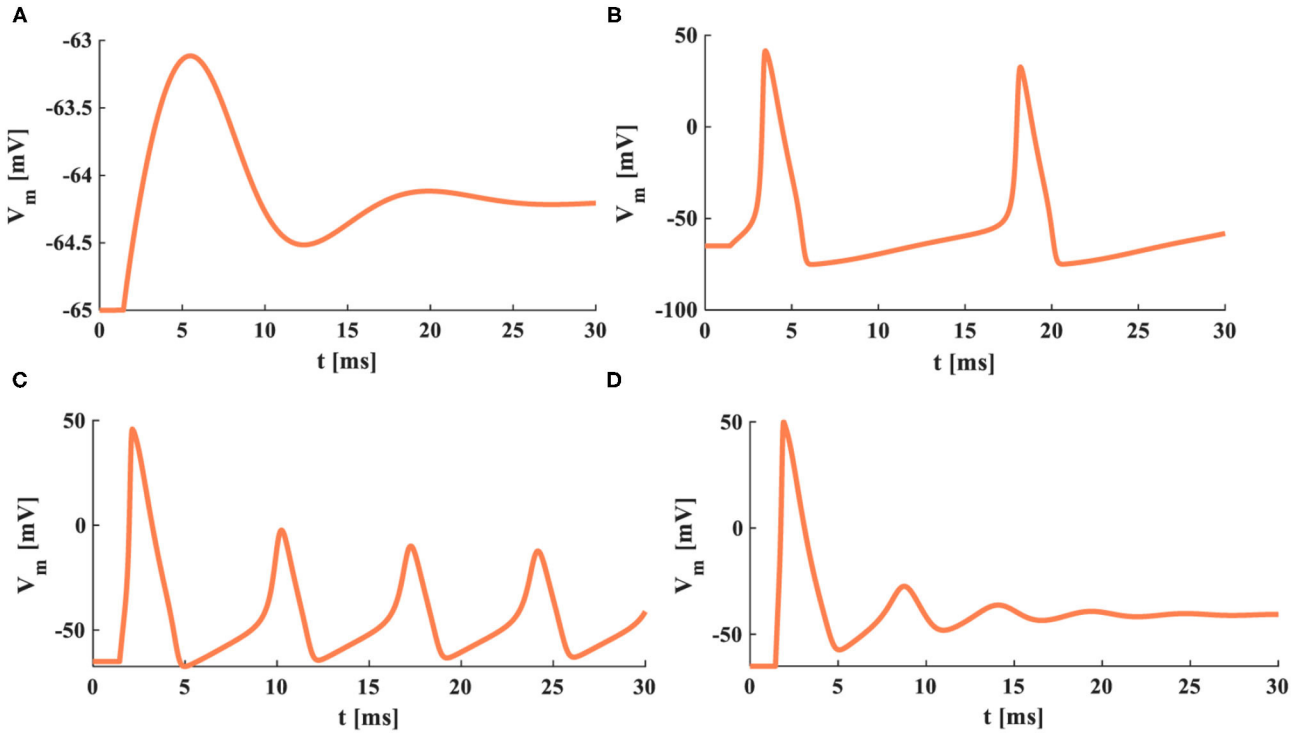


FIGURE 6 | The distinct stimuli are applied to the HH model. **(A)** The external stimulus is 0.001 $mAcm^{-2}$. **(B)** The external stimulus is 0.01 $mAcm^{-2}$. **(C)** The external stimulus is 0.09 $mAcm^{-2}$. **(D)** The external stimulus is 0.2 $mAcm^{-2}$.

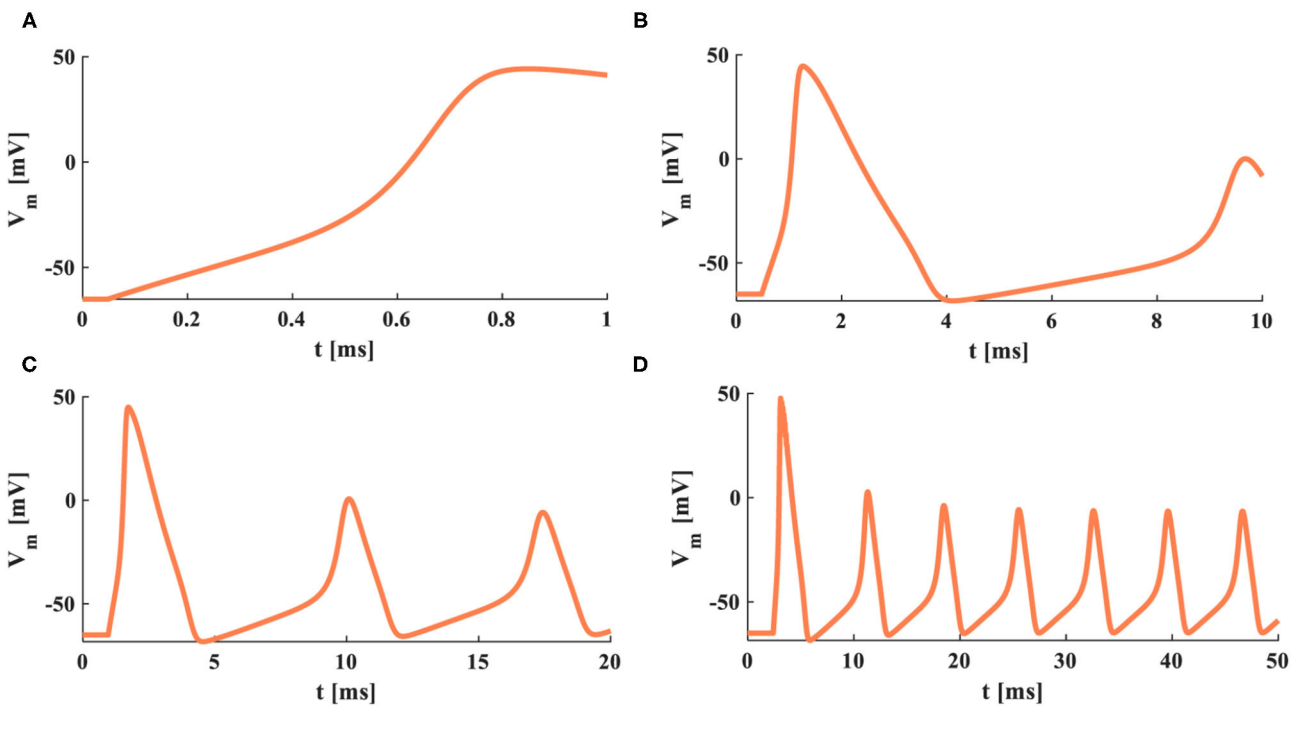


FIGURE 7 | The firing behaviors under the various action time of the external stimulus. **(A)** The action time is 1 ms. **(B)** The action time is 10 ms. **(C)** The action time is 20 ms. **(D)** The action time is 50 ms.

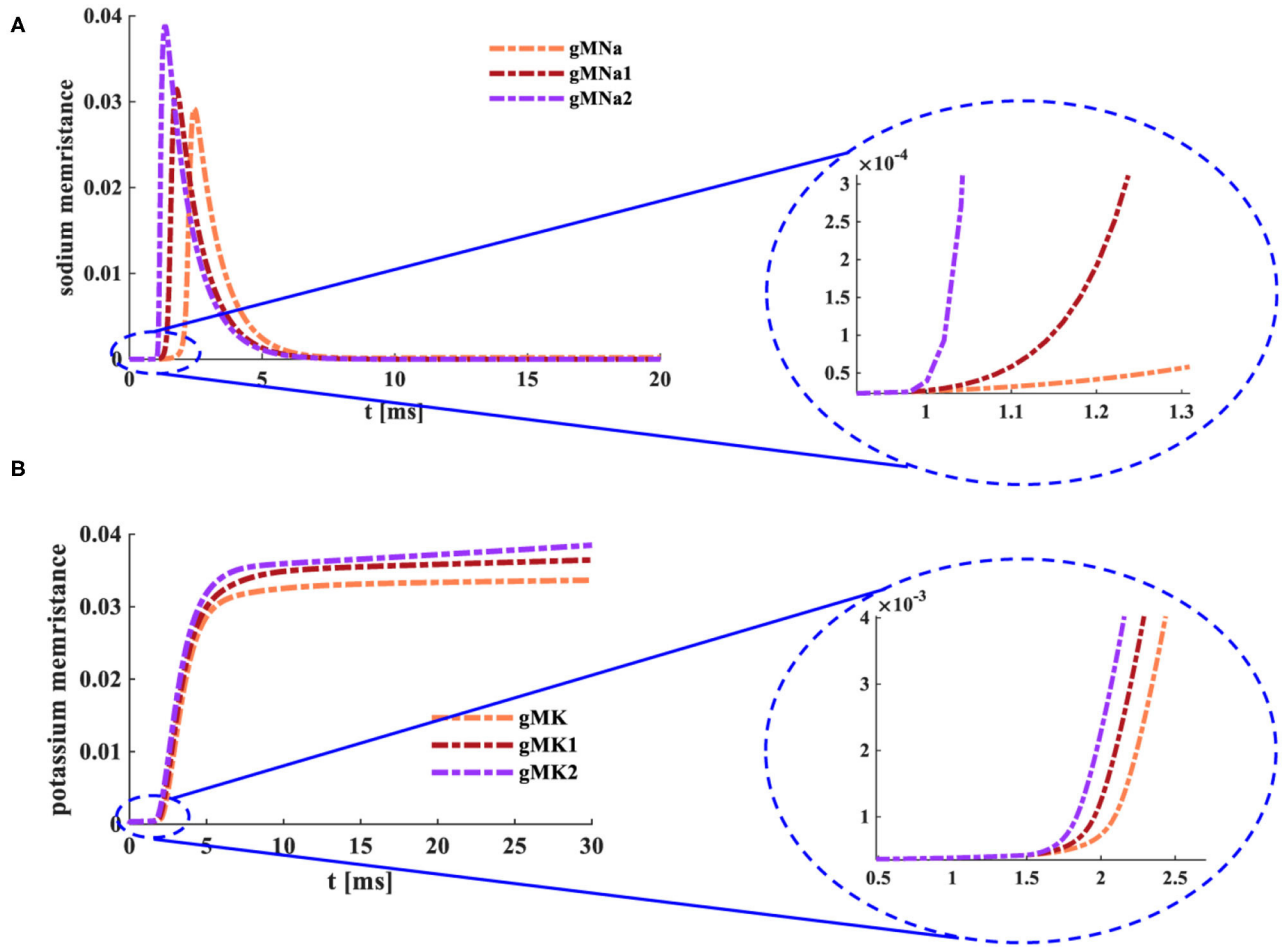


FIGURE 8 | Changes of the memristance under the distinct input currents. **(A)** The variations of the sodium memristance. **(B)** The variations of the potassium memristance.

A transient increase of sodium ions in the cell leads to the depolarization of the action potential. The waveforms of the two models change in the same way when the clamp voltage is 80 mV (the membrane potential is 15 mV). We take the MHH model as an example and make a vertical comparison (**Figures 10B,D**). With the increase of clamp voltage, the current densities of sodium and potassium increase significantly. The value of gate variable n changes from 0.5 to 1, and the value of gate variable m varies from 0.4 to 1. The potassium memductance changes from 1.26 to 8 $mOhm^{-1}cm^{-2}$, and the sodium memductance changes from 1.88 to 30 $mOhm^{-1}cm^{-2}$.

When the clamp-voltage value is 80 mV, the HH and MHH spiking models can produce the action potential. The gate variables n and m change with the identical waveforms. The current densities, the potassium conductance, and the sodium conductance are different. The maxima of J_{MNa}, J_{MK}, J_L, and J_M are -1.059, 0.74, 0.0297, and -0.97 $mAcm^{-2}$ (the right-upper plot in **Figure 10D**), which are larger than those of the HH spiking model (**Figure 10C**). The variation ranges of potassium conductance and sodium conductance for the MHH spiking model are [0 8], [0 30] less than those [0 29], [0 37] in the HH spiking model. The higher the voltage-clamp value, the larger the variable values, the smaller the conductance variation range.

5. THE COMPARISON BETWEEN TWO MODELS UNDER THE DIFFERENT STIMULI

5.1. The Individual Current Pulse Stimulus

The forward stimulus $J_{ext} = 0.1mAcm^{-2}$ (the pulse width is 0.1 ms) is applied to the HH spiking model and the MHH spiking model, the temperature is selected as 18.5°C, and the response time of the model is 5 ms. The initial value of the membrane potential is the resting potential, $V_{rest} = -65$ mV.

Here, J_{ext} is the external stimulus, J_{Na} (J_{MNa}) is the sodium current (the coral color curve), J_K (J_{MK}) is the potassium current (the blue curve), J_L (J_{ML}) is the leaky current (the green curve), and J_m (J_M) is the total current (the purple curve) flowing through the cell membrane in the HH (MHH) spiking model. V

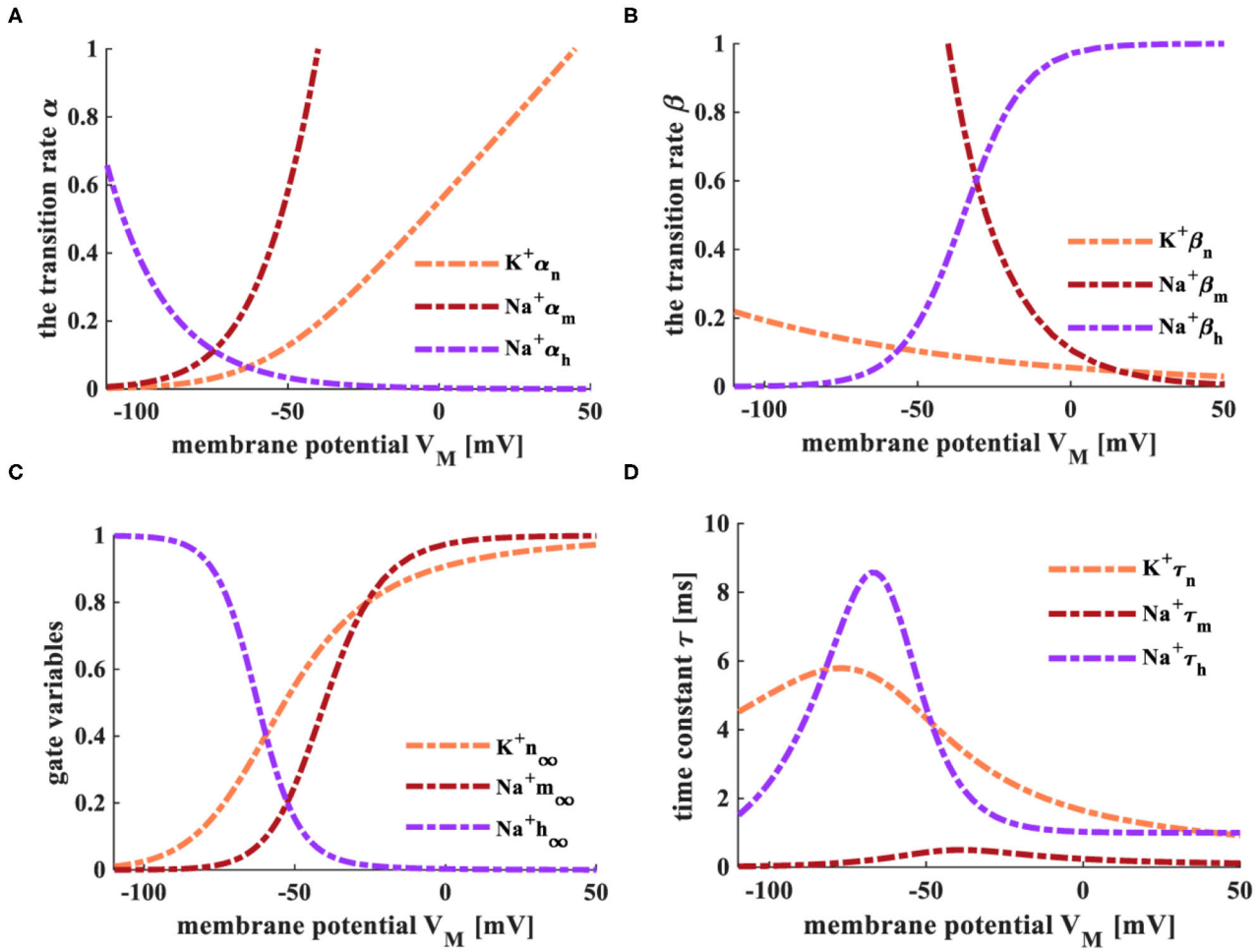


FIGURE 9 | The transition rate, gate variables, and time constant of the MHH spiking model. **(A)** The variation process of the transition α. **(B)** The variation process of the transition rate β. **(C)** The variation process of gate variables. **(D)** The variation of the time constant τ.

(V_M) is the action potential generated by the HH (MHH) spiking model. g_{Na} (g_{MNa}) is the sodium conductance (the sodium memductance), and g_K (g_{MK}) is the potassium conductance (the potassium memductance) in the HH (MHH) spiking model.

The HH and MHH spiking models receive the external stimuli and produce the corresponding current densities of the ion channels. The sodium current is negative because the sodium ions move from the outside to the inside of the cell. In contrast, the potassium current is positive because the potassium ions flow from intracellular to extracellular. The potassium and total current densities (the peak values: J_K = 0.82 $mAcm^{-2}$, J_m= -0.51 $mAcm^{-2}$) generated by the HH spiking model are larger than those (the peak values: J_{MK} = 0.4 $mAcm^{-2}$, J_M = -0.53 $mAcm^{-2}$) in the MHH spiking model. The sodium and leaky current densities (the peak values: J_{Na}= -0.7 $mAcm^{-2}$, J_L= 0.024 $mAcm^{-2}$) generated by the HH spiking model are smaller than those (the peak values: J_{MNa}= -0.6 $mAcm^{-2}$, J_L= 0.026 $mAcm^{-2}$) in the MHH spiking model. The sodium current of the MHH model has a smooth perturbation at around t = 1.072 s, and the sodium current of the HH model has an obvious perturbation at around t = 1.279 s. The perturbation is caused by the rapid variation of potassium conductance (potassium memductance). The curves formed by the MHH model (the left side plot in **Figure 11B**) are smoother than those in the HH model (the left side plot in **Figure 11A**) because the memristor has a unique time-varying property.

The HH spiking model and the MHH spiking model can perform the action potential. The membrane potential peak value (V_M = 38.33 mV at 1.188 ms) of the MHH model (the middle plot in **Figure 11B**) is stronger than that (V_m = 28.31 mV at 1.366 ms) of the HH model (the central plot in **Figure 11A**). Meanwhile, the MHH spiking model takes a short time to produce the action potential. After generating the action potential, both models return to the equilibrium state (the resting state, V_{rest} = -65 mV).

The HH spiking model takes 1.354 ms to reach the maximum value of g_{Na} (23.53 $mOhm^{-1}cm^{-2}$) and needs 1.715 ms to get the peak value of g_K (12.45 $mOhm^{-1}cm^{-2}$; the right side plot in **Figure 11A**). Therefore, the MHH spiking model takes 1.134

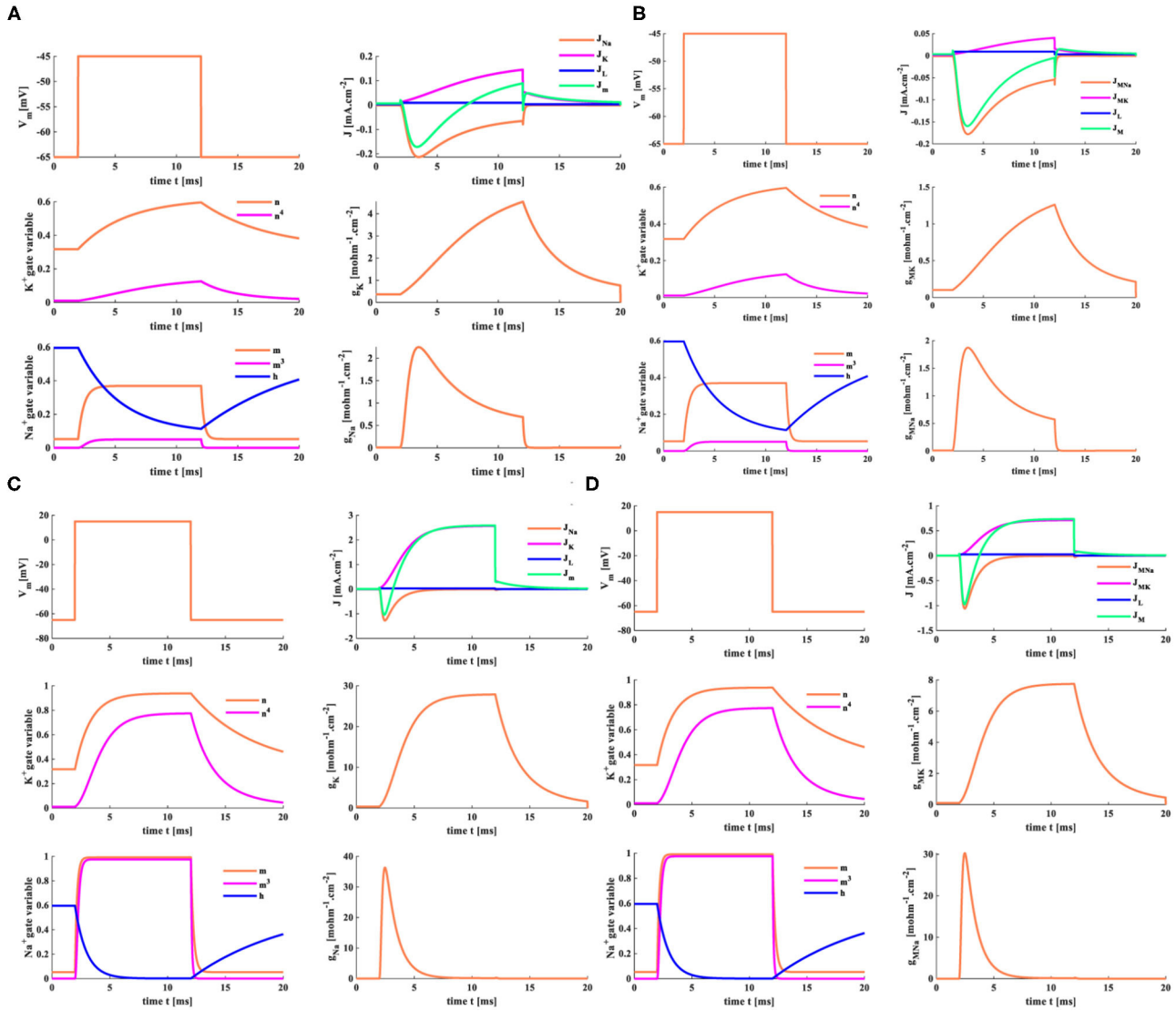


FIGURE 10 | The distinct clamp voltages are applied to the HH spiking model and the MHH spiking model. **(A)** The HH model with V_{clamp} = 20 mV. **(B)** The MHH model with V_{clamp} = 20 mV. **(C)** The HH model with V_{clamp} = 80 mV. **(D)** The MHH model with V_{clamp} = 80 mV.

ms to attain the maximum value of g_{MNa} (20.81 $mOhm^{-1}cm^{-2}$) and needs 1.673 ms to reach the peak value of g_{MK} (5.196 $mOhm^{-1}cm^{-2}$) (the right side plot in **Figure 11B**). The rise in sodium conductance (sodium memductance) is faster than potassium conductance (potassium memductance). The MHH spiking model utilizes less time than the HH model to activate the change of the memductance; however, the obtained memductane is small. Because the variation in the memductance is slight in a short time (5 ms), it maintains a large memristance.

5.2. The Reverse Single Current Pulse Stimulus

The reverse stimulus ($J_{ext} = -0.1\ mAcm^{-2}$, the pulse width is 0.1 ms) acts on the HH spiking model and the MHH spiking model, the temperature is 18.5°C, and the response time of the model is 5 ms.

There are not enough ions to move from intracellular (extracellular) to extracellular (intracellular); therefore, the sodium current and the potassium current cannot be produced (the left-side plots in **Figures 11C,D**). The significant variation of the conductance causes the generation of potassium and sodium currents. The sodium conductance (sodium memductance) is close to zero (the right-side plots in **Figures 11C,D**). The potassium conductance (potassium memductance) decreases from 0.37 $mOhm^{-1}cm^{-2}$ (0.36 $mOhm^{-1}cm^{-2}$) to 0.17 $mOhm^{-1}cm^{-2}$ (0.14 $mOhm^{-1}cm^{-2}$) and then increases to 0.35 $mOhm^{-1}cm^{-2}$ (0.26 $mOhm^{-1}cm^{-2}$). The HH and MHH spiking models are unable to generate the action potential, and the membrane potentials become hyperpolarization

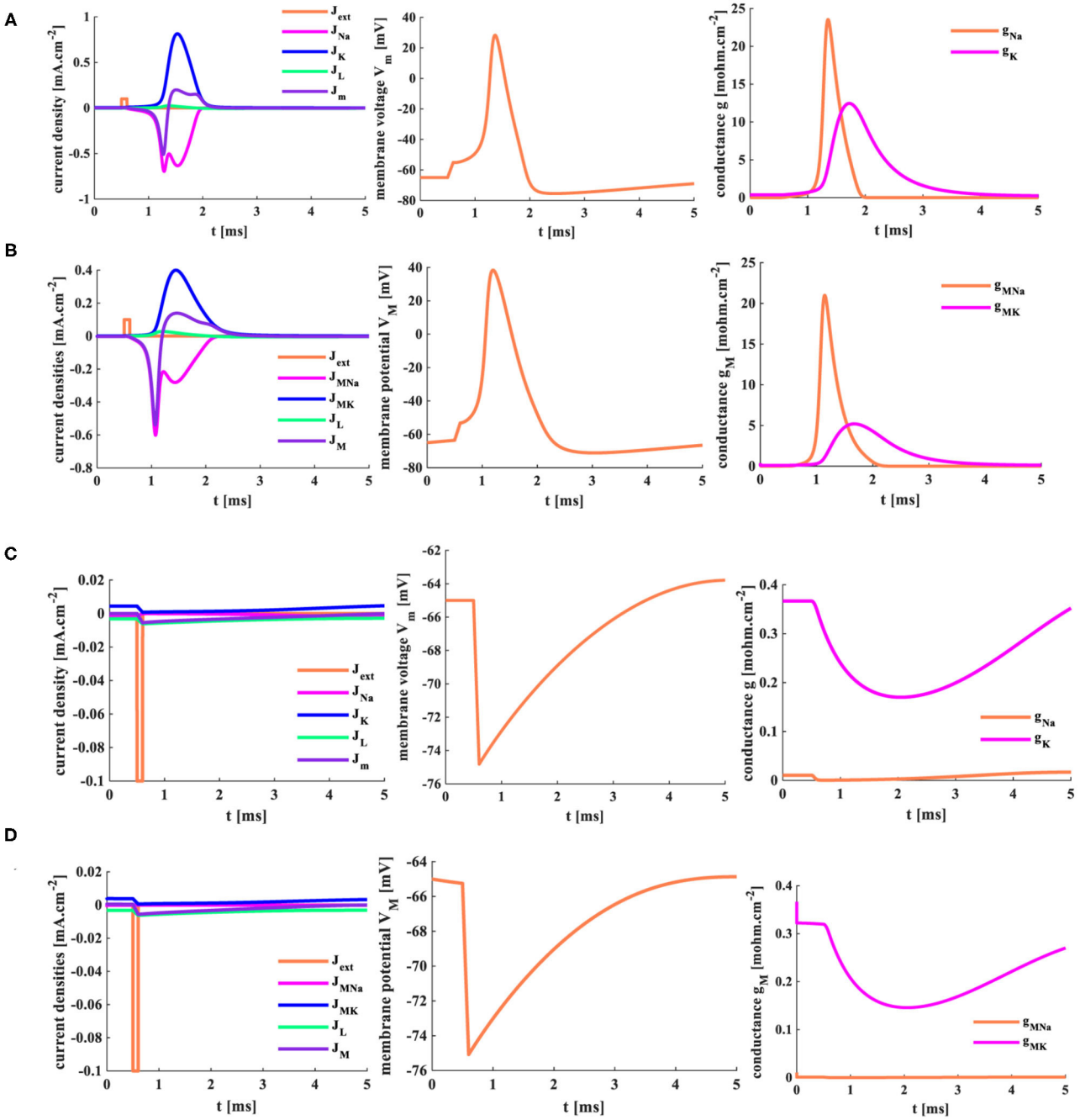


FIGURE 11 | The single pulse and the reverse pules are applied to two models. **(A)** The single current pulse is injected into the HH spiking model. **(B)** The single current pulse is injected into the MHH spiking model. **(C)** The reverse current pulse is injected into the HH spiking model. **(D)** The reverse current pulse is injected into the MHH spiking model.

before returning to their resting states (the middle plots in **Figures 11C,D**.

5.3. The Three External Stimuli With Different Intensity

The external stimuli J_{ext1} = 0.5 $mAcm^{-2}$, J_{ext2} =1 $mAcm^{-2}$, and J_{ext3} =2 $mAcm^{-2}$ are injected into the HH spiking model and the MHH spiking model, the temperature is 18.5°C, the response time is 5 ms.

When the small external stimulus (J_{ext1} = 0.5 $mA.cm^{-2}$) is applied to the HH spiking model, the action potential cannot be produced. The membrane potential has a slight rise (V_m = −60 mv) and then returns to the resting potential (−65 mv) at 3 ms (the second plot in **Figure 12A**). The current density

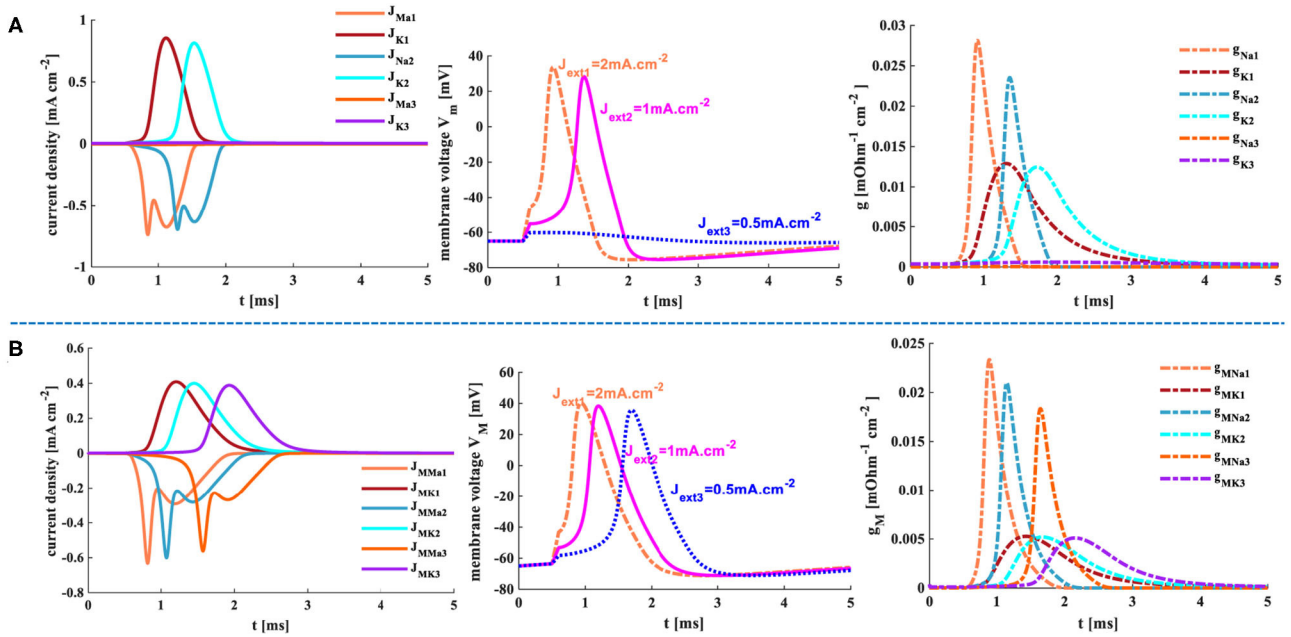


FIGURE 12 | The three different current pulses are applied to the HH and MHH spiking models. **(A)** The HH spiking model. **(B)** The MHH spiking model.

is zero (the first plot in **Figure 12A**). There is only a slight change in the conductance, which can be ignored (the third plot in **Figure 12A**). However, when the MHH spiking receives the stimulus J_{ext1} = 0.5 $mA.cm^{-2}$, the action potential is obtained (the second plot in **Figure 12B**). The changes in current densities and the memductance are noticeable. When the external stimuli increase to J_{ext2} = 1 $mA.cm^{-2}$ and J_{ext1} = 2 $mA.cm^{-2}$, the values in current density, membrane potential, and conductances strengthen gradually (**Figure 12**).

The larger the external stimulus, the faster the action potential is produced, the higher the peak value is generated, the more significant change in conductances, and the greater the current density. The smaller the external stimulus, the longer time it takes to produce the action potential. The peak value of membrane potential in the MHH model (the middle plot in **Figure 12B**) is greater than that of the HH model (the middle plot in **Figure 12A**). The maximum values of current densities and conductances in the MHH spiking model (the first and third plots in **Figure 12B**) are lower than those in the HH spiking model (the first and third plots in **Figure 12A**).

5.4. A Series of Pulse Stimuli

When a series of pulses (J_{ext}(n)= 1$mAcm^{-2}$, n = 1,2,.....,18, the temperature is 18.5°C.) act on the HH and MHH spiking models, the action potentials are achieved. However, not every single pulse can cause the generation of the action potential (the first plots in **Figures 13A,B**). Only when the action potential generated by the previous pulse has enough time to return to its resting state, another action potential will be generated. The MHH spiking model [six action potentials (the second plot in **Figure 13B**)] generates more action potentials than the HH spiking model (five action potentials (the second plot in **Figure 13A**)). Meanwhile, the action potential performs two oscillation behaviors in the MHH spiking model (inside the blue ellipse in **Figure 13B**), and the action potential shows three oscillation behaviors in the HH spiking model (inside the blue ellipse in **Figure 13A**). The memductances in the MHH model (the third plot in **Figure 13B**) are smaller than those in the HH model (the third plot in **Figure 13A**), which causes the current density produced by the MHH model (the first plot in **Figure 13B**) to be lower than the HH model (the first plot in **Figure 13A**).

The action time of the external stimulus is extended to 100 ms, and two models can produce more action potentials than **Figures 13A,B**. The MHH spiking model generates more action potentials (the middle plot in **Figure 13C**) than the HH spiking model (the middle plot in **Figure 13D**).

The action time is increased to 200 ms, the doublet currents (Shigaki et al., 2020) are generated in the MHH spiking model, one is large, the other is small (the enlarged plot inside the left ellipse in **Figure 13F**). Meanwhile, the action potential is produced before the current pulse comes in the MHH model because the memristor has an initial charge even though it is very small (the enlarged plot inside the right ellipse in **Figure 13F**). The current intensity, the voltage peak value, and conductances in the HH spiking model (**Figure 13E**) are larger than the simulation results in **Figure 13F**.

With the increasing of time length, the conductance (or memductance) and the current density of sodium and potassium increase dramatically. The more time we give, the more

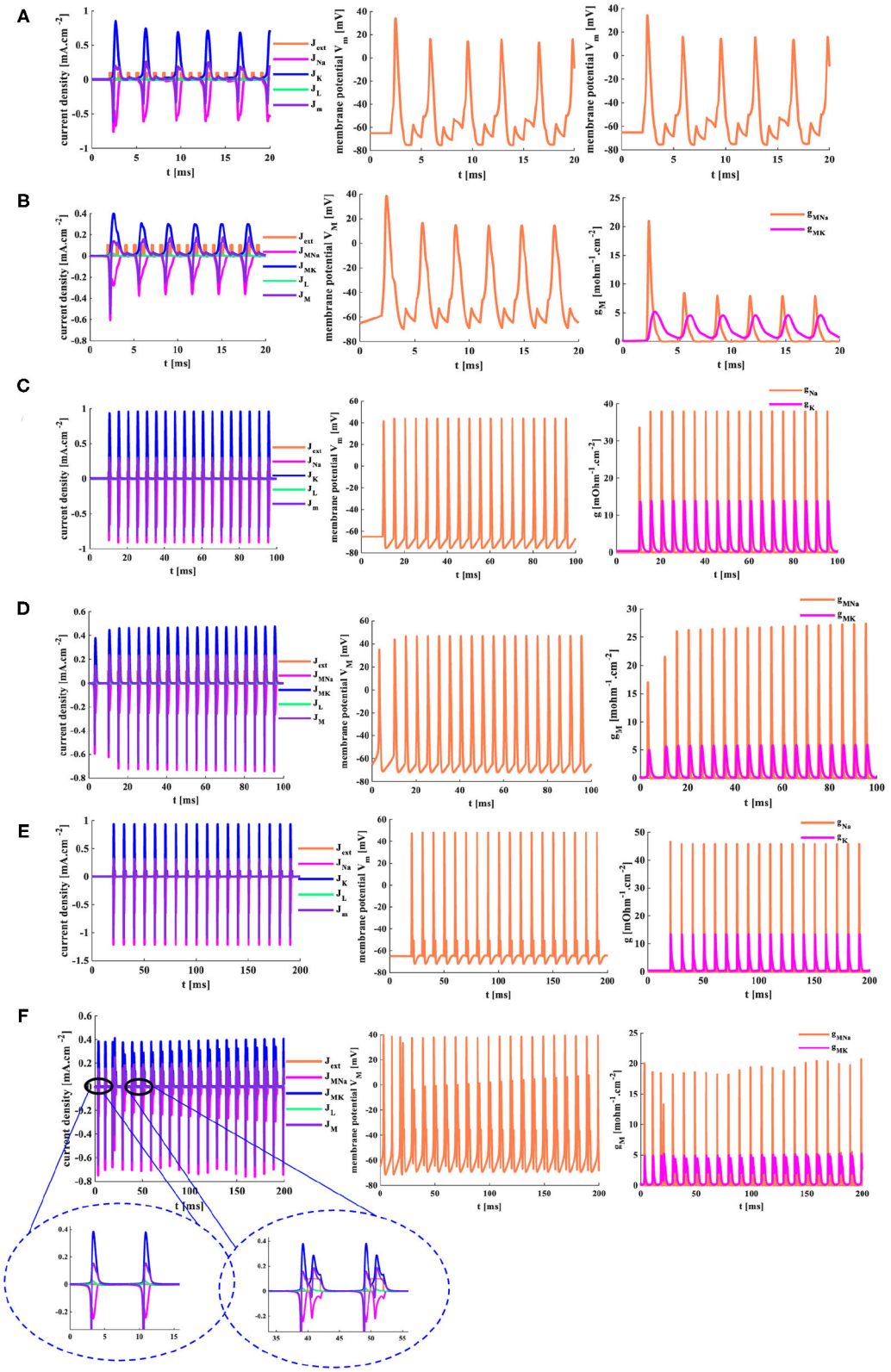


FIGURE 13 | The distinct action time of the external stimulus is set for the two models. **(A)** The HH spiking model with 20 ms action time. **(B)** The MHH spiking model with 20 ms action time. **(C)** The HH spiking model with 100 ms action time. **(D)** The MHH spiking model with 100 ms action time. **(E)** The HH spiking model with 200 ms action time. **(F)** The MHH spiking model with 200 ms action time.

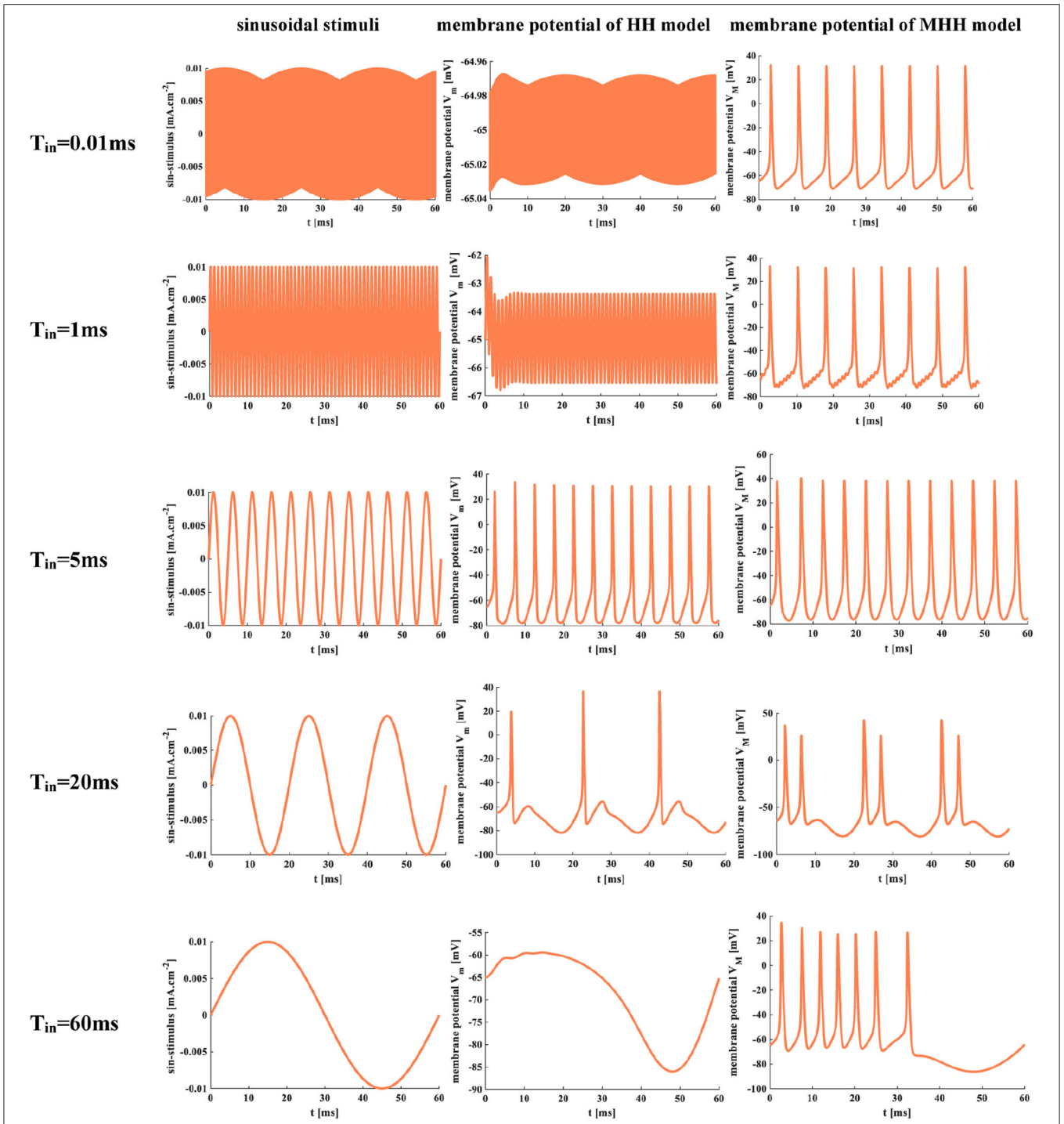


FIGURE 14 | The time periods of the sinusoidal signal. T_{in} = 0.01 ms, T_{in} = 1 ms, T_{in} = 5 ms, T_{in} = 20 ms, T_{in} = 60 ms.

action potentials are generated, the larger the peak values of current densities, conductances (or memductances), and action potentials. However, the action time length should not be too long; otherwise, the function of neurons cannot be replicated effectively (Chen et al., 2019).

5.5. The Sinusoidal Current Stimulus

The sinusoidal stimulus [$J_{ext} = J_{extm} \times \sin(2\ t/T_{in})$, J_{extm}=0.01 $mA.cm^{-2}$] is a positive-negative periodic signal with a single-frequency component. T_{in} is the time period of input signals, and the temperature is 18.5°C.

When T_{in} = 0.01 ms and T_{in} = 1 ms, the sinusoidal stimuli are applied to the HH spiking model. The action potential cannot be obtained because there is not enough time for the neuron to depolarize. But the MHH model generates action potentials under the same conditions. The frequency of the sinusoidal stimulus affects the generation of the action potential. When the frequency is low, there is sufficient time to depolarize, and the action potential occurs (**Figure 14**). When T_{in} = 5 ms, the HH and MHH spiking models produce the action potentials, their spiking patterns belong to tonic spikes in pyramidal neurons. When T_{in} = 20 ms, the MHH model generates the repetitive bursts with doublet spikes, and the HH model performs the tonic spiking. When the value of T_{in} is increased to 60 ms, the action potential cannot be produced in the HH spiking model but can be obtained in the MHH model. The frequency range of the sinusoidal stimulus in the MHH spiking model is wider than that of the HH spiking model. The various spiking patterns can be obtained by appropriately adjusting the frequency of the sinusoidal signal.

6. CONCLUSION

The biological neuron is expressed adequately by the classic HH spiking model. It is sensitive to the temperature, the strength of the external stimulus, and the action time of the stimulus. The MHH spiking model successfully simulates the generation of the action potential in a neuron. When the different external stimuli are applied to the HH and MHH spiking models, the action potential is produced, and various spiking patterns are achieved. The MHH spiking model has advantages in generating the action potential through the comparison with the HH spiking model. The waveforms with smaller perturbations formed by the MHH spiking model are smooth. The higher frequency of the external stimulus, the more action potentials generated. The response speed of the MHH spiking model is faster than that of the HH spiking model. The various spiking behaviors are obtained by adjusting the signal frequency in the MHH spiking model. And meanwhile, the combination between neuron models and a memristor provides the possibility to scale down the neuron circuit and gives a novel way to replicate the functions of the biological neuron.

DATA AVAILABILITY STATEMENT

The original contributions generated for the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

XF built models and simulations, carried out the experimental analysis, and prepared the manuscript in this work. SD and LW supervised the content of the article and the results of the simulations. All authors contributed to the article and approved the submitted version.

FUNDING

Project supported by the National Key R and D Program of China (Grant No. 2018YFB1306600), the National Natural Science Foundation of China (Grant Nos. 62076207, 62076208, U20A20227), the Fundamental Science and Advanced Technology Research Foundation of Chongqing, China (Grant No. cstc2017jcyjBX0050).

REFERENCES

Beck, M. E., Shylendra, A., Sangwan, V. K., Guo, S., Gaviria Rojas, W. A., and Yoo, H., et al. (2020). Spiking neurons from tunable gaussian heterojunction transistors. *Nat. Commun.* 11:1565. doi: 10.1038/s41467-020-15378-7

Berdan, R., Khat, A., Papavassiliou, C., and Prodromakis, T. (2014). "Qualitative spice modeling accounting for volatile dynamics of TIO2 memristors," in *IEEE International Symposium on Circuits and Systems (ISCAS)* (Melbourne, VIC), 2033–2036.

Berdan, R., Vasilaki, E., Khiat, A., Indiveri, G., Serb, A., and Prodromakis, T. (2016). Emulating short-term synaptic dynamics with memristive devices. *Sci. Rep. Nat.* 6:18639. doi: 10.1038/srep18639

Börgers (2017). An introduction to modeling neuronal dynamics. *Texts Appl. Math.* 66, 1–42. doi: 10.1007/978-3-319-51171-9

Buhry, L., Grassia, F., Giremus, A., Grivel, E., Renaud, S., and Saghi, S. (2011). Automated parameter estimation of the Hodgkin-Huxley model using the differential evolution algorithm: application to neuromimetic analog integrated circuits. *Neural Comput.* 23, 2599–2625. doi: 10.1162/NECO_a_00170

Chen, B. S., and Li, C. W. (2010). On the noise-enhancing ability of stochastic hodgkin-huxley neuron systems. *Neural Comput.* 22, 1737–1763. doi: 10.1162/neco.2010.07-09-1057

Chen, S., Zou, Y., and Zhang, X. (2019). An efficient method for hopf bifurcation control in fractional-order neuron model. *IEEE Access* 7, 77490–77498. doi: 10.1109/ACCESS.2019.2920007

Chua, L. (1971). Memristor–the missing circuit element. *IEEE Trans. Circuit Theory* 18, 507–519. doi: 10.1109/TCT.1971.1083337

Corinto, F., and Forti, M. (2017). Memristor circuits: bifurcations without parameters. *IEEE Trans. Circuits Syst. I* 64, 1540–1551. doi: 10.1109/TCSI.2016.2642112

Corinto, F., Gilli, M., and Forti, M. (2018). Flux-charge description of circuits with non-volatile switching memristor devices. *IEEE Trans. Circuits Syst. II* 65, 642–646. doi: 10.1109/TCSII.2018.2825447

Dröscher, C. (1998). Camillo golgi and the discovery of the Golgi apparatus. *Histochem. Cell Biol.* 109, 425–430. doi: 10.1007/s004180050245

Emili, Schiavoni, Francavilla, Roselli, and Sorrentino (2003). Computation of electromagnetic field inside a tissue at mobile communications frequencies. *IEEE Trans. Microw. Theory Techn.* 51, 178–186. doi: 10.1109/TMTT.2002.806899

Ghavami, S., Rahmati, V., Lahouti, F., and Schwabe, L. (2018). Neuronal synchronization can control the energy efficiency of inter-spike interval coding. *IEEE Trans. Mol. Biol. Multi-Scale Commun.* 4, 221–236. doi: 10.1109/TMBMC.2019.2937291

Häusser, M. (2000). The Hodgkin-Huxley theory of the action potential. *Nat. Neurosci.* 3, 1165–1165. doi: 10.1038/81426

Hill, B. (1992). *Ionic Channels of Excitable Membranes*. Sunderlan, MA: Sinauer Assoociates.

Hodgkin, A. L., and Huxley, A. F. (1952). The components of membrane conductance in the giant axon of loligo. *J. Physiol.* 116, 473–496. doi: 10.1113/jphysiol.1952.sp004718

Hodgkin, A. L., and Huxley, A. F. (1989). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Bull. Math. Biol*. 52, 25–71. doi: 10.1016/S0092-8240(05)80004-7

James, A. P. (2019). A hybrid memristor-CMOS chip for AI. *Nat. Electron*. 2, 268–269. doi: 10.1038/s41928-019-0274-6

Le, Z., Sangho, S., and Sung-Mo, K. (2015). "Memristor-based synapses and neurons for neuromorphic computing," in *IEEE International Symposium on Circuits and Systems (ISCAS)* (Lisbon), 1150–1153.

Li, C., Wang, Z., Rao, M., Belkin, D., Song, W., Jiang, H., et al. (2019). Long short-term memory networks in memristor crossbar arrays. *Nat. Mach. Intell*. 1, 49–57. doi: 10.1038/s42256-018-0001-4

Maheshwar, P. S., Kim, H., and Chua, L. O. (2014). Brains are made of memristors. *Circuits Syst. Mag. IEEE* 14, 12–36. doi: 10.1109/MCAS.2013.2296414

Mandal, S. and Saha, A. (2016). "Memristors act as synapses in neuromorphic architectures," in *IEEE International Conference on Communication and Electronics Systems (ICCES)* (Coimbatore), 1–4. doi: 10.1109/CESYS.2016.7889862

Mokhtar, S. M. A., Abdullah, W. F. H., Kadiran, K. A., Rifin, R., and Omar, M. (2017). "Programmable delay element using memristor and case study in delay lock loop," in *Control & System Graduate Research Colloquium* (Shah Alam), 17–21. doi: 10.1109/ICSGRC.2017.8070560

Petras, I. (2010). Fractional-order memristor-based Chua's circuit. *IEEE Trans. Circuits Syst. II* 57, 975–979. doi: 10.1109/TCSII.2010.2083150

Pu, S., and Thomas, P. J. (2020). Fast and accurate Langevin simulations of stochastic Hodgkin-Huxley dynamics. *Neural Comput*. 32, 1775–1835. doi: 10.1162/neco_a_01312

Raviola, E., and Mazzarello, P. (2011). The diffuse nervous network of camillo golgi: facts and fiction. *Brain Res. Rev*. 66, 425–430. doi: 10.1016/j.brainresrev.2010.09.005

Saïgai, S., Bornat, Y., Tomas, J., Le Masson, G., and Renaud, S. (2011). A library of analog operators based on the hodgkin-huxley formalism for the design of tunable, real-time, silicon neurons. *IEEE Trans. Biomed. Circuits Syst*. 5, 3–19. doi: 10.1109/TBCAS.2010.2078816

Sayyaparaju, S., Amer, S., and Rose, G. S. (2018). "A bi-memristor synapse with spike-timing-dependent plasticity for on-chip learning in memristive neuromorphic systems," in *IEEE19th International Symposium on Quality Electronic Design (ISQED)* (Santa Clara, CA), 69–74. doi: 10.1109/ISQED.2018.8357267

Shigaki, S., Ohashi, H., Sakurai, T., Shimizu, M., Hosoda, K., and Kurabayashi, D. (2020). Real-time odor discrimination using single antenna of insect. *IEEE Sensors Lett*. 4, 1–4. doi: 10.1109/LSENS.2020.3024606

Strukov, D. B., Snider, G. S., Stewart, D. R., and Williams, R. S. (2008). The missing memristor found. *Nature* 453, 80–83. doi: 10.1038/nature06932

Truong, S. C., Pham, K. V., and Yang, W. (2016). "Memristor circuits and systems for future computing and bio-inspired information processing," in *Biomedical Circuits & Systems Conference* (Shanghai), 456–459. doi: 10.1109/BioCAS.2016.7833830

Tuckwell, H. C., and Ditlevsen, S. (2016). The space-clamped Hodgkin-Huxley system with random synaptic input: Inhibition of spiking by weak noise and analysis with moment equations. *Neural Comput*. 28, 2129–2161. doi: 10.1162/NECO_a_00881

Wang, L., Drakakis, E., and Duan, S. (2012). Memristor model and its application for chaos generation. *Int. J. Bifurc. Chaos* 22:1250205. doi: 10.1142/S0218127412502057

Xia, Q., and Yang, J. J. (2019). Memristive crossbar arrays for brain-inspired computing. *Nat. Mater*. 18, 309–323. doi: 10.1038/s41563-019-0291-x

Yang, J. J., Strukov, D. B., and Stewart, D. R. (2013). Memristive devices for computing. *Nat. Nanotech*. 8, 13–24. doi: 10.1038/nnano.2012.240

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.

frontiers in Nanotechnology

ORIGINAL RESEARCH
published: 06 October 2021
doi: 10.3389/fnano.2021.734121

TCAD Modeling of Resistive-Switching of HfO_2 Memristors: Efficient Device-Circuit Co-Design for Neuromorphic Systems

Andre Zeumault [1,2], Shamiul Alam [1], Zack Wood [1], Ryan J. Weiss [1], Ahmedullah Aziz [1] and Garrett S. Rose [1]*

[1]Min H. Kao Department of Electrical Engineering and Computer Science, The University of Tennessee, Knoxville, TN, United States, [2]Department of Materials Science and Engineering, The University of Tennessee, Knoxville, TN, United States

OPEN ACCESS

Edited by:
Huanglong Li, Tsinghua University, China

Reviewed by:
Bin Gao, Tsinghua University, China
Fernando Garcia Redondo, Arm Ltd., United Kingdom

***Correspondence:**
Andre Zeumault
azeumault@utk.edu

Specialty section:
This article was submitted to Nanodevices, a section of the journal Frontiers in Nanotechnology

Received: *30 June 2021*
Accepted: *20 September 2021*
Published: *06 October 2021*

Citation:
Zeumault A, Alam S, Wood Z, Weiss RJ, Aziz A and Rose GS (2021) TCAD Modeling of Resistive-Switching of HfO_2 Memristors: Efficient Device-Circuit Co-Design for Neuromorphic Systems. Front. Nanotechnol. 3:734121. doi: 10.3389/fnano.2021.734121

In neuromorphic computing, memristors (or "memory resistors") have been primarily studied as key elements in artificial synapse implementations, where the memristor provides a variable weight with intrinsic long-term memory capabilities, based on its modifiable resistive-switching characteristics. Here, we demonstrate an efficient methodology for simulating resistive-switching of HfO_2 memristors within Synopsys TCAD Sentaurus—a well established, versatile framework for electronic device simulation, visualization and modeling. Kinetic Monte Carlo is used to model the temporal dynamics of filament formation and rupture wherein additional band-to-trap electronic transitions are included to account for polaronic effects due to strong electron-lattice coupling in HfO_2. The conductive filament is modeled as oxygen vacancies which behave as electron traps as opposed to ionized donors, consistent with recent experimental data showing p-type conductivity in HfO_x films having high oxygen vacancy concentrations and ab-initio calculations showing the increased thermodynamic stability of neutral and charged oxygen vacancies under conditions of electron injection. Pulsed IV characteristics are obtained by inputting the dynamic state of the system—which consists of oxygen ions, unoccupied oxygen vacancies, and occupied oxygen vacancies at various positions—into Synopsis TCAD Sentaurus for quasi-static simulations. This allows direct visualization of filament electrostatics as well as the implementation of a nonlocal, trap-assisted-tunneling model to estimate current-voltage characteristics during switching. The model utilizes effective masses and work functions of the top and bottom electrodes as additional parameters influencing filament dynamics. Together, this approach can be used to provide valuable device- and circuit-level insight, such as forming voltage, resistance levels and success rates of programming operations, as we demonstrate.

Keywords: memristor, neuromorphic, nanoelectronics, non-volatile memory, RRAM, Monte Carlo, TCAD Sentaurus

1 INTRODUCTION

In recent years, memristor devices have shown great potential for neuromorphic computing due to their resistive-switching dynamics and electrical behavior resembling that of biological synapses (Chua, 1971; Xia and Yang, 2019; Strukov et al., 2008). Memristors are resistors with memory whose resistance level can be controlled either through an applied voltage (i.e., flux-linkage) or current (i.e., charge-fluence). Oxide memristors tend to be voltage-controlled, having a metal-oxide-metal device structure identical to a parallel-plate capacitor. Since the oxide thickness tends to be thin (~ 2–5 nm) (Pi et al., 2019) and the switching speed can be very fast (<1 ns) (Choi et al., 2016), a small amount of energy is required for programming resistance states. With these unique features, in addition to non-volatility, they have shown the great promise for building energy and area efficient memristive crossbar arrays (1T1R arrays) to form neural networks for a wide range of applications including robotics, computer vision, and speech recognition (Yao et al., 2017; Li et al., 2018b,a; Hu et al., 2016). A 1T1R crossbar array (**Figure 1A**) offers added benefits due to the use of a transistor in each resistive RAM (RRAM) memory cell. The transistor plays a major role in mitigating the sneak current path and programming disturbance associated with resistive (i.e., 1R) crossbar arrays (Manem et al., 2012; Yao et al., 2015). Furthermore, the transistor's gate terminal in the 1T1R cell allows for better control over the current through the memristive device. It also provides more resilience to the switching voltage magnitude and attains better uniformity (Liu et al., 2014).

Electroforming, a one-time forming or initialization process, is often required in transition metal oxide (TMO) memristors (Strukov et al., 2008), which have been widely used for memristive crossbar arrays, including those used in neuromorphic systems. However, for forming, the voltage often needs to be higher than the nominal supply voltage of modern CMOS processes, which causes significant design and integration challenges (Amer et al., 2017b,a). Memristors with high forming voltages require dedicated circuitry capable of tolerating such high voltage levels for executing the in-field forming. Furthermore, the area constraints associated with the in-field forming circuitry undermines the density benefits of the crossbars. In addition, such transistors are generally large compared to the regular devices to accommodate these high forming voltages (**Figure 1B**). For example, for a 65 nm CMOS process (Beckmann et al., 2016; Amer et al., 2017a) used to prepare 80 nm × 80 nm memristor areas, the minimum length of the transistor used in the 1T1R cell could be as much as 0.5 μm to endure forming voltages up to 3.3 V. In contrast, the minimum length of the regular transistor used for peripheral circuitry is 60 nm with the nominal voltage 1.2 V or 1.0 V, depending on the process. Thus, researchers have focused on lowering the forming voltages to a level of operation that allows for better exploitation of memristive crossbar density (Govoreanu et al., 2011; Koveshnikov et al., 2012; Huang et al., 2013a; Chen, 2013; Kim et al., 2016; Amer et al., 2017c).

The electroforming process, in addition to reset and set operations, can be simulated from a condensed set of rate equations that define all possible changes of state of the system using Kinetic Monte Carlo (KMC). Rate equations used to model filament dynamics are typically based on the following physical transitions: oxygen ion (i.e., O^{2-}) and vacancy (i.e., V_O^{2+}) diffusion, and the generation and recombination of Frenkel pairs $\{V_O^{2+}, O^{2-}\}$. These have been implemented, successfully, by several authors for both 2D and 3D filaments with fitting capability to experimental data (Sementa et al., 2017; Aldana et al., 2018, 2020; Loy et al., 2020). The large difference in diffusion rates of oxygen ions and oxygen vacancies tends to favor filament growth along preexisting vacancy sites or positions in which the local electric field is high (e.g., grain boundaries or point defects) according to the thermochemical model of dielectric breakdown (McPherson et al., 2003). As a consequence, resulting filaments obtained from previous KMC approaches consist of positively charged oxygen vacancies resulting from the repetition of: 1) breaking Hf–O bonds and 2) the formation of Frenkel pairs consisting of nearly stationary oxygen vacancies and relatively diffuse, oxygen ions at interstitial sites. In other words, forming/set operations are thus determined by the local electric field—producing dendritic filament growth—whereas reset is determined by the coincidence of oxygen ion diffusion and recombination.

Despite growing evidence to the contrary, few modelling approaches allow the charge state of the oxygen vacancy (i.e., +2) to change during forming. In effect, oxygen vacancies are modeled as fixed charges for the purpose of determining filament evolution, yet as electron traps for the purpose of calculating current, which is based on nonlocal multiple-phonon trap-assisted tunneling in which trap occupancy dynamics is fundamentally important. This inconsistency greatly limits the utility of existing approaches to provide increased physical insight into HfO_x switching behavior–beyond that which existing compact models already provide (Bianchi et al. (2020); Yu and Wong (2010); Huang et al. (2013b); Guan et al. (2012b); Jiang et al. (2014))–which can be extended to device design, circuits and systems-level refinements (e.g., reducing forming voltage). The assumption of a static positive charge contradicts experimental evidence showing p-type conductivity in highly defective HfO_x films (Hildebrandt et al., 2011)—suggesting that oxygen vacancies are deep acceptor-like traps (>3 eV from the conduction band edge). Moreover, ab-inito calculations have shown the thermodynamic stability of neutral and negatively charged vacancy states increases in conditions of electron injection (i.e., current flow) due to electron capture (Bradley et al., 2015). This is consistent with experimental work using *in-situ* TEM electron holography and EELS in which oxygen-vacancy filaments were observed, spatially, as regions of negative space-charge (Li et al., 2017). Unlike previous KMC approaches, together these observations are, in fact, self-consistent with the physical assumptions used to model current flow in HfO_2 based memristive devices, in which, conduction occurs through a nonlocal, trap-assisted tunneling process involving electron capture and emission—appropriate for insulators having point defects (e.g., TaO_x, HfO_x, ZrO_x, NbO_x).

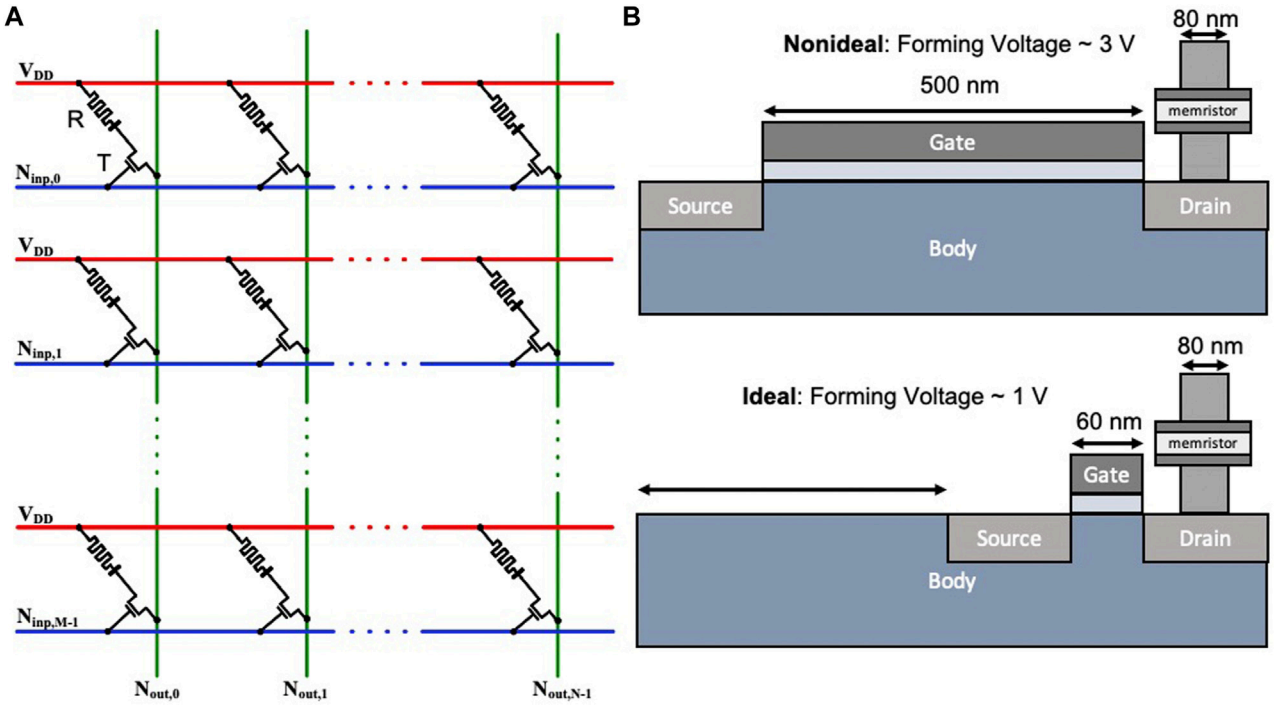


FIGURE 1 | (A) A *M* × *N* crossbar array of one-transistor, one-memristor (1T1R) devices for neuromorphic applications. **(B)** Illustration of the size of transistors needed for memristors having high and low forming voltages.

Here, using a simple 2D model, we show that, in addition to the conventional set of rate equations (i.e., Frenkel pair generation/recombination and diffusion) filament evolution in HfO_x can be modeled self-consistently as the result of band-to-trap electron capture and emission processes between the electrodes and oxygen vacancies (**Figure 2**). In this way, the conductive filament consists of occupied oxygen vacancy electron traps, which lower their energy upon electron capture due to strong coupling between ionized defects and the lattice in HfO_x (Huang-Rhys factor, S = 17) as depicted in **Figure 2**. The primary benefit in this approach is that additional parameters associated with the electrodes (e.g., work function, effective mass) and those of the oxygen vacancy states (e.g., trap energy level, capture cross-section, thermal barrier and binding energy) are intimately linked to resistive switching behavior, as we show. Not only do these additional parameters provide more depth in terms of physical insight and modeling capability, they are readily accessible experimentally or through ab-initio estimates. Using TCAD Sentaurus (Synopsys, 2019), we demonstrate that the common forming, reset and set characteristics can be successfully reproduced and visualized. In particular, we show that certain regions within the filament have a negative potential–stemming from a negative space charge due to electron capture. This is consistent with recent experimental work describing the filament as a *negative potential synapse* (Li et al., 2017). Next, we couple our device model with a phenomenological compact model to bring in physics-based insights to the circuit-level simulation of a memristor-based synapse topology. Finally, to underscore the unique strength of our model, we investigate a device-circuit co-design strategy powered by Monte-Carlo simulations with different levels of initial oxygen vacancy volume faction.

2 METHODS

2.1 Material Specifications and Device Geometry

A complete list of parameters used to specify the HfO_2 layer are provided in **Table 1**. Of these, the parameters related to hafnium-oxygen bond energy and polarization were obtained from the thermochemical model (McPherson et al., 2003), assuming a 100% monoclinic phase composition. This makes clear the assumptions regarding crystalline phase of the HfO_2 thin film and the activation enthalpy required for breaking the hafnium-oxygen bond–which differs due to differences in polarization and the number of bonds required to be broken about the Hf central atom. It should be noted that, in practice, a mixture of monoclinic and tetragonal phases are present in varying ratios–the monoclinic phase has a nominal breakdown field of 6.7 MV cm^{-1} whereas the tetragonal phase has a breakdown field of 3.9 MV cm^{-1} (McPherson et al., 2003). Empirically, the breakdown field is found to vary between 3 and 5 MV cm^{-1} (Sire et al., 2007) for atomic layer deposition (ALD) grown HfO_2 films on TiN of comparable thickness used in this work.

As indicated in **Figure 3A**, the memristor is modeled as a two-dimensional top-electrode (TE)/HfO_x/bottom-electrode (BE)

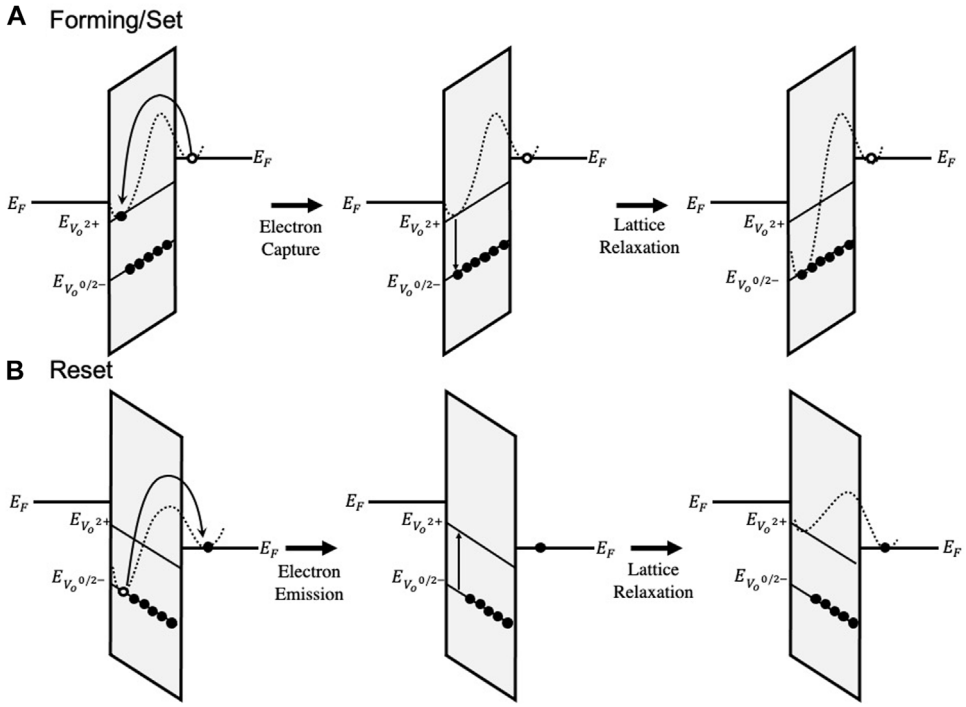


FIGURE 2 | Depiction of electron capture and emission as band-to-trap and trap-to-band, respectively for a trap located below the Fermi level of the cathode (i.e. $\Delta E_{TF} < 0$). **(A)** The forming and set process is facilitated by electron capture and lattice relaxation. **(B)** The reset process is facilitated by electron emission and lattice relaxation.

TABLE 1 | Summary of nominal materials parameters used in this work unless otherwise stated.

Parameter	Description	Value and unit
T	Lattice Temperature	300 K
ϵ_r	Relative permittivity	21
f_{ph}	Attempt-to-escape frequency	10 THz
$E_{a,V_O^{2+}}$	Activation energy for V_O^{2+} diffusion	1.5 eV
$E_{a,O^{2-}}$	Activation energy for bulk O^{2-} diffusion	0.7 eV
$E_{a,O^{2-}}$	Activation energy for interfacial O^{2-} diffusion	0.375 eV
$E_{a,g,bulk}$	Activation energy for Frenkel pair generation (bulk)	4.50 eV
$E_{a,g,pair}$	Activation energy for Frenkel pair generation (pair)	2.97 eV
$E_{a,r,bulk}$	Activation energy for Frenkel pair recombination (bulk)	0.2 eV
$E_{a,r,pair}$	Activation energy for Frenkel pair recombination (pair)	0.83 eV
$E_{a,t}$	Activation energy for V_O^{2+} capture cross-section	0.1 eV
$E_{a,get.}$	Activation energy for gettering of oxygen at TE/oxide interface	0.1 eV
σ_0	capture cross-section for V_O^{2+}	1×10^{-16} cm^2
n_{be}	electron concentration of bottom electrode	1×10^{23} cm^{-3}
Δr_i	Jump distance for V_O^{2+} and O^{2-}	3 Å
$\vec{p}_0$	HfO_2 molecular dipole moment	11×10^{-10} Cm
S	Huang-Rhys factor	17
$\hbar\omega_0$	Optical phonon energy	0.07 eV
E_g	HfO_2 Bandgap energy	5.9 eV
E_t	Trap level of V_O^{2+} relative to conduction-band edge	3.0 eV
ΔE_t	Trap level reduction due to lattice relaxation	0.2 eV
χ	HfO_2 electron affinity	2 eV
x_0	Initial volume-fraction of V_O^{2+} defects	0.0002
N_i	Concentration of oxygen vacancies and oxygen ions	1×10^{18} cm^{-3}

TABLE 2 | Summary of transitions rates modelled using Kinetic Monte Carlo procedure and their parameters. Lattice coordinates are listed as relative positions to a given lattice point at (*i*, *j*) following the convention of the lattice gas model (Jansen, 2012). * = site, O^{2-} = oxygenion, V_O^{2+} = positively charged oxygen vacancy (unoccupied), V_O^{2-} = negatively c harged oxygen vacancy (occupied)

Transition	Reaction	Parameters
Oxygen Vacancy Diffusion	$(0,0),(\pm1,0): V_O^{2+} * \rightarrow * V_O^{2+}$ $(0,0),(0,\pm1): V_O^{2+} * \rightarrow * V_O^{2+}$	$f_{ph}, E_{a,V_O^{2+}}, \Delta r_{V_O^{2+}}$
Oxygen Ion Diffusion	$(0,0),(\pm1,0): O_V^{2+} * \rightarrow * O_V^{2+}$ $(0,0),(0,\pm1): O_V^{2+} * \rightarrow * O_V^{2+}$	$f_{ph}, E_{a,O^{2-}}, \Delta r_{O^{2-}}$
Frenkel Pair Generation	$(0,0),(\pm1,0): * * \rightarrow V_O^{2+} O^{2-}$ $(0,0),(0,\pm1): * * \rightarrow V_O^{2+} O^{2-}$	$f_{ph}, E_{a,g}, \vec{p}_0, \epsilon_r$
Frenkel Pair Recombination	$(0,0),(\pm1,0): V_O^{2+} O^{2-} \rightarrow * *$ $(0,0),(0,\pm1): V_O^{2+} O^{2-} \rightarrow * *$	f_{ph}, $E_{a,r}$
Electron Capture	$(0,0): V^{2+} \rightarrow V^{2-}$	$\sigma_0, m^*_{n,cath.}, n_{cath.}, E_t, E_{a,t} \Phi_{cath.}$
Electron Emission	$(0,0): V^{2-} \rightarrow V^{2+}$	$\sigma_0, m^*_{n,cath.}, n_{cath.}, E_t, E_{a,t} \Phi_{cath.}$
Oxygen Gettering at TE/Oxide	$(0,0): O^{2-} \rightarrow *$	f_{ph}, $E_{a,get.}$

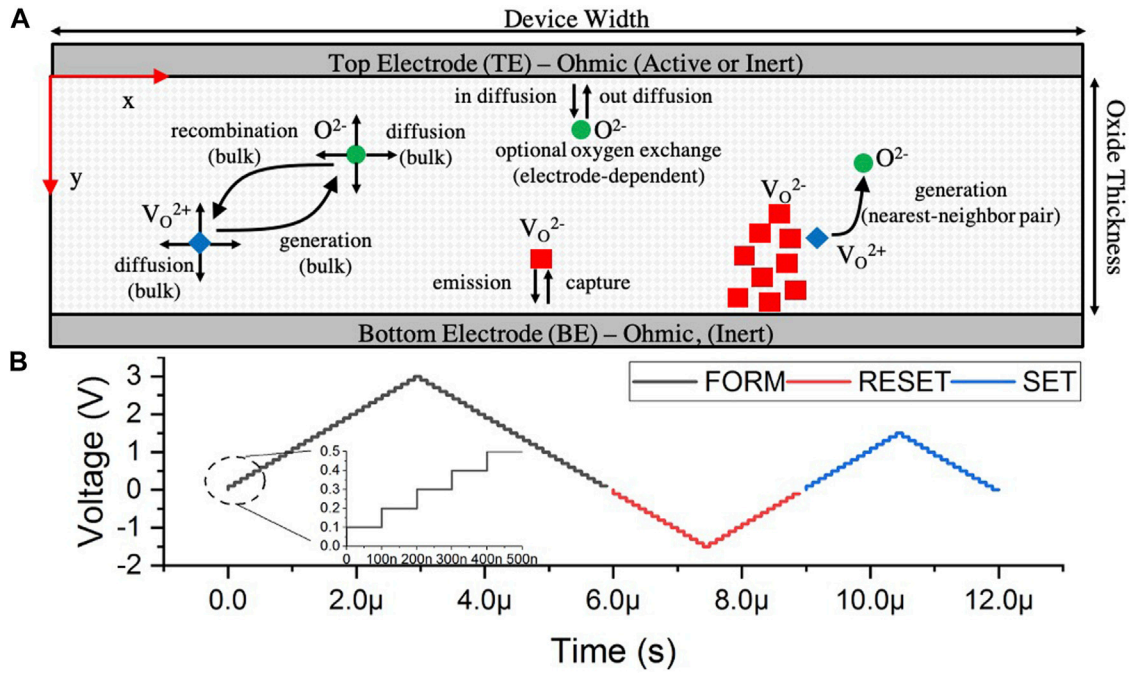


FIGURE 3 | (A) Illustration showing device geometry and description of the processes modeled using Kinetic Monte Carlo. **(B)** Example voltage waveform used to perform *forming*, *reset*, and *set* operations in sequence using a stepped voltage ramp with a KMC simulation time of 100 ns. The resulting state of the system at the end of each voltage increment is used as input to Synopsys TCAD Sentaurus.

structure on a square grid. The HfO_x thickness is 5 nm and the device width is 40 nm. The Ti (TE) and TiN (BE) electrodes are modeled as ideal, Ohmic contacts with a 0 Ω series resistance. The work function of the Ti and TiN layers were set to 4.33 and 4.5 eV respectively. The electron effective masses of the Ti and TiN layers were set to 3.2 and 2.0 respectively according to literature (Lima et al., 2012). Following the lattice-gas model, a grid point represents the smallest physical unit considered by this simulation, capable of representing either an empty "site" (for diffusion or the formation of a Frenkel pair), a positively charged oxygen vacancy (V_O^{2+}), a negatively charged oxygen ion interstitial (O^{2-}) or a negatively charged oxygen vacancy (V_O^{2-}) which also represents the conductive filament. Thus, field-independent transitions (e.g. Frenkel pair recombination) occur over nearest-neighbor distances whereas field-dependent transitions (e.g. Frenkel pair generation, ion diffusion) interact over many grid points through the screened Coulomb potential.

The initial state of the system can be defined by randomizing the location of oxygen vacancies and oxygen ions (needed to ensure charge-neutrality)–to represent an amorphous film. Alternatively, since it is known that ALD-deposited HfO_2 thin films exhibit a columnar grain

morphology (≈8 nm grain size (Ho et al., 2003)), oxygen vacancies can be placed along grain boundaries due to the reduced formation energy–to represent a polycrystalline film. The initial concentration of vacancies is determined by a variable volume fraction parameter which we nominally set to 0.0002 (i.e., 0.02 at. %). Our focus here is to demonstrate the key differences and advantages of our physical model incorporating additional electronic transitions to the constitutive rate equations describing filament dynamics and its implementation in TCAD Sentaurus.

2.2 Filament Evolution Under Voltage Stress

Filament evolution during forming, set and reset operations are described using a simple set of rate equations corresponding to the following physical processes, as outlined in **Table 2**:

- Electron capture or emission by oxygen vacancies
- Oxygen vacancy and ion diffusion
- Frenkel pair generation and recombination (isolated bulk and nearest neighbor pairs)
- Oxygen gettering by Ti

These processes are implemented via a classical KMC selection algorithm applied to a chosen initial state of the system (top electrode, bottom electrode, oxygen ion, oxygen vacancy and filament), and updated in time according to Poisson statistics and the time scale of each *selected* mechanism. The details of the dynamical monte carlo algorithm and the meaning of simulation time (Fichthorn and Weinberg, 1991), and its application to the formation of 2D/3D conductive filaments (metallic and oxygen vacancy) has been discussed elsewhere (Sementa et al., 2017; Aldana et al., 2018, 2020; Loy et al., 2020). Here, we provide a minimal outline of the essential aspects, assumptions and parameters of the rate equations we've implemented. In particular, we highlight the physical assumptions that establish consistency between phenomenological models of filament evolution and models of electric conduction—needed for efficient and accurate device-circuit co-design.

2.2.1 Filament Changes Mediated by Electron Capture/Emission

We model filament precipitation (dissolution) as the net result of Frenkel pair generation (recombination) and electron capture (emission) by oxygen vacancy electron traps. For simplicity, we couple oxygen vacancy traps to the conduction band in the bottom-electrode, which permits straightforward evaluation of rate equations within the electron trap picture. Conventionally, capture and emission rates are defined in terms of a thermally activated capture-cross section, a tunneling coefficient, and a field-dependent trap barrier that depends on the relative energy difference between the trap level and the Fermi level of the bottom electrode. In other words, the forming, set and reset operations are described as band-to-trap (or trap-to-band) electronic transitions within the Wentzel-Kramers-Brillouin (WKB) approximation **Eqs 1–3**.

$$R_c \approx \sigma_0 v_{th} n_{be} \exp\left(-\frac{y_t}{y_0}\right) \exp\left(-\frac{E_{a,t}}{k_B T}\right) \begin{cases} \exp\left(-\frac{E_{tf} - q\mathscr{E}_y y_t}{k_B T}\right) & E_{tf} > 0 \\ \exp\left(\frac{q\mathscr{E}_y y_t}{k_B T}\right) & E_{tf} < 0 \end{cases} \tag{1}$$

$$R_e \approx \sigma_0 v_{th} n_{be} \exp\left(-\frac{y_t}{y_0}\right) \exp\left(-\frac{E_{a,t}}{k_B T}\right) \begin{cases} \exp\left(-\frac{q\mathscr{E}_y y_t}{k_B T}\right) & E_{tf} > 0 \\ \exp\left(-\frac{-E_{tf} + q\mathscr{E}_y y_t}{k_B T}\right) & E_{tf} < 0 \end{cases} \tag{2}$$

$$E_{tf} \equiv E_t - E_{f,be} \tag{3}$$

These expressions are derived in the **Supplementary Material**. Here, y_t represents the y-coordinate of the trap relative to the electrode, y_0 is a parameter related to the wavefunction overlap between the electronic state in the trap and in the electrode. Within the WKB approximation, after applying the triangular barrier approximation for the bands we have the following:

$$y_0 = \left(2\int_0^{t_{ox}} \frac{\sqrt{2m^* q E_y y}}{\hbar} dy\right)^{-1} \approx \frac{3q\hbar E_y}{4\sqrt{2m^*}\left(\Phi_{be} - \chi_{ox}\right)^{3/2}} \tag{4}$$

The quantity $\Phi_{be} - \chi_{ox}$ represents the conduction band offset between the bottom electrode and the oxide, in terms of the work function of the bottom electrode Φ_{be} and the electron affinity of the oxide χ_{ox}.

In **Eqs 1–3**, it is assumed that the electric field has a symmetric influence on transition rates, that is, the barrier lowering in the forward direction is equal and opposite that of the reverse in order to maintain steady-state equilibrium. The case statements in **Eqs 1**, **2** exist since the trap may be higher (i.e., $E_{tf} > 0$) or lower ($E_{tf} < 0$) than the Fermi level in the bottom electrode. Parameters which depend on the bottom electrode are the electron concentration, n_{be} and the effective-mass, $m^*_{n,be}$, which enters through the thermal velocity (**Eq. 5**):

$$v_{th} = \sqrt{\frac{3k_b T}{m^*_{n,be}}} \tag{5}$$

We note that the time scale of electron capture and emission depends on the product of the carrier concentration in the electrodes, the thermal velocity and the capture cross section of oxygen vacancies as shown in **Eqs 1–3** through a common exponential prefactor. Using values listed in **Table 1**, the ratio of the exponential prefactors for electronic ($\sigma_0 n_B E v_{th}$) and atomic processes (f_{ph}) is evaluated to be 8.25, so electronic processes are expected to occur much faster than atomic ones. However, the rate of electronic transitions also depends on the local electric field and the position of the trap relative to the electrode and so the above estimate only reflects those traps that are close to the bottom electrode. Therefore, the relative rates of electronic and

atomic processes are expected to differ (generally reducing) as one moves from the BE to the TE under forming/set and from the TE to the BE under reset due to the change in voltage polarity.

The initial system consists of the electrodes and an initial concentration of positively charged, unoccupied oxygen vacancies. It should be noted that, although we have assumed a +2 charge state, the +2 oxygen vacancy is unstable in the presence of interstitial oxygen and/or conditions of electron injection. This is supported by ab-initio calculations suggesting that Frenkel pairs stabilize through the formation of neutral and/or negatively charged oxygen vacancies—facilitated by electron capture (Bradley et al., 2015). Experimentally, this is supported by the observation of p-type conductivity in highly defective HfO_x films, suggesting that oxygen vacancies interact and stabilize as deep acceptors (Hildebrandt et al., 2011) as opposed to shallow donors. Thus, the formation of conductive oxygen vacancy filaments can be regarded as thermodynamically driven by the increase in binding energy due to electron capture of single vacancies and expected to subsequently stabilize due to defect aggregation (i.e. filament growth). We account for these effects by lowering the energy level of an oxygen vacancy, E_t upon electron capture, by an amount equal to the increase in binding energy of neutral and negatively charged vacancies (≈0.2–0.9 eV) as predicted by ab-initio calculations (Bradley et al., 2015; Sementa et al., 2017). In general, this is a lattice relaxation process involving the emission/absorption of multiple phonons as described by several authors (Englman and Jortner, 1970; Henry and Lang, 1977; Nasyrov et al., 2004; Nasyrov and Gritsenko, 2011). Here, for simplicity, we assume the lattice relaxation coincides with electron capture (or emission). The assumed electron capture and emission processes are illustrated in **Figure 2**.

The significance of this model of filament growth is that it is more consistent with the nonlocal trap-assisted tunneling processes associated with electron conduction, which we later implement in TCAD Sentaurus to calculate the current-voltage characteristics as a more rigorous extension of these assumptions.

2.2.2 Oxygen Ion and Vacancy Diffusion

The rate of diffusion of oxygen ion and oxygen vacancy species is described as an Arrhenius **Eq. 6**. Here, the important parameters are the thermal barrier, ionic charge, and jump distance for each diffusing species.

$$R_{d,i} = f_{ph} \exp\left(-\frac{E_{a,i} - q_i \vec{\mathscr{E}} \cdot \Delta r_i}{k_B T}\right); \quad i = \{O^{2-}, V_O^{2+}\} \tag{6}$$

2.2.3 Frenkel Pair Generation

Oxygen vacancy formation is achieved through the production of Frenkel pairs, requiring the breaking of metal-oxygen bonds. A thermochemical description of dielectric breakdown exists (McPherson et al., 2003), in which the activation energy for breakdown is lowered by the local electric field projection along a polarizable bond axis. This expression is common in describing dielectric breakdown in thin insulators and is common in oxide-reliability studies **Eq. 7**.

$$R_g = f_{ph} \exp\left(\frac{-E_{a,g} - \vec{\mathscr{E}} \cdot \overrightarrow{p_0}\left(\frac{2+\epsilon_r}{3}\right)}{k_B T}\right) \tag{7}$$

Within this model, breakdown is expected to begin at an electric field that lowers the *effective* thermal barrier to zero.

$$|\vec{\mathscr{E}}_{bd}| = \frac{E_{a,g}}{|\overrightarrow{p_0}|}\left(\frac{2+\epsilon_r}{3}\right) \tag{8}$$

The value of $E_{a,g}$ determines the minimum voltage needed for forming, and is therefore an important consideration for the design of memristor circuits, as previously discussed. Using typical values for HfO_2 ($p_0 = 11 \times 10^{-10}$ Cm, $E_{a,g} = 4.5$ eV, $\epsilon_r = 21$), the breakdown field $|\mathcal{E}_{bd}| = 5.3$ MV cm^{-1}. This value corresponds to a nominal forming voltage roughly equal to half the HfO_2 thickness when measured in nanometers (i.e., $V_{form} \approx 2.5$ V for a 5 nm film). Empirically, $E_{a,g}$ is found to reduce with increased volume fraction of oxygen vacancies, as is commonly observed in highly defective HfO_x films, providing an empirical means for reducing forming voltage through the controlled introduction of defects. For example, choice of precursor (Hazra et al., 2019) and reaction time (Hazra et al., 2020) for the atomic-layer deposition of HfO_x films have a profound influence on forming voltage and pre-forming high-resistance levels. Furthermore, recent work have incorporated a model having a large (68%) reduction in the activation energy of forming oxygen vacancies at the Ti/HfO_2 interface as opposed to the bulk (Xu et al., 2020). This may be anticipated, since the net energy cost of breaking Hf-O bonds is lowered by the large driving force of oxidation (gettering) in the Ti (Stout and Gibbons, 1955). These factors imply that $E_{a,g}$ is a spatially-dependent parameter, essential to filament formation dynamics. According to ab-initio work, which also includes the effect of electron injection, we assume a 36% reduction in the activation barrier to Frenkel pair generation in the vicinity of existing nearest-neighbor Frenkel defect pairs (Bradley et al., 2015). Phenomenologically speaking, this accounts for the accelerating effect that point defects have on dielectric breakdown, as is well-known from oxide reliability studies. Additionally, this attempts to account for the formation of stable clusters of oxygen vacancies in regions where the binding energy is high and may be a potential source of retention failure in addition to existing theories based on oxygen diffusion (Raghavan et al., 2015; Kumar et al., 2017).

2.2.4 Frenkel Pair Recombination

As previously discussed, upon formation, charged Frenkel pairs in HfO_2 are unstable, requiring additional electrons from the conduction band to neutralize the oxygen vacancy and prevent rapid recombination. It is therefore expected that the thermal barrier to recombination is small relative to other processes, producing a rapid recombination rate, which we model using a simple field-independent Arrhenius **Eq. 9**. We use a value of 0.2 eV, according to previous work (Larcher et al., 2012), though this value becomes most relevant during reset operations, when the recombination rate of Frenkel pairs and electron emission (filament precipitation) becomes comparable for deep level vacancy states.

$$R_r = f_{ph} \exp\left(-\frac{E_a}{k_B T}\right) \tag{9}$$

2.3 Synopsys TCAD Sentaurus Modeling of Electric Current

2.3.1 Simulation Domain and Defect Modeling

Device geometry is defined in Synopsys TCAD Sentaurus with mesh refined using a maximum element size of 1 Å. Concentration profiles for each species (i.e., oxygen ions, unoccupied/occupied vacancies) were defined as point defects having Gaussian shape with decay length of 3 Å, corresponding to the minimum ion jump distance and grid spacing in our KMC model. Positions for each species were obtained from the output of the KMC simulation at each voltage step. Oxygen ions are modeled as negative fixed charges, with concentration as a parameter chosen to compensate the charge density of oxygen vacancies. Unoccupied oxygen vacancies are modeled as donors located 3 eV below the conduction band edge. Occupied oxygen vacancies are modeled as acceptors located 0.2–0.9 eV below the donor level, depending on the binding energy parameter (ΔE_t). As mentioned previously, the energy level difference between unoccupied/occupied oxygen vacancies reflects the increase in binding energy upon electron capture due to the large lattice coupling of ionized vacancies in HfO_x. To model effects due to disorder, the energy levels of oxygen vacancies were defined having Gaussian energy broadening (σ = 0.33 eV) consistent with similar approaches (Jiménez-Molinos et al., 2002).

2.3.2 Electric Current

Electric current was calculated using an electron barrier-tunneling model that couples each trap to the conduction band of the top and bottom electrodes through nonlocal, multiphonon-assisted inelastic and elastic transitions. Steady-state conditions were assumed. The rate of inelastic electron capture is described in terms of a maximum transition rate multiplied by the WKB tunneling probability $T_{i,j} = |\frac{\Psi(y_i)}{\Psi(y_j)}|^2$ and the phonon transition probability $M_{i,j} = \frac{(S-l)^2}{S} e^{-S(2f_B+1)+\frac{l\hbar\omega_0}{2kT}} I_l(z)$ for a transition between two states denoted i and j located at y_i and y_j.

$$c^n_{i,j} = \tau_0^{-1} T_{i,j} M_{i,j} \tag{10}$$

Sentaurus uses the asymptotic (large order) approximation to the conventional expression for $I_l(z)$, the modified Bessel function of order l contained within $M_{i,j}$, and is therefore appropriate when the number of phonons emitted during a transition is large (Schenk and Heiser, 1997). Under this approximation, the capture rate for an electron in the conduction band of an electrode at $y = 0$ to a trap located at y_t can be written as:

$$c^n = \tau_0^{-1} \left|\frac{\Psi(y_t)}{\Psi(0)}\right|^2 \frac{(S-l)^2}{S} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\chi}} \left(\frac{z}{l+\chi}\right)^l F_{1/2}\left(\frac{E_F - E_C(0)}{kT}\right) e^{\left(-S(2f_B+1)+\frac{l\hbar\omega_0}{2kT}+\chi\right)} \tag{11}$$

$$\tau_0^{-1} \equiv \frac{\sqrt{m^*_{n,be} m_0^3 k^3 T^3}}{\hbar^3 \sqrt{\chi}} g_c V_T S \omega_0 \tag{12}$$

$$\chi \equiv \sqrt{l^2 + z^2} \tag{13}$$

$$z \equiv 2S\sqrt{f_b(f_b+1)} \tag{14}$$

$$f_b = \frac{1}{\exp\left(\frac{\hbar\omega_0}{kT}\right) - 1} \tag{15}$$

$$l \equiv \frac{|E_C(0) - E_T|}{\hbar\omega_0} \tag{16}$$

The emission rate is then computed from the capture rate using the principle of detailed balance.

$$e^n = c^n \exp\left(-\frac{E_C(0) - E_T}{kT}\right) \tag{17}$$

Parameters for the model were defined as follows. The Huang-Rhys factor, S, was set to 17, the phonon energy to 0.07 eV, and the electron effective mass of HfO_x was set to the band mass 0.1 according to similar reports (Guan et al., 2012a). Nonlocal tunneling paths were considered, extending outwards from each electrode towards the opposite electrode along the oxide thickness. Electrodes were treated as Ohmic, with a variable work function with nominal values defined to mimic realistic device structures having a Titanium top-electrode (Ψ_{TE} = 4.33 eV, $m^*_{n,te}$ = 3.2) and titanium nitride bottom-electrode (Ψ_{BE} = 4.55 eV, $m^*_{n,be}$ = 2.0) (Lima et al., 2012). Trap volumes were estimated according to the effective mass of electrons in HfO_x and trap level relative to the conduction band edge (Palma et al., 1997; Jiménez-Molinos et al., 2002).

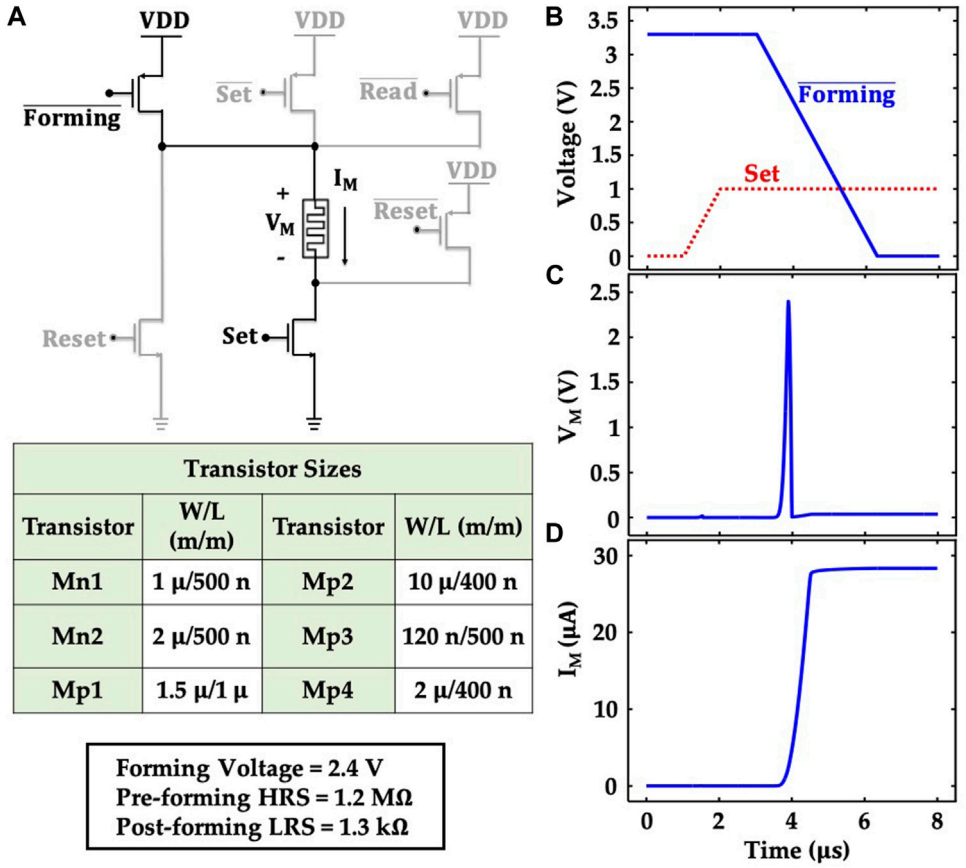
$$V_T = \left(\frac{\sqrt{4\pi/3}\,\hbar}{\sqrt{2m^*_{n,ox}|E_C(0) - E_T|}}\right)^3 \tag{18}$$

2.4 HSPICE Transient Simulations

To better utilize the insights provided by the device model, we couple the KMC + TCAD framework with a phenomenological compact model (Verilog-A) (Amer et al., 2017c) for the memristor to facilitate circuit-level simulations (in HSPICE). This compact model assumes a piecewise linear current-voltage (I-V) behavior in pre-forming and post-forming states of the memristor. The compact model considers the resistance behavior of the memristor (during the forming process) as follows:

$$R_M = \begin{cases} R_{Pre-forming} & V_M < V_{forming} \\ R_{Post-forming} & V_M > V_{forming} \end{cases} \tag{19}$$

Here, V_M and R_M are the voltage and resistance of the memristor (respectively). Clearly, forming voltage ($V_{forming}$), pre-forming ($R_{Pre-forming}$) and post-forming ($R_{Post-forming}$) resistance levels are the three necessary parameters for the forming operation of the memristor. This piecewise linear compact model considers ohmic



Transistor Sizes			
Transistor	W/L (m/m)	Transistor	W/L (m/m)
Mn1	1 μ/500 n	Mp2	10 μ/400 n
Mn2	2 μ/500 n	Mp3	120 n/500 n
Mp1	1.5 μ/1 μ	Mp4	2 μ/400 n

Forming Voltage = 2.4 V
Pre-forming HRS = 1.2 MΩ
Post-forming LRS = 1.3 kΩ

FIGURE 4 | (A) Schematic of the memristor-based synapse circuit that can control forming, set, reset and read operations in a memristor device. In this work, the set, reset and readout portions of the circuit are greyed out (unused). **(B)** Time dynamics of the input signals (*Forming* and Set) to control the forming process. Time dynamics of **(C)** memristor voltage (V_M), and **(D)** memristor current (I_M) for the applied input signals.

current-voltage relation before and after forming. We extract the parameters for this compact model using extensive analysis powered by the KMC + TCAD framework. We simulate the DC I-V characteristics for 25 device instances with an idealized ramp voltage across the memristor and formulate distributions of device resistance and forming voltage.

We simulate the forming operation for a memristor-based synapse circuit (shown in **Figure 4A**) in HSPICE to obtain the time dynamics of the voltage across the memristor. This synapse circuit can control all the operations of a memristor such as forming, set, reset and read. In this work, we only utilize the forming portion of the circuit. Therefore, the connections corresponding to other three operations are greyed out. Here, M_{p1} transistor controls the forming of the memristor and M_{n1} transistor is used to set the compliance current limit during the forming process.

To simulate the forming process in HSPICE, we calibrate the compact model with the relevant parameters (forming voltage, pre-forming HRS, and post-forming LRS), extracted from the KMC + TCAD simulations. Note, the ideal approach to calibrate the compact model would be to use iterations between the circuit level simulations and KMC + TCAD simulations. The approach would include running the KMC + TCAD simulations with the results from the circuit level simulation and then again simulating the circuit with the KMC + TCAD results. But, this would require a compact model that could capture the non-linear behavior observed in the KMC + TCAD simulations. Here, to calibrate the model for circuit simulations, we choose the mean values of the pre-forming HRS (1.2 $M\Omega$), and post-forming LRS (1.3 $k\Omega$) to set up the compact model. As for the forming voltage, we choose the maximum value of the corresponding distribution (2.4 V), to capture the worst-case scenario. For the transistors, we use the DGXFET NMOS/PMOS models for the IBM 65 nm 10LPe process.

Figures 4B–D illustrate the simulated transient characteristics of the synapse circuit (only the forming process), bearing the signature of the circuit-level interactions of the memristor. We govern the forming process with appropriately designed control signals (*Forming* and Set). We first turn on the M_{n1} transistor by applying an appropriate gate voltage (Set). Then, we turn on the M_{p1} transistor to control the forming of the memristor. The voltage across the memristor (V_M) gradually increases when the forming operation begins (**Figure 4C**). The memristor takes the prime share of the supply voltage (V_{DD}) when the two series transistors (M_{p1} and M_{n1}) turn ON. Subsequently, after successful

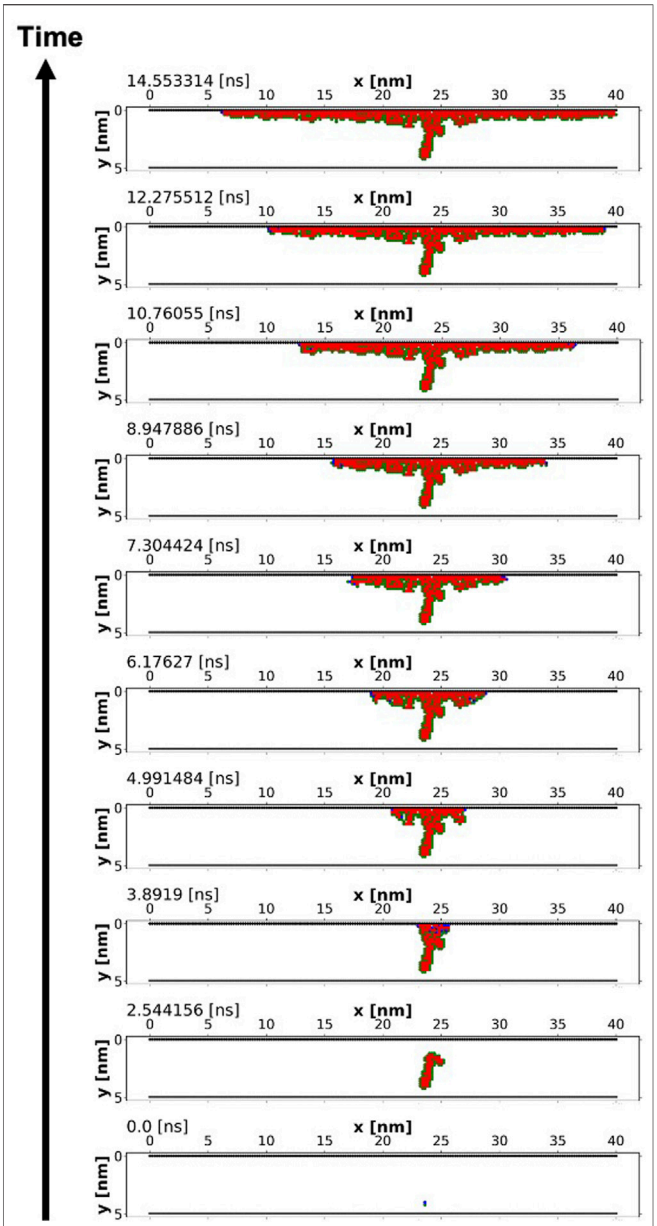


FIGURE 5 | An example of filament formation dynamics with the top electrode at 3 V and the bottom electrode at 0 V. The red squares are the filament, the green circles are oxygen ions and the blue circles are unoccupied oxygen vacancies.

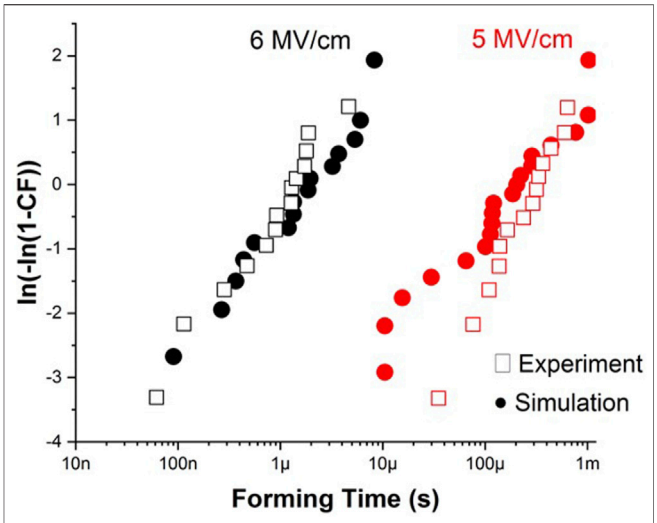


FIGURE 6 | Comparison of forming time distributions obtained from our simulations and experimental work (Lorenzi et al., 2013) for a TiN/HfO_2/Pt device at an electric field of 5 MV cm^{-1} and 6 MV cm^{-1}.

forming, the resistance of the memristor drastically reduces and so does the voltage across it. This circuit topology serves as the baseline for the Monte-Carlo simulations (discussed later).

3 RESULTS

3.1 Filament Formation Dynamics

Figure 5 shows the filament formation dynamics for a single device simulation at a constant voltage of 3 V applied to the top electrode and 0 V applied to the bottom electrode. For this simulation, the initial defect volume fraction was chosen to correspond with 1 oxygen vacancy within the simulation domain. New defects tend to form in the vicinity of pre-existing defects due to: 1) a lower formation energy in the presence of pre-existing Frenkel pairs; and 2) a higher electric field in the vicinity of a charged filament. As time progresses, filament growth proceeds towards the top electrode, where the electric field is highest. Once the electrodes have bridged, additional filament growth occurs along the width of the top electrode. This occurs due to: 1) a higher lateral electric field, 2) the effect of oxygen gettering by the Ti top electrode, which readily removes oxygen ion interstitials, and 3) screening of the electric field seen by points closer to the bottom electrode by charged vacancies near the top electrode.

Since bond breaking is modeled as a statistical process, it is necessary to evaluate the behavior of more than one device. Thus, we further investigate forming dynamics by assessing the statistical distribution of forming times for repeated device simulations. Here, we focus on the forming time–the time required to form, which we estimate by halting the simulation once a predefined number of filament states are formed. We use a volume fraction of 0.1, which should correspond to a concentration of the order of $\approx 1 \times 10^{21}$ cm^{-3}. This is comparable to what has been observed experimentally in HfOx films having high oxygen vacancy concentrations (Hildebrandt et al., 2011), reported as high as 6×10^{21} cm^{-3}.

We note that since the forming process is modeled using the thermochemical model of dielectric breakdown, the *forming time* is synonymous with the time-dependent-dielectric-breakdown (TDDB), and is therefore an experimental observable which is straightforward to measure and, most importantly, can be used to validate kinetic monte carlo simulation models. Previous authors have compared the effect of top and bottom electrodes on the

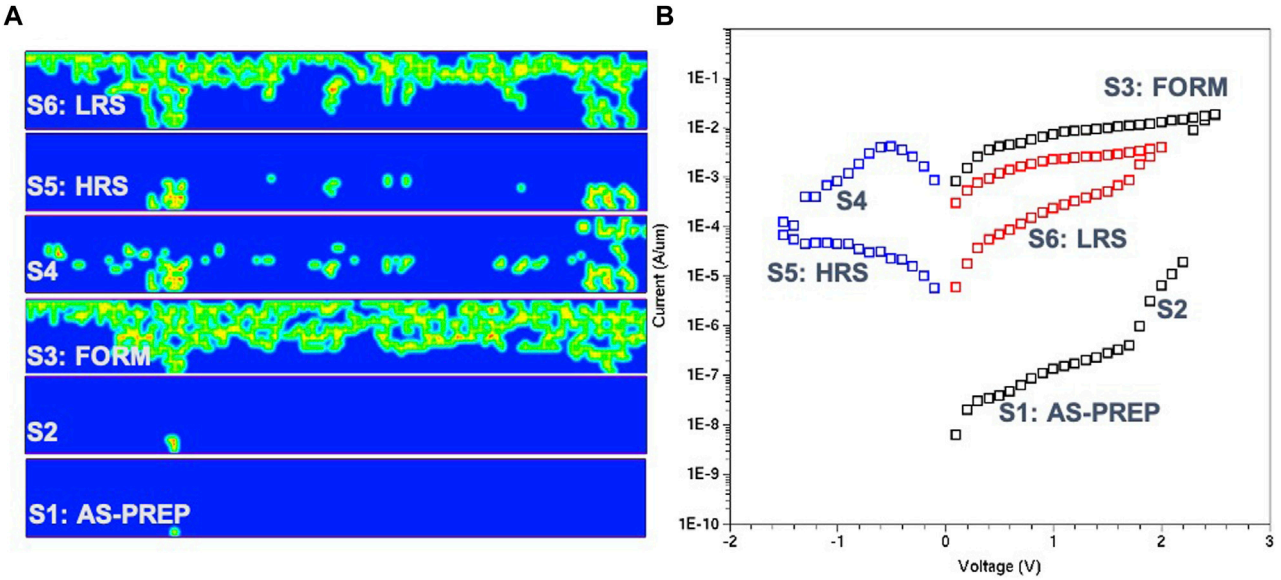


FIGURE 7 | (A) Filament positions corresponding to different points along the programming sequence (forming, reset, set). **(B)** Pulsed IV characteristics for the complete programming sequence.

forming time for nearly stoichiometric 10 nm thick HfO_2 thin films deposited by atomic layer deposition (Lorenzi et al., 2013; Cagli et al., 2011). **Figure 6** shows a Weibull distribution of forming times obtained from our simulation and those of Lorenzi et al., 2013 for a TiN/HfO_2/Pt device at an electric field of 5 MV cm^{-1} and 6 MV cm^{-1}. Simulation results at both values of electric field show reasonable quantitative agreement to experiment. At the lower electric field of 5 MV cm^{-1}, there appears to be a deviation from a Weibull distribution, however the forming times are similar in magnitude. These results illustrate the similarity between forming time and TDDB, and provide a potential route towards empirical validation of simulation models—from which the activation energy of Frenkel pair generation, $E_{a,g}$ and the field-acceleration factor, γ can be derived (McPherson and Mogul, 1998).

$$\ln(TDDB) \propto \frac{E_{a,g}}{k_B T} - \gamma \mathscr{E} \tag{20}$$

This is especially important to establish agreement to experimental results, since HfOx films can exhibit mixed crystalline phases and variable oxygen content depending on deposition conditions—both of which are expected to modify $E_{a,g}$ and γ.

3.2 Current-Voltage Characteristics

Figure 7 shows the complete current-voltage (IV) characteristics for a forming, reset and set programming cycle. In order to obtain IV characteristics, snapshots of the state of the system are taken at the end of each voltage step shown in **Figure 3B**. In this case, the time step is 1 μs and the voltage step is 0.1 V. In **Figure 7A**, intermediate filament states are shown at different stages of programming. It can be seen that, at forming, a large volume fraction of the device consists of oxygen vacancy filament with a structure that extends laterally near the top electrode. The spatial extent of the filament (i.e., volume fraction of vacancies) will ultimately be controlled by the flux linkage (forming voltage × time), which is set by the compliance current in practice.

Several of the well-known aspects of oxide memristors are captured by this simulated result shown in **Figure 7B**: 1) the high-resistance initial state of the as-prepared thin-film; 2) A low resistance state after forming; 3) A gradual *reset* behavior with programmable analog high-resistance levels; and 4) A low-resistance state following *set* of the order of kilo-Ohms. The ability to quantify and visualize the increase in current upon *set* (potentiation) and the decrease in current upon *reset* (depression) is a key benefit of **Figure 7B** incorporating TCAD Sentaurus for modeling synaptic behavior.

3.3 Filament Electrostatics

Figure 8 shows key aspects of the filament electrostatics. In **Figure 8A**, the electrostatic potential and x and y components of the electric field throughout the device are shown. In regions where the filament bridges the top and bottom electrode, the voltage drop across this region is large enough such that the filament region has a net negative potential near the bottom electrode. This agrees well with experimental results which relied on *in-situ* TEM electron holography measurements (Li et al., 2017), in which they described the filament as a "negative potential synapse." Our results show that this stems directly from the negative space-charge associated with the filament, assumed to be due to electron capture ($V_O^{2+} \rightarrow V_O^{2-}$), and supported by both experimental and theoretical insights. We show this explicitly, by comparing two different line plots—outside the filament and within the filament—in **Figures 8B,C**. A dashed line at a potential of zero is added as

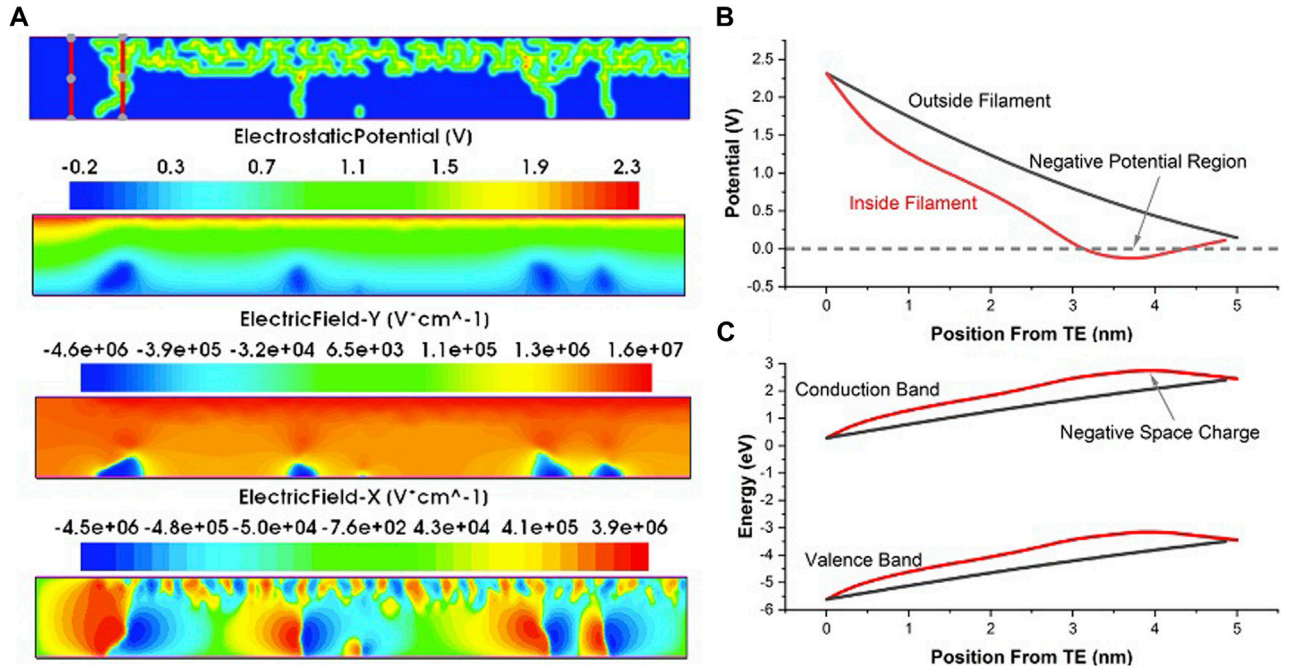


FIGURE 8 | (A) Filament electrostatics after forming, including the electrostatic potential, and electric field components. **(B)** Comparison of electrostatic potential inside and outside of a filament region as indicated in the red lines in **(A)**. The filament region shows a region of negative potential and can be likened to a "negative potential synapse." **(C)** Comparison of the energy band diagram inside and outside of the filament. Within the filament, a negative curvature indicates that the filament is a negative space-charge region.

a visual aid, clearly indicating a negative potential within the filament. This is also reflected in the energy band diagram in **Figure 8C**, which shows a negative curvature as expected for a negative space charge.

3.4 Monte Carlo Analysis of Synapse Forming Circuit

Finally, we use the unique capability of the KMC + TCAD model to investigate a device-circuit co-design strategy. We test memristor characteristics for different levels of initial oxygen vacancy volume faction (x), a design variable that can be easily controlled during the fabrication process. **Figure 9A** shows the pre-forming HRS for different values of x for 25 devices (each) obtained from the KMC + TCAD simulations. Considering the circuit-level scenario of a synapse, the variations in the characteristics of multiple transistors need to be superposed with the inherent device-level variations of the memristor. To account for all these variations concurrently, we run 1000-point (3σ) Monte-Carlo simulations for the forming circuit using the data obtained from the KMC + TCAD framework. We utilize the dependence of the pre-forming HRS on the initial oxygen vacancy volume faction and the threshold voltage variation of the PMOS and NMOS transistors to set the input distributions for the Monte-Carlo simulations. **Figure 9A** shows the pre-forming HRS for different values of x for 25 devices (each) obtained from the KMC + TCAD simulations. Without any loss of generality, we run the Monte-Carlo simulation for two values of x (0.02 and 0.04%). To incorporate the threshold voltage variation of the transistors, we use a gaussian distribution with a mean value equal to the nominal threshold voltage (0.65 V for 65 nm DGXFET transistors) and standard deviation of 20 mV (shown in the table of **Figure 9B**). We also run the Monte-Carlo simulations with different levels of current compliance, controlled by applying appropriate gate voltage (Set) to M_{n1}.

Figures 9C,D show the scatter plots for the average current through the memristor and average power of the forming circuit, respectively. Each of these metrics have been reported for different levels of compliance currents (different levels of Set). To ensure a fair comparison, we allow a constant time for the forming process for all cases. Naturally, we observe that for lower compliance limits, many instances of the Monte-Carlo simulations exhibit "unsuccessful" forming. Note, a compliance limit may lead to a different level of post-forming LRS and hence might be treated as successful, if the post-forming LRS is known during the design stage. However, if such changes in the post-forming LRS occurs dynamically and randomly, those will lead to read/sense failure. Therefore, we simplify our analysis by tagging such cases as 'Forming Failure'. Higher compliance limit allows most memristor instances to successfully form and hence leads to a larger average current level (**Figure 9C**). If the memristor can successfully form, it goes to post-forming LRS ($1.3k\omega$ in our simulation). Otherwise, it remains in the pre-forming HRS which is much larger compared to LRS. Therefore, the increase in the value of Set increases the average current of the memristor due to the increase in the

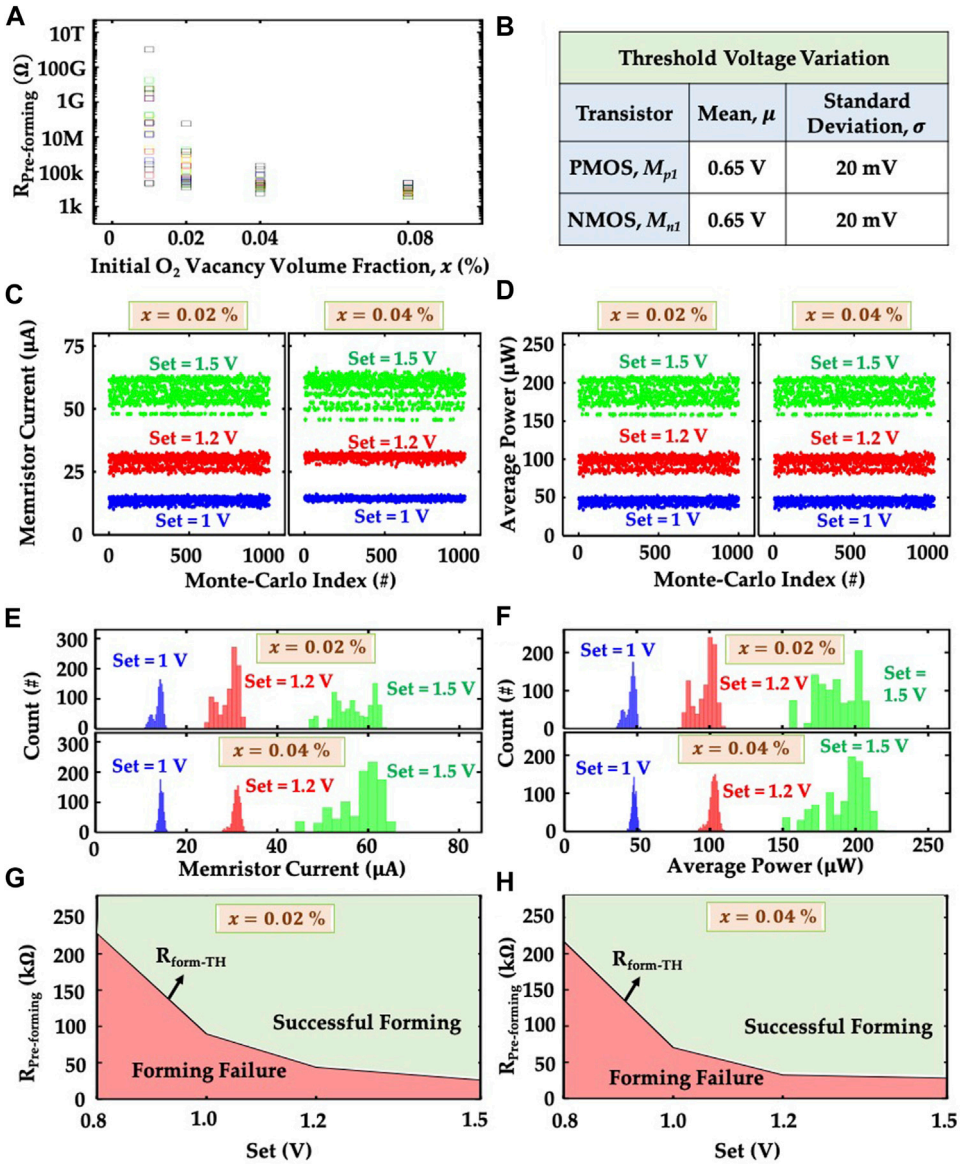


FIGURE 9 | (A) Dependence of pre-forming HRS on the initial oxygen vacancy volume fraction (x). These results are obtained from KMC-TCAD simulation for 25 devices. **(B)** Table shows the values of mean and standard deviation of the threshold voltage distribution of M_{p1} and M_{n1} transistors used for the Monte-Carlo simulation. Scatter plot of the **(C)** memristor current, and **(D)** average power consumption of the forming circuit obtained from the 1,000 point Monte-Carlo simulation for three different values of Set (1 *V*, 1.2 and 1.5 *V*) and the pre-forming HRS values for two different values of x (0.02 and 0.04%). Histogram plot of the **(E)** memristor current, and **(F)** average power consumption of the forming circuit for the data shown in the scatter plots of **(C)** and **(D)** respectively. Dependence of forming success and failure on the values of Set for the pre-forming HRS values obtained for (g) x = 0.02%, and (h) x = 0.04%.

number of formed memristors. Since, the average power of the forming circuit is very closely related to the memristor current, the average forming power shows the same trend like the average memristor current (**Figure 9D**). The initial oxygen vacancy volume factions lead to similar results, with different levels of mean value and standard deviation for the memristor current and forming power (**Figures 9C,D**). **Figure 9A** shows that the pre-forming HRS for x = 0.02% has a larger standard deviation compared to that for x = 0.04%. Therefore, the memristor current and forming power obtained from the Monte-Carlo simulation show larger standard deviation for x = 0.02% compared to the case of x = 0.04%. But only for Set = 1.5 V, the smallest value of pre-forming HRS of the memristors that cannot form becomes comparable to the value of the post-forming LRS. Therefore, the effect of x on the standard deviation of the memristor current and forming power gets suppressed. **Figures 9E,F** show the histogram plot for Monte-Carlo results of memristor current and forming power shown in the scatter plots (**Figures 9C,D**).

Based on these Monte-Carlo simulations, we correlate the compliance limits (controlled by the Set pulse) with the pre-

forming HRS ($R_{Pre-forming}$) level of the memristor. **Figure 9G** illustrates the combinations of Set and $R_{Pre-forming}$ that lead to successful forming (and vice versa). Clearly, for a given compliance limit, the pre-forming HRS level of a memristor needs to be higher than a critical threshold ($R_{form-TH}$), illustrated (in **Figure 9G**) as a line separating the successful and unsuccessful forming cases. **Figure 9H** shows similar trends for a different initial oxygen vacancy volume faction. Our analysis shows a pathway to optimize the synapse circuit by correlating the material and circuit-level design knobs.

4 CONCLUSION

The ability to design and implement fast, scalable and robust neuromorphic systems relies heavily upon our fundamental understanding of memristor switching. Oxide memristors, envisioned for RRAM-based neuromorphic systems, exhibit changes in resistance state through multiple synergistic effects involving electronic and atomic degrees of freedom, often modelled as separate influences. One of the main purposes of this work was to establish a more direct connection between the two in order to: 1) provide a unified view of filament evolution and electronic conduction; 2) to implement this description within a state-of-the-art TCAD framework for modeling electric conduction; and 3) gain circuit-level insight.

Here, we have argued the use of a simple model of filament evolution that makes explicit use of Fermi-Dirac statistics, coupling the rate of defect generation and recombination to electronic transitions associated with conduction and lattice relaxation. By combining Synopsys TCAD Sentaurus with Kinetic Monte Carlo simulations of filament evolution, we have shown the ability to quantify both the common and subtle aspects of resistive-switching behavior of HfO_x memristors. Quasi-static *snapshots* of the device state—consisting of positive/negative oxygen vacancies, and oxygen ions—were taken at various voltages to obtain IV characteristics under stepped voltage ramp conditions. Electric conduction in oxygen vacancy filaments is modeled as trap-to-band transitions between occupied and unoccupied electronic states assisted by multiphonon absorption and emission. According to Fermi-Dirac statistics, such processes are expected to occur within a band of energies in the vicinity of the Fermi level wherein both occupied and unoccupied states are probable. Thus, the occupancy of a trap and its relation to the Fermi level is fundamentally related to transition rates associated with electronic conduction. The use of TCAD Sentaurus provides a powerful framework for modeling these and other conduction processes as well as visualizing filament electrostatics, as we've shown. In particular, we have obtained results that are consistent with experimental observations of a negative space charge and potential associated with a vacancy-rich filament. Our approach will enable the more efficient evaluation of memristor device behavior and circuit performance, stemming from physics-based modeling, having a direct impact and benefit on the fields of neuromorphic computing, memory design and dynamical systems.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

AZ conceived of and designed the Kinetic Monte Carlo algorithm and TCAD Sentaurus modeling framework. SA performed the transient simulations, and the circuit-level Monte Carlo analysis. ZW assisted with device-level simulations and parameter influences. RW designed the synapse circuit. AA analyzed the circuit simulations and Monte Carlo data. AZ AA, and GR jointly supervised the work and analyzed the results. All authors contributed to writing the article and approving the submitted version.

FUNDING

This material is based in part on research sponsored by Air Force Research Laboratory under agreement number FA8750-19-1-0025. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Air Force Research Laboratory or the U.S. Government.

ACKNOWLEDGMENTS

The authors would like to acknowledge helpful conversations with Dr. Nathaniel Cady of SUNY Polytechnic Institute at Albany, NY.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnano.2021.734121/full#supplementary-material

REFERENCES

Aldana, S., Garcia-Fernandez, P., Romero-Zaliz, R., Gonzalez, M. B., Jimenez-Molinos, F., Campabadal, F., et al. (2018). "A Kinetic Monte Carlo Simulator to Characterize Resistive Switching and Charge Conduction in Ni/HfO2/Si RRAMs," in *Proceedings Of The 2018 12th Spanish Conference On Electron Devices (Cde)*. Editors J. Mateos and T. Gonzalez (NEW YORK, NY 10017 USA: IEEESpanish Conference on Electron DevicesBackup Publisher: IEEE; Univ Salamanca ISSN), 2163–4971. Type: Proceedings Paper.

Aldana, S., Garcia-Fernandez, P., Romero-Zaliz, R., Gonzalez, M. B., Jimenez-Molinos, F., Gomez-Campos, F., et al. (2020). Resistive Switching in HfO2 Based Valence Change Memories, a Comprehensive 3D Kinetic Monte Carlo Approach. *JOURNAL PHYSICS D-APPLIED PHYSICS* 53. doi:10.1088/1361-6463/ab7bb6

Amer, S., Hasan, M. S., and Rose, G. S. (2017a). Analysis and Modeling of Electroforming in Transition Metal Oxide-Based Memristors and its Impact on Crossbar Array Density. *IEEE Electron. Device Lett.* 39, 19–22.

Amer, S., Rose, G. S., Beckmann, K., and Cady, N. C. (2017b). Design Techniques for In-Field Memristor Forming Circuits. In 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS). IEEE, 1224–1227. doi:10.1109/mwscas.2017.8053150

Amer, S., Sayyaparaju, S., Rose, G. S., Beckmann, K., and Cady, N. C. (2017c). A Practical Hafnium-Oxide Memristor Model Suitable for Circuit Design and Simulation. In 2017 IEEE International Symposium on Circuits and Systems (ISCAS), 1–IEEE4. doi:10.1109/iscas.2017.8050790

Beckmann, K., Holt, J., Manem, H., Van Nostrand, J., and Cady, N. C. (2016). Nanoscale Hafnium Oxide Rram Devices Exhibit Pulse Dependent Behavior and Multi-Level Resistance Capability. *Mrs Adv.* 1, 3355–3360. doi:10.1557/adv.2016.377

Bianchi, S., Pedretti, G., Munoz-Martin, I., Calderoni, A., Ramaswamy, N., Ambrogio, S., et al. (2020). A Compact Model for Stochastic Spike-timing-dependent Plasticity (STDP) Based on Resistive Switching Memory (RRAM) Synapses. *IEEE Trans. Electron. Devices* 67, 2800–2806. doi:10.1109/TED.2020.2992386

Bradley, S. R., Shluger, A. L., and Bersuker, G. (2015). Electron-Injection-Assisted Generation of Oxygen Vacancies in MonoclinicHfO2. *Phys. Rev. Appl.* 4, 064008. doi:10.1103/PhysRevApplied.4.064008

Cagli, C., Buckley, J., Jousseaume, V., Cabout, T., Salaun, A., Grampeix, H., et al. (2011). Experimental and Theoretical Study of Electrode Effects in HfO2 Based RRAM. In 2011 International Electron Devices Meeting. 28. 7.1–28.7.4. doi:10.1109/IEDM.2011.6131634.ISSN:2156-017X

Chen, A. (2013). Area and Thickness Scaling of Forming Voltage of Resistive Switching Memories. *IEEE Electron. Device Letters* 35, 57–59.

Choi, B. J., Torrezan, A. C., Strachan, J. P., Kotula, P. G., Lohn, A. J., Marinella, M. J., et al. (2016). High-Speed and Low-Energy Nitride Memristors. *Adv. Funct. Mater.* 26, 5290–5296. doi:10.1002/adfm.201600680

Chua, L. (1971). Memristor-the Missing Circuit Element. *IEEE Trans. Circuit Theor.* 18, 507–519. doi:10.1109/tct.1971.1083337

Englman, R., and Jortner, J. (1970). The Energy gap Law for Radiationless Transitions in Large Molecules. *Mol. Phys.* 18, 145–164. doi:10.1080/00268970000100171

Fichthorn, K. A., and Weinberg, W. H. (1991). Theoretical Foundations of Dynamical Monte Carlo Simulations. *J. Chem. Phys.* 95, 1090–1096. doi:10.1063/1.461138

Govoreanu, B., Kar, G., Chen, Y., Paraschiv, V., Kubicek, S., Fantini, A., et al. (2011). 10× 10nm 2 Hf/hfo X Crossbar Resistive Ram with Excellent Performance, Reliability and Low-Energy Operation. In 2011 International Electron Devices Meeting. IEEE, 31–36.

Guan, X., Yu, S., and Wong, H.-S. P. (2012b). A Spice Compact Model Of Metal Oxide Resistive Switching Memory With Variations. *IEEE Electron. Device Lett.* 33, 1405–1407. doi:10.1109/LED.2012.2210856

Guan, X., Yu, S., and Wong, H.-S. P. (2012a). On the Switching Parameter Variation of Metal-Oxide RRAM-Part I: Physical Modeling and Simulation Methodology. *IEEE Trans. Electron. Devices* 59, 1172–1182. doi:10.1109/TED.2012.2184545

Hazra, J., Liehr, M., Beckmann, K., Rafiq, S., and Cady, N. (2020). Impact of Atomic Layer Deposition Co-reactant Pulse Time on 65nm CMOS Integrated Hafnium Dioxide-Based Nanoscale RRAM Devices. In 2020 IEEE International Integrated Reliability Workshop (IIRW). South Lake Tahoe, CA, USA: IEEE, 1–4. doi:10.1109/IIRW49815.2020.9312877

Hazra, J., Liehr, M., Beckmann, K., Rafiq, S., and Cady, N. (2019). Improving the Memory Window/Resistance Variability Trade-Off for 65nm CMOS Integrated HfO2 Based Nanoscale RRAM Devices. In 2019 IEEE International Integrated Reliability Workshop (IIRW). South Lake Tahoe, CA, USA: IEEE, 1–4. doi:10.1109/IIRW47491.2019.8989872

Henry, C. H., and Lang, D. V. (1977). Nonradiative Capture and Recombination by Multiphonon Emission in GaAs and GaP. *Phys. Rev. B* 15, 989–1016. doi:10.1103/PhysRevB.15.989

Hildebrandt, E., Kurian, J., Müller, M. M., Schroeder, T., Kleebe, H.-J., and Alff, L. (2011). Controlled Oxygen Vacancy Induced P-type Conductivity in HfO2−x Thin Films. *Appl. Phys. Lett.* 99, 112902. doi:10.1063/1.3637603

Ho, M.-Y., Gong, H., Wilk, G. D., Busch, B. W., Green, M. L., Voyles, P. M., et al. (2003). Morphology and Crystallization Kinetics in HfO2 Thin Films Grown by Atomic Layer Deposition. *J. Appl. Phys.*, 93, 1477–1481. doi:10.1063/1.1534381

Hu, M., Strachan, J. P., Li, Z., and Stanley, R. (2016). Dot-product Engine as Computing Memory to Accelerate Machine Learning Algorithms. In 2016 17th International Symposium on Quality Electronic Design (ISQED). IEEE, 374–379. doi:10.1109/isqed.2016.7479230

Huang, P., Deng, Y., Gao, B., Chen, B., Zhang, F., Yu, D., et al. (2013a). Optimization of Conductive Filament of Oxide-Based Resistive-Switching Random Access Memory for Low Operation Current by Stochastic Simulation. *Jpn. J. Appl. Phys.* 52, 04CD04. doi:10.7567/JJAP.52.04CD04

Huang, P., Liu, X. Y., Chen, B., Li, H. T., Wang, Y. J., Deng, Y. X., et al. (2013b). A Physics-Based Compact Model of Metal-Oxide-Based RRAM DC and AC Operations. *IEEE Trans. Electron. Devices* 60, 4090–4097. doi:10.1109/ted.2013.2287755

Jansen, A. P. J. (2012). *An Introduction to Kinetic Monte Carlo Simulations of Surface Reactions*, 856. Springer.

Jiang, Z., Yu, S., Wu, Y., Engel, J. H., Guan, X., and Wong, H.-S. P. (2014). Verilog-A Compact Model for Oxide-Based Resistive Random Access Memory (RRAM). In 2014 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD). Yokohama, Japan: IEEE, 41–44. doi:10.1109/SISPAD.2014.6931558

Jiménez-Molinos, F., Gámiz, F., Palma, A., Cartujo, P., and López-Villanueva, J. A. (2002). Direct and Trap-Assisted Elastic Tunneling through Ultrathin Gate Oxides. *J. Appl. Phys.* 91, 5116–5124. doi:10.1063/1.1461062

Kim, W., Wouters, D. J., Menzel, S., Rodenbücher, C., Waser, R., and Rana, V. (2016). Lowering Forming Voltage and Forming-free Behavior of Ta 2 O 5 Reram Devices. In 2016 46th European Solid-State Device Research Conference (ESSDERC). IEEE, 164–167.

Koveshnikov, S., Matthews, K., Min, K., Gilmer, D., Sung, M., Deora, S., et al. (2012). Real-time Study of Switching Kinetics in Integrated 1t/hfo X 1r Rram: Intrinsic Tunability of Set/reset Voltage and Trade-Off with Switching Time. In 2012 International Electron Devices Meeting. IEEE, 20–24.

Kumar, S., Wang, Z., Huang, X., Kumari, N., Davila, N., Strachan, J. P., et al. (2017). Oxygen Migration during Resistance Switching and Failure of Hafnium Oxide Memristors. *Appl. Phys. Lett.* 110, 103503. doi:10.1063/1.4974535

Larcher, L., Padovani, A., Pirrotta, O., Vandelli, L., and Bersuker, G. (2012). Microscopic Understanding and Modeling of HfO2 RRAM Device Physics. In 2012 International Electron Devices Meeting. 20. 1.1–20.1.4. doi:10.1109/IEDM.2012.6479077.ISSN:2156-017X

Li, C., Belkin, D., Li, Y., Yan, P., Hu, M., Ge, N., et al. (2018a). Efficient and Self-Adaptive *In-Situ* Learning in Multilayer Memristor Neural Networks. *Nat. Commun.* 9, 2385–2388. doi:10.1038/s41467-018-04484-2

Li, C., Gao, B., Yao, Y., Guan, X., Shen, X., Wang, Y., et al. (2017). Direct Observations of Nanofilament Evolution in Switching Processes in HfO2-Based Resistive Random Access Memory by *In Situ* TEM Studies. *Adv. Mater.* 29, 1602976. doi:10.1002/adma.201602976

Li, C., Hu, M., Li, Y., Jiang, H., Ge, N., Montgomery, E., et al. (2018b). Analogue Signal and Image Processing with Large Memristor Crossbars. *Nat. Electron.* 1, 52–59. doi:10.1038/s41928-017-0002-z

Lima, L. P. B., Diniz, J. A., Doi, I., and Godoy Fo, J. (2012). Titanium Nitride as Electrode for MOS Technology and Schottky Diode: Alternative Extraction Method of Titanium Nitride Work Function. *Microelectronic Eng.* 92, 86–90. doi:10.1016/j.mee.2011.04.059

Liu, H., Lv, H., Yang, B., Xu, X., Liu, R., Liu, Q., et al. (2014). Uniformity Improvement in 1t1r Rram with Gate Voltage Ramp Programming. *IEEE Electron. Device Lett.* 35, 1224–1226. doi:10.1109/led.2014.2364171

Lorenzi, P., Rao, R., and Irrera, F. (2013). Forming Kinetics in $\hbox{HfO}_{2}$ -Based RRAM Cells. *IEEE Trans. Electron. Devices* 60, 438–443. doi:10.1109/TED.2012.2227324

Loy, D. J. J., Dananjaya, P. A., Chakrabarti, S., Tan, K. H., Chow, S. C. W., Toh, E. H., et al. (2020). Oxygen Vacancy Density Dependence with a Hopping Conduction Mechanism in Multilevel Switching Behavior of HfO2-Based Resistive Random Access Memory Devices. *ACS Appl. Electron. Mater.* 2, 3160–3170. Place: 1155 16TH ST, NW, WASHINGTON, DC 20036 USA Publisher: AMER CHEMICAL SOC Type: Article. doi:10.1021/acsaelm.0c00515

Manem, H., Rajendran, J., and Rose, G. S. (2012). Design Considerations for Multilevel CMOS/Nano Memristive Memory. *J. Emerg. Technol. Comput. Syst.* 8, 1–22. doi:10.1145/2093145.2093151

McPherson, J., Kim, J.-Y., Shanware, A., and Mogul, H. (2003). Thermochemical Description of Dielectric Breakdown in High Dielectric Constant Materials. *Appl. Phys. Lett.* 82, 2121–2123. doi:10.1063/1.1565180

McPherson, J. W., and Mogul, H. C. (1998). Underlying Physics of the Thermochemical E Model in Describing Low-Field Time-dependent Dielectric Breakdown in SiO2 Thin Films. *J. Appl. Phys.* 84, 1513–1523. doi:10.1063/1.368217

Nasyrov, K. A., and Gritsenko, V. A. (2011). Charge Transport in Dielectrics via Tunneling between Traps. *J. Appl. Phys.* 109, 093705. doi:10.1063/1.3587452

Nasyrov, K. A., Gritsenko, V. A., Novikov, Y. N., Lee, E.-H., Yoon, S. Y., and Kim, C. W. (2004). Two-bands Charge Transport in Silicon Nitride Due to Phonon-Assisted Trap Ionization. *J. Appl. Phys.* 96, 4293–4296. doi:10.1063/1.1790059

Palma, A., Godoy, A., Jiménez-Tejada, J. A., Carceller, J. E., and López-Villanueva, J. A. (1997). Quantum Two-Dimensional Calculation of Time Constants of Random Telegraph Signals in Metal-Oxide-Semiconductor Structures. *Phys. Rev. B* 56, 9565–9574. doi:10.1103/PhysRevB.56.9565

Pi, S., Li, C., Jiang, H., Xia, W., Xin, H., Yang, J. J., et al. (2019). Memristor Crossbar Arrays with 6-nm Half-Pitch and 2-nm Critical Dimension. *Nat. Nanotech* 14, 35–39. doi:10.1038/s41565-018-0302-0

Raghavan, N., Frey, D. D., Bosman, M., and Pey, K. L. (2015). Statistics of Retention Failure in the Low Resistance State for Hafnium Oxide RRAM Using a Kinetic Monte Carlo Approach. *Microelectronics Reliability* 55, 1422–1426. doi:10.1016/j.microrel.2015.06.090

Schenk, A., and Heiser, G. (1997). Modeling and Simulation of Tunneling through Ultra-thin Gate Dielectrics. *J. Appl. Phys.* 81, 7900–7908. doi:10.1063/1.365364

Sementa, L., Larcher, L., Barcaro, G., and Montorsi, M. (2017). Ab Initio modelling of Oxygen Vacancy Arrangement in Highly Defective HfO2 Resistive Layers. *Phys. Chem. Chem. Phys.* 19, 11318–11325. doi:10.1039/C7CP01216K

Sire, C., Blonkowski, S., Gordon, M. J., and Baron, T. (2007). Statistics of Electrical Breakdown Field in HfO2 and SiO2 Films from Millimeter to Nanometer Length Scales. *Appl. Phys. Lett.* 91, 242905. doi:10.1063/1.2822420

Stout, V. L., and Gibbons, M. D. (1955). Gettering of Gas by Titanium. *J. Appl. Phys.* 26, 1488–1492. doi:10.1063/1.1721936

Strukov, D. B., Snider, G. S., Stewart, D. R., and Williams, R. S. (2008). The Missing Memristor Found. *nature* 453, 80–83. doi:10.1038/nature06932

Synopsys (2019). TCAD Sentaurus.

Xia, Q., and Yang, J. J. (2019). Memristive Crossbar Arrays for Brain-Inspired Computing. *Nat. Mater.* 18, 309–323. doi:10.1038/s41563-019-0291-x

Xu, X., Rajendran, B., and Anantram, M. P. (2020). Kinetic Monte Carlo Simulation of Interface-Controlled Hafnia-Based Resistive Memory. *IEEE Trans. Electron. Devices* 67, 118–124. doi:10.1109/TED.2019.2953917

Yao, P., Wu, H., Gao, B., Eryilmaz, S. B., Huang, X., Zhang, W., et al. (2017). Face Classification Using Electronic Synapses. *Nat. Commun.* 8, 1–8. doi:10.1038/ncomms15199

Yao, P., Wu, H., Gao, B., Zhang, G., and Qian, H. (2015). The Effect of Variation on Neuromorphic Network Based on 1t1r Memristor Array. In 2015 15th Non-Volatile Memory Technology Symposium (NVMTS). IEEE, 1–3. doi:10.1109/nvmts.2015.7457492

Yu, S., and Wong, H.-S. P. (2010). A Phenomenological Model for the Reset Mechanism of Metal Oxide RRAM. *IEEE Electron. Device Lett.* 31, 1455–1457. doi:10.1109/LED.2010.2078794

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.

frontiers in Nanotechnology

ORIGINAL RESEARCH
published: 08 December 2021
doi: 10.3389/fnano.2021.765947

TReMo+: Modeling Ternary and Binary ReRAM-Based Memories With Flexible Write-Verification Mechanisms

Shima Hosseinzadeh, Mehrdad Biglari and Dietmar Fey*

Department Computer Science, Chair of Computer Architecture, Friedrich-Alexander-Universitat Erlangen-Nürnberg (FAU), Erlangen, Germany

Non-volatile memory (NVM) technologies offer a number of advantages over conventional memory technologies such as SRAM and DRAM. These include a smaller area requirement, a lower energy requirement for reading and partly for writing, too, and, of course, the non-volatility and especially the qualitative advantage of multi-bit capability. It is expected that memristors based on resistive random access memories (ReRAMs), phase-change memories, or spin-transfer torque random access memories will replace conventional memory technologies in certain areas or complement them in hybrid solutions. To support the design of systems that use NVMs, there is still research to be done on the modeling side of NVMs. In this paper, we focus on multi-bit ternary memories in particular. Ternary NVMs allow the implementation of extremely memory-efficient ternary weights in neural networks, which have sufficiently high accuracy in interference, or they are part of carry-free fast ternary adders. Furthermore, we lay a focus on the technology side of memristive ReRAMs. In this paper, a novel memory model in the circuit level is presented to support the design of systems that profit from ternary data representations. This model considers two read methods of ternary ReRAMs, namely, serial read and parallel read. They are extensively studied and compared in this work, as well as the write-verification method that is often used in NVMs to reduce the device stress and to increase the endurance. In addition, a comprehensive tool for the ternary model was developed, which is capable of performing energy, performance, and area estimation for a given setup. In this work, three case studies were conducted, namely, area cost per trit, excessive parameter selection for the write-verification method, and the assessment of pulse width variation and their energy latency trade-off for the write-verification method in ReRAM.

Keywords: memristor, ternary system, analytical circuit model, ReRAM, ternary memory model, non-volatile memory, write-verification programming

OPEN ACCESS

Edited by:
Huanglong Li,
Tsinghua University, China

Reviewed by:
Rui Yang,
Shanghai Jiao Tong University, China
Zhong Sun,
Peking University, China

****Correspondence:***
Shima Hosseinzadeh
shima.hosseinzadeh@fau.de

Specialty section:
This article was submitted to Nanodevices, a section of the journal Frontiers in Nanotechnology

Received: *27 August 2021*
Accepted: *25 October 2021*
Published: *08 December 2021*

Citation:
Hosseinzadeh S, Biglari M and Fey D (2021) TReMo+: Modeling Ternary and Binary ReRAM-Based Memories With Flexible Write-Verification Mechanisms. Front. Nanotechnol. 3:765947. doi: 10.3389/fnano.2021.765947

1 INTRODUCTION

Ever since the creation of the digital computing systems, the base of two has mostly been utilized for information processing and communication. Nevertheless, it is long known that a ternary representation of data, i.e., for each digit d_i of a number holds, e.g., $d_i \in \{-1, 0, 1\}$ or $d_i \in \{0, 1, 2\}$, offers advantages over the binary system in some aspects (Metze and Robertson, 1959;

Avizienis, 1961; Parhami, 1470). One of the most attractive merits of using the ternary system is its capability of carrying out an addition operation in two steps, i.e., in O (1), regardless of the operand length [an example can be found in the work of Fey (2014)]. Using a binary data representation, this can be done only in O (log(N)) with a reasonable hardware effort. Furthermore, neural networks with ternary weights are much better than ones with binary weights and not much worse than ones with floating-point weights concerning the recognition accuracy, and they require much less storage capacity than neural networks with floating-point weights (Yonekawa et al., 2018).

However, realizing ternary states with binary storage elements requires two binary storage elements, e.g., flip-flops (Rath, 1975), making such designs immensely expensive. With the emergence of CMOS-compatible multi-bit capable memristive resistive random access memories (ReRAMs)[1], this situation changed. This is achievable by ReRAMs because their resistive window can be splitted into quantized levels for having multilevel states (El-Slehdar et al., 2013). The idea of programming memristive devices into several resistance states was proposed, e.g., in the work of Kinoshita et al. (2007), in which the authors analyzed the application of a thin-film memristor as an N-level ReRAM element. Another approach, which was introduced by Junsangsri et al. (2014), uses two memristors to obtain three different states to handle ternary states instead of multiple quantized memristive levels but loses the advantage of saving one storage cell compared to multi-bit approach.

Using memristive devices for ternary arithmetic was first investigated by Fey (2014). On the basis of the work of Fey et al. (2016), the improvement in the energy-delay product and area for a ternary adder circuitry using multi-bit registers based on memristors compared to SRAM-based solutions was shown. The architecture can be further enhanced by using memristor-based pipeline registers that make it possible to use homogeneous pipelines for not only the addition operation but also the subtraction and multiplication operations instead of superscalar pipelines that use different pipeline paths for various operations (Fey, 2015). Although various proposals for ternary memristive circuits are now available in the literature, there is still a lack of sufficient ternary modeling at the circuit level to be able to use such components systematically and more easily than today in one's own circuits.

Memory modeling enables architectural exploration and system integration of different memory technologies and design approaches. To ease the process of memory modeling, a need for a comprehensive modeling tool seems to be evident. Luckily, some high-precision open-source modeling tools such as CACTI (Wilton and Jouppi, 1996; Thoziyoor and Ahn, 2008), NVSim (Xiangyu Dong et al., 2012), and Destiny (Mittal et al., 2017) enable designers not only to utilize them with their original offered toolsets but also to build upon the current features for state-of-the-art modeling, which, in our case, is ternary memory modeling.

Research and development on non-volatile memories (NVMs) either require prototype chips, which are limited to a small portion of the entire design area, or a simulation tool that estimates energy, area, and performance of NVMs with different design specifications before the real chip fabrication. When designing a ternary system, researchers cannot benefit from any of the aforementioned solutions because there are no ternary memory chip fabrications and appropriate simulation tools have yet to be developed. Although the current most popular NVM simulation tools offer some design and estimation features, they still have limitations with respect to ternary memory design and accurate evaluation.

In this work, for the first time, to the best of our knowledge, a new ternary model has been developed that utilizes different reading and writing methods. Moreover, a comprehensive simulation tool for ternary memory modeling has been developed, which uses the NVSim (Xiangyu Dong et al., 2012) as its base. The main contributions of this work are as follows:

- Development of a comprehensive simulation tool for ternary memory modeling called "TReMo+". On the one hand, the TReMo+ benefits from the methods and feature sets used by the most well-known memory simulation tools, namely, NVSim (Xiangyu Dong et al., 2012) and Destiny (Mittal et al., 2017), and on the other hand, it adds some more features for the first time ever.
- One of the unique features of the TReMo+ is that it supports the generic write-verification method for both reset and set operation, with the capability of overwriting average iteration, and different pulse width and voltage or current amplitude for consecutive pulses. This write method is made available for both the binary and ternary memory models.
- For the first time, TReMo+ introduced two new read methods, namely, serial and novel parallel read, which are configurable based on the desire of the user. The serial read method was adapted from the work of Mittal et al. (2017), and the novel parallel read approach was introduced in our previous work (Hosseinzadeh et al., 2020).
- Because the TReMo+ supports not only binary but also ternary memory modeling, the tool now enables users to choose optimization target for ternary memory (alongside with binary memory modeling), which could be area, latency, and energy.

Furthermore, we demonstrated the application of our model with three case studies. In addition to area cost per trit evaluation studied in our previous work (Hosseinzadeh et al., 2020), we present two further case studies in this work, namely, excessive parameter selections for the write-verification method and programming pulse width assessment. In the first case study, the impact of Incremental Step Pulse and Verify Algorithm (ISPVA) on delay and energy consumption is investigated to achieve more reliable writing operations and compared it to other known methods. These comparisons between different write schemes including the overhead and enhancements (as a trade-off analysis) are possible by using the presented model using the TReMo+ tool that we developed.

The TReMo+ modeling tool can assist researchers who are modeling systems in architecture-level tools such as gem5 (Binkert et al., 2011) by estimating performance, energy, latency, and area of the ternary ReRAM-based memory models. This tool

[1]The term memristor and ReRAM are used in this paper interchangeably.

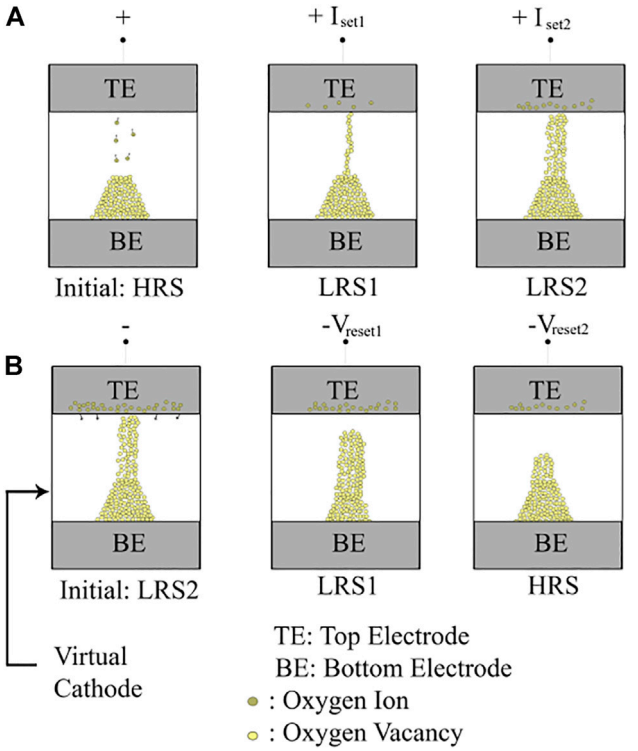


FIGURE 1 | Multi-level switching in ternary ReRAM: **(A)** H2L and **(B)** L2H programming.

also gives memory designers the ability to employ ternary logic based on ReRAM in their designs. The benefits of this work are not only limited to stand-alone ternary logic but also include exploiting new storage mechanisms and architectures. In other words, TReMo+ supports the use of innovative computing storage technology in own CMOS-based designs.

The rest of the paper is structed as follows: In **Section 2**, some basic information about the ReRAM will be presented, and different reading and writing methodologies on this memory will be studied. In **Section 3**, after having a deep overview of the state-of-the-art memory simulation tools, a thorough comparison among them will be reported. **Section 4** is about implementation of the novel read and write methods in ReRAM devices, followed by **Section 5**, in which the results will be presented. Last, in **Section 6**, three case studies will be elaborated, and a brief conclusion will be presented in **Section 7**.

2 PRELIMINARIES

2.1 ReRAM

Many of the NVM technologies, such as PCRAM and STTRAM, are designed on the basis of electrically inferred resistive switching effects. ReRAM is implemented by utilizing electro- and thermochemical effects, resulting in the resistance change of a memory architecture, in which a metal/oxide/metal layer stack is used to store data (Hosseinzadeh et al., 2020). In our confined variation, which is a bipolar ReRAM, a metal oxide layer (e.g., TiO_2, HfO_2) is sandwiched between two metal electrodes to store data. The value stored in the memory is dependent on the oxygen vacancy concentration of the metal oxide layer. When a voltage is applied to the two electrodes, conductive filaments (CFs) are either formed or ruptured, depending upon the voltage polarity. In case of CF formation inside the metal oxide, the top and the bottom of the electrodes are bridged, and the current can flow inside the CF. In this situation, the cell is considered to be in a low-resistance state (LRS), representing the value of "1". Oppositely, when the CF is ruptured, the top and the bottom electrodes are disconnected and thus result in a high-resistance state (HRS) representing the value of "0" (Yang et al., 2008).

It has been proven by Xu et al. (2013) that the size of the CF has a direct relation with the value of the current, meaning that the cell resistance can be controlled by changing the strength of the CF. Therefore, it would be possible to program the middle-level resistance of ReRAM between the HRS and the LRS by manipulating the programming current and to establish by the multi-bit capability.

Figure 1 represents the physical behavior of a bipolar ternary ReRAM memory. As it can be seen in **Figure 1A**, by increasing the size of the CFs, the resistance is decreased, resulting in two distinct LRSs, namely, LRS1, and LRS2. On the other hand, as it can be seen in **Figure 1B**, by decreasing the size of the CFs, the resistance is increased, resulting in the HRS. Programming to intermediate states can be started from either the highest-resistance state (H2L programming) or the lowest-resistance state (L2H programming).

2.2 Read Methodologies in ReRAM

The normal read operations in ReRAM and many other NVM technologies are identical. The read operation can be done in two ways, in which both of them take advantage of the fact that NVMs have different resistances in LRS and HRS states. In the first method, a small voltage is applied on the bitline attached to NVM storage cell, and the current moving through the cell is measured. In the second method, a small current is sent out in the bitline, and in return, the voltage across the memory cell is measured. The methods are known as current sensing or voltage sensing, respectively. The response back from the cell comes in the form of voltage (or current), and afterward, it is compared against a reference voltage (or current). The comparison is done by utilizing a sense amplifier (SA) (Xiangyu Dong et al., 2012).

Depending on the resistance levels stored in one cell, the number of SAs varies. In the case of SLC (single-level cell or 1 bit per cell), it is sufficient to use one SA for the read operation (Xiangyu Dong et al., 2012), whereas in non-SLCs, the number of the SAs should be more than one, depending on whether the read operation is done in serial or in parallel.

2.2.1 Serial Read

Serial sensing for the non-SLC memories can be done by two methods. In our case, non-SLC memories consist of MLC (multi-level cell or 2 bits per cell for storing four states), TLC (triple-level cell or 3 bits per cell for storing eight states), and ternary (three states in 2 bits per cell) memories. The first method is the sensing

model, which is based on the multi-step "sequential single reference". This method is based on the non-linearity nature of charging and discharging resistance of the NVMs. Within the resistance change time, the SA captures samples from it (Xu et al., 2013). The second method is the binary search read out model, in which the number of read out iterations is based on the number of stored bits in the cell (Mittal et al., 2017).

2.2.2 Parallel Read

In the parallel sensing method, only a single step is needed, but the current (or voltage) is compared with multiple current (or voltage) references (Xu et al., 2013). On the basis of the work of Xu et al. (2013), the MLC parallel read circuitry is associated with seven sets of SA.

In our work, ternary read circuitry could be the binary search readout method or the parallel read method. We carry out two comparisons using the binary read approach for the ternary memory. For distinguishing the resistances in ternary memories with the parallel read approach, two SAs are enough.

2.3 Write Methodologies in ReRAM

A set operation is defined as switching between HRS and LRS, and reset is vice versa (Biglari et al., 2018). Because there is a large resistance variation, cell programming with verification could add an extra level of reliability (Higuchi et al., 2012). To control the cell programming in intermediate states, either the DC sweep (Grossi et al., 2016), write-verification (Higuchi et al., 2012; Song et al., 2013), or ISPVA (Higuchi et al., 2012) is applied, which could start from lowest- to highest-resistance state (L2H) or vice versa (H2L).

The ISPVA is based on a chain of increasing voltage pulses on the drain electrode during set operation, whereas during reset operation, this sequence of pulses is applied to the source terminal. After applying each pulse, a read verification is done to check whether the read current has reached the threshold value for the set and the reset operation. The algorithm stops when the threshold is reached (Pérez et al., 2018). Although single pulse benefits from shorter forming time by using high compliance and voltage parameters (Grossi et al., 2016), ISPVA offers a wide range of advantages including improvements in spatial process variation, more reliable writing, and higher endurance (Pérez et al., 2018; Pérez et al., 2019).

2.4 Trade-offs in Writing Parameter Selections

The high cycle-to-cycle and device-to-device variability in switching characteristics of ReRAM devices will result in excessive electrical stress on ReRAM cells during the worst case–based programming (Biglari et al., 2019). This contributes to a higher energy consumption as well as reduced reliability (Yu et al., 2012) and endurance (Song et al., 2013). To tackle this problem, novel structures have been proposed that intrinsically reduce this stress at the cell level (Linn et al., 2010; Biglari and Fey, 2017). Write-verification (Song et al., 2013; Higuchi et al., 2012) and feedback-based programming (Lee et al., 2017; Biglari et al., 2018) terminate the write operation after detecting that the device has reached the desired state.

In write-verification programming, this detection is done by reading the device between programming steps, whereas in feedback-based programming, the resistive state of the cell is monitored at real time during programming. The ISPVA method mentioned in the previous section is in the category of write-verification method. Both methods also enable multi-level programming of the ReRAM cells (Lieske et al., 2018; Puglisi et al., 2015). This work models a write-verification method that is the most common practice for memory design.

Although bearing the extra cost of the write-verification method is undeniable, it can be seen in other experiments that the ISPVA method was utilized both for SLC and MLC types of memory, mainly due its numerous advantages mentioned above (Pérez et al., 2018; Pérez et al., 2019).

A key capability of a memory model is to demonstrate how the observed behavior of a memory cell (in this case, ReRAM) at the device level will affect the overall behavior and performance characteristics of complete memories constructed with it.

In this case study, we study how write-verification parameter selection affects delay and energy consumption of the realized memory in relation to its endurance and reliability properties.

3 SIMULATION TOOL

3.1 The NVSim Tool

To investigate early phases of NVM design, a simulator for ReRAM circuit level design is needed, so that the evaluation without any real-chip fabrication can be done. Among existing tools used in industry and academia for NVM estimation, the NVSim (Xiangyu Dong et al., 2012) and Destiny (Mittal et al., 2017) are the most popular ones.

The NVSim simulates some of non-volatile memristor–based memory technologies, such as phase-change memory, spin-transfer torque random access memory, and ReRAM. As an input, the NVSim takes device parameters and optimizes the circuit design and, as an output, evaluates the area, energy, and performance with the given design specification.

NVSim organizes chips using three main building blocks: bank, mat, and subarray. As shown in **Figure 2**, the top level building block in the hierarchy is the bank, and each bank consists of some mats, and last, subarrays are designed inside mats as the basic structure of a memory, in which they contain memory arrays and peripheral circuitry. The peripheral circuitry has SAs, a multiplexer (Mux), a decoder, and an output driver, and the overall cell layout is controlled by the access transistor. In **Figure 3**, the peripheral circuitry associated with the bitline of the subarray, used by NVSim, is depicted (Xiangyu Dong et al., 2012).

The NVSim only models the SLC memories with regard to the submitted code. A more recent fork of NVSim, called Destiny, introduced a design evaluation of MLCs. In our work, a novel design for ternary memory simulation is implemented by heavily modifying the original NVSim code. The main focus of our work is internal sensing, and changes for ternary modification are done in the subarray level, especially in the peripheral circuitry, and then, the effects are evaluated in higher levels of the cell design. Needless to say, this work

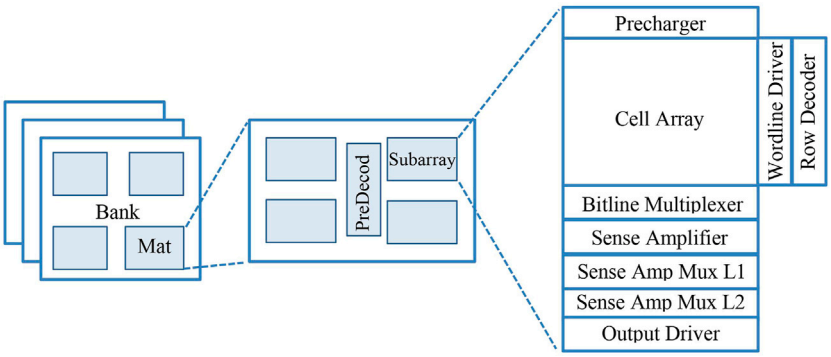


FIGURE 2 | Memory array organization in NVSim (Xiangyu Dong et al., 2012).

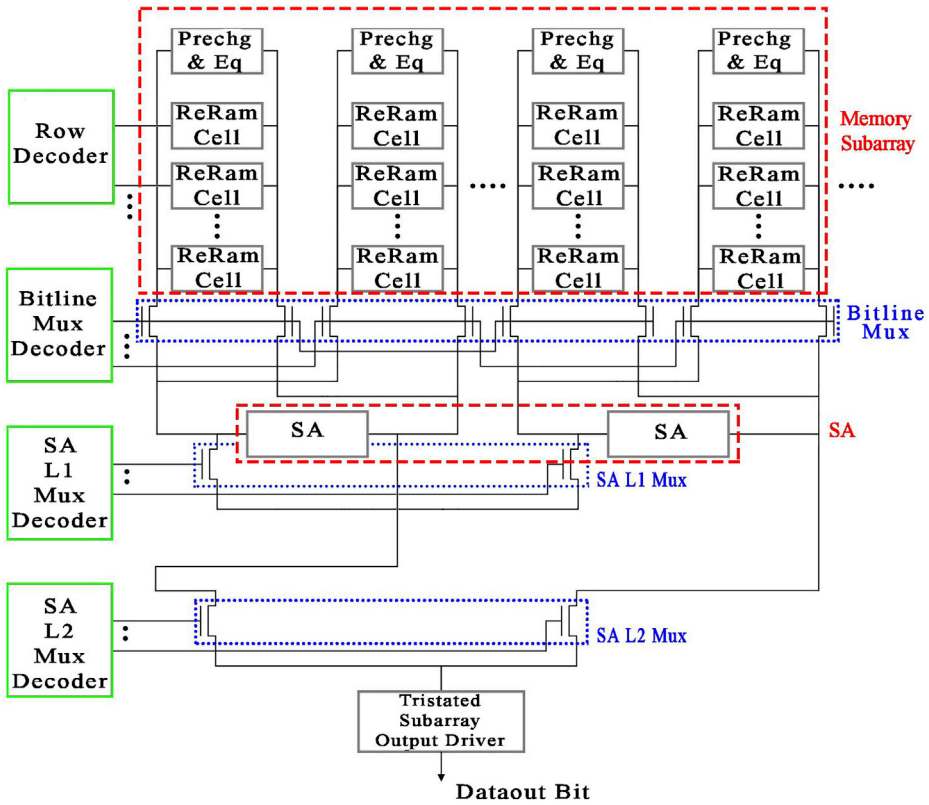


FIGURE 3 | Peripheral circuitry associated with bitlines in the NVSim.

focuses on modeling of memories and not designing circuits. Therefore, we modeled the ternary model on the basis of the block diagrams. Describing the details regarding the building block of our models is out of scope of this work. However, for circuit detail of every module, the base of most of modules is described in the manuscript and guideline of CACTI (Wilton and Jouppi, 1996; Thoziyoor and Ahn, 2008) and some small parts in NVSim (Xiangyu Dong et al., 2012). Furthermore, our solution differs in further features that are outlined next.

3.2 Simulation Tools Comparison

Among many NVM simulation tools, NVSim (Xiangyu Dong et al., 2012) and Destiny (Mittal et al., 2017) are the ones offering the richest features. However, these tools lack certain essential features for more accurate results and for maintaining the fast-paced NVM technology. The present work addressed some of these issues by adding the missing features to the toolset.

For instance, NVSim only supports SLC design, whereas Destiny included MLC, allowing a cell to store 1 bit, 2 bits, 3

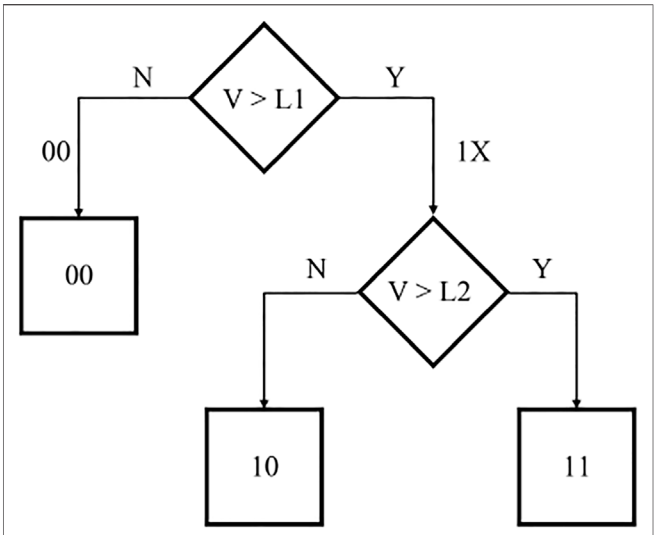


FIGURE 4 | Serial read model for ternary memory.

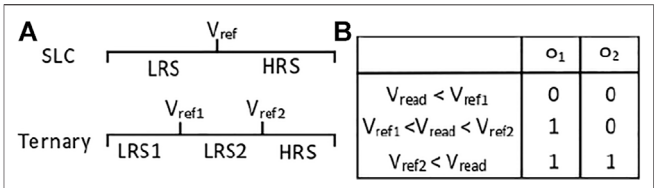


FIGURE 5 | (A) Resistance level in ternary memory. **(B)** Truth table for parallel ternary read.

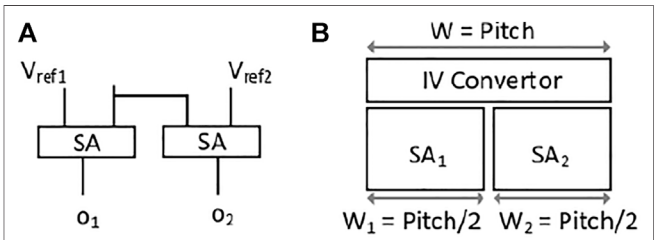


FIGURE 6 | (A) SA circuitry for ternary memory in parallel read. **(B)** The ternary SA block layout. Pitch is the maximum allowed width for one SA layout in NVSim.

bits, etc. The present work introduced the support for ternary memory cells considering three states for the first time.

The support for the generic write-verification–based method that is capable of variant pulse width and variant current or voltage amplitude for both set and reset operation is another feature added in TReMo+, which was entirely absent in the NVSim. In addition, there are no verifications done when writing data, neither in reset before set nor in set before reset in NVSim, whereas in TReMo+, the verification is possible for both cases. In Destiny, only the write-verification method with fixed voltage and current is supported. Although not directly mentioned in the Destiny paper, it is evident that, in latency and energy calculation formulae, time pulses for voltage and current are equal, which is not the common write method for memories. For more realistic and accurate results, in TReMo+, we added the enhanced variant of write-verification method for both reset and set operations, namely, 1) with the dynamic voltage levels and pulse widths and 2) current levels with the variant pulse widths. Moreover, TReMo+ has two read methods, namely, serial and the novel parallel read methods, whereas in Destiny, only serial read is available.

4 IMPLEMENTATION AND METHODOLOGY

4.1 Sense Amplifier Read Circuitry

To adapt the NVSim SA read circuitry to the ternary memories, it is necessary to take both read methods into consideration, specifically the serial read and the parallel read.

In the serial read, there are no modifications needed to the internal SA block of the original NVSim. However, the total number of SAs are halved, due to the halved number of columns in ternary memory. It is notable to mention that, for the serial read, as shown in **Figure 4**, the maximum number of read iterations should be two times.

In contrast, the parallel read requires some adjustments in the NVSim SA read circuitry. To store three values in one cell, it is necessary to have a trit cell; we therefore added an extra SA coupled with the existing SA so that it would be possible to read from a cell concurrently. Storing ternary data requires at least three distinguished levels of resistance. To accomplish this, at least two sense SAs with different voltage references are required. Therefore, one bitline should be connected to two SAs.

The general idea of dividing and distinguishing the resistance level in parallel read circuitry is demonstrated in **Figure 5**. As a result, V_{ref1} and V_{ref2} should adhere to the following rules: 1). To prove this design, the truth table is shown in **Figure 5B**.

$$V_{LRS1} < V_{ref1} < V_{LRS2}, \quad V_{LRS2} < V_{ref2} < V_{HRS} \tag{1}$$

The original SA in NVSim is based on the SA used in the CACTI (Thoziyoor and Ahn, 2008) tool, which is voltage-based. Therefore, we kept this module unchanged. In case of current sensing, an I-V converter is needed, which is responsible for converting the current running in the bitlines to voltage before passing through the SAs. Because two SAs work simultaneously, one I-V converter is sufficient to be shared among two SAs in case of current sensing depicted in **Figure 6B**.

4.2 Write Operation Modeling for Single-Level Cell and Non–Single-Level Cells

4.2.1 Single Write

In non-crossbar structures, the write pulse is applied once to the cell, assuming that the cell will be written in only one single pulse. Writing to the cells is performed in two steps. First, the row

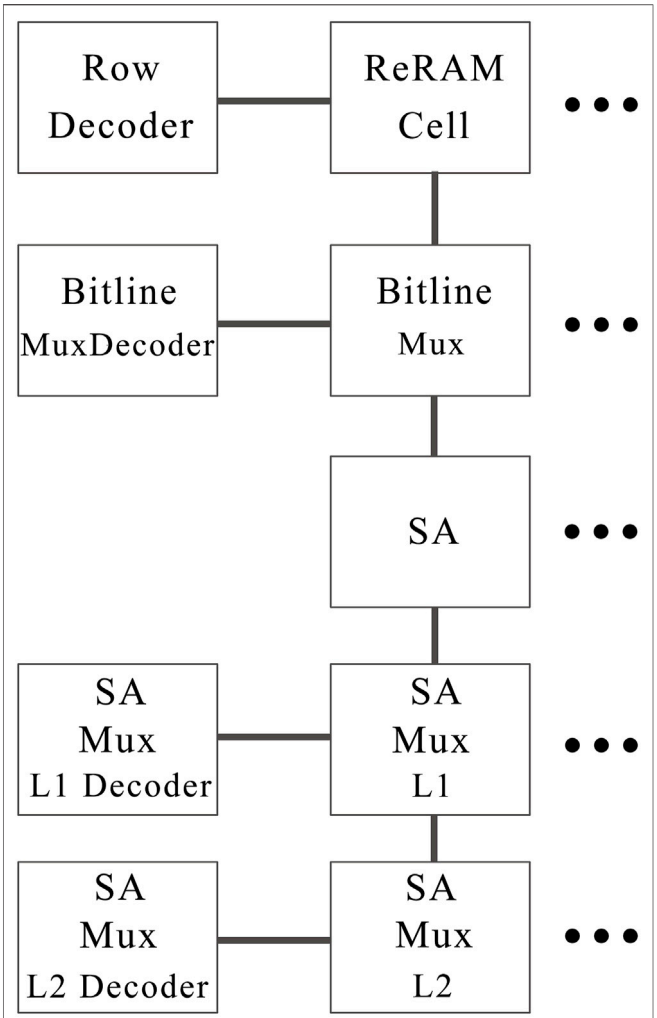


FIGURE 7 | The write path for one cell at the subarray level.

decoder applies the row address to latch the data from a row of the ReRAM subarray module into the SAs. Second, after the data become latched, the column address is applied, and the read or write access will be performed. **Figure 7** shows the write path for a single cell. The latency is calculated by summing the worst-case latency of reset and set pulse, the maximum value among decoder latency, and the summation of the column decoder latency (calculated by summation of latency of bitline Mux decoder, SA Mux decoder level 1 and level 2 modules) and the latency of other modules in the write path (calculated by summation of latency of bitline Mux, SA Mux level 1 and level 2).

In crossbar structures, because set and reset operations cannot be performed simultaneously, two methods for write operations are available; first, having a separated set and reset operation called "reset before set" or "set before reset" method and, second, in which all the cells in the selected row are erased before a selective set operation is carried out. This method is called the "erase before set" or "erase before reset" method (Xiangyu Dong et al., 2012).

4.2.2 Verification After Single Write

Another write scheme that is utilized and modeled in this work is the verification after single write. The latency of any cell type (crossbar or non-crossbar) with the "write and verification" scheme is higher than the "without verification" one. The read latency itself comes from the latency of every module in the read path sequentially, for instance, SAs, bitline Muxes, and different multiplexers that come after the SAs or the decoders. From the write energy perspective, in the "write and verification scheme", the energy consumed for the verification, specifically in the cell and the SA, are added to the write energy. Therefore, the write energy in this scheme is higher than that of the "write without verification" scheme.

4.2.3 The Write-Verification Method

There are two variants of write-verification methods that have been modeled in this work. The first variant is based on the write method used by Xu et al. (2013). On the basis of this variant of write-verification method, first, the device is initialized to reset state by a single pulse followed by an iterative sequence of set and verification pulses until the device has reached the desired resistive level (**Figure 8A**) or vice versa. The second variant is based on the write method used by Pérez et al. (2017). On the basis of this variant of write-verification method, first, the device is being initialized to reset state with a sequence of iterative reset and verification pulses. Then, it is programmed to the desired resistive state by a sequence of iterative set and verification pulses (**Figure 8B**).

The average energy for single-pulse–based reset (first variant) is calculated by the following:

$$E_{reset} = V_{reset} \times \left(V_{reset} - V_{drop,reset}\right)/R_{LRS} \times t_{reset} \tag{2}$$

The amounts of energy consumed during the sequence of program and verification pulses for the reset operation in second variant and the set operation for both variants are calculated by either (3) or (4) as follows. It is considered that the average number of iterations for set and reset operations is assigned to variable "n" and "m", respectively.

If the set or reset operation holds the current source, then the energy is calculated by the following:

$$E_{set|reset} = \sum_{i=1}^{n|m} \left((vdd \times PI[i] \times PT[i]) + \left(\left(V_{read} - V_{drop,read}\right)/R_{LRS} \times vdd \times t_{read}\right)\right) \tag{3}$$

If the set operation holds the voltage source, then energy for the set operation is calculated by the following:

$$E_{set|reset} = \sum_{i=1}^{n|m} \left(PV[i] \times \left(PV[i] - V_{drop,set}\right)/ R_{LRS} \times PT[i] + V_{read} \times I_{read} \times t_{read}\right) \tag{4}$$

PV = $[v_1, v_2, \ldots, v_n|v_m]$ consists of a sequence of voltages in the write-verification method for the set or reset procedure.

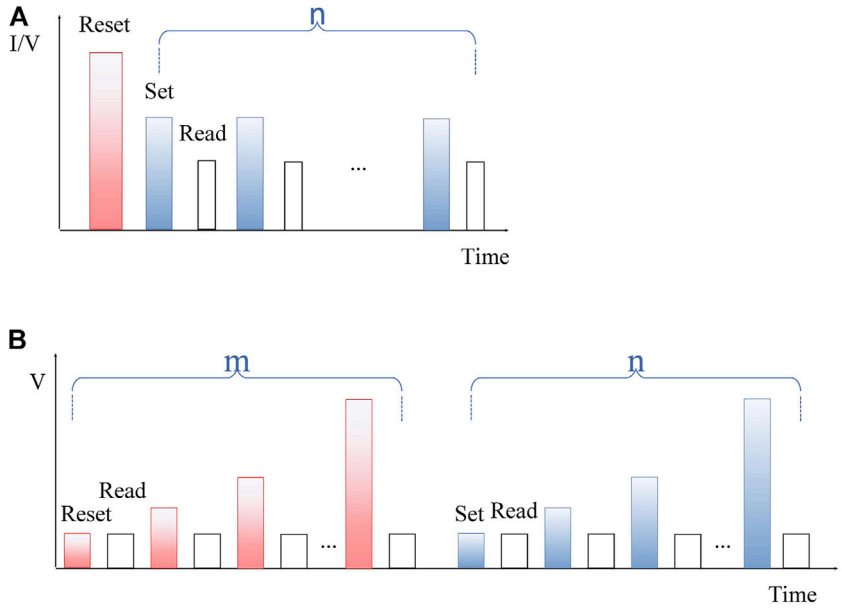


FIGURE 8 | (A) The first variant of write-verification model (reset before set). **(B)** The second variant of write-verification model (ISPVA).

PI = $[I_1, I_2, \ldots, I_n|I_m]$ consists of a sequence of currents in the write-verification method for the set or reset procedure.
PT = $[t_1, t_2, \ldots, t_n|t_m]$ consists of a sequence of pulse widths of current or voltage pulses for the set or reset procedure.

V_{drop} is the voltage dropping on the device due to the transistor connected to the cell while reading or writing.

The total required energy for writing is as follows:

$$E_{write} = E_{reset} + E_{set} \tag{5}$$

The total latency for writing for the first variant write-verification is calculated by the following:

$$Latency_{write} = n \times t_{read} + t_{reset} + \sum_{i=1}^{n} PT[i] \tag{6}$$

The total latency for writing for the second variant write-verification is calculated by the following:

$$Latency_{write} = (n+m) \times t_{read} + \sum_{i=1}^{m} PT[i] + \sum_{i=1}^{n} PT[i] \tag{7}$$

4.3 Analysis of Single-Level Cell and Parallel Ternary

In this section, an architecture for ternary memory in the subarray level is modeled and evaluated in terms of area, latency, and dynamic energy.

4.3.1 Ternary Area Consumption

Figure 9 shows the SLC and the ternary memory in *parallel read models* in one frame. Given an SLC memory with the capacity specified by the product of the number of rows and columns, represented by the color black, a ternary memory with the same capacity is compared to it using the color red.

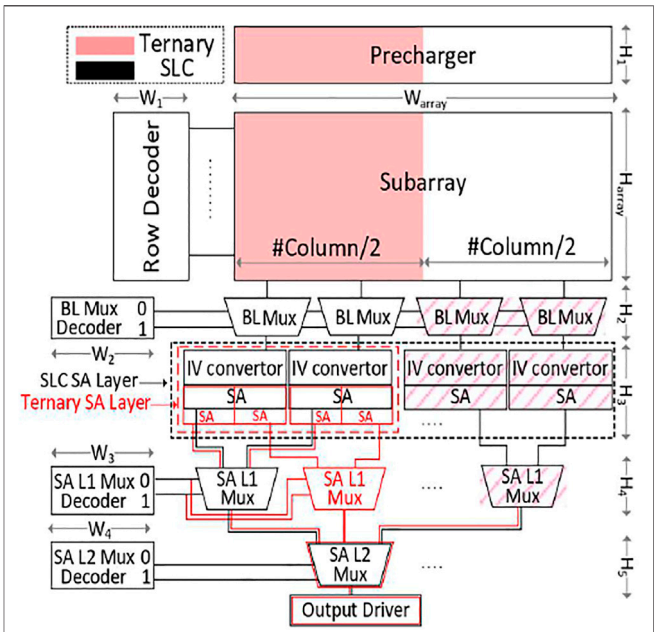


FIGURE 9 | The peripheral circuitry for SLC memory (black bordered modules) and the proposed ternary memory (red bordered modules) in parallel read.

It can be seen that, in the parallel ternary memory, there has been some modifications, when compared to the SLC memory. The first change is the number of columns is halved in the

subarray because each cell is capable of storing a trit. Subsequently, the width of the precharger is also halved for the same reason mentioned above. Moving forward to the layers below, the number of bitline multiplexers is halved, caused by the reduced number of columns. In the SA layer, the total number of internal SAs is kept unchanged, whereas the width of the SA layer is halved. It is also notable to mention that the height of each internal SA is slightly longer than that of SLC, but the effect of this is neglectable in the SA layer because the height of the I-V converter is dominant. In the next layer, namely, SA multiplexer level 1, the number of multiplexers has not changed, obviously because the number of SAs in the previous layer was kept constant. The same logic applies to the SA multiplexer level 2, and therefore, the total number is kept unchanged.

To estimate the area of each peripheral circuitry component, each component is delved into the actual gate-level logic design considering the height and width of each gate as it is also done in NVSim and CACTI. The height and width of each gate is dependent on the optimization target as we have three different types of transistors (latency-optimized, balances, and area-optimized) with different sizes. When calculating the total area at the subarray level in SLC memory, the following formulas are used (8) (9).

$$H = \sum_{i=1}^{5} H_i + H_{Array} \tag{8}$$

$$W = MAX\left(\sum_{i=1}^{4} W_i\right) + W_{Array} \tag{9}$$

H_1, H_2, H_3, H_4, H_5, and H_{array} are the height of precharger, bitline Mux, SA layer, SA Mux level 1, SA MUX level 2, and subarray modules, respectively, as depicted in **Figure 9**. W_1, W_2, W_3, W_4, and W_{array} are the width of row decoder, bitline Mux decoder, SA Mux decoder level 1, SA Mux decoder level 2, and subarray modules, respectively, as depicted in **Figure 11**. The total area is calculated by multiplication of the total height and the width. When evaluating the area consumption of the ternary memory with parallel read mode, all the heights and widths measurements are the same, except for W_{array}, which is halved, and the height of the SA. As a result, the total subarray area of the SLC memory is almost two times greater than that of the ternary memory.

In the case of ternary memory with *serial read mode*, there are some minor changes when compared to ternary memory with parallel read mode. First, the number of SAs is halved, whereas the width of each SA is doubled, keeping the total width of the SA layer unchanged. Second, the number of SA multiplexer level 1 and level 2 is halved because of the decreased number of SAs in the previous layer.

4.3.2 Ternary Latency

The latency calculated for the components is based on RC analysis and the simplified version of Horowitz's timing model that is used in the NVSim tool (Xiangyu Dong et al., 2012) .

$$Delay = \tau\sqrt{\left(ln\frac{1}{2}\right)^2 + \alpha\beta} \tag{10}$$

In this formula, α is the slope of the input, $\beta = g_m R$ is the normalized input transconductance by the output resistance, and τ is the RC time constant. When comparing the latency of the SLC memory cell with the ternary memory cell, some differences in the latency of each component can be found.

Row decoder is the first component with halved latency. The reason is that the number of subarray columns is halved, which results in halved wordline capacitance that is loaded to the row decoder.

Bitline multiplexer decoder is another component with halved latency, when compared to that of SLC. The capacitance loaded to this module comes from the capacitance of the wordline and the pass transistors of the multiplexers, and they are both halved. The same reason applies to the SA multiplexer level 1 and level 2 decoders. The total decoder latency is calculated by finding the maximum latency of the modules mentioned above, resulting in halved decoder latency.

When reading from the memory cell, the total read latency is the summation of the decoder latency, the bitline delay, and the delay of multiplexers through the read path. In the case of ternary memory with parallel read mode, the total read latency is less than that of SLC due to the halved decoder latency mentioned above, leaving the other latency values unchanged. However, in ternary memory with serial read mode, in addition to the halved decoder latency, SAs latency is doubled because binary search reading should be done at least twice. The comparison between the ternary memory with serial read mode and parallel read mode is also an interesting matter, because the read latency of the ternary memory with parallel read mode is lower than that of serial one. It can be justified that parallel read sensing is done in parallel with the use of the SA, whereas in serial read mode, two times more comparisons are needed in the worst-case scenario.

If we put the write latency under scrutiny, then we realize that the latency of the ternary cell is higher than that of SLC because the programming ternary cells need more write iterations than SLC. When comparing the write latency of the ternary memory with parallel read mode with that of the serial one, the write latency in the parallel mode is lower due to lower read latency in the parallel mode during the writing program by write-verification compared to the serial read.

4.3.3 Ternary Energy Consumption

The energy consumption comparison is done in this section between the ternary memory and the SLC memory type. The dynamic energy and leakage power consumption can be modeled as follows:

$$\begin{aligned} E_{dynamic} &= C \times V_{DD}^2 \\ P_{leakage} &= V_{DD} \times I_{Leak} \end{aligned} \tag{11}$$

The dynamic energy of the precharger is halved because of the halved number of columns, which is caused by dividing the capacitance of the wordline by two. Dynamic energy of other

modules including the decoders of the bitline multiplexers, row decoder, and SA level 1 and level 2 decoders is also halved for the exact same reasons mentioned above. The read dynamic energy consumed in SAs is the same in both cases because they have the same number of SAs. The dynamic energy of bitline multiplexer is unchanged; although the load capacitance of the two SAs connected parallel to it is doubled, the number of columns is halved.

Cell read energy is also lower in ternary memory, because, first, read pulse is not considered in the calculation and, second, the number of columns in the cell is halved compared to that of SLC. When reading from the ternary memory cell, the read dynamic energy is calculated by adding all dynamic energy of the active components mentioned above and cell read energy in the read path. In parallel read, sensing is done in parallel with the use of SA, whereas in serial read, two times more comparisons are needed in the worst-case scenario. As a result, the read energy consumption during serial read is higher than that of the parallel variant.

For a write operation on the ternary memory cell, the write dynamic energy is the sum of all active modules mentioned above plus write dynamic energy of the write path depending upon the writing method. When the write-verification method is used for writing data on a ternary cell, it will definitely need more iterations compared to the single-pulse method, resulting in higher write energy in ternary memory. It can therefore be concluded that the total dynamic energy in ternary is greater than SLC, despite the number of columns in SLC being two times more. It is worth to consider that the write energy for the ternary memory with parallel read mode is higher than that of the ternary memory with serial read mode because the two SAs are used for concurrent sensing passing through the bitline multiplexer that doubles the capacitance.

If the reset dynamic energy and the set dynamic energy were analyzed separately assuming the first variant of write-verification, the reset dynamic energy in SLC would be greater than in the ternary memory because the number of columns in SLC is higher than in the ternary memory. However, because the number of iterations in ternary memory is higher than SLC, the set dynamic energy in this memory is greater than SLC, outweighing the number of columns in ternary.

Regarding the cell leakage, the total leakage in SLC is higher than that of the ternary memory. The reason behind this is the effect of the dominant precharger leakage in SLC on the total leakage.

5 RESULTS

5.1 Single-Level Cell ReRAM With Write-Verification

The motivation for this section is to demonstrate some additional enhancements on SLC memory models, previously modeled in NVSim, such as the verification after single write or first variant of write-verification method by considering the overhead of the verification controller as input.

In cases of verification-based write method, a finite state machine (FSM) is required to control the write scheme. For instance, when a write voltage is applied, the state machine is utilized to verify the current whether the write was successful followed by iteration termination or the voltage should still be increased. The overhead values of this state machine, including the energy, latency, and area, are technology dependent; therefore, these values can be estimated by the synthesis results of the desired controller. The FSM overhead values, including the area, latency, and energy overhead, are then given as an input to the simulator for a more accurate result. Therefore, TReMo+ is capable of getting the overhead of write driver as an input to make estimated values closer to real fabricated chip values. However, our evaluations for the memory arrays are based on the IHP cell settings, and they also do not have a write-driver to produce the pulse trains. The pulse trains in IHP company are produced with a computer-based system called RIFLE SE. Therefore, even the IHP researchers do not have the overhead values for producing the consecutive pulses, and as a result, the overhead of the control circuitry is not considered in the results.

The SLC ReRAM model used for this section is based on a 0.18-μm 4-Mb MOS-accessed ReRAM prototype chip (Sheu et al., 2011). According to Xu et al. (2013), the set and reset pulse duration were set to 5 ns

Table 1 contains a thorough comparison of 1T1R and crossbar architecture, each with and without verification after the writing scheme with different underline physics.

As it can be seen, the verification method after writing to the cells has increased the write energy and the latency based on the explained reasons in **Section 4.3**. The write latency has increased at least by 42%.

5.2 Ternary Memory With Serial and Parallel Modes Vs. Single-Level Cell Memory

The experimental results shown in **Table 2** are based on the prototype chip of Sheu et al. (2011) with the first variant of write-verification method for different memory models including SLC and ternary memroy with serial and parallel mode. It is assumed that the average number of iterations set in the first variant of write-verification method for SLC and ternary model are 5 and 12, respectively. In addition, the projected results for the ternary memory in serial and parallel modes are compared with the SLC memory in **Table 2**. As it can be seen in **Table 2**, the parallel read has a lower read latency in comparison with the serial read while keeping the overhead to a minimum level.

In MLC mode, the ReRAM prototype chip has a write latency of 160 (Sheu et al., 2011). Using first variant of write-verification method with number of set iterations as 12 in TReMo+, we observe the write latency for ternary memory with serial and parallel modes that are 122.965 and 122.960 ns, respectively, as shown in **Table 2**. Thus, our ternary memory projected to have lower write latency than the MLC version of the prototype chip as expected within an acceptable error rate.

TABLE 1 | The effect of single verification and multiple verification on latency and energy of crossbar and SLC 1T1R architecture. Item number 1 does not have the verification method after writing, whereas item number 2 has only one iteration of verification. Last, item number 3 has verification with five times iteration. The reason behind different numbers of iteration is due to different underlying physics.

Num	Cell Type	Verification	Avg No. Itr	Write Latency (ns)	Write Energy (nJ)
1	SLC Crossbar	N	0	14.236	3.391
2	SLC Crossbar	Y	1	19.389	3.493
3	SLC 1T1R	N	0	12.256	1.143
4	SLC 1T1R	Y	1	18.496	1.144
5	SLC 1T1R	Y	5	64.956	21.957

TABLE 2 | 1T1R SLC memory vs. ternary memory with serial and parallel read based on the first variant of write-verification scheme.

Cell Level	Read Method	Avg No. Itr	Verification	Total Area (mm^2)	Read Latency (ns)	Write Latency (ns)	Read Energy (nJ)	Reset Energy (nJ)	Write Energy (nJ)	Set Energy (nJ)
SLC 1T1R	Normal	5	Y	74.045	10.96	66.512	3.895	17.785	41.370	31.218
Ternary 1T1R	Serial	12	Y	37.374	7.713	122.965	1.438	8.383	39.860	34.275
Ternary 1T1R	Parallel	12	Y	37.489	5.234	122.960	1.470	8.398	39.864	34.281

6 CASE STUDIES

6.1 Write-Verification Parameter Settings Trade-offs

The work described in this subsection models the first variant of the write-verification method, which is explained in **Section 4.2.3**, and investigates the trade-off, which is explained in **Section 2.4**. The write-verification setting determines the write energy and write latency.

In this case study, we examine how the selection of the write-verification parameter affects the delay and the energy consumption of the realized memory in relation to its endurance and reliability properties. For the evaluation, the results from Pérez et al. (2017) are used as reference for our simulated data with TReMo+. The paper contains measurement data concerning the average number of programming iterations, the set voltage, and the voltage step acquired from various experiments on real ReRAM devices programmed using ISPVA. These devices were made by IHP[2]. The known device configuration from Pérez et al. (2017) served as inputs to our simulation tool. The write latency, the write energy, and the set energy at the chip level were collected from the output of the simulator. As shown in **Table 3**, by incrementing the voltage step, both the write latency and the spent energies for set and resetting of the device decrease subsequently. As a result, this study shows how the advancement of the device level, e.g., a still sufficient lower iteration number, can actually affect the actual design of memories. Therefore, the conveyed idea is that, for the minimum write latency and energy, the voltage step should be high. However, this is not the ultimate consideration because cell reliability and endurance after writing should also be examined, and these features can be negatively affected by large voltage steps.

TABLE 3 | The impact of ISPVA settings on latency and energy of the total memory.

V_{Step}	Avg No. Itr	Write Latency (ns)	E_{Write} (nJ)	E_{Set} (nJ)
0.05	14	280003.507	2334.976	1757.618
0.1	8	170003.507	1926.703	1349.345
0.2	5	110003.507	1624.092	1046.724
0.4	3	70003.507	1243.907	666.549

With regard to the experiment done by Pérez et al. (2017), two important results were presented: 1) On the basis of **Figure 10**, by incrementing the voltage step, the number of cells willing to be set within the expected current threshold (the current threshold is the current threshold condition for the set operation in ISPVA) will decrease from ~80%–90% to ~60%. In other words, cells will be set with only one current peak when the voltage step is low, whereas in the opposite case when the voltage step is high, two current peaks appear (**Figure 10**). Quantization of the conduction is inherent to the CF, and therefore, it is always there. However, this behavior of the memory cell is due to the increase of the voltage step, and the overstress on the sample makes the conduction "jumping" to the next level of quantization, which means to a conduction level coherent with two CFs as it was observed and found by Pérez et al. (2017). It was demonstrated by Pérez et al. (2017) that, in lower-voltage steps, only one CF forms in the cell, whereas in higher-voltage steps, two separate CFs are formed or in other words the device is overset. 2) The carried-out cycling experiment on programmed cells with various voltage steps shows that the cells that are set with lower-voltage steps

[2]Innovations for High Performance Microelectronics.

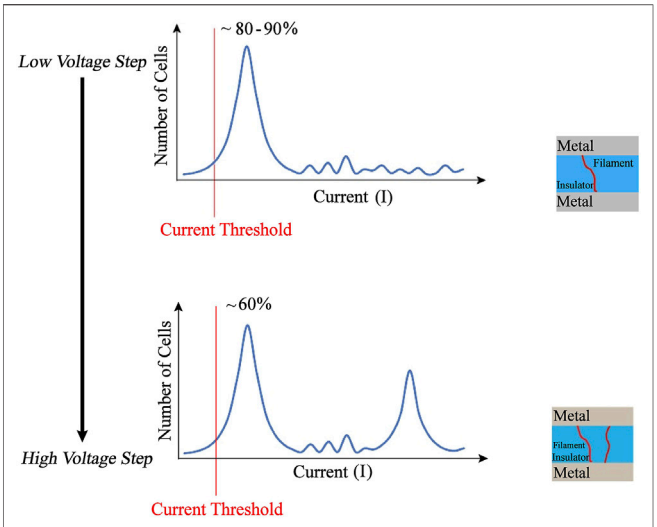


FIGURE 10 | Current distribution for set operation with incremental voltage step. In the top curve, a lower incremental step is used as in the bottom. This figure is depicted based on **Figure 6** in the work of Pérez et al. (2017).

tend to be more stable than those with higher-voltage steps, making them only partially stable. The reason behind this instability is that, in higher voltages, two filaments are involved (or overset behavior) in the process of switching, making the reliability fragile (Pérez et al., 2017).

It can be concluded that, for writing with the write-verification method (in this experiment, for the ISPVA method), there should be a trade-off when choosing an appropriate voltage step. The voltage step should not be too large to jeopardize the stability and, at the same time, should not be too low to increase the cells energy consumption.

6.2 Programming Pulse Width Assessment Trade-off

The work presented in this subsection models the second variant of the write-verification method, which is explained in **Section 2.4**. In this case study, we first verify that the results from TReMo+ correspond to the data at the cell level from IHP, and then, we examine how the programming of different pulse widths at the cell level affects the write energy and write latency at the chip level.

For this study, the results at the cell, such as the average iteration number for set and reset operation and the reset and set voltage for different pulse widths at the cell level, are extracted from the work of Perez et al. (2020). Besides, those data at the cell level and the IHP device configuration, such as 4 Kbit, read pulse width, and HRS and LRS values, given by Perez et al. (2020), are used as input to the simulation tool. As a result, the energy and latency at the cell and chip levels are collected from the output of the simulator.

For the first assessment, five different pulse widths—50 ns, 100 ns, 500 ns, 1 µs, and 10 µs—for both reset and set in ISPVA operation were utilized. **Figure 11** depicts the trend of energy at the cell level for set (S_E_Cell), reset (RS_E_Cell), and read energy (RD_E_Cell_for_Rs, RD_E_Cell_for_Rs). Furthermore, we show in **Figure 11** the read energy on the chip level for read (RD_E_total), reset (RS_E_total), and set (S_E_total). The data from TReMo+ at the cell match with the data at the cell level from IHP (Perez et al., 2020). Read energy at the cell and chip levels for set and reset operation is independent of the set and reset pulse width. However, reset and set energy at the cell and chip levels are increasing by the growth of pulse width. It is also evident that the reset energy is higher than the set energy both at the cell level and the chip level. It is validated that TReMo+'s result matches that of IHP's at the cell level. In addition, TReMo+ also estimates the write latency and write energy at the chip level.

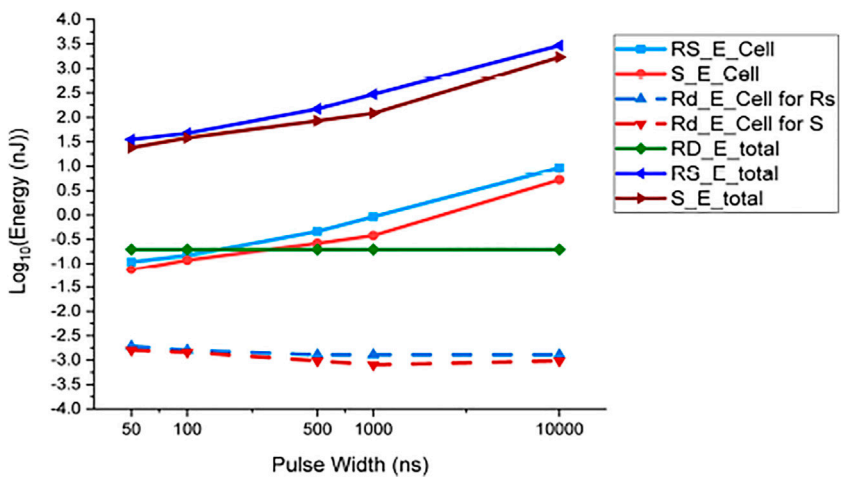


FIGURE 11 | Average energy required to do reset (blue dots) and set (red dots) operations, read energy for set and reset, and total read on a single 1T1R ReRAM cell and 4 Kbit ReRAm memory arrays.

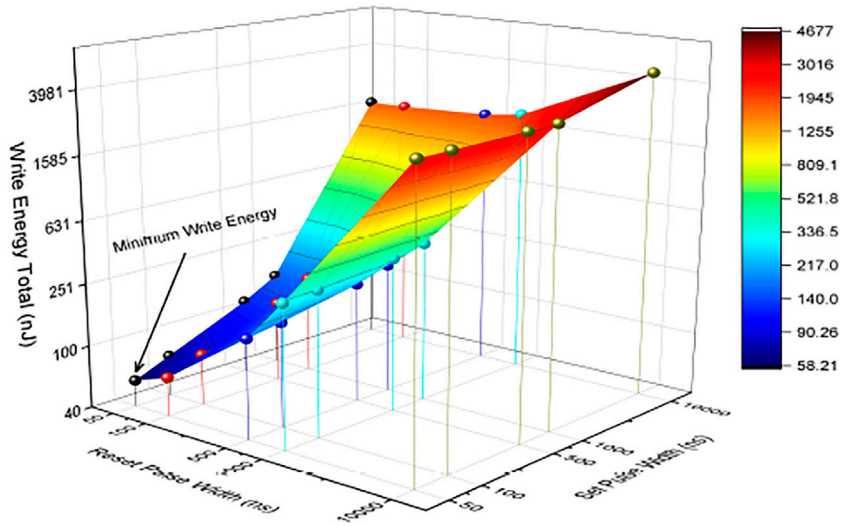


FIGURE 12 | Write energy over different pulse widths.

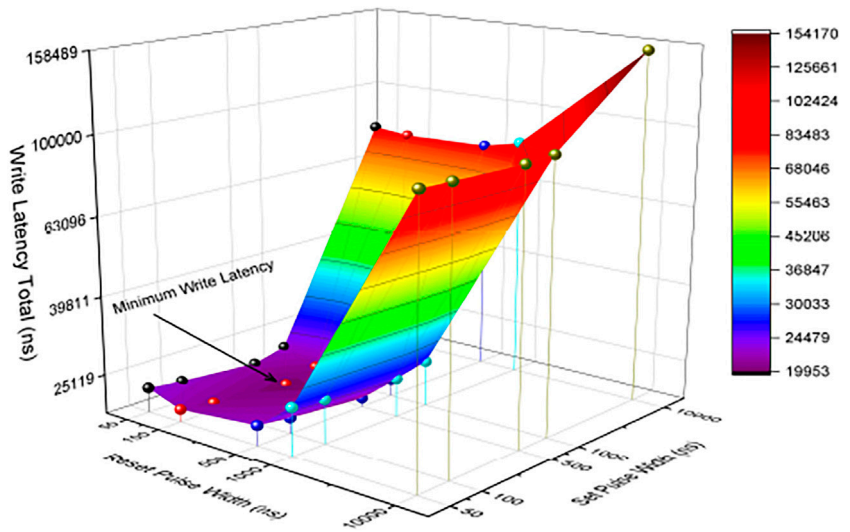


FIGURE 13 | Write latency over different pulse widths.

TABLE 4 | Cost per trit.

Cell Level	Total Area (mm^2)	Total No. Cells	Area Cost Per Trit (μm^2)
Parallel Ternary	3.294	4194304	0.7846
SLC	6.491	8388608	0.7738

In the second assessment, TReMo+ was executed using different ordered pairs of the reset and set pulse widths. The ordered pairs are consist of the total combination of 50 ns, 100 ns, 500 ns, 1 µs, and 10 µs for set and 50 ns, 100 ns, 500 ns, 1 µs, and 10 µs for reset, making 25 cell configurations. These are used to evaluate the effect of different pulse widths on the energy and latency at the chip level and the best points in terms of the write energy and write

latency. Needless to say, TReMo+ is capable of evaluating any pulse width given as an input. Therefore, there is no limitation to use our tool for any pulse width. Furthermore, for this case study, we utilized the experimental data at the cell level from the IHP company available in the work of Perez et al. (2020). Because of some limitation in producing a pulse width smaller than 50 ns for their experiments, they did not assess pulse width smaller than 50 ns in their analysis.

As it is depicted in **Figure 12**, the best point from write energy perspective belongs to 50 ns for reset and set pulse width. However, the lowest write latency belongs to 100 ns for reset pulse width and 500 ns for set pulse width, as depicted in **Figure 13**. Furthermore, as it can be seen in **Figure 12**, when increasing the set pulse width while keeping the reset pulse width fixed, the write latency will grow in each iteration. However, in this situation, the write energy will fall up to the third point but then starts to increase from the fourth point onward as depicted in **Figure 12**. As a result, the lowest write energy point among every five points is the third one. This shows the obvious trade-off between write latency and energy latency.

On the basis of the retention and the reliability test in the work of Perez et al. (2020), there is no reliability issue for different combinations of reset and pulse width, except for that of 50 and 50 ns for reset and set pulse width, due to longer ending tail in **Figure 3** in the work of Perez et al. (2020). That means, on the basis of the experimental results, although 50 ns for both reset and set pulse width shows the best write energy, 100 ns for reset and 50 ns for set pulse width with the second minimum write energy seems to be the best point for programming the cell with ISPVA with no reliability issue. Having discussed this, still a trade-off would exist to determine which programming pulse width ensures the lowest energy and the most reliable operation.

6.3 Area Cost per Trit

Cost per bit is one of the most important aspects when modeling a novel memory technology. Some memory design goals, such as technology scaling, chip yield enhancement, and cell structure modernization, all point toward reducing cost per bit of a memory chip.

When adapting the MLC memory for ternary memory design, the issue of area arises in a sense that is based on **Section 4.1**. Ternary memory does not require any decoders for the reading operation, whereas MLC needs at least seven sets of SA and an extra decoder (Xu et al., 2013). The results in **Table 4** also prove that the area per trit in the ternary memory is the most optimal case.

According to Xu et al. (2014), to calculate cost per trit, area and fabrication costs are the most important factors. On the basis of the above explanation and assuming that fabrication costs in MLC and ternary memory are the same, a lower cost per trit in comparison to the MLC counterpart is given because the ternary memory has a smaller area. The experimental results shown in **Table 4** are based on the same settings utilized in **Section 6.2** but for 1-Mb memory chip capacity. The number of cells calculated in **Table 4** is total number of cells of simulated complete array for the given setting.

7 CONCLUSION

In this paper, a new memristor-based ternary memory model was modeled that benefits from optimized reading and writing methods. Alongside the serial read method, the parallel read for the ternary memory model was modeled for the first time, which made the read latency lower than its rival and, at the same time, kept the overhead to a minimum. The writing method of choice in this paper was the write-verification method, which offered more reliable writing operation, compared with the single-pulse method.

Moreover, some case studies were presented for proving the usefulness and versatility of the model, including parameter selection for write-verification method and their ramifications on energy and latency, programming pulse width assessment and its trade-off in energy and latency, and a study on area cost per trit proving that the ternary case offers the most optimal solution in terms of area consumption.

Finally, to ease the process of ternary memory development by researchers and manufacturers, a comprehensive tool was developed that is capable of performing energy, performance, and area estimation for a given setting.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

Conceptualization: SH and MB; Implementation and experimental evaluation: SH; Investigation: SH; Methodology: SH; Visualization: SH; Writing - original draft: SH; Manuscript revision, review, and editing: SH, MB, and DF.

ACKNOWLEDGMENTS

Authors would like to thank Eduardo Pérez from IHP for his valuable feedback and suggestions.

REFERENCES

Avizienis, A. (1961). "Signed-digit Numbe Representations for Fast Parallel Arithmetic," in IEEE Trans. Electron. Comput. (IEEE), 389–400. doi:10.1109/tec.1961.5219227EC-10

Biglari, M., and Fey, D. (2017). "Memristive Voltage Divider: A Bipolar Reram-Based Unit for Non-volatile Flip-Flops," in Proceedings of the International Symposium on Memory Systems, 217–222. doi:10.1145/3132402.3132432

Biglari, M., Lieske, T., and Fey, D. (2018). "High-endurance Bipolar Reram-Based Non-volatile Flip-Flops with Run-Time Tunable Resistive States," in Proceedings of the 14th IEEE/ACM International Symposium on Nanoscale Architectures, 19–24. doi:10.1145/3232195.3232217

Biglari, M., Lieske, T., and Fey, D. (2019). Reducing Hibernation Energy and Degradation in Bipolar ReRAM-Based Non-volatile Processors. *IEEE Trans. Nanotechnology* 18, 657–669. doi:10.1109/tnano.2019.2922363

Binkert, N., Beckmann, B., Black, G., Reinhardt, S. K., Saidi, A., Basu, A., et al. (2011). The Gem5 Simulator. *SIGARCH Comput. Archit. News* 39, 1–7. doi:10.1145/2024716.2024718

El-Slehdar, A. A., Fouad, A. H., and Radwan, A. G. (2013). "Memristor-based Balanced Ternary Adder," in 2013 25th International Conference on Microelectronics (ICM), 1–4. doi:10.1109/icm.2013.6735002

Perez, E., González Ossorio, Ó., Dueñas, S., García, H., and Wenger, C., "Programming Pulse Width Assessment for Reliable and Low-Energy Endurance Performance in Al: Hfo2-Based Rram Arrays," *Electronics*, 2020.

Fey, D., Reichenbach, M., Söll, C., Biglari, M., Röber, J., and Weigel, R. (2016). "Using Memristor Technology for Multi-Value Registers in Signed-Digit Arithmetic Circuits," in Proceedings of the Second International Symposium on Memory Systems, 442–454. doi:10.1145/2989081.2989124

Fey, D. (2015). Ternary Arithmetic Pipeline Architectures Using Multi-Bit Memristors. *Future Comput.*, 1–6.

Fey, D. (2014). Using the Multi-Bit Feature of Memristors for Register Files in Signed-Digit Arithmetic Units. *Semicond. Sci. Technol.* 29, 104008. doi:10.1088/0268-1242/29/10/104008

Grossi, A., Zambelli, C., Olivo, P., Miranda, E., Stikanov, V., Walczyk, C., et al. (2016). Electrical Characterization and Modeling of Pulse-Based Forming Techniques in RRAM Arrays. *Solid-State Electro.* 115, 17–25. doi:10.1016/j.sse.2015.10.003

Higuchi, K., Iwasaki, T., and Takeuchi, K. (2012). "Investigation of Verify-Programming Methods to Achieve 10 Million Cycles for 50nm Hfo2 Reram," in 2012 4th IEEE International Memory Workshop, 1–4. doi:10.1109/imw.2012.6213665

Hosseinzadeh, S., Biglari, M., and Fey, D. (2020). "Tremo: A Model for Ternary Reram-Based Memories with Adjustable Write-Verification Capabilities," in 2020 23rd Euromicro Conference on Digital System Design (DSD), 44–48.

Junsangsri, P., Lombardi, F., and Han, J. (2014). "A Memristor-Based Tcam (Ternary Content Addressable Memory) Cell," in 2014 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH), 1–6. doi:10.1109/nanoarch.2014.6880478

Kinoshita, K., Tsunoda, K., Sato, Y., Noshiro, H., Yamazaki, Y., Fukano, T., et al. (2007). "Reduction of Reset Current in Nio-Reram Brought about by Ideal Current Limiter," in 2007 22nd IEEE Non-Volatile Semiconductor Memory Workshop, 66–67. doi:10.1109/nvsmw.2007.4290583

Lee, A., Lo, C.-P., Lin, C.-C., Chen, W.-H., Hsu, K.-H., Wang, Z., et al. (2017). A ReRAM-Based Nonvolatile Flip-Flop with Self-Write-Termination Scheme for Frequent-OFF Fast-Wake-Up Nonvolatile Processors. *IEEE J. Solid-state Circuits* 52, 2194–2207. doi:10.1109/jssc.2017.2700788

Lieske, T., Biglari, M., and Fey, D. (2018). "Multi-level Memristive Voltage Divider: Programming Scheme Trade-Offs," in Proceedings of the International Symposium on Memory Systems, 259–268. doi:10.1145/3240302.3240430

Linn, E., Rosezin, R., Kügeler, C., and Waser, R. (2010). Complementary Resistive Switches for Passive Nanocrossbar Memories. *Nat. Mater* 9, 403–406. doi:10.1038/nmat2748

Metze, G., and Robertson, J. E. (1959). *IFIP Congress*.Elimination of Carry Propagation in Digital Computers.

Mittal, S., Wang, R., and Vetter, J. (2017). Destiny: A Comprehensive Tool with 3d and Multi-Level Cell Memory Modeling Capability. *JLPEA* 23.

Parhami, B. "Carry-free Addition of Recoded Binary Signed-Digit Numbers," in IEEE Transactions on Computers (IEEE), 1470–1476. doi:10.1109/12.872211

Pérez, E., Grossi, A., Zambelli, C., Mahadevaiah, M. K., Olivo, P., Wenger, Ch., et al. (2018). "Temperature Impact and Programming Algorithm for Rram Based Memories," in IMWS-AMP, 1–3. doi:10.1109/IMWS-AMP.2018.8457132

Pérez, E., Grossi, A., Zambelli, C., Olivo, P., and Wenger, C. (2017). "Impact of the Incremental Programming Algorithm on the Filament Conduction in Hfo2-Based Rram Arrays," in IEEE Journal of the Electron Devices Society, 64–68. doi:10.1109/jeds.2016.2618425

Pérez, E., Zambelli, C., Kalishettyhalli Mahadevaiah, M., Olivo, P., and Wenger, C. (2019). "Toward Reliable Multi-Level Operation in Rram Arrays: Improving post-algorithm Stability and Assessing Endurance/data Retention," in IEEE Journal of the Electron Devices Society, 740–747.

Puglisi, F. M., Wenger, C., and Pavan, P. (2015). A Novel Program-Verify Algorithm for Multi-Bit Operation in Hfo2 Rram. *IEEE Electron. Device Lett.* 36 (10), 1030–1032. doi:10.1109/led.2015.2464256

Rath, S. S. (1975). A Ternary Flip-Flop Circuit. *Int. J. Electro.* 38 (1), 41–47. doi:10.1080/00207217508920376

Sheu, S., Chang, M., Lin, K-F., Wu, C-W., Chen, Y-S., Chiu, P-F., et al. (2011). "A 4mb Embedded Slc Resistive-Ram Macro with 7.2ns Read-Write Random-Access Time and 160ns Mlc-Access Capability," in 2011 IEEE International Solid-State Circuits Conference, 200–202. doi:10.1109/isscc.2011.5746281

Song, Y. L., Meng, Y., Xue, X. Y., Xiao, F. J., Liu, Y., Chen, B., et al. (2013). "Reliability Significant Improvement of Resistive Switching Memory by Dynamic Self-Adaptive Write Method," in VLSIT (IEEE), T102–T103.

Thoziyoor, N. M. S., and Ahn, J. H. (2008). Cacti 5.1 Technical Report. *Tech. Rep*. HPL-2008-20.

Wilton, S. J. E., and Jouppi, N. P. (1996). CACTI: an Enhanced Cache Access and Cycle Time Model. *IEEE J. Solid-state Circuits* 31, 677–688. doi:10.1109/4.509850

Xiangyu Dong, X., Yuan Xie, C., and Jouppi, N. P. (2012). Nvsim: A Circuit-Level Performance, Energy, and Area Model for Emerging Nonvolatile Memory. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* 31, 994–1007. doi:10.1109/tcad.2012.2185930

Xu, C., Niu, D., Muralimanohar, N., Jouppi, N. P., and Xie, Y. (2013). "Understanding the Trade-Offs in Multi-Level Cell Reram Memory Design," in Proceedings of the 50th Annual Design Automation Conference on - DAC, 1–6.

Xu, C., Niu, D., Yu, S., and Xie, Y. (2014). "Modeling and Design Analysis of 3d Vertical Resistive Memory — a Low Cost Cross-point Architecture," in 2014 19th Asia and South Pacific Design Automation Conference (ASP-DAC), 825–830. doi:10.1109/aspdac.2014.6742992

Yang, J. J., Pickett, M. D., Li, X., Ohlberg, D. A. A., Stewart, D. R., and Williams, R. S. (2008). Memristive Switching Mechanism for Metal/oxide/metal Nanodevices. *Nat. Nanotech* 3, 429–433. doi:10.1038/nnano.2008.160

Yonekawa, H., Sato, S., and Nakahara, H. (2018). "A Ternary Weight Binary Input Convolutional Neural Network: Realization on the Embedded Processor," in IEEE 48th International Symposium on Multiple-Valued Logic (ISMVL) (IEEE), 174–179. doi:10.1109/ismvl.2018.00038

Yu, S., Guan, X., and Wong, H.-S. P. (2012). On the Switching Parameter Variation of Metal Oxide RRAM-Part II: Model Corroboration and Device Design Strategy. *IEEE Trans. Electron. Devices* 59, 1183–1188. doi:10.1109/ted.2012.2184544

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership