Incorporating Variability of Resistive RAM in Circuit Simulations using the Stanford–PKU Model

John Reuben, *Member, IEEE*, Mehrdad Biglari, and Dietmar Fey

***Abstract*—Intrinsic variability observed in resistive-switching devices (cycle-to-cycle and device-to-device) is widely recognised as a major hurdle for widespread adoption of Resistive RAM technology. While physics-based models have been developed to accurately reproduce the resistive-switching behaviour, reproducing the observed variability behavior of a specific RRAM has not been studied. Without a properly fitted variability in the model, the simulation error introduced at the device-level propagates through circuit-level to system-level simulations in an unpredictable manner. In this work, we propose an algorithm to fit a certain amount of variability to an existing physics-based analytical model (Stanford-PKU model). The extent of variability exhibited by the device is fitted to the model in a manner agnostic to the cause of variability. Further, the model is modified to better reproduce the variations observed in a device. The model, fitted with variability can well reproduce cycle-to-cycle, as well as device-to-device variations. The significance of integrating variability into RRAM models is underscored using a sensing example.**

***Index Terms*—Resistive RAM (RRAM), physics-based models, cycle-to-cycle variability, device-to-device variability, Stanford model, memristor, sense amplifier, resistive-switching, 1T-1R**

I. Introduction

RESISTIVE RAMs (RRAMs) are two terminal devices capable of changing their resistance in response to voltage stress. Initially RRAM was researched as an emerging Non-Volatile Memory (NVM), and, recently, RRAM has also extended its influence beyond memory to logic [1] and computing [2]. Consequently, research in RRAM-based memories and RRAM-based computing circuits are very active and ever increasing [3]. Variability in RRAM devices is widely recognised as a serious concern hindering its industry adoption and commercialization [4]–[8]. In RRAM technology, variability in switching behavior can be temporal (cycle-to-cycle) and spatial (device-to-device). The variability in the switching behaviour manifests as variation in the programmed resistance of the device (low resistance state, LRS and high resistance state, HRS) and variation in the voltage at which the device switches ($HRS \rightarrow LRS$ occurring at V_{SET} and vice versa at V_{RESET}). It is evident that such stochasticity jeopardizes the operation of memory ($e.g.$ large HRS variation can result in erroneous read-out by the sense amplifier of the memory array) and logic ($e.g.$ large V_{SET} variation can result in failure of NOR gates implemented in memory array [9]). While variations have been utilized for some applications like stochastic learning and physical unclonable functions [10]–[12], they remain a hurdle for memory and other deterministic computing applications.

In the past, researchers attempted to eradicate variability in resistive switching behavior by device engineering, $e.g.$ the introduction of an additional Al_2O_3 layer [14], Germanium layer [15], $TiOx$ layer [16] and other such techniques. The subject of variability and its causes is still a matter of intense research. However, past efforts in device engineering seem to suggest that variability can be reduced, but not completely eliminated. There is a certain amount of variability which is intrinsic to Resistive RAM, $i.e.$ variability is due to the stochastic nature of formation and rupture of the conductive filament [5], [6], [17]. While variability needs to be mitigated at the device level, the exhibited variability of a device needs to be included in RRAM models so that circuit/system designer can observe in their simulations a behaviour close to reality. If the variability observed in a device is not included in its model, the error introduced in the simulation of a single RRAM cell propagates through circuit level (memory array) to system-level simulations. As noted in [18], the impact of device-level variations on realized circuits and systems can not be predicted in a deterministic manner, resulting in an estimate far from reality. Without incorporating variations, there cannot be a proper assessment of the functionality and yield of RRAM-based ICs, which will result in a series of expensive trial-and-errors. Pessimistic design approaches which allocate huge safety margins to accommodate variability are not recommended since they sacrifice design properties like energy, delay, and area. Consequently, there is an exigent

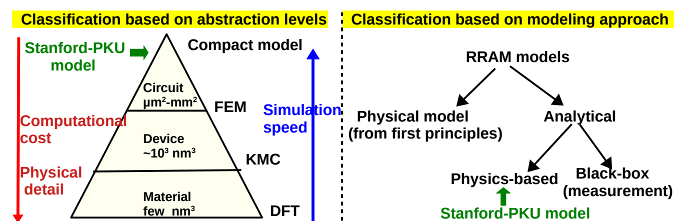


Fig. 1: Based on degree of detail, a model for RRAM can range from atomistic (Density Functional Theory (DFT)) to Kinetic Monte Carlo (KMC) to Finite Element Method (FEM) to compact model [13]. From another perspective, a model for RRAM can be derived from first principles or simplifications.

Authors are with Chair of Computer Science 3 – Computer Architecture, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 91058 Erlangen, Germany. (email:johnreubenp@gmail.com, mehrdad.biglari@fau.de, dietmar.fey@fau.de)

Manuscript received xx 2020; revised yy, 2020 . Copyright (c) 2020 IEEE. Personal use of this material is permitted. However, permission to use this material for any other other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication. The final version of record is available at http://dx.doi.org/10.1109/TNANO.2020.3004666
Copyright (c) 2020 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

need to incorporate variability precisely into RRAM models for reliable circuit/system evaluations.

RRAM (memristor[1]) modeling has made significant progress in recent years, and, at the time of this writing, 15-20 distinct models have been reported in literature. A classification of models based on the degree of detail is presented in [13] (Fig. 1). While atomistic models capture intricate details like ion/atom diffusion and migration mechanisms at atomistic scale (few nm^3), compact models capture macroscopic details like geometry of the conductive filament and temperature. Based on the modeling approach, they can be classified as– physical models and analytical models (Fig. 1). Physical models try to model resistive switching behavior from first principles, $i.e.$ by modeling the fundamental cause for change in resistance, from the perspective of device-physics. On the other hand, analytical models are formulated as mathematical equations, which match the device's experimentally-observed behaviour. The physical principles approach and analytical approach are two extremes, and in reality, a model for RRAM is a compromise between these two extremes [19]. Analytical approaches can be further classified as being either physics-based or black-box. In a physics-based analytical model, the physics of resistive switching is modeled by appropriate equations $e.g.$ a physics-based analytical model may simplify resistive switching to the formation and rupture of conductive filament under voltage stress. In a black-box model, a measurement approach is followed and the approach is agnostic to the device structure or switching mechanism $e.g.$ a device is subjected to different stimulus and its response studied; then mathematical equations are formulated which obey this behavior (VTEAM model [20] is a classic example of black-box approach).

In addition to the aforementioned classification schemes, some models are specifically proposed for a RRAM switching mechanism. Examples of these would be [21] and [22] that specifically model Valence Change Mechanism (VCM) and Electrochemical Metallization (ECM) devices, respectively. More details on RRAM models and their classifications can be found in [13], [17], [19], [23]–[26]. Since the focus of this work is variability modeling, we discuss a few models which have the capability to reproduce variability (in section V) to make our contribution clear.

Physics-based analytical models have proven to be more accurate in reproducing the resistive switching behaviour when compared to black-box analytical models [13]. Further, some physics-based models have the capability to mimic the variability observed in resistive-switching behaviour. However, the reliability of these models comes at a cost– the process to fit a physics-based analytical model to a device is strenuous. The authors in [27] have proposed an algorithm to fit different devices (HfO_x, SiO_x, TaO_x) to a physics-based analytical model. However, to the best of our knowledge, there is no work reported on fitting experimentally observed variations into a RRAM model.

In this work, we focus on incorporating the variability behavior of RRAM devices in circuit simulations using a physics-based model. Our main contributions are:

1) We present a methodology to fit device variations ($i.e.$ measured from fabricated RRAM devices) to a well-established physics-based model presented in [28].
2) We propose modifications to the model that enable better fitting of variations to fabricated devices.

Our emphasis in this work is modeling the quantum of variability, $i.e.$ analysing the degree of variability in LRS, HRS, V_{SET}, and V_{RESET} cycle-to-cycle and device-to-device; and including it in the model. Therefore, our approach is agnostic to the cause of variability[2], and, the inherent correlation of variability observed in devices (V_{SET}, V_{RESET} exhibited correlation between consecutive cycles [29] and a similar phenomenon was observed in HRS [30]). We pursued this approach due to the following reasons:

1) It is the amount of variations that decides correct functionality at the circuit and system-level.
2) Incorporating variability in a physics-based approach by modeling its cause will increase the computational complexity, making the model inappropriate for circuit/system-level simulations where numerous devices are interconnected.

For example, the cause of variability at HRS is due to variation of the number of particles in the narrowest current-controlling part of the filament [31], while the cause of variability at LRS is due to variation in the morphology of the filament, $i.e.$, it's shape varies from cycle-to-cycle [13]. Therefore, to model variability at HRS and LRS from its cause will increase the simulation time drastically since the resolution needed will be that of an atomistic or Kinetic Monte Carlo model for RRAM. It must be noted that our approach still models resistive-switching in a physics-based manner and only the variations are modeled in a black-box approach. Therefore our approach retains the low computational complexity of a compact model and reliability of a physics-based model.

The remaining of this paper is organized as follows. Section II describes the Stanford-PKU model and its capabilities to model variations. Section III describes our algorithm to fit the model to exhibit cycle-to-cycle variations. We first fit the Stanford-PKU model to exhibit the variations observed in devices manufactured in a process from IHP[3] and analyse the error in fitting (Section III-A). In Section III-B, we improve the Stanford-PKU model to minimize the error in fitting. The following section (Section IV-A) describes how device-to-device variability can be incorporated in simulation by considering a 1T1R memory array. We also discuss the implications of device variability at the circuit level by examining the design of a sense amplifier for the memory array. Section V discusses our contribution in light of other works and Section VI concludes our contribution.

[1]The terms memristor and RRAM are used interchangeably by researchers in this field.

[2]The cause of variability is being studied by device researchers and is beyond the scope of this research.

[3]Innovations for High Performance Microelectronics, Germany

II. Stanford-PKU RRAM model and variability

A. Stanford-PKU model in a nutshell

The Stanford-PKU is a physics-based (analytical), compact model developed for metal oxide bipolar RRAMs [28], [32]–[34]. The model was well characterized on HfO_2 and HfO_x/TiO_x bilayer devices [28]. This model simplifies the resistive switching behaviour to the growth ($HRS \rightarrow LRS$) and rupture ($LRS \rightarrow HRS$) of a single dominant conductive filament. The gap distance, g (between the tip of the filament and the counter electrode) is the crucial parameter which determines the resistive state. The parameter g is programmable between gap_{min} and gap_{max} with the device being in HRS at gap_{max} and in LRS at gap_{min}. Fig. 2 lists the equations governing the resistive switching process. The key equation (shaded grey in Fig. 2) describes the current through the RRAM (I) as a function of voltage across it (V) and the gap in the conductive filament (g). The current has an exponential dependence on g, which together with hyperbolic dependence on the V, implements the sudden increase (or decrease) of the current resulting in a transition to LRS (or HRS). The reader is referred to [28] for an elaborate description of the model.

$$g = \int \left(\frac{dg}{dt} + \delta_g \times \chi(t) \right) . dt$$

$$\delta_g(T) = \frac{\delta_g^0}{1 + \exp\left[\frac{(T_{CRIT} - T)}{T_{SMTH}} \right]}$$

$$I = I_o \times \exp\left(-\frac{g}{g_0} \right) \times \sinh\left(\frac{V}{V_0} \right)$$

$$\frac{dg}{dt} = -\upsilon_0 \times \exp\left(-\frac{E_a}{kT} \right) \times \sinh\left(\gamma \times \frac{a_0}{t_{ox}} \times \frac{qV}{kT} \right)$$

$$T = T_0 + V \times I \times R_{TH}$$

$$\gamma = \gamma_0 - \beta \times g^3$$

Fig. 2: Equations of Stanford-PKU RRAM model: Cycle-to-cycle variation is introduced by multiplying δ_g by a Gaussian noise χ(t) and adding to the instantaneous value of the gap, g. $\delta_g(T)$ is the variation in the gap as a function of current temperature, T (shaded yellow).

The parameters E_a (activation energy), a_0 (atomic spacing of the switching oxide), t_{ox} (thickness of the switching oxide), T_0 (environment temperature) and R_{TH} (thermal resistance) are determined by device structure, material properties and test environment. Let them be called 'process parameters' since they are dictated by the fabrication aspects of the device. The parameters I_0, g_0, V_0, υ_0, γ_0 and β are called 'switching parameters' or 'conductance parameters' by the model developers, and they determine the median switching characteristics. They are used to tune the median switching characteristics to a specific RRAM, as elaborated in the fitting algorithm proposed for this model in [27]. δ_g^0 is the fitting parameter for variations in the gap. T_{CRIT} denotes the threshold temperature, above which significant variations in the gap occurs and T_{SMTH} is the variations smoothing parameter [28]. These three parameters (δ_g^0, T_{CRIT}, T_{SMTH}) are related to variations in resistive-switching and they can be used to incorporate the experimentally observed variability in the Stanford-PKU model.

B. Variability

We first define terms to quantify variability. Variability in V_{SET}/ V_{RESET} is quantified as σ_{SET}/σ_{RESET}, the standard deviation from the mean, μ_{SET}/μ_{RESET}. Variability in HRS/LRS is also expressed as standard deviation, σ_{HRS}/σ_{LRS} from the mean, μ_{HRS}/μ_{LRS}. While the V_{SET} and V_{RESET} voltages of RRAMs are almost of the same order, the HRS can be up to three orders of magnitude higher than LRS. Therefore, the resistance variability is normalised by mean resistance and quantified by Co-efficient of Variation (CV), σ_{LRS}/μ_{LRS} and σ_{HRS}/μ_{HRS}. Stanford-PKU model [28] has the capability to reproduce variability, and, the modeled variability can be configured using the parameters: δ_g^0, T_{CRIT}, T_{SMTH}. Although the model can be configured to introduce variability in simulation, the developers of the model [28] do not suggest any steps or procedure to fit certain amount of variability (σ_{HRS}, σ_{SET} etc) to their model. To introduce a particular amount of variability in LRS, HRS, V_{SET}, and V_{RESET}, how should the parameters δ_g^0, T_{CRIT}, T_{SMTH} be tuned? As can be deciphered from the equations in Fig. 2, the three parameters together decide the variation in the gap, $\delta_g(T)$, which is multiplied by a random Gaussian noise χ(t), and added to the instantaneous value of the gap, g. Since the gap g is the key state variable, variation in g affects all aspects of resistive-switching – threshold voltage at which the device switches and also the programmed resistive states. The mean switching characteristics of a specific RRAM were fitted to the Stanford-PKU model in [27]. This was done by varying the switching(conductance) parameters, one at a time, and the predominant role of each parameter in the switching process was analyzed and tuned accordingly [27]. However, such an approach could not be used to fit variations to the model because

1) The three parameters together decide variability and they are fused, $i.e.$ it was not possible to find a direct relation between one of these parameters and one facet of variability ($e.g.$ effect of δ_g^0 on σ_{HRS} or effect of T_{CRIT} on σ_{SET}).
2) The three parameters have a statistical effect on resistive-switching, $i.e.$ it requires numerous simulations of resistive switching to determine the effect of the parameter set (δ_g^0, T_{CRIT}, T_{SMTH}) on (σ_{HRS}, σ_{LRS}, σ_{SET}, σ_{RESET})

Therefore, fitting variability is not straightforward and our purpose in the first part of this work is to formulate a methodology to fit experimentally observed variations into the model, $i.e.$ given ($\sigma_{HRS}, \sigma_{LRS}, \sigma_{SET}, \sigma_{RESET}$) from RRAM characterization experiments, our goal is to find (δ_g^0, T_{CRIT}, T_{SMTH}) which will produce that variability during resistive-switching simulations.

III. Variation fitting methodology

A. Fitting variations to the Stanford-PKU model

We propose a variability fitting algorithm (Fig. 4) which is generic and can be applied for any RRAM device. We shall elucidate the variability fitting algorithm of Fig. 4 using the 1T1R devices manufactured at IHP (lower left corner

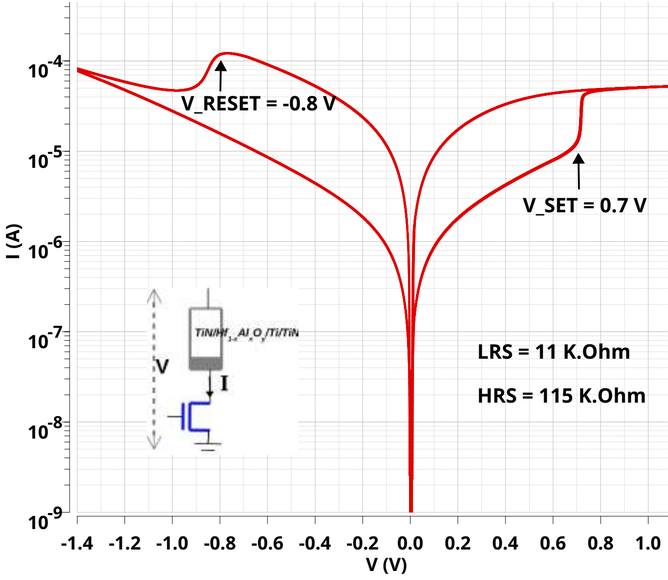


Fig. 3: Stanford-PKU model fitted to mean characteristics' of IHP's RRAM using the fitting algorithm presented in [27]. Fitting parameters: I_0 = 1.55e^{-4}, g_0 = 0.25 e^{-9}, V_0 = 0.325, υ_0 = 1.5, γ_0 = 24.25, β = 10, t_{ox}= 6 nm, gap_{min} = 0.01 nm, gap_{max} = 1.1 nm, E_a = 0.6 eV, a_0 = 2.5e^{-10}, R_{TH} = 2100.

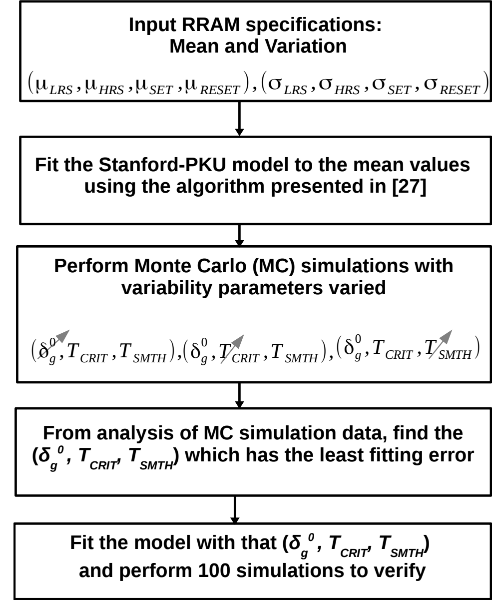


Fig. 4: Generic algorithm to fit variations observed in a RRAM to Stanford-PKU model.

of Fig. 3). The 1T1R is constituted by NMOS transistor manufactured in IHP's 250 nm CMOS technology, whose drain is connected to the bottom electrode of the RRAM. The RRAM is a $TiN/Hf_{1-x}Al_xO_y/Ti/TiN$ stack integrated on the metal line 2 of the CMOS process (Fig. 3). The median switching characteristics are: μ_{SET} = 0.7 V, μ_{RESET}= -0.8 V, μ_{LRS} = 11 KΩ and μ_{HRS} = 115 KΩ. The $SET/RESET$ voltage variability of IHP's RRAMs are extracted from [35]: σ_{SET} = 100 mV and σ_{RESET}= 75.3 mV. Since variability at LRS/HRS is not reported in [35], we considered a variability, (σ_{LRS}/μ_{LRS}) of 12.6 % and (σ_{HRS}/μ_{HRS}) of 20.9 %, which is the statistically reported variability for a similar HfO_x device [5]. This translates to σ_{LRS} of 1.4 KΩ and σ_{HRS} of 24 KΩ[4]. To fit the model with these variations, we first fit the mean characteristics of the device. This is achieved by tuning (I_0, g_0, V_0, υ_0, γ_0, β) to match median switching characteristics, as elaborated in [27]. The I-V curve produced by the fitted model matches the mean switching characteristics (Fig. 3). As observed in section II-B, the effect of the triplet (δ_g^0, T_{CRIT}, T_{SMTH}) on resistive switching behaviour is statistical. Numerous simulations with a particular (δ_g^0, T_{CRIT}, T_{SMTH}) needs to be performed, and the corresponding standard deviation σ_{HRS}, σ_{LRS}, σ_{SET} and σ_{RESET} needs to be calculated. $e.g$ the standard deviation in HRS is,

$$\sigma_{HRS} = \sqrt{\left(\frac{1}{N-1}\right)\sum_{i=1}^{N}(HRS_i - \mu_{HRS})^2} \tag{1}$$

[4]Both normal and log-normal distribution are used in literature for reporting the variability in resistive states [12], [31], [36], [37]. While we fit normally distributed resistive states here, the same procedure can be applied to fit the model to a log-normal distribution of HRS and LRS (Appendix A)

where HRS_i denotes the HRS to which the device gets programmed in the ith cycle for a particular triplet (we call the set (δ_g^0, T_{CRIT}, T_{SMTH}) a triplet). μ_{HRS} is constant for any triplet. Therefore, we resorted to Monte Carlo simulations.

We varied each of the three variability parameters around their default values mentioned in the model release [28]. δ_g^0 was varied from 0.5$\times$10^{-3} to 7$\times$10^{-3}, in steps of 0.5$\times$10^{-3}. T_{CRIT} and T_{SMTH} were varied from 100 to 1000, in steps of 150. This corresponds to 686 triplets. Each triplet was subject to 100 resistive switching cycles (N = 100 in Eq. 1). We chose N = 100 since repeated sets of 100 resistive switching cycles resulted in almost the same standard deviation in (σ_{HRS}, σ_{LRS}, σ_{SET}, σ_{RESET}) for a particular triplet. For each cycle, HRS and LRS values were measured with the read voltage of 200 mV. For V_{SET} and V_{RESET} values, we observed the internal state of the RRAM which is modeled by the filament gap (g). During switching operations, this gap varies between gap_{min} (corresponding to LRS) and gap_{max} (corresponding to HRS). V_{SET} and V_{RESET} were extracted by capturing the voltage at which the filament gap crosses a certain threshold in the direction of LRS and HRS, respectively (0.9 nm and 0.75 nm). Simulations were performed in Cadence ADE using Spetre simulator. The four standard deviations were analysed for the 686 triplets in MATLAB, and the triplet which minimizes the error was chosen as the fitting parameter.

Based on the analysis of the data obtained from Monte Carlo simulations, the following insights were obtained. Although (δ_g^0, T_{CRIT}, T_{SMTH}) together decide the amount of variability, their effect on different facets of variability was of different degrees. δ_g^0 had the highest effect on the amount of variability, followed by T_{CRIT}, while T_{SMTH} had the least effect on variability. Since multi-dimensional plots obscure analysis, we analyzed the effect of the triplet on variability, in two sets of plots. In the first set of plots (Fig. 5), we analyse the effect

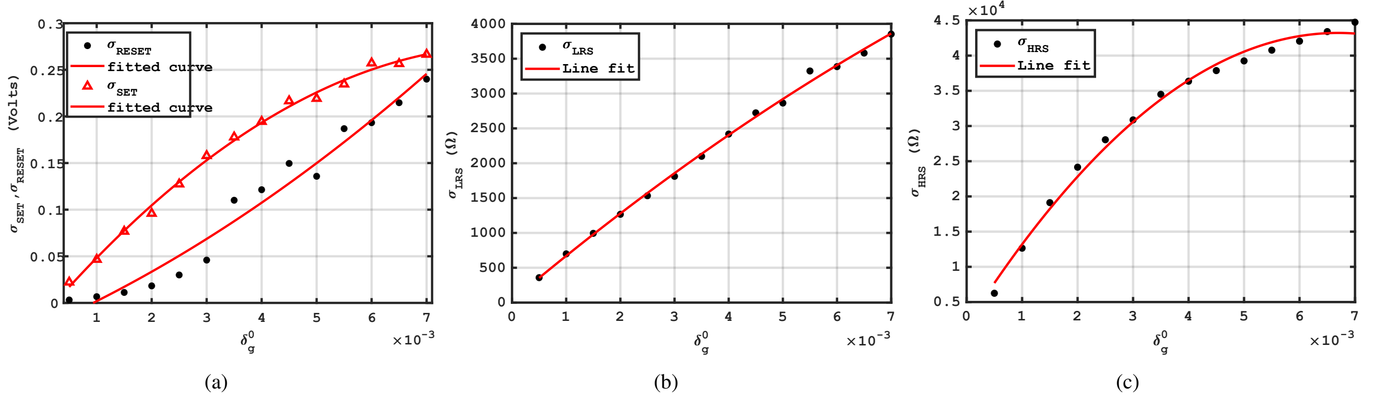


Fig. 5: Effect of δ_g^0 on (σ_{HRS}, σ_{LRS}, σ_{SET}, σ_{RESET}) for a constant (T_{CRIT} = 400, T_{SMTH} = 400) and N = 100. δ_g^0 clearly increases variability in all facets of resistive switching behaviour. (a) δ_g^0 affects σ_{SET} and σ_{RESET} in the same manner, but to different extent (b) δ_g^0's effect on σ_{LRS} is almost linear (c) δ_g^0's effect on σ_{HRS} is also linear, except for a saturating effect at high δ_g^0.

of δ_g^0 (T_{CRIT} and T_{SMTH} were fixed) on variability since its effect was significantly higher ($\sim 50\times$) compared to the effect of T_{CRIT} and T_{SMTH}. In the second set of plots (Fig. 6), we analyse the effect of T_{CRIT} and T_{SMTH} on variability, simultaneously (with δ_g^0 fixed).

As is evident from Fig. 5, δ_g^0 affects all aspects of variability in a linear manner. However, the extent to which it affects each variability is different. For example, δ_g^0 of 3×10^{-3} results in a σ_{SET} of 150 mV, while the same δ_g^0 results in a σ_{RESET} of 50 mV. This will be a disadvantage while fitting a certain amount of variability if the experimental variability does not follow the same pattern. Such a disparity is also observed in Fig. 6, where σ_{SET} is much higher than σ_{RESET} for the same (T_{CRIT}, T_{SMTH}). Therefore, error is inevitable while fitting variations, $i.e.$ it may not be possible to find a triplet which perfectly satisfies (σ_{HRS}, σ_{LRS}, σ_{SET}, σ_{RESET}) of a particular RRAM. Based on the results of Monte Carlo simulations, we found the triplet that satisfied the required variations in an over-cautious approach. If (σ_{HRS} = 24 KΩ, σ_{LRS} = 1.4 K Ω, σ_{SET} = 100 mV, σ_{RESET} = 75 mV) is the required variability to be fitted, we found the triplet which satisfied ($\sigma_{HRS} \geq$ 24 KΩ, $\sigma_{LRS} \geq$ 1.4 K Ω, $\sigma_{SET} \geq$ 100 mV, $\sigma_{RESET} \geq$ 75 mV). In other words, the accuracy of the fitting was compromised to fit the minimum required variability in all the four aspects. This was because, as depicted in Fig. 5 and Fig. 6, the triplet affected each variability in (σ_{HRS}, σ_{LRS}, σ_{SET}, σ_{RESET}) to different degrees. The triplet (δ_g^0 = 3×10^{-3}, T_{CRIT} = 400, T_{SMTH} = 550) which results in a variability (σ_{HRS} = 31.47 KΩ, σ_{LRS} = 1.87 KΩ, σ_{SET} = 163 mV, σ_{RESET} = 102 mV) is the most reasonable fitting that can be obtained in this situation. If we define variability fitting error with equal weights to all errors, the error in fitting is,

$$\Delta = \frac{\Delta_{HRS} + \Delta_{LRS} + \Delta_{SET} + \Delta_{RESET}}{4} \tag{2}$$

where

$$\Delta_{HRS} = \frac{\sigma_{HRS}^{fitted} - \sigma_{HRS}^{device}}{\sigma_{HRS}^{device}} \tag{3}$$

For the present fitting, (Δ_{HRS} = 31.12 %, Δ_{LRS} = 33.5 %, Δ_{SET} = 63 %, Δ_{RESET} = 36 %), resulting in Δ of 41 %.

B. Introducing double region δ_g^0 to the Stanford-PKU model

While fitting variations (observed in IHP devices) to the Stanford-PKU model, we faced a conflict – the disparity between σ_{SET} and σ_{RESET} introduced by the model for a particular triplet (see Fig. 5-(a) and Fig. 6). This is because variability in SET and $RESET$ operations are modeled by the same mechanism (adding stochasticity to the gap, g) using the same parameters (δ_g^0, T_{CRIT}, T_{SMTH}). This reduces the flexibility to fit it to devices with different variation during SET and $RESET$ processes. To enable better fitting of variations to different devices, modifications were necessary. The Stanford-PKU model, as presented in [28], is flexible yet stable. This made the model capable of being accurately fitted to the mean switching characteristics of different RRAMs (HfO_x, SiO_x, TaO_x) [27]. Therefore, we modified only the variability part of the model (the deterministic part of the model which corresponds to the mean resistive-switching behaviour was unaltered) to minimize the error in fitting variations. To this end, we introduced two different δ_g^0 – one for positive polarity ($P_\delta_g^0$) and one for negative polarity ($N_\delta_g^0$) of the voltage across the RRAM. This adds a new degree of flexibility while fitting experimentally observed variations, which will usually have some disparity. Following the procedure of Section III-A, we varied the set ($P_\delta_g^0$, $N_\delta_g^0$, T_{CRIT}, T_{SMTH}) and analysed its effect on different aspects of variability. As evident from the analysis of Section III-A, $P_\delta_g^0$ and $N_\delta_g^0$ had the prominent effect on variability and, therefore, we analyse their effect meticulously. Fig. 7-(a)-(d) depicts the effect of $N_\delta_g^0$ on all the four aspects of variability, and each curve in the plot corresponds to a particular $P_\delta_g^0$. σ_{LRS} is strongly influenced by $P_\delta_g^0$. Similarly, σ_{HRS} is a strong function of $N_\delta_g^0$ since at higher $N_\delta_g^0$, different $P_\delta_g^0$ produce almost the same σ_{HRS}. σ_{SET} has a clear linear dependence on both $P_\delta_g^0$ and $N_\delta_g^0$, while we could not identify any clear difference between σ_{RESET}'s dependence

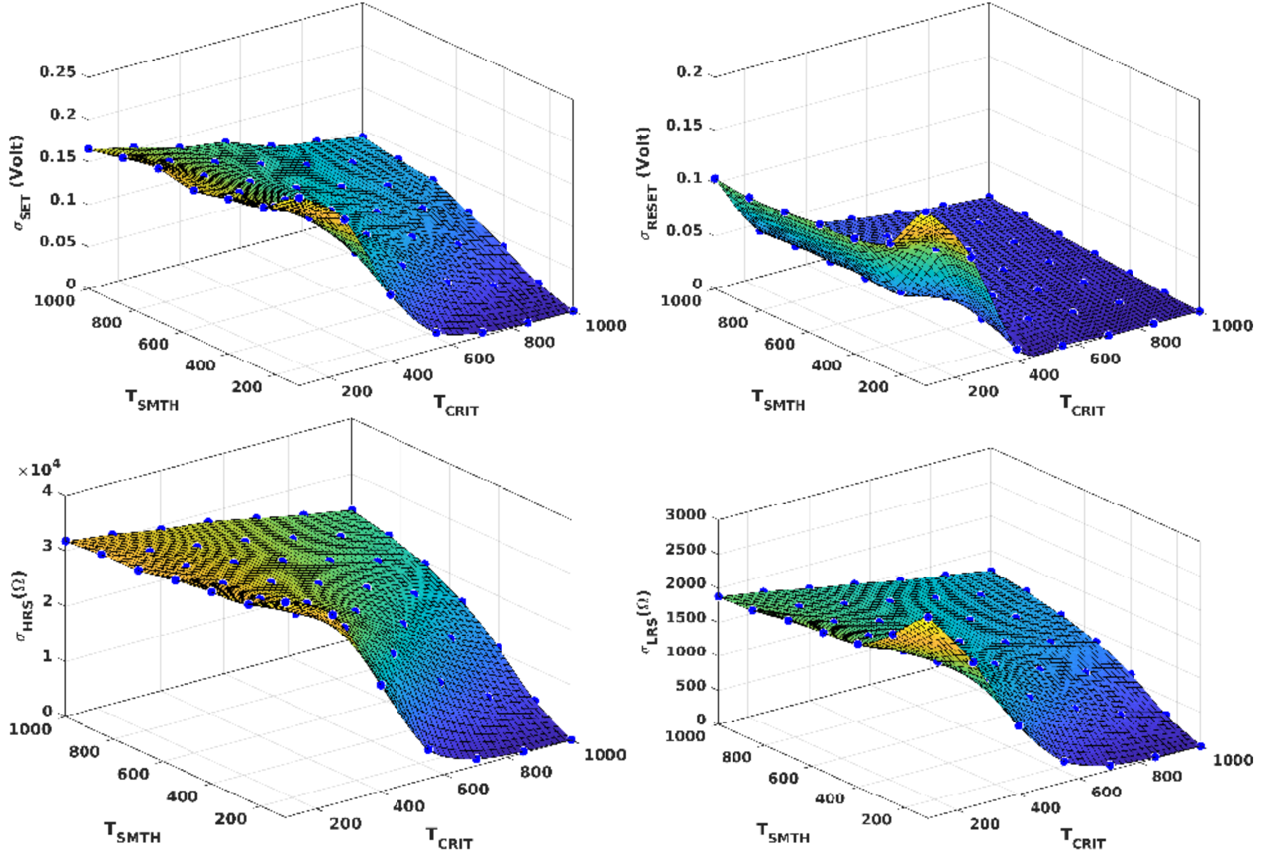


Fig. 6: Effect of T_{CRIT} and T_{SMTH} on (σ_{HRS}, σ_{LRS}, σ_{SET}, σ_{RESET}) for a constant δ_g^0 of 2.5×10^{-3} . No clear trend could be observed, except that the variation is uniformly high when T_{CRIT} is low.

on $P_\delta_g^0$ and $N_\delta_g^0$. However, it is clear that by introducing another dimension to δ_g^0, it is possible to fit variability more efficiently, $e.g$ even if one parameter $N_\delta_g^0$ has to be fixed at some value to satisfy a certain amount of σ_{RESET}, $P_\delta_g^0$ can be varied to satisfy the required variability in σ_{SET}. This was not possible in the fitting presented in section III-A, where $P_\delta_g^0$ was always equal to $N_\delta_g^0$. Based on the analysis of Monte Carlo simulation data, the fitting which minimized error was ($P_\delta_g^0$ = 2.5×10^{-3}, $N_\delta_g^0$ = 3.5×10^{-3} T_{CRIT} = 400, T_{SMTH} = 250). This results in a variability (σ_{HRS} = 31.2 KΩ, σ_{LRS} = 1.42 K Ω, σ_{SET} = 147 mV, σ_{RESET} = 97 mV). The fitting error was (Δ_{HRS} = 30 %, Δ_{LRS} = 1.4 %, Δ_{SET} = 47 %, Δ_{RESET} = 29.86 %), resulting in Δ of 27 %. We performed 100 cycles of resistive-switching with the Stanford-PKU model fitted with this variability parameters. Fig. 7-(e) depicts the variability observed in resistive states and Fig. 7-(f) the variability observed in $SET/RESET$ voltages.

IV. Significance of Modeling Variability: Effects at the Circuit Level

A. Device-to-device variability

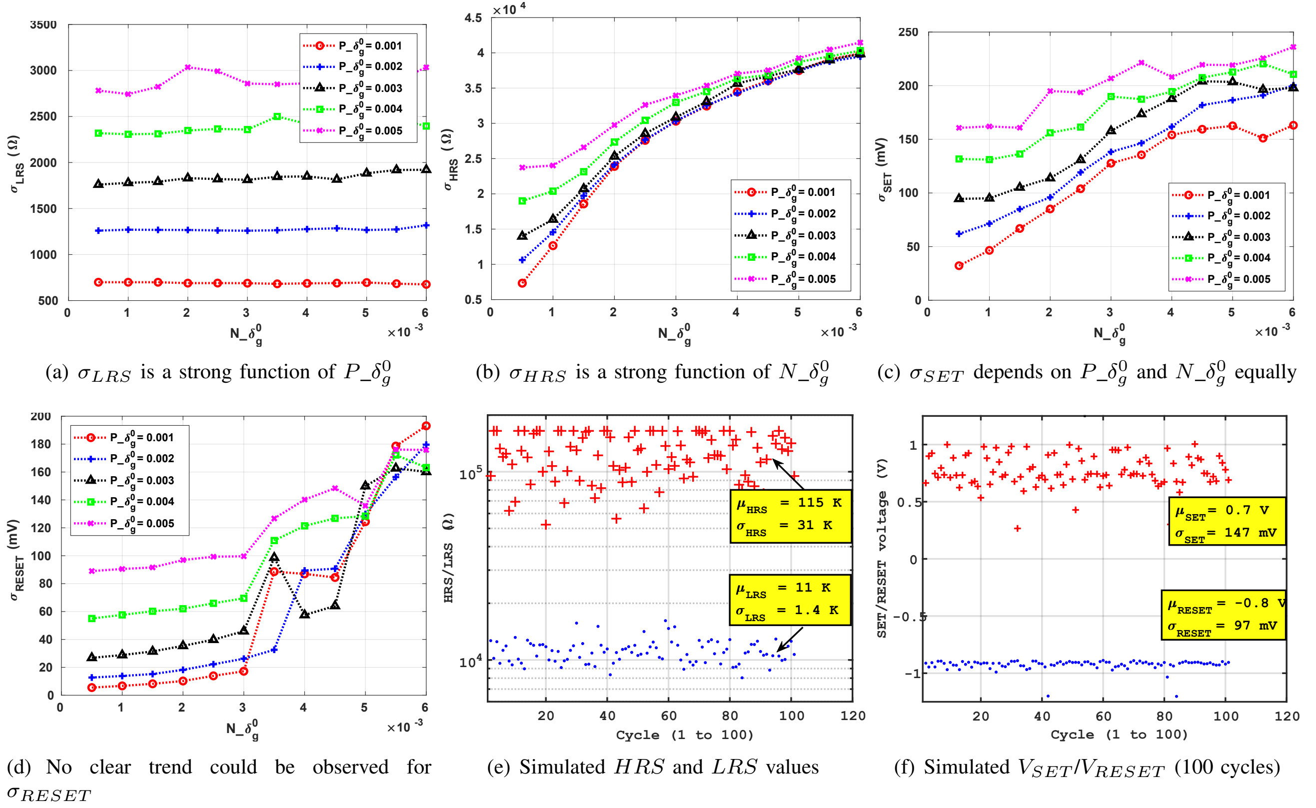
So far, we focused on incorporating variability into a single RRAM cell. As plotted (Fig. 7-(e),(f)), we are able to fit variations in cycle-to-cycle simulations. When RRAM cells are fabricated in an array, the cells exhibit device-to-device variations, $i.e.$ when observed in a common time window, there is disparity in resistive switching behaviour of the cells. Origin of device-to-device variability is attributed to discrepancies in the fabrication processes such as variation in the switching oxide thickness, surface roughness of the electrodes, etching damages, etc., as well as the lack of precise control over the defect generation and filament formation during the 'forming' step of a pristine device [5]. As noted in section II-B, variability is included by adding a stochastic part (δ_g multiplied with $\chi(t)$) to the instantaneous value of the gap, g (the deterministic part).

$$g = \int \left(\frac{dg}{dt} + \delta_g \cdot \chi(t) \right) \quad (4)$$

where $\chi(t)$ is Gaussian distribution. In Verilog-A, $\chi(t)$ is generated by the command

$$\chi(t) = \$rdist_normal(seed, \mu, \sigma) \quad (5)$$

where $seed$ is an integer which is used initialize the process of generating random numbers. Mean, μ was set to 0 and standard deviation, σ was set to 1 in the Stanford-PKU model, while $seed$ is a random integer chosen at the start of simulation. This implies that choosing a different seed can produce a different resistive switching behaviour, although the standard deviation (σ_{SET},σ_{HRS} etc) for a large number of cycles will be the same for a particular (δ_g^0, T_{CRIT}, T_{SMTH}). Therefore, two RRAM cells (simulated with different $seeds$) will have different ($V_{SET}, V_{RESET}, HRS, LRS$) in the same cycle.



(a) σ_{LRS} is a strong function of $P_\delta_g^0$

(b) σ_{HRS} is a strong function of $N_\delta_g^0$

(c) σ_{SET} depends on $P_\delta_g^0$ and $N_\delta_g^0$ equally

(d) No clear trend could be observed for σ_{RESET}

(e) Simulated HRS and LRS values

(f) Simulated V_{SET}/V_{RESET} (100 cycles)

Fig. 7: Effect of $P_\delta_g^0$ and $N_\delta_g^0$ on (σ_{HRS}, σ_{LRS}, σ_{SET}, σ_{RESET}) for a constant (T_{CRIT} = 400, T_{SMTH} = 400).

Thus, by using different seeds for different cells, device-to-device variations can be incorporated in simulations. In other words, the RRAM model incorporated with a certain amount of cycle-to-cycle variation will exhibit similar device-to-device variations, when observed in a specific time window. Fig. 9-(a) depicts the cycle-to-cycle variation (50 cycles) observed in a single RRAM cell for a particular fitting, with $seed$ = 1234. To observe the device-to-device variations of the 1T1R array in Fig. 8, we simulated the 1T1R structure with 100 different seeds and the variability was analysed. Interestingly, the observed device-to-device variability (in row 3) was the same amount of cycle-to-cycle variability observed for a particular fitting, $i.e.$ σ_{SET}, σ_{RESET}, σ_{HRS}, σ_{LRS} observed in a device over 100 cycles was not different from that observed over 100 devices in a particular cycle for a ($P_\delta_g^0$, $N_\delta_g^0$, T_{CRIT}, T_{SMTH}). This is reasonable since we fitted variations in the temporal domain and it manifests in the spatial domain when observed at a specific time window. The IV curves of device-to-device variability of four devices of a row, at an arbitrary cycle is plotted in Fig. 9-(b).

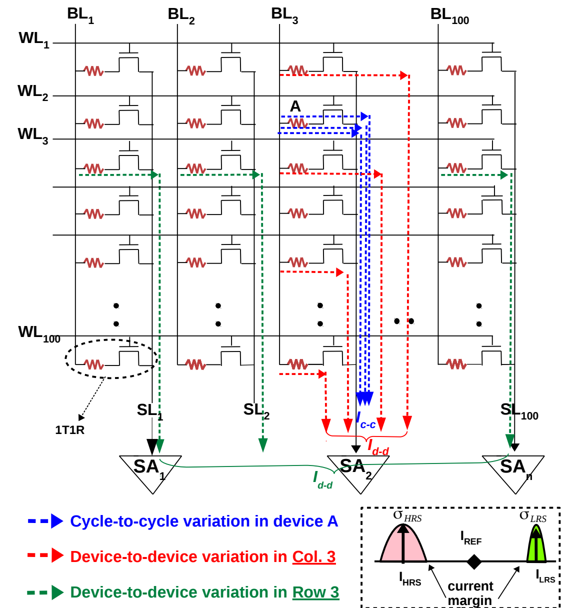


Fig. 8: Illustration of the impact of cycle-to-cycle and device-to-device variations on the design of the Sense Amplifier. While sensing, the current margin is degraded in the presence of variations in resistive states.

B. Effect on variations on the Sense Amplifier Design

Variations in RRAM cell affect the design of the peripheral circuitry. As an example of the effect of variations at the circuit-level, the design of Sense Amplifier (SA) for a 1T1R memory array is analysed. In Fig. 8, consider a device 'A' located in (2,3) of a 100×100 1T1R array. Whenever the cell 'A' is written into, the device gets programmed to a different resistance, due to cycle-to-cycle variability. This results in a different current every time the cell 'A' is read, denoted I_{c-c} (blue lines in Fig. 8). Sense amplifier SA_2 has to distinguish

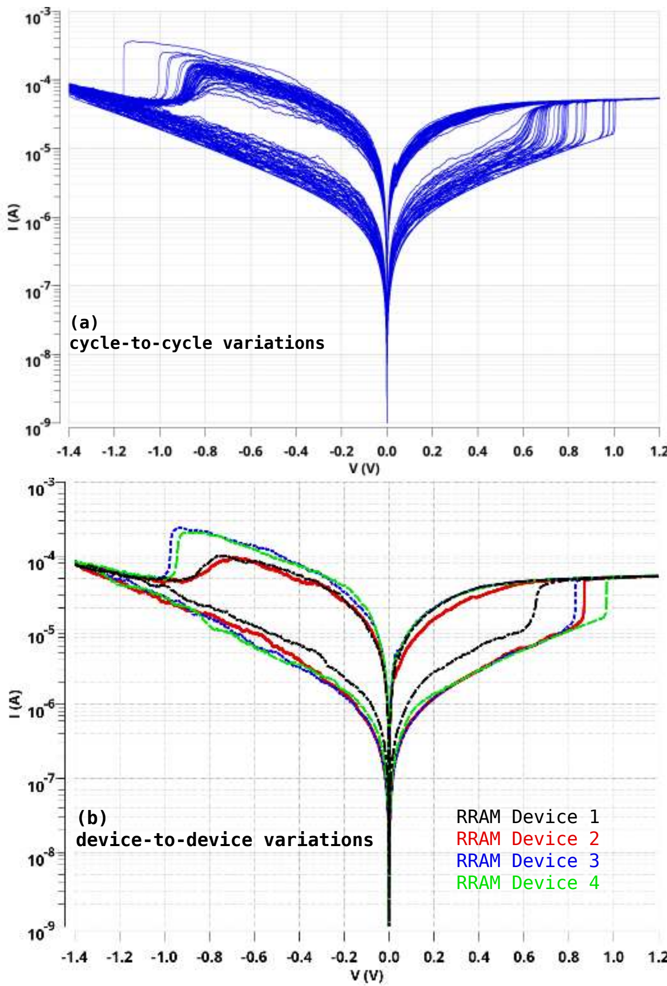


Fig. 9: (a) Cycle-to-cycle variations (50 cycles) reproduced by the Stanford-PKU model fitted for (σ_{HRS} = 31.2 KΩ, σ_{LRS} = 1.42 K Ω, σ_{SET} = 147 mV, σ_{RESET} = 97 mV) (b) Device-to-device variations with the same fitting; Only 4 devices depicted for clarity of presentation.

between LRS and HRS stored in 'A', over different cycles. The same SA has to be tolerant of device-to-device variations (red lines in Fig. 8) since it is common to all the cells in column 3. When different cells in column 3 are read (at different instances), it will result in different currents, denoted I_{d-d} (even if all the cells in that column are programmed to the same state, HRS). Therefore SA_2 has to be designed to tolerate the higher of the two variations (cycle-to-cycle and device-to-device). Since all the sense amplifiers in the peripheral circuitry of the array are identical in structure (same design, W/L), they must be designed to tolerate device-to-device variations across the array (green lines in Fig. 8). Device-to-device variability is usually reported across the array and not across a row/column. For example, the reported cycle-to-cycle variability for a Al/Ge/TaOx/Pt device is: (σ_{SET} = 0.48 V, σ_{RESET} = 0.25 V, $(\sigma/\mu)_{LRS}$ = 25%, $(\sigma/\mu)_{HRS}$ = 80%). The device-to-device variability across the array is (σ_{SET} = 0.39 V, σ_{RESET} = 0.29 V, $(\sigma/\mu)_{LRS}$ = 5%, $(\sigma/\mu)_{HRS}$ = 26%) [15]. Clearly, the cycle-to-cycle dispersion is greater than device-to-device dispersion for both HRS and LRS for this device. In this case, the SA for such devices must be designed to tolerate cycle-to-cycle variations.

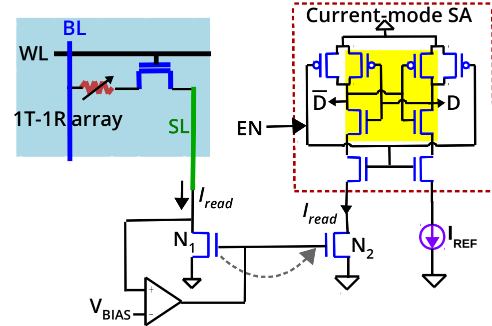


Fig. 10: Current-mode SA [38], [39] used to evaluate the effect of RRAM variations. Current from the RRAM cell I_{READ} and I_{REF} are compared. 'D' and '$\overline{D}$' are precharged to V_{DD} when EN is low. When EN goes high, one of them discharges at a faster rate, which is reinforced by the positive feedback formed by cross-coupled inverters (shaded yellow).

For IHP's devices fitted in this paper, mean I_{HRS} and I_{LRS} will be 1.8 μA and 18 μA, respectively. We adopted the current-mode SA proposed in [38], to differentiate between HRS and LRS. As shown in Fig. 10, the current I_{READ} from the 1T–1R array is mirrored by N_1-N_2 pair and compared with I_{REF} in a current-mode SA. The op-amp biases the drain of transistor N_1 at a constant voltage, V_{BIAS} to ensure that N_1 is in saturation. To read from a cell, V_{BIAS} of 0.8 V was used and 1 V was applied at BL, resulting in an effective voltage V_{READ} of 0.2 V (the SL is held at 0.8 V by the op-amp, Fig. 10). The SA of Fig. 10 was designed in 250 nm CMOS technology and simulated to verify read-out of HRS (1.8 μA) and LRS (18 μA). I_{REF} was chosen to be 10 μA, midway between the mean I_{LRS} and I_{HRS}. Bit Error Rate (BER) is one of the important performance metrics in the design of SA and is typically evaluated by performing Monte Carlo (MC) simulations. We performed 4000 MC simulations first without RRAM variations $i.e.$. the effect of CMOS variations (process and mismatch) on the BER was analysed. The BER was found to be 0 (0 errors in 4000 simulations). We then performed another set of 4000 MC simulations with RRAM variations of σ_{HRS} = 31.2 KΩ and σ_{LRS} = 1.4 KΩ. The BER of the SA in the presence of RRAM variations was 0.4% (16 errors in 4000 simulations). It must be noted that the difference between the two BERs is not huge in this case since the SA 'accommodates' variations, $i.e.$ 1.8 μA $\pm$ 2 μA is still sensed as HRS and only significant deviations from mean I_{HRS} are sensed erroneously. Nevertheless, including RRAM variations did give a realistic estimate of the BER and will be even more useful in analysing other RRAM-based circuits where variations affect circuit-performance more linearly. Depending on whether the application can tolerate such a BER, the SA needs to be re-designed to be more error-tolerant ($e.g.$ by increasing size of transistors) or separate error detection/correction circuit needs to be included. This will increase the area and hardware complexity of the read-out

circuitry. In this manner, incorporating variations in RRAM models can enable variability-aware design of the SA and also aid in determining the effects of variability at the circuit-level (e.g. increased area *etc.*).

V. Prospects of the Proposed Variability Fitting Methodology to Other Physics-Based Models

In this work, for the first time, we elucidate the fitting of the variability measured from fabricated RRAM devices to a physics-based model. Stanford-PKU RRAM model was used as the base physics-based model for incorporating variations. The literature on RRAM models is growing rapidly and there are many other models proposed in recent years. Some of them are capable of reproducing variability in simulations. In this section, we discuss four other relevant physics-based models which are capable of producing variability. We throw light on how each of them models variability, and discuss the prospects of fitting variability to them.

The work in [18] studies the effect of device-level variations on circuit and system. This work [18] builds on the physics-based model presented in [40] and incorporates variations by adding a variable resistor (called 'Monte Carlo resistor') between the top and bottom electrode of the RRAM. However, the underlying physics-based model to which variations were augmented, is a two-dimensional filament (both length and width of the filament). Therefore, the number of model parameters were twice the number of model parameters in the model we adopted [28], which is a single dimensional filament (length). Consequently, using [18]'s approach, the process of fitting the model to the median characteristics of a RRAM (which has to performed before fitting variations) becomes challenging. Further, [18] incorporates variability in resistive states (by fitting the variable resistor to experimentally observed variations) and does not report anything about variability in V_{SET}/V_{RESET}.

The work in [41] also models variability in a physics-based model. The approach of this work is radically different as variability is incorporated by introducing fluctuations in E_a, the energy barrier for ion migration. The energy barrier was randomly generated from a uniform distribution between 0.7 and 1.7 eV (with a mean of 1.2 eV) and, could reproduce the variability observed in a HfO_x RRAM.

The model proposed in [42], [43] uses the switching energy as the threshold which triggers the change of state ($HRS \leftrightarrow LRS$). The model abstracts resistive switching to a conduction module (to model the dynamic resistance of the RRAM) and state module (to model the RRAM's state). Both modules together constitute a compact model. Variability is included in the model by multiplying the conduction function $f(v, s)$ by the probability density function $r_j(v, s)$ [42]. Moreover, this work has the capability to capture how variability evolves over time (numerous cycles) by using an additional variable in its state module, a capability not supported by the Stanford-PKU model.

The compact model presented for OxRAM devices in [44] is similar to Stanford-PKU model in the sense that the resistive switching is modeled by the growth and shrinkage of a conductive filament. But the key state variable is the radius of the filament. In this model, device-to-device variability is modeled by introducing stochasticity in charge transfer coefficient (α) and the thickness of the oxide (L_x). The work does not report anything about cycle-to-cycle variability.

To summarize, all these models have the capability to reproduce variations in simulations. But no procedure is proposed by these model developers to fit certain variability to their models, *i.e.* the relation between variation in E_a and σ_{HRS} for the model in [41], how much variation in the radius of the filament is needed to produce a particular σ_{LRS} for the model in [44] *etc.* Our contribution meets this need by proposing a generic algorithm to fit a certain amount of variability[5] by first fitting the median switching characteristics to a physics-based model [27] and, then fitting variations by tuning the variability parameters. Furthermore, the manner in which resistive-switching is modeled and the key parameter tweaked to include variability, is different in all the models. Therefore, our variability fitting methodology cannot be directly applied to these models. However, it can be adapted to these models by incorporating stochasticity in the key state variable in a controlled manner *e.g.* for the model in [41], E_a can be Gaussian distributed with a mean μ_{Ea} and standard deviation σ_{Ea} and a relation between the produced variability σ_{HRS} and σ_{Ea} can be formulated. In others words, the amount of variability reproduced by the model can be varied by tuning the parameter σ_{Ea} (to fit to σ_{HRS} exhibited by the device). Depending on how σ_{Ea} affects the four aspects of variability (σ_{HRS}, σ_{LRS}, σ_{SET}, σ_{RESET}), additional variables or modifications may be required to reduce the fitting error.

VI. Conclusion

Variability observed in resistive-switching behaviour is a key impediment for commercialization of RRAM technology. In this work, we have proposed an algorithm to fit certain amount of variability (obtained from RRAM characterization) to the Stanford-PKU model. Further, we have modified the model to enable better fitting of variations (reducing the fitting error from 41 % to 27 %). It was also verified that the model incorporated with cycle-to-cycle variability can also reproduce device-to-device variability, in simulations. The proposed methodology will enable circuit and system designers to fit variability to the Stanford-PKU model. Although the presented methodology was verified only on the Stanford-PKU model, the insights gained in this work can be useful to fit variability to other analytical RRAM models that exist in literature. If variability in RRAMs cannot be eradicated, at best, they must be modeled accurately to facilitate the design of variability-aware circuits and systems.

Acknowledgment

This research was funded by Deutsche Forschungsgemeinschaft (DFG) -Integrierte Memristor-Basierte Rechner-Architekturen (IMBRA) (Project number 389549790).

[5] amount of variability observed is different in different RRAMs depending on the switching oxide, bilayer/single layer *etc.*

APPENDIX A
FITTING LOG-NORMAL DISTRIBUTION OF HRS AND LRS

The log-normal distribution is a probability distribution whose logarithm has a normal distribution. Suppose the random variable X has a normal distribution with expected value μ and standard deviation σ. The density function of this distribution is given by the function f , where

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}.e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{6}$$

We denote X: $N(\mu, \sigma^2)$ as a normal distribution. Then, if

$$Y = e^X \tag{7}$$

Y is said to be a log-normal random variable i.e. $ln(Y)$ is normally distributed ($ln(Y) = X$). The distribution of Y is called a log-normal distribution with mean μ_{log} and standard deviation σ_{log}.

$$\mu_{log} = \frac{1}{N}\sum_{i=1}^{N} ln(y_i) \tag{8}$$

$$\sigma_{log} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}[ln(y_i - \mu_{log})]^2} \tag{9}$$

The density function of the log-normal distribution Y is given by the function g, where

$$g(y) = \frac{1}{y.\sigma_{log}.\sqrt{2\pi}}.e^{\frac{-(ln(y)-\mu_{log})^2}{2\sigma_{log}^2}} \tag{10}$$

While analysing RRAM variability data, we noticed inconsistency in the way the resistance variability was reported by researchers. Some report that both HRS and LRS are normally distributed [5]; some report that LRS variation follows a normal distribution, while HRS a log-normal distribution [12], [36] and, some report that both are log-normally distributed [31], [37]. To demonstrate the applicability of our fitting methodology to a log-normal distribution, we elucidate the fitting methodology presented in Section III-A and III-B considering a log-normal distribution for HRS and LRS (in principle, our fitting methodology can be applied even when only the HRS is log-normally distributed). We used the following procedure to demonstrate fitting log-normal variations (in resistive states) to the Stanford-PKU model.

1) Since the raw data for HRS and LRS distribution was not available, we generated a sample population for HRS and LRS with the size of 1000 based on the distribution presented in III-A. Then we fitted this data to a log-normal distribution in order to find the target fitting point. The target fitting point was: μ_{log}^{HRS} = 11.6, σ_{log}^{HRS} = 0.22, μ_{log}^{LRS} = 9.3, and σ_{log}^{LRS} = 0.13.
2) We followed the procedure of Section III-A/Section III-B by calculating σ_{HRS} and σ_{LRS} based on Eq. 9 instead of Eq. 1.
3) The triplet (δ_g^0, T_{CRIT}, T_{SMTH}) was varied and the new standard deviation (σ_{log}^{HRS}, σ_{log}^{LRS}, σ_{SET}, σ_{RESET}) was collected and analysed. The triplet which minimizes the error in fitting was chosen as the variability fitting parameter.

For the single region fitting (Section III-A), the fitting was (δ_g^0 = 3×10^{-3}, T_{CRIT} = 400, T_{SMTH} = 400) and for the two-region fitting (Section III-B), the fitting was ($P_\delta_g^0$ = 3×10^{-3}, $N_\delta_g^0$ = 4×10^{-3} T_{CRIT} = 850, T_{SMTH} = 850). In this manner, our proposed algorithm can be used to fit a quantum of variability in the resistive states (the distribution may be normal or log-normal) to the Stanford-PKU RRAM model.

APPENDIX B
CHANGE IN THE VERILOG-A CODE OF STANFORD-PKU RRAM MODEL TO IMPLEMENT TWO DIFFERENT δ_g^0

The change to the Stanford-PKU model was done by introducing two new input parameters ($P_\delta_g^0$ and $N_\delta_g^0$) that determine the δ_g^0 value based on the polarity of the applied voltage on RRAM:

...
parameter real $P_\delta_g^0 = 0.0025$*;*
parameter real $N_\delta_g^0 = 0.0035$*;*
...
real δ_g^0*;*
...
if(Vtb $>$ *0)*
    δ_g^0 = $P_\delta_g^0$;
else
    δ_g^0 = $N_\delta_g^0$;

where Vtb denoted the voltage between top and bottom electrode.

REFERENCES

[1] J. Reuben, R. Ben-Hur, N. Wald, N. Talati, A. Ali, P.-E. Gaillardon, and S. Kvatinsky, "Memristive logic: A framework for evaluation and comparison," in *Power And Timing Modeling, Optimization and Simulation (PATMOS)*, September 2017, pp. 1–8.
[2] D. Ielmini and H.-S. P. Wong, "In-memory computing with resistive switching devices," *Nature Electronics*, vol. 1, pp. 333 – 343, 2018.
[3] J. Reuben, N. Talati, N. Wald, R. Ben-Hur, A. H. Ali, P.-E. Gaillardon, and S. Kvatinsky, *A Taxonomy and Evaluation Framework for Memristive Logic*. Cham: Springer International Publishing, 2019, pp. 1065–1099.
[4] G. C. Adam, A. Khiat, and T. Prodromakis, "Challenges hindering memristive neuromorphic hardware from going mainstream," *Nature Communications*, vol. 9, 2018, doi: 10.1038/s41467-018-07565-4.
[5] A. Prakash and H. Hwang, "Multilevel cell storage and resistance variability in resistive random access memory," *Physical Sciences Reviews*, vol. 1, no. 6, pp. –, 2016.
[6] A. Fantini, L. Goux, R. Degraeve, D. J. Wouters, N. Raghavan, G. Kar, A. Belmonte, Y. . Chen, B. Govoreanu, and M. Jurczak, "Intrinsic switching variability in hfo2 rram," in *2013 5th IEEE International Memory Workshop*, May 2013, pp. 30–33.
[7] V. Parmar and M. Suri, *Exploiting Variability in Resistive Memory Devices for Cognitive Systems*. New Delhi: Springer India, 2017, pp. 175–195, doi: 10.1007/978-81-322-3703-7_9.
[8] Y.-F. Kao, W. C. Zhuang, C.-J. Lin, and Y.-C. King, "A study of the variability in contact resistive random access memory by stochastic vacancy model," *Nanoscale Research Letters*, vol. 13, no. 1, p. 213, Jul 2018.
[9] M. Escudero, I. Vourkas, A. Rubio, and F. Moll, "Memristive logic in crossbar memory arrays: Variability-aware design for higher reliability," *IEEE Transactions on Nanotechnology*, vol. 18, pp. 635–646, 2019.

[10] S. Yu, B. Gao, Z. Fang, H. Yu, J. Kang, and H.-S. P. Wong, "Stochastic learning in oxide binary synaptic device for neuromorphic computing," *Frontiers in Neuroscience*, vol. 7, p. 186, 2013. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnins.2013.00186

[11] W. Wang, Y. Li, M. Wang, L. Wang, Q. Liu, W. Banerjee, L. Li, and M. Liu, "A hardware neural network for handwritten digits recognition using binary rram as synaptic weight element," in *2016 IEEE Silicon Nanoelectronics Workshop (SNW)*, 2016, pp. 50–51.

[12] A. Chen, "Utilizing the variability of resistive random access memory to implement reconfigurable physical unclonable functions," *IEEE Electron Device Letters*, vol. 36, no. 2, pp. 138–140, 2015.

[13] D. Ielmini and V. Milo, "Physics-based modeling approaches of resistive switching devices for memory and in-memory computing applications," *J. Comput. Electron.*, vol. 16, no. 4, pp. 1121–1143, Dec. 2017.

[14] E. Shahrabi, C. Giovinazzo, J. Sandrini, and Y. Leblebici, "The key impact of incorporated al2o3 barrier layer on w-based reram switching performance," in *2018 14th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME)*, July 2018, pp. 69–72.

[15] V. Y. Zhuo, Y. Jiang, R. Zhao, L. P. Shi, Y. Yang, T. C. Chong, and J. Robertson, "Improved switching uniformity and low-voltage operation intao_x-based rram using ge reactive layer," *IEEE Electron Device Letters*, vol. 34, no. 9, pp. 1130–1132, Sept 2013.

[16] A. Hardtdegen, C. La Torre, F. Cüppers, S. Menzel, R. Waser, and S. Hoffmann-Eifert, "Improved switching stability and the effect of an internal series resistor in hfo2/tioxbilayer reram cells," *IEEE Transactions on Electron Devices*, vol. 65, no. 8, pp. 3229–3236, Aug 2018.

[17] M. Lanza, H.-S. P. Wong, E. Pop, D. Ielmini, D. Strukov, B. C. Regan, L. Larcher, M. A. Villena, J. J. Yang, L. Goux, A. Belmonte, Y. Yang, F. M. Puglisi, J. Kang, B. Magyari-Köpe, E. Yalon, A. Kenyon, M. Buckwell, A. Mehonic, A. Shluger, H. Li, T.-H. Hou, B. Hudec, D. Akinwande, R. Ge, S. Ambrogio, J. B. Roldan, E. Miranda, J. Suñe, K. L. Pey, X. Wu, N. Raghavan, E. Wu, W. D. Lu, G. Navarro, W. Zhang, H. Wu, R. Li, A. Holleitner, U. Wurstbauer, M. C. Lemme, M. Liu, S. Long, Q. Liu, H. Lv, A. Padovani, P. Pavan, I. Valov, X. Jing, T. Han, K. Zhu, S. Chen, F. Hui, and Y. Shi, "Recommended methods to study resistive switching devices," *Advanced Electronic Materials*, vol. 5, no. 1, p. 1800143, 2019.

[18] H. Li, P. Huang, B. Gao, X. Liu, J. Kang, and H. . Philip Wong, "Device and circuit interaction analysis of stochastic behaviors in cross-point rram arrays," *IEEE Transactions on Electron Devices*, vol. 64, no. 12, pp. 4928–4936, Dec 2017.

[19] S. Kumar, G. Gibson, C. E. Graves, M. D. Pickett, J. P. Strachan, and R. Stanley Williams, *The Art and Science of Constructing a Memristor Model: Updated*. Cham: Springer International Publishing, 2019, pp. 267–285.

[20] S. Kvatinsky, M. Ramadan, E. G. Friedman, and A. Kolodny, "Vteam: A general model for voltage-controlled memristors," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 62, no. 8, pp. 786–790, Aug 2015.

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

[22] S. Menzel and R. Waser, "Analytical analysis of the generic set and reset characteristics of electrochemical metallization memory cells," *Nanoscale*, vol. 5, pp. 11 003–11 010, 2013. [Online]. Available: http://dx.doi.org/10.1039/C3NR03387B

[23] D. Panda, P. P. Sahu, and T. Y. Tseng, "A collective study on modeling and simulation of resistive random access memory," *Nanoscale Research Letters*, vol. 13, no. 1, p. 8, Jan 2018.

[24] S. Menzel, "Comprehensive modeling of electrochemical metallization memory cells," *Journal of Computational Electronics*, vol. 16, no. 4, pp. 1017–1037, Dec 2017.

[25] E. Linn, A. Siemon, R. Waser, and S. Menzel, "Applicability of well-established memristive models for simulations of resistive switching devices," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 8, pp. 2402–2410, Aug 2014.

[26] B. Hajri, H. Aziza, M. M. Mansour, and A. Chehab, "Rram device models: A comparative analysis with experimental validation," *IEEE Access*, vol. 7, pp. 168 963–168 980, 2019.

[27] J. Reuben, D. Fey, and C. Wenger, "A modeling methodology for resistive ram based on stanford-pku model with extended multilevel capability," *IEEE Transactions on Nanotechnology*, vol. 18, pp. 647–656, 2019.

[28] 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 Transactions on Electron Devices*, vol. 63, no. 5, pp. 1884–1892, May 2016.

[29] J. B. Roldán, F. J. Alonso, A. M. Aguilera, D. Maldonado, and M. Lanza, "Time series statistical analysis: A powerful tool to evaluate the variability of resistive switching memories," *Journal of Applied Physics*, vol. 125, no. 17, p. 174504, 2019.

[30] E. Miranda, A. Mehonic, W. H. Ng, and A. J. Kenyon, "Simulation of cycle-to-cycle instabilities in sio $_x$ -based reram devices using a self-correlated process with long-term variation," *IEEE Electron Device Letters*, vol. 40, no. 1, pp. 28–31, Jan 2019.

[31] R. Degraeve, A. Fantini, N. Raghavan, L. Goux, S. Clima, B. Govoreanu, A. Belmonte, D. Linten, and M. Jurczak, "Causes and consequences of the stochastic aspect of filamentary rram," *Microelectronic Engineering*, vol. 147, pp. 171 – 175, 2015, insulating Films on Semiconductors 2015.

[32] X. Guan, S. Yu, and H. . P. Wong, "A spice compact model of metal oxide resistive switching memory with variations," *IEEE Electron Device Letters*, vol. 33, no. 10, pp. 1405–1407, Oct 2012.

[33] S. Yu, B. Gao, Z. Fang, H. Yu, J. Kang, and H. . P. Wong, "A neuromorphic visual system using rram synaptic devices with sub-pj energy and tolerance to variability: Experimental characterization and large-scale modeling," in *2012 International Electron Devices Meeting*, Dec 2012, pp. 10.4.1–10.4.4.

[34] Z. Jiang, S. Yu, Y. Wu, J. H. Engel, X. Guan, and H. . P. Wong, "Verilog-a compact model for oxide-based resistive random access memory (rram)," in *2014 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Sept 2014, pp. 41–44.

[35] E. Pérez, D. Maldonado, C. Acal, J. Ruiz-Castro, F. Alonso, A. Aguilera, F. Jiménez-Molinos, C. Wenger, and J. Roldán, "Analysis of the statistics of device-to-device and cycle-to-cycle variability in tin/ti/al:hfo2/tin rrams," *Microelectronic Engineering*, vol. 214, pp. 104 – 109, 2019.

[36] D. Veksler and G. Bersuker, "Advances in rram technology: Identifying and mitigating roadblocks," *International Journal of High Speed Electronics and Systems*, vol. 25, no. 01n02, p. 1640006, 2016.

[37] W. Wang, G. Pedretti, V. Milo, R. Carboni, A. Calderoni, N. Ramaswamy, A. S. Spinelli, and D. Ielmini, "Learning of spatiotemporal patterns in a spiking neural network with resistive switching synapses," *Science Advances*, vol. 4, no. 9, 2018.

[38] W. Bae, K. J. Yoon, C. S. Hwang, and D.-K. Jeong, "A crossbar resistance switching memory readout scheme with sneak current cancellation based on a two-port current-mode sensing," *Nanotechnology*, vol. 27, no. 48, p. 485201, oct 2016.

[39] J. Reuben, "Binary addition in resistance switching memory array by sensing majority," *Micromachines*, vol. 11, no. 5, 2020.

[40] H. Li, P. Huang, B. Gao, B. Chen, X. Liu, and J. Kang, "A spice model of resistive random access memory for large-scale memory array simulation," *IEEE Electron Device Letters*, vol. 35, no. 2, pp. 211–213, Feb 2014.

[41] S. Ambrogio, S. Balatti, A. Cubeta, A. Calderoni, N. Ramaswamy, and D. Ielmini, "Statistical fluctuations in hfox resistive-switching memory: Part i - set/reset variability," *IEEE Transactions on Electron Devices*, vol. 61, no. 8, pp. 2912–2919, Aug 2014.

[42] F. García-Redondo, R. P. Gowers, A. Crespo-Yepes, M. López-Vallejo, and L. Jiang, "Spice compact modeling of bipolar/unipolar memristor switching governed by electrical thresholds," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 63, no. 8, pp. 1255–1264, 2016.

[43] F. García-Redondo, M. López-Vallejo, and C. L. Barrio, "Advanced integration of variability and degradation in rram spice compact models," in *2017 14th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD)*, 2017, pp. 1–4.

[44] M. Bocquet, D. Deleruyelle, H. Aziza, C. Muller, J. Portal, T. Cabout, and E. Jalaguier, "Robust compact model for bipolar oxide-based resistive switching memories," *IEEE Transactions on Electron Devices*, vol. 61, no. 3, pp. 674–681, 2014.

John Reuben received B.E (Hon's) degree from BITS, Pilani in 2004 and Masters and PhD from VIT University, Vellore, India in 2008 and 2015, respectively. He was a post-doctoral researcher at Technion, Israel from January 2017–January 2018. He is currently working as a post-doctoral researcher at Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). He was a recipient of Viterbi fellowship (Technion, Israel) and Emerging Talents Initiative (FAU, Germany). His research interests are Resistive RAMs, memristive logic and in-memory computing.

Mehrdad Biglari received his B.Sc. degree (2012) from the University of Tehran, Tehran, Iran, and his M.Sc. degree (2015) from Sharif University of Technology, Tehran, Iran. Currently, he is a Ph.D. candidate at the Chair of Computer Architecture at Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany. His research interests include energy-efficient embedded system design with emphasis on the introduction of emerging memory technologies in modern IoT devices.

Dietmar Fey is a full Professor of Computer Science at Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) where he leads the Chair for Computer Architecture since 2009. Before he was associate professor from 2001-2009 for Computer Engineering at University Jena. His research interests are in parallel computer architectures, memristive computing, parallel programming environments, and embedded systems. He was and is involved in several national and international research projects and initiatives on parallel and embedded computing. He authored or co-authored over 140 articles including 3 books, and about 20 papers in journals. He is a member of German Computer Society and and of HiPEAC (European Network of Excellence on High Performance and Embedded Architecture and Compilation) and a contributor for the HiPEAC roadmap.