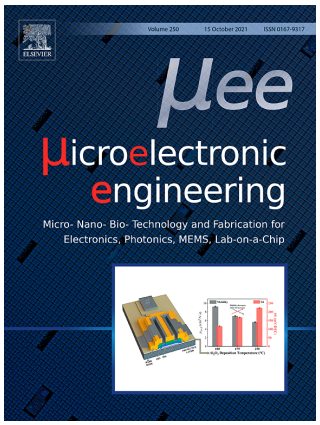
Analysis of the statistics of device-to-device and cycle-to-cycle variability in TiN/Ti/Al:HfO2/TiN RRAMs

- Eduardo Pérez, David Maldonado, Christian Acal, Juan Eloy Ruiz-Castro, F.J. Alonso, Ana María Aguilera, Francisco Jiménez-Molinos, Juan Bautista Roldán, Christian Wenger

- Analysis of the statistics of device-to-device and cycle-to-cycle variability in TiN/Ti/Al:HfO2/TiN RRAMs
- *Microelectronic Engineering Volume 214, 1 June 2019, Pages 104-109*
- DOI: https://doi.org/10.1016/j.mee.2019.05.004



Microelectronic Engineering 214 (2019) 104–109

Contents lists available at ScienceDirect

Microelectronic Engineering

journal homepage: www.elsevier.com/locate/mee

Research paper

Analysis of the statistics of device-to-device and cycle-to-cycle variability in TiN/Ti/Al:HfO_2/TiN RRAMs

E. Pérez[a,*], D. Maldonado[b], C. Acal[c], J.E. Ruiz-Castro[c], F.J. Alonso[c], A.M. Aguilera[c], F. Jiménez-Molinos[b], Ch. Wenger[a,d], J.B. Roldán[b]

[a] IHP-Leibniz-Institut für innovative Mikroelektronik. Frankfurt (Oder), Germany
[b] Departamento de Electrónica y Tecnología de Computadores, Universidad de Granada, Granada, Spain
[c] Departamento de Estadística e Investigación Operativa, Universidad de Granada, Granada, Spain
[d] Brandenburg Medical School Theodor Fontane, Neuruppin, Germany

ARTICLE INFO

Keywords:
RRAM Array
Variability
Weibull distribution
Phase-type distribution

ABSTRACT

In order to study the device-to-device and cycle-to-cycle variability of switching voltages in 4-kbit RRAM arrays, an alternative statistical approach has been adopted by using experimental data collected from a batch of 128 devices switched along 200 cycles. The statistical distributions of switching voltages have been usually studied by using the Weibull distribution. However, this distribution does not work accurately on Al:HfO_2-based RRAM devices. Therefore, an alternative approach based on phase-type distributions is proposed to model the forming, reset and set voltage distributions. Experimental results show that in general the phase-type analysis works better than the Weibull one.

1. Introduction

Emerging resistive memories or Resistive Random Access Memories (RRAMs) are a promising technology that can be implemented for embedded non-volatile memory (NVM) applications on micro-controllers or some secure products within the cryptographic hardware realm [1,2]. In addition, RRAMs are also gaining momentum working as memristive artificial synapses interconnections, which are the key components in hardware neural networks [3]. Some of the advantages of oxide-based RRAMs over FLASH devices, the current technology used in massive NVMs, are lower read time, faster write performance and more scaling capacity [1].

However, there are hurdles to overcome to make this technology enter to the massive production lines. The operation of HfO_2-based RRAMs is based on the stochastic nature of resistive switching (RS) processes that create and disrupt conductive filaments (CFs), consisting of oxygen vacancies (V_O), that changes drastically the device resistance [4]. Therefore, the programmed resistive states and their related RS parameters usually vary in a large scale, which makes the variability a relevant issue that has to be addressed [5,6]. In order to study this stochastic behavior thoroughly, two types of variability have to be taken into account: device-to-device (DTD) and cycle-to-cycle (CTC) variability. The comparison is interesting since DTD variability is analyzed considering independent data, namely, the data correspond to different devices, whereas for CTC variability we consider dependent data coming from RS series measured for just one device (the remnants of CFs after a reset process influence the next set event and, therefore, the subsequent reset process, and so on) [7]. Finding the right statistical model to describe the distribution of RS voltages is a critical requirement that ensures a robust design of RRAM-based devices.

In this work, such a statistical study was performed by making use of experimental data collected during the programming of the RRAM devices several times in the high resistive state (HRS) and in the low resistive state (LRS). The RS voltages during forming, reset and set operations of several RRAM cells integrated in 4-kbit arrays were considered for the DTD variability analysis, while the RS voltages during 200 reset-set cycles were considered for the CTC variability analysis. The usual statistical distribution used for this kind of devices is the Weibull distribution (WD) [8,9], well known in the reliability context [10,11]. However, it was found that WD does not fit accurately the voltage distributions considered [12,13]. Therefore, a new approach based on the phase-type distribution (PHD) was used to characterize the voltage variability of Al:HfO_2-based RRAM devices [13]. This distribution is known to be able to approximate whatever classical positive distribution due to its versatility by considering different intermediate states, phases, and the transition intensities between them.

* Corresponding author.
E-mail address: perez@ihp-microelectronics.com (E. Pérez).

https://doi.org/10.1016/j.mee.2019.05.004
Received 5 April 2019; Accepted 3 May 2019
Available online 13 May 2019

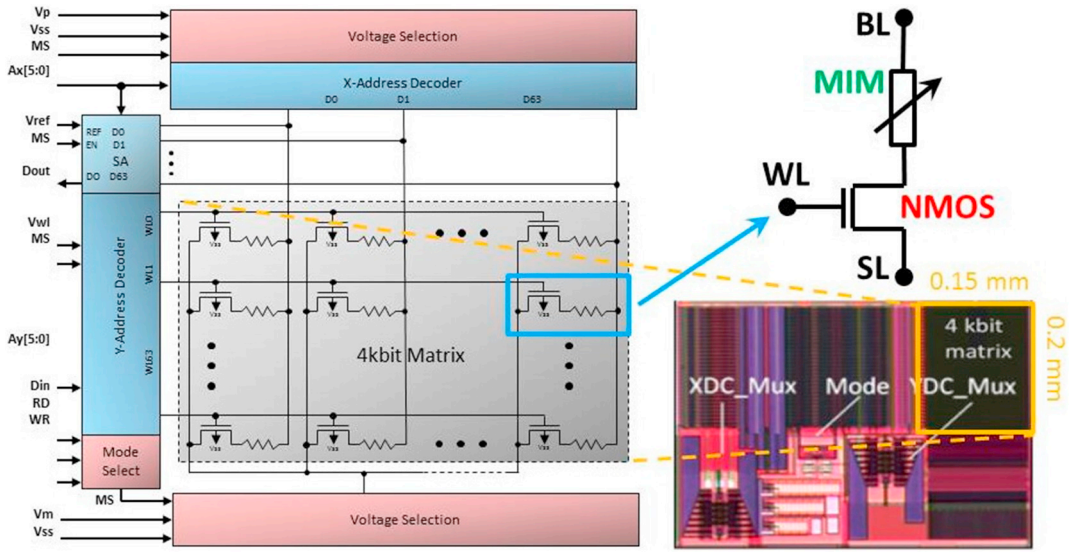


Fig. 1. Block diagram (left) and micrograph (right bottom) of the 4-kbit memory array and schematic of the 1T1R cells (right top) integrated into the array.

Therefore, it allows the study of intermediate probabilistic states in the evolution of statistical systems; in our case, intermediate states in forming, reset or set processes.

2. Device fabrication and measurement

The samples characterized in this study are 1-transistor-1-resistor (1T1R) RRAM cells integrated in 4-kbit arrays, as shown in Fig. 1. Each cell is constituted by a select NMOS transistor, manufactured in the 0.25 µm CMOS technology, whose drain is connected in series to a metal-insulator-metal (MIM) stack placed on the metal line 2 of the CMOS process. The variable MIM resistor is composed by 150 nm TiN top and bottom electrode layers, a 7 nm Ti layer (under the TiN top electrode), and a 6 nm Al-doped HfO_2 layer with an Al content of about 10%. Metal layers were deposited by using magnetron sputtering, whereas the dielectric layer was deposited by using atomic layer deposition (ALD). The resistor is patterned with an area of about 0.4 μm^2 and encapsulated by depositing a thin Si_3N_4 layer to protect it from the environment [14].

In order to accomplish the accurate control of the conductivity of the RRAM cells needed for the implementation of artificial synapses in neuromorphic networks, an optimized programming procedure is required. The incremental step pulse with verify algorithm (ISPVA) has shown to be an effective strategy to reduce the DTD variability by defining threshold current values (I_{th}) [15]. By using this programming algorithm, a sequence of increasing voltage pulses is applied on the bit line (BL), connected to the MIM resistor, during forming and set operations, whereas this sequence is applied on the source line (SL), connected to the transistor source, during reset operations [16]. After every pulse a read-verify operation is carried out. The programming operation is stopped when the read-out current reaches the corresponding I_{th}.

In order to statistically model the RS voltage distributions, a criterion should be established to define what is considered a RS event during forming, reset and set operations. As shown in Fig. 2, multiple criteria can be established for each operation. For set operations (as well as for the forming operation) there are two possibilities: the voltage amplitude at which the maximum derivative of the read-out current curve takes place (V_{set_slp}) or the voltage amplitude at which the set I_{th} is overcome (V_{set_th}). For reset operations two types of reset should be considered [17]. During the 1st Reset (Reset1) after the forming operation, the read-out currents feature a significant increase before performing the reset switch. Hence, two definitions were considered:

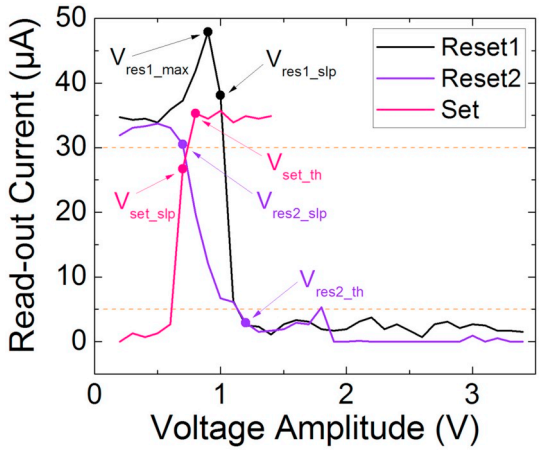


Fig. 2. Schematic illustration of the criteria definition of RS voltages on the read-out current evolution during the ISPVA. The dotted lines represent the I_{th} values.

the voltage amplitude at which the maximum current value is achieved (V_{res1_max}) or the voltage amplitude at which the highest current drop takes place (V_{res1_slp}). During the subsequent reset operations (Reset2) such a current increase is suppressed. For these reset operations, V_{res1_slp} definition remains still useful (V_{res2_slp}), however, one alternative definition to V_{res1_max} was considered: the voltage amplitude at which the reset I_{th} is crossed down (V_{res2_th}).

3. Results and discussion

3.1. DTD variability in forming operation

In order to activate the RS behavior, a preliminary forming operation is required to bring the RRAM devices from pristine state to the LRS [18,19]. To study the DTD variability of the forming voltages, the forming operation was performed on a batch of 128 1T1R devices integrated in the 4-kbit array. The amplitude of the voltage pulses applied by the ISPVA was swept in the range of 2–5 V with a voltage step of 0.01 V. The LRS was defined by using a I_{th} value of 30 µA together with a compliance current limitation imposed by a gate voltage (V_g) of 1.4 V applied on the word line (WL).

The evolution of the read-out current values measured after each programming pulse during the voltage amplitude sweep is shown in

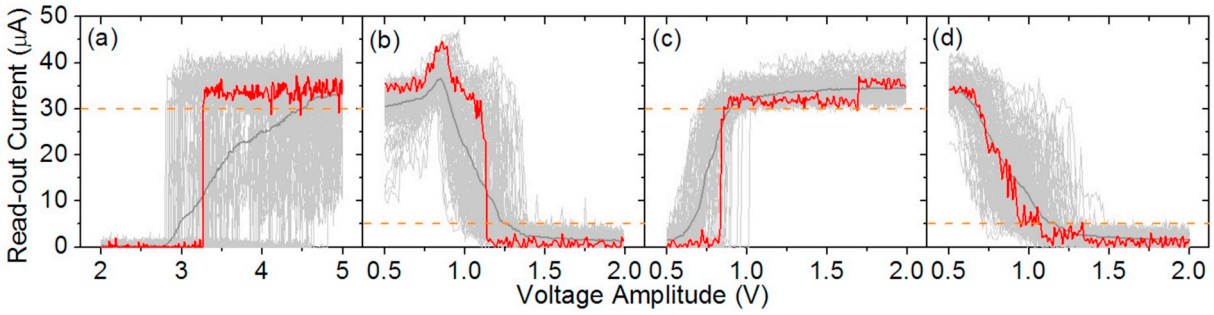


Fig. 3. Evolution of read-out current values during ISPVA for forming (a), the 1st Reset (b), set (c) and second reset (d) operations on 128 devices (grey), its average (dark grey) and on one specific device (red). The dotted orange lines represent the I_{th} values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

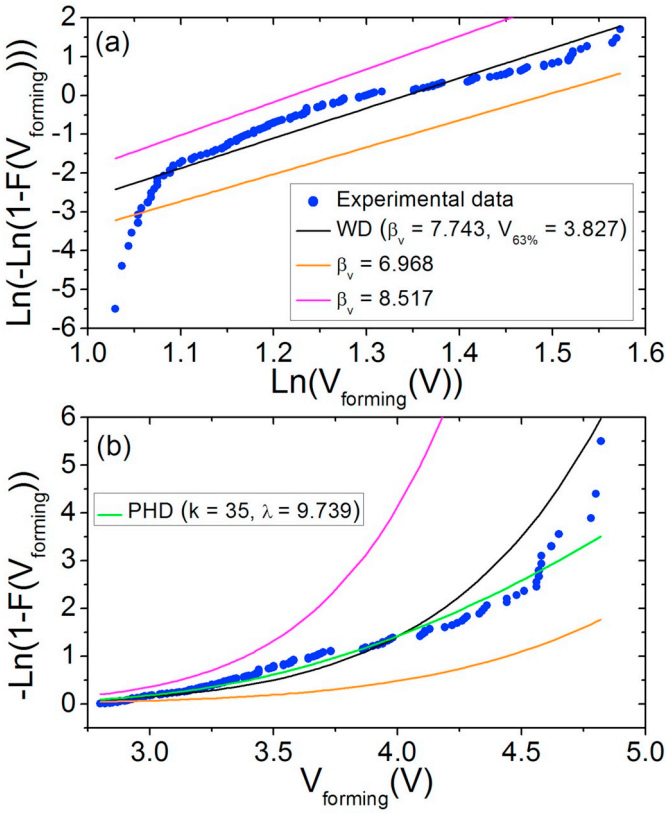


Fig. 4. Statistical fits of the forming voltages ($V_{forming}$) for the 128 RRAM devices under consideration: WD linear fit (a) and fit of the cumulative hazard rate (CHR) calculated as -ln(1-F) (b). The WD fit is obtained with the black line and a reduction (increase) of a 10% in the β_v parameter was assumed in the orange (magenta) line. The PHD fit is obtained with the green line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

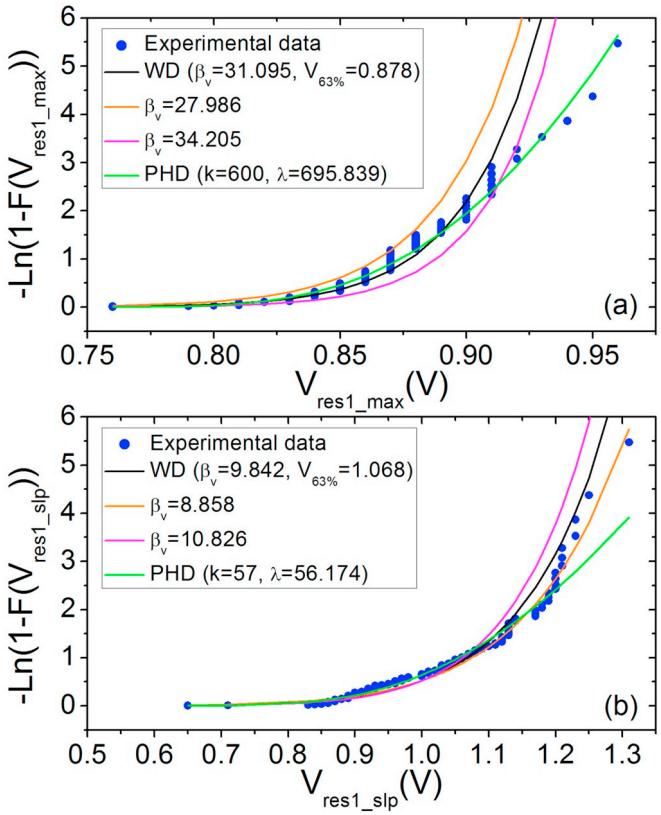


Fig. 5. Statistical fits of the CHR of the reset voltages extracted from the Reset1 operation: V_{res1_max} (a) and V_{res1_slp} (b); for the 128 RRAM devices under consideration. The WD fit is obtained with the black line and a reduction (increase) of a 10% in the β_v parameter was assumed in the orange (magenta) line. The PHD fit is obtained with the green line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 3(a). Based on the V_{set_slp} definition, the forming voltages ($V_{forming}$) were obtained and the statistical analysis was performed starting with the WD, as shown in Fig. 4(a), which is defined by the following expression:

$$F(V) = 1 - \exp\left[-\left(\frac{V}{V_{63\%}}\right)^{\beta_v}\right] \quad (1)$$

where $F(V)$ is the cumulative distribution function of voltages, β_v is the shape parameter and $V_{63\%}$ is the voltage value by which 63% of the samples perform the RS. As it can be seen clearly in Fig. 4(a), this distribution does not fit the experimental data. Therefore, other alternative distributions should be tried. The PHDs constitute a class of non-negative distributions that makes possible to model complex problems with well-structured results. Caused by their valuable properties, this type of distributions has been used in reliability studies in diverse science fields. Some particular cases of PHDs are the exponential, Erlang, generalized Erlang, hyper-geometric and Coxian distributions, among others. The versatility and advantages of PHDs can provide a better analysis of the voltages distribution. Details such as the matrix structure and different representations of PHD can be seen in [13]. As shown in Fig. 4(b), the analysis based on PHD, where the k parameter stands for the number of phases, fits the experimental data much better than in the case of the WD. In particular, after the PHD analysis, an Erlang distribution function works well because of the features of the

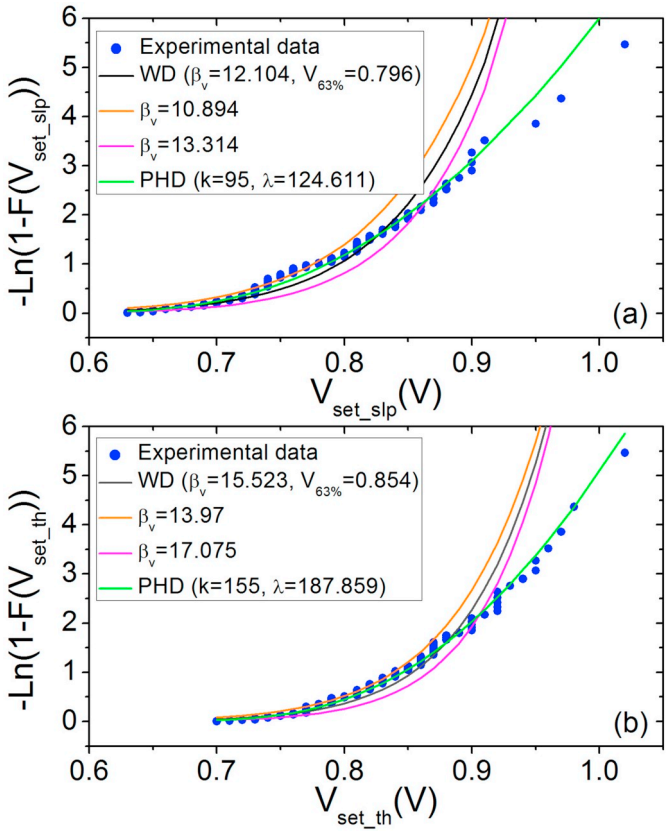


Fig. 6. Statistical fits of the CHR of the set voltages extracted from the set operation: V_{set_slp} (a) and V_{set_th} (b); for the 128 RRAM devices under consideration. The WD fit is obtained with the black line and a reduction (increase) of a 10% in the β_v parameter was assumed in the orange (magenta) line. The PHD fit is obtained with the green line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

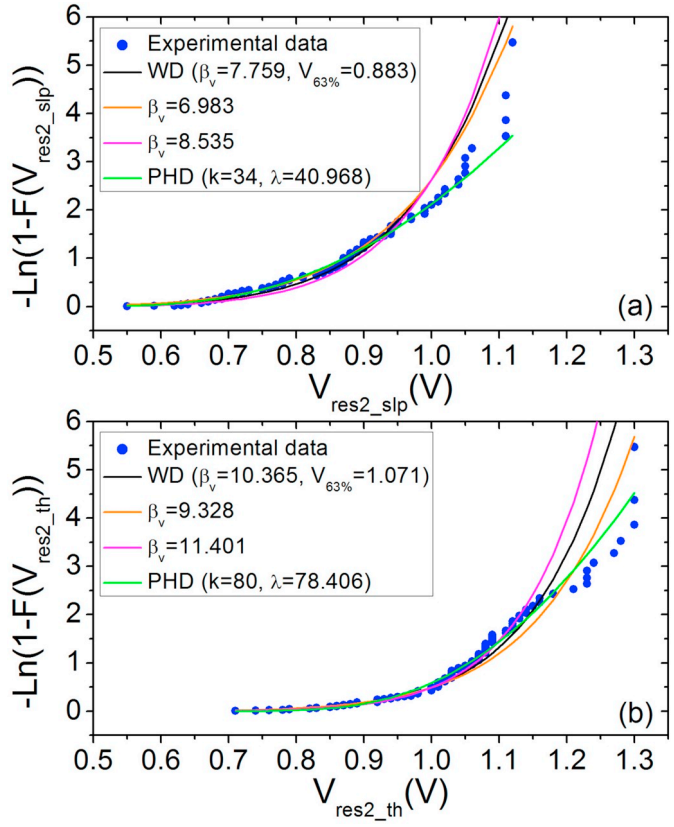


Fig. 7. Statistical fits of the CHR of the reset voltages extracted from the Reset2 operation: V_{res2_slp} (a) and V_{res2_th} (b); for the 128 RRAM devices under consideration. The WD fit is obtained with the black line and a reduction (increase) of a 10% in the β_v parameter was assumed in the orange (magenta) line. The PHD fit is obtained with the green line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

transitions between the states obtained. The Erlang distribution is defined by the expression:

$$F(V) = 1 - \sum_{j=0}^{k-1} \frac{1}{j!} \exp(-\lambda V)(\lambda V)^j \tag{2}$$

where k is the shape parameter and λ is the rate parameter. The phase-type representation is described in [13].

3.2. DTD variability in reset and set operations

After the forming operation, the batch of 128 1T1R devices was switched to the HRS by means of the 1st Reset operation (Reset1). Afterwards it was switched back to the LRS by means of a set operation and finally switched again to the HRS by means of a reset operation (Reset2). The RS voltage distributions obtained during these three consecutive operations were studied in terms of DTD variability. During these programming operations the amplitude of the voltage pulses applied by the ISPVA was swept between 0.5 and 2.0 V with a voltage step of 0.01 V. The I_{th} and V_g values during the set operation were defined as 30 µA and 1.4 V, respectively; whereas they were defined as 5 µA and 2.7 V, respectively, during the reset operations.

The evolution of the read-out current values measured after each programming pulse during the Reset1 operation is shown in Fig. 3(b). The current peak featured by the read-out current characteristic during the 1st Reset operation is clearly visible [17]. As illustrated in Fig. 5, the voltage distributions obtained during the Reset1 operation are strongly impacted by the definition chosen as reset switching criteria (Fig. 2). By

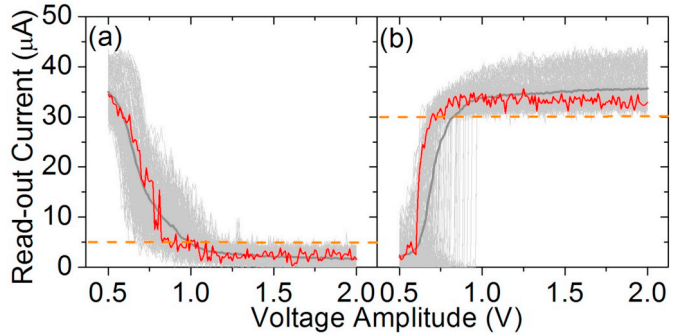


Fig. 8. Evolution of read-out current values during ISPVA for reset (a) and set (b) operations during 200 reset-set cycles (grey), its average (dark grey) and at one specific cycle (red). The dotted orange lines represent the I_{th} values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

using V_{res1_max}, the voltage distribution ranges between 0.75 and 1.00 V, which is much narrower than the distribution obtained by using V_{res1_slp}, namely, between 0.6 and 1.3 V. Additionally, the former can be well fitted with the PHD analysis (Erlang), whereas the later can be better fitted with the WD.

In Fig. 3(c) the evolution of the read-out currents measured after each ISPVA programming pulse during the set operation is depicted. As shown in Fig. 2, two criteria were taken into account in order to define the RS event during the set operation. The statistical analyses depicted

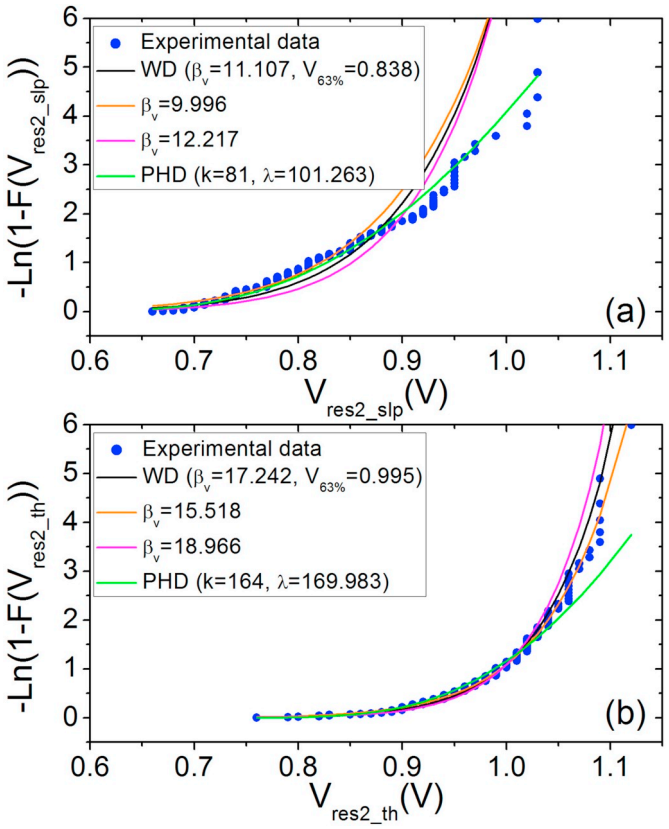


Fig. 9. Statistical fits of the CHR of the reset voltages extracted from the reset operations: V_{res2_slp} (a) and V_{res2_th} (b); during the 200 endurance cycles. The WD fit is obtained with the black line and a reduction (increase) of a 10% in the β_v parameter was assumed in the orange (magenta) line. The PHD fit is obtained with the green line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

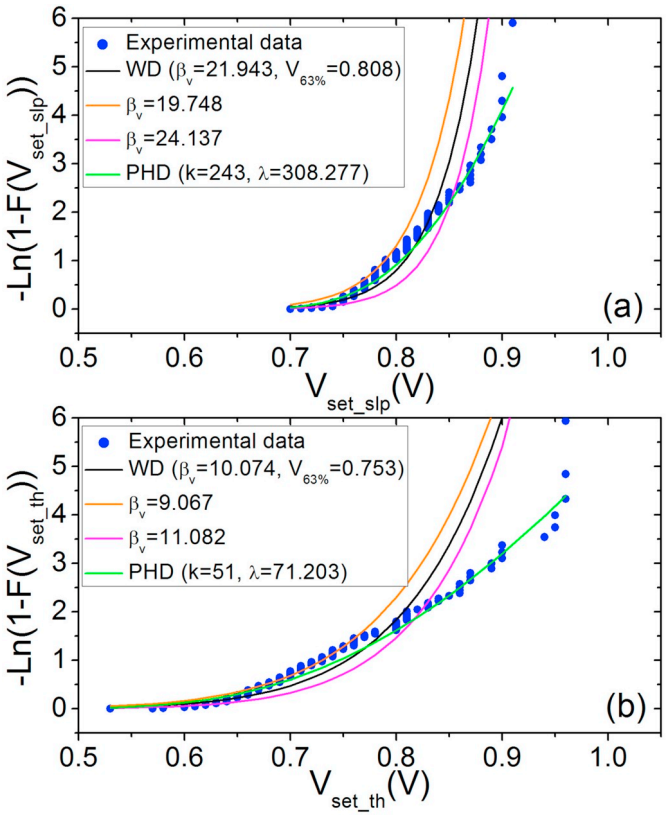


Fig. 10. Statistical fits of the CHR of the set voltages extracted from the set operations: V_{set_slp} (a) and V_{set_th} (b); during the 200 endurance cycles. The WD fit is obtained with the black line and a reduction (increase) of a 10% in the β_v parameter was assumed in the orange (magenta) line. The PHD fit is obtained with the green line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in Fig. 6 does not show a significant difference regarding the criterion used, namely, V_{set_slp} or V_{set_th}. Both voltage distributions can be well fitted by the PHD analysis, in particular, by the Erlang distribution. The most remarkable difference between the two distributions is a slight reduction of the voltage range from about 0.6–1.0 V in V_{set_slp} to about 0.7–1.0 V in V_{set_th}.

Finally, the evolution of the read-out current values measured during the Reset2 operation is shown in Fig. 3(d). Compared to the current characteristic of Reset1 (Fig. 3(b)), the current increase featured before the reset switching is suppressed during the Reset2 operation. This absence of a current peak during reset operations endures along the subsequent RS cycles [17]. Similarly to the set operation, the statistical analysis shown in Fig. 7 does not show significant differences regarding the criterion used to define the reset switching, namely, V_{res2_slp} or V_{res2_th}. An Erlang distribution function works well after the PHD analysis with both voltage distributions. In this particular case the range of the voltage distribution is shifted from about 0.5–1.1 V in V_{res2_slp} to about 0.7–1.3 V in V_{res2_th}.

3.3. CTC variability in reset and set operations

After the first forming-reset-set sequence, 200 reset-set cycles (where the reset operation of the first cycle is the Reset2 operation) were performed to statistically analyze the CTC variability of the RS voltages. The ISPVA parameters used during this endurance test were the same as those used during the study of the DTD variability. In addition, the RS criteria are once again those defined in Fig. 2: V_{res2_slp} and V_{res2_th} for reset operations; and V_{set_slp} and V_{set_th} for set operations.

The evolution of the read-out current values measured during the

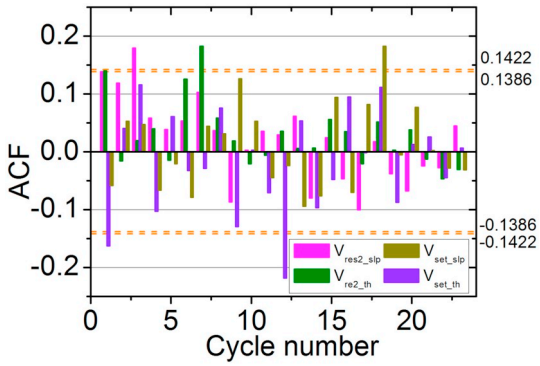


Fig. 11. Autocorrelation function (ACF) versus cycle lag (distance apart in cycles within a series of RS events) for V_{res2_slp}, V_{res2_th}, V_{set_slp} and V_{set_th} series. The ACF minimum threshold bounds are shown with dashed lines.

200 reset operations is shown in Fig. 8(a). The voltage distributions obtained during this cycling are depicted in Fig. 9. The RS criterion selected in this analysis has a strong impact on the result. The distribution obtained by using the definition of V_{res2_slp} ranges between about 0.65 and 1.05 V and is well fitted by the PHD analysis, whereas by using V_{res2_th} it ranges between about 0.75 and 1.10 V and is better fitted by the WD. These results contrast with those in Fig. 7 for the DTD analysis, where both distributions were well fitted by the PHD analysis.

In Fig. 8(b), the evolution of the read-out currents measured during the 200 set operations is illustrated. As shown in Fig. 10, the PHD analysis fits well both voltage distributions, namely, V_{set_slp} and V_{set_th}, like in Fig. 6 for the DTD study. However, the rage of voltages covered

by these distributions is quite different. By using V_{set_slp}, a voltage range of about 0.7–0.9 V is covered, whereas by using V_{set_th} the range is about 0.50–0.95 V.

Finally, the autocorrelation function (ACF) between cycles has been calculated along the endurance cycling for the V_{res2_slp}, V_{res2_th}, V_{set_slp} and V_{set_th} series. In study the correlation between a certain cycle and the previous cycles within the same RS series is calculated. In Fig. 11 the dependence of the data in the CTC variability case (up to a cycle lag of 25) is clearly illustrated. The threshold values shown, namely, 0.1386 and 0.1422, are used in the Time Series analysis to detect the need of regressed terms in modeling a variable with respect to the values of the previous values within a statistical series. This result is in line with the idea that the remnants of CFs after a reset operation influence the subsequent RS operations [7].

4. Conclusions

In this work, the DTD and CTC variability of the RS voltages were statistically studied in Al:HfO_2-based RRAM devices integrated in 4-kbit arrays during forming, reset, and set operations. The statistical fitting most commonly used, namely, the WD, proved not to be an accurate option in most of the analysis. In contrast, the PHD analysis showed a greater versatility fitting most of the voltage distributions, which turned out to follow an Erlang distribution. It deserves to be mentioned that the choice of a specific criterion to define the RS event can influence strongly the distributions obtained. The correlation analysis performed on the CTC data exhibited a clear dependency of the switching behavior of one cycle on previous cycles, caused by the remnants of CFs after reset operations.

Acknowledgements

We thank the support of the Spanish Ministry of Science and the FEDER program through projects TEC2017-84321-C4-3-R and MTM2017-88708-P as well as of the German Research Foundation (DFG) in the frame of research group FOR2093.

References

[1] D.J. Wouters, R. Waser, M. Wuttig, Phase-change and redox-based resistive switching memories, Proc. IEEE 103 (2015) 1274–1288.
[2] A. Chen, Utilizing the variability of resistive random access memory to implement reconfigurable physical unclonable functions, IEEE Electron. Dev. Lett. 36 (2015) 138–140.
[3] S. Yu, Y. Wu, R. Jeyasingh, D. Kuzum, H.-S. Philip Wong, An electronic synapse device based on metal oxide resistive switching memory for neuromorphic computation, IEEE Trans. Electron Dev. 58 (2011) 2729–2737.
[4] Y.S. Lin, F. Zeng, S.G. Tang, H.Y. Liu, C. Chen, S. Gao, Y.G. Wang, F. Pan, Resistive switching mechanisms relating to oxygen vacancies migration in both interfaces in Ti/HfO_x/Pt memory devices, J. Appl. Phys. 113 (2013) 064510.
[5] S. Yu, X. Guan, H.-S. Philip Wong, On the switching parameter variation of metal-oxide RRAM - part II: model corroboration and device design strategy, IEEE Trans. Electron Dev. 59 (2012) 1183–1188.
[6] A. Grossi, E. Nowak, C. Zambelli, C. Pellissier, S. Bernasconi, G. Cibrario, K. El Hajjam, R. Crochemore, J.F. Nodin, P. Olivo, L. Perniola, Fundamental variability limits of filament-based RRAM, Proc. IEEE Int. Electron Dev. Meeting, 2016 (4.7.1-4.7.4.).
[7] C. Baeumer, R. Valenta, C. Schmitz, A. Locatelli, T.O. Mentes, S.P. Rogers, A. Sala, N. Raab, S. Nemsak, M. Shim, C.M. Schneider, S. Menzel, R. Waser, R. Dittmann, Subfilamentary networks cause cycle-to-cycle variablity in memristive devices, ACS Nano 11 (2017) 6921–6929.
[8] S. Long, C. Cagli, D. Ielmini, M. Liu, J. Suñe, Analysis and modeling of resistive switching statistics, J. Appl. Phys. 111 (2012) 074508.
[9] N.J. Western, I. Perez-Wurfl, S.R. Wenham, S.P. Bremmer, Point-contacting by localized dielectric breakdown with breakdown fields described by the Weibull distribution, IEEE Trans. Electron Dev. 62 (2015) 1826–1830.
[10] J. Suñe, New physics-based analytic approach to the thin-oxide breakdown statistics, IEEE Electron. Dev. Lett. 22 (2001) 296–298.
[11] M. Ohring, L. Kasprzak, Reliability and Failure of Electronic Materials and Devices, 2nd ed., Academic, Elsevier, 2014.
[12] Y. Li, H. Sun, M. Zhang, S. Long, Q. Liu, H. Lv, J. Suñe, M. Liu, Statistical analysis of "tail bits" phenomena with defect clustering in RESET switching process of RRAM devices, Proc. IEEE Int. Symp. Phys. Fail. Anal. Integr. Circuits, 2017.
[13] C. Acal, J.E. Ruiz-Castro, A.M. Aguilera, F. Jimenez-Molinos, J.B. Roldan, Phase-type distributions for studying variability in resistive memories, J. Comput. Appl. Math. 345 (2019) 23–32.
[14] P.-Y. Gu, Y.-S. Chen, H.-Y. Lee, P.-S. Chen, W.-H. Liu, W.-S. Chen, Y.-Y. Hsu, F. Chen, M.-J. Tsai, Scalability with silicon nitride encapsulation layer for Ti/HfOx pillar RRAM, Proc. Int. Symp. Technol. Syst. Appl. 2010, pp. 146–147.
[15] E. Perez, A. Grossi, C. Zambelli, P. Olivo, R. Roelofs, Ch. Wenger, Reduction of the cell-to-cell variability in $Hf_{1-x}Al_xO_y$ based RRAM arrays by using program algorithms, IEEE Electron. Dev. Lett. 38 (2017) 175–178.
[16] A. Grossi, C. Zambelli, P. Olivo, E. Miranda, V. Stikanov, C. Walczyk, Ch. Wenger, Electrical characterization and modeling of pulse-based forming techniques in RRAM arrays, Solid State Electron. 115 (2016) 17–25.
[17] E. Perez, M.K. Mahadevaiah, Ch. Wenger, C. Zambelli, P. Olivo, The role of the bottom and top interfaces in the 1st reset operation in HfO_2 based RRAM devices, Proc. Joint Int. EUROSOI Workshop and Int. Conf. Ult. Integr. Silicon, 2018.
[18] G. Bersuker, D.C. Gilmer, D. Veksler, P. Kirsch, L. Vandelli, A. Padovani, L. Larcher, K. McKenna, A. Shluger, V. Iglesias, M. Porti, M. Nafria, Metal oxide resistive memory switching mechanism based on conductive filament properties, J. Appl. Phys. 110 (2011) 124518.
[19] A. Padovani, L. Larcher, O. Pirrotta, L. Vandelli, G. Bersuker, Microscopic modeling of HfO_x RRAM operations: from forming to switching, IEEE Trans. Electron. Dev. 62 (2015) 1998–2006.