

Testing and Fault Diagnosis for Multi-level Resistive Random-Access Memory in Monolithic 3D Integration*

Shao-Chun Hung[†], Partho Bhounik^{‡§}, and Krishnendu Chakrabarty^{‡§}

[†]Department of Electrical and Computer Engineering, Duke University

[‡]School of Electrical, Computer and Energy Engineering, Arizona State University

[§]ASU Center for Semiconductor Microelectronics (ACME), Arizona State University

Abstract—The integration of resistive random-access memory (RRAM) with emerging monolithic 3D (M3D) integration technology offers the promise of lower cost per bit and higher memory density than current technologies. Multiple bits per RRAM cell, i.e., multi-level cell (MLC), further enhances memory density and reduces power consumption. However, MLCs are susceptible to errors due to power supply noise (PSN) within the power delivery network. In addition, monolithic inter-tier vias (MIVs) in M3D-integrated MLC designs are prone to manufacturing defects as they penetrate through a device tier between memory cells and access memories. These defects impact the voltage level at the inputs of an MLC, resulting in incorrect read/write operations. Therefore, testing and fault diagnosis for M3D-integrated MLC RRAM is important to detect defective cells and facilitate yield learning. We present a detailed analysis of RRAM faulty behaviors in the presence of PSN and MIV defects. We also introduce a March sequence to detect faults due to PSN noise and quantify the magnitude of noise and defects. Experimental results for MLC RRAM cells demonstrate the effectiveness of the proposed test and diagnosis solution.

I. INTRODUCTION

As Moore's Law approaches its physical limits, monolithic 3D (M3D) has emerged as a promising alternative to continue power, performance, and area improvements. One of the major applications of M3D lies in the adoption of compute-in-memory architectures featuring non-volatile memory, multiple device layers, and inter-layer vias. Among emerging non-volatile memory technologies, resistive random-access memory (RRAM) has been demonstrated to be compatible with M3D to significantly reduce the cost per bit and increase the memory density [1]. Furthermore, multi-level cells (MLCs) can be implemented with RRAM technology by employing conductance relaxation, which enables the storage of multiple bits in a single memory device [2].

However, the advantages of MLC RRAM are accompanied by new challenges. During the manufacturing process, RRAM devices are exposed to process variations [3]. These variations are manifested as differences in oxide thickness and doping concentrations among RRAM devices within an array,

leading to deviation from intended behaviors. Power supply noise (PSN) is another critical issue for RRAM devices. In conventional planar RRAM devices, functional operations can experience up to 10% PSN-induced voltage droop [4] [5]. 3D-integrated devices are more susceptible to PSN than 2D counterparts due to the long detour from power sources to local receivers and the increased current density [6]–[8]. The PSN-induced voltage droop increases the complexity of reliably achieving multiple conductive levels in an MLC RRAM.

Moreover, interconnects between device tiers in an M3D design, i.e., monolithic inter-tier vias (MIVs), can suffer from high defect rates as they need to traverse through inter-tier dielectrics. Surface roughness within the dielectrics can create voids in MIVs during fabrication, leading to open defects. A diagnosis strategy for distinguishing MIV defects from process variations in an M3D-integrated RRAM has been proposed in [9]; however, the impacts of MIV defects on MLC RRAM devices remain unexplored. To fully exploit the potential of M3D-integrated MLC RRAM architectures, there is a need for a testing scheme that can detect and diagnose issues related to both PSN and fabrication processes.

In this paper, we propose a new testing framework for M3D-integrated MLC RRAM. Our method is able to detect all faults arising from manufacturing defects and PSN. Moreover, root-cause fault origins can be identified based on the output signatures obtained from the proposed algorithm. The main contributions of this paper are as follows:

- We provide a comprehensive characterization of MLC RRAM under the impacts of PSN-induced voltage droop and MIV defects.
- We propose a March testing algorithm that ensures the detection of all faults induced by PSN and fabrication defects.
- We demonstrate how the outputs generated by our solution can be harnessed to pinpoint the origins of faults and assess the size of MIV defects.

The rest of the paper is organized as follows. Section II provides background regarding M3D integration and MLC RRAM switching mechanisms. Section III presents characterizations and fault modeling of M3D-integrated MLC RRAM in the presence of PSN and MIV defects. In Section IV, the proposed March testing algorithm is provided and a discussion

*This research was supported in part by the National Science Foundation under grant CCF-2309822 and in part by CHIMES, one of the seven centers in JUMP 2.0, a Semiconductor Research Corporation (SRC) program sponsored by DARPA.

is presented on how to utilize output signatures for diagnostic purposes. Finally, Section V concludes the paper.

II. BACKGROUND

A. M3D Integration

M3D integration is a promising technology that provides substantial improvements in computing power and energy efficiency compared to conventional 2D devices. In an M3D device, all tiers are fabricated sequentially on the same wafer; this is achieved through recent breakthroughs in low-temperature manufacturing processes [10]. These processes are necessary to ensure the fabrication of upper tiers without causing damage to devices and wires in lower tiers. M3D implementation relies on fine-grained MIVs to establish connections among the components within each device tier, which allows the incorporation of multiple tiers within a single chip [11]. Compared to through-silicon vias (TSVs) in today's 3D integration technologies, MIVs have a much smaller physical footprint with negligible induced capacitance. These advantages facilitate the close packing of device tiers [12], leading to a significant reduction in total wirelength and a subsequent decrease in power consumption compared to TSV-based 3D integration.

M3D integration has been explored in memory-on-logic architectures to achieve improvements in energy efficiency and memory density [1] [2]. Among emerging non-volatile memory technologies, RRAM has been shown to be ideal with M3D integration due to its simple structure and compatibility with atomic-layer deposition techniques [13]. However, M3D-integrated RRAM suffers from high defect rates due to immature manufacturing processes. Furthermore, an M3D architecture experiences severe PSN-induced voltage droop due to the need for current to traverse upper-tier power delivery networks (PDNs) to supply lower-tier devices. The increased conduction path and high current demand lead to excessive IR-drop [6]. These unique defects in M3D integration are ignored by prior work as they do not exist in conventional planar devices. This motivates us to develop a new test and diagnosis strategy for MLC RRAM with M3D integration. Such a framework is key to identifying defective RRAM devices and facilitating yield learning.

B. MLC RRAM

In an RRAM device, a memristor serves as the primary storage component. The fundamental structure of a memristor consists of a metal-insulator-metal (MIM) stack, where a thin dielectric layer is situated between the top electrode (TE) and the bottom electrode (BE). To form a functional RRAM device, the memristor is usually integrated with an access transistor to create a one-transistor-one-resistor (1T1R) structure within a crossbar array, as shown in Fig. 1. When the RRAM is selected, the corresponding access transistor becomes active, allowing current to flow through the memristor for write and read operations. Conversely, when the RRAM element is not selected, the access transistor is deactivated to prevent sneak-path current and reduce power consumption.

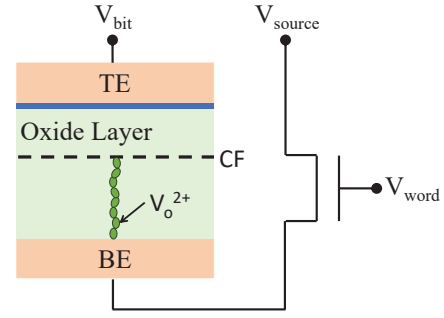


Fig. 1: Schematic of a 1T1R RRAM device, where V_o^{2+} is the positive-charged oxygen vacancy, CF is the conduction filament, and V_{bit} , V_{word} , and V_{source} are the applied voltages at the bit line, the word line, and the source line, respectively. This figure is adapted from [9] and redrawn.

After fabrication, a memristor exhibits an exceptionally high equivalent resistance value. To make the RRAM cell ready for normal operations, a forming process is utilized. In this process, a positive voltage is applied across the TE and BE. This voltage knocks out oxygen atoms from the lattice, leading to the accumulation of oxygen vacancies within the dielectric [14]. Such an accumulation creates the formation of a conduction filament (CF), which facilitates the flow of current through the insulating layer. Therefore, the RRAM cell transitions into a low-resistance state (LRS). This process is referred to as the “SET” operation. Conversely, when a negative voltage is applied across the RRAM, the oxygen ions recombine with the oxygen vacancies. This recombination process ruptures the CF; therefore, less current can flow through the device, causing the RRAM to enter a high-resistance state (HRS), denoted as the “RESET” operation. For reading the data stored in an RRAM, a small pulse, which is insufficient to alter the resistance state within the cell, is applied to the device.

The storage of multiple bits within a single RRAM device to create an MLC RRAM is made possible by precisely controlling the position of the CF. Strategies for achieving multi-level switching are outlined as follows [14]:

- **Controlling compliance current:** For a 1T1R RRAM device, the manipulation of multiple resistance states can be accomplished by regulating the current compliance during write operations. This control involves adjusting the voltage applied to the gate contact of the access transistor.
- **Controlling write voltage:** During write operations, the speed at which the CF moves is directly proportional to the voltage applied across the device. Therefore, within a fixed pulse width, intermediate resistance states can be obtained by modifying the write voltage applied to the TE and the BE.
- **Controlling pulse width:** In an RRAM device, CF traverses the insulator to complete the state transition. By precisely controlling the pulse width, the CF can be retained in the desired location after the write duration, leading to intermediate resistance values.

TABLE I: Parameters used in SPICE simulation for multi-level switching with the memristor model in [15].

Parameter	Value	Description
t_{ox}	12 nm	Oxide thickness
F_{min}	1.4×10^9 V/m	Minimum electric field to switch state
R_{th}	2.1×10^3 K/W	Thermal resistance
E_a	0.6 eV	Average activation energy for oxygen vacancy generation
a_0	0.25 nm	Atom spacing
T_{write}	225 ns	Pulse width for write operations
T_{read}	50 ns	Pulse width for read operations
R_{wire}	12.78 Ω	Wire resistance [16]
C_{wire}	0.046 fF	Wire capacitance [16]
V_{SET}	3.6 V	Voltage for SET operation
V_{RESET}	-3.6 V	Voltage for RESET operation
V_{Read}	0.8 V	Voltage for read operation
V_g^{w0}	2.7 V	Gate voltage for w0 operation
V_g^{w1}	2.2 V	Gate voltage for w1 operation
V_g^{w2}	2.0 V	Gate voltage for w2 operation
V_g^{w3}	3.0 V	Gate voltage for w3 operation

III. FAULT MODELING OF M3D-INTEGRATED MLC RRAM

In this paper, we modulate compliance current by adjusting gate voltages to achieve multi-level switching. We carry out SPICE simulations using the experimentally calibrated memristor model from [15] with 3x3 arrays in both tiers of an M3D-integrated RRAM. We use the partitioning method for [2], where access transistors are divided into two tiers and memristors are fabricated in the top-tier back-end-of-line. Without loss of generality, we discretize each memristor into four logic levels, with each logic level representing two bits of data: 00, 01, 10, and 11. To simplify the notation, we employ a compact representation where logic levels are denoted by indices from 0 to 3. We utilize a scheme that evenly distributes the memristor state based on the location of the CF to achieve uniform state assignment in MLC RRAM. This approach helps in maximizing the margin between adjacent logic levels [4]. Let L be the distance between TE and BE and $CF(t)$ be the depth of CF at time t . In our scheme, the reading of logic-0 (L0) occurs when $1 \geq \frac{CF(t)}{L} > 0.75$; logic-1 (L1) is read when $0.75 \geq \frac{CF(t)}{L} > 0.5$; logic-2 (L2) is detected when $0.5 \geq \frac{CF(t)}{L} > 0.25$; logic-3 (L3) is identified when $0.25 \geq \frac{CF(t)}{L} \geq 0$.

Table I presents the parameters employed in our SPICE simulations to achieve multi-level switching, where “w0” corresponds to the write operation for L0, “w1” pertains to the write operation for L1, “w2” is associated with the write operation for L2, and “w3” denotes the write operation for L3. Note that in an MLC RRAM, the operations w0, w1, and w2 start with an initial SET operation to switch the memristor into the LRS, followed by a subsequent RESET operation with the appropriate gate voltage. In contrast, the operation w3 is equivalent to a single SET operation.

Fig. 2 illustrates the resistance values of the RRAM during various write operations. Clearly, after the write time period, the RRAM can enter four distinct logic levels corresponding to their respective write operations. However, due to the division

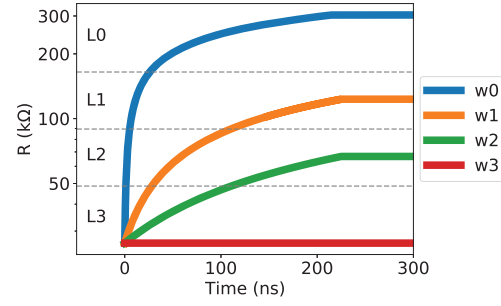


Fig. 2: Multi-level switching for an MLC RRAM, where R is the resistance value of the RRAM device.

of the resistance range into four levels, the separation between adjacent logic states is closer compared to a conventional single-bit device. This makes an MLC RRAM more sensitive to voltage fluctuations and PSN inside the PDN. Moreover, MIV defects in an M3D-integrated RRAM impact the voltage experienced across the memristor, leading to faulty behaviors. These defects are unique to M3D integration and have not been addressed in prior work. In this section, we perform a comprehensive characterization and fault modeling of M3D-integrated RRAMs in the presence of PSN and MIV defects.

A. PSN-induced Voltage Droop

During write operations, PSN leads to voltage droop in the PDN. This phenomenon has a direct impact on the voltage across the memristor and the voltage received at the gate contact of the access transistor. Conventional single-bit RRAM designs typically ensure resilience against PSN-induced voltage droop of up to 10% [4]. However, this level of voltage droop can lead to erroneous behaviors in an MLC RRAM. To simulate and analyze this scenario, we systematically vary the voltages applied to the bit line (i.e., V_{SET} and V_{RESET}) and the gate voltage during various write operations. The variations range from 0% to 10% of nominal values, with increments of 0.1%.

Fig. 3 presents the characterization of an RRAM in the presence of PSN-induced voltage droop following different write operations. Fig. 3(a) displays the resistance value distribution after the w0 operation with concurrent voltage droop in V_{RESET} and V_g^{w0} . In all cases with voltage droop up to 10%, the resistance values remain within the L0 range without causing any faulty behavior. This outcome aligns with expectations because the w0 operation is functionally equivalent to the RESET operation in a single-bit RRAM device, which is designed to withstand such levels of voltage droop. Similarly, the PSN-induced voltage droop does not induce any faulty behavior during w3 operations. In every scenario, the RRAM successfully attains its LRS, as demonstrated in Fig. 3(d).

However, faulty behaviors are manifested during both the w1 and w2 operations, as shown in Fig. 3(b) and Fig. 3(c). To assess the impacts of PSN-induced voltage droop in both the RESET voltage and the gate voltage, we plot the logic levels in the presence of voltage droop in these voltages. Fig. 4(a) and Fig. 4(b) demonstrate the logic levels after the w1 and

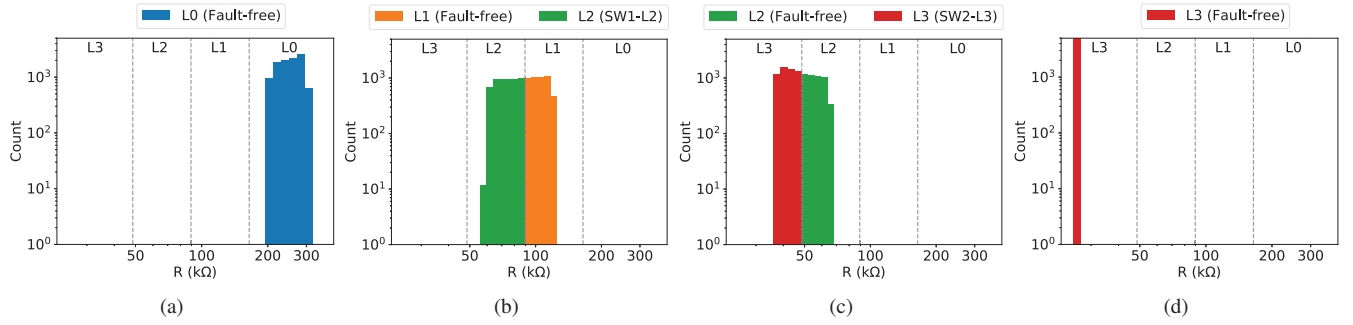


Fig. 3: Resistance value distributions in the presence of PSN-induced voltage droop: (a) resistance distribution after w0; (b) resistance distribution after w1; (c) resistance distribution after w2; (d) resistance distribution after w3.

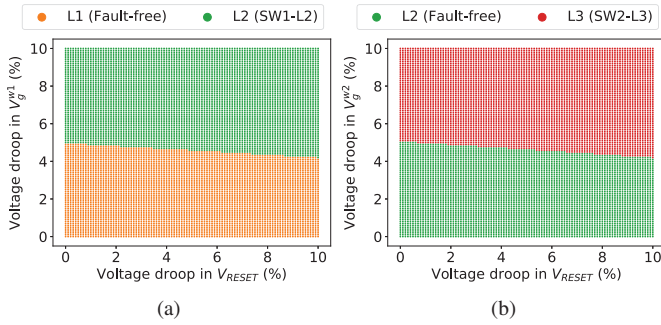


Fig. 4: Logic levels in the presence of PSN-induced voltage droop: (a) logic levels after w1; (b) logic levels after w2.

w2 operations, respectively. When V_g^{w1} and V_g^{w2} experience voltage droop exceeding 5% of nominal values, the compliance current is insufficient to complete the desired state transition within the write time window, even if no voltage droop exists in V_{RESET} . When the voltage droop in gate voltages is near 5%, a reduction in V_{RESET} decreases the voltage difference across the memristor, impeding the movement of the CF. These defects make the RRAM unable to reach the intended state with the write strength supplied by the internal circuitry, leading to slow-write (SW) faults.

The above characterization of RRAMs clearly shows that PSN-induced voltage droop can cause the memristor to transition into an unintended state, yielding undesired output signals during read operations. These faulty behaviors can be modeled by extending the concept of SW faults in single-bit devices in multi-level manners. For the w1 operation, voltage droops in V_g^{w1} and V_{RESET} make the RRAM unable to reach L1 and stay in L2. This situation can be modeled as a SW1-L2 fault. Similarly, a SW2-L3 fault can occur due to PSN-induced voltage droop occurring during the w2 operation. These fault models are of considerable importance in the design and implementation of a dependable MLC RRAM array. Therefore, such faults should be taken into account when designing testing and diagnosis algorithms.

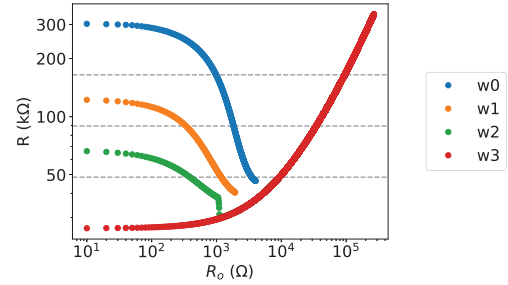


Fig. 5: RRAM characterization in the presence of MIV defects.

B. MIV Defects

MIVs in an M3D design are susceptible to defects because they penetrate through the inter-tier dielectric. The presence of voids in the dielectric can lead to open defects within these MIVs. In M3D-integrated RRAM, an MIV is used to establish a connection between the drain contact of an access transistor and the BE of a memristor. Therefore, an MIV open defect is functionally equivalent to the introduction of an additional resistance, denoted as R_o , in series with the memristor. This additional resistance can impact the voltage across the memristor during write operations due to charge sharing.

To characterize RRAM behaviors in the presence of MIV defects, we carry out SPICE simulations with different write operations and vary R_o with step increments of 10Ω . Note that in these simulations, we maintain nominal values for V_{SET} , V_{RESET} , and gate voltages. This is because if R_o with nominal voltages leads to a fault, R_o under conditions where PSN-induced voltage droop occurs would exacerbate the issue and also result in faulty behaviors. Therefore, simulating with nominal values is sufficient for analyzing the effects of MIV defects.

Fig. 5 demonstrates the RRAM characterization results with MIV defects after different write operations. Note that the x-axis is scaled logarithmically; therefore, discrete points appear to be continuous when R_o is spread over several orders of magnitude. In our simulations, we apply all write operations with memristors initially set to the LRS, which is the default state following the forming process. For the w0, w1, and w2

TABLE II: Faulty behaviors in the presence of MIV defects.

Operation	Size of R_o	Fault model
w0	$1830 \Omega \geq R_o > 980 \Omega$	SW0-L1
	$3510 \Omega \geq R_o > 1830 \Omega$	SW0-L2
	$4010 \Omega \geq R_o > 3510 \Omega$	SW0-L3
	$R_o > 4010 \Omega$	Stuck-at LRS
w1	$1180 \Omega \geq R_o > 330 \Omega$	SW1-L2
	$1950 \Omega \geq R_o > 1180 \Omega$	SW1-L3
	$R_o > 1950 \Omega$	Stuck-at LRS
w2	$1120 \Omega \geq R_o > 380 \Omega$ $R_o > 1120 \Omega$	SW2-L3 Stuck-at LRS
w3	$58040 \Omega \geq R_o > 9360 \Omega$	Stuck-at L2
	$94920 \Omega \geq R_o > 58040 \Omega$	Stuck-at L1
	$R_o > 94920 \Omega$	Stuck-at L0

operations, the introduction of R_o decelerates the movement of the CF, and it cannot reach the intended location within the write time. This scenario becomes severe as the magnitude of R_o increases, causing CF to stay at an unintended logic level and lead to a SW fault. The w0 operation exhibits the highest resilience to the effects of R_o because it benefits from the most robust write strength, provided by the highest gate voltage. However, even for the w0 operation, a SW fault is manifested when R_o is larger than 980Ω . When R_o reaches a sufficiently large magnitude, the internal circuitry's write strength is not able to displace the CF. Therefore, the memristor becomes perpetually stuck in the LRS.

Table II shows the summary of faulty behaviors caused by MIV defects. When R_o reaches a magnitude that traps the CF in the LRS for the w0, w1, and w2 operations, the resistance level of the RRAM becomes dominated by the magnitude of R_o , which can be observed from the outcomes after the w3 operation. This scenario can be explained by the fact that the equivalent resistance of a 1T1R cell is composed of the on-resistance of the access transistor, the MIV open R_o , and the inherent resistance in the memristor. The on-resistance of the access transistor is minimized to have a negligible impact on normal operations, while the memristor remains fixed at the LRS, which is relatively low in resistance compared to R_o . Therefore, the logic level of a defective RRAM predominantly depends on the size of R_o . As the magnitude of R_o increases, the defective RRAM can manifest a range of faulty behaviors, including being stuck at logic levels spanning from L3 to L0.

IV. FAULT DETECTION AND DIAGNOSIS

A. Proposed Algorithm

In this section, we provide the details of the proposed testing algorithm that can be implemented using memory built-in self-test with March testing schemes, a widely recognized approach in memory testing known for its high fault coverage. The March testing schemes apply a systematic sequence of operations to a given memory device before transitioning to the next device. Note that in addition to faulty behaviors discussed in Section III, conventional fault models for planar 1T1R devices remain applicable to the testing of MLC RRAM [17] [18]. These fault models include stuck-at fault, deep fault, SW fault, fast-write (FW) fault, and coupling fault. We consider all of these fault models and faulty behaviors in the formulation

of the proposed March-type algorithm, which is outlined as follows:

$$\left\{ \begin{array}{l} \uparrow (r3, w3, w0); \uparrow (r0, w0, w3) \\ \downarrow (r3, w0); \downarrow (r0, w3); \Downarrow (w1, r1, w3, w2, r2) \end{array} \right\} \quad (1)$$

where $r0, r1, r2$, and $r3$ denote the read operations, with each corresponding to the expected logic level L0, L1, L2, and L3, respectively. The symbols \uparrow (\downarrow) indicate the increasing (decreasing) address order, while \Downarrow denotes that the address order may follow either an increasing or a decreasing order.

B. Fault Detection with the Proposed Algorithm

The proposed algorithm consists of five March elements, each enclosed within parentheses. We assign a label " M_a " to each March element a and describe how these elements detect each fault type as follows:

- **Stuck-at faults:** Stuck-at logic- i faults are detectable by performing a w_j operation and a subsequent r_j operation, where $i \neq j$. These faults can be sensitized by write operations in M_1, M_2, M_3, M_5 , and they can be detected by read operations in M_2, M_3, M_4, M_5 .
- **SW faults and FW faults:** SW faults and FW faults with the intended logic level i are detectable by performing a write operation w_i and conducting a subsequent read operation r_i . SW and FW faults can be sensitized by the write operations in M_1, M_2, M_3, M_5 and be detected by read operations in M_2, M_3, M_4, M_5 . Note that FW faults are not manifested during w0 and w3 operations. This is because the resistance states remain unaltered in the intended logic level when they reach the HRS and the LRS during w0 and w3, respectively, even if these states are reached early during the write period.
- **Deep faults:** Deep faults can occur at logic levels characterized by the highest and lowest resistance values, i.e., L0 and L3, respectively. Deep faults are induced by variations in the length and cross-sectional area of the memristor, making it unable to transition within the write time [17]. In the proposed algorithm, Deep-L0 faults are sensitized by two consecutive w0 operations in M_1 and M_2 , followed by the w3 operation in M_2 ; the read operation $r3$ in M_3 serves to detect Deep-L0 faults. For Deep-L3 faults, as the RRAM is initially in the LRS state following the forming process, operations $\{w3, w0\}$ in M_1 are adequate for sensitizing these faults; subsequently, read operation $r0$ in M_2 can detect Deep-L3 faults.
- **Coupling faults:** Coupling faults are manifested when an undesired state change occurs in a victim cell due to a write operation applied to another aggressor cell. We represent coupling faults as $CpF(x, y)\uparrow$ ($CpF(x, y)\downarrow$), where x denotes the write operation w_x executed on the aggressor cell, and y is the initial state of the victim cell. The symbols \uparrow (\downarrow) indicate whether the address of the victim cell is higher (lower) than the address of the aggressor cell. In the proposed algorithm, $CpF(0,3)\uparrow$ is sensitized by $\{w3, w0\}$ in M_1 and detected by $r3$ in M_1 ; $CpF(3,0)\uparrow$ is sensitized by $\{w0, w3\}$ in M_2 and detected by $r0$ in M_2 ; $CpF(0,3)\downarrow$

TABLE III: Fault-detection dictionary of the proposed March testing algorithm.

Fault types/origins	M ₁	M ₂	M ₃	M ₄	M ₅	
	r3	r0	r3	r0	r1	r2
Stuck-at L0	0	0	1	0	1	1
Stuck-at L1	0	1	1	1	0	1
Stuck-at L2	0	1	1	1	1	0
Stuck-at L3	0	1	0	1	1	1
SW0	0	1	0	1	0	0
SW1 & FW1	0	0	0	0	1	0
SW2 & FW2	0	0	0	0	0	1
SW3	0	0	1	0	0	0
Deep-L0	0	0	1	0	0	0
Deep-L3	0	1	0	0	0	0
CpF(0,3)↑	1	0	0	0	0	0
CpF(3,0)↑	0	1	0	0	0	0
CpF(0,3)↓	0	0	1	0	0	0
CpF(3,0)↓	0	0	0	1	0	0
PSN-induced voltage droop	0	0	0	0	1	1
MIV defects	1	1	1	1	1	1

is sensitized by w3 and w0 in M₂ and M₃, respectively, and detected by r3 in M₃; CpF(3,0)↓ is sensitized by w0 and w3 in M₃ and M₄, respectively, and detected by r0 in M₄. Note that the detection of these four CpFs is adequate to identify all coupling faults, as the aggressor is written to a logic level farthest from the state of the victim. These CpFs cover all scenarios where the aggressor and victim cells are in intermediate states.

- **PSN-induced voltage droop:** The PSN-induced voltage droop can cause a defective RRAM to stay in a neighboring logic level after w1 or w2 operations. The tests for SW faults cover these types of faulty behaviors.
- **MIV defects:** As demonstrated in Table II, MIV defects can lead to either SW faults or stuck-at faults depending on the magnitude of R_o . Such fault types are covered by the proposed March algorithm.

The details of fault detection with the proposed algorithm are presented in Table III, where “1” and “0” denote whether the respective fault types can or cannot be identified through the associated read operation.

C. Fault Diagnosis

In addition to fault detection, the output signatures generated by the proposed March algorithm can be used for diagnosis based on the RRAM characteristics elaborated in Section III. For PSN-induced voltage droop, concurrent variations in V_{SET} , V_{RESET} , and gate voltages lead to SW faults that are manifested during w1 and w2 operations, with no faults emerging after writing the RRAM to L0 and L3, as shown in Fig. 3. Furthermore, RRAM cells affected by PSN-induced voltage droop typically remain at a logic level one step lower than the intended logic level following w1 and w2 operations.

Leveraging these behavioral patterns, we can pinpoint the root cause using the output signatures based on Equation (1). Let R_1 - R_6 denote the logic levels detected during the six read operations in the proposed algorithm. When $(R_1, R_2, R_3, R_4, R_5, R_6) = (L3, L0, L3, L0, L2, L3)$, the defective cell is susceptible to severe PSN-induced voltage droop. Therefore, when a large number of RRAM cells within

TABLE IV: Output signatures in the presence of MIV defects.

Size of R_o	$(R_1, R_2, R_3, R_4, R_5, R_6)$
$380 \Omega \geq R_o > 330 \Omega$	(L3, L0, L3, L0, L2 , L2)
$980 \Omega \geq R_o > 380 \Omega$	(L3, L0, L3, L0, L2 , L3)
$1180 \Omega \geq R_o > 980 \Omega$	(L3, L1 , L3, L1 , L2 , L3)
$1830 \Omega \geq R_o > 1180 \Omega$	(L3, L1 , L3, L1 , L3 , L3)
$3510 \Omega \geq R_o > 1830 \Omega$	(L3, L2 , L3, L2 , L3 , L3)
$9360 \Omega \geq R_o > 3510 \Omega$	(L3, L3 , L3, L3 , L3 , L3)
$58040 \Omega \geq R_o > 9360 \Omega$	(L2 , L2 , L2 , L2 , L2 , L2)
$94920 \Omega \geq R_o > 58040 \Omega$	(L1 , L1 , L1 , L1 , L1 , L1)
$R_o > 94920 \Omega$	(L0 , L0, L0 , L0, L0 , L0)

an M3D-integrated array exhibits such a set of output signatures, it indicates that the PDN design is not sufficiently robust to support MLC switching. This insight is invaluable for providing early feedback to the foundry to aid in optimizing both the PDN design and the manufacturing processes.

For RRAM devices with MIV defects, different sizes of R_o can lead to varied behaviors during each write operation. Table IV lists the output signatures when MIV defects are present, where signatures highlighted in red indicate that faults are successfully detected by the corresponding read operation. Note that the range of $R_o \leq 330 \Omega$ is not listed as no fault occurs within this range.

From Table IV, it is clear that each range of R_o yields its unique output signature. Note that RRAM cells experiencing an MIV defect in the range of $980 \Omega \geq R_o > 380 \Omega$ exhibit the same output signature as those affected by PSN-induced voltage droop. However, a key distinction lies in the fact that PSN has no impact on the RRAM after the w3 operation, while an MIV defect introduces additional resistance to the RRAM. Such additional resistance influences the resistance value being read out during read operations.

Let R_{ref} be the equivalent resistance of a defective RRAM with $R_o = 380 \Omega$. By performing an additional w3 operation, followed by a read operation, the resistance value of the RRAM, denoted as R , can be compared to R_{ref} . If $R > R_{ref}$, an MIV defect is presented. Otherwise, defective RRAM suffers from PSN-induced voltage droop. With this approach and the distinctive signatures for other ranges of R_o , the outputs of the proposed March algorithm can effectively pinpoint the root cause and the size of the defects when MIV defects are present.

V. CONCLUSION

We have presented a comprehensive characterization of an M3D-integrated MLC RRAM in the presence of PSN-induced voltage droop and MIV defects. We have demonstrated that the PDN designed to be robust for single-bit RRAM devices can lead to unintended state transitions during MLC switching. We have also shown that different sizes of MIV defects can cause various types of faulty behaviors during write operations. We have proposed a March testing algorithm to detect every fault type and faulty behavior. We have demonstrated that by leveraging the output signatures obtained from the proposed algorithm, the root cause can be successfully pinpointed and the size of defects can be accurately determined.

REFERENCES

- [1] C. Xu, D. Niu, Y. Zheng, S. Yu, and Y. Xie. Impact of cell failure on reliable cross-point resistive memory design. *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, 20(4):1–21, 2015.
- [2] E. Esmanhotto, L. Brunet, N. Castellani, D. Bonnet, T. Dalgaty, L. Grenouillet, D. R. B. Ly, C. Cagli, C. Vizios, N. Allouti, F. Laulagnet, O. Gully, N. Bernard-Henriques, M. Bocquet, G. Molas, P. Vivet, D. Querlioz, JM. Portal, S. Mitra, F. Andrieu, C. Fenouillet-Beranger, E. Nowak, and E. Vianello. High-density 3D monolithically integrated multiple 1T1R multi-level-cell for neural networks. In *IEEE International Electron Devices Meeting (IEDM)*, pages 36–5, 2020.
- [3] D. Niu, Y. Chen, C. Xu, and Y. Xie. Impact of process variations on emerging memristor. In *Proceedings of the Design Automation Conference*, pages 877–882, 2010.
- [4] M. Zangeneh and A. Joshi. Design and optimization of nonvolatile multibit 1T1R resistive ram. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 22(8):1815–1828, 2014.
- [5] M. N. I. Khan and S. Ghosh. Test of supply noise for emerging non-volatile memory. In *IEEE International Test Conference (ITC)*, pages 1–10, 2018.
- [6] S. K. Samal, K. Samadi, P. Kamal, Y. Du, and S. K. Lim. Full chip impact study of power delivery network designs in monolithic 3D ICs. In *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 565–572, 2014.
- [7] A. Todri-Sanial, S. Kundu, P. Girard, A. Bosio, L. Dilillo, and A. Virazel. Globally constrained locally optimized 3-D power delivery networks. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 22(10):2131–2144, 2014.
- [8] S.-C. Hung, A. Chaudhuri, and K. Chakrabarty. Test-point insertion for power-safe testing of monolithic 3D ICs using reinforcement learning*. In *IEEE European Test Symposium (ETS)*, pages 1–6, 2023.
- [9] S.-C. Hung, A. Chaudhuri, S. Banerjee, and K. Chakrabarty. Fault diagnosis for resistive random-access memory and monolithic inter-tier vias in monolithic 3D integration. In *IEEE International Test Conference (ITC)*, pages 118–127, 2022.
- [10] L. Brunet, C. Fenouillet-Beranger, P. Batude, S. Beaurepaire, F. Ponthenier, N. Rambal, V. Mazzocchi, J-B. Pin, P. Acosta-Alba, S. Kerdiles, P. Besson, H. Fontaine, T. Lardin, F. Fournel, V. Larrey, F. Mazen, V. Balan, C. Morales, C. Guerin, V. Jousseau, X. Federspiel, D. Ney, X. Garros, A. Roman, D. Scevola, P. Perreau, F. Kouemini-Tchouake, L. Arnaud, C. Scibetta, S. Chevalliez, F. Aussenac, J. Aubin, S. Reboh, F. Andrieu, S. Maitrejean, and M. Vinet. Breakthroughs in 3D sequential technology. In *IEEE International Electron Devices Meeting (IEDM)*, pages 7.2.1–7.2.4, 2018.
- [11] S. Panth, S. Samal, Y. S. Yu, and S. K. Lim. Design challenges and solutions for ultra-high-density monolithic 3D ICs. In *SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*, pages 1–2, 2014.
- [12] S. Datta, S. Dutta, B. Grisafe, J. Smith, S. Srinivasa, and H. Ye. Back-end-of-line compatible transistors for monolithic 3-D integration. *IEEE Micro*, 39(6):8–15, 2019.
- [13] C. Xu, D. Niu, Y. Zheng, S. Yu, and Y. Xie. Impact of cell failure on reliable cross-point resistive memory design. *ACM Trans. Des. Autom. Electron. Syst.*, 20(4), 2015.
- [14] F. Zahoor, T. Z. Azni Zulkifli, and F. A. Khanday. Resistive random access memory (RRAM): an overview of materials, switching mechanism, performance, multilevel cell (mlc) storage, modeling, and applications. *Nanoscale Research Letters*, 15(1):90, 2020.
- [15] Z. Jiang, S. Yu, Y. Wu, J. H. Engel, X. Guan, and H.-S. P. Wong. Verilog-A compact model for oxide-based resistive random access memory (RRAM). In *International Conference on Simulation of Semiconductor Processes and Devices*, pages 41–44, 2014.
- [16] IEEE Electronics Packaging Society. Heterogeneous integration roadmap. <https://eps.ieee.org/technology/heterogeneous-integration-roadmap.html>, 2023.
- [17] S. Kannan, J. Rajendran, R. Karri, and O. Sinanoglu. Sneak-path testing of crossbar-based nonvolatile random access memories. *IEEE Transactions on Nanotechnology*, 12(3):413–426, 2013.
- [18] Y.-X. Chen and J.-F. Li. Fault modeling and testing of 1T1R memristor memories. In *IEEE VLSI Test Symposium (VTS)*, pages 1–6, 2015.