# A Behavioral Model of Digital Resistive Switching for Systems Level DNN Acceleration

Jason K. Eshraghian, *Member, IEEE*, Qi Lin, Xiaoyuan Wang, Herbert H.C. Iu, *Senior Member, IEEE*, Qing Hu, and Hao Tong

***Abstract*—The deployment of IoT has brought on the generation of massive amounts of data in need of analysis. In recent times, resistive switching-based crossbar arrays have been presented as a viable candidate for the acceleration of neural network inference, in pushing beyond the limit of CMOS process scaling so as to keep pace with the ever-growing complexity of computation. While plenty of empirical and physically descriptive models exist, their simulation run times become inconvenient for users when used in large scale crossbar arrays. In this paper, we first present a behavioral model of digital resistive switching devices, demonstrated on experimental PCM data to exhibit generality which can see useful implementation in circuit analysis methods for compute-in-memory applications. This model is based on a pair of nonlinear ordinary differential equations that request switching time and threshold voltage inputs from the user, which are the most important concerns for binarized weights in crossbar arrays. By stripping the model of detailed physical characteristics that is not required at the systems level, we demonstrate an improvement of computational run time of up to 20-fold over state-of-the-art physics-based models, and 1.3 times over the most commonly used empirically driven models.**

***Index Terms*—behavioral model, crossbar, neural network, memristor, RRAM**

## I. Introduction

In response to the proliferating amount of data in our IoT-driven world, the types of algorithms being developed for inferential processing and training of deep neural networks (DNNs) are increasing, as are the types of hardware capable of processing these algorithms. In fact, many advanced AI solutions tend to use a hybrid of technologies. While CPUs, GPUs and FPGAs (often remotely accessible via cloud-based options) are well established of their DNN machine learning capabilities, we have witnessed the introduction of multiple application specific integrated circuits (ASICs) that have the capacity to outperform conventional hardware in terms of speed, size and power consumption.

In addition, an experimental candidate which has shown good promise for inference acceleration are Resistive RAM (RRAM) crossbar arrays, thanks to its fast switching speed ($<$10 ns), low switching threshold ($<$3 V) and ease of integrability with CMOS [1]–[8]. As memristor crossbars are increasingly integrated with CMOS cells and peripheral circuitry, it is essential for RRAM models to keep pace with the needs of circuit designers to simulate very large scale arrays with a high level of computational efficiency. Circuit simulation speed and accuracy is critical for timely design, and this can be implemented by avoiding expensive math functions. Where internal nodes are used, the circuit simulator should solve for the quantities on the node itself.

Many models that are being developed seek to accurately capture the physical phenomena that gives rise to the conduction mechanisms of low and high resistance states. This is often performed by identifying the physical variables that define the resistance state in question. In valence change mechanisms of RRAM, this includes nucleation, ion hopping, electron-transfer reactions and temperature and electric field acceleration. The contributions of these various mechanisms determine set and reset switching times, and switching thresholds [9]. Such physics-driven models are essential for determining process corners through parametric variations, such as Monte Carlo simulations.

As memristor crossbar arrays increase in scale [10]–[12], there is a need for simplified models that capture the essential features of resistive switching relevant to systems level designers that are operating at a higher layer of abstraction [13]. Compact models are a computationally efficient description of the terminal properties of a device as a function of the terminal voltages. The main challenge of developing compact models is the requirement of balancing the need for physical scaling and accuracy with computational speed. At present, unified methodologies of memristor modeling in electronic design automation (EDA) and layout tools are sparse. At the highest level of abstraction, RRAM devices are treated as simplistic switches in hardware description languages (HDLs). The next layer of complexity often treats them as simple resistors for read-out, and beyond that, memristors are represented with more complex macromodels [14]–[16], or physics based analytical models [9], [17]–[20].

Behavioral models in HDLs only consider characteristics essential for systems level implementation. At this stage, industrial use of RRAM tends to be limited to digital-domain, single-bit switching due to ensure reliability [**?**]. In this paper, we present a generalized RRAM model that strips the device of physically-based descriptors that give rise to analog behaviors,

J. K Eshraghian is with the Department of Electrical Engineering and Computer Science, University of Michigan Ann Arbor, MI 48109, USA. (Corresponding author: Jason K. Eshraghian, e-mail: jaosnesh@umich.edu)

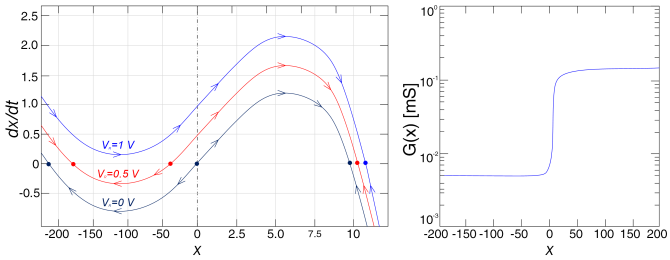
Q. Lin, Q. Hu and H. Tong are with the School of Optical and Electornic Information, Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430074, China

X. Wang is with the School of Electronics and Information, Hangzhou Dianzi University, Hangzhou 310018, China

H.H.C. Iu is with the School of Electrical, Electronic and Computer Engineering, The University of Western Australia, Perth, WA 6009, Australia

J. K. Eshraghian's contribution was supported by the Endeavour Research Leadership Award from the Australian Government.

Manuscript received February 2, 2020.




(a) (b)

Fig. 1. Behavioral digital RRAM model (a) Phase plane of the model with voltage variation dynamic route map (note the asymmetry along the x-axis caused by large disparity of switching time between set and reset processes in phase-change memory) (b) conductance variation with state.

and prioritize computational efficiency by including information that is relevant only at the behavioral level. This is wholly limited to switching thresholds, voltage-dependent switching times, on and off conductances, and transition times. As we will demonstrate, it can be used to characterize both bipolar and unipolar switching characteristics. The simplistic form of the model enables real-time simulation on a modern system of large-scale crossbar arrays, whilst capturing the behavioral parameters most relevant to a systems level designer.

## II. Generalized RRAM Model

In deriving the dynamical system model of RRAM, a number of conditions are imposed by virtue of the set and reset processes and their associated switching characteristics. The variable description and parameterization of each constituent property is summarized in Table I, with accompanying examples of parameters found in phase change memory devices (PCM). We use PCM as the guiding example to show generality to any sort of resistive switch intended for operation in either of two modes. A general form of our conductance-based dynamical system model is characterized by (1) and (2) below:

$$\frac{dx}{dt} = \begin{cases} ax(x+b) \pm v_M, & \text{if } x < 0 \\ -cx(x-d) \pm v_M, & \text{if } x \geq 0 \end{cases} \tag{1}$$

$$i(x) = v_M\big(\frac{g_s}{1+e^{sx}} + g_r\big). \tag{2}$$

Equation (1) is known as the 'dynamical system' or 'phase plane' model which describes the rate of change of the internal state variable, $x$, and (2) follows a state-dependent Ohm's Law to express current. The two above equations are illustrated in Fig. 1, and (2) is in terms of the bracketed conductance term.

The design behind the form of (1) is the need for bistability to ensure the device is always operating in one of two states outside the programming phase. The model can take a number of forms, with a single cubic function being the most obvious alternative, though it becomes simple to analytically solve when it is approximated by a piecewise square function as we have used. The variable $v_M$ is the voltage across the terminals of the device, where during read it is a small signal voltage applied at the word line, and during write processes it is the difference of the voltages applied at the word and bit lines. When the device is set, $v_M$ is added to (1), and when the device is reset it is subtracted from (1). For unipolar devices, such as PCM, the absolute value of $v_M$ must be taken instead. The constants $a$, $b$, $c$ and $d$ are device-dependent values calculable using pre-determined switching characteristics, where the following two conditions must be satisfied for set:

$$t_s = \frac{2tan^{-1}\big[\frac{2ab}{\sqrt{4a|v_T|-b^2}}\big]}{\sqrt{4a|v_T|-b^2}}, \tag{3}$$

$$v_T = \frac{b^2(1-2a)}{4a}, \tag{4}$$

where switching time in (3) is derived by multiplying both sides of (1) by $dt$, dividing both sides by the quadratic term, and evaluating the definite integral from $x = -b$ to $x = 0$, and has the condition of $4av_T > b^2$ imposed upon it. Voltage threshold in (4) is derived by solving for $x$ at the zero-crossing of the time derivative of (1), and substituting this back into (1). These two equations can be solved simultaneously, making use of known experimental values for switching time and threshold to find the constants $a$ and $b$. This identical procedure can be repeated for reset to find $c$ and $d$.

The bracketed conductance term in (2) is much simpler to derive than (1). The logistic function is used where $g_s$ is the upper limit of conductance in the set (or crystalline for PCM) state, $g_r$ is the lower limit in the reset (or amorphous) state and $s$ is the switching-slope characteristic, the greater it is the steeper the switching slope. This form is chosen as it is a straightforward way to enable state-induced $x$ current (or conductance) switching.

Graphically, with respect to Fig. 1, $v_M$ has the effect of shifting $\frac{dx}{dt}$ upwards in set and downwards in reset. During set, if $v_M$ exceeds $v_T$ for a sufficiently long pulse width, then the position on the phase-plane is on the positive half of $\frac{dx}{dt}$. This causes $x$ to shift to the right. Once $x$ crosses the y-axis, the write voltage can be turned off as the state will tend to the right-hand point of stability, and the device has successfully been set. The same principle applies in reverse for reset, with the curve shifting downwards and thus requiring a higher voltage (which corresponds to the need for a higher programming current).

Specifically, a memristive system is mathematically defined as a pair of dynamical state equations $I = g(x,V)V$ and $\frac{dx}{dt} = f(x,V)$. These are in the same form as that presented in (1) and (2). The functional flow of the model is depicted in Fig. 2, where rather than fitting V-I characteristic curves to the model, they instead use empirical set, reset and conductance level data the device is intended to be used in.

With the operation of the RRAM model now elucidated, the next section will functionally use it in a series of simulations to demonstrate how it can optimize computational performance.

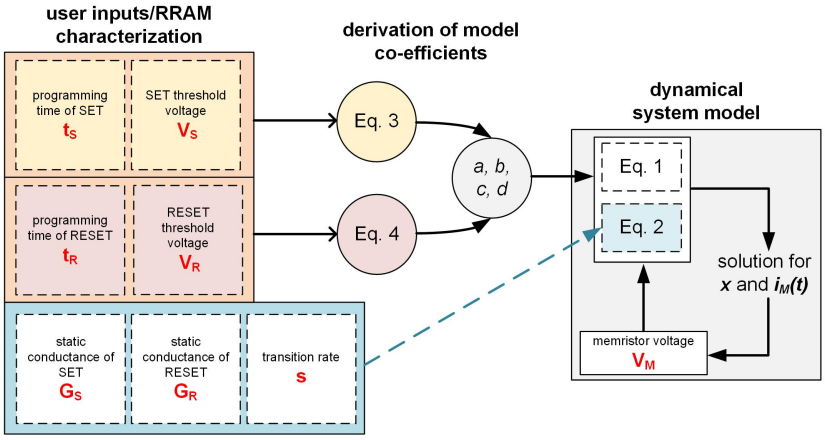



Fig. 2. Using behavioral parameters to solve for the dynamical system model of RRAM. Values for $t_S$, $V_S$, $t_R$ and $V_R$ are used to solve for $a$, $b$, $c$, and $d$ in (3)–(4), which are substituted into (1)–(2) along with $g_S$, $g_R$ and $s$.

## III. Simulation Results

### A. Phase Change Memory Model

We first validate a single unipolar PCM device model as against experimental data [21], and on in-house fabricated $Ge_2Sb_2Te_5$ materials in a conventional mushroom structure, using WTi as the bottom electrode. This device has been chosen as it is most widely commercially used. The I-V curve in Fig. 3 shows a readout performed at low bias (i.e., in the read region). In order to reach the set and reset programming regions, the bias is raised above the switching threshold $v_T$, which is the mechanism leading to phase-change. As shown in the I-V curve, by using the parameters in Table I our model is capable of successfully exhibiting realistic set and reset switching times and thresholds, and can thus be used in numerical and analytical circuit simulations with good reliability for the relevant regions of operation in single-bit mode. This model is not intended to retain accuracy in intermediary regions, and presumes volatility in such regions. Thus, for the purposes of large-scale inferential acceleration in crossbars we prefer to reduce computational complexity by entirely removing specificity in the modes of operation that are not meant to be used. We have additionally included model simulation data of phase-plane dependence on voltage in Fig. 4, as this information is not captured by the voltage sweep in Fig. 3(b).

Fig. 5 shows simulation results of a 3×3 crossbar array performed in Simulink and displays the current readout of a selected and unselected pair of PCM devices during both read and write processes as a function of time. During write, the unselected cell sees half the voltage as seen by the selected device due to the V/2 write scheme using values of 0.8V for 100ns to set and 1.25V for 10ns to reset. Therefore, the state of the unselected cell always reaches switching failure due to insufficient voltage (i.e., the dynamic route map from Fig. 1 does not translate vertically enough in order for $x$ to cross the y-axis). The state shifts back to the original stable position in accordance to (1), while the selected cell

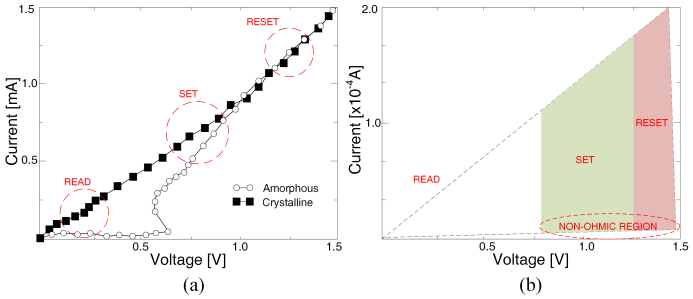



Fig. 3. PCM I-V curve (a) Experimental data adapted from [21] in the crystalline and amorphous states (b) presented PCM model simulated in MATLAB using parameters from Table I.

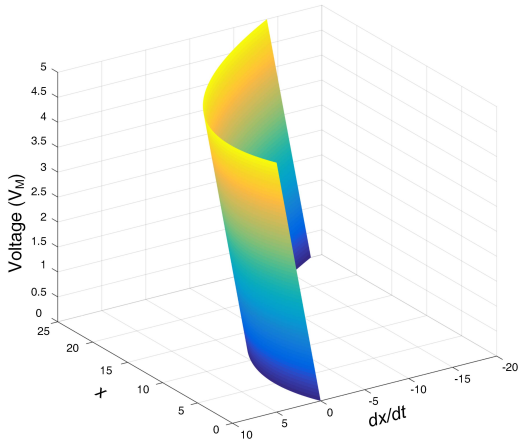
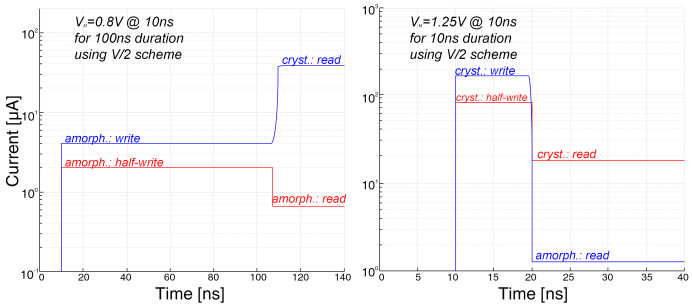



Fig. 4. Voltage-dependence of PCM phase plane during reset process.

fully switches. We note these simulations are performed under idealized conditions where no line losses occur in a small 3×3 crossbar. For our simple proof-of-concept simulations, this will not deviate from experimental work, though line resistance may need to be considered in larger arrays.




(a) (b)

Fig. 5. Model simulation, reading and writing to a pair of PCM devices. Relevant parameters are given in Table I (a) Set operation (b) Reset operation.

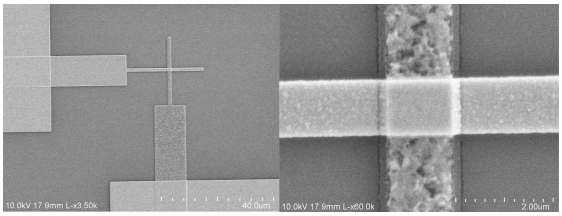


Fig. 6. W/WOx/Pd RRAM device on a 2 $\mu$m×2 $\mu$m crossbar structure shown in the SEM images.

### B. Valence Change Mechanism Model

The RRAM device we verify our model on is shown in Fig. 6. The fabrication process is as follows: 1. the bottom electrode is patterned by lithography; 2. tungsten is sputtered, 3. the surface of the tungsten is oxidized to get WOx, 4. perform lift-off, 5. pattern the top electrode by lithography, 6. evaporate Pd, and 7. perform lift-off. Table II summarizes the parameters that we use to successfully emulate digital switching in the device using the same process as described for PCM devices, this time without taking the absolute value of voltage in (1). The device is characterized in Table II across several modes of digital operation between set and reset where current is read out using a 0.2 V voltage pulse to measure the conductance, and these values are adopted into Eqs. (1) and (2) in order to derive their respective models using the provided measured characteristics. We note that transition rate $s$ has not been provided as this is not a physically characterized phenomena, and is instead used as a way to control the abruptness of the conductance shift.

## IV. Discussion

The RRAM model in (1) and (2) presents a small system of two dynamical state equations which are analytically simple to solve, and numerically computationally inexpensive to perform. In comparison to the most recent state-of-the-art physically based models [9], our model's average execution time for a current-based set process is only 5.8% of other RRAM models that are intended for analog-domain modeling. This was calculated by averaging execution time across 20 trials in MATLAB using the implicit trapezoidal Euler method

TABLE I
Model Parameterization [22]

| Symbol | Parameter | Value |
|---|---|---|
| $R_s$ | static resistance of set; $R_s = \frac{1}{g_s}$ | 7KΩ |
| $i_s$ | programming current of set | 600$\mu$A |
| $t_s$ | programming time of set | 100ns |
| $R_r$ | static resistance of reset; $R_r = \frac{1}{g_r}$ | 200KΩ |
| $i_r$ | programming current of reset | 1700$\mu$A |
| $t_r$ | programming time of reset | 10ns |
| $v_T$ | threshold voltage | 0.78V |
| $s$ | transition rate | 100 |

TABLE II
Device Parameters

| | $V_s$ (V) | $t_s$ (ms) | $G_s$ ($\mu$S) | $V_r$ (V) | $t_r$ (ms) | $G_r$ ($\mu$S) |
|---|---|---|---|---|---|---|
| VCM: $WO_x$ | 1.5 | 1.00 | 31.5 | -1 | 1.00 | 12.35 |
| | 2 | 0.10 | 32.55 | -1.5 | 0.10 | 9.75 |
| | 2.5 | 5.00 | 113.1 | -1.5 | 5.00 | 7.75 |
| T-Type GST | 1.6 | 1.00E-3 | 45.4 | 2 | 1.00E-6 | 0.897 |
| | 1.6 | 1.00E-3 | 38.0 | 2.2 | 1.00E-6 | 0.636 |
| | 1.6 | 1.00E-3 | 3.02 | 2.8 | 1.00E-6 | 0.434 |

for both cases. This substantial improvement in run time is to be expected, as our model has been optimized and simplified for speed, and not for physical accuracy across the analog spectrum of states. Therefore, our model would not be used as a substitute for physical models unless operating digitally, without the need for accurate predictions beyond $g_S$ and $g_R$.

Therefore, a more fair comparison is the VTEAM model, which depends less on physical characteristics in favor of generalizability [23]. In this case, the system of equations are of a relatively similar form and achieved an execution speed improvement of approximately 30%. However, the VTEAM model is substantially more accurate for analog domain characteristics as it is developed based on curve-fitting using gradient descent for optimization across a wide spectrum of values, whereas our behavioral model is derived analytically by using 7 parameters for switching (see (3) and (4) and Fig. 2), and is therefore restricted in accuracy to the information fed into the system of equations. As this model prioritizes simplicity and speed, it is not a substitute for other models that accurately capture empirical, analog and physical characteristics of RRAM, but it may be more desirable to use in large crossbar arrays where programming and read voltages are consistent, and each device is restricted to single-bit use. In addition, switching characteristic data tends to be more readily available than detailed V-I data at various driving frequencies which makes the co-efficients of (1) intuitive and straightforward to derive, without the need for any optimization procedures.

It is important to stress that, as a result of computational simplicity, this is a relatively unphysical model in comparison with prior state-of-the-art models. Physically based models are employed to generate accurate simulations when parameters are varied due to process variations. In the case of this model, however, any process variations that give rise to switching parameters must be implemented directly into switching time and thresholding parameters. Whilst Monte Carlo simulations

are possible, they will not be based on physical process and temperature variations, rather the distribution of switching times and threshold variation should be derived instead. If device variability is required, it would be preferable to use compact models that are accurate beyond single-bit operation unless large-scale simulations are being executed on limited computational resources. One final consideration of our model, in the form presented in (1), $\frac{dx}{dt}$ never perfectly reaches the x-axis and will only approach the limit indefinitely. Although convergence is guaranteed, there will be a large dependence on absolute or relative tolerances in SPICE simulations that determine how long it takes for a solution to be calculated.

## V. Conclusion

In this paper, we present a generalized model of resistive memory that captures only the behavioral characteristics of switching time, threshold, and transition time in the regions of operation intended for use in digital switching. We verified its operation on both PCM and VCM classes of RRAM. This simplification has shown an improvement in computational speed by approximately 20-fold over physically-driven models, and 30% over empirically-driven generalized models.

Models are only appropriate within their limit of intended use. As a note of caution, this model is inappropriate for implementing physical quantities. It is intended to capture digital switching characteristics that would aid behavioral EDA tools that are only concerned with gate-level parameters of bipolar and unipolar switching. It avoids expensive math functions that are typically required in descriptive models intended for analog applications, and achieves large computational benefits by assuming the RRAM devices are used as single-bit switches, as well as a very high degree of simplicity by automating the calculation of parameterization by taking only the known, measurable parameters of programming amplitude and time.

## Acknowledgements

This work was supported by iDataMap Corporation.

## References

[1] S. Yu, "Overview of resistive switching memory (RRAM) switching mechanism and device modeling", 2014 IEEE Int. Symp. Circuits and Syst. (ISCAS), Melbourne, VIC, Australia, pp. 2017–2020, June 2014.

[2] J. K. Eshraghian, S. M. Kang, S. Baek, G. Orchard, H. H. C. Iu and W. Lei, "Analog weights in ReRAM DNN accelerators", 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), Hsinchu, Taiwan, pp. 267–271, March 2019.

[3] J. K. Eshraghian, K.R. Cho, H.H.C. Iu, T. Fernando, N. Iannella, S.-M. Kang and K. Eshraghian, "Maximization of Crossbar Array Memory Using Fundamental Memristor Theory", IEEE Trans. Circuits and Syst. II: Exp. Briefs, vol. 64, no. 12, pp. 1402–1406, December 2017.

[4] J. Lee, J. K. Eshraghian, M. Jeong, F. Shan, H.H.C. Iu and K. Cho, "Nano-programmable logics based on double-layer anti-facing memristors", Journal of Nanoscience and Nanotechnology, vol. 19, no. 3, pp. 1295–1300, March 2019.

[5] O. Krestinskaya, K. N. Salama, A. P. James, "Analog backpropagation learning circuits for memristive crossbar neural networks", 2018 IEEE Int. Symp. on Circuits and Syst. (ISCAS), Florence, Italy, May 2018.

[6] J. K. Eshraghian, K. Cho, C. Zheng, M. Nam, H.H.C. Iu, W. Lei and K. Eshraghian, "Neuromorphic vision hybrid RRAM-CMOS architecture", IEEE Trans. on Very Large Scale Integration (VLSI) Syst., vol. 26, no. 12, pp. 2816–2829, May 2018.

[7] S. Baek, J. K. Eshraghian, S. H. Ahn, A. James and K. Cho, "A memristor-CMOS multiplier array for arithmetic pipelining", 2019 26th IEEE International Conf. Electronics, Circuits and Systems, pp. 735–738, November 2019.

[8] M. R. Azghadi, *et al.*, "CMOS and Memristive Hardware for Neuromorphic Computing", Advanced Intelligent Syst., *in press*.

[9] C. La Torre, A. F. Zurhelle, T. Breuer, R. Waser and S. Menzel, "Compact modeling of complementary switching in oxide-based ReRAM devices", IEEE Trans. on Electron-Devices, vol. 66, no. 3, pp. 1268–1275, January 2019.

[10] Q. Xia and J. J. Yang, "Memristive crossbar arrays for brain-inspired computing", Nat. Materials, vol. 18, no. 4, p. 309, April 2019.

[11] C. Li, *et al.*, "Long short-term memory networks in memristor crossbar arrays", Nat. Machine Intelligence, vol. 1, no. 1, p. 49, January 2019.

[12] F. Cai, *et al.*, "A fully integrated reprogrammable memristor-CMOS system for efficient multiply-accumulate operations", Nat. Electronics, vol. 2, no. 7, pp. 290–299, July 2019.

[13] C. Zheng, D. Yu, H.H.C. Iu, T. Fernando, T. Sun, J.K. Eshraghian and H. Guo, "A novel universal interface for constructing memory elements for circuit applications", IEEE Trans. Circuits and Syst. I: Reg. Papers, vol. 66, no. 12, pp. 4793–4806, September 2019.

[14] À. Ràk and G. Cserey, "Macromodeling of the memristor in SPICE", IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems, vol. 29, no. 4, pp. 632–636, March 2010.

[15] S. Benderli and T. A. Wey, "On SPICE macromodelling of $TiO_2$ memristors", Electronics Letters, vol. 45, no. 7, pp. 377-379, March 2009.

[16] J. K. Eshraghian, H. H. C. Iu, T. Fernando, D. Yu and Z. Li, "Modelling and characterization of dynamic behavior of coupled memristor circuits", 2016 IEEE Int. Symp. on Circuits and Syst. (ISCAS), pp. 690–693, May 2016.

[17] S. Menzel, S. Tappertzhofen, R. Waser and I. Valov, "Switching kinetics of electrochemical metallization memory cells", Physical Chemistry Chemical Physics, vol. 15, no. 18, pp. 6945–6952, 2013.

[18] A. Hardtdegen, C. La Torre, F. Cüppers, S. Menzel, R. Waser and S. Hoffman-Eifert, "Improved switching stability and the effect of an internal series resistor in $HfO_2$/$TiO_x$ bilayer ReRAM cells", IEEE Trans. on Electron Devices, vol. 65, no. 8, pp. 3229, 2018.

[19] Z. Jiang, Y. Wu, S. Yu, L. Yang, K. Song, Z. Karim and H.-P. Wong, "A compact model for metal-oxide resistive random access memory with experiment verification", IEEE Trans. Electron Devices, vol. 63, pp. 1884–1892, 2016.

[20] J. P. Strachan, A. C. Torrezan, F. Miao, M. D. Pickett, J. J. Yang, W. Yi, G. Medeiros-Ribeiro and R. S. Williams, "State dynamics and modeling of tantalum oxide memristors", IEEE Trans. Electron Devices, vol. 60, pp. 2194–2202, 2013.

[21] C. Lammie, O. Krestinskaya, A. James and M. R. Azghadi, "Variation-aware binarized memristive networks", arXiv preprint arXiv:1910.05920, October 2019.

[22] A. Pirovano, A. L. Lacaita, F. Pellizzer, S. A. Kostylev, A. Benvenuti and R. Bez, "Low-field amorphous state resistance and threshold voltage drift in chalcogenide materials", IEEE Trans. Electron Devices, vol. 51, no. 5, pp. 714–719, May 2004.

[23] H.-S. P. Wong, S. Raoux, S. B. Kim, J. Liang, J. P. Reifenberg, B. Rajendran *et al.*, "Phase Change Memory", Proc. of the IEEE, vol. 98, no. 12, pp. 2201–2227, October 2010.

[24] S. Kvatinsky, M. Ramadan, E. G. Friedman and A. Kolodny, "VTEAM: A general model for voltage-controlled memristors", IEEE Trans. Circuits and Syst. II: Exp. Briefs, vol. 62, no. 8, pp. 786–790, May 2015.

[25] F. Bedeschi, R. Fackenthal, C. Resta, E. M. Donze, M. Jagasivamani, E. C. Buda, *et al.*, "A bipolar-selected phase change memory featuring multi-level cell storage", IEEE J. of Solid-State Circuits, vol. 44, no. 1, pp. 217–227, January 2009.