



**HAL**  
open science

## Design Considerations Towards Zero-Variability Resistive RAMs in HRS State

Hassen Aziza, K. Coulie, W. Rahajandraibe

► **To cite this version:**

Hassen Aziza, K. Coulie, W. Rahajandraibe. Design Considerations Towards Zero-Variability Resistive RAMs in HRS State. 2021 IEEE 22nd Latin American Test Symposium (LATS), Oct 2021, Punta del Este, France. pp.1-5, 10.1109/LATS53581.2021.9651758 . hal-03511410

**HAL Id: hal-03511410**

**<https://hal.science/hal-03511410v1>**

Submitted on 4 Jan 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Design Considerations Towards Zero-Variability Resistive RAMs in HRS State

H. Aziza, K. Coulié, W. Rahajandraibe  
IM2NP, UMR CNRS 7334, Aix-Marseille Université, 38 rue Joliot Curie  
F-13451- Marseille, France  
hassen.aziza@univ-amu.fr

**Abstract**— Resistive RAM (RAM) intrinsic variability is widely recognized as a major hurdle for widespread adoption of the technology. Moreover, the deeper we go into the High Resistance State (HRS), the higher the variability. In this context, this paper proposes circuit level design strategies to mitigate HRS variability. During the RESET operation, the programming current is strictly controlled while the voltage across the RRAM cell is regulated. From a design standpoint, a write termination circuit is used to constantly sense the programming current and stop the RESET pulse when the preferred RESET current is reached. The write termination is combined with a voltage regulator which provides a strict control of the RESET voltage. The paper first reviews the RRAM variability phenomenon. Then, an optimized programming scheme is developed to control the HRS state to approach zero-variability. Compared to the classical fixed-pulse programming scheme, variability is reduced by 99%.

**Keywords**— Resistive RAM, RRAM, Oxide-based RAM, OxRAM, variability, current control, voltage regulation.

## I. INTRODUCTION

Emerging technologies such as Resistive RAMs (RRAMs) are attracting considerable attention due to their tempting characteristics such as high scalability, CMOS compatibility and non-volatility [1]. Bipolar Oxide-based Resistive RAM (so-called OxRAM) is focusing strong interest as a candidate for densely packed cross-point 3D-arrays [2], high performance Storage Class Memory (SCM) [3], as well as a new component able to unlock the full potential of disruptive computing architecture such as edge-AI [4]. While aggressive scalability and easy manufacturability of HfO<sub>2</sub>-based RRAM has been recently demonstrated [5], the stochastic variability in the device operation still needs mitigation schemes to enable RRAM commercialization. Therefore, there is an urgent need for robust design strategies at the circuit level able to alleviate the impact of variability on RRAM circuit's performances.

While improving the device variability is mostly a manufacturing matter [6], the key factor for minimizing its impact on reliability is designing variation-tolerant circuits. Typically, memories use peripheral or assist circuits to improve the chip-level reliability [7]. These circuits are not only meant to mitigate the existing manufacturing reliability such as time-zero process variability, but also to track device degradation over time. For RRAMs, write current control remains one of the main factors for improving key performance markers of the technology. Indeed, it has been demonstrated that high endurance and retention are closely linked to the programming currents since excessive and insufficient currents can result in failure [8-9]. Thus, not only new test schemes have been opposed [10] but also new programming algorithms and write assist circuits which play a crucial role on performance and yield optimization [11]. Although several write assist circuits have appeared in the literature [12-14], they still suffer from several issues related

to a strict control of the operating currents, flexibility, and integration cost. Furthermore, typical memory circuits include bit-line voltage regulators to deal with problems related to write voltage regulation, which is a crucial parameter affecting the programming operation. Indeed, voltage drops due to an important load combined with bit-line capacitances can lead to weak or imprecise program operations [15].

In this context, this paper introduces design techniques to achieve a strict control of currents/voltages involved during the RRAM RESET operation. To the authors' knowledge, this is the first work combining a "lightweight" (i.e., a dozen of transistors) write termination for current control combined with a regulator for voltage control.

The paper is organized as follows: Section II analyses RRAM variability based on experimental data. Section III develops the proposed design scheme to mitigate the variability at the circuit level. In section IV, the impact of variability on a RRAM memory array is analysed to demonstrate the robustness of our approach. Finally, Section V concludes the paper.

## II. RRAM VARIABILITY

### A. Experimental evidence

The application of an external voltage pulse across the RRAM cell enables a transition of the device from a High Resistance State (HRS)/Low Resistance State (LRS) after a SET/RESET (RST) operation respectively. In RRAM, variability encompasses temporal (cycle to cycle or C2C) and spatial (device to device or D2D) variability and leads to poor uniformity of HRS and LRS resistances [9].

To investigate the impact of variability on the RRAM resistance states uniformity, an 8x8 fabricated 1T-1R array presented in Fig. 1a is considered for measurements. Word Lines ( $WL_X$ ) are used to select the active row, Bit-Lines ( $BL_X$ ) are used to select active columns during a SET operation and Source Lines ( $SL_X$ ) are used to RST a whole memory word or a specific cell. Fig. 1b presents the micrograph of the memory array test chip fabricated in a 130 nm CMOS technology.

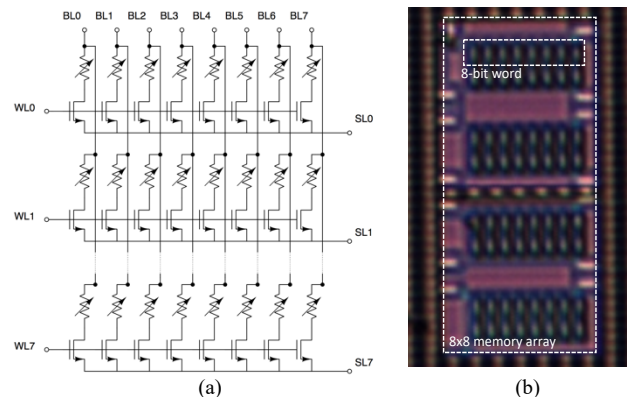


Fig. 1. (a) 8x8 OxRAM array test chip and (b) corresponding micrograph.

Experiments are performed using a Keysight B1500 semiconductor parameter analyzer. The memory array is first formed. Then, memory cells are RST one by one to extract the HRS resistance. After RST, cells are SET to extract the LRS resistance. HRS and LRS variability is demonstrated after a RST/SET cycle applied two times to each of the 64 cells of the memory to catch D2D variability (Fig. 2a) and after a RST/SET cycle applied 128 times to a single cell across the memory array (located at BL3-WL3) to catch C2C variability (Fig. 2b). Note that both D2D and C2C variability impacts HRS and LRS resistances, with a much more pronounced impact for the HRS state, mainly when D2D is concerned. The effect of variability (combining D2D and C2C) can be seen more clearly in the cumulative probability plot shown in Fig. 2c obtained after 500 consecutive RST/SET cycles applied to the memory array (500x64 cells). These experimental results make clear that variability mitigation circuits are needed to counterbalance variability, and foremost HRS variability.

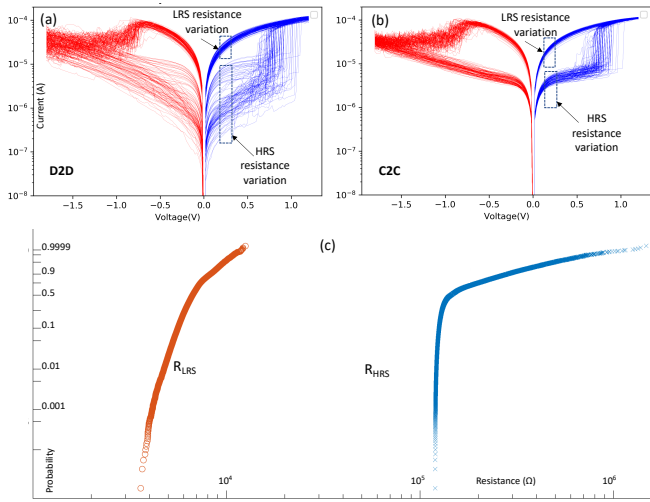


Fig. 2. RRAM I-V characteristics highlighting (a) D2D variability and (b) C2C variability. (c) HRS and LRS resistance distributions.

### B. OxRAM model

To assess the impact of variability at the circuit level, a compact OxRAM model [16] calibrated on measurements proposed in section II.A is used. The model accurately reproduces the stochastic switching nature of OxRAM cells. Indeed, as already revealed, RRAM operation is ruled by stochastic mechanisms (ion generation, ion migration in the resistive layer, etc.), leading to intrinsic variability. When a voltage is applied across the OxRAM cell, depending upon the voltage polarity, one or more Conductive Filaments (CFs) made out of oxygen vacancies are either formed or ruptured. Once the CFs are formed inside the metal oxide to bridge the top and bottom electrodes, the cell is switched in LRS and current can flow through the CFs. When CFs are ruptured, the cell switches in HRS. However, incomplete destruction of CFs can lead to different HRS levels, which is believed to be the main reason for HRS variability [17].

## III. VARIABILITY MITIGATION SCHEME

### A. High level architecture implementation

Fig. 3 shows the high level architecture of the variability mitigation scheme. We consider a regular OxRAM memory

array including word line ( $WL_x$ ), bit-line ( $BL_x$ ), source line ( $SL_x$ ) drivers, and sense amplifiers. The drivers select active SLs, BLs and WLs during a memory operation, while the sense amplifiers convert a read current to a logical value. Eight memory cells are grouped together in a word. The blocks highlighted in gray are the changes applied to the memory to integrate the variability mitigation scheme. We add one RST termination circuit per BL driver to monitor the RST current and a voltage regulator at the SL driver level to maintain a constant RST voltage. We also modify the control logic to stop the RST operation once the cell current equals a predefined reference current. The core element of our design scheme is the RST termination circuit that strictly controls the RST current in order to obtain a well-defined final HRS level, while the RST voltage is maintained constant through the voltage regulator.

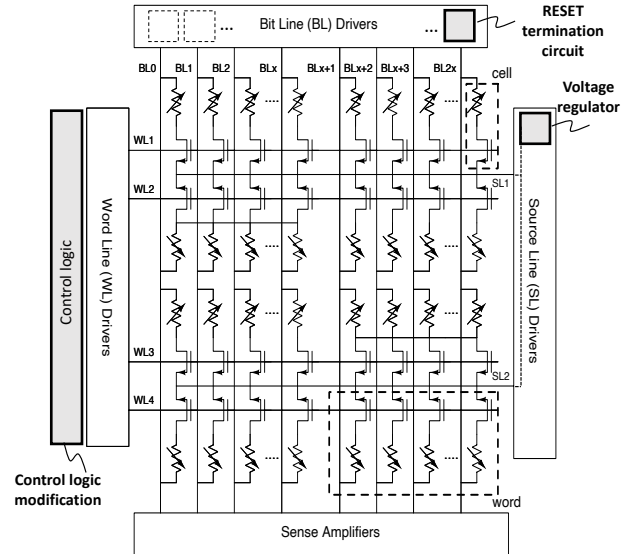


Fig. 3. Memory architecture including the modifications required for the implementation of our variability mitigation scheme.

### B. RESET Current control

Fig. 4a shows the transistor level implementation of the proposed RST termination circuit. During the RST operation, the RRAM cell current  $I_{cell}$  is copied by a n-MOS current mirror (M1, M2). The current mirror (M3, M4) is used to mirror the reference current  $I_{refR}$  (provided by M5, M6) which feeds the input of inverter I1. If  $(I_{cell} - I_{refR}) > 0$ , the inverter input  $A$  is set low and the comparator output  $out$  is set high. If  $(I_{cell} - I_{refR}) < 0$ , input  $A$  is set high and  $out$  is set low to terminate the RST operation (i.e., the RST operation is terminated when  $I_{cell}$  decreases down to  $I_{refR}$ ). It is worth noticing that  $I_{refR}$  current is derived from a ‘golden’ current source provided by a bandgap voltage reference circuit [18] included in the design to achieve stability over process, voltage and temperature (PVT). Fig. 4b shows the usage of the termination circuit in the memory architecture. For clarity, we only show the current copy stage of the RST termination circuit. The RST operation is performed by biasing the memory cell through the  $SL$  driver while  $WL0$  is activated.  $BL0$  connects to the current copy stage of Fig. 4a and sinks the cell current. When  $I_{cell}$  equals  $I_{refR}$  (i.e.,  $out$  signal is set low), the control logic triggers a *stop* pulse to the  $SL$  driver to terminate the RST operation. Note that each RST

termination needs to be included in every SL driver with a “twelve transistors” area overhead.

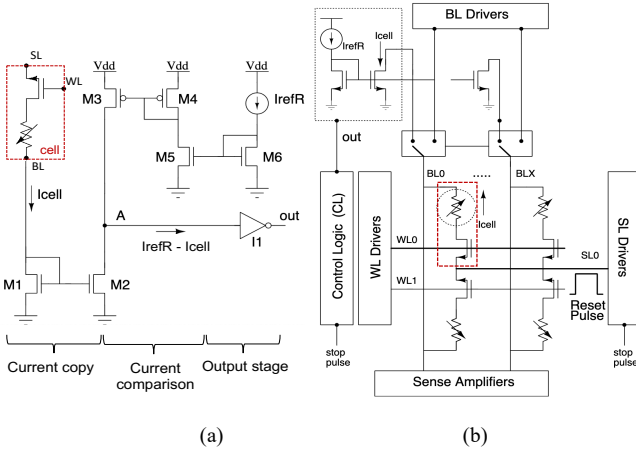


Fig. 4. (a) Self-terminating write circuit for RST operations (b) RST write termination implementation.

### C. RESET Voltage regulation

The RST voltage has to be regulated for the RST process to stop at a fixed voltage ( $V_{REG}$ ) so as to achieve a steady final HRS resistance regardless variability ( $R_{HRS} = V_{REG}/I_{refR}$ ). In order to obtain a flexible regulated voltage, the PMOS Low Drop voltage Regulator (LDO) architecture presented in Fig. 5 is considered [19]. A feedback loop made of resistances  $R_1$  and  $R_2$  feeds an error amplifier. The error voltage  $V_{error}$  controls the drain-to-source resistance of the PMOS device. Assuming the circuit consists of ideal components (i.e., no offset and high gain), the regulator output is given by (1):

$$V_{REG} = (1 + R_1/R_2) \cdot V_{ref} \quad (1)$$

The precision of  $V_{REG}$  depends essentially on the precision of the voltage reference  $V_{ref}$  since integrated resistors can be made with good matching.  $V_{ref}$  being provided by the bandgap circuit. The latter generates a fixed (constant) voltage independent of PVT variations. Also, to meet low-power requirements, large resistors  $R_1$  and  $R_2$  are used in the regulation loop. The proposed regulator architecture is suitable for parallel programming (byte-based) as the current fed to the load can reach  $600 \mu A$ . The latter matches the worst case load corresponding to the maximal current drawn by 8 LRS cells reset in parallel. The load is made by the bit-line capacitor (denoted by  $C_{BL}$  or  $C_L$ ) and the memory cells equivalent resistance (denoted by  $R_{cell}$  or  $R_L$ ), associated with the access transistor resistances. Fig. 5 inset shows the LDO targeted values for the considered application.

The use of an LDO circuit requires a stability analysis since it forms a closed loop system. The frequency response is defined by the dominant pole (100 Hz) along with two additional poles (1 KHz and 10 KHz) that can lead to an unstable condition. Thus, a frequency compensation was conducted following the methodology presented in [19] where a source follower stage is introduced in between the input voltage and the LDO to improve the closed-loop system stability.

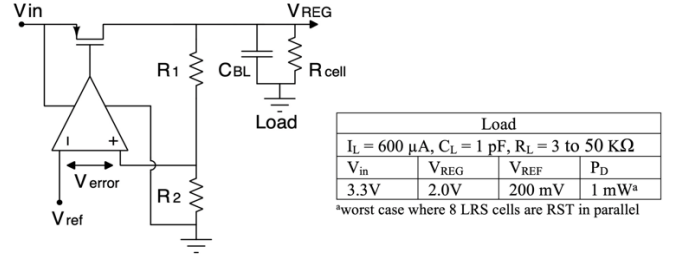


Fig. 5. Low Drop voltage Regulator architecture.

## IV. SIMULATION RESULTS

### A. Simulation setup

We implemented the memory circuit presented in Fig. 3 using a 130 nm High Voltage CMOS technology offering a 3.3 V supply voltage. A 3.3 V technology is required as the RRAM forming operation involves high voltages [20]. To verify the operation of our design scheme, SPICE simulations are performed using the Eldo simulator. In order to accurately evaluate the benefits of our proposed scheme on large memory arrays, BL and WL lengths have been modelled to mimic a 1 Mbyte array (made of 1024 WLs and 1024x8-bits BLs). As a BL is characterized by a parasitic capacitance distributed through its length, a 1 pF bit-line capacitance is used according to the targeted technology [21]. Additionally, parasitic resistances [22] distributed along BLs/WLs have been inserted, following the methodology developed in [23].

### B. Simulation results

Transient simulation results are presented in Fig. 6 after a RST operation associated with a compliance current  $I_{refR}$  equal to  $9 \mu A$ . The cell current  $I_{cell}$  gradually decreases down to  $I_{refR}$ . Beyond this point, the RST pulse is terminated by the write termination circuit, limiting the HRS resistance value to  $171 k\Omega$  with a write latency equal to  $2.1 \mu s$ . The standard RST pulse  $V_{RST\_std}$  is also reported. Adopting this standard pulse leads to a final HRS resistance value close to  $338 k\Omega$ . Also, using the conventional one-pulse programming approach without controlling the RST current/voltage can result in an uncontrolled spread in the HRS and subsequently in the LRS distributions. Note that the standard RST pulse width is set to  $3.8 \mu s$  to cover the worst cases during RST (i.e., tail bits in the switching parameter distributions).

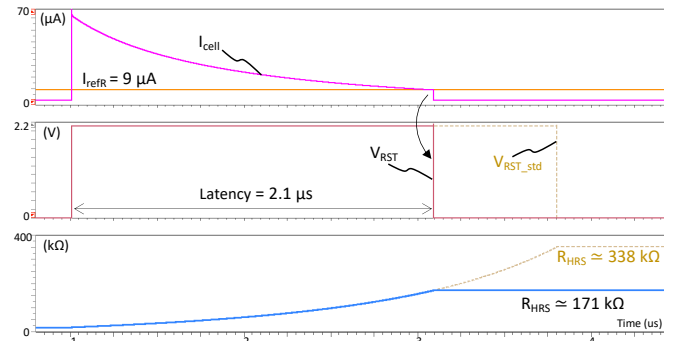


Fig. 6. Transient simulation results after a RST operation associated with a reference current  $I_{refR}$  equals to  $9 \mu A$ .

To demonstrate the robustness of our approach, a Monte Carlo (MC) analysis is conducted. In this analysis, only actual possible variations are reported, since cell variability is

generated based on a targeted OxRAM technology. Moreover, the variability (including transistor mismatch [24] [25]) also targets the CMOS subsystem and especially the memory cell access transistor as its impact on the memory cell electrical characteristics is dominant [26]. Process variation parameters used for CMOS transistors are provided by ST-Microelectronics (Crolles facility, France). For each simulation run, the MC analysis calculates every parameter randomly according to statistical distribution models, provided for active devices as well as for passive devices and cover corner cases. In this study,  $I_{refR}$  is set to 9  $\mu\text{A}$  to target an HRS resistance ( $\sim 170 \text{ k}\Omega$ ) one decade higher than the LRS resistance ( $\sim 17 \text{ k}\Omega$ ).

Fig. 7a presents the impact of variability on HRS distributions in the form of box plots after 1000 statistical runs following a RST operation performed in three different configurations. The first configuration referred to as “NO-CTRL” performs a RST operation using the fixed-duration pulse signal. The second configuration referred to as “I-CTRL” leverages on the write termination to provide a strict control of the RST current. The third configuration referred to as “I/V-CTRL” leverages on the LDO circuit to provide a regulated RST voltage in addition to current control. As expected, without control over current/voltage, the final HRS resistance is not limited ( $\mu = 337 \text{ k}\Omega$ ) and exhibits a large spread ( $\sigma = 69.5 \text{ k}\Omega$ ). As highlighted in Fig. 7b, with a strict control of the RST current, the final mean HRS value reaches  $171.2 \text{ k}\Omega$  with a standard deviation of  $676 \Omega$ , resulting in a variability reduction of 99%. The standard deviation is further reduced to  $630 \Omega$  in the “I/V-CTRL” configuration (99.1% reduction).

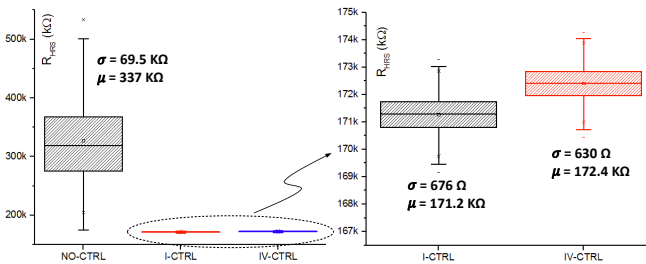


Fig. 7. (a) HRS box plots obtained after 1000 MC runs obtained for the “NO-CTRL”, “I-CTRL” and “I/V-CTRL” configurations. (b) Expanded view of “I-CTRL” and “I/V-CTRL” box plots.

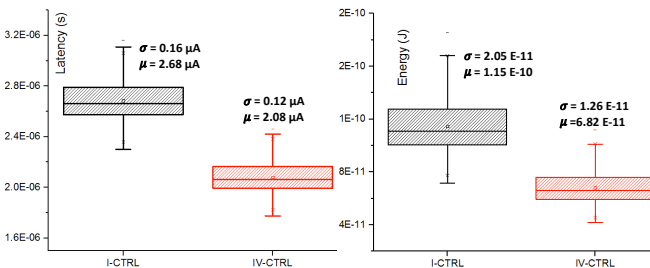


Fig. 8. (a) Latency and (b) Energy/cell box plots obtained after 1000 MC runs for “I-CTRL” and “I/V-CTRL” configurations.

Regarding the programming latency (Fig. 8a), for the “I-CTRL” configuration, the latency mean value is equal to  $2.68 \mu\text{s}$  versus  $2.08 \mu\text{s}$  for the “IV-CTRL” one, resulting in a 22% reduction. This difference is due to the regulation of the RST voltage in the “IV-CTRL” configuration, which improves the programming efficiency (i.e., voltage drop reduced). Regarding the energy (Fig. 8b), results are correlated with

latency results. The higher the latency, the higher the energy. For the “I-CTRL” configuration, the energy mean value is equal to  $115 \text{ pJ/cell}$  and only  $68.2 \text{ pJ/cell}$  for the “IV-CTRL” configuration (21% reduction). However, the proposed results do not include the LDO contribution in the “IV-CTRL” configuration ( $250 \text{ pJ/cell}$ ), which gives a clear advantage to the “I-CTRL” configuration as far as energy is concerned. To summarize, it has been demonstrated that current control during the RST operation has a major impact on variability, while RST voltage regulation plays a minor role. However, voltage regulation reduces the RST latency (i.e., improves the RST speed and the overall programming speed as the SET latency is around  $100 \text{ ns}$  [13]).

## V. CONCLUSION

This paper proposes a simple and feasible solution to mitigate one of the most challenging issues in RRAMs namely, the variability. A new design scheme based on voltage/current control during the RST operation is introduced. MC simulation results show a 99% reduction of the variability, demonstrating the robustness of our approach. It is demonstrated that current control during the RST operation has a major impact on variability, while RST voltage regulation plays a minor role in variability reduction. Also, it has been demonstrated that voltage regulation reduces the RST latency, which is highly desirable for high speed applications.

## REFERENCES

- [1] H. Aziza (2020). Embedded Memories. Silicon Systems For Wireless Lan, 22, 199.
- [2] Y. Bai et al., “Study of Multi-Level Characteristics for 3D Vertical Resistive Switching Memory,” Scientific Reports, vol. 4, no. 1, 2014.
- [3] J.-M. Portal et al., “Design and simulation of a 128 kb embedded nonvolatile memory based on a hybrid RRAM (HfO<sub>2</sub>)/28 nm FDSOI CMOS technology,” IEEE Tr. on Nanotech. 16.4 (2017): 677-686.
- [4] H. Aziza et al., “A capacitor-less CMOS neuron circuit for neuromemristive networks,” IEEE NEWCAS, 2019.
- [5] C. Ho et al., “Integrated HfO<sub>2</sub>-RRAM to achieve highly reliable, greener, faster, cost-effective, and scaled devices,” 2017 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2017, pp. 2.6.1-2.6.4.
- [6] P. Pouyan et al., “RRAM variability and its mitigation schemes,” IEEE PATMOS, Bremen, Germany, 2016, pp. 141-146.
- [7] H. Aziza et al., “An Energy-Efficient Current-Controlled Write and Read Scheme for Resistive RAMs (RRAMs),” IEEE Access, 2020.
- [8] Ning S. et al., “50 nm AlxOy ReRAM program 31% energy, 1.6× endurance, and 3.6× speed improvement by advanced cell condition adaptive verify- reset”, SSE journal, vol. 103, pp. 64-72, 2015.
- [9] Liu H. et al., “Uniformity Improvement in 1T1R RRAM With Gate Voltage Ramp Programming”, IEEE Electron Device Letters, vol. 35, pp. 1224-1226, 2014.
- [10] M. Fieback et al., “Device-Aware Test: A New Test Approach Towards DPPB Level,” 2019 IEEE International Test Conference (ITC), Washington, DC, USA, 2019, pp. 1-10.
- [11] Wang G. M. et al., “Improving resistance uniformity and endurance of resistive switching memory by accurately controlling the stress time of pulse program operation”, Applied Physics Letters, vol. 106, 2015.
- [12] A. Chen et al., “Variability of Resistive Switching Memories and its Impact on Crossbar Array Performance,” IEEE IPRS, pp.10.1–10.4, 2011.
- [13] F. T. Chen, “Fast-Write Resistive RAM (RRAM) for Embedded Applications,” IEEE Des. Test Comput., vol. 28, no. 1, pp. 64–71, 2011.
- [14] R. Schemes et al., “A High-Speed 7.2-ns Read-Write Random Access 4-Mb Embedded Resistive RAM ( ReRAM ) Macro,” IEEE JSSC, vol. 48, no. 3, pp. 878–891, 2013.

- [15] C. Flocchi et al., "Program load adaptive voltage generator for flash memories," in *IEEE Journal of Solid-State Circuits*, vol. 32, no. 1, pp. 100-104, Jan. 1997.
- [16] M. Bocquet et al., "Compact Modeling Solutions for Oxide-Based Resistive Switching Memories (RRAM)," *Journal of Low Power Electronics and Applications*, vol. 4, no 1, pp. 1-14, 2014.
- [17] H. Basma et al., "RRAM device models: a comparative analysis with experimental validation," *IEEE Access* 7, 2019.
- [18] H. Banba et al., "A CMOS bandgap reference circuit with sub-1-V operation", in *IEEE Journal of Solid-State Circuits*, vol. 34, no. 5, pp. 670-674, 1999.
- [19] Crepaldi Paulo C. et al., "Structural Design of a CMOS Voltage Regulator for an Implanted Device," *Current trends and challenges in RFID* (2011).
- [20] H. Aziza et al., "True Random Number Generator Integration in a Resistive RAM Memory Array Using Input Current Limitation," in *IEEE Transactions on Nanotechnology*, vol. 19, pp. 214-222, 2020.
- [21] L. Welter et al., "Embedded high-precision frequency-based capacitor measurement system," *IEEE 19th International On-Line Testing Symposium (IOLTS)*, pp. 116-121, 2013.
- [22] J. Liang et al., "Effect of wordline/bitline scaling on the performance, energy consumption, and reliability of cross-point memory array", *ACM JETC*, Vol. 9, No. 1, 2013.
- [23] H. Aziza et al., "ReRAM ON/OFF resistance ratio degradation due to line resistance combined with device variability in 28nm FDSOI technology," *IEEE EUROSOCI-ULIS*, 2017, pp. 35-38.
- [24] Y. Joly et al., "Impact of hump effect on MOSFET mismatch in the sub-threshold area for low power analog applications," *ISICT* 2010.
- [25] Y. Joly et al., "Matching degradation of threshold voltage and gate voltage of NMOSFET after Hot Carrier Injection stress." *microelectronics reliability* 51.9-11 (2011): 1561-1563.
- [26] H. Aziza et al., "Evaluation of OxRAM cell variability impact on memory performances through electrical simulations", *NVMTS*, 2011.