Intra-cell Resistive-Open Defect Analysis on a Foundry 8T SRAM-based IMC Architecture

Lila Ammoura, Marie-Lise Flottes, Patrick Girard, Jean-Philippe Noël, Arnaud Virazel

Intra-cell Resistive-Open Defect Analysis on a Foundry 8T SRAM-based IMC Architecture

L. Ammoura [1] M.-L. Flottes [1] P. Girard [1] J.-P. Noel [2] A. Virazel [1]

[1] LIRMM – Univ. of Montpellier / CNRS
F-34392 Montpellier
<name>@lirmm.fr

[2] Univ. Grenoble Alpes, CEA, LIST
F-38000 Grenoble
jean-philippe.noel@cea.fr

***Abstract*— The adoption of In-Memory Computing (IMC) architectures is one of the promising approaches to efficiently solve the Von Neumann bottleneck problem. In addition to arithmetic operations, IMC architectures aim at integrating additional logic operations directly in the memory array or/and at the periphery for saving time and power consumption. In this paper, a comprehensive model of a 128x128 bitcell array based on a 28nm FD-SOI process technology has been considered to analyze the behavior of IMC 8T SRAM bitcells in the presence of resistive-open defects injected in the read port. A hierarchical analysis including a detailed study of each defect was performed in order to determine their impact both in memory and computing modes, both locally on the defective bitcell and globally on the array. Experimental results show that the IMC mode offers the most effective detectability of resistive-open defects.**

Keywords*— *In-Memory Computing, 8T SRAM cell, resistive-open defect, Test.

I. Introduction

The Von Neumann paradigm is the basis for most computer systems dedicated to data-intensive applications, which severely reduces performance acceleration, increases power consumption and limits system scalability [1]. These limitations are mainly due to the frequent and large data transfers between the main memory and the processor, which is known as the Von Neumann bottleneck [2]. Therefore, data storage and processing are the most critical challenges of the new Big Data paradigm. In high performance computing system design, a new data-centric approach could be adopted, instead of the conventional model that is primarily computationally focused [3]. This approach involves minimizing data transfer by performing processing closer to where the data is stored in the system memory. Two memory–centric alternative architectures have been considered so far for better performance and lower power consumption. One alternative is called "Near-memory" computing, which aims at processing data close to where it is located. Another promising approach to efficiently solve the problem of data-intensive applications is the adoption of IMC (In-Memory Computing) architectures, which consist in introducing processing operations directly into the memory array. IMC architectures go beyond classical operations and aim at integrating additional logic into the memory array in order to provide close computing abilities and efficiently overcome the Von Neumann bottleneck problem [4].

Defect analysis and, potentially, dedicated test solutions is however mandatory in order to allow a large deployment of these new computing paradigms and related architectures. Two solutions have been proposed in [5-6] for testing correct operations of an 8T SRAM-based IMC architectures in computing mode. These solutions mainly consist in adding computing operations to the original March test algorithm. As shown in [7], these tests however do not cover all potential defects in the targeted IMC architecture. In particular, defects in the memory read port are not covered.

This paper presents a study on resistive-open defects located in the read port of an 8T SRAM bitcell designed using a 28nm FD-SOI process technology. The study has been carried out in order to determine the impact of each of these intra-cell resistive-open defects in both memory (Read/Write operations) and computing modes, locally, on the defective bitcell as well as, globally, on other cells of the array. Impact on sensitization test sequences is also discussed. Experiments show that the operation in IMC mode improves the detectability of resistive-open defects and allows the detection of smaller size defects when considering all the cells of the same column for the computing operation.

The remainder of this paper is organized as follow. Section II presents the memory model considered in this study. In Section III, the framework and the defect injection approach are first described. Then, a qualitative analysis of the considered defects is presented in Section V. Experimental results are reported in Section IV. Finally, Section V concludes the paper and gives future perspectives.

II. Considered IMC SRAM Array

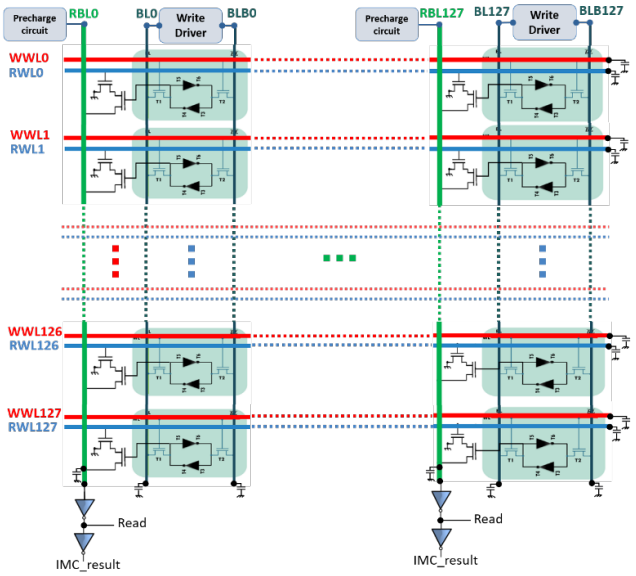

Figure 1. Considered 128x128 matrix model with layout extraction of parasitic capacitances

To characterize the electrical behavior under realistic conditions, our study was conducted on the model presented

in Fig. 1, which is a 128x128 bitcell array designed using a 28 nm FD-SOI process technology [8]. The model is made of write drivers ensuring the writing operations in the bitcells and of precharge circuits, which maintain the *RBL* signals at Vdd necessary to perform the read operations and the computation at array level. The model also includes parasitic capacitances (obtained by a layout extraction) of the main signals (i.e., *BL*/*BLB*, *RBL* and *RWL*) that reinforces the realistic aspect of the model and allows achieving results that closely approximate to those that can be achieved in a real circuit.

III. Resistive-Open Defect Analysis

A. Resistive-open defect injection framework

Performing a computation in memory is ultimately equivalent to performing a Read operation on at least two bitcells of the same column. So, ensuring that the read operation operates correctly is essential for any IMC architectures. 8T SRAM bitcells are the most suitable for SRAM-based IMC because they have a read port isolated from the write port. It ensures that the read operation does not interfere with the data content of the bitcell, even if several *RWLs* are activated simultaneously. This behavior makes these cells useful in the IMC context. Therefore, our goal is to analyze the impact of open defects in the read port of 8T SRAM bitcells. Three resistive-open defects are considered for each of the two transistors as shown in Fig. 2.

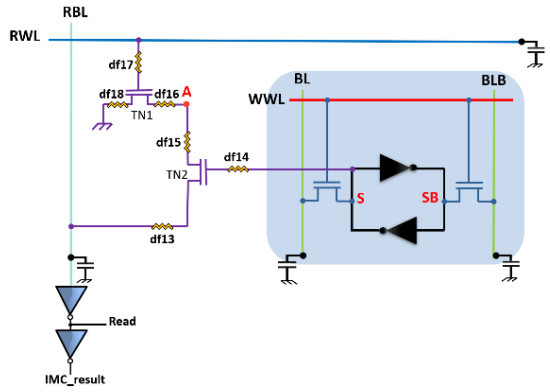

Figure 2. Resistive-open defect injection in the read port of an 8T SRAM cell

To proceed with the injection of resistive-open defects, a monitoring bitcell (*i;j*) (i.e., a bitcell located at row *i* and column *j*) is targeted by a single defect injected at its read port. To analyze each injected defect, we set-up an approach to monitor each time the state of the faulty bitcell (i.e., aggressor cell "c_a"), the states of the neighboring bitcells (i.e., victim cell "c_v" on the same column or row), and the computation results between the faulty bitcell and at least one fault-free bitcell on the same column. The purpose is to reveal the potential impact of each defect on the read/write/computing operations. The defect analysis is hierarchically performed as follows:

- Stand Alone Analysis (SA_Analysis): local impact on the defective bitcell itself during memory mode operations on that bitcell.
- Neighborhood Analysis (N_Analysis): It is done in two steps: i) impact on defect-free surrounding bitcells during memory mode operations on the faulty bitcell, and ii) local impact on the defective bitcell during memory mode operations performed on fault-free surrounding bitcells only.
- Computation Analysis (C_Analysis): It is done in two steps: i) impact on computing mode operations performed between the defective bitcell and at least a fault-free one in the same column i.e., NOR(c_a;c_v), and ii) impact on computing mode operations performed between at least two defect-free bitcells located in the same column than the defective one, i.e., NOR(c_v;c_v).

This hierarchical analysis allows a thorough study of each defect to identify their impact in both memory and computing modes locally on the defective bitcell as well as globally on the array. Moreover, it enables the definition of a Fault Primitive (FP) [9] for each considered defect as follows:

- **<S/F/R>** when a single cell is involved; the cell c_v (victim cell) is used to sensitize a fault where it appears. **S** describes the Sensitizing Operation Sequence (SOS) that sensitizes the fault; **S** $\in$ {0, 1, w0, w1, w↑, w↓, r0, r1}.
- **<Sa,Sv/F/R>** when two cells are involved; **Sa** describes the sensitizing operation or state of the aggressor cell, while **Sv** describes those of the victim cell; Si $\in$ {0, 1, X, w0, w1, w↑, w↓, r0, r1} (i$\in${a, v}), where X is the don't care value X$\in${0, 1}.

In both notations**, F** describes the value or the behavior of the faulty cell; **F** $\in$ {0, 1, ↑, ↓, -} where ↑ (resp. ↓) means the faulty cell undergoes a transition. **R** describes the logic output level of a read operation in case **S** contains read operations. Generally, it takes one of the values {0, 1, -}, where '-' is used when no read operation is required for the SOS.

B. Qualitative resistive-open defects analysis

Each injected resistive-open defect induces a faulty behavior during the memory mode as well as computing mode. Note that we assume the presence of a single defect for each analysis because the occurrence of multiple defects is unlikely. As case study, faulty behaviors produced by the resistive-open defect df13 (see Fig. 2) is described below.

- SA_Analysis: Since the read port is isolated from the write port, the write operation is not disturbed by df13. The presence of this defect produces a delay, which is directly related to the RBL, which may disturb its discharge during a R1 operation. Let us consider the example of a defective bitcell storing a logic '1'. In order to read its contents, the RWL signal is activated. In the case of a proper operation, the RBL discharges through the two transistors TN1 and TN2. In presence of df13, the discharge of the RBL is delayed and, the greater the resistance value of the defect, the greater the delay produced. So, at a certain value of the resistive-open defect the value read will not be captured at the output of the read port.
- N_Analysis: The operations performed on the defective cell do not affect the functioning of the neighboring cells (same row and same column). All read/write operations on the surrounding defect-free bitcells are performed correctly and vice versa.

- C_Analysis: The operation in computing mode between defect-free bitcells of the same column is not affected by df13. However, the computation performed with bitcells including the defective one is affected by the delay produced by df13. Let us consider the example of a defective bitcell storing a logic '1' and another defect-free bitcell in the same column storing a logic '0' implying a NOR (1;0) operation. As seen previously, the presence of df13 produces a delay which slows down the discharge of the RBL so that the data is not captured at the output of the read port. For high resistance values of df13, the RBL does not discharge. So, the IMC_result output of the read port provides a logic '1' instead of a logic '0'.

All the injected resistive-open defects have been analyzed in the same way as detailed for df13. Table 1 summarizes the results of the analysis and reports at each analysis step (i.e., SA_Analysis, N_Analysis, C_Analysis) the operations affected ("RX"/"WX" when the read/write operation, respectively, is affected, "-" if no operation is affected).

Table 1. Summary of the qualitative resistive-open defect analysis

Defect	SA_Analysis	N_Analysis		C_Analysis	
	Operation	Same Row	Same Column	IMC NOR(c_a;c_v)	IMC NOR(c_v;c_v)
df13	R1	-	-	NOR(1;0)	-
df14	R1	-	-	NOR(1;0)	-
df15	R1	-	-	NOR(1;0)	-
df16	R1	-	-	NOR(1;0)	-
df17	R1	-	-	NOR(1;0)	-
df18	R1	-	-	NOR(1;0)	-

IV. Experimental Results

A. Experimental Setup

All the electrical simulations of the injected defects have been performed using the XA simulator from Synopsys [10] considering the 128x128 bitcell matrix model designed in 28nm FD-SOI process technology depicted in Fig. 1. The simulations have been carried out by applying sequences of operations deduced from the previous subsection with the aim of sensitizing each defect.

All injected resistive-open defects cause a read delay that may induce faulty behaviors. This produced delay depends on the size of the injected defect resistance and the discharge time of the RBL (i.e., parasitic capacitors). Therefore, it is necessary to define a read time T_{read}, at which, we ensure that the data at the output of the read port is captured in the meantime. Thus, T_{read} is deduced from the maximum time T0, which is required by the RBL to completely discharge at the operating environment of a typical process corner, 1V supply voltage and 125°C temperature. Then, T0 is added to a margin of T0/2 as the necessary delay for the data to pass through the read inverters. Based on simulations performed at the operating environment described above, T0 is measured at about 700ps, so T_{read} is selected at 1ns for the rest of the defect injection campaign. To extract the minimum resistance value of each resistive-open defect, a threshold is defined at the level of the RBL voltage which is 30% of Vdd (i.e., 300mv) at the T_{read} instant.

In the following, the read operation "RX" and the IMC operation "NOR(X;Y)" are checked by the behavior of the RBL signal. The RBL behavior is considered correct if the response is provided during the Tread≈1ns. Beyond this read time, the data is not captured at the output of the read port, so the operations are not executed properly which leads to a faulty behavior.

B. Resistive-open defect simulation results

Table 2 summarizes all the results obtained for all the injected resistive-open defects. It details the two categories where the operations affected by the defects appear (SA_Analysis and C_Analysis). The first category (SA_Analysis) is represented in the second part of Table 2 detailing the minimum size of defects that lead to this faulty behavior (i.e., R_{min}). For the category C_Analysis, it is divided into 3 groups (cf. the third the third part of Table 2), the computation is performed between 2 bitcells (i.e., N=1), then 16 bitcells (i.e., N=15) and between all the cells of the column (i.e., N=127), while specifying each time the minimum value of the resistance of the defect which leads to this faulty behavior. For each injected defect, the sequence of operations allowing its sensitization is determined (last column of Table 2) according to the operation that generates the minimum resistance value, i.e., the critical resistance value "$R_c = \min\{R_{min}\}$" of the defect, in order to cover the largest range of these resistive-open defects.

For example, in the case of df13, the minimum defect size is achieved with the C_Analysis with N=127 (i.e., R_c=26.8kΩ) that corresponds to a NOR(1;0^{127}) computing operation. So, the sequence <1,0^{127} NOR(1;0^{127})/0^{127}/1> (detailed below) will be applied considering all the bitcells of the column where the defective bitcell is located as follows:

<1,0^N NOR(1;0^N)/0^N/1>

where a logic '1' is initially stored in the defective bitcell and logic '0' in the N bitcells of the same column as the defective one. Then, a NOR(1;0^N) operation is performed between all

Table 2. Summary of resistive-open defect simulation results

Defect	SA_Analysis		C_Analysis				<S/F/R>/ <Sa,Sv/F/R>
	Operation	R_{min} Ω	IMC Operation NOR(c_a;c_v)	R_{min} Ω N=1	R_{min} Ω N=15	R_{min} Ω N=127	
df13	R1	≈31k	NOR(1;0^N)	≈31k	≈29.8k	≈ 26.8k	<1,0^N NOR(1;0^N)/0^N/1>
df14	R1	≈16.79M	NOR(1;0^N)	≈16.89M	≈16.59M	≈14.56M	<1,0^N NOR(1;0^N)/0^N/1>
df15	R1	≈23.2k	NOR(1;0^N)	≈23.2k	≈22.9k	≈20k	<1,0^N NOR(1;0^N)/0^N/1>
df16	R1	≈23.2k	NOR(1;0^N)	≈23.2k	≈22.9k	≈20k	<1,0^N NOR(1;0^N)/0^N/1>
df17	R1	≈4.39M	NOR(1;0^N)	≈4.44M	≈4.42M	≈4.35M	<1,0^N NOR(1;0^N)/0^N/1>
df18	R1	≈21k	NOR(1;0^N)	≈20.6k	≈20.6k	≈17.7k	<1,0^N NOR(1;0^N)/0^N/1>

the selected bitcells. The N bitcells remains at logic '0'. The output level of the logical operation is a logic '1'.

Waveforms in Fig. 3 present the SPICE simulation performed using the 128x128 bitcell array using the sequence of operations allowing the sensitization of the resistive-open defect df13, i.e., $<1,0^{127}$ NOR$(1;0^{127})/0^{127}/1>$ at its minimum detectable resistance. Thus, resistances below this critical value R_c=26.8kΩ lead to a correct computing operation (i.e., NOR$(1;0^N)$=0; N=127) and resistances higher than R_c lead to an incorrect behavior.

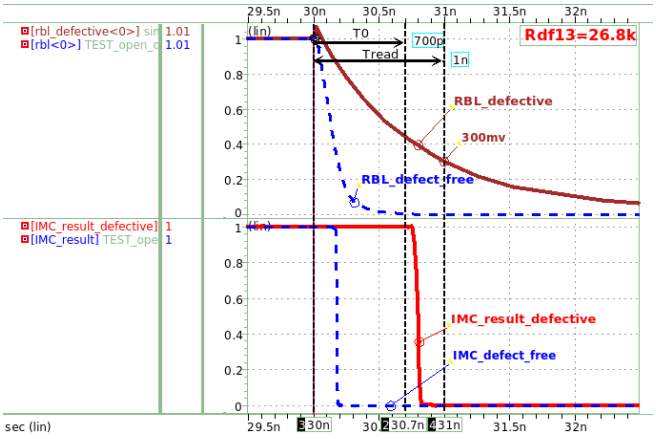

Figure 3. Waveforms of the sensitization sequence "$<1,0^N$ NOR$(1;0^N)/0^N/1>$, with N=127" with df13 size set at 26.8kΩ.

The aggressor bitcell initially contains a logic '1' and the victim bitcells contain a logic '0'. Then, a NOR is performed on the whole column by activating all the RWLs signal simultaneously at t=30ns. The RBL signal starts to discharge until it reaches 300mv at T_{read} time (i.e., 31ns), which are the two considered limits for extracting the minimum resistance of detectability. The red signal in Fig. 3 represents the result of the NOR operation at the output of the read port, i.e., the IMC_result signal in Fig. 1. Note that the blue dotted lines in Fig. 4 represent the defect-free behavior (i.e., when no resistive-open defect are injected).

C. Discussion

According to the defect behavior presented so far for an IMC 8T SRAM bitcell with resistive-open defects at the read port, some detectability conditions can be deduced. As shown in the results reported in Table 2, the IMC mode implying all the bitcells of the same column offers a better detectability of all the injected defects, i.e., an improvement of up to 13.8% for df13 to df16, of 15.7% for df18 and of 1% for df17. Note that the improvement in resistance values is different depending on the location of each defect. Moreover, different ranges of critical resistance have been found. The minimum resistance of defects that are connected to the gates of transistors is in the MΩ range (i.e., df14 and df17), while the minimum resistance for the other defects (i.e., df13, df15, df16, df18) is in the kΩ range.

From these results, the main conclusion is that the IMC mode improves the detectability of resistive-open defects and allows the detection of smaller size defects by involving all bitcells of the same column for a computing operation. Consequently, the row decoder must be adapted in order to make possible a computing operation involving all bitcells of each column for detecting and covering smaller sizes of resistive-open defects.

V. Conclusion

In this paper, we first detailed the operating principle of 8T SRAM bitcells in their two operation modes. Then, we presented the comprehensive memory model considered in our study (128x128 bitcell array in 28nm FD-SOI process technology). We highlighted the fact that the algorithms proposed in the literature to test 8T SRAM-based IMC architectures do not completely cover the defects that can affect the read port of 8T SRAM memory bitcells. Then, we presented our analysis for a thorough study of intra-cell resistive-open defects injected into the read port. Impacts in both memory and computation modes were identified, both locally (on the defective bitcell), and globally (on the array). Then, we reported results obtained during the simulation campaigns based on the qualitative analysis by specifying the critical size of the defects for which they are detectable. The obtained results show that the IMC mode improves the detectability of the injected resistive-open defects.

Our future work will consist, in a first step, in analyzing the inter-cell resistive-open and resistive-short defects so that, in a second step, it will be possible to develop an effective test and design-for-test solutions that allow to cover all the defects that can affect the IMC 8T SRAM architectures.

References

[1] A. Jaiswal, I. Chakraborty, A. Agrawal and K. Roy, "8T SRAM Cell as a Multibit Dot-Product Engine for Beyond Von Neumann Computing," in IEEE Transactions on Very Large Scale Integration Systems, vol. 27, no. 11, pp. 2556-2567, Nov. 2019.

[2] M. Horowitz, "Computing's energy problem (and what we can do about it)," 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014.

[3] S. Hamdioui et al., "Memristor based computation-in-memory architecture for data-intensive applications," 2015 Design, Automation & Test in Europe Conference & Exhibition, 2015, pp. 1718-1725.

[4] A. Agrawal, A. Jaiswal, C. Lee and K. Roy, "X-SRAM: Enabling In-Memory Boolean Computations in CMOS Static Random Access Memories," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 65, no. 12, pp. 4219-4232, Dec. 2018.

[5] T.L. Tsai, J.F. Li, C.-L. Hsu and C.-T. Sun, "Testing of In-MemoryComputing 8T SRAMs," Proc. IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems, 2019.

[6] J.-F. Li, T.-L. Tsai, C.-L. Hsu and C.-T. Sun, "Testing of Configurable 8T SRAMs for In-Memory Computing," Proc. IEEE Asian Test Symposium, 2020.

[7] L. Ammoura, M.-L. Flottes, P. Girard and A. Virazel, "Preliminary Defect Analysis of 8T SRAM Cells for In-Memory Computing Architectures," 16th International Conference on Design & Technology of Integrated Systems in Nanoscale Era, 2021, pp. 1-4.

[8] N. Planes et al., "28nm FDSOI technology platform for high-speed low-voltage digital applications," 2012 Symposium on VLSI Technology, 2012, pp. 133-134.

[9] A.J. van de Goor and Z. Al-Ars, "Functional Memory Faults: A Formal Notation and a Taxonomy", VLSI Test Symposium, pp. 281-289, 2000.

[10] "XA User Guide."https://www.synopsys.com/content/dam/synopsys/implementation&signoff/datasheets/primesim-xa-ds.pdf