frontiers in Electronics

MINI REVIEW
published: 25 March 2022
doi: 10.3389/felec.2022.825077

Exploiting Non-idealities of Resistive Switching Memories for Efficient Machine Learning

Victor Yon[1,2]*, *Amirali Amirsoleimani*[3], *Fabien Alibart*[2,4], *Roger G. Melko*[5,6], *Dominique Drouin*[1,2] *and Yann Beilliard*[1,2]

[1]Institut Interdisciplinaire d'Innovation Technologique (3IT), Université de Sherbrooke, Sherbrooke, QC, Canada, [2]Laboratoire Nanotechnologies Nanosystèmes (LN2)—CNRS 3463, Université de Sherbrooke, Sherbrooke, QC, Canada, [3]Department of Electrical Engineering and Computer Science, Lassonde School of Engineering, York University, Toronto, ON, Canada, [4]Institute of Electronics, Microelectronics and Nanotechnology (IEMN), Villeneuve d'Ascq, France, [5]Department of Physics and Astronomy, University of Waterloo, Waterloo, ON, Canada, [6]Perimeter Institute for Theoretical Physics, Waterloo, ON, Canada

Novel computing architectures based on resistive switching memories (also known as memristors or RRAMs) have been shown to be promising approaches for tackling the energy inefficiency of deep learning and spiking neural networks. However, resistive switch technology is immature and suffers from numerous imperfections, which are often considered limitations on implementations of artificial neural networks. Nevertheless, a reasonable amount of variability can be harnessed to implement efficient probabilistic or approximate computing. This approach turns out to improve robustness, decrease overfitting and reduce energy consumption for specific applications, such as Bayesian and spiking neural networks. Thus, certain non-idealities could become opportunities if we adapt machine learning methods to the intrinsic characteristics of resistive switching memories. In this short review, we introduce some key considerations for circuit design and the most common non-idealities. We illustrate the possible benefits of stochasticity and compression with examples of well-established software methods. We then present an overview of recent neural network implementations that exploit the imperfections of resistive switching memory, and discuss the potential and limitations of these approaches.

Keywords: resistive switching memories, memristor, in-memory computing, hardware non-idealities, artificial neural networks, bayesian neural networks, probabilistic computing

OPEN ACCESS

Edited by:
Aida Todri-Sanial, de Robotique et de Microélectronique de Montpellier (LIRMM), France

Reviewed by:
Zongwei Wang, Peking University, China

***Correspondence:**
Victor Yon
victor.yon@usherbrooke.ca

Specialty section:
This article was submitted to Nano- and Microelectronics, a section of the journal Frontiers in Electronics

Received: *08 December 2021*
Accepted: *07 March 2022*
Published: *25 March 2022*

Citation:
Yon V, Amirsoleimani A, Alibart F, Melko RG, Drouin D and Beilliard Y (2022) Exploiting Non-idealities of Resistive Switching Memories for Efficient Machine Learning. Front. Electron. 3:825077. doi: 10.3389/felec.2022.825077

INTRODUCTION

The recent success of machine learning has essentially arisen from breakthroughs in learning algorithms, data availability, and computing resources. The latter benefited from the consistent increase in the number of transistors in complementary metal-oxide-semiconductor (CMOS) microchips and the development of highly parallel and specialized hardware (Reuther et al., 2019) such as graphics processing units (GPUs) (Keckler et al., 2011) and tensor processing units (TPUs) (Jouppi et al., 2017). Nevertheless, even with such optimizations, the traditional computing architecture is inappropriate for the implementation of modern machine learning algorithms due to the intensive data transfer requirement between memory and processing units (Horowitz, 2014; Ankit et al., 2017; Sze et al., 2017). Novel architectures based on resistive switching memories (RSMs) can take advantage of non-volatile in-memory computing (Mutlu et al., 2019) to efficiently perform vector-matrix multiplications (VMMs) (Gu et al., 2015; Amirsoleimani et al.,

2020), the most critical operation of neural networks (NNs) inference. This technology is also an excellent candidate for the implementation of membrane potential and activation functions for brain-inspired spiking neural networks (SNNs) (Xia and Yang, 2019; Wang et al., 2020a; Yang et al., 2020a; Agrawal et al., 2021). Therefore, this emerging technology offers an opportunity to tackle current limitations of traditional computing hardware, such as energy efficiency, computation speed, and integration footprint (Chen, 2020; Marković et al., 2020; Christensen et al., 2021).

However, in the context of in-memory computing, RSM technology is still in its infancy compared to the much more mature CMOS technology at the heart of traditional von Neumann computers. RSMs are subject to variability and performance issues (Adam et al., 2018; Wang et al., 2019a; Krestinskaya et al., 2019; Chakraborty et al., 2020; Zahoor et al., 2020; Zhao et al., 2020; Xi et al., 2021), which currently restrict their usage to small and noisy NN hardware implementations that can solve only simple problems such as classification of the MNIST digit database (Ambrogio et al., 2018; Hu et al., 2018; Li et al., 2018; Wang et al., 2019b; Lin et al., 2019; Liu et al., 2020b; Yao et al., 2020; Zahari et al., 2020).

Since a wide variety of resistive memory technologies are still at an early stage of development (Zahoor et al., 2020; Christensen et al., 2021), it may be that a better understanding of resistive switching mechanisms combined with improvements in fabrication processes in the future will rectify some of their non-idealities (e.g., device-to-device variability and programming nonlinearity). But, even with technological maturity, other imperfections may remain due to their inseparable relationship with the physics of the device/circuit (e.g., programming variability and interconnect resistance). Numerous methods have therefore been proposed to mitigate the effect of these non-idealities for NN applications (Chen et al., 2015; Lim et al., 2018; He et al., 2019; Liu et al., 2020a; Wang et al., 2020b; Mahmoodi et al., 2020; Pan et al., 2020; Zhang et al., 2020; Xi et al., 2021). Although a mitigation approach can significantly increase the performance of RSM-based NNs, none of these methods are able to achieve the accuracy of their software counterparts. An alternative strategy consists in harnessing device imperfections rather than fighting them, which would enable highly efficient probabilistic and approximate computing hardware. This strategy is particularly appealing since software and biological NNs already take advantage of randomness to enhance information processing (McDonnell and Ward, 2011). The goal of this article is to review the latest progress in the exploitation of hardware imperfections by RSM-based NNs.

PRIOR DESIGN CHOICES

Design choices can influence the overall performance of RSM-based systems and induce different types and amounts of non-idealities. Starting with the choice of resistive switching mechanisms (Sung et al., 2018; Xia and Yang, 2019; Zahoor et al., 2020; Christensen et al., 2021) (e.g., valence change, electrochemical metallization, phase-change, ferroelectricity, magnetoresistivity), which could all be implemented with numerous materials and fabrication processes. For example, magnetoresistive memories are known to be especially durable [$> 10^{14}$ cycles (Kan et al., 2016)], whereas ferroelectric memories benefit from highly linear resistance programming (Tian et al., 2020), and valence change memories usually offer a high on/off ratio [$> 10^{7}$ (Chen et al., 2017)]. Despite these many options, an ideal RSM that combines all of these desirable properties has yet to be discovered. Until this is found, the design of an RSM-based computing system will always involve a trade-off between non-idealities (Siegel et al., 2020).

The design of the circuit architecture is also a critical step of the system conception. RSMs are typically organized in the form of an array when they represent the synaptic weights of a NN. Therefore, the size of this array limits the dimension of the VMM operation that can be computed in a single step. Large-scale implementations have been reported (Mochida et al., 2018; Ishii et al., 2019); for example, the system designed by Ambrogio et al. (2018) contained 524 k cells and could implement 204,900 synapses of a fully connected 4-layer NN. There are at least three strategies for selecting each of these cells (Chen and Yu, 2015; Wang et al., 2019a): 1) active selectors, 1-Transistor-n-Resistor (1TnR), in which one transistor allows access to one or a group of RSMs; 2) passive selectors, 1-Selector-1-Resistor (1S1R), which is usually a diode vertically stacked with the resistive material; 3) and no selector, 1-Resistor (1R).

Finally, the peripheral circuit used to address and control each RSM must not be neglected, since it will constrain system performance and impact the speed and energy efficiency (Li et al., 2016; Kadetotad et al., 2017; Amirsoleimani et al., 2020; Li and Ang, 2020). In particular, the limited resolution of the digital-to-analog converter (DAC) and the analog-to-digital converter (ADC) will restrict the input/output precision and produce a significant energy overhead. This external circuit is also used to send writing pulses using a specific programming scheme that affects the performance (Woo et al., 2016; Stathopoulos et al., 2017; Chen et al., 2020). A *write-verify* scheme (Papandreou et al., 2011; Yi et al., 2011; Alibart et al., 2012; Perez et al., 2017; Pan et al., 2020) typically neglects energy consumption and memory lifetime for the sake of accuracy.

NON-IDEAL CHARACTERISTICS

Device Level

Modern fabricated RSM-based systems are imperfect at many levels (**Figure 1**). At the device level, one of the most critical characteristics for NN applications is the resistance range (Yu, 2018), since this property constrains the possible values of the model's parameters. This metric is often expressed as the ratio between the minimal and maximal resistance values. Moreover, by considering the smallest writing value addressable with the peripheral circuit, we can estimate the total number of intermediate resistance states in the memory. This number can vary from 2 states for binary RSMs (Wong et al., 2012; Bocquet et al., 2018; Zahari et al., 2020) to more than 128 for the most

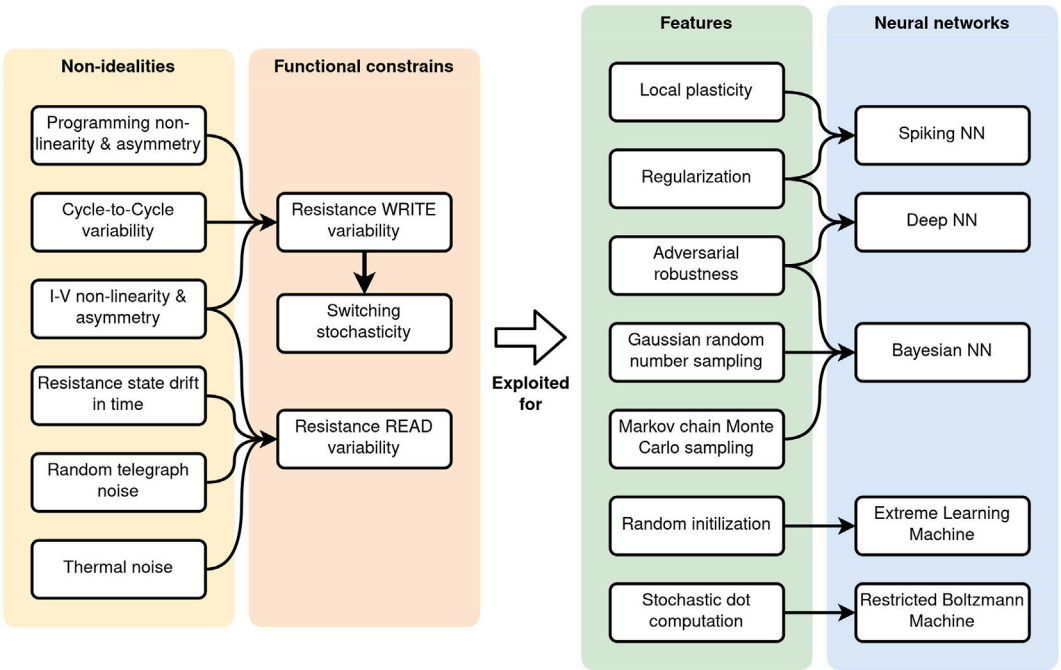

FIGURE 1 | Device-level non-idealities influence on RSM functional constraints and their possible usage for different NN architectures. The list of non-idealities and functional constraints given here is limited to those reported in the literature for these use cases, thus the list is not exhaustive.

accurate devices (Gao et al., 2015; Li et al., 2017; Ambrogio et al., 2018; Wu et al., 2018). However, to estimate the quantity of information that a RSM can store, only reproducible and distinguishable states should be considered, which reduces this number to around 64-level (equivalent to 6-bits) in the best case (Li et al., 2017; Stathopoulos et al., 2017).

Thus, even with the best control circuit, it appears to be impossible to program a resistance state under an arbitrary precision threshold, usually between 1–5% of the resistance range (Adam et al., 2018; Xia and Yang, 2019; Xi et al., 2021). This writing variability is most likely attributable to the local environment [room temperature (Abunahla et al., 2016; Bunnam et al., 2020; Roldán et al., 2021), humidity (Messerschmitt et al., 2015; Valov and Tsuruoka, 2018)] and the internal state of the RSM at the atomic scale. In valence change memories, for example, the conductive filament may break abruptly and result in a state that is more resistive than expected (Gao et al., 2009; González-Cordero et al., 2017; Wiefels et al., 2020). The reading process is also affected by the variability, and several phenomena such as random telegraph noise (Ielmini et al., 2010; Lee et al., 2011; Veksler et al., 2013; Claeys et al., 2016) and thermal noise (Bae and Yoon, 2020) can disturb the measured resistance value and lead to inaccurate outputs.

The dynamics of the resistance programming can also pose a challenge for NN training (Sidler et al., 2016; Woo and Yu, 2018). The same writing pulse can lead to a different outcome depending on the current resistance value (programming nonlinearity) (Jacobs-Gedrim et al., 2017) and the update direction (programming asymmetry). However, if the behavior of the RSM is well characterized, its programming dynamics can be partially anticipated and taken into account during the NN training (Chang et al., 2018; Lim et al., 2018; Pan et al., 2020).

Finally, a RSM is never perfectly stable over time, data retention can vary from a few seconds (Oh et al., 2019) to more than 10 years (Wei et al., 2008), depending on the resistive technology and the local environment (Gao et al., 2011; Subhechha et al., 2016; Kang et al., 2017; Zhao et al., 2019). Moreover, the number of writing operations is also limited [up to 10^{10} (Yang et al., 2010)], each change in resistance will slightly degrade the characteristics of the RSM, which will decay until breakdown is eventually reached.

Array Level

To efficiently compute VMM, RSMs must be arranged in large arrays surrounded by CMOS-based control electronics. Although this structure takes advantage of the two terminals of this device to maximize the integration density and offer a good scaling perspective, this integration (whether in two or three dimensions) faces technical issues (Li and Ang, 2020), including the inevitable resistance of the interconnections (Mahmoodi et al., 2020). Hence, even if two RSMs have the same internal state, the resistance of the metallic lines will affect the total resistance depending on the RSMs position in the crossbar. Furthermore, in 1R arrays (and to a lesser extent in 1S1R arrays), the sneak path current could also be a concern (Cassuto et al., 2013; Chen et al., 2021b). This phenomenon occurs when the electric current follows an unexpected path, leading to corruption of the final

output. These two technical challenges motivated the development of tile-based architectures (Shafiee et al., 2016; Nag et al., 2018), in which a crossbar is split into several smaller ones.

The fabrication processes of RSMs typically induce significant variability between devices, meaning that each characteristic identified in **Section 3.1** may differ for each RSM in a given crossbar. In extreme cases, the resistive state may even be stuck at its maximal or minimal value. The current best fabrication techniques can approach a yield of 99% (Li et al., 2017; Ambrogio et al., 2018). Although variability and faults will strongly impact the performance of a NN in the case of *ex-situ* (offline) training (Boquet et al., 2021), a NN trained *in-situ* (online) can mitigate these imperfections to a certain extent (Alibart et al., 2013; Li et al., 2018; Wang et al., 2019b; Romero et al., 2019).

Writing and reading operations in a crossbar can also induce unwanted disturbances in the resistance state of the RSMs (Yan et al., 2017; Wang et al., 2018; Amirsoleimani et al., 2021). The writing operation is usually achieved by applying a tension V to the device we want to program, where V is the threshold voltage required to change the resistance value. However, the surrounding devices that share the bottom or top electrode with the target receive a tension $V/2$, which can alter the RSM state after a large number of these operations, even though the tension is below the theoretical writing threshold. The same side effect can be observed for the read operation if the reading tension is not low enough to guarantee no disturbance.

The device and array level non-idealities listed in this section impact the overall quality of the NNs. The training of these models typically relies on numerous parameter updates, computed by gradient descent, which is strongly affected by inconsistent writing operation. A NN trained on such imperfect hardware will either reach a suboptimal state or simply fail to converge to a solution. While the drift of resistance in time and the non-idealities that damage the read operation (**Figure 1**) will lead to inaccuracy during the inference.

TAKING ADVANTAGE OF NON-IDEALITIES

Software Methods

With 32-bits floating-point variables, noiseless computation and very large-scale integration, there are solid arguments for using traditional computers for machine learning computation. Nevertheless, this impressive accuracy becomes a curse when the parameters of the model are so numerous and precise that they are able to extract the residual variation of the training data. This overfitting issue is a real challenge for modern NNs, but fortunately, many solutions now exist to mitigate this problem. Surprisingly, the most common methods are similar to what we would consider non-idealities in the field of RSMs.

The most popular approaches are probably dropout (Hinton et al., 2012) and drop connect (Wan et al., 2013). These regularization techniques consist in randomly omitting a subset of activation units or weights during NNs training. This turns out to be very effective to prevent complex co-adaptations of the units that usually lead to overfitting. Another counterintuitive but efficient strategy is to purposely add noise at different stages of the training process (An, 1996; McDonnell and Ward, 2011; Qin and Vucinic, 2018), in particular to the gradient value (Neelakantan et al., 2015), activation functions (Gulcehre et al., 2016), model parameters (He et al., 2019), and layers input (Liu et al., 2017; Creswell et al., 2018; Rakin et al., 2018). The injection of an appropriate level of noise can improve generalization, reduce training losses, and increase the robustness of these models against adversarial attacks.

The success of quantization techniques (Guo, 2018; Mishra et al., 2020; Chen et al., 2021a) also indicate that very accurate parameters are not mandatory to implement reliable NNs. Some works have achieved compression from 32 to 16-bits without sacrificing the final accuracy (Gupta et al., 2015; Micikevicius et al., 2017), and down to 8 or 3-bits with acceptable performance loss depending on the targeted application (Holt and Baker, 1991; Anwar et al., 2015; Shin et al., 2015). This idea can be extended to binarized parameters (Courbariaux et al., 2015) and activation functions (Zhou et al., 2016), which drastically reduce the computational cost although with a significant loss of accuracy.

Although regularization methods based on stochastic processes are now well-established in the machine learning community, their implementation on von Neumann computers is very inefficient, due to the serial and deterministic nature of the hardware. In particular, the high power consumption of random number generation (Cai et al., 2018; Gross and Gaudet, 2019) with standard CMOS technologies calls for the development of novel hardware that natively implements stochasticity. In the next section, we will see that RSMs have attracted much attention over recent years in this regard.

RSM-Based Methods

RSM-based NNs have been successfully used on hardware to solve simple data-driven problems. The state-of-the-art 1T1R array can classify MNIST digits with >96% accuracy (Ambrogio et al., 2018; Yao et al., 2020). Those high scores are often made possible by mitigating the negative impact of non-idealities with specific weight mapping schemes and learning strategies (Chen et al., 2015; Wu et al., 2017; Gong et al., 2018; Cai et al., 2020b; Wang et al., 2020b; Pan et al., 2020; Zanotti et al., 2021), such as committee machines (Joksas et al., 2020) or the *write-verify* update loop procedure. This mitigating approach is promising, but does not seem to be sufficient to fill the accuracy gap between the state-of-art RSM-based NNs and their software counterparts [99.7% accuracy (Ciresan et al., 2012; Mazzia et al., 2021) for the same task]. A different strategy is therefore required to overcome this limitation, one possibility is to accept the imperfections of the hardware and take advantage of them. This can be done by imitating software regularization methods or implementing NNs that rely on stochastic mechanisms (**Figure 1**).

TABLE 1 | Summary of reported RSM-based NNs that exploit non-idealities in an explicit way to improve the accuracy or efficiency of the model. *Sim.* and *Exp.* stand for *Simulation* and *Experimental*, Conductive-Bridging Random-Access Memory (CBRAM), Oxide-based Random-Access Memory (OxRAM), Magnetoresistive Random-Access Memory (MRAM).

Ref	Type of Network	Non-idealities Exploited	Benefits	Sim. Exp	RSM Type	Training Type	RSM Usage
Lin et al. (2019)	BNN	Random Telegraph Noise; Read variability	Energy-efficient stochastic sampling giving adversarial robustness and accuracy comparable to software implementation	Exp.	OxRAM	*Ex-situ*	Synapse
Yang et al. (2020b)	BNN	Thermal noise; Switching stochasticity	Energy-efficient Gaussian random number sampling	Sim.	MRAM	*Ex-situ*	Synapse
Malhotra et al. (2020)	BNN	Programming variability	Energy-efficient Gaussian random number sampling	Sim.	OxRAM	*Ex-situ*	Synapse
Dalgaty et al. (2021a)	BNN; Bayesian perceptrons	Programming variability	Energy-efficient implementation of in-memory Markov chain Monte Carlo sampling	Both	OxRAM	*In-situ*	Synapse
Dalgaty et al. (2021b)	BNN	Programming variability	Native representation of complex probability distribution that can be used to quantify model uncertainty	Exp.	OxRAM	*Ex-situ*	Synapse
Suri et al. (2015)	RBM	HRS and LRS programming variability	Native stochastic activation function	Sim.	OxRAM	*In-situ*	Synapse Neuron
Mahmoodi et al. (2019)	RBM	Thermal noise	Implementation of scalable, versatile, and efficient stochastic dot computation	Both	OxRAM	*In-situ*	Synapse
Yu et al. (2013)	Spiking NN	Switching stochasticity	Allows trading multi-state memory to stochastic binary memory with comparable performance	Sim.	OxRAM	*In-situ*	Synapse
Wijesinghe et al. (2018)	Spiking CNN	Switching stochasticity	Probabilistic activation function makes the network more robust to synaptic non-idealities	Sim.	CBRAM	*Ex-situ*	Synapse Neuron
Dalgaty et al. (2019)	Spiking Recurrent NN	Programming variability	Neuronal intrinsic local plasticity for low-power temporal data processing	Sim.	OxRAM	*In-situ*	Synapse Neuron
Bhattacharjee and Panda (2020)	Deep NN	Number of states; Programming variability; Interconnect resistance	Adversarial robustness 10–20% better than software baseline	Sim.	N/A	*Ex-situ*	Synapse
Wang et al. (2019b)	CNN; long short-term memory	Programming variability	Natural regularization, reduced difference between the training and the test accuracy	Exp.	OxRAM	*In-situ*	Synapse
Suri and Parmar (2015)	Extreme Learning Machine	HRS programming variability	Low-power and low-footprint implementation of the fixed hidden layer of an ELM	Sim.	CBRAM OxRAM	N/A	Synapse
Cai et al. (2020a)	Hopfield NN	Interconnect resistance; Programming variability; Finite ON/OFF ratio; I-V non-linearity	Improved convergence to optimal solution for combinatorial optimization problems	Both	OxRAM	*In-situ*	Synapse
Lin et al. (2018)	Generative Adversarial Network	Programming variability; Reading variability	Reasonable amount of noise increase the diversity of generated patterns	Both	OxRAM	*In-situ*	Synapse

Over the last decade, the unpredictable behavior of RSMs has been demonstrated to be an efficient way to create true random number generators (Shen et al., 2021; Gaba et al., 2013; Yang et al., 2020b; Hu et al., 2016; Faria et al., 2018; Bao et al., 2020) that can be used for security purposes (Khan et al., 2021; Pang et al., 2019; Lv et al., 2020). For machine learning applications, the Gaussian nature of the stochastic distribution turns out to be an efficient way to implement probabilistic computing in hardware (**Table 1**). Indeed, for multi-level RSMs, the programming error around the targeted conductance state can be used as a regularization method if the device encodes the synaptic weights. Wang et al. (2019b) showed that a standard deviation of 10 μS of Gaussian writing noise helped to avoid overfitting for the MNIST classification task using a RSM-based convolutional neural network (CNN). In the case of binary memories, a consistent stochastic switch can be triggered by a programming writing voltage that is below the device threshold value. With such weak programming conditions, it is possible to set a RSM from a high resistive state (HRS) to a low resistive state (LRS) with a given probability of 50%. This

behavior has successfully been exploited to implement a neuron activation function (Wijesinghe et al., 2018), a stochastic learning rule (Yu et al., 2013; Payvand et al., 2019; Zahari et al., 2020), and controllable weight sampling (Yang et al., 2020b).

Some types of NNs are particularly suitable to take advantage of RSM non-idealities, and of these, Bayesian neural networks (BNNs), restricted Boltzmann machines (RBMs), and spiking neural networks (SNNs) have received special attention.

BNNs are difficult to use for real-world problems because of their prohibitive computation cost on traditional computers, which is mainly due to the expensive random sampling of parameters. Nevertheless, the uncertainty measurement provided by this NN is valuable for many applications, such as healthcare (McLachlan et al., 2020) and autonomous vehicles (McAllister et al., 2017). An array of RSMs can provide an elegant solution to this issue by exploiting the writing (Malhotra et al., 2020; Dalgaty et al., 2021a,b; Yang et al., 2020b) or the reading (Lin et al., 2019) variability to efficiently sample the network weights in parallel while computing the VMM in place at the same time.

RBMs have non-deterministic activation functions, and usually have a relatively small number of parameters, which fits well with the small and noisy RSM crossbars that are currently available. In the same way as BNNs, the probabilistic aspect of RBMs is not very desirable for CMOS chips, whereas RSMs offer new design perspectives (Kaiser et al., 2022). For example, Suri et al. (2015) suggested using the HRS and LRS variability to build a stochastic activation function an RBM and Mahmoodi et al. (2019) experimentally demonstrated the benefits of thermal noise to realize a stochastic dot product computation.

Finally, the spiking approach seems to be a promising candidate for RSM-based NNs, as they naturally adapt to variability (Maass, 2014; Neftci et al., 2016; Leugering and Pipa, 2018). SNNs share many similarities with biological NNs, which are known to rely heavily on stochastic mechanisms (Stein et al., 2005; Deco et al., 2009; Rolls and Deco, 2010; Yarom and Hounsgaard, 2011) such as Poisson process or short-term memory. In this context, RSMs are well suited to implement the synaptic weights (Yu et al., 2013; Naous et al., 2016; Payvand et al., 2019; Wang et al., 2020a) and the components of a neuron (Al-Shedivat et al., 2015; Naous et al., 2016; Wijesinghe et al., 2018; Dalgaty et al., 2019; Li et al., 2020), in particular, the membrane potential and the activation function.

DISCUSSION

In the early 2010's, GPUs played an essential role in the rebirth of artificial intelligence as a research field by offering an efficient alternative to CPUs for VMM. Although RSM-based electronics have the potential to give rise to a similar hardware revolution, this transition is much more challenging since machine learning models face new constraints. Several works have shown through simulation and experimental results that harnessing the imperfections of RSMs is a viable option for tackling this problem and getting closer to the performance of software NNs.

However, this approach is subject to some limitations. While a reasonable quantity of stochasticity can improve the robustness of the model, larger amounts will be beneficial only to NNs that intrinsically rely on stochasticity, such as BNNs, RBMs, or SNNs. But, even in these cases, non-idealities must be kept under control to obtain a specific probability distribution shape or a consistent switching probability. The covariance between the resistance and the standard deviation is one example of the constraints that must be considered. To implement efficient probabilistic or approximate computing on RSMs, we may have to use unconventional approaches such as aggregating devices (Dalgaty et al., 2021a) or applying continuous writing operations (Yang et al., 2020b; Malhotra et al., 2020), which will alter the global energy efficiency and reduce the device lifetime.

Moreover, several RSM non-idealities have not yet been exploited for non-conventional computing schemes, such as device-to-device variability, sneak path current, state drift in time, or read and write disturbances. Further studies should explore novel circuit designs, encoding methods, and learning techniques benefiting from these characteristics for future hardware-based NNs.

The exploitation of non-idealities seems desirable, if not necessary, to accelerate the development of large-scale artificial NNs at state-of-the-art performance with competitive energy and area efficiency. The benefits are twofold in the case of stochastic (Dalgaty et al., 2021c) (BNNs and RBMs) or asynchronous event-based (Wang et al., 2020a; Agrawal et al., 2021) (SNNs) models, for which von Neumann CMOS computers are especially inefficient. This approach could be combined with the mitigation of harmful non-idealities, hardware-software co-design, and optimization of fabrication techniques to reach the full potential of RSM technology.

AUTHOR CONTRIBUTIONS

All authors conceived the review topic. VY wrote the first draft. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financed by the EU: ERC-2017-COG project IONOS (#GA 773228), the Natural Sciences and Engineering Research Council of Canada (NSERC) HIDATA project 506289-2017, CHIST-ERA UNICO project, the Fond de Recherche du Québec Nature et Technologies (FRQNT), and the Canada First Research Excellence Fund (CFREF). RM is supported by NSERC, the Canada Research Chair program, and the Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Economic Development, Job Creation and Trade.

REFERENCES

Abunahla, H., Mohammad, B., Homouz, D., and Okelly, C. J. (2016). Modeling Valance Change Memristor Device: Oxide Thickness, Material Type, and Temperature Effects. *IEEE Trans. Circuits Syst.* 63, 2139–2148. doi:10.1109/tcsi.2016.2622225

Adam, G. C., Khiat, A., and Prodromakis, T. (2018). Challenges Hindering Memristive Neuromorphic Hardware from Going Mainstream. *Nat. Commun.* 9. doi:10.1038/s41467-018-07565-4

Agrawal, A., Roy, D., Saxena, U., and Roy, K. (2021). Embracing Stochasticity to Enable Neuromorphic Computing at the Edge. *IEEE Des. Test.* 38, 28–35. doi:10.1109/mdat.2021.3051399

Al-Shedivat, M., Naous, R., Cauwenberghs, G., and Salama, K. N. (2015). Memristors Empower Spiking Neurons with Stochasticity. *IEEE J. Emerg. Sel. Top. Circuits Syst.* 5, 242–253. doi:10.1109/jetcas.2015.2435512

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. doi:10.1038/ncomms3072

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

Amirsoleimani, A., Alibart, F., Yon, V., Xu, J., Pazhouhandeh, M. R., Ecoffey, S., et al. (2020). In-Memory Vector-Matrix Multiplication in Monolithic Complementary Metal-Oxide-Semiconductor-Memristor Integrated Circuits: Design Choices, Challenges, and Perspectives. *Adv. Intell. Syst.* 2, 2000115. doi:10.1002/aisy.202000115

Amirsoleimani, A., Liu, T., Alibart, F., Eccofey, S., Chang, Y.-F., Drouin, D., et al. (2021). "Mitigating State-Drift in Memristor Crossbar Arrays for Vector Matrix Multiplication," in *Memristor - an Emerging Device for Post-Moore's Computing and Applications* (London, UK: IntechOpen). doi:10.5772/intechopen.100246

An, G. (1996). The Effects of Adding Noise during Backpropagation Training on a Generalization Performance. *Neural Comput.* 8, 643–674. doi:10.1162/neco.1996.8.3.643

Anwar, S., Hwang, K., and Sung, W. (2015). "Fixed point Optimization of Deep Convolutional Neural Networks for Object Recognition," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (South Brisbane, QLD, Australia: IEEE). doi:10.1109/icassp.2015.7178146

Bae, W., and Yoon, K. J. (2020). Comprehensive Read Margin and BER Analysis of One Selector-One Memristor Crossbar Array Considering thermal Noise of Memristor with Noise-Aware Device Model. *IEEE Trans. Nanotechnology* 19, 553–564. doi:10.1109/tnano.2020.3006114

Bao, L., Wang, Z., Yu, Z., Fang, Y., Yang, Y., Cai, Y., et al. (2020). "Adaptive Random Number Generator Based on RRAM Intrinsic Fluctuation for Reinforcement Learning," in 2020 International Symposium on VLSI Technology, Systems and Applications (VLSI-TSA) (Hsinchu, Taiwan: IEEE). doi:10.1109/vlsi-tsa48913.2020.9203571

[Dataset] Bhattacharjee, A., and Panda, P. (2020). *Rethinking Non-idealities in Memristive Crossbars for Adversarial Robustness in Neural Networks*.

Bocquet, M., Hirztlin, T., Klein, J.-O., Nowak, E., Vianello, E., Portal, J.-M., et al. (2018). "In-memory and Error-Immune Differential RRAM Implementation of Binarized Deep Neural Networks," in 2018 IEEE International Electron Devices Meeting (IEDM) (San Francisco, CA, USA: IEEE). doi:10.1109/iedm.2018.8614639

Boquet, G., Macias, E., Morell, A., Serrano, J., Miranda, E., and Vicario, J. L. (2021). "Offline Training for Memristor-Based Neural Networks," in 2020 28th European Signal Processing Conference (EUSIPCO) (Amsterdam, Netherlands: IEEE). doi:10.23919/eusipco47968.2020.9287574

Bunnam, T., Soltan, A., Sokolov, D., Maevsky, O., Degenaar, P., and Yakovlev, A. (2020). "Empirical Temperature Model of Self-Directed Channel Memristor," in 2020 IEEE SENSORS (Rotterdam, Netherlands: IEEE). doi:10.1109/sensors47125.2020.9278602

Cai, F., Kumar, S., Van Vaerenbergh, T., Sheng, X., Liu, R., Li, C., et al. (2020a). Power-efficient Combinatorial Optimization Using Intrinsic Noise in Memristor Hopfield Neural Networks. *Nat. Electron.* 3, 409–418. doi:10.1038/s41928-020-0436-6

Cai, R., Ren, A., Liu, N., Ding, C., Wang, L., Qian, X., et al. (2018). VIBNN. *SIGPLAN Not.* 53, 476–488. doi:10.1145/3296957.3173212

Cai, Y., Wang, Z., Yu, Z., Ling, Y., Chen, Q., Yang, Y., et al. (2020b). "Technology-array-algorithm Co-optimization of RRAM for Storage and Neuromorphic Computing: Device Non-idealities and thermal Cross-Talk," in 2020 IEEE International Electron Devices Meeting (IEDM) (San Francisco, CA, USA: IEEE). doi:10.1109/iedm13553.2020.9371968

Cassuto, Y., Kvatinsky, S., and Yaakobi, E. (2013). "Sneak-path Constraints in Memristor Crossbar Arrays," in 2013 IEEE International Symposium on Information Theory (Istanbul, Turkey: IEEE). doi:10.1109/isit.2013.6620207

Chakraborty, I., Ali, M., Ankit, A., Jain, S., Roy, S., Sridharan, S., et al. (2020). Resistive Crossbars as Approximate Hardware Building Blocks for Machine Learning: Opportunities and Challenges. *Proc. IEEE* 108, 2276–2310. doi:10.1109/jproc.2020.3003007

Chang, C.-C., Chen, P.-C., Chou, T., Wang, I.-T., Hudec, B., Chang, C.-C., et al. (2018). Mitigating Asymmetric Nonlinear Weight Update Effects in Hardware Neural Network Based on Analog Resistive Synapse. *IEEE J. Emerg. Sel. Top. Circuits Syst.* 8, 116–124. doi:10.1109/jetcas.2017.2771529

Chen, J., Wu, H., Gao, B., Tang, J., Hu, X. S., and Qian, H. (2020). A Parallel Multibit Programing Scheme with High Precision for RRAM-Based Neuromorphic Systems. *IEEE Trans. Electron. Devices* 67, 2213–2217. doi:10.1109/ted.2020.2979606

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 2015 IEEE/ACM International Conference on Computer-Aided Design (ICCAD) (Austin, TX, USA: IEEE). doi:10.1109/iccad.2015.7372570

Chen, P.-Y., and Yu, S. (2015). Compact Modeling of RRAM Devices and its Applications in 1t1r and 1s1r Array Design. *IEEE Trans. Electron. Devices* 62, 4022–4028. doi:10.1109/ted.2015.2492421

Chen, W., Qiu, H., Zhuang, J., Zhang, C., Hu, Y., Lu, Q., et al. (2021a). Quantization of Deep Neural Networks for Accurate Edge Computing. *J. Emerg. Technol. Comput. Syst.* 17, 1–11. doi:10.1145/3451211

Chen, Y.-C., Chang, Y.-F., Wu, X., Zhou, F., Guo, M., Lin, C.-Y., et al. (2017). Dynamic Conductance Characteristics in HfOx-Based Resistive Random Access Memory. *RSC Adv.* 7, 12984–12989. doi:10.1039/c7ra00567a

Chen, Y.-C., Lin, C.-C., and Chang, Y.-F. (2021b). Post-moore Memory Technology: Sneak Path Current (SPC) Phenomena on RRAM Crossbar Array and Solutions. *Micromachines* 12, 50. doi:10.3390/mi12010050

Chen, Y. (2020). ReRAM: History, Status, and Future. *IEEE Trans. Electron. Devices* 67, 1420–1433. doi:10.1109/ted.2019.2961505

Ciresan, D., Meier, U., and Schmidhuber, J. (2012). "Multi-column Deep Neural Networks for Image Classification," in 2012 IEEE Conference on Computer Vision and Pattern Recognition (IEEE). doi:10.1109/cvpr.2012.6248110

Claeys, C., de Andrade, M. G. C., Chai, Z., Fang, W., Govoreanu, B., Kaczer, B., et al. (2016). "Random Telegraph Signal Noise in Advanced High Performance and Memory Devices," in 2016 31st Symposium on Microelectronics Technology and Devices (SBMicro) (Belo Horizonte, Brazil: IEEE). doi:10.1109/sbmicro.2016.7731315

Courbariaux, M., Bengio, Y., and David, J.-P. (2015). "Binaryconnect: Training Deep Neural Networks with Binary Weights during Propagations," in NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems (Cambridge, MA, USA: MIT Press), 3123–3131. Vol. 2 of NIPS'15.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., and Bharath, A. A. (2018). Generative Adversarial Networks: An Overview. *IEEE Signal. Process. Mag.* 35, 53–65. doi:10.1109/msp.2017.2765202

Dalgaty, T., Castellani, N., Turck, C., Harabi, K.-E., Querlioz, D., and Vianello, E. (2021a). *In Situ* learning Using Intrinsic Memristor Variability via Markov Chain Monte Carlo Sampling. *Nat. Electron.* 4, 151–161. doi:10.1038/s41928-020-00523-3

Dalgaty, T., Esmanhotto, E., Castellani, N., Querlioz, D., and Vianello, E. (2021b). *Ex Situ* Transfer of Bayesian Neural Networks to Resistive Memory-Based Inference Hardware. *Adv. Intell. Syst.* 3, 2000103. doi:10.1002/aisy.202000103

Dalgaty, T., Payvand, M., Moro, F., Ly, D. R. B., Pebay-Peyroula, F., Casas, J., et al. (2019). Hybrid Neuromorphic Circuits Exploiting Non-conventional

Properties of RRAM for Massively Parallel Local Plasticity Mechanisms. *APL Mater.* 7, 081125. doi:10.1063/1.5108663

Dalgaty, T., Vianello, E., and Querlioz, D. (2021c). "Harnessing Intrinsic Memristor Randomness with Bayesian Neural Networks," in 2021 International Conference on IC Design and Technology (ICICDT) (Dresden, Germany: IEEE). doi:10.1109/icicdt51558.2021.9626535

[Dataset] Ankit, A., Sengupta, A., Panda, P., and Roy, K. (2017). *Resparc: A Reconfigurable and Energy-Efficient Architecture with Memristive Crossbars for Deep Spiking Neural Networks*.

[Dataset] Christensen, D. V., Dittmann, R., Linares-Barranco, B., Sebastian, A., Gallo, M. L., Redaelli, A., et al. (2021). 2022 Roadmap on Neuromorphic Computing and Engineering. *Neuromorphic Comput. Eng.*

[Dataset] Guo, Y. (2018). *A Survey on Methods and Theories of Quantized Neural Networks*.

[Dataset] Gupta, S., Agrawal, A., Gopalakrishnan, K., and Narayanan, P. (2015). *Deep Learning with Limited Numerical Precision*.

[Dataset] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). *Improving Neural Networks by Preventing Co-adaptation of Feature Detectors*.

[Dataset] Liu, X., Cheng, M., Zhang, H., and Hsieh, C.-J. (2017). *Towards Robust Neural Networks via Random Self-Ensemble*.

[Dataset] Mazzia, V., Salvetti, F., and Chiaberge, M. (2021). *Efficient-capsnet: Capsule Network with Self-Attention Routing*. doi:10.1038/s41598-021-93977-0

[Dataset] Micikevicius, P., Narang, S., Alben, J., Diamos, G., Elsen, E., Garcia, D., et al. (2017). *Mixed Precision Training*.

[Dataset] Mishra, R., Gupta, H. P., and Dutta, T. (2020). *A Survey on Deep Neural Network Compression: Challenges, Overview, and Solutions*.

[Dataset] Neelakantan, A., Vilnis, L., Le, Q. V., Sutskever, I., Kaiser, L., Kurach, K., et al. (2015). *Adding Gradient Noise Improves Learning for Very Deep Networks*.

[Dataset] Qin, M., and Vucinic, D. (2018). *Noisy Computations during Inference: Harmful or Helpful?*

[Dataset] Rakin, A. S., He, Z., and Fan, D. (2018). *Parametric Noise Injection: Trainable Randomness to Improve Deep Neural Network Robustness against Adversarial Attack*.

[Dataset] Shin, S., Hwang, K., and Sung, W. (2015). "Fixed-point Performance Analysis of Recurrent Neural Networks," in 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). doi:10.1109/MSP.2015.2411564

[Dataset] Zhou, S., Wu, Y., Ni, Z., Zhou, X., Wen, H., and Zou, Y. (2016). *Dorefa-net: Training Low Bitwidth Convolutional Neural Networks with Low Bitwidth Gradients*.

Deco, G., Rolls, E. T., and Romo, R. (2009). Stochastic Dynamics as a Principle of Brain Function. *Prog. Neurobiol.* 88, 1–16. doi:10.1016/j.pneurobio.2009.01.006

Faria, R., Camsari, K. Y., and Datta, S. (2018). Implementing Bayesian Networks with Embedded Stochastic MRAM. *AIP Adv.* 8, 045101. doi:10.1063/1.5021332

Gaba, S., Sheridan, P., Zhou, J., Choi, S., and Lu, W. (2013). Stochastic Memristive Devices for Computing and Neuromorphic Applications. *Nanoscale* 5, 5872. doi:10.1039/c3nr01176c

Gao, B., Sun, B., Zhang, H., Liu, L., Liu, X., Han, R., et al. (2009). Unified Physical Model of Bipolar Oxide-Based Resistive Switching Memory. *IEEE Electron. Device Lett.* 30, 1326–1328. doi:10.1109/led.2009.2032308

Gao, B., Zhang, H., Chen, B., Liu, L., Liu, X., Han, R., et al. (2011). Modeling of Retention Failure Behavior in Bipolar Oxide-Based Resistive Switching Memory. *IEEE Electron. Device Lett.* 32, 276–278. doi:10.1109/led.2010.2102002

Gao, L., Wang, I.-T., Chen, P.-Y., Vrudhula, S., Seo, J.-s., 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

Gong, N., Idé, T., Kim, S., Boybat, I., Sebastian, A., Narayanan, V., et al. (2018). Signal and Noise Extraction from Analog Memory Elements for Neuromorphic Computing. *Nat. Commun.* 9. doi:10.1038/s41467-018-04485-1

González-Cordero, G., Jiménez-Molinos, F., Roldán, J. B., González, M. B., and Campabadal, F. (2017). In-depth Study of the Physics behind Resistive Switching in TiN/ti/HfO2/w Structures. *J. Vacuum Sci. Technol. B, Nanotechnology Microelectronics: Mater. Process. Meas. Phenomena* 35, 01A110. doi:10.1116/1.4973372

Gulcehre, C., Moczulski, M., Denil, M., and Bengio, Y. (2016). "Noisy Activation Functions," in Proceedings of The 33rd International Conference on Machine Learning. Editors M. F. Balcan and K. Q. Weinberger (New York, New York, USA: PMLR), 3059–3068. vol. 48 of Proceedings of Machine Learning Research.

He, Z., Lin, J., Ewetz, R., Yuan, J.-S., and Fan, D. (2019). "Noise Injection Adaption," in Proceedings of the 56th Annual Design Automation Conference 2019 (ACM). doi:10.1145/3316781.3317870

Holt, J. L., and Baker, T. E. (1991). "Back Propagation Simulations Using Limited Precision Calculations," in IJCNN-91-Seattle International Joint Conference on Neural Networks (Seattle, WA, USA: IEEE). doi:10.1109/ijcnn.1991.155324

Horowitz, M. (2014). "1.1 Computing's Energy Problem (And what We Can Do about it)," in 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC) (San Francisco, CA, USA: IEEE). doi:10.1109/isscc.2014.6757323

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., Wang, Y., Wen, W., Wang, Y., and Li, H. (2016). Leveraging Stochastic Memristor Devices in Neuromorphic Hardware Systems. *IEEE J. Emerg. Sel. Top. Circuits Syst.* 6, 235–246. doi:10.1109/jetcas.2016.2547780

Ielmini, D., Nardi, F., and Cagli, C. (2010). Resistance-dependent Amplitude of Random Telegraph-Signal Noise in Resistive Switching Memories. *Appl. Phys. Lett.* 96, 053503. doi:10.1063/1.3304167

Ishii, M., Shin, U., Hosokawa, K., BrightSky, M., Haensch, W., Kim, S., et al. (2019). "On-chip Trainable 1.4m 6t2r PCM Synaptic Array with 1.6k Stochastic LIF Neurons for Spiking RBM," in 2019 IEEE International Electron Devices Meeting (IEDM) (San Francisco, CA, USA: IEEE). doi:10.1109/iedm19573.2019.8993466

Jacobs-Gedrim, R. B., Agarwal, S., Knisely, K. E., Stevens, J. E., van Heukelom, M. S., Hughart, D. R., et al. (2017). "Impact of Linearity and Write Noise of Analog Resistive Memory Devices in a Neural Algorithm Accelerator," in 2017 IEEE International Conference on Rebooting Computing (ICRC) (Washington, DC, USA: IEEE). doi:10.1109/icrc.2017.8123657

Joksas, D., Freitas, P., Chai, Z., Ng, W. H., Buckwell, M., Li, C., et al. (2020). Committee Machines-A Universal Method to deal with Non-idealities in Memristor-Based Neural Networks. *Nat. Commun.* 11. doi:10.1038/s41467-020-18098-0

Jouppi, N. P., Young, C., Patil, N., Patterson, D., Agrawal, G., Bajwa, R., et al. (2017). "In-datacenter Performance Analysis of a Tensor Processing Unit," in Proceedings of the 44th Annual International Symposium on Computer Architecture (ACM). doi:10.1145/3079856.3080246

Kadetotad, D., Chen, P.-Y., Cao, Y., Yu, S., and Seo, J.-s. (2017). "Peripheral Circuit Design Considerations of Neuro-Inspired Architectures," in *Neuro-inspired Computing Using Resistive Synaptic Devices* (Berlin, Germany: Springer International Publishing), 167–182. doi:10.1007/978-3-319-54313-0_9

Kaiser, J., Borders, W. A., Camsari, K. Y., Fukami, S., Ohno, H., and Datta, S. (2022). Hardware-aware *In Situ* Learning Based on Stochastic Magnetic Tunnel Junctions. *Phys. Rev. Appl.* 17. doi:10.1103/physrevapplied.17.014016

Kan, J. J., Park, C., Ching, C., Ahn, J., Xue, L., Wang, R., et al. (2016). "Systematic Validation of 2x Nm Diameter Perpendicular MTJ Arrays and MgO Barrier for Sub-10 Nm Embedded STT-MRAM with Practically Unlimited Endurance," in 2016 IEEE International Electron Devices Meeting (IEDM) (San Francisco, CA, USA: IEEE). doi:10.1109/iedm.2016.7838493

Kang, J., Yu, Z., Wu, L., Fang, Y., Wang, Z., Cai, Y., et al. (2017). "Time-dependent Variability in RRAM-Based Analog Neuromorphic System for Pattern Recognition," in 2017 IEEE International Electron Devices Meeting (IEDM) (San Francisco, CA, USA: IEEE). doi:10.1109/iedm.2017.8268340

Keckler, S. W., Dally, W. J., Khailany, B., Garland, M., and Glasco, D. (2011). GPUs and the Future of Parallel Computing. *IEEE Micro* 31, 7–17. doi:10.1109/mm.2011.89

Khan, M. I., Ali, S., Al-Tamimi, A., Hassan, A., Ikram, A. A., and Bermak, A. (2021). A Robust Architecture of Physical Unclonable Function Based on Memristor Crossbar Array. *Microelectronics J.* 116, 105238. doi:10.1016/j.mejo.2021.105238

Krestinskaya, O., Irmanova, A., and James, A. P. (2019). "Memristive Non-idealities: Is There Any Practical Implications for Designing Neural Network Chips?," in 2019 IEEE International Symposium on Circuits and Systems (ISCAS) (Sapporo, Japan: IEEE). doi:10.1109/iscas.2019.8702245

Lee, J.-K., Lee, J.-W., Park, J., Chung, S.-W., Roh, J. S., Hong, S.-J., et al. (2011). Extraction of Trap Location and Energy from Random Telegraph Noise in Amorphous TiOx Resistance Random Access Memories. *Appl. Phys. Lett.* 98, 143502. doi:10.1063/1.3575572

Leugering, J., and Pipa, G. (2018). A Unifying Framework of Synaptic and Intrinsic Plasticity in Neural Populations. *Neural Comput.* 30, 945–986. doi:10.1162/neco_a_01057

Li, B., Gu, P., Wang, Y., and Yang, H. (2016). Exploring the Precision Limitation for RRAM-Based Analog Approximate Computing. *IEEE Des. Test.* 33, 51–58. doi:10.1109/mdat.2015.2487218

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. doi:10.1038/s41467-018-04484-2

Li, C., Hu, M., Li, Y., Jiang, H., Ge, N., Montgomery, E., et al. (2017). Analogue Signal and Image Processing with Large Memristor Crossbars. *Nat. Electron.* 1, 52–59. doi:10.1038/s41928-017-0002-z

Li, X., Tang, J., Zhang, Q., Gao, B., Yang, J. J., Song, S., et al. (2020). Power-efficient Neural Network with Artificial Dendrites. *Nat. Nanotechnol.* 15, 776–782. doi:10.1038/s41565-020-0722-5

Li, Y., and Ang, K.-W. (2020). Hardware Implementation of Neuromorphic Computing Using Large-Scale Memristor Crossbar Arrays. *Adv. Intell. Syst.* 3, 2000137. doi:10.1002/aisy.202000137

Lim, S., Bae, J.-H., Eum, J.-H., Lee, S., Kim, C.-H., Kwon, D., et al. (2018). Adaptive Learning Rule for Hardware-Based Deep Neural Networks Using Electronic Synapse Devices. *Neural Comput. Applic* 31, 8101–8116. doi:10.1007/s00521-018-3659-y

Lin, Y., Hu, X. S., Qian, H., Wu, H., Zhang, Q., Tang, J., et al. (2019). "Bayesian Neural Network Realization by Exploiting Inherent Stochastic Characteristics of Analog RRAM," in 2019 IEEE International Electron Devices Meeting (IEDM) (San Francisco, CA, USA: IEEE). doi:10.1109/iedm19573.2019.8993616

Lin, Y., Wu, H., Gao, B., Yao, P., Wu, W., Zhang, Q., et al. (2018). "Demonstration of Generative Adversarial Network by Intrinsic Random Noises of Analog RRAM Devices," in 2018 IEEE International Electron Devices Meeting (IEDM) (San Francisco, CA, USA: IEEE). doi:10.1109/iedm.2018.8614483

Liu, C., Yu, F., Qin, Z., and Chen, X. (2020a). "Enabling Efficient ReRAM-Based Neural Network Computing via Crossbar Structure Adaptive Optimization," in Proceedings of the ACM/IEEE International Symposium on Low Power Electronics and Design (ACM). doi:10.1145/3370748.3406581

Liu, Q., Gao, B., Yao, P., Wu, D., Chen, J., Pang, Y., et al. (2020b). "33.2 a Fully Integrated Analog ReRAM Based 78.4tops/w Compute-In-Memory Chip with Fully Parallel MAC Computing," in 2020 IEEE International Solid- State Circuits Conference - (ISSCC) (San Francisco, CA, USA: IEEE). doi:10.1109/isscc19947.2020.9062953

Lv, S., Liu, J., and Geng, Z. (2020). Application of Memristors in Hardware Security: A Current State-of-the-Art Technology. *Adv. Intell. Syst.* 3, 2000127. doi:10.1002/aisy.202000127

Maass, W. (2014). Noise as a Resource for Computation and Learning in Networks of Spiking Neurons. *Proc. IEEE* 102, 860–880. doi:10.1109/jproc.2014.2310593

Mahmoodi, M. R., Prezioso, M., and Strukov, D. B. (2019). Versatile Stochastic Dot Product Circuits Based on Nonvolatile Memories for High Performance Neurocomputing and Neurooptimization. *Nat. Commun.* 10, 5113. doi:10.1038/s41467-019-13103-7

Mahmoodi, M. R., Vincent, A. F., Nili, H., and Strukov, D. B. (2020). Intrinsic Bounds for Computing Precision in Memristor-Based Vector-By-Matrix Multipliers. *IEEE Trans. Nanotechnology* 19, 429–435. doi:10.1109/tnano.2020.2992493

Malhotra, A., Lu, S., Yang, K., and Sengupta, A. (2020). Exploiting Oxide Based Resistive RAM Variability for Bayesian Neural Network Hardware Design. *IEEE Trans. Nanotechnology* 19, 328–331. doi:10.1109/tnano.2020.2982819

Marković, D., Mizrahi, A., Querlioz, D., and Grollier, J. (2020). Physics for Neuromorphic Computing. *Nat. Rev. Phys.* 2, 499–510. doi:10.1038/s42254-020-0208-2

McAllister, R., Gal, Y., Kendall, A., van der Wilk, M., Shah, A., Cipolla, R., et al. (2017). "Concrete Problems for Autonomous Vehicle Safety: Advantages of Bayesian Deep Learning," in Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (Messe Wien, Vienna, Austria: International Joint Conferences on Artificial Intelligence Organization). doi:10.24963/ijcai.2017/661

McDonnell, M. D., and Ward, L. M. (2011). The Benefits of Noise in Neural Systems: Bridging Theory and experiment. *Nat. Rev. Neurosci.* 12, 415–425. doi:10.1038/nrn3061

McLachlan, S., Dube, K., Hitman, G. A., Fenton, N. E., and Kyrimi, E. (2020). Bayesian Networks in Healthcare: Distribution by Medical Condition. *Artif. Intelligence Med.* 107, 101912. doi:10.1016/j.artmed.2020.101912

Messerschmitt, F., Kubicek, M., and Rupp, J. L. M. (2015). How Does Moisture Affect the Physical Property of Memristance for Anionic-Electronic Resistive Switching Memories? *Adv. Funct. Mater.* 25, 5117–5125. doi:10.1002/adfm.201501517

Mochida, R., Kouno, K., Hayata, Y., Nakayama, M., Ono, T., Suwa, H., et al. (2018). "A 4m Synapses Integrated Analog ReRAM Based 66.5 TOPS/w Neural-Network Processor with Cell Current Controlled Writing and Flexible Network Architecture," in 2018 IEEE Symposium on VLSI Technology (Honolulu, HI, USA: IEEE). doi:10.1109/vlsit.2018.8510676

Mutlu, O., Ghose, S., Gómez-Luna, J., and Ausavarungnirun, R. (2019). Processing Data where it Makes Sense: Enabling In-Memory Computation. *Microprocessors and Microsystems* 67, 28–41. doi:10.1016/j.micpro.2019.01.009

Nag, A., Balasubramonian, R., Srikumar, V., Walker, R., Shafiee, A., Strachan, J. P., et al. (2018). Newton: Gravitating towards the Physical Limits of Crossbar Acceleration. *IEEE Micro* 38, 41–49. doi:10.1109/mm.2018.053631140

Naous, R., AlShedivat, M., Neftci, E., Cauwenberghs, G., and Salama, K. N. (2016). Memristor-based Neural Networks: Synaptic versus Neuronal Stochasticity. *AIP Adv.* 6, 111304. doi:10.1063/1.4967352

Neftci, E. O., Pedroni, B. U., Joshi, S., Al-Shedivat, M., and Cauwenberghs, G. (2016). Stochastic Synapses Enable Efficient Brain-Inspired Learning Machines. *Front. Neurosci.* 10. doi:10.3389/fnins.2016.00241

Oh, S., Huang, Z., Shi, Y., and Kuzum, D. (2019). The Impact of Resistance Drift of Phase Change Memory (PCM) Synaptic Devices on Artificial Neural Network Performance. *IEEE Electron. Device Lett.* 40, 1325–1328. doi:10.1109/led.2019.2925832

Pan, W.-Q., Chen, J., Kuang, R., Li, Y., He, Y.-H., Feng, G.-R., et al. (2020). Strategies to Improve the Accuracy of Memristor-Based Convolutional Neural Networks. *IEEE Trans. Electron. Devices* 67, 895–901. doi:10.1109/ted.2019.2963323

Pang, Y., Gao, B., Lin, B., Qian, H., and Wu, H. (2019). Memristors for Hardware Security Applications. *Adv. Electron. Mater.* 5, 1800872. doi:10.1002/aelm.201800872

Papandreou, N., Pozidis, H., Pantazi, A., Sebastian, A., Breitwisch, M., Lam, C., et al. (2011). "Programming Algorithms for Multilevel Phase-Change Memory," in 2011 IEEE International Symposium of Circuits and Systems (ISCAS) (Rio de Janeiro, Brazil: IEEE). doi:10.1109/iscas.2011.5937569

Payvand, M., Nair, M. V., Müller, L. K., and Indiveri, G. (2019). A Neuromorphic Systems Approach to In-Memory Computing with Non-ideal Memristive Devices: from Mitigation to Exploitation. *Faraday Discuss.* 213, 487–510. doi:10.1039/c8fd00114f

Peng Gu, P., Boxun Li, B., Tianqi Tang, T., Yu, S., Yu Cao, Y., Wang, Y., et al. (2015). "Technological Exploration of RRAM Crossbar Array for Matrix-Vector Multiplication," in The 20th Asia and South Pacific Design Automation Conference (Chiba, Japan: IEEE). doi:10.1109/aspdac.2015.7058989

Perez, E., Grossi, A., Zambelli, C., Olivo, P., Roelofs, R., and Wenger, C. (2017). Reduction of the Cell-To-Cell Variability in Hf1-xAlxOyBased RRAM Arrays by Using Program Algorithms. *IEEE Electron. Device Lett.* 38, 175–178. doi:10.1109/led.2016.2646758

Reuther, A., Michaleas, P., Jones, M., Gadepally, V., Samsi, S., and Kepner, J. (2019). "Survey and Benchmarking of Machine Learning Accelerators," in 2019 IEEE High Performance Extreme Computing Conference (HPEC) (Waltham, MA, USA: IEEE). doi:10.1109/hpec.2019.8916327

Roldán, J. B., González-Cordero, G., Picos, R., Miranda, E., Palumbo, F., Jiménez-Molinos, F., et al. (2021). On the thermal Models for Resistive Random Access Memory Circuit Simulation. *Nanomaterials* 11, 1261. doi:10.3390/nano11051261

Rolls, E. T., and Deco, G. (2010). *The Noisy BrainStochastic Dynamics as a Principle of Brain Function*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199587865.001.0001

Romero, L. P., Ambrogio, S., Giordano, M., Cristiano, G., Bodini, M., Narayanan, P., et al. (2019). Training Fully Connected Networks with Resistive Memories: Impact of Device Failures. *Faraday Discuss.* 213, 371–391. doi:10.1039/c8fd00107c

Shafiee, A., Nag, A., Muralimanohar, N., Balasubramonian, R., Strachan, J. P., Hu, M., et al. (2016). ISAAC. *SIGARCH Comput. Archit. News* 44, 14–26. doi:10.1145/3007787.3001139

Shen, W., Huang, P., Fan, M., Zhao, Y., Feng, Y., Liu, L., et al. (2021). A Seamless, Reconfigurable, and Highly Parallel In-Memory Stochastic Computing Approach with Resistive Random Access Memory Array. *IEEE Trans. Electron. Devices* 68, 103–108. doi:10.1109/ted.2020.3037279

Sidler, S., Boybat, I., Shelby, R. M., Narayanan, P., Jang, J., Fumarola, A., et al. (2016). "Large-scale Neural Networks Implemented with Non-volatile Memory as the Synaptic Weight Element: Impact of Conductance Response," in 2016 46th European Solid-State Device Research Conference (ESSDERC) (Lausanne, Switzerland: IEEE). doi:10.1109/essderc.2016.7599680

Siegel, S., Baeumer, C., Gutsche, A., Witzleben, M., Waser, R., Menzel, S., et al. (2020). Trade-Off between Data Retention and Switching Speed in Resistive Switching ReRAM Devices. *Adv. Electron. Mater.* 7, 2000815. doi:10.1002/aelm.202000815

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. doi:10.1038/s41598-017-17785-1

Stein, R. B., Gossen, E. R., and Jones, K. E. (2005). Neuronal Variability: Noise or Part of the Signal? *Nat. Rev. Neurosci.* 6, 389–397. doi:10.1038/nrn1668

Subhechha, S., Govoreanu, B., Chen, Y., Clima, S., De Meyer, K., Van Houdt, J., et al. (2016). "Extensive Reliability Investigation of A-VMCO Nonfilamentary RRAM: Relaxation, Retention and Key Differences to Filamentary Switching," in 2016 IEEE International Reliability Physics Symposium (IRPS) (Pasadena, CA, USA: IEEE). doi:10.1109/irps.2016.7574568

Sung, C., Hwang, H., and Yoo, I. K. (2018). Perspective: A Review on Memristive Hardware for Neuromorphic Computation. *J. Appl. Phys.* 124, 151903. doi:10.1063/1.5037835

Suri, M., and Parmar, V. (2015). Exploiting Intrinsic Variability of Filamentary Resistive Memory for Extreme Learning Machine Architectures. *IEEE Trans. Nanotechnology* 14, 963–968. doi:10.1109/tnano.2015.2441112

Suri, M., Parmar, V., Kumar, A., Querlioz, D., and Alibart, F. (2015). "Neuromorphic Hybrid RRAM-CMOS RBM Architecture," in 2015 15th Non-Volatile Memory Technology Symposium (NVMTS) (Beijing, China: IEEE). doi:10.1109/nvmts.2015.7457484

Sze, V., Chen, Y.-H., Emer, J., Suleiman, A., and Zhang, Z. (2017). "Hardware for Machine Learning: Challenges and Opportunities," in 2017 IEEE Custom Integrated Circuits Conference (CICC) (IEEE). doi:10.1109/cicc.2017.7993626

Tian, B.-B., Zhong, N., and Duan, C.-G. (2020). Recent Advances, Perspectives, and Challenges in Ferroelectric Synapses*. *Chin. Phys. B* 29, 097701. doi:10.1088/1674-1056/aba603

Valov, I., and Tsuruoka, T. (2018). Effects of Moisture and Redox Reactions in VCM and ECM Resistive Switching Memories. *J. Phys. D: Appl. Phys.* 51, 413001. doi:10.1088/1361-6463/aad581

Veksler, D., Bersuker, G., Vandelli, L., Padovani, A., Larcher, L., Muraviev, A., et al. (2013). "Random Telegraph Noise (RTN) in Scaled RRAM Devices," in 2013 IEEE International Reliability Physics Symposium (IRPS) (Monterey, CA, USA: IEEE). doi:10.1109/irps.2013.6532101

Wan, L., Zeiler, M., Zhang, S., Le Cun, Y., and Fergus, R. (2013). "Regularization of Neural Networks Using Dropconnect," in Proceedings of the 30th International Conference on Machine Learning. Editors S. Dasgupta and D. McAllester (Atlanta, Georgia, USA: PMLR), 1058–1066. Vol. 28 of Proceedings of Machine Learning Research.

Wang, C., Feng, D., Tong, W., Liu, J., Li, Z., Chang, J., et al. (2019a). Cross-point Resistive Memory. *ACM Trans. Des. Autom. Electron. Syst.* 24, 1–37. doi:10.1145/3325067

Wang, C., Wu, H., Gao, B., Zhang, T., Yang, Y., and Qian, H. (2018). Conduction Mechanisms, Dynamics and Stability in ReRAMs. *Microelectronic Eng.* 187-188, 121–133. doi:10.1016/j.mee.2017.11.003

Wang, W., Song, W., Yao, P., Li, Y., Van Nostrand, J., Qiu, Q., et al. (2020a). Integration and Co-design of Memristive Devices and Algorithms for Artificial Intelligence. *iScience* 23, 101809. doi:10.1016/j.isci.2020.101809

Wang, Y., Wu, S., Tian, L., and Shi, L. (2020b). SSM: a High-Performance Scheme for *In Situ* Training of Imprecise Memristor Neural Networks. *Neurocomputing* 407, 270–280. doi:10.1016/j.neucom.2020.04.130

Wang, Z., Li, C., Lin, P., Rao, M., Nie, Y., Song, W., et al. (2019b). *In Situ* training of Feed-Forward and Recurrent Convolutional Memristor Networks. *Nat. Mach Intell.* 1, 434–442. doi:10.1038/s42256-019-0089-1

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 2008 IEEE International Electron Devices Meeting (IEEE). doi:10.1109/iedm.2008.4796676

Wiefels, S., Bengel, C., Kopperberg, N., Zhang, K., Waser, R., and Menzel, S. (2020). HRS Instability in Oxide-Based Bipolar Resistive Switching Cells. *IEEE Trans. Electron. Devices* 67, 4208–4215. doi:10.1109/ted.2020.3018096

Wijesinghe, P., Ankit, A., Sengupta, A., and Roy, K. (2018). An All-Memristor Deep Spiking Neural Computing System: A Step toward Realizing the Low-Power Stochastic Brain. *IEEE Trans. Emerg. Top. Comput. Intell.* 2, 345–358. doi:10.1109/tetci.2018.2829924

W. J. Gross and V. C. Gaudet (Editors) (2019). *Stochastic Computing: Techniques and Applications* (Berlin, Germany: Springer). doi:10.1007/978-3-030-03730-7

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. (2016). 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., and Yu, S. (2018). Resistive Memory-Based Analog Synapse: The Pursuit for Linear and Symmetric Weight Update. *IEEE Nanotechnology Mag.* 12, 36–44. doi:10.1109/mnano.2018.2844902

Wu, H., Yao, P., Gao, B., Wu, W., Zhang, Q., Zhang, W., et al. (2017). "Device and Circuit Optimization of RRAM for Neuromorphic Computing," in 2017 IEEE International Electron Devices Meeting (IEDM) (San Francisco, CA, USA: IEEE). doi:10.1109/iedm.2017.8268372

Wu, W., Wu, H., Gao, B., Yao, P., Zhang, X., Peng, X., et al. (2018). "A Methodology to Improve Linearity of Analog RRAM for Neuromorphic Computing," in 2018 IEEE Symposium on VLSI Technology (Honolulu, HI, USA: IEEE). doi:10.1109/vlsit.2018.8510690

Xi, Y., Gao, B., Tang, J., Chen, A., Chang, M.-F., Hu, X. S., et al. (2021). In-memory Learning with Analog Resistive Switching Memory: A Review and Perspective. *Proc. IEEE* 109, 14–42. doi:10.1109/jproc.2020.3004543

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

Yan, B., Yang, J., Wu, Q., Chen, Y., and Li, H. (2017). "A Closed-Loop Design to Enhance Weight Stability of Memristor Based Neural Network Chips," in 2017 IEEE/ACM International Conference on Computer-Aided Design (ICCAD) (Irvine, CA, USA: IEEE). doi:10.1109/iccad.2017.8203824

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, 232102. doi:10.1063/1.3524521

Yang, J. Q., Wang, R., Ren, Y., Mao, J. Y., Wang, Z. P., Zhou, Y., et al. (2020a). Neuromorphic Engineering: From Biological to Spike-Based Hardware Nervous Systems. *Adv. Mater.* 32, 2003610. doi:10.1002/adma.202003610

Yang, K., Malhotra, A., Lu, S., and Sengupta, A. (2020b). All-spin Bayesian Neural Networks. *IEEE Trans. Electron. Devices* 67, 1340–1347. doi:10.1109/ted.2020.2968223

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

Yarom, Y., and Hounsgaard, J. (2011). Voltage Fluctuations in Neurons: Signal or Noise? *Physiol. Rev.* 91, 917–929. doi:10.1152/physrev.00019.2010

Yi, W., Perner, F., Qureshi, M. S., Abdalla, H., Pickett, M. D., Yang, J. J., et al. (2011). Feedback Write Scheme for Memristive Switching Devices. *Appl. Phys. A.* 102, 973–982. doi:10.1007/s00339-011-6279-2

Yu, S., Gao, B., Fang, Z., Yu, H., Kang, J., and Wong, H.-S. P. (2013). Stochastic Learning in Oxide Binary Synaptic Device for Neuromorphic Computing. *Front. Neurosci.* 7. doi:10.3389/fnins.2013.00186

Yu, S. (2018). Neuro-inspired Computing with Emerging Nonvolatile Memorys. *Proc. IEEE* 106, 260–285. doi:10.1109/jproc.2018.2790840

Zahari, F., Pérez, E., Mahadevaiah, M. K., Kohlstedt, H., Wenger, C., and Ziegler, M. (2020). Analogue Pattern Recognition with Stochastic Switching Binary CMOS-Integrated Memristive Devices. *Sci. Rep.* 10. doi:10.1038/s41598-020-71334-x

Zahoor, F., Azni Zulkifli, T. Z., and Khanday, F. A. (2020). Resistive Random Access Memory (RRAM): an Overview of Materials, Switching Mechanism, Performance, Multilevel Cell (Mlc) Storage, Modeling, and Applications. *Nanoscale Res. Lett.* 15. doi:10.1186/s11671-020-03299-9

Zanotti, T., Puglisi, F. M., and Pavan, P. (2021). "Low-bit Precision Neural Network Architecture with High Immunity to Variability and Random Telegraph Noise Based on Resistive Memories," in 2021 IEEE International Reliability Physics Symposium (IRPS) (Monterey, CA, USA: IEEE). doi:10.1109/irps46558.2021.9405103

Zhang, B., Murshed, M. G. S., Hussain, F., and Ewetz, R. (2020). "Fast Resilient-Aware Data Layout Organization for Resistive Computing Systems," in 2020 IEEE Computer Society Annual Symposium on VLSI (ISVLSI) (Limassol, Cyprus: IEEE). doi:10.1109/isvlsi49217.2020.00023

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

Zhao, M., Gao, B., Xi, Y., Xu, F., Wu, H., and Qian, H. (2019). Endurance and Retention Degradation of Intermediate Levels in Filamentary Analog RRAM. *IEEE J. Electron. Devices Soc.* 7, 1239–1247. doi:10.1109/jeds.2019.2943017

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.