Impact of RRAM Read Fluctuations on the Program-Verify Approach

David M. Nminibapiel, Dmitry Veksler, Pragya R. Shrestha, Jason P. Campbell, Jason T. Ryan, Helmut Baumgart, Kin. P. Cheung, *Senior Member*, *IEEE*

***Abstract*— The stochastic nature of the conductive filaments in oxide-based resistive memory (RRAM) represents a sizeable impediment to commercialization. As such, program-verify methodologies are highly alluring. However, it was recently shown that program-verify methods are unworkable due to strong resistance state relaxation after SET/RESET programming. In this paper, we demonstrate that resistance state relaxation is not the main culprit. Instead, it is fluctuation-induced false-reading (triggering) that defeats the program-verify method, producing a large distribution tail immediately after programming. The fluctuation impact on the verify mechanism has serious implications on the overall write/erase speed of RRAM.**

***Index Terms*—Fluctuations, Instability, Program-Verify, Relaxation, RRAM**

I. Introduction

The well-publicized stochastic nature of Resistive Random Access Memory (RRAM) filamentary resistance states presents a formidable obstacle [1]–[3]. The stochasticity results in long resistance distribution tails which greatly limit the memory window. Subsequent efforts to control these distribution tails have relied heavily on "program-verification" methods [4]–[6]. It was recently shown that even careful implementation of "program-verify" can result in widely distributed HRS values whose distributions can broaden with time [4]. The broadening of the resistance distribution acts to close the memory window and effectively nullifies the efficiency of program-verify.

In this work, we revisit this troubling observation by studying the filament dynamics with enhanced measurement time resolution. Similar to [4], we observe large tails in the lower percentile of the High Resistance State (HRS) distribution that extend with time. However, we attribute these large initial tails to stochastic resistance state fluctuations which falsely trigger the verification mechanism and not rapid relaxation [4], [7] or other slowly changing mechanisms [8]–[10]. Recently it was shown that the occurrence of resistance fluctuations, or fluctuation probability, decays to a constant (non-zero) level within the first hundred microseconds after programming [11]. More importantly, this fluctuation probability decays without a concomitant decay in the average fluctuation amplitude. Without a change in fluctuation amplitude, the fluctuation-induced resistance distribution will not change with time. Thus, the increase in resistance dispersion with time suggests an underlying relaxation which in our case plays a minor role in the failure of the program-verify algorithm.

Manuscript received February 28, 2017; revised April 06, 2017; accepted April 15, 2017.

D. M. Nminibapiel is with the Engineering Physics Division at National Institute Standards and Technology (NIST), Gaithersburg, MD 20899 USA and also with the Department of Electrical and Computer Engineering at Old Dominion University, Norfolk, VA 23529 USA. (email: david.nminibapiel@nist.gov)

P. R. Shrestha is with Theiss Research, La Jolla, CA USA and the Engineering Physics Division at NIST, Gaithersburg, MD 20899 USA

D. Veksler, J.-H. Kim, J. P. Campbell, J. T. Ryan and K. P. Cheung are with the Engineering Physics Division at NIST, Gaithersburg, MD 20899 USA (email: kin.cheung@nist.gov).

H. Baumgart is with the Department of Electrical and Computer Engineering at Old Dominion University, Norfolk, VA 23529 USA.

II. Experimental Details

Experiments were done on 200 nm x 200 nm crossbar RRAM with TiN\HfO_2(5.8 nm)\Ti\TiN stacks using programming pulses with fixed width (≈ 100 ps) and amplitude. Spurning convention [4]–[6], the fast pulses are applied with a 50 Ω terminated probe with *no* current-limiting elements [12] as shown in Fig 1 (a). We rely on the ultra-short pulse timing to ensure that the energy delivered per pulse is small [13] and well controlled. Initial forming was performed with the same ultra-short pulses ranging in amplitude from +3 V to +4 V. Switching pulse amplitudes of ≤ |2| V were used for SET/RESET.

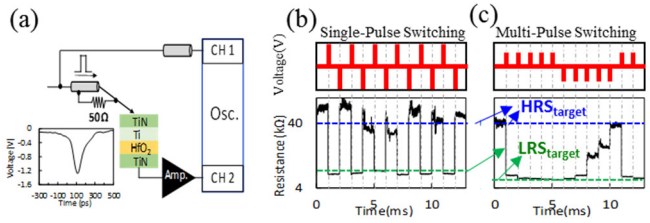

Fig. 1. (a) High speed measurement setup with 50 Ω termination at the probe tip (inset: 100 ps RESET pulse). Notice the absence of a series current limiting element. Varying the CUSPP voltage pulse amplitude or resistance target levels allows for resistance switching via (b) larger energy single pulses or (c) smaller energy multiple pulses.

The resistance of the RRAM states is monitored continuously by a low DC voltage, V_{READ}, (-8 mV). Instead of a read pulse, this approach is adopted to provide a continuous recording of the resistance evolution during programming. In this study, the read-out duration is fixed at 1 s, unless otherwise stated. The speed at which the resistance can be measured is limited only by the measurement circuit RC time constant of 30 ns.

A fast comparator with response time less than 5 ns automatically (1) compares the instantaneous state of the RRAM versus the target resistance and (2) orchestrates cycling by switching the pulse polarity after reaching the target resistance. This Compliance-free Ultra-Short Smart Pulse Programming (CUSPP) approach allows for cycling as fast as 5 MHz and supports the tuning of the pulse amplitude to allow the investigation of switching/forming via single pulse or

multiple pulses as shown in Fig 1 (b) and Fig 1 (c). The CUSPP functionality has been demonstrated on multiple RRAM cells (not shown here) and the experimental results presented in this manuscript should be viewed as representative for an array of RRAM elements.

III. Results and Discussion

Fig. 2 shows the distributions of the low resistance state (LRS) and high resistance state (HRS) values for a representative cell measured over 13,000 switching cycles at four different post-programming time windows (10 µs, 100 µs, 1 ms and 1 s; all resistance measurements are performed with 25 µs integration time). Target LRS and HRS resistances were set to 10 kΩ and 100 kΩ, respectively; unless otherwise stated. It is apparent that the resistance distribution measured 10 µs after programming extends past the target levels (dashed lines in Fig. 1b and 1c) for both LRS and HRS, therefore, reducing the target (10x) resistance window. In addition, similar to [4], the distributions broaden with time, leading to a complete window closure (HRS and LRS distributions overlap) 1 s after the programming event. The remainder of this work focuses on the HRS distribution, as it dominates the window closure.

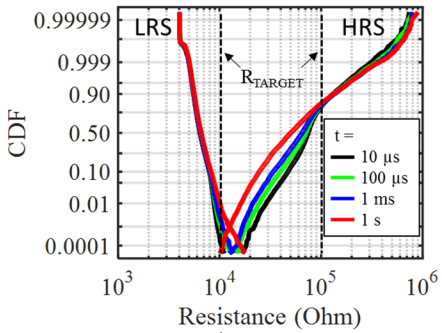

Fig. 2. Cumulative distribution (10^4 cycles) of LRS and HRS measured 10 µs, 100 µs, 1 ms and 1 s post-programming. A significant tail extends beyond the target resistances and is observed even at 10 µs after programming. The distribution tail does widen further with time, but the degree of this widening is a much smaller in comparison with the initial measured distribution width (measured at 10 µs) which is driven by fluctuations of state resistances.

Because a constant V_{READ} operation is utilized during cycling, a continuous recording of the filament evolution post programming is acquired and analyzed to directly elucidate the resistance state behavior. Fig. 3(a) is a representative temporal trace of the resistance evolution during a 3-pulse RESET process followed by a single pulse SET process. Immediately following each voltage pulse (pulse location indicated by markers), there is a large transient resistance fluctuation. These resistance fluctuations are not measurement system artifacts and are completely absent in calibration measurements of thin film reference resistors of similar values. There are, of course, many published results of random telegraph noise (RTN) like fluctuations which impact the retention of HfOx-based RRAM devices over tens of seconds [14] – [16]. While there may be some commonality with our observations of large post-programming transient fluctuations, we note that the time scales differ by approximately 6 orders of magnitude. Thus, if they are of a common origin, then the mechanism must span this rather wide time range.

These large transient fluctuations unsurprisingly force the verification to occur on unstable resistance states. In our measurements, the transient fluctuations trigger the fast comparator somewhat pre-maturely on unstable resistance states. This results in broadening of the resistance distributions, with the distribution tails penetrating deeply below/above (for HRS/LRS) the target resistance threshold (see Fig. 2). These distribution tails may lead to the incorrect perception that the CUSSP program-verification procedure has failed since there is no verification-mandated state bounding. However, this is not the case as the high fidelity continuous read operation of the measurement setup provides clear evidence of fluctuations that exceed the trigger point and initiate the start of HRS (Fig. 3 (a)). At the start of HRS, defined here as t = 0, both LRS and HRS distributions are bounded, however stochastic fluctuations almost immediately after programming degrade the boundaries as shown in Fig. 2.

We note that a slower comparator (verification operation) will not avoid the fluctuations completely. Recent analysis [11] of the fluctuation behavior post programming show that fluctuations continue to persist for an entire millisecond and while the occurrence of fluctuations decreases, their amplitudes remain unchanged. Therefore, a post-programming wait/delay still encounters fluctuations large enough to false trigger program verification. This calls into question whether this instability is related to resistance state relaxation or is a natural feature of the stochastic switching mechanism.

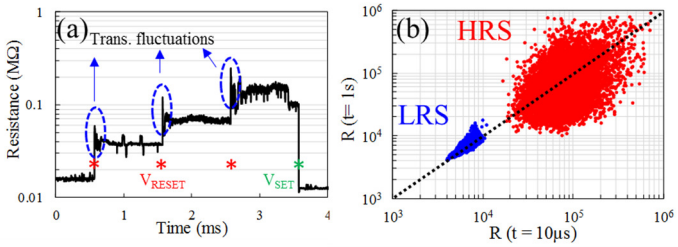

Fig. 3. (a) Measured resistance (V_{READ} = -8 mV) as a function of time (with 100 ns resolution) during programming. The first three pulses perform a RESET operation, while the last pulse performs a SET operation. Note, SET and RESET targets are 20 kΩ and 200 kΩ, respectively. The target is reached after the third V_{RESET} pulse, signifying the start of HRS. Notice transient resistance fluctuations immediately after programming pulse. (b) Time lag representation of the resistance states taken at 10 µs and 1 s post-programming (25 µs averaging window). The black line is drawn to show perfect correlation. For both LRS and HRS, the initial (10 µs) resistance value can either randomly decrease (below line) or increase (above line) with longer time delay.

Fig. 3 (b) shows a time lag representation of the resistance states taken 10 µs and 1 s post programming to examine any possible tendency of the state instability. Assuming a resistance state maintains the same value for an entire second, the data should follow the dashed line (black). Therefore, the dashed line represents perfect stability of the post programmed state. In this representation, a deviation above or below the line is an indication that the initial (10 µs) resistance has increased or decreased, respectively after 1s. It is evident that for both LRS and HRS there is instability in the post-programmed state that

causes the resistance to randomly increase and decrease. However, 72.3% of HRS tend to decrease while 65.4% of the LRS tend to increase after 1 s, leading to severe window closure. HRS dominates window closure due to larger magnitude of the deviations. The trend towards lower values for high resistance states is consistent with the observation that majority of large fluctuations tend to decrease the initial HRS value [11].

Still, it is incorrect to assign the instability shown in Fig. 3(b) to fluctuations only. As one would expect, the filament configuration (resistance state) changes in each cycle (large resistance distribution). It is conceivable that the resistance may increase or decrease based on the specific thermodynamics of the conductive filament. This cycle dependent relaxation would produce the same time lag plot. To differentiate between stochastic fluctuations and relaxation, we examine the resistance trends at intermediate times between the initial (10 µs) and final (1 s) read.

Since, the broad HRS distributions arise from unstable resistance state verification, we can *artificially* null out this effect by normalizing the distribution to a more stable value (10 µs post-programming). Fig. 4 (a) presents the probability distributions in such an exercise where we take the ratio of the resistance measured at some time, *t*, post-programming, *R(t)* to the resistance measured 10 µs post programming, *R(10µs)*. The choice of R(10µs) as the reference is arbitrary, since choosing other time delays (not shown here) did not change the results of Fig. 4. As such, at *t* = 10 µs, the ratio *R(t)/R(10µs)* results in the vertical line (black) located at value 1 and serves as a reference. The distributions corresponding to ratios taken 100 µs, 1 ms and 1 s post-programming extend above and below this reference line, indicating a shift of the initial resistance value towards higher and lower resistance values, respectively. We note that both cycle to cycle fluctuations and relaxations would result in deviations from the reference line (1) shown in Fig. 4 (a).

We sorted the distribution to select all the final (t = 1s) resistance states (red curve Fig. 4(a)) that ended either below (*(R(1s)/R(10µs)* < 1) or above (*(R(1s)/R(10µs)* > 1) the reference line. We then examined these distributions at intermediate read times to test if the states consistently trended towards these final read values. In other words, did the resistance state relax over the course of the read? If so, any intermediate read value should trend towards its final read value. This is our criteria for a relaxation process. These two subsets (below and above the reference line) are illustrated in Fig. 4(b) and Fig. 4(c), respectively. Note that, all the final distributions for Fig. 4(b) are below the reference line, while all the final distributions for Fig. 4(c) are above the reference line. Thus, Fig. 4b illustrates a distribution of all HRS cycles which decreased in resistance after 1 s and Fig. 4 (c) illustrates the distribution of all HRS cycles which increased in resistance after 1 s. If these distributions were indeed due to a relaxation effect, then we should observe the same distribution trends for any randomly chosen read-time window post-programming. As shown in Figs. 4b and 4c, this is clearly not the case. Fig. 4 (b) only contains those cycles in which the 1 s resistances finish below the 10 µs reference value. Yet at intermediary times (100 µs and 1 ms), the distributions span both above and below the reference line. Similarly, Fig. 4 (c) only contains those cycles in which the 1 s resistances finish above the 10 µs reference line. Yet at intermediary times (100 µs and 1 ms), the distributions span both above and below the reference line. While inconsistent with a relaxation effect, this data is notably consistent with a system dominated by random fluctuations.

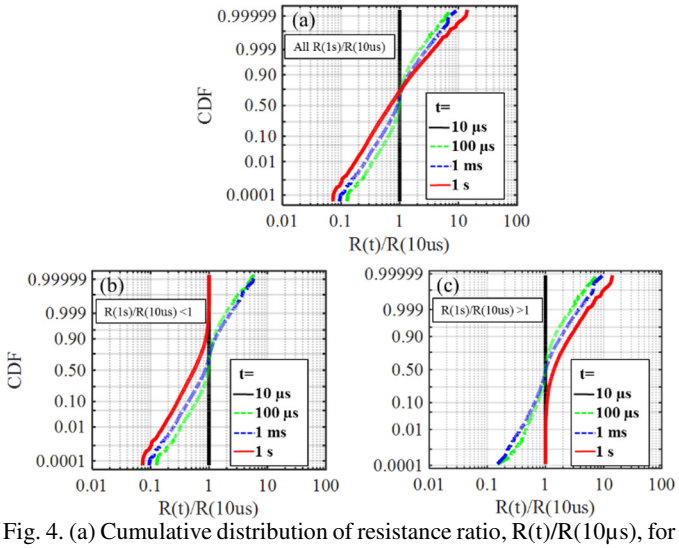

Fig. 4. (a) Cumulative distribution of resistance ratio, R(t)/R(10µs), for t = 10 µs, 100 µs, 1 ms and 1 s. The reference distribution (t = 10 µs) widens with time which may be attributed to fluctuations or relaxation. Distributions for all HRS which, after 1 s, end up lower (b) or higher (c) than the reference distribution. The appearance of values above and below the references at intermediate times (t = 100 µs and t = 1s) suggest that fluctuations dominate the dispersion.

From this analysis, fluctuations seem to be the major culprit of window closure. As mentioned above, no noticeable decay in fluctuation amplitude was observed up to 1 ms. Indeed, one may even infer from data in [11] that fluctuation amplitude remains similar out to 1 s, otherwise the fluctuation amplitude decay will compensate the relaxation by tightening the resistance distribution. These observations suggest that a simple delayed read strategy does not eliminate the appearance of a wide distribution post-programming.

The observed minor impact of relaxation in these measurements is markedly different than in earlier works [4]. The reduced relaxation role may be attributed to a variety of entities including: (1) differences in RRAM materials, (2) differences in the thermodynamics surrounding the conductive filament and the insulating matrix which are afforded by our non-compliance ultra-fast programming approach, and (3) the differences associated with pulsed versus continuous read operations. Regardless of these differences, it is clear that fluctuations dominate our observations.

IV. Conclusions

We identify that the failure of program-verify is largely due to stochastic resistance fluctuations. These fluctuations function to trigger the verification mechanism on unstable resistance states, resulting in a broad unbounded resistance distribution immediately (10 µs) after programming. By examining the instability behavior with time, we determined that the relaxation process is dominated by these large stochastic fluctuations.

V. REFERENCES

[1] R. Waser, R. Dittmann, G. Staikov, and K. Szot "Redox based resistive switching memories Nanoionic mechanisms, prospects, and challenges," Adv. Mater., vol. 21, pp. 2632-2663, Jul. 2009 doi 10.1002/adma.200900375

[2] H. -S. P. Wong, H. -Y. Lee, S. Yu, Y. -S Chen, Y. Wu, P. -S. Chen, B. Lee, F. T. Chen and M.-J. Tsai, "Metal-oxide RRAM", *Proc. IEEE*, vol. 100, pp. 1951-1970, May 2012, doi 10.1109/JPROC.2012.2190369

[3] S. M. Yu, X. Guan and H. -S. Wong, et al., "On the stochastic nature of resistive switching in metal oxide RRAM: Physical modeling, monte carlo simulation, and experimental characterization" in *IEEE Int. Electron Devices Meeting (IEDM),* pp., Dec. 2011, doi 10.1109/IEDM.2011.6131572

[4] A. Fantini, G. Gorine, R. Degraeve, L. Goux, C.Y. Chen, A. Redolfi, S. Clima, A. Cabrini, G. Torelli and M. Jurczak, "Intrinsic program instability in HfO_2 RRAM and consequences on program algorithms," in *IEEE Int. Electron Devices Meeting (IEDM)*, pp. 7.5.1-7.5.4, Dec. 2015, doi: 10.1109/IEDM.2015.7409648.

[5] F.M. Puglisi, C. Wenger, and P. Pavan, "A novel program-verify algorithm for multi-bit operation in HfO2 RRAM", *IEEE Electron Device Lett.*, vol. 36, pp. 1030 - 1032, Aug. 2015, doi 10.1109/LED.2015.2464256

[6] K. Higuchi, T. Iwasaki and K. Takeuchi. "Investigation of verify-programming methods to achieve 10 million cycles for 50nm HfO2 ReRAM", *IEEE Int. Memory Workshop (IMW),* pp. 1-4, May 2012, doi: 10.1109/IMW.2012.6213665.

[7] C. Wang, H. Wu, B. Gao, L. Dai, N. Deng, D.C. Sekar, Z. Lu, M. Kellam, G. Bronner, and H. Qian, "Relaxation effect in RRAM arrays: demonstration and characteristics", *IEEE Electron Device Lett.*, vol. 37, pp. 182-185, Dec. 2015, doi: 10.1109/LED.2015.2508034

[8] S. Clima, Y. Y. Chen, A. Fantini, L. Goux, R. Degraeve, B. Govoreanu, G. Pourtois and M. Jurczak, "Intrinsic tailing of resistive states distribution in amorphous HfOx and TaOx based resistive random access memories", *IEEE Electron Device Lett.*, vol. 37, pp. 769-771, Aug. 2015, doi: 0.1109/LED.2015.2448731

[9] S. Ambrogio, S. Balatti, V. McCaffrey, D. C. Wang, and D. Ielmini, "Noise-induced resistance broadening in resistive switching memory─Part II: Array Statistics," *IEEE Trans. Electron Devices*, vol. 62, no. 11, pp. 3812-3819, Nov. 2015, doi:10.1109/TED.2015.2477135.

[10] Y. Pan, Y. Cai, Y. Liu, Y. Fang, M. Yu, S. Tan and R. Huang "Microscopic origin of read current noise in TaOx-based resistive switching memory by ultra-low temperature measurement," *Appl. Phys. Lett*., vol. 108, p. 153504, 2016, doi:10.1063/1.4945790.

[11] D. M. Nminibapiel, D. Veksler, P. R. Shrestha, J. -H. 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, Jan. 2017, doi: 10.1109/LED.2017.2656818.

[12] P. R. Shrestha, D. Nminibapiel, J. H. Kim, J. P. Campbell, K. P. Cheung, S. Deora, G. Bersuker, and H. Baumgart, "Energy control paradigm for compliance-free reliable operation of RRAM," *IEEE Int. Rel. Physics Symp. (IRPS)*, pp. 1-5, Jun. 2014, doi: 10.1109/IRPS.2014.6861164.

[13] P. Shrestha, D. M. Nminibapiel, J.-H. Kim, H. Baumgart, K. P. Cheung and J. P. Campbell, "Compliance-free Pulse Forming of Filamentary RRAM", ECS Transactions, vol. 75, pp. 81 – 92, 2016, doi: 10.1149/07513.0081ecst

[14] F. M. Puglisi, L. Larcher, A. Padovani and P. Pavan, "A complete statistical investigation of RTN in HfO2-based RRAM in High Resistive State", IEEE Trans. Electron Devices, vol. 62, pp. 2606-2613, Aug. 2015, doi 10.1109/TED.2015.2439812

[15] N. Raghavan, R. Degraeve, A. Fantini, L. Goux, S. Strangio, B. Govoreanu, D. J. Wouters, G. Groeseneken and M. Jurczak, "Microscopic origin of random telegraph noise fluctuations in aggressively scaled RRAM and its impact on read disturb variability" *IEEE Int. Reliab. Phys. Symp. (IRPS)*, pp. 5E.3.1-5E.3.7, Jun. 2013, doi: 10.1109/IRPS.2013.6532042

[16] S. Balatti, S. Ambrogio, A. Cubeta, A. Calderoni, N. Ramaswamy, and D. Ielmini, "Voltage-dependent random telegraph noise (RTN) in HfO_X resistive RAM," in *IEEE Int. Reliab. Phys. Symp. (IRPS)*, pp. MY. 4.1-MY.4.6, Jun. 2014 doi: 10.1109/IRPS.2014.6861159.