Density Enhancement of RRAMs using a RESET Write Termination for MLC Operation

H. Aziza*, S. Hamdioui†, M. Fieback†, M. Taouil†, M. Moreau*

*IM2NP, UMR CNRS 7334, Aix-Marseille Université, 38 rue Joliot Curie, F-13451, Marseille, France

†Computer Engineering Laboratory, Delft University of Technology, Mekelweg 4, 2628CD, Delft, The Netherlands

***Abstract*— Multi-Level Cell (MLC) technology can greatly reduce Resistive RAM (RRAM) die sizes to achieve a breakthrough in cost structure. In this paper, a novel design scheme is proposed to realize reliable and uniform MLC RRAM operation without the need of any read verification. MLC is implemented based on a strict control of the cell programming currents of 1T-1R HfO_2-based RRAM cells. Specifically, a self-adaptive write termination circuit is proposed to control the RRAM RESET current. Eight different resistance states are obtained by varying the compliance current which is defined as the minimal current allowed by the termination circuit in the RESET direction.**

Keywords*— Multi-level cell, MLC, Resistive RAM, RRAM, Oxide-based RAM (OxRAM), variability, current control

I. Introduction

Different alternative memory concepts have been explored in the last twenty years aiming to overcome the major limitations of existing semiconductor memories, i.e. the volatility of RAM's and the slow programming and limited endurance of Flash [1]. Among these emerging technologies, Resistive RAMs (referred to as RRAM or ReRAM) are believed to be a good choice due to the advantages of simple structures offering low manufacturing costs, fast switching speed (~10 ns), small feature sizes (<10 nm), compatibility with current CMOS technology, and low voltage operation [2]. In an attempt to gain grounds in these highly competitive emerging memory market, Non Volatile Memories (NVMs) vendors are trying to squeeze more and more capacity into constantly shrinking silicon dies, thereby optimizing both storage density and cost benefits. Among the various technique used to increase the density of RRAMs [3-5], MLC operation is considered as one of the most promising properties of RRAM as it can increase density without much change to current technologies.

The MLC storage characteristics of RRAM have been reported in many studies showing that different resistance states can be achieved [3-7]. MLC can be implemented by varying the RRAM compliance current during the SET programming operation, or by varying the voltage during the RESET (RST) operation, or by varying the programming pulse widths and amplitudes during SET or RST operations. However, the related prior work has the following shortcomings: MLC operation is validated at the device level and design implications for MLC RRAM at the circuit and system levels remain to be explored. In particular, programming currents of the order of 500 µA [8] or 1 mA [9] have been reported at a device level which is incompatible with low power RRAM applications. Moreover, RRAM variability at the memory array level is not accounted for as well as the impact of the peripheral circuits on the cell performances. Only a few studies in the prior art explore applications of MLC RRAM at the circuit level and most of the work focuses on read-out circuits [10-12].

The proposed study advances the state of the art by proposing a new design scheme that addresses MLC programming properties that are missing or poorly achieved in other previously proposed works, including:

- A novel 3 bit/cell MLC architecture based on compliance current control during the RST operation, allowing a tight control of post-programming resistances.
- An implementation at a circuit level with a minimal area overhead.
- A validation at a circuit level considering trade-offs related to robustness, energy consumption and latency.

The remainder of this paper is organized as follows. Section II presents the RRAM technology along with conventional MLC approaches. In Section III, the MLC design scheme implementation is presented. Section IV presents simulation results. Section V discusses the proposed MLC strategy. Finally, Section VI concludes this paper.

II. OxRAM technology vs MLC Modes

Oxide-based RRAMs memories (so-called OxRAMs) are considered in this study [13]. Fig. 1a shows the basic 1T-1R memory cell where one MOS transistor is connected in series with an OxRAM cell made of a Metal-Insulator-Metal (MIM) structure (TiN/10nm-HfO_2/10nm-Ti/TiN stack). Fig. 1b presents a typical 1T-1R OxRAM I-V characteristic in logarithmic scale. Based on the I-V curve, the memory cell operation can be seen as follows: after an initial electro-FORMING (FMG) step [13], the memory element can be reversibly switched between the Low Resistance State (LRS) and High Resistance State (HRS). The resistance change is triggered by applying specific biases across the 1T-1R cell, i.e., V_{SET} to switch to LRS and V_{RST} to switch to HRS. The maximum current allowed by the select transistor is called the compliance current and is referred to as I_C in Fig. 1b. I_C controls the LRS resistance value in the SET state as well as the maximal RST current I_{reset}, and thus plays an essential role in limiting power consumption. Table I presents the different voltage levels used during the different operating stages.

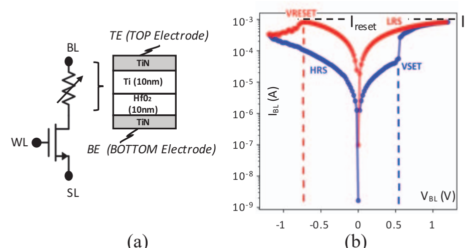

Fig. 1. (b) 1T-1R cell (b) OxRAM I-V characteristic in log scale.

TABLE I. STANDARD OPERATING VOLTAGES (CELL LEVEL)

	FMG	RST	SET	READ
WL	2V	3V	2V	3V
BL	2V	0V	1.2V	0.2V
SL	0V	1.2V	0V	0V

A. *OxRAM Variability*

Although OxRAM-based devices have shown encouraging properties, challenges remain, among which device variability is the main. Indeed, the resistance variance from cycle to cycle (C2C) and from device to device (D2D) can be very large. This inherent drawback of the technology has to be investigated because of its impact on MLC operation. Fig. 2a shows cumulative probability distributions obtained after 500 consecutive RST/SET cycles applied to an 8x8 elementary OxRAM memory array (500x64 cells). Experiments are performed using a Keysight B1500 semiconductor parameter analyzer and a 0.2 V READ bias voltage is used to extract R_{LRS} and R_{HRS} resistance distributions. The HRS distribution spread is more pronounced compared to the LRS spread, which is a common feature of the considered OxRAM technology [13]. These experimental results clearly indicate that a strict control of the HRS resistance is required to implement a reliable MLC HRS scheme.

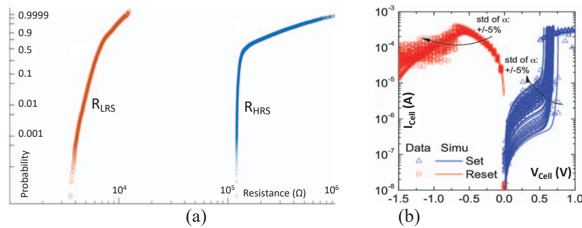

Fig. 2. I-V (a) HRS and LRS distribution measurement results. (b) Measured and corresponding simulated I-V characteristic obtained from TiN/HfO_2/Ti/TiN devices showing R_{LHS} and R_{HRS} variations after RST/SET.

For memory array simulations, a compact OxRAM model [14] calibrated on these measurements is used. The model accurately reproduces the stochastic switching nature of OxRAM cells. The variation is chosen to fit experimental data as presented in Fig. 2b where the model (lines) is consistent with experimental data (symbols) for SET (blue) and RST (red) operations. To mitigate the impact of variability on HRS/LRS resistances, it has been demonstrated that multi-step programming helps tolerate both temporal and spatial variations to obtain uniform intermediate states. However, the approach is energy inefficient as it involves a sequence of programming-and-verify operations [15].

III. MLC DESIGN SCHEME

A. *RST write termination circuit implementation*

The core element of our MLC design scheme is a self-terminating RST circuit that strictly controls the RST current to obtain eight different HRS levels (i.e., 3 bits/cell). During a RST operation, the circuit constantly compares the cell current to the reference current of the desired HRS level. Once these currents are equal, the driver terminates the RST operation. Fig. 3a shows the transistor level implementation of the proposed RST termination circuit. During the RST operation, the RRAM cell current I_{cell} is copied by an n-MOS current mirror (M1, M2). The current mirror (M3, M4) is used to mirror the reference current I_{refR} 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 to high. If (I_{cell}-I_{refR}) < 0, input *A* is set high and *out* is set to low to terminate the RST operation (i.e., the RST operation is terminated when I_{cell} decreases down to I_{refR}). I_{refR} is derived from a bandgap voltage reference circuit that is also included in a regular memory architecture [16]. Typically, I_{refR} is set to a few micro amperes to limit the HRS resistance to a few hundred kilo ohms. Note that the RST process is a negative feedback mechanism: as the current flows, the resistance of the cell increases, causing current to reduce. Fig. 3b 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. 3a 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.

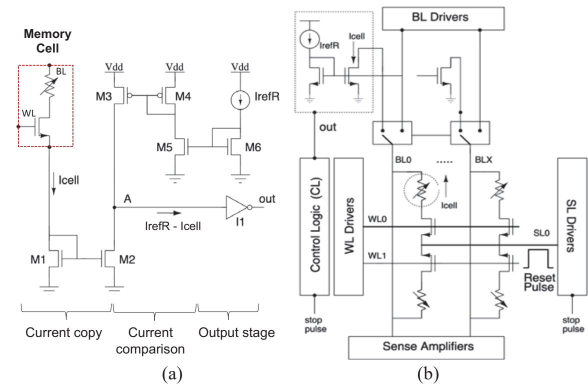

Fig. 3. (a) Self-terminating write driver for RST operations. (b) RST write termination implementation.

IV. CIRCUIT LEVEL EVALUATION

It is possible to define a relationship between the variation of the RST compliance current and the HRS resistance as presented in Fig. 4a. Fig. 4b is a subset of Fig. 4a in log scale showing the pseudo-linear variation of the resistance, with compliance currents ranging from 9 μA to 30 μA and resistance values ranging from 46.6 kΩ to 171 kΩ. These current and resistance ranges are considered for the MLC operation implementation. The deeper we go in the HRS state, the higher the variability and the RST latency. Hence, the maximal HRS value is limited to 195 kΩ. Regarding the minimal resistance, its value is set to 48 kΩ to maintain compliance currents below 30 μA.

Given the minimum HRS and maximum HRS resistances and the number of levels required, there are different schemes in determining the resistance values, including ISO-ΔR where the resistance is linearly spaced and ISO-ΔI where the programming current (inverse of resistance) is linearly spaced as described in [3]. The ISO-ΔI approach is adopted as the proposed MLC scheme is based on RST current control. Table II presents the 8 different states along with the corresponding compliance currents I_{refR} and HRS resistances.

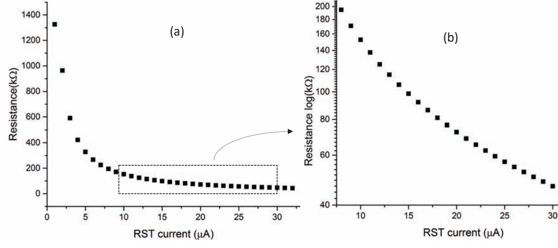

Fig 4. (a) Evolution of the HRS resistance versus the RST compliance. (b) HRS resistance in log scale for I_{refR} ranging from 9 µA to 30 µA.

TABLE II. ALLOCATION OF THE 8 RESISTANCE LEVELS RANGING FROM 46.6KΩ TO 171MΩ

State	111	110	101	100	011	010	001	000
I_{refR} (µA)	9	12	15	18	21	24	27	30
R_{HRS} (kΩ)	171	138	98.5	81.3	68.1	59.5	52.3	46.6

At the OxRAM device level, the resistance allocation strategy can be seen as a segmentation of the I-V plan by several I-V characteristics as shown in Fig. 5. Each characteristic being associated with a resistance state with a slope of $1/R_x$, with x the number of HRS states ranging from *0* to *n*. The READ operation is accomplished by applying a gate voltage to the memory cells (V_{Read}) and comparing the current drawn by the cell to currents provided by a set of reference current sources denoted by I_{Refx}, where x ranges from *0* to *n-1*. If 8 resistance states are targeted, *7* current references are required. Moreover, the DC value of each current reference needs to be located between the current provided by two consecutive memory states which are separated by a resistance margin ΔR. Note that ΔR takes into account the variability of the *n* resistance states. The latter is represented by the shaded area encompassing each characteristic.

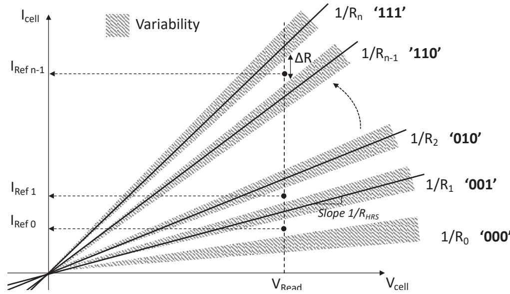

Fig. 5. MLC allocation strategy with the current sensing approach: the cell is read at V_{Read}, and the cell current is compared to fixed reference currents.

A. Simulation setup

We designed an 8-bit word elementary memory array circuit in a 0.13 µm High Voltage CMOS technology offering a 3.3 V supply voltage. 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 and the array architecture. Additionally, parasitic resistances distributed along BLs and WLs have been inserted in the design, following the methodology developed in [17]. Based on the proposed simulation setup, after SET, a RST pulse with different compliance currents is applied to the memory array. Then, HRS resistances, RST switching time (latency) as well as the energy consumption per cell value are extracted.

B. Transient simulations

Word programming is performed in 2 steps. Once an 8-bit word is addressed, each memory word is RST in parallel through the SL with a predefined compliance current set according to the data bus values. Transient simulation results are presented in Fig. 6 after a RST operation associated with a 9 µA compliance current. The cell current I_{cell} gradually decreases down to I_{refR} set to 9 µA. Beyond this point, the RST pulse is terminated by the write termination circuit, limiting the HRS resistance value to 171 kΩ with a write latency equal to 2.7 µs and an energy consumption/cell of 0.58 pJ. The standard (i.e., without the RST termination) RST pulse V_{RST_std} is also reported. Adopting this standard pulse would lead to a final HRS resistance value close to 1 MΩ.

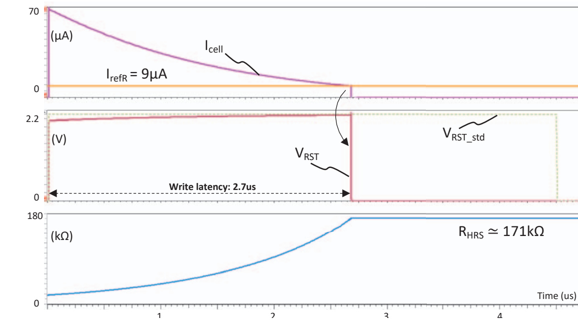

Fig. 6. Transient simulation results after a RST operation associated with a reference current value I_{RefR} equals to 9µA.

C. Monte Carlo (MC) analysis

To assess the robustness of our MLC design scheme, a Monte Carlo (MC) analysis is conducted. Variability (including transistor mismatch [18]) is chosen to target the OxRAM cells as well as the CMOS subsystem [19]. Process variation parameters used for CMOS transistors are provided by ST-Microelectronics (Crolles facility, France). Fig. 7a presents the impact of variability on HRS distributions in the form of box plots after 300 statistical runs following RST operations performed with compliance currents I_{refR} defined in Table II. The uniformity of the HRS is well-controlled. Indeed, having a strict control of the RST pulse through the RST compliance current contains the HRS resistance variation. However, when smaller I_{refR} are considered, the variability of the HRS state noticeably increases, but without causing distribution overlaps. Fig. 7b presents the RST latency evolution versus I_{refR}. The worst case scenario in terms of speed is associated with low I_{refR} values (the maximum latency reaches 3.39 µs for 9 µA). The energy/cell distributions are also reported in Fig. 8a showing that low compliance currents result in higher energy dissipation due to larger RST pulses. If the contribution of the current references is accounted during the RST operation (Fig. 8b), the impact of the energy dissipation/cell increases (shift in the order of 10 to 25 pJ).

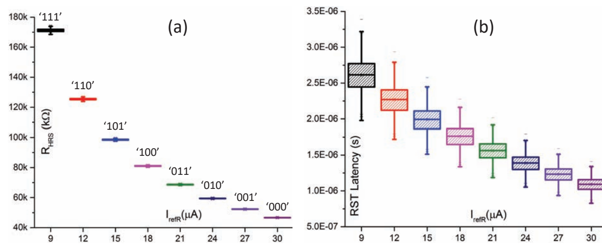

Fig. 7. (a) HRS and (b) RST latency box plots obtained after MC simulation for different RST compliance currents ranging from 9 to 30 μA.

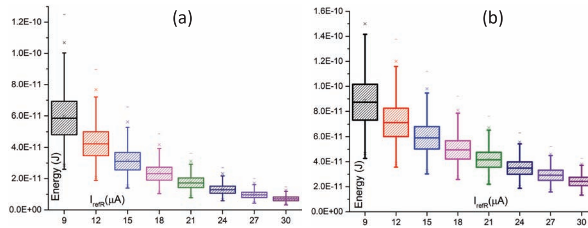

Fig. 8. (a) Energy/cell and (b) Energy/cell including I_{refR} current source contributions for different RST compliance currents.

V. Discussion

The eight resistance levels range from 46.6 kΩ to 171 kΩ. All the states have resistance values above 45 kΩ, limiting the READ current below 5 μA for most of the time. Therefore, this scheme is preferred from the energy standpoint. Regarding the latency, results provided in Fig. 8b do not reflect the SET operations preceding each RST operation. This is explained by the fact that the standard SET pulse is constant and common to any RST operation. The SET pulse is very short (~100 ns), which is a common feature of the considered technology and contributes 20 pJ/cell to the total energy dissipation [20]. Hence, in the worst case, the total energy/cell associated with a SET/RST cycle can reach 45 pJ. If the area is considered, the proposed implementation based on a write termination offers a minimal area overhead along with a minimal number of control signals compared to state-of-the-art write termination implementations [21] as only 8 transistors are required for each bit line (see Fig. 3a). The resistance margin between two consecutive states is remarkable. According to Fig. 7a, it ranges from 6.1 kΩ (between states '000' and '001') to 47.3 kΩ (between states '110' and '111'). Table III summarizes the proposed MLC design scheme and compares it to state-of-the-art.

TABLE III. STATE-OF-THE-ART OF MLC IMPLEMENTATIONS

	RRAM Device	# states	MLC mode	Design
[4]	$Pt/TaOx/Ta_2O_5/Pt$	4 HRS	V_{RST} & Pulse	Device
[7]	$TiN/Ti/HfO_2/TiN$	3 LRS/1 HRS	I_C SET	Device
[8]	$TiN/HfO_x/Pt$	8 HRS	V_{RST} & Pulse	Device
[9]	$Cu/HfO_2/Cu/Pt$	3 LRS/1 HRS	I_C SET	Device
[12]	$Ti/HfO_x/Ti/TiN$	3 LRS/1 HRS	I_C SET	Circuit
[22]	Pt/W/ TaOx / Pt	7 HRS/1 LRS	V_{RST}	Device
Work	***TiN/Ti/HfO$_x$/TiN***	***8 HRS***	***I_C RST***	***Circuit***

Comparison metrics include the targeted RRAM technology, the number of resistance states, the MLC operation mode and the design level (i.e., device or circuit level). Storing 8 states has only been reported in [8][22] at the device level, mainly by varying RST voltages (V_{RST}) and programming pulses. Note that our methodology is the first one leveraging on compliance current (I_C) control in the RST direction.

VI. Conclusion

MLC RRAM research is still in an early stage and most studies are focused on the device level. In this context, a MLC operation design scheme based on RST current control is proposed at the circuit level to achieve robust MLC operation without the need of read-verify operations. The proposed MLC implementation can target different HRS resistance ranges to optimize energy and latency in addition to allow remarkable resistance margins.

References

[1] Y. Shimeng, P.-Y. Chen, "Emerging memory technologies: Recent trends and prospects," IEEE SSC Magazine, vol. 8, no. 2, 2016.

[2] D. Ielmini, "Resistive switching memories based on metal oxides: mechanisms, reliability and scaling," Semiconductor Science and Technology, 2016.

[3] C. Xu et al., "Understanding the trade-offs in multi-level cell ReRAM memory design," IEEE DAC conference, 2013.

[4] S. R. Lee et al., "Multi-level switching of triple-layered TaOx RRAM with excellent reliability for storage class memory," IEEE VLSIT conference, 2012.

[5] M.H. Lee et al., "Studies on ReRAM Conduction Mechanism and the Varying-bias Read Scheme for MLC and Wide Temperature Range TMO ReRAM," IEEE ICSICT conference, 2018.

[6] X. Zhu et al., "Observation of conductance quantization in oxide-based resistive switching memory," Advanced Materials, 2012.

[7] B. Chakrabarti et al., "Multilevel Switching in Forming-Free Resistive Memory Devices With Atomic Layer Deposited HfTiOx Nanolaminate," EDL, 2013.

[8] L. Zhao et al., "Improved multi-level control of RRAM using pulse-train programming," IEEE VLSI-TSA conference, 2014.

[9] Y. Wang et al., "Investigation of resistive switching in Cu-doped HfO_2 thin film for multilevel non-volatile memory applications," Nanotechnology, 2010.

[10] M. Zangeneh and J. Ajay, "Design and optimization of nonvolatile multibit 1T1R resistive RAM," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2013.

[11] J. Xing, et al., "Practical considerations of read-out circuits for passive, multi-level ReRAM arrays," IEEE 3M-NANO, 2016.

[12] J. Reuben and F. Dietmar, "A Time-based Sensing Scheme for Multi-level Cell (MLC) Resistive RAM," IEEE NORCAS, 2019.

[13] A. Grossi et al., "Fundamental variability limits of filament-based RRAM," IEDM, 2016.

[14] H. Basma et al., "RRAM device models: a comparative analysis with experimental validation," IEEE Access 7, 2019.

[15] L. Zhao et al., "Multi-level control of conductive nano-filament evolution in HfO_2 ReRAM by pulse-train operations", Nanoscale, 2014.

[16] H. Banba et al., "A CMOS bandgap reference circuit with sub-1-V operation", IEEE Journal of Solid-State Circuits, 1999.

[17] H. Aziza et al., "ReRAM ON/OFF resistance ratio degradation due to line resistance combined with device variability in 28nm FDSOI technology," IEEE EUROSOI-ULIS conference, 2017.

[18] Y. Joly et al., "Impact of hump effect on MOSFET mismatch in the sub-threshold area for low power analog applications," IEEE ICSICT conference, 2010.

[19] H. Aziza et al., "Evaluation of OxRAM cell variability impact on memory performances through electrical simulations", IEEE NVMTS conference, 2011.

[20] H. Aziza et al., "An Energy-Efficient Current-Controlled Write and Read Scheme for Resistive RAMs (RRAMs)," IEEE Access, 2020.

[21] A. Levisse et al., "Write Termination Circuits for RRAM: A Holistic Approach From Technology to Application Considerations," IEEE Access, 2020.

[22] W. Kim et al., "3-Bit multilevel switching by deep reset phenomenon in Pt/W/TaOx /Pt-ReRAM devices," EDL, 2016.