This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LED.2018.2821707, IEEE Electron Device Letters

Understanding and Optimization of Pulsed SET Operation in HfO_x-based RRAM Devices for Neuromorphic Computing Applications

Andrea Padovani, *Member, IEEE*, Jiyong Woo, Hyunsang Hwang, *Senior Member, IEEE*, and Luca Larcher, *Member, IEEE*

***Abstract*—We use experiments and device simulations to investigate pulsed SET operation of HfO_2-based RRAM devices for their possible use as electronic synapses. The application of a train of identical pulses only allows for an abrupt change of the device current, which is not suitable for synaptic devices. By using simulations, we link the microscopic properties and changes of the Conductive Filament (CF) during pulsed operation to the measured conductance and its dependence on pulse voltage, width and number. The results allow to derive guidelines that we use to design optimized SET pulses (or pulse trains) allowing extending the conventional binary operation of HfO_2-based RRAMs to the Multi-Level Cell (MLC) operation required by electronic synapses.**

***Index Terms*— Electronic synapse, resistive random access memory (RRAM), device simulations, HfO_2.**

I. Introduction

Transition Metal Oxide (TMO) based resistive memories (OxRAMs) are considered one of the most viable solutions to implement artificial synapses in neuromorphic computing systems [1]-[3]. These applications require the capability to achieve multi-level cell (MLC) characteristics by tuning the RRAM memory state with high precision. This *analog switching* behavior has been previously demonstrated in synaptic devices comprised of multiple memory cells [4] or operated under particular programming schemes (e.g. compliance current modulation) [5]. However, both solutions come at the cost of increased area and/or power consumption, and the analog switching should be achieved by electrical pulses for practical applications. More recently, pulsed programming has been proposed as a promising alternative [6]-[9], with the potential to provide good analog switching characteristics at the same (area) cost of a 1T-1R RRAM device [10]. Nevertheless, a critical optimization of the programming pulses enabling synaptic like operation is needed, which requires to understand the physical mechanisms and processes governing pulsed operation.

In this paper we use experiments and device simulations to investigate the response of TiN/HfO_2/Ti/TiN RRAM stacks to pulsed SET operation for their possible use as synaptic devices in neuromorphic computing systems. Besides providing fundamental insights on the physics of pulsed SET operation, the results of this study clearly demonstrate that optimizing the amplitude/width/number of the SET pulses allows achieving MLC characteristics, making the studied RRAM cells a viable option as electronic synapses.

II. Experiments and Simulations

Kbit synapse array was comprised of RRAM device with TiN/6-nm-thick HfO_2/Ti/TiN stack with an active area having a diameter of 400 nm and a transistor fabricated with 0.35μm technology, resulting in 1T-1R configuration. RRAM operations are simulated using the Ginestra™ software package [11], based on a multi-scale approach allowing connecting the microscopic/atomic properties of the materials (e.g. defects, morphology, stoichiometry) to the device electrical behavior [12]. The simulation framework includes all the physical mechanisms that are relevant for the operations (forming and switching) and reliability (endurance and retention) of RRAM devices: carriers trapping and transport (including direct/Fowler Nordheim and trap assisted tunneling, and drift across defect sub-bands [12]), the induced power dissipation and the associated temperature increase, generation, recombination and diffusion of oxygen vacancies (V_0) and ions (O^-). All these mechanisms are implemented self-consistently using a multi-scale approach [12], which allows to avoid some important assumptions (e.g. the CF characteristics) with respect to other RRAM models mainly focused on the switching operations [13]–[16]. Statistical simulations are performed by randomly generating every device (and its pre-existing defects) and by using a Monte Carlo technique to account for the stochastic nature of charge transport, degradation, and ion transport processes. The poly-crystalline nature of the HfO_x film is taken into account by considering randomly generated grain boundary (GB) regions of cylindrical shape as preferential locations for charge transport and degradation [17].

Manuscript received XX August 2017. This research was supported by the MOTIE (10067794) and KSRC support program for the development of the future semiconductor device.

A. Padovani is with MDLab s.r.l., Via Sicilia 31, 42122, Reggio Emilia, Italy and with MDLSoft Inc., 5201 Great America Parkway, Suite 320, Santa Clara, CA 95054 (e-mail: andrea.padovani@mdlab-software.it; andrea.padovani@mdlsoft.com).

J. Woo and H. Hwang are with the Department of Materials Science and Engineering, Pohang University of Science and Technology, Pohang, 790-784, South Korea (e-mail: hwanghs@postech.ac.kr).

L. Larcher is with Dipartimento di Scienze e Metodi dell'Ingegneria, Universita di Modena e Reggio Emilia, 41125, Reggio Emilia, Italy.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LED.2018.2821707, IEEE Electron Device Letters

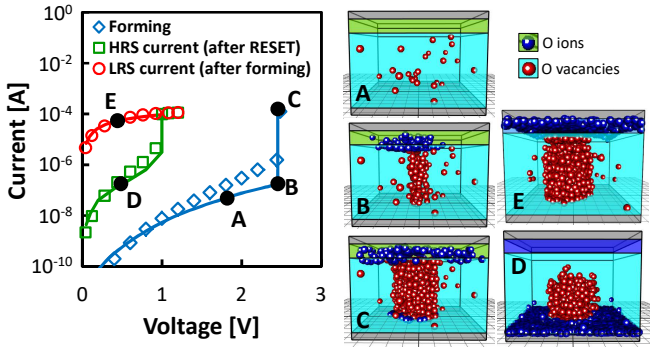

Fig. 1. Measured (symbols) and simulated (lines) forming, LRS (after SET) and HRS (after RESET) IV characteristics and evolution of O^- and V_0 distributions during forming (A)-(C), after RESET (D) and after SET (E). The simulated RRAM consists of the 6-nm-thick HfOx (light blue) and of a thin TiO_x interfacial layer acting as the oxygen reservoir.

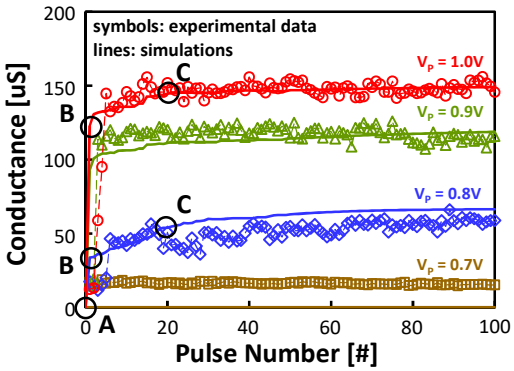

Fig. 2. Conductance (symbols) measured and (lines) simulated during SET pulse trains as a function of the pulse voltage (t_P= 1ms). Data are averaged on 10 statistical trials. Initial forming was done with a 100μA compliance.

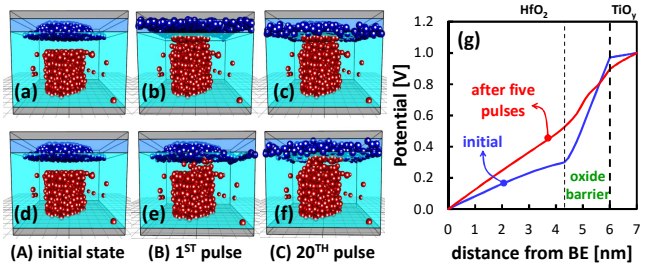

Fig. 3. 3D maps of V_0 (red spheres) and O^- (blue spheres) distributions corresponding to the device states after the application of 0, 1, and 20 pulses (marked as A, B, and C in Fig. 2) with V_P of (a)-(c) 1V and (d)-(f) 0.8V. (g) 1D profile along the center of the conductive filament as extracted form the simulation at the beginning of the SET pulse sequence (blue line) and after five pulses (red line) with V_P=1V.

III. Forming, SET and RESET Simulations

We first used simulations to reproduce the forming (blue diamonds and line in Fig. 1), SET (green squares and line in Fig. 1) and RESET operation (not shown for brevity) in order to determine the properties of the conductive filament (CF) in the low (LRS) and high resistance (HRS) states. The evolution of the generated ion and vacancy distributions during the simulated 100μA compliance forming, Fig. 1 (A)-(C), shows the clear formation of the CF at a GB site (B) followed by its enlargement in the surrounding region (C) up to a diameter of ~4.5nm. The subsequent RESET operation leads to the formation of a ~1.5nm-thick oxygen barrier close to the top electrode, Fig. 1(D), whose breakdown during SET recreates the full CF, Fig. 1(E), and is responsible for the current jump observed at the set voltage of 1V, see Fig. 1 [12]. The CFs obtained through the simulation of the forming and RESET operations allow to nicely reproduce the LRS and HRS currents, respectively, see Fig 1.

IV. Pulse Train Simulations

We now focus on the simulation of consecutive identical SET pulses considering the HRS device state in Fig. 1(D) as the starting point. Pulse amplitudes (V_P) in the range of the device SET voltage (0.7V-1V) and a wide range of pulse widths (t_P) from 1μs to 100ms are considered. A read pulse with voltage V_R=0.1V and duration (t_R) fixed at 1ms is applied immediately after each SET pulse.

A. Effect of Pulse Voltage

Figure 2 shows the conductance measured (lines) and simulated (symbols) during pulse train experiments performed at different pulse voltages (with t_P = 1ms). Significant changes in the conductance are observed when increasing the pulse amplitude to 0.8V and above. Interestingly, the first pulse leads to an abrupt current jump followed by a quick saturation indicating that subsequent pulses do not affect significantly the CF. Figure 2 also shows that the conductance saturation level increases with the applied V_P, suggesting that pulses with higher amplitude allow recreating a stronger CF. All these experimental trends are nicely reproduced by simulations that can thus be used to gain insights on the processes controlling the pulsed SET operation.

We analyzed the evolution of the conductive filament during simulated pulse trains with low (0.8V) and high (1V) V_P. Figures 3(a)-(f) show the 3D maps of V_0/O^- distributions after 0, 1 and 20 pulses (marked as A, B and C in Fig. 2). In the case of a high pulse amplitude, Fig. 3(a)-(c), simulations reveal the existence of a self-limited process: the full filament is completely reconstructed already after the first pulse, Fig. 3(b), which explains the large initial jump of the conductance in Fig. 2. Only few additional vacancies are generated during the subsequent pulses, Fig. 3(c), consistently with the saturation of the conductance observed in Fig. 2. This behavior is determined by the redistribution of the potential along the conductive filament. At the beginning of the pulse train most of the applied voltage drops across the oxide barrier, Fig. 3(g), leading to a massive V_0 generation, Fig. 3(b). Once the oxide barrier is broken (i.e. after the first pulse), the potential drop becomes lower and more uniform across the metallic-like CF, Fig. 3(g), strongly reducing the probability to generate additional V_0.

The evolution of the CF during subsequent pulses with low V_P is quite different as reflected by the different evolution of the conductance, Fig. 2. Results show that only a portion of the CF is reconstructed after the first pulse, see Fig. 3(e). Therefore, the redistribution of the electric field across the CF is limited and insufficient to inhibit the generation of additional V_0, which is indeed observed in the initial part of the pulse train [Fig. 3(f)]. This well explains the smaller initial jump and the continuous increase of the conductance observed in both experiments and simulations at low V_P, Fig. 2.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LED.2018.2821707, IEEE Electron Device Letters

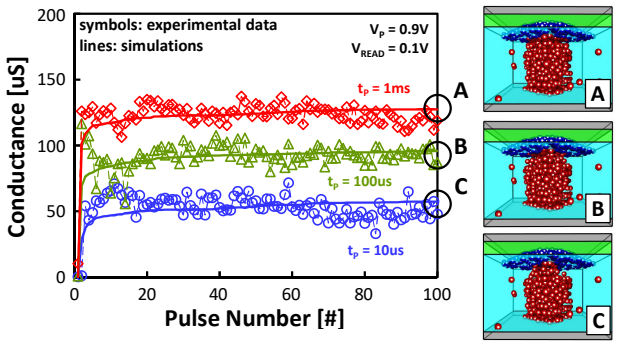

Fig. 4. Conductance (symbols) measured and (lines) simulated during SET pulse trains with V_P=0.9V and different t_P. 3D V_0/O^- distributions of the device states at the end of the simulated 100 consecutive pulses (A, B, and C) are also shown. Data are averaged on 10 statistical trials. Initial forming was done with a 100μA compliance.

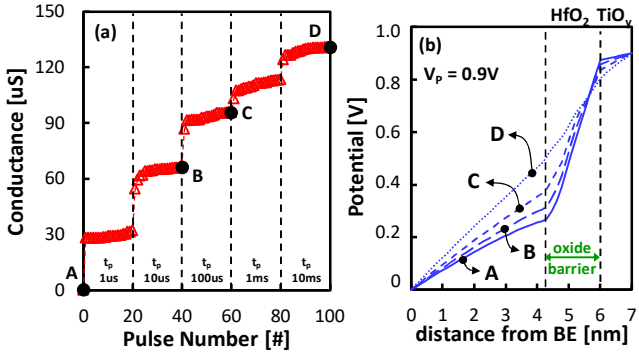

Fig. 5. (a) simulated conductance modulation obtained by applying successive pulse sequences with increasing t_P. (b) Evolution of the 1D potential profile along the center of the conductive filament as extracted from simulations at the points A, B, C and D in (a).

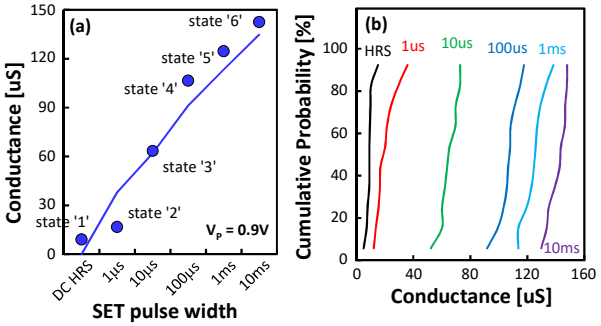

Fig. 6. (a) (symbols) Measured and (line) simulated conductance as obtained by applying the optimized SET pulse train in 5(a). (b) Experimental cumulative probability distributions of the states in (a). Initial forming conditions: 1V/s voltage ramp; 100μA compliance.

B. *Effect of Pulse Width*

Figure 4 shows the conductance measured (lines) and simulated (symbols) during pulse train experiments performed with a fixed V_P of 0.9V and pulse widths of 10μs, 100μs and 1ms. Similarly to what observed when varying the amplitude of pulse trains, the conductance exhibit an abrupt jump after the first pulse and then saturates to a current level that increases when increasing t_P. Simulations reproduce these experimental observations, allowing once again to connect the microscopic characteristics of the CF reconstructed by the SET operation to the measured conductance. To this purpose, Fig. 4 shows the 3D maps of V_0/O^- distributions at the end of the simulated train of 100 SET pulses for the considered t_P of 1ms, 100μs and 10μs (respectively marked as A, B and C). As can be seen, varying the pulse width allows controlling the recreation of the CF: a smaller amount of vacancies is generated when applying short pulses (10μs – state C in Fig. 4), whereas the full CF is recreated when using long pulses (1ms – state A in Fig. 4).

C. *Pulse Optimization for MLC Operation*

The results discussed in the previous sections allow identifying the physical mechanisms determining the observed conductance changes (initial jump and subsequent saturation) and how they depend on the applied SET pulse sequence. On one end, they demonstrate that using a sequence of subsequent identical pulses is an ineffective way of modulating device conductance (independently on the pulse width/amplitude considered). On the other hand, the results in Figs. 2 and 4 clearly indicate that the conductance at the end of a pulse sequence is strongly affected by V_P and/or t_P, since they allow to efficiently control the recreation of the CF during pulsed SET operation. We used these information to design optimized SET pulse trains allowing extending the conventional binary operation of HfO-based RRAMs into MLC operation. Figure 5(a) shows the conductance simulated for an optimized SET pulse train consisting of subsequent sequences of 20 pulses with increasing t_P (from 1μs to 10ms). At the end of each pulse sequence the conductance saturates to a different level, which enables MLC operation (the number of pulses is indeed chosen to obtain a good saturation). Despite requiring a significant number of relatively long pulses, the optimized pulse train is suitable for the operation of electronic synapses. Nevertheless, the overall programming time can be reduced using ultra-short pulse switching [18], [19]. The behavior in Fig. 5(a) can be understood by looking at the dynamics of the CF recreation. As new V_0 are generated, the potential drop across the oxide barrier reduces, Fig. 5(b), which in turns reduces the probability to generate additional vacancies. Increasing t_P gives more time to create new vacancies within the pulse sequence allowing reaching higher conductance levels, Fig. 5(a). Six well defined states are obtained, Fig. 6, a number that can be increased using a finer t_P modulation. The corresponding cumulative distributions exhibit good uniformity and separation between the states, Fig. 6(b). Simulations and experiments also show that the variability can be reduced by increasing the pulse amplitude above 0.9V or the number of sequence pulses (not shown for brevity). These results clearly demonstrate that optimizing the pulse scheme is crucial to achieve the MLC capabilities required for neuromorphic computing applications.

V. Conclusions

We used dedicated experiment and accurate physics-based RRAM simulations to investigate pulsed SET operation in TiN/HfO_2/Ti/TiN RRAM devices. Simulation results provide fundamental insights on the mechanisms governing pulsed SET operation and allow linking the microscopic properties of the recreated CF to the measured electrical characteristics. The acquired knowledge is then used to achieve MLC operation by means of optimized SET pulse sequences, providing a pathway for the use of TiN/HfO_2/Ti/TiN RRAM devices as electronic synapses in neuromorphic computing systems.

REFERENCES

[1] D. Garbin, E. Vianello, O. Bichler, Q. Rafhay, C. Gamrat, G. Ghibaudo, B. DeSalvo, and L. Perniola, "HfO_2-Based OxRAM Devices as Synapses for Convolutional Neural Networks," *IEEE Trans. on Electron Devices*, vol. 62, no. 8, pp. 2494-2501, Aug 2015, doi: 10.1109/TED.2015.2440102.

[2] P. A. Merolla, J. V. Arthur, R. A. Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, D. S. Modham, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science*, vol. 345, 6197, pp, 668-673, 2014, doi: 10.1126/science.1254642.

[3] D. Kuzum, S. Yu, and H.-S. Philip Wong, "Synaptic electronics: materials, devices and applications," *Nanotechnology* vol. 24, p. 382001, 2013, doi: 10.1088/0957-4484/24/38/382001.

[4] M. Suri, O. Bichler, D. Querlioz, O. Cueto, L. Perniola, V. Sousa, D. Vuillaume, C. Gamrat, B. DeSalvo, "Phase change memory as synapse for ultra-dense neuromorphic systems: Application to complex visual pattern extraction," *IEEE IEDM Tech. Dig.*, Dec. 2011, pp. 4.4.1–4.4.4, doi: 10.1109/IEDM.2011.6131488.

[5] S. Yu, Y. Wu, R. Jeyasingh, D. Kuzum, and H.-S.P. Wong, "An electronic synapse device based on metal oxide resistive switching memory for neuromorphic computation," *IEEE Trans. on Electron Devices*, vol. 58, no. 8, pp. 2729-2737, Aug 2011, doi: 10.1109/TED.2011.2147791.

[6] D. Ielmini, F. Nardi and S. Balatti, "Evidence for Voltage-Driven Set/Reset Processes in Bipolar switching RRAM," *IEEE Trans. on Electron Devices*, vol. 59, no. 8, pp. 2049-2056, August 2012, doi: 10.1109/TED.2012.2199497

[7] S. Yu, B. Gao, Z. Fang, H. Yu, J. Kang and H. S. P. Wong, "A neuromorphic visual system using RRAM synaptic devices with Sub-pJ energy and tolerance to variability: Experimental characterization and large-scale modeling," *IEEE IEDM Tech. Dig.*, Dec. 2012, pp. 10.4.1-10.4.4, doi: 10.1109/IEDM.2012.6479018.

[8] E. Covi, S. Brivio, A. Serb, T. Prodromakis, M. Fanciulli and S. Spiga, "HfO2-based memristors for neuromorphic applications, " *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2016, pp. 393-396. doi: 10.1109/ISCAS.2016.7527253.

[9] J. Woo, A. Padovani, K. Moon, M. Kwak, L. Larcher, and H. Hwang, "Linking Conductive Filament Properties and Evolution to Synaptic Behavior of RRAM Devices for Neuromorphic Applications," *IEEE Electron Device Lett.*, vol. 38, pp.1220-1223, September 2017, doi: 10.1109/LED.2017.2731859.

[10] J. Woo, K. Moon, J. Song, S. Lee, M. Kwak, J. Park, and H. Hwang, "Improved Synaptic Behavior Under Identical Pulses Using AlO_x/HfO_2 Bilayer RRAM Array for Neuromorphic Systems," *IEEE Electron Device Lett.*, vol. 37, pp.994-997, Aug 2016, doi: 10.1109/LED.2016.2582859.

[11] Ginestra™ software: www.mdlsoft.com.

[12] A. Padovani, L. Larcher, O. Pirrotta, L. Vandelli, and G. Bersuker, "Microscopic modeling of HfO_x RRAM operations: from forming to switching," *IEEE Trans. on Electron Devices*, vol. 62, no. 6, pp. 1998-2006, 2015, doi: 10.1109/TED.2015.2418114.

[13] D. Ielmini, "Modeling the universal set/reset characteristics of bipolar RRAM by field- and temperature-driven filament growth," *IEEE Trans. Electron Devices*, vol. 58, no. 12, pp. 4309–4317, Dec. 2011, doi: 10.1109/TED.2011.2167513.

[14] S. Long, X. Lian, C. Cagli, L. Perniola, E. Miranda, M. Liu, J. Suñé, "A model for the set statistics of RRAM inspired in the percolation model of oxide breakdown," *IEEE Electron Device Lett.*, vol. 34, no. 8, pp. 999–1001, Aug. 2013, doi: 10.1109/LED.2013.2266332.

[15] R. Degraeve, A. Fantini, N. Raghavan, L. Goux, S. Clima, Y. Y. Chen, A. Belmonte, S. Cosemans, B. Govoreanu, D. J. Wouters, Ph. Roussel, G. S. Kar, G. Groeseneken, and M. Jurczak, "Hourglass concept for RRAM: A dynamic and statistical device model", *21st IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)*, pp. 245-249, 2014, doi: 10.1109/IPFA.2014.6898205.

[16] N. Raghavan, "application of the defect clustering model for forming, SET and RESET statistics in RRAM devices," *Microel. Rel.*, vol. 64, pp. 54–58, 2016, doi: 10.1016/j.microrel.2016.07.139

[17] M. Lanza, "A Review on Resistive Switching in High-k Dielectrics: A Nanoscale Point of View Using Conductive Atomic Force Microscope," *Materials*, vol. 7, pp. 2155-2182, 2014, doi: 10.3390/ma7032155.

[18] D. M. Nminibapiel, D. Veksler, P. R. Shrestha, Ji-Hong Kim, J. P. Campbell, J. T. Ryan, H. Baumgart, and K. P. Cheung, "Characteristics of Resistive Memory Read Fluctuations in Endurance Cycling," *IEEE Electron Device Lett.*, vol. 38, pp. 326-329, March 2017, doi: 10.1109/LED.2017.2656818.

[19] D. M. Nminibapiel, D. Veksler, P. R. Shrestha, J. P. Campbell, J. T. Ryan, H. Baumgart, and K. P. Cheung, "Impact of RRAM Read Fluctuations on the Program-Verify Approach," *IEEE Electron Device Lett.*, vol. 38, pp. 736-739, June 2017, doi: 10.1109/LED.2017.2696002.