Investigations of Faulty DRAM Behavior Using Electrical Simulation Versus an Analytical Approach

Zaid Al-Ars[1] Jörg Vollrath[2] Said Hamdioui[1]

[1]Delft University of Technology
Faculty of EE, Mathematics and CS
Laboratory of Computer Engineering
Mekelweg 4, 2628 CD Delft, The Netherlands

[2]Infineon Technologies AG
Product Engineering Group 3
Balanstr. 73, 81541 Munich, Germany

E-mail: z.e.al-ars@ewi.tudelft.nl

Abstract: *Fabrication process improvements and technology scaling results in modifications in the characteristics and in the behavior of manufactured memory chips, which also modifies the faulty behavior of the memory. This paper introduces an analytical (equation-based) method to give a rough analysis of the faulty behavior of cell opens in the memory, that simplifies the understanding and identifies the major factors responsible for the faulty behavior. Having these factors makes it easier to optimize the circuit and allows extrapolation of the behavior of future technologies. The paper also compares the results of the analytical approach with those from the simulation-based analysis and discusses the advantages and disadvantages of both.*

Key words: *DRAMs, faulty behavior, defect simulation, analytical evaluation, memory testing.*

1 Introduction

Technology scaling and the continuous increase in integration density results in industrial modification of the fabrication process of IC devices, which in turn induces gradual changes in the way these IC devices behave. The changes induced in circuit behavior impact the way a circuit fulfills its design criteria, but also in the way a circuits *fails* to operate according to specifications. It is industrially very desirable to come up with a rough prediction of the way the faulty behavior of a circuit changes as a result of changes in the characteristics of the fabrication process.

Previous studies have shown that it is possible to analyze the faulty behavior of DRAM devices by electrical simulation using a Spice model of the memory [Al-Ars02]. The used simulation model includes a description of the electronic circuits of the memory, including technology related parasitics to ensure proper correspondence between the simulated behavior and the one reflected by silicon.

Simulation-based fault analysis cannot be used for an early prediction of the impact of technology changes on the faulty behavior, since the simulation model only becomes available very late in the design process. It is therefore important to be able to perform a simulation-free evaluation of the faulty behavior using a generic and simple description of the memory rather than the full electrical analysis needed in the simulation-based approach. In this paper, we identify the most important parameters in a memory simulation model and introduce such a simple description where technology impact can easily be evaluated, and then compare the results with those acquired from the full simulation model and discuss the advantages and disadvantages of both.

This paper is organized as follows. Section 2 presents the electrical DRAM model used to perform memory simulations, and later used to construct the analytical (equation-based) memory model. Section 3 discusses the simulation-based analysis method used to analyze the faulty behavior of DRAMs. Section 4 identifies the model parameters most relevant to the faulty behavior and derives the analytical (equation-based) memory model. Then, Section 5 validates the model by applying it to evaluate the faulty behavior of the memory and comparing the results obtained from simulation with those proposed by the analytical model. Section 6 discusses the advantages and disadvantages of both models. Section 7 ends with the conclusions.

2 DRAM simulation model

The simulation model used in this paper is based on a design-validation model of an actual DRAM produced by Infineon Technologies. Since the time needed for simulating a complete memory model is excessively long, the simulation model used in our analysis is simplified, taking two factors into consideration in order to preserve model accuracy and usefulness. First, removed components should be electrically compensated, and second, the resulting simplified circuit should describe enough of the memory to en-

Proceedings of the 14th Asian Test Symposium (ATS '05)
1081-7735/05 $20.00 © 2005 IEEE

IEEE COMPUTER SOCIETY

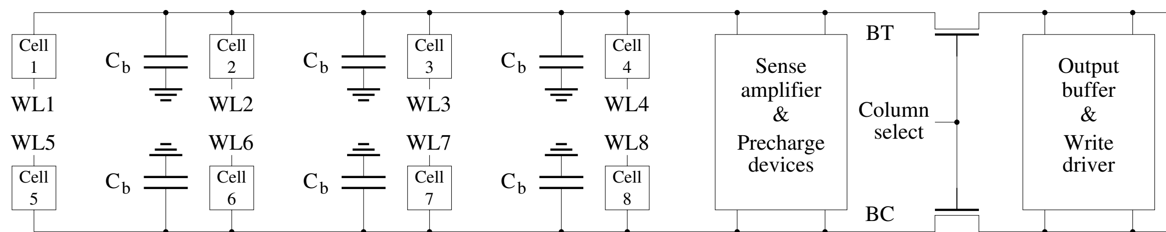

Figure 1. Block diagram of the bit line pair used for simulation.

able injecting and simulating the defects of interest.

Figure 1 shows a block diagram of the bit line pair to be simulated. This simplified simulation model contains a 4×2 cell array with nMOS access transistors, in addition to a sense amplifier and precharge devices. The removed memory cells are compensated for by capacitances of different values distributed along the bit lines. External to the bit line pair, the simulation model contains one data output buffer needed to examine data on the output, and a write driver needed to perform write operations.

At the beginning of each simulation run, cell capacitances (C_c) are initialized to the voltage level corresponding to the logic value they are supposed to store; bit line capacitances (C_b) are set to the precharge voltage (equals $\frac{V_{dd}}{2}$); and the data output buffer is forced to contain a logic 1 at the true side.

3 Simulation method

This section describes the simulation-based analysis method to evaluate the dynamic faulty behavior of defective DRAMs with cell defects [Al-Ars02]. Consider the defective DRAM cell shown in Figure 2, where a resistive open (R_{op}) between the cell capacitor and the pass transistor limits the ability to control and observe the voltage across the cell capacitor (V_c). The open is injected into Cell 1 and simulated as part of the reduced memory model shown in Figure 1. The analysis takes a range of possible open resistances 100 k$\Omega \leq R_{op} \leq$ 100 MΩ, and a range of possible cell voltages (GND $\leq V_c \leq V_{dd}$) into consideration.

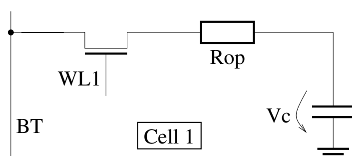

Figure 2. Open injected into Cell 1.

Two different (V_c, R_{op}) result planes are generated, one for the write 0 (Wr0) and one for the write 1 (Wr1) operation. These result planes describe the impact of successive Wr0 and successive Wr1 operations on V_c, for a given value of R_{op}. Write operations mentioned here refer to DRAM-style write operations, where a cell is accessed and successively written a number of times without being disconnected from the bit line. There are the following different DRAM-specific operations: Act, Wr0, Wr1, Rd, Pre and Nop [Vollrath00]. Figure 3 shows an automatically generated result plane corresponding to a Wr0 operation, while Figure 4 shows the result plane corresponding to a Wr1 operation, for the open R_{op} shown in Figure 2.

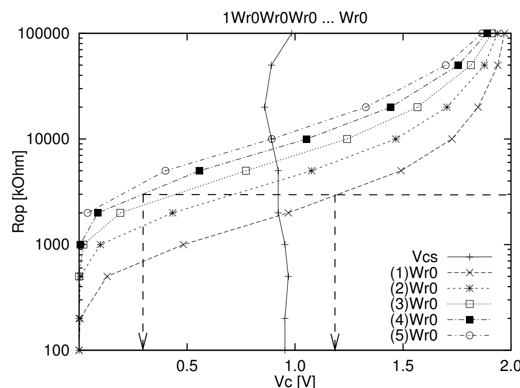

Figure 3. Result plane corresponding to Wr0.

Plane of Wr0: This result plane is shown in Figure 3. To generate this figure, the floating cell voltage V_c is initialized to V_{dd} (because a Wr0 operation is performed) and then the operation sequence 1Wr0Wr0 ... Wr0 is applied to the cell. The net result of this sequence is the gradual decrease (depending on the value of R_{op}) of V_c toward GND. The voltage level after each Wr0 operation is recorded on the result plane, resulting in a number of curves. The curves are numbered as (n)Wr0, where n is the number of Wr0 operations needed to get to the curve. For example, the arrows in the figure indicate that, for R_{op} = 5 MΩ, a

1081-7735/05 $20.00 © 2005 IEEE

IEEE COMPUTER SOCIETY

single Wr0 operation represented by (1)Wr0 pulls V_c from V_{dd} to about 1.2 V, while four Wr0 operations represented by (4)Wr0 pull V_c to about 0.3 V. We stop performing the Wr0 sequence when the voltage change ΔV_c, as a result of Wr0 operations, becomes $\Delta V_c \leq 0.05$ V, which results in identifying up to 5 different Wr0 curves in the plane. Initially, an arbitrary small value for ΔV_c is selected, which can be reduced afterwards if it turns out that more than 5 Wr0 operations are needed to describe the faulty behavior.

The sense-threshold cell voltage (V_{cs}), shown as a solid line that runs across the center of the figure, is the cell voltage above which the sense amplifier reads a 1, and below which the sense amplifier reads a 0. This curve is generated by performing a read operation for a number of V_c values and recursively identifying the V_c border that distinguishes a 1 and a 0 on the output.

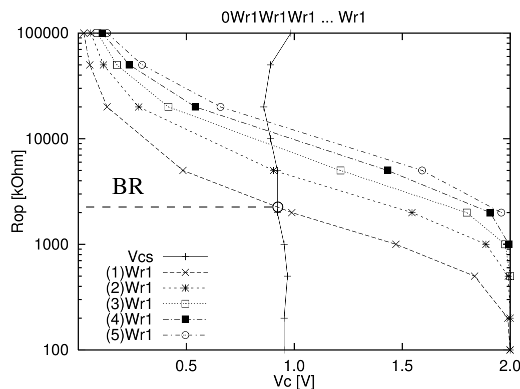

Figure 4. Result plane corresponding to Wr1.

Plane of Wr1: This result plane is shown in Figure 4, and is generated in the same way as the result plane of Wr0. First, V_c is initialized to GND and then the operation sequence 0Wr1Wr1 ... Wr1 is applied to the cell. The result is a gradual increase of V_c toward V_{dd}. The voltage level after each Wr1 operation is recorded on the result plane, which gives a number of curves in the plane. We stop the Wr1 sequence when ΔV_c becomes small enough (0.05 V in this example). V_{cs} is also shown in the figure as a solid line.

It is possible to use the result planes to analyze a number of important aspects of the faulty behavior [Al-Ars02]. One such aspect relevant to this paper is the *border resistance (BR)*, which is the R_{op} value where the cell *starts* to cause faults on the output. BR is derived on the result planes as the resistive value of the intersection point of the first write curve, either (1)Wr1 or (1)Wr0, and the V_{cs} curve. For the faulty behavior shown in Figures 3 and 4, both (1)Wr1 and (1)Wr0 intersect the V_{cs} curve at about 2 MΩ, which also means that BR $\approx$ 2 MΩ. Close inspection shows that (1)Wr0 crosses V_{cs} slightly below 2 MΩ, while (1)Wr1 crosses it slightly above 2 MΩ.

Another important aspect relevant to this paper is generating a test that detects the faulty behavior of the defect. Since BR has a value of 2 MΩ, a test should be able to detect faults for defects with a resistance value above, but as close as possible to, BR. If we assume that V_c is initialized to V_{dd}, then a fault is detected by performing a sequence of Wr0 Pre Act Rd0, while a V_c initialized to GND causes the sequence Wr1 Pre Act Rd1 to fail. In order to initialize V_c to V_{dd}, a sequence of Wr1 operations is used, and to initialize V_c to GND, a sequence of Wr0 operations is used. Inspecting the result planes at $R_{op} = 2$ MΩ indicates that in the worst case a sequence of 4 to 5 Wr1 operations charges V_c up to about 1.9 V, while a sequence of 4 to 5 Wr0 operations discharges V_c to about 0.1 V. It is possible to achieve higher and lower cell voltages by increasing the number of write operations in the initialization sequence, but the more operations are used the less the effect of each added operation becomes. Therefore, there are two possible conditions for detecting R_{op}:

- Cond1: $\Updownarrow$(..., Wr0, Wr0, Wr0, Wr0, Wr1, Pre, Act, Rd1, ...)
- Cond0: $\Updownarrow$(..., Wr1, Wr1, Wr1, Wr1, Wr0, Pre, Act, Rd0, ...)

But since (1)Wr0 crosses V_{cs} at a resistance below that of (1)Wr1, Cond0 is expected to detect defects with lower R_{op} values than Cond1.

4 Analytical operation model

The simulation-based fault analysis described in Section 3 requires a fairly well developed simulation model and the results of the analysis give limited insight into the reasons behind the simulated behavior due to the complexity of the model. This section introduces a simplified analytical model of the memory, where we only attempt to include the most relevant parameters to properly reflect the faulty behavior. Proper evaluation of the faulty behavior for the open defect shown in Figure 2 is based on a proper evaluation of the way the defective memory behaves when sensing the voltage in the cell and when performing a write operation.

4.1 Sense simplification

Sensing depends mainly on the way the sense amplifiers (SAs) detect the voltage difference present across their terminals. Although there are many types of sense amplifiers,

Proceedings of the 14th Asian Test Symposium (ATS '05)
1081-7735/05 $20.00 © 2005 IEEE
IEEE COMPUTER SOCIETY

each with its own principle of operation, most of them operate according to the same concept: a positive voltage differential is amplified by the sense amplifier to a full voltage high, while a negative voltage differential is amplified to a full voltage low.

Based on this simple concept of operation one can define a so-called *ideal sense amplifier* as shown in Figure 5, where only two different parameters (the input voltage, V_{in}, and the reference voltage, V_{ref}) on the input define the resulting voltage on the output (V_{out}) according to the following relation:

$$V_{out} = \begin{cases} V_{dd} & : \quad V_{in} > V_{ref} \\ \text{GND} & : \quad V_{in} < V_{ref} \end{cases} \tag{1}$$

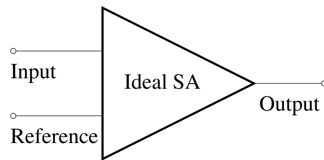

Figure 5. Simplified representation of a sense amplifier.

The ideal sense amplifier is perfectly balanced around its reference voltage. This is not the case for a real sense amplifier, which has a number of different sources of imbalance that cause a biased sensing of a 0 or a 1 on the output. Causes of sense amplifier imbalance include [Sarpeshkar91]:

1. transconductance of the sense amplifier transistors
2. gate-source parasitic capacitance of transistors
3. threshold voltage of transistors
4. bit line capacitance

4.2 Write simplification

Write operations depend mainly on the way a cell is accessed through the pass transistor and subsequently charged up (in case of Wr1) or discharged (in case of Wr0), which means that proper simplification of the pass transistor is important for write operations. Figure 6 shows a simplified representation of the memory cell, where the pass transistor is replaced with a switch in combination with the conduction resistance or *on-resistance* of the transistor (R_{on}).

In Figure 6, a high voltage on the WL turns the switch on, while a low voltage turns the switch off. In order to derive a mathematical expression for R_{on}, we assume that

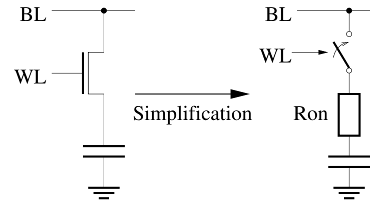

Figure 6. Simplified representation of the memory cell.

the WL voltage (V_W) is boosted to a level higher than $V_{dd} + V_T$, where V_T is the threshold voltage of the pass transistor. Such a boosted WL voltage makes the pass transistor operate in the linear region when accessed, both during charging up and discharging the cell capacitor. As a result, the drain current (I_D) can be calculated as follows:

$$I_D = \beta V_{DS}(V_{GS} - V_T - \frac{1}{2}V_{DS}) \tag{2}$$

where β is the transconductance parameter of the pass transistor, while V_{DS} and V_{GS} are the drain-source and the gate-source voltages, respectively. From Equation 2, we can calculate the on-resistance as the differential of V_{DS} with respect to I_D as follows:

$$R_{on} = \frac{\partial V_{DS}}{\partial I_D} = \frac{1}{\beta(V_{GD} - V_T)} \tag{3}$$

where V_{GD} is the gate-drain voltage of the pass transistor.

In an nMOS transistor, the drain current always flows from the drain node toward the source node of the transistor. This means that when the cell is charging up (BL voltage $>$ cell voltage), the transistor node connected to the bit line is considered to be the drain, while when the cell is discharging (BL voltage $<$ cell voltage), the transistor node connected to the cell is considered to be the drain. This results in the following expression for the on-resistance:

$$R_{on} = \begin{cases} \frac{1}{\beta(V_W - V_B - V_T)} & : \quad V_B > V_c \\ \frac{1}{\beta(V_W - V_c - V_T)} & : \quad V_B < V_c \end{cases} \tag{4}$$

where V_B is the BL voltage and V_c is the cell voltage. Equation 4 shows that the on-resistance stays constant when the cell is charging up, while it decreases gradually as the cell is discharged toward a value of $\frac{1}{\beta(V_W - V_T)}$ when $V_c = 0$ V. Depending on V_B and V_c, the value of R_{on} ranges between 500 Ω and 1.5 kΩ. These changes are negligible since we always analyze the faulty cell behavior with the presence of an $R_{op} > 10\text{ k}\Omega$ connected in series with R_{on}. In the rest of this paper, we consider the value of R_{on} to be constant

$$R_{on} = 1\text{k}\Omega \tag{5}$$

1081-7735/05 $20.00 © 2005 IEEE
IEEE COMPUTER SOCIETY

5 Model validation

In order to ensure the ability of the simplified analytical model discussed in Section 4 to represent the faulty behavior of the memory, this section applies the analytical model to predict the faulty behavior of the memory and compares the outcome with the results of the detailed simulation-based fault analysis discussed in Section 3. Here, we consider the defective DRAM cell shown in Figure 7, where a resistive open (R_{op}) and the on-resistance (R_{on}) limit the ability of the memory to control and observe the voltage across V_c. The analysis takes a range of possible open resistances 10 k$\Omega \leq R_{op} \leq$ 10 MΩ, and a range of possible cell voltages (GND $\leq V_c \leq V_{dd}$) into consideration. Since R_{on} is connected in series to R_{op}, which is at least about 10 time larger than R_{on}, we consider R_{on} to be constant and equal to 1 kΩ [see Equation 5].

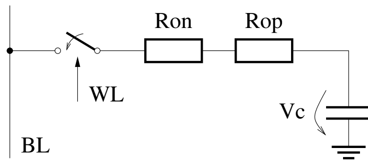

Figure 7. R_{op} injected into simplified cell model.

When a read operation is performed the BL of the cell is connected to the ideal SA of Figure 5 with $V_{ref} = \frac{V_{dd}}{2}$. Therefore, the output of a read operation depends on V_c as follows [see Equation 1]:

$$\text{Read output} = \begin{cases} 0 & : \quad V_c < \frac{V_{dd}}{2} \\ 1 & : \quad V_c > \frac{V_{dd}}{2} \end{cases} \tag{6}$$

When a write 0 operation is performed, the BL of the cell is forced to GND, while a write 1 forces the BL to V_{dd}. To find out the way V_c develops with time during charge up and discharge, we need to solve the differential equation of the cell:

$$C_c \frac{\partial V_c}{\partial t} = \frac{V_B - V_c}{R_{on} + R_{op}} \tag{7}$$

where C_c is the cell capacitance. Solving Equation 7 results in:

$$\text{Wr1} \quad : \quad V_c(t) = V_{dd}\,(1 - e^{\frac{-t}{C_c(R_{on}+R_{op})}}) \tag{8}$$

$$\text{Wr0} \quad : \quad V_c(t) = V_{dd}\, e^{\frac{-t}{C_c(R_{on}+R_{op})}} \tag{9}$$

where $V_B =$ GND when writing a 0, and $V_B = V_{dd}$ when writing a 1.

Using Equations 6–9, two different (V_c, R_{op}) result planes are generated, one for the Wr0 and one for the Wr1 operation. These result planes describe the impact of successive Wr0 and successive Wr1 operations on V_c, for a given value of R_{op}. Write operations mentioned here refer to DRAM-style write operations, where a cell is accessed and successively written a number of times without being disconnected from the bit line. Figure 8 shows the analytically generated result plane corresponding to Wr0 operations, while Figure 9 shows the result plane corresponding to Wr1 operations.

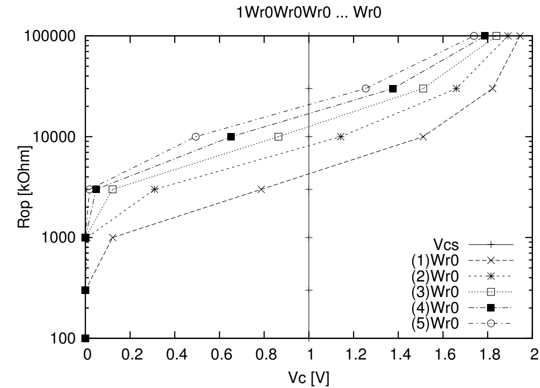

Figure 8. Analytically generated result plane for Wr0.

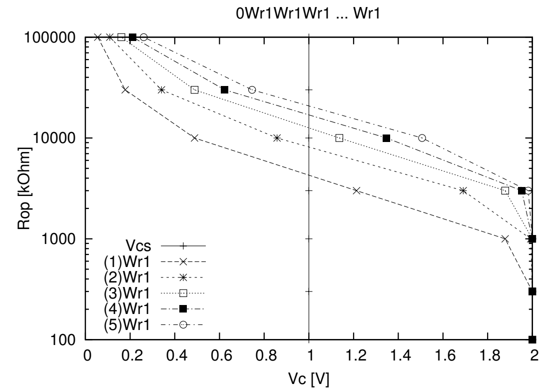

Figure 9. Analytically generated result plane for Wr1.

The figures show strong correspondence with the simulation-based result planes shown in Figures 3 and 4. Write 0 and write 1 operation curves, as well as the V_{cs} curve, all have similar dependence on R_{op} as those from the simulation-based analysis. The faulty behavior described by the analytical approach is quite similar too. The two figures are mirror images of each other, which means that both Wr1 and Wr0 have an exactly complementary faulty behavior. The value of the border resistance according to these figures is about 4 MΩ, compared to 2 MΩ

1081-7735/05 $20.00 © 2005 IEEE
IEEE COMPUTER SOCIETY

in the simulation-based analysis. On the other hand, the test required to detect the faulty behavior in the figure is identical to that required by the simulation-based analysis, $\Updownarrow$(..., Wr0, Wr0, Wr0, Wr0, Wr1, Pre, Act, Rd1, ...) or its complementary to ensure detection. This analysis also indicates that it is possible to achieve higher coverage by increasing the number of write operations in the initialization sequence, but the more operations are used the less the effect of each added operation becomes. This means that the analytical model, despite its extreme simplicity, is very capable of providing a good deal of useful information about the faulty behavior of the memory.

On the other hand, the figures show clear differences in the details they unveil about the faulty behavior of the injected open. One such difference is with regard to the shape of the V_{cs} curve, which is analytically shown as a straight vertical line at $V_c = 1$ V, while the simulation-based analysis indicates that it deviates from this vertical line toward GND. This difference can be attributed to the ideal nature of the analytical SA of Figure 5, which lacks any source of sensing imbalance present in the simulation-based SA (such as transconductance mismatch or threshold voltage differences) [see Section 4.1]. A second difference is with regard to the perfect symmetry of the Wr0 and Wr1 curves around the $V_c = 1$ V line resulting from the analytical model, whereas the simulation-based results show that the Wr0 curves are faster in writing a 0 into the cell than the Wr1 curves. This difference can be attributed to the simplifying assumption made in the analytical model that R_{on} is constant during a Wr0 operation, when in fact it decreases gradually as the cell discharges, thereby accelerating the Wr0 process [see Section 4.2].

6 Advantages and disadvantages

The analytic mathematical equations, derived as shown in Section 4, can be used as a tool to check the understanding of the faulty behavior. The comparison of the result from the simple equations with the results from the electrical simulations points out deficits in the mathematical model and thus the understanding of the behavior. Analyzing these differences can lead to a refinement of the analytical model and subsequently an increase in the understanding.

Advantages—The analytical approach has the following advantages over the simulation-based approach:

- The analytical approach gives more insight into the causes of the faulty behavior as a result of the limited number of model parameters contributing to the observed behavior.
- It needs a simple set of design and technology parameters that can be acquired well in advance of any design process, thereby enabling an upfront evaluation of the faulty behavior associated with a specific technology.
- Fast and simple to perform, which makes it easy to evaluate the impact of different parameters (changes to temperature and timing, for example) on the faulty behavior, by simply including the impact of each parameter on the value of the R_{on} resistance.

Disadvantages—The discussion also shows that the analytical approach has some disadvantages:

- The limited accuracy of the analytical model means that it is unreliable to identify the impact of sensitive parameters (such as parasitics and capacitive coupling) on the faulty behavior.
- The analytical model includes only those parameters that we assume to have the biggest influence the faulty behavior, which means that we only get to see the effects we expect to see in the faulty behavior.
- The derivation of the analytical model is mathematically involved, and each defect requires a totally different analytical model to be derived, whereas the simulation-based analysis simply injects a defect into the same model to perform the analysis.

7 Conclusions

This paper discussed an analytical approach to analyze the faulty behavior of memory devices. This approach is characterized by its simplicity and the full control we have on the model. The paper evaluated the capabilities of the analytical approach using an example of open within the cell, and comparing the results with those acquired from the simulation-based analysis approach. The comparison between the two approaches show strong correspondence between them, and validates the ability of the analytical approach to produce a good deal of useful information, despite its extreme simplicity. The comparison also indicates that each approach has its own advantages during the fabrication process, where simulation has the advantage in the precise identification of new types of fault effects, while the analytical approach has the advantage in evaluating the expected changes in the faulty behavior as a result of technology scaling or changes in the fabrication process.

References

[Al-Ars02] Z. Al-Ars and A.J. van de Goor, "Approximating Infinite Dynamic Behavior for DRAM Cell Defects," *in Proc. IEEE VLSI Test Symp.*, 2002, pp. 401–406.

[Naik93] S. Naik, F. Agricola and W. Maly, "Failure Analysis of High Density CMOS SRAMs," *in IEEE Design and Test of Computers*, vol. 10, no. 2, 1993, pp. 13–23.

[Sarpeshkar91] R. Sarpeshkar *et al.*, "Mismatch Sensitivity of a Simultaneously Latched CMOS Sense Amplifier," *in IEEE J. of Solid-State Circuits*, vol. 26, no. 10, 1991, pp. 1413–1422.

[Vollrath97] J.E. Vollrath, "Cell Signal Measurement for High-Density DRAMs," *in Proc. IEEE Int'l Test Conf.*, 1997, pp. 209–215.

[Vollrath00] J. Vollrath, "Tutorial: Synchronous Dynamic Memory Test Construction, A Field Approach," *in Proc. IEEE Int'l Workshop Memory Technology, Design and Testing*, 2000, pp. 59–64.

IEEE COMPUTER SOCIETY