Japanese Journal of Applied Physics 61, SM1007 (2022)
https://doi.org/10.35848/1347-4065/ac6a3b

REGULAR PAPER

In-depth characterization of switching dynamics in amorphous HfO_2 memristive arrays for the implementation of synaptic updating rules

Eduardo Perez[1*], Mamathamba K. Mahadevaiah[1], Emilio Perez-Bosch Quesada[1], and Christian Wenger[1,2]

[1]IHP-Leibniz-Institut für Innovative Mikroelektronik, Frankfurt (Oder) 15236, Germany
[2]BTU Cottbus-Senftenberg, Cottbus 03046, Germany

*E-mail: perez@ihp-microelectronics.com

Received December 22, 2021; revised March 9, 2022; accepted April 24, 2022; published online June 17, 2022

Accomplishing truly analog conductance modulation in memristive arrays is crucial in order to implement the synaptic plasticity in hardware-based neuromorphic systems. In this paper, such a feature was addressed by exploiting the inherent stochasticity of switching dynamics in amorphous HfO_2 technology. A thorough statistical analysis of experimental characteristics measured in 4 kbit arrays by using trains of identical depression/potentiation pulses with different voltage amplitudes and pulse widths provided the key to develop two different updating rules and to define their optimal programming parameters. The first rule is based on applying a specific number of identical pulses until the conductance value achieves the desired level. The second one utilized only one single pulse with a particular amplitude to achieve the targeted conductance level. In addition, all the results provided by the statistical analysis performed may play an important role in understanding better the switching behavior of this particular technology.

1. Introduction

Artificial neural networks (ANNs) accomplished one of the most important achievements within the field in 2016 when Google AlphaGo defeated the Go world champion Lee Sedol.[1)] Up to that moment, the implementation of such kind of systems had been carried out by using pure software approaches executed in supercomputers with thousands of CPU/GPUs. Nevertheless, the need of frequent transferences of data between the main memory and the processing unit in the von Neumann architecture, the so called von Neumann bottleneck,[2)] limits drastically the throughput and efficiency of ANN systems. In order to get rid of this limitation, one obvious strategy is to rearrange the hardware architecture to truly mimic the neural network of the brain. Nowadays, the most promising implementation of this computing paradigm relies on the idea of in-memory computing, in which calculations are carried out in situ where the data is stored.[3)] This approach automatically eliminates the von Neumann bottleneck by suppressing the memory-processor communications and lays the foundations for an intrinsically parallel paradigm of computing.[4)]

In the last few years, memristive arrays have emerged as one of the best technologies to implement the huge batches of electronic synapses demanded by such hardware-based ANNs.[5,6)] The most widespread method used to emulate the synaptic plasticity on memristive devices is to adopt the multi-level conductance approach by tuning the current compliance during set operations.[7–11)] However, this strategy actually requires complex and time/energy-consuming programming algorithms.[12–15)] There is a different strategy that is gaining momentum in overcoming the intrinsic limitations associated to implement true analog update of the conductive state of filamentary-based memristors. Such a strategy exploits the inherent stochastic nature of the processes driving the resistive switching operation within the memristors to mimic the synaptic plasticity.[16–20)] In order to achieve this aim, the conductance update stochasticity of the memristive devices has to be statistically modeled. From this kind of analysis it is also possible to develop alternative learning rules for the training phase of the ANNs.[16,18,19,21–23)]

In previous studies,[19,24)] we already contributed to this purpose by analyzing the switching probabilities of memristive devices based on polycrystalline HfO_2 dielectric layers as a function of the number of identical voltage pulses applied during potentiation/depression operations and their amplitude. Based on these results, a novel learning rule was implemented in an ANN intended for classifying the MNIST dataset.[19,21)] Similar statistical studies can be found elsewhere,[25–27)] however every particular technology needs to be analysed in order to obtain the specific characteristics of its stochastic switching. In this work, the switching behavior of memristive devices based on amorphous HfO_2 layers, which are integrated in 4 kbit arrays, was characterized as a function of the amplitude, the width and the number of identical programming pulses applied on batches of 128 1-transistor-1-resistor (1T1R) cells. Afterwards, the experimental measurements were statistically modeled in order to provide methods to develop synaptic updating procedures as well as alternative on-line learning rules in stochastic hardware-based ANNs.

2. Methods

The samples employed to obtain experimental characteristics of the stochastic switching are memristive devices integrated into 4 kbit arrays [see Fig. 1(b)], which consist of 64 rows, each with 64 memristive cells. These cells follow the 1T1R structure [see Fig. 1(a)], which is constituted by a NMOS transistor (manufactured in 250 nm CMOS technology) connected in series to a metal–insulator–metal (MIM) structure placed on metal line 2 of the CMOS process. The MIM resistor consists of a $TiN/HfO_2/Ti/TiN$ stack with 150 nm TiN top and bottom electrodes (TE and BE, respectively) deposited by magnetron sputtering, a 7 nm Ti scavenging layer (under the TiN TE), and a 8 nm amorphous HfO_2 switching layer grown by atomic layer deposition (ALD). After patterning the MIM stack with an area of $600 \times 600\,nm^2$, a SiON layer was deposited to encapsulate and protect the memristive cell.

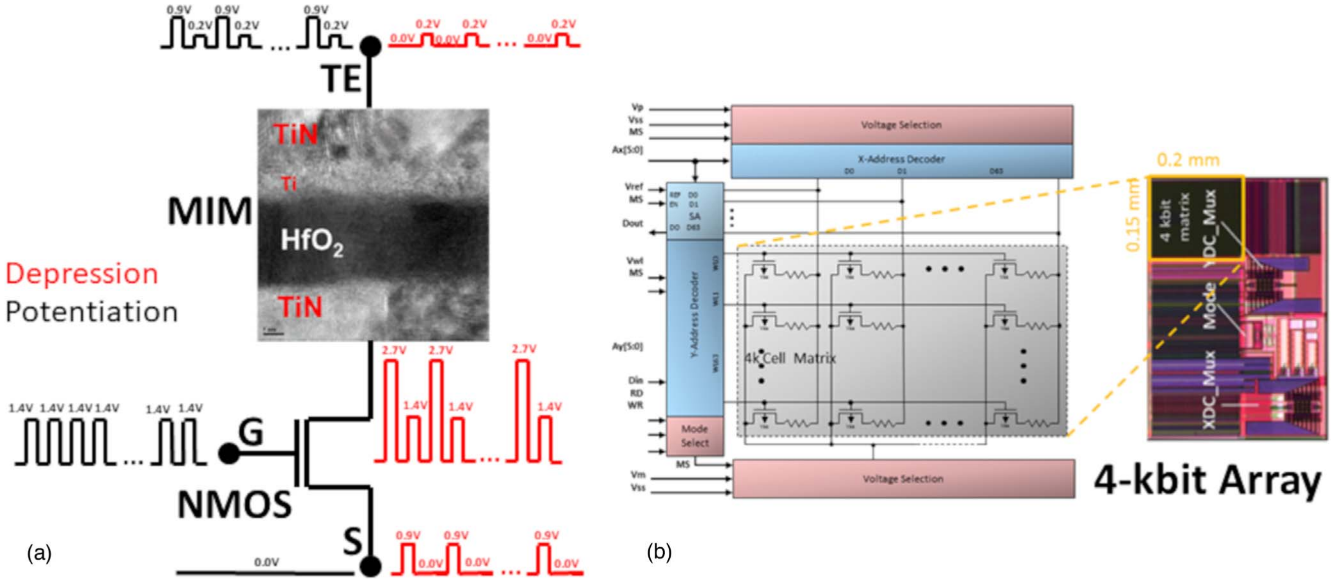

Fig. 1. (Color online) Circuit schematic of the 1T1R memristive cells with the voltage waveforms applied on each terminal during potentiation (black) and depression (red) operations and cross-sectional TEM image of the TiN/HfO_2/Ti/TiN MIM stack (a). Schematic and micrograph of the 4 kbit array (b).

The algorithm used to explore the stochastic nature of the switching behavior of the HfO_2-based memristive devices consists of applying trains of 100 voltage pulses with the same amplitude and pulse width either on the TE or on the S terminal of the 1T1R cell [see Fig. 1(a)] for potentiation or depression operations, respectively. Seven different voltage amplitudes, namely, from 0.6 to 1.2 in steps of 0.1 V, as well as three pulse widths, namely, 10 μs, 1 μs, and 100 ns, were employed on a total of 42 batches of 128 memristive devices (two rows of the array). After each programming pulse a read-out operation at 0.2 V is carried out to track the evolution of the conductive state of the stressed samples. According to Milo et al.,[8] such amplitude value is low enough to have no impact on the switching characteristics. In 1T1R devices the G terminal [see Fig. 1(a)] has also to be properly biased. During potentiation operations the G voltage (1.4 V) controls the maximum current that flows through the memristive device, avoiding its hard breakdown. During depression operations, the G voltage (2.7 V) is selected to minimize the series resistance of the select transistor. For the read-out operation the G voltage is equal to 1.4 V.

3. Results

The memristive technology employed in the present work requires a preliminary electroforming operation to activate the resistive switching behavior.[28,29] In particular, as show in Perez et al.,[30] an electroforming procedure in three steps has shown to result in a more stable and reliable conductive filament (CF). All three steps carried out in this stage, namely, forming, reset and set, are based on the incremental step pulse with verify algorithm (ISPVA).[31] Prior to applying the 100 programming pulses, every batch of 128 memristive devices has been also programmed through the ISPVA into either the low resistive state (LRS) or the high resistive state (HRS) depending on the synaptic operation to be characterized, namely, depression or potentiation,

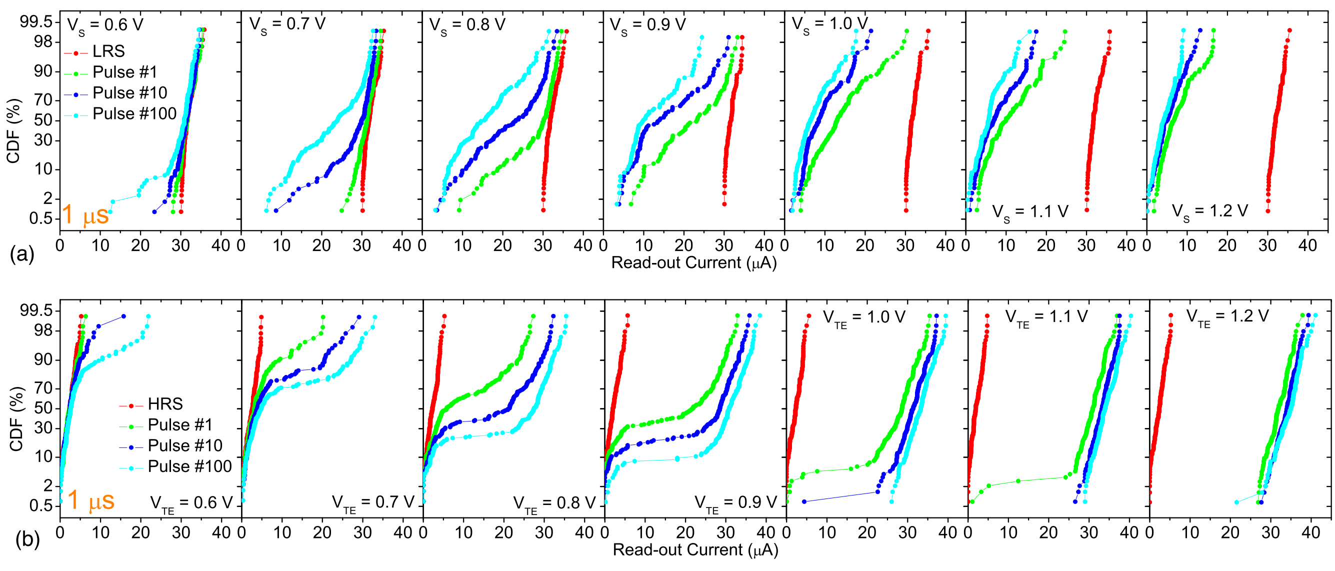

Fig. 2. (Color online) Evolution of the read-out current CDFs measured after applying 1, 10, and 100 programming pulses with the seven voltage amplitudes starting from LRS during depression (a) and from HRS during potentiation (b) for the case of a pulse width equal to 1 μs.

respectively. The current thresholds (I_{trg}) used to program the LRS and the HRS are 30 μA and 5 μA, respectively, as depicted by the red cumulative distribution functions (CDFs) in Figs. 2(a)–2(b).

The evolution of the read-out currents collected while applying the train of 100 programming pulses as a function of the voltage amplitude and pulse width is illustrated in Figs. 3 and 4 for depression and potentiation operations, respectively. Figure 3 shows a significant impact of the voltage amplitude on the modulation of the conductivity whereas the impact of the pulse width is somewhat unexpected. The larger the voltage amplitude the faster the depression process, starting from no conductance change at 0.6 V until switching to LRS in almost one single step at 1.2 V. It is possible to slow down the depression process by working with a pulse width of 100 ns in comparison to 10 and 1 μs. However there is not a noticeable difference between the modulation performed by using a pulse width of 10 μs and a pulse width of 1 μs. Somehow the modulation of the CF takes place at the beginning of the voltage pulse: faster than 1 μs but taking longer than 100 ns. Figure 4 also shows a significant impact of the voltage amplitude on the modulation of the conductivity during potentiation operations. The larger the voltage amplitude, the faster the potentiation process, starting from no conductance change at 0.6 V until switching to LRS in almost one single step at 1.2 V. In contrast to depression operations, the pulse width has the expected strong impact on the modulation of the conductivity. The larger the pulse width, the faster the potentiation process. In this case the modulation of the CF seems to take place during the whole pulse width regardless of the value employed.

In order to understand better how the modulation of the conductivity takes place along the 100 programming pulses for the 128 memristive devices in each batch, the shape of the read-out current CDFs in Figs. 2(a) and 2(b) is analyzed in detail. Only the CDFs regarding the pulse width equal to 1 μs will be considered in the present analysis since that explanation can be easily extrapolated to the other two pulse width values. On the one hand, for the depression operation in Fig. 2(a), when the voltage amplitude is low, the CDF starts to tail toward the HRS, increasing with the number of pulses. At 0.8 V this lower tail achieves the HRS and from 0.9 V onwards the upper tail of the distribution shifts towards the HRS with increasing number of pulses. Such a shift is performed faster with larger voltage amplitudes. On the other hand, for the potentiation operation in Fig. 2(b), the CDF splits mainly into two distributions. This is specially clear for voltage amplitudes larger than 0.7 V. The upper part settles on the LRS, while the lower part remains on the HRS. The number of devices on the upper part increases with the voltage amplitude and the number of pulses. When the voltage amplitude is large enough, namely, larger than 1.1 V, all the devices switch toward LRS even after applying only one single pulse.

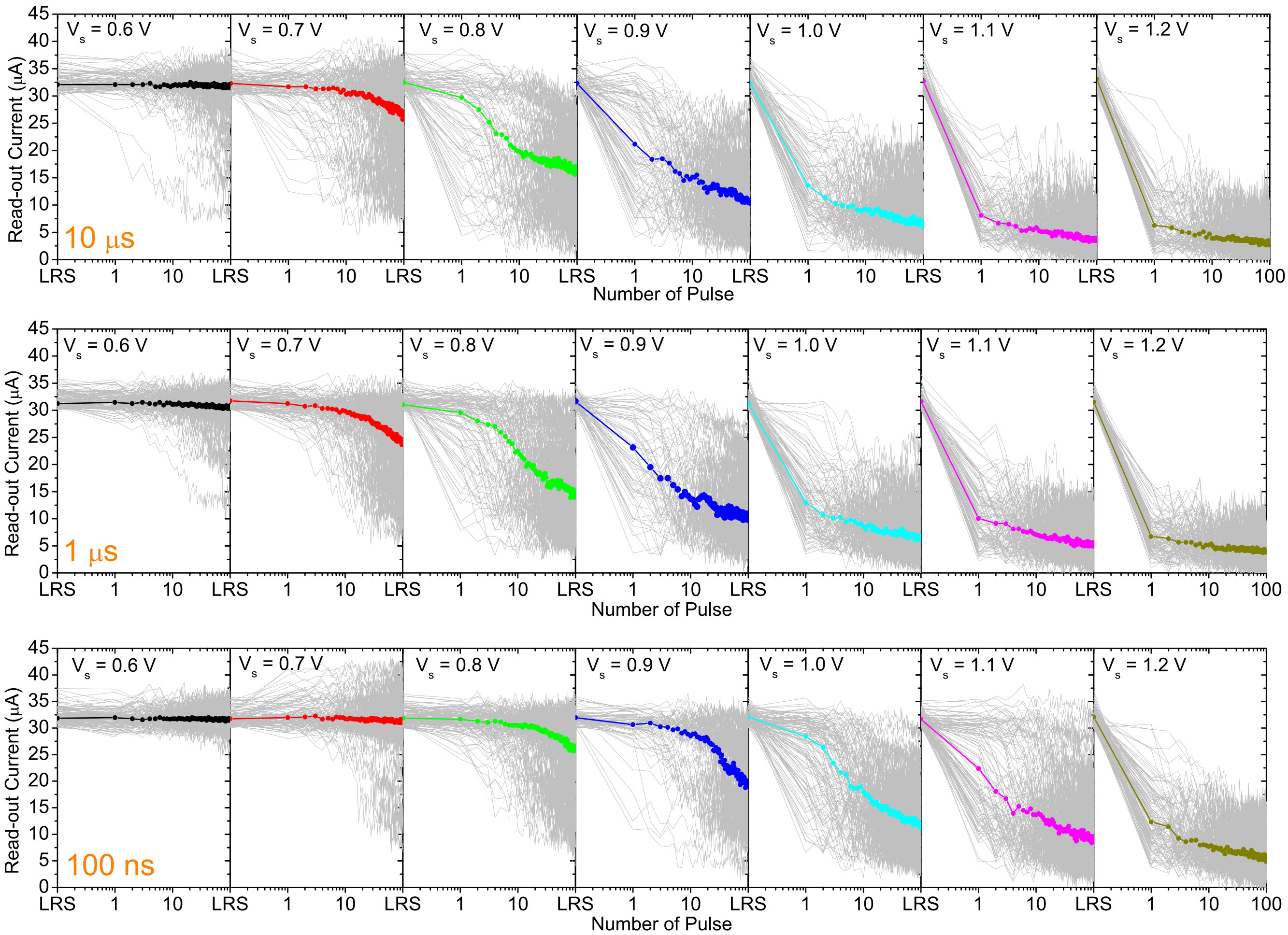

Fig. 3. (Color online) Evolution of the read-out currents measured along the sequence of 100 depression pulses with the three pulse widths and the seven different amplitudes. Grey lines represent each of the 128 devices while the median value is highlighted in different colours.

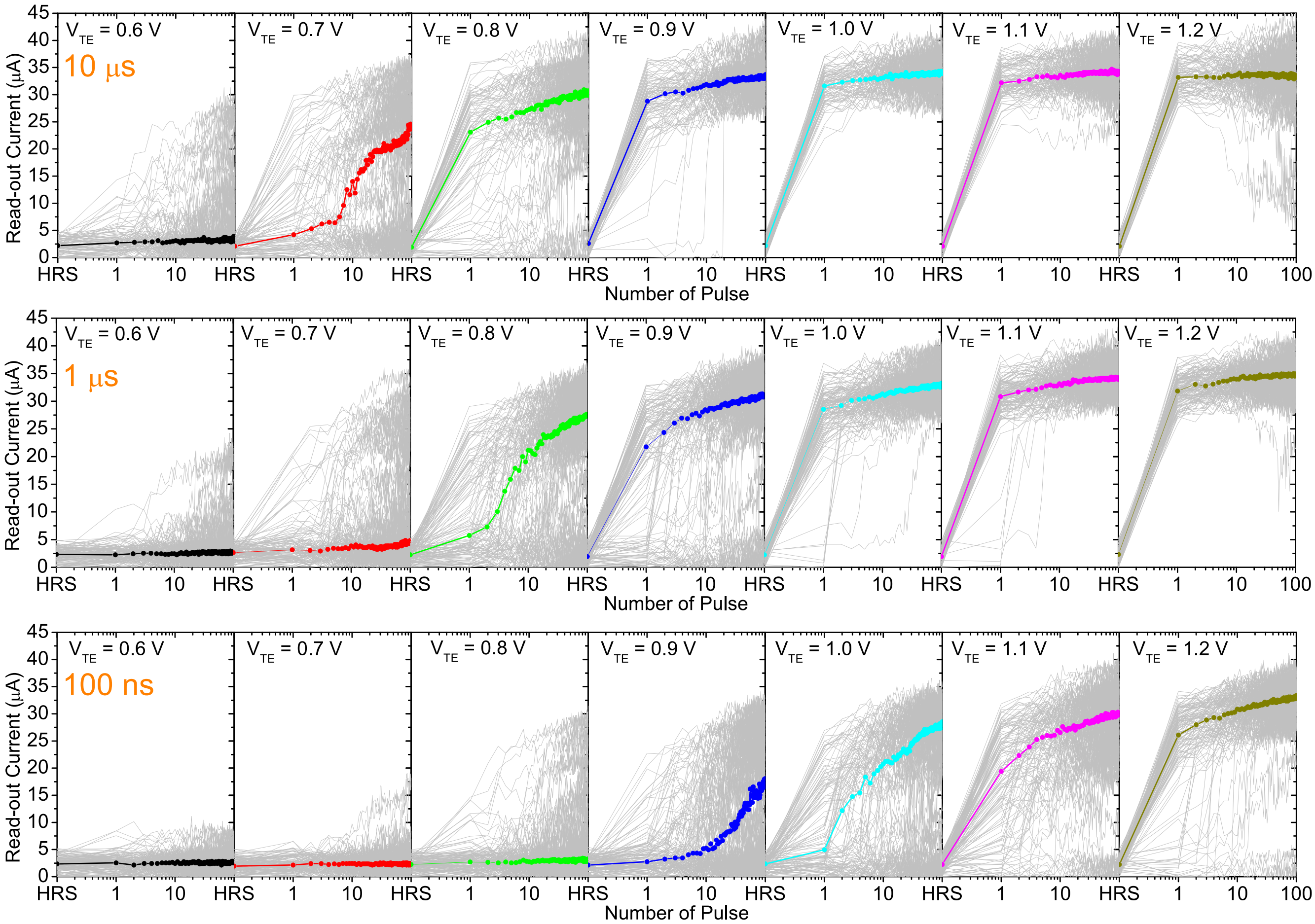

Fig. 4. (Color online) Evolution of the read-out currents measured along the sequence of 100 potentiation pulses with the three pulse widths and the seven different amplitudes. Grey lines represent each of the 128 devices while the median value is highlighted in different colours.

4. Discussion

In order to design synaptic updating rules or even new learning methodologies based on the stochastic conductance modulation phenomenon, a thorough statistical analysis of the experimental characteristics shown in the previous Sect. is required. By taking the median values of the read-out current CDFs, as shown in Figs. 2(a) and 2(b), it is possible to define transference functions relating input voltage amplitudes with the corresponding modulation effect on the memristive conductance, depending on the number of identical pulses applied. The results of this analysis are shown in Figs. 5 and 6 for depression and potentiation operations, respectively, for all three pulse width values considered in this work after applying 1, 10, and 100 programming pulses. The transference functions illustrated can be well fitted by utilizing the sigmoid function, which was already used in Wenger et al.,[19] Zahari et al.,[21] and Mahadevaiah et al.[24] to model the switching probability of HfO_2-based memristive devices.

Figures 5 and 6 show again in a quite clear way that larger voltage amplitudes and larger number of programming pulses lead to stronger conductance modulation on memristive devices. It is also possible to observe the strange dependency already explained regarding the impact of the pulse width during depression operations. Transference functions for 10 and 1 μs pulse widths in Fig. 5 overlap, which was an unexpected result. Considering a theoretical equivalent cumulative effect of programming pulses on the modulation of the read-out current, that is, 1 pulse of 10 μs, 10 pulses of 1 μs, and 100 pulses of 100 ns, the curves shown in Fig. 7 are obtained. Figure 7(b) shows that the conductance modulation during depression for 1 pulse of 10 μs pulse width is weaker than the modulation performed by 10 pulses of 1 μs pulse width, as already explained. It would be expected that 100 pulses of 100 ns pulse width had the same modulating effect as either of the other two, however surprisingly it was obtained something in between. In addition, the results in Fig. 7(a) for potentiation are also remarkable. As expected, the curve resulting from applying 1 pulse of 10 μs pulse width overlaps with the curve resulting from applying 10 pulses of 1 μs pulse width. However the curve resulting from applying 100 pulses of 100 ns pulse width is shifted toward larger voltages, that is, larger voltage amplitudes are required than the theoretically expected for a particular amount of conductance modulation when using a pulse width of 100 ns. Such unexpected behaviors require further research in future works.

The statistical study can not be considered complete if the switching stochasticity is not analyzed in terms of the device-to-device (DTD) variability of the read-out current modulation. In Figs. 8 and 9 the evolution of the read-out current during depression and potentiation operations, respectively, is depicted in terms of the median value versus the standard deviation (σ) calculated from batches of 128 memristive devices as a function of both the voltage amplitude and the

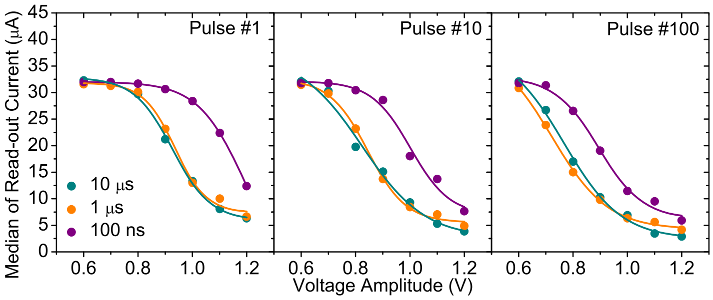

Fig. 5. (Color online) Transference functions connecting the voltage amplitude of the programming pulse and the resulting modulation of the read-out current during depression operations for the three pulse widths depending on the number of pulses applied, in particular: 1, 10, and 100. Experimental (dots) and modeled (solid lines) data are shown.

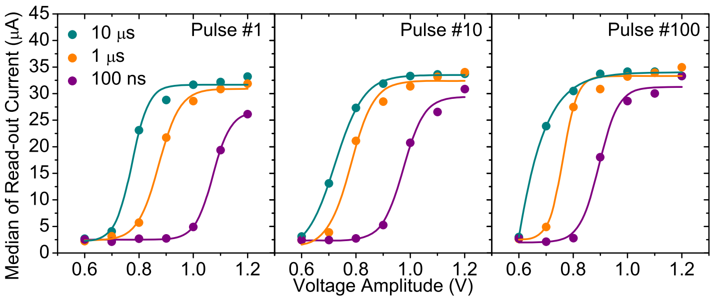

Fig. 6. (Color online) Transference functions connecting the voltage amplitude of the programming pulse and the resulting modulation of the read-out current during potentiation operations for the three pulse widths depending on the number of pulses applied, in particular: 1, 10, and 100. Experimental (dots) and modeled (solid lines) data are shown.

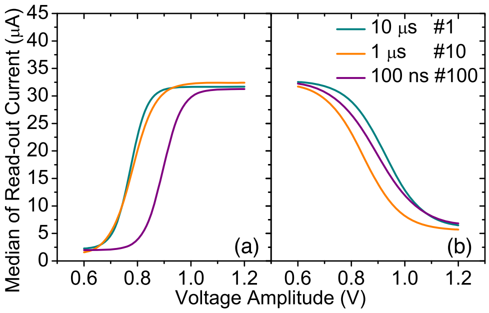

Fig. 7. (Color online) Transference functions with the same theoretical equivalent cumulative modulation effect, namely, 1 pulse of 10 μs pulse width (green line), 10 pulses of 1 μs pulse width (orange line), and 100 pulses of 100 ns pulse width (purple line), for potentiation (a) and depression (b) operations.

number of programming pulses for the three different pulse widths. The behavior already explained for the read-out current CDFs in Fig. 2 can also be observed here. For depression operations in Fig. 8, at low voltage amplitudes σ increases with slight median value decrease from LRS (orange arrow) as the number of pulses rises, which corresponds to the CDF tailing. At larger voltage amplitudes, both σ and median values decrease gradually (purple arrow) as the number of pulses goes on, in correspondence to the shift of the upper tail of the CDF toward the HRS. For potentiation operations in Fig. 9, at low voltage amplitudes σ increases with almost constant median value (purple arrow) as the number of pulses rises, which corresponds to the CDF split into two sub-distributions. At medium voltage amplitudes, σ slightly increases while the median value is strongly increased (black arrow) with increasing number of pulses, which is linked to the transition of the median value from the lower part of the CDF (at HRS) toward the upper part (at LRS). Finally, at high voltage amplitudes σ decreases with almost constant median value (orange arrow) as the number of pulses goes on, which corresponds to the progressive merging of both sub-distributions into a single one at the LRS.

All the insightful knowledge provided by the experimental measurements and statistical analyses previously presented

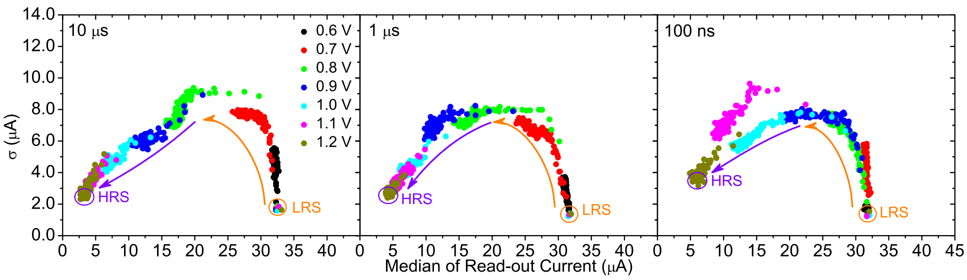

Fig. 8. (Color online) Evolution of median values versus σ values of the read-out currents during depression calculated from batches of 128 memristive devices as a function of the voltage amplitude (each colour) and the number of programming pulses (each dot) for the three different pulse widths: 1 μs, 10 μs, and 100 ns.

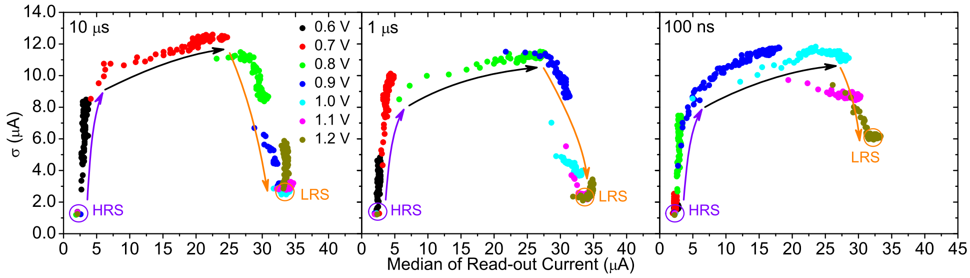

Fig. 9. (Color online) Evolution of median values versus σ values of the read-out currents during potentiation calculated from batches of 128 memristive devices as a function of the voltage amplitude (each colour) and the number of programming pulses (each dot) for the three different pulse widths: 1 μs, 10 μs, and 100 ns.

about our memristive technology can be used now to fulfil the original goal of this work. Regarding the design of synaptic updating rules, two different approaches are considered, namely, conductance modulation based on the number of programming pulses and based on the voltage amplitude of one single programming pulse. For the first one, Figs. 3 and 4 should focus the attention. The programming parameters that provide the most gradual and controllable conductance modulation during depression operations are either a pulse width of 1 or 10 μs with a voltage amplitude of 0.9 V or a pulse width of 100 ns with a voltage amplitude of 1.1 V. The programming parameters that optimize potentiation operations in the same terms are either a pulse width of 1 μs with a voltage amplitude of 0.8 V or a pulse width of 100 ns with a voltage amplitude of 1.0 V. Finding the right combination of programming parameters was more challenging for potentiation since its behavior is well known to be much more abrupt than the depression one.[32,33] By using the second approach, it is possible to obtain the desired conductance modulation by applying just one single pulse with the appropriate voltage amplitude, as shown in Figs. 5 and 6 (Pulse #1). The optimal programming parameters in this case are those that place the transference function on the most centered position within the voltage span. In particular, 1 and 10 μs for depression and 1 μs for potentiation. The second approach has the advantage of consuming less time (just one pulse) but at the cost of a much lower linearity. Regarding the development of new learning methodologies, unfortunately, it is too early and further research will be carried out in this direction in future works.

5. Conclusions

In this study, the stochasticity inherent to the switching dynamics of amorphous HfO_2 memristive devices was experimentally characterized by means of trains of 100 identical depression/potentiation pulses as a function of the voltage amplitude and the pulse width in batches of 128 1T1R cells integrated within 4 kbit arrays. Afterwards, the resulting characteristics were statistically analyzed in order to develop gradual conductance updating methods that mimic the synaptic plasticity feature. Two different approaches were proposed, namely, conductance modulation as a function of the number of programming pulses applied and conductance modulation as a function of the voltage amplitude featured by one single programming pulse. The former provides a more linear control of the conductance update at the cost of larger latency and energy consumption than the later. Based on this insightful knowledge, future works will focus on developing on-line learning rules to be implemented in hardware-based ANNs intended to classify the MNIST dataset, which will be used as standard benchmark.

Acknowledgments

This work was supported in part by the Deutsche Forschungsgemeinschaft (German Research Foundation) under Project 434434223-SFB1461 and in part by the Federal Ministry of Education and Research of Germany under Grant 16ME0092.

ORCID iDs

Eduardo Perez https://orcid.org/0000-0001-7545-9420

1) F. Y. Wang, J. J. Zhang, X. Zheng, X. Wang, Y. Yuan, X. Dai, J. Zhang, and L. Yang, IEEE/CAA J. Autom. Sin. **3**, 113 (2016).
2) N. R. Mahapatra and B. Venkatrao, XRDS: Crossroads ACM Mag. Stud. **5**, 2 (1999).
3) M. Di Ventra and Y. V. Pershin, Nat. Phys. **9**, 200 (2013).
4) J. Welser, J. W. Pitera, and C. Goldberg, IEDM Tech. Dig.1.3.1 (2018).
5) S. Yu, Proc. IEEE **106**, 260 (2018).
6) D. Ielmini and H.-S. Philip Wong, Nat. Electron. **1**, 333 (2018).
7) A. Prakash, J. Park, J. Song, J. Woo, E.-J. Cha, and H. Hwang, IEEE Electron Device Lett. **36**, 32 (2015).
8) V. Milo, C. Zambelli, P. Olivo, E. Perez, M. K. Mahadevaiah, O. G. Ossorio, C. Wenger, and D. Ielmini, APL Mater. **7**, 081120 (2019).
9) P. Yao, H. Wu, B. Gao, J. Tang, Q. Zhang, W. Zhang, J. J. Yang, and H. Qian, Nature **577**, 641 (2020).
10) E. Perez, A. J. Perez-Avila, R. Romero-Zaliz, M. K. Mahadevaiah, E. Perez-Bosch Quesada, J. B. Roldan, F. Jimenez-Molinos, and C. Wenger, Electronics **10**, 1084 (2021).
11) Y. Feng, P. Huang, Y. Zhao, Y. Shan, Y. Zhang, Z. Zhou, L. Liu, X. Liu, and J. Kang, Electron Device Lett. **42**, 1168 (2021).
12) X. Sheng, C. E. Graves, S. Kumar, X. Li, B. Buchanan, L. Zheng, S. Lam, C. Li, and J. P. Strachan, Adv. Electron. Mater. **5**, 1800876 (2019).
13) J. Chen, H. Wu, B. Gao, J. Tang, X. S. Hu, and H. Qian, IEEE Trans. Electron Devices **67**, 2213 (2020).
14) W. Shim, J.-S. Seo, and S. Yu, Semicond. Sci. Technol. **35**, 115026 (2021).
15) E. Perez, M. K. Mahadevaiah, E. Perez-Bosch Quesada, and C. Wenger, IEEE Trans. Electron Devices **68**, 2693 (2021).
16) S. Yu, B. Gao, Z. Fang, H. Yu, J. Kang, and H.-S. Philip Wong, Front. Neurosci. **7**, 186 (2013).
17) Z. Wang, S. Ambrogio, S. Balatti, and D. Ielmini, Front. Neurosci. **8**, 438 (2015).
18) J. Doevenspeck, R. Degraeve, A. Fantini, P. Debacker, D. Verkest, R. Lauwereins, and W. Dehaene, IEDM Tech. Dig. 20.5.1 (2018).
19) C. Wenger, F. Zahari, M. K. Mahadevaiah, E. Perez, I. Beckers, H. Kohlstedt, and M. Ziegler, IEEE Electron Device Lett. **40**, 639 (2019).
20) M. Ignatov, M. Ziegler, M. Hansen, and H. Kohlstedt, Sci. Adv. **3**, e1700849 (2017).

21) F. Zahari, E. Perez, M. K. Mahadevaiah, H. Kohlstedt, C. Wenger, and M. Ziegler, Sci. Rep. **10**, 14450 (2020).
22) D. Garbin, O. Bichler, E. Vianello, Q. Rafhay, C. Gamrat, L. Perniola, G. Ghibaudo, and B. DeSalvo, IEDM Tech. Dig. 28.4.1 (2014).
23) C. Bengel, F. Cueppers, M. Payyand, R. Dittmann, R. Waser, S. Hoffmann-Eifert, and S. Menzel, Front. Neurosci. **15**, 661856 (2021).
24) M. K. Mahadevaiah, E. Perez, C. Wenger, A. Grossi, C. Zambelli, P. Olivo, F. Zahari, H. Kohlstedt, and M. Ziegler, IRPS, 2019.
25) F. Cueppers, S. Menzel, C. Bengel, A. Hardtdegen, M. von Witzieben, U. Boettger, R. Waser, and S. Hoffmann-Eifert, APL Mater. **7**, 091105 (2019).
26) Y. Liao, B. Gao, F. Xu, P. Yao, J. Chen, W. Zhang, J. Tang, H. Wu, and H. Qian, IEEE Trans. Electron Devices **67**, 1593 (2020).
27) J. Hazra, M. Liehr, K. Beckmann, M. Abedin, S. Rafq, and N. Cady, IMW, 2021.
28) G. Bersuker et al., J. Appl. Phys. **110**, 124518 (2011).
29) B. Traore, P. Blaise, E. Vianello, L. Perminola, B. De Salvo, and Y. Nishi, IEEE Trans. Electron Devices **63**, 360 (2016).
30) E. Perez, M. K. Mahadevaiah, C. Zambelli, P. Olivo, and C. Wenger, Solid State Electron **159**, 51 (2019).
31) A. Grossi, C. Zambelli, P. Olivo, E. Miranda, V. Stikanov, C. Walczyk, and C. Wenger, Solid State Electron. **115**, 17 (2016).
32) S. Yu, X. Guan, and H.-S. Philip Wong, IEDM Tech. Dig. 17.3.1 (2011).
33) S. Ambrogio, S. Balatti, D. C. Gilmer, and D. Ielmini, IEEE Trans. Electron Devices **61**, 2378 (2014).