Non-Algorithmic Stress Optimization Using Simulation for DRAMs

Zaid Al-Ars Said Hamdioui

Laboratory of Computer Engineering, Faculty of EE, Mathematics and CS

Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands

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

Abstract: *Stress optimization for memory devices is a complex process due to the continuous space of possible optimization values for relevant parameters. This paper uses a method based on electrical Spice simulation to perform this optimization process for DRAM devices. The paper presents a case-study performed in Qimonda to optimize the timing and temperature stresses for the strap problem in defective memory cells. The paper also considers the impact of bit line coupling effects on the faulty behavior and identifies the worst case coupling background needed to detect the faulty cells.*

Keywords: DRAM testing, non-algorithmic stresses, stress optimization, Spice simulation, strap problem

1 Introduction

Throughout the manufacturing process of memory devices, memory testing procedures (called *test insertions*) have to be performed at a number of critical stages of the process to ensure the proper functionality of the produced memory devices. The two main test insertions found in any memory manufacturing process are the following [Falter00]:

1. *wafer-level* testing, which takes place before individual memory chips are diced and packaged
2. *device-level* testing, which takes place on individual memory chips after packaging.

At each of these stages, a different set of test requirements is needed to enable the effective evaluation of the behavior of each product. For wafer-level testing in particular, it is required to extensively make use of non-algorithmic stresses to bring the chips closer to failure. The non-algorithmic stresses most widely used by test engineers are timing, voltage and temperature [Vollrath00]. Using these stresses during wafer-level testing makes it possible to identify weak chips that will not function properly later on during device-level testing.

The task of optimizing the different non-algorithmic stresses is commonly done by experienced test engineers based on their general test development expertise and their knowledge of the product under test. However, using these stresses is rather complex due to the continuous space of possible optimization values of relevant parameters, which leads to an infinite amount of different usable stress combinations. This makes it impractical to empirically identify optimal stress values by straight forward stress deployment on a defective chip.

This paper uses a simulation-based method to perform stress optimization [Al-Ars08], which provides the test engineer with valuable insights into the faulty behavior of defective chips, and the way this behavior changes as a result of a specific applied stress. Electrical simulation has been successfully applied before to analyze memory faults [Azimane04, Lin00, Mohammad03]. The paper uses the simulation method to optimize the needed stresses for a DRAM device produced by Qimonda, where the cells suffer from an elevated strap resistance problem.

The paper is organized as follows. Section 2 discusses the electrical Spice model used for simulating the behavior of the memory, along with a method to model the defect to be simulated. Section 3 describes the simulation-based fault analysis method to be used to evaluate the faulty behavior of the memory. Then, Section 4 uses the concept of critical resistance to enable stress optimization for the defect under test. Section 5 ends with the conclusions.

2 Memory simulation model

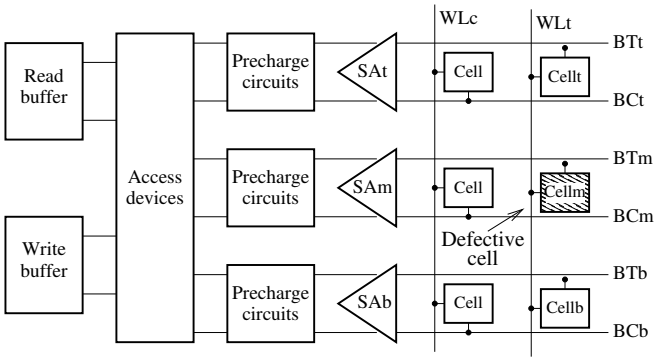


Figure 1. Spice simulation model used for fault analysis.

The electrical memory simulation model used to perform the simulations for the strap problem is a design validation model in the 0.11μm technology. To reduce simula-

tion time, the memory model is reduced to include only those parts of the memory needed to perform the fault analysis. Figure 1 shows a block diagram of the memory model. The model has three *bit line (BL)* pairs: top true BL (BTt) and top complement BL (BCt); middle true BL (BTm) and middle complement BL (BCm); and bottom true BL (BTb) and bottom complement BL (BCb). Each BL pair is connected to 2 memory cells, one to BT and the other to BC. In addition, the model has 3 sense amplifiers (SAt, SAm and SAb), precharge circuits and access devices. A write buffer is included to enable simulating a write operation, in addition to a read buffer for simulating a read operation.

The model has two *word line (WLs)*, each is connected to three memory cells: WLt is connected to three cells on BT, while WLc is connected to three cells in BC. The fault analysis described in this paper is performed on Cellm (the memory cell connected to WLt and BTm). The behavior of cells connected to BC is the complementary to that of cells connected to BT (i.e., with all 0s replaced with 1s, and vice versa).

The top and bottom BL pairs are included in the model in order to simulate the impact of *coupling background (CB)*patterns on the faulty behavior. Each BL is connected to the two adjacent BLs by parasitic capacitances. When Cellm is accessed, Cellt and Cellb are accessed at the same time (since all are connected to WLt), thereby influencing the behavior of the operations performed on Cellm. To simulate the impact of different CBs, the simulation analysis is performed for different voltages stored in Cellt and Cellb. There are four different CBs:

1. **CB 00**—This refers to 0s stored in Cellt on BTt and Cellb on BTb.
2. **CB 10**—This refers to a 1 stored in Cellt on BTt and a 0 stored in Cellb on BTb.
3. **CB 11**—This refers to 1s stored in Cellt and Cellb.
4. **CB 01**—This refers to a 0 stored in Cellt and a 1 stored in Cellb.

The special open defect analyzed in this paper models an increase in the resistive value of what is called the **strap connection**. The strap connection is a conductive path between the drain region of the pass transistor of the memory cell and the trench capacitor [Adler95]. Figure 2(a) gives a schematic representation of the layout of the cell and the position of the strap in the cell. In addition, the figure show the WL connection, the BL connection, as well as the trench capacitor of the memory. Due to imperfections in the fabrication process, the strap may take up any resistive value according to the statistical distribution of the fabrication process.

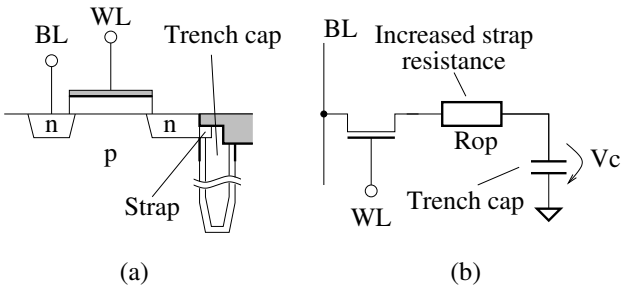


Figure 2. Modeling the strap at (a) the layout, and (b) electrical level.

Ideally, the memory is designed such that the strap should be manufactured with a predefined target resistance value. An *increase* in the strap resistance can be electrically modeled as an added series resistance (R_{op}) along the conductive path between the pass transistor and the trench capacitor in the cell, as shown in Figure 2(b). An increase in R_{op} reduces the ability of the memory to control the voltage stored across the cell capacitor, which leaves the stored voltage in the cell (V_c) floating to a certain extent.

From a physical point of view, the modeled increase in the strap resistance R_{op} can be attributed to a number of factors, such as a change in the doping concentration of the strap, or a geometrical misalignment in the positioning or sizing of the strap.

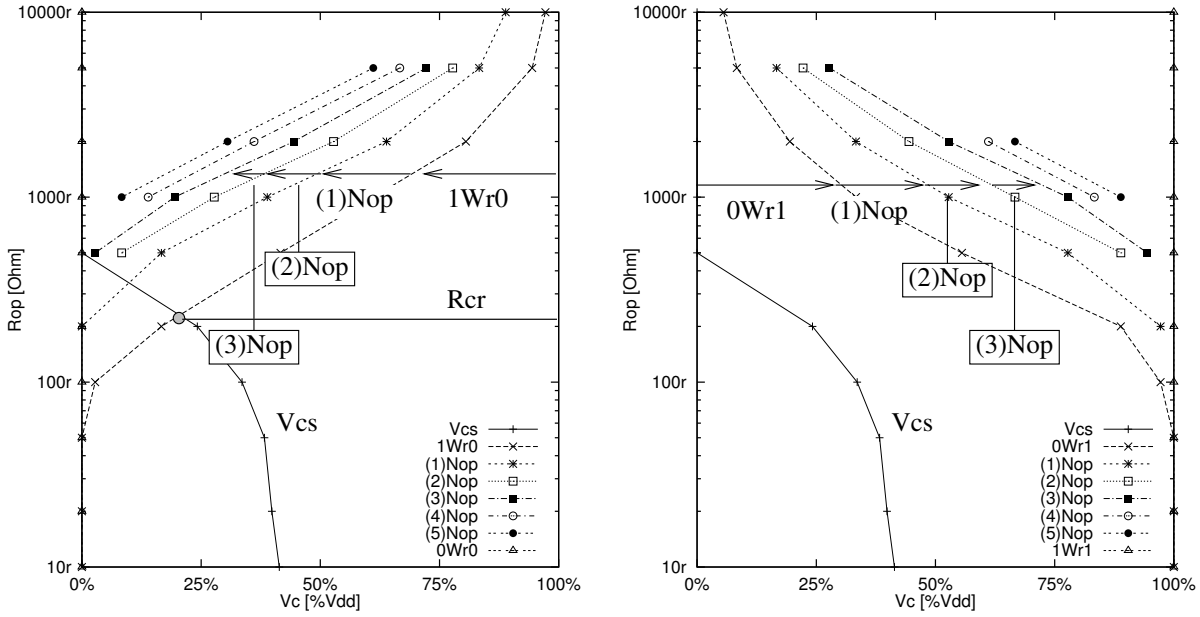
3 Analysis methodology

In this section, we describe the simulation-based fault analysis methodology by discussing the analysis performed using CB 00.

3.1 Simulated sequences

Since there is an infinite number of possible operation sequences to simulate, it is impossible to simulate all of them. The solution is to simulate a limited number of sequences, called *basic sequences*, and then use those to *approximate* the behavior of any other sequence. For DRAMs, basic sequences should enable us of approximating any functional sequence composed of the five basic DRAM commands: activate (Act), read (Rd), write (Wr), precharge (Pre) and no operation (Nop). Simulations of the strap problem indicate that only the following sequences are enough to analyze the total faulty behavior of the cell [Al-Ars08]:

- Sequence of Wr0: `Wr0 Nop Nop ... Nop`
- Sequence of Wr1: `Wr1 Nop Nop ... Nop`
- Sequence of Act: `Act Nop Nop ... Nop`



(a) Plane of Wr0

(b) Plane of Wr1

Figure 3. Result planes at nominal stress and with CB 00, for the sequences (a) Wr0, and (b) Wr1.

3.2 Background 00

Figures 3 and 4 show the simulation results at nominal stresses (according to the specifications of the memory) and with CB 00 (0 is stored in cells on the adjacent BL pairs). The results are divided into three different result planes, one for each analyzed basic sequence. Each result plane describes the impact of performing successive commands on the stored voltage within the cell (V_c) for a given value of the open resistance (R_{op}), as shown in Figure 2(b). The x-axes in the result planes represents V_c, while the y-axis represents the value of R_{op}. The value of V_c is not given in absolute voltage levels, but as percentages of V_{dd}. In the same way, a scaled value of R_{op} is shown on the y-axis using the scale factor r.

Plane of Wr0: This result plane is shown in Figure 3(a). To generate this figure, the floating cell voltage V_c is initialized to the two worst case voltages, V_{dd} and GND, and then the sequence Wr0 Nop ... Nop is applied to the cell. With an initial $V_c = V_{dd}$, the sequence results in the gradual decrease (depending on the value of R_{op}) of V_c towards GND. With an initial $V_c =$ GND, the value of V_c remains at GND. The voltage level after each command in the sequence is recorded on the result plane, which results in a number of curves in the plane. All curves have names, and some of them are indicated by an arrow pointing in the direction of voltage change. The 1Wr0 curve identifies the impact of Wr0 on a cell voltage initialized to V_{dd}, while the 0Wr0 curve (the last entry in the legend) identifies the impact of Wr0 on a cell voltage initialized to GND. The curves numbered as (n)Nop indicate the impact of no operations on V_c following a 1Wr0, where n is the number of Nops needed to get to the indicated curve. The figure also shows the cell sense-threshold curve (V_{cs}), above which the sense amplifier senses a 1 and below which the sense amplifier senses a 0. The V_{cs} curve is copied from the plane of the Act sequence, which is explained in detail below [see "Plane of Act" below]. This plane enables evaluating the effect of any Wr0 on the defective cell.

Plane of Wr1: This result plane is shown in Figure 3(b). To generate this figure, V_c is initialized to the two worst case voltages V_{dd} and GND and then the sequence Wr1 Nop ... Nop is applied to the cell. With an initial $V_c =$ GND, the result is the gradual increase of V_c towards V_{dd}, while an initial V_{dd} remains as it is in the cell. The voltage level after each command in the sequence is recorded on the result plane, which produces a number of curves in the plane. These curves are indicated in the same way as for the curves in the plane of Wr0 above. This curve enables evaluating the effect of any Wr1 on the defective cell.

Plane of Act: This result plane is shown in Figure 4. To generate this figure, first we identify the threshold voltage within the cell that determines the sense amplifier output V_{cs} (the cell voltage above which the sense amplifier detects a 1, and below which it detects a 0). Then, the sequence Act Nop ... Nop is applied twice: first for V_c that is initially marginally lower than V_{cs}, and a second time for V_c that is marginally higher than V_{cs}. After each command, V_c is recorded on the result plane, which results in a number of curves on the plane. The +Act curve indicates the impact of performing Act with V_c marginally higher than V_{cs}, while -Act indicates the impact of performing Act with V_c marginally lower than V_{cs}. The other curves indicated the impact of the nth Nop following the initial Act.This plane enables evaluating the effect of any Act on the defective cell.

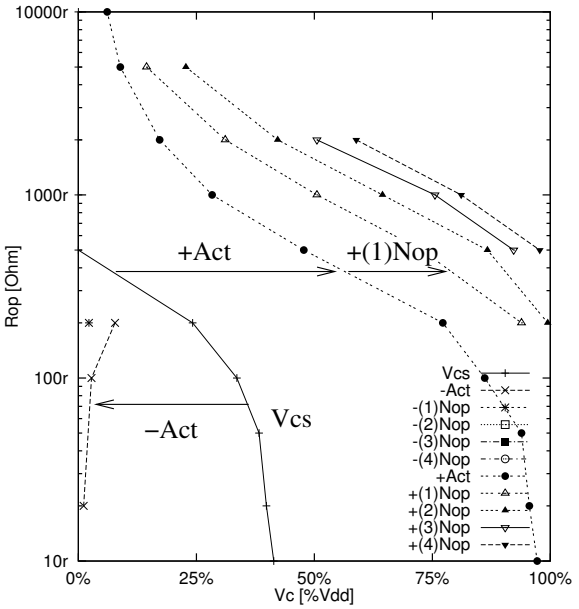


Figure 4. Result plane at nominal stress and with CB 00, for the Act sequence.

Using the result curves in Figure 3, we can analyze the following aspects of the faulty behavior:

1. Identify the *critical resistance* (R_{cr}), which is the R_{op} value where the cell *starts* to cause faults on the output, for *any* sequence of operations.
2. Generate a test that detects the faulty behavior of the defect for any resistance value and any initial floating voltage.

(1) For the fault analysis shown in Figure 3, the memory behaves properly for any operation sequence as long as $R_{op} < 210r\Omega$. To understand why, note that a fault would only be detected when a Wr1 command fails to charge V_c up above V_{cs}, or a Wr0 fails to discharge V_c to below V_{cs} (V_{cs} is indicated by a curve in Figure 3). In both situations, trying to read after performing the write would detect the faulty behavior. Note that for $R_{op} > 210r\Omega$, Wr0 fails to discharge V_c to the value needed by Act to sense a 0. This is indicated in Figure 3(a) as a dot at the intersection between the 1Wr0 curve and the V_{cs} curve. Furthermore, note that the curve 0Wr1 in Figure 3(b) does not intersect the V_{cs} curve, which means that Wr1 never fail no matter how high R_{op} becomes!

(2) Now, the result planes are used to generate a detection condition that detects the faulty behavior caused by any defect resistance for any initial floating voltage, in case a fault can be detected. Figure 3(a) shows that faults can be detected with $R_{op} \geq 210r\Omega$. Inspecting the figure shows that with $R_{op} \geq 210r\Omega$, and with any initial voltage V_c, the sequence Wr1 Nop Nop Wr0 will sensitize a fault. This can be validated by checking Figure 3(b) for $R_{op} = 210r\Omega$, and noting that performing Wr1 Nop Nop charges V_c up from any voltage (GND or higher) to approximately V_{dd}. With $V_c = V_{dd}$, performing Wr0 sensitizes the fault which can then be detected as discussed in point (1) above. Therefore to detect the fault, the detection condition $\Updownarrow$(..., Wr1, Nop, Nop, Wr0, Pre, Act, Rd0, ...) is sufficient.

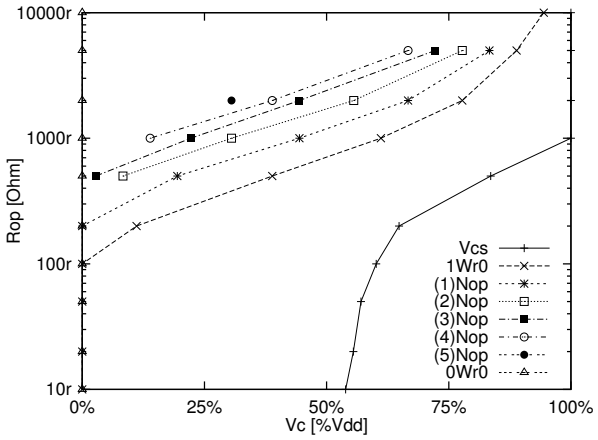
3.3 Backgrounds 10, 11 and 01

Figure 5 shows the analysis results for CB 11. Figures 5(a) and (b) give the results for Wr0 and Wr1, respectively. The figures show that the V_{cs} curve changed significantly with CB 11, compared to CB 00. The sense amplifier is now biased towards detecting a stored 0 instead of detecting a stored 1.

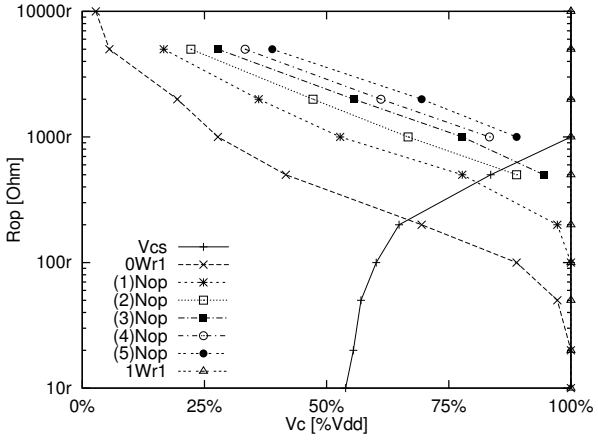
Inspecting Figure 5(a) reveals that the V_{cs} curve does not intersect the 1Wr0 or any of the (n)Nop curves, which means that the Wr0 sequence never fails. However, the V_{cs} curve does intersect the 0Wr1 curve in Figure 5(b) at about $R_{op} = 205r\Omega$. This indicates that the Wr1 sequence starts to fail with $R_{op} \geq 205r\Omega$.

A detection condition to detect this fault is $\Updownarrow$(..., Wr0, Nop, Nop, Wr1, Pre, Act, Rd1, ...). It is interesting to note that this detection condition has a similar sequence of commands as the detection condition derived for CB 00, with the exception that the data used in this detection condition is complementary to that used in the condition for CB 00 (i.e., 1s are replaced with 0s, and vice versa).

Analysis of the behavior of CB 10 and CB 01 shows that they behave in a very similar way to CB 00 and CB 11, respectively. For CB 10, we note a small increase in the range of failing R_{op} values, which indicates that CB



(a) Wr0



(b) Wr1

Figure 5. Analysis results with CB 11 for (a) Wr0 and (b) Wr1.

10 ($R_{cr} = 205r\Omega$) is slightly more effective in stressing the test than CB 00 ($R_{cr} = 210r\Omega$). In a similar way, CB 01 ($R_{cr} = 200r\Omega$) is slightly more effective in stressing the test than CB 11 ($R_{cr} = 205r\Omega$).

4 Optimizing test stresses

This section discusses the problem of test stress optimization for the strap problem. Stress optimization for a test means properly adjusting the stresses (timing, temperature and voltage) such that a higher coverage of a given test can be achieved. Simulation-based optimization of stresses is made possible using the concept of the critical resistance (R_{cr}) of a defect [Al-Ars05]. This important piece of information can be used for optimizing any test stress, as follows:

> A change in a given stress should modify the value of R_{cr} in that direction which maximizes the range of a detectable functional fault.

In the following, two different stresses are analyzed: timing represented by the clock cycle time (t_{cyc}), and temperature. The optimization of voltages is not discussed here, because of the confidential nature of this information.

4.1 Optimizing t_{cyc}

In this section, the clock cycle time is optimized by inspecting the impact of a number of t_{cyc} values on the resulting R_{cr}. The fault analysis results of Section 3 indicate that R_{cr} is specified by the intersection point of the first Wr command curve and the V_{cs} curve [see Figure 3(a), for example]. As a result and in order to identify R_{cr} for every t_{cyc}, it is sufficient to trace the value of R_{cr} by generating the Wr command curve and the V_{cs} curve, and subsequently tracing their point of intersection.

Figure 6 shows the critical resistance as a function of t_{cyc}. The x-axis in the figure lists t_{cyc}, while the y-axis in the figure lists the value of R_{cr}. There are four curves in the figure, one for each CB. The first important conclusion one can derive from the figure is that t_{cyc} is a *decisive* stress for the strap problem, which means that it is possible to use timing to induce a detectable fault in any cell (defective as well as functional). This is indicated by the linear decline of R_{cr} toward 0 $r\Omega$ with decreasing t_{cyc}.

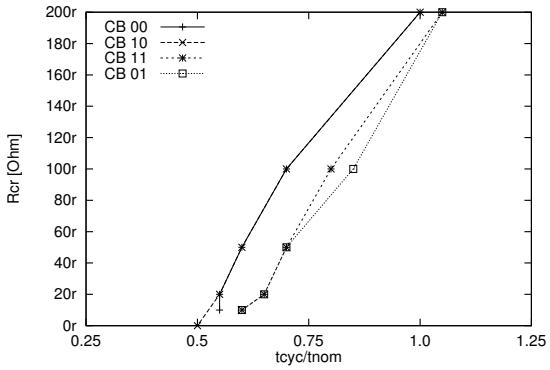


Figure 6. Critical resistance as a function of cycle time.

Furthermore, the following conclusions can be derived from the figure:

- In general, the critical resistances with CB 00 and CB 10 have values that are close to each other. The same is true for the critical resistance values with CB 11 and CB 01.
- As t_{cyc} decreases below its nominal value, the critical resistances with CB 00 and 10 start to deviate in a clear way from those with CB 11 and 01.

- With the exception of very low values of t_{cyc}, the critical resistance with CB 10 and 01 have either a lower or exactly the same value as that with CB 00 and 11, respectively.

- For a given t_{cyc} that is below t_{nom}, CB 11 and CB 01 are more stressful than CB 00 and CB 10.

4.2 Optimizing temperature

In this section, an analysis is performed of the temperature effect on the faulty behavior. The objective is to find the temperature that is most stressful for our test. In the following, the critical resistance (R_{cr}) is evaluated at $-50°$ C, $-10°$ C, $27°$ C, $87°$, $110°$ C and $150°$ C.

The analysis has been performed for all 4 CBs and with nominal voltages and using $t_{cyc} = 0.75t_{nom}$.

Figure 7 shows the R_{cr} value as a function of temperature with all CBs. According to the figure, the value of R_{cr} increases with increasing temperature. This is expected since the value of the defect resistance (R_{op}) decreases with increasing temperature. This means that a higher temperature reduces the impact of the defect on the faulty behavior, thereby increasing R_{cr}.

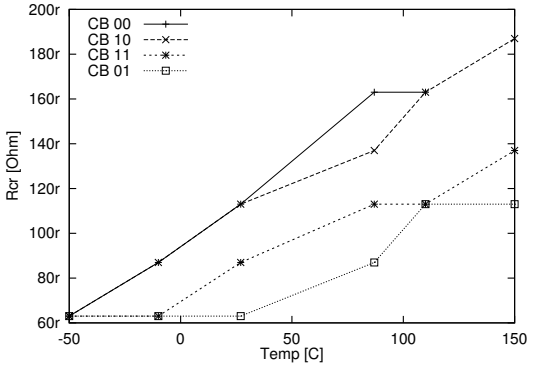


Figure 7. Critical resistance as a function of temperature.

The first important conclusion one can derive from the figure is that temperature is an *indecisive* stress, which means that it is *not* possible to use temperature to induce a detectable fault in any cell (defective as well as functional). This is indicated by the very slow decline in R_{cr} with decreasing temperature, which reaches a minimum of about $60\ r\Omega$ within the analyzed temperature range. An even lower temperature may result in a lower R_{cr}, but such extreme temperatures are both expensive and unreasonable, falling well beyond the bounds of industrial standards.

The figure shows similar trends to those observed with the optimization analysis of the cycle time. Two groups of R_{cr} traces can be identified, CB 00 and 10 and CB 11 and 01, where the traces in each group change in the same way and remain close together. In addition, the CBs 10 and 01 are still more effective than their counterparts CB 00 and 11, respectively. The most stressful CB is 01 along the whole analyzed temperature range.

5 Conclusions

This paper presented a case study to apply the simulation-based fault analysis method in analyzing the faulty behavior of the elevated strap problem. The analysis results make it possible to both generate test patterns to detect the faulty behavior, as well as optimize those tests with respect to various memory stresses. The paper discussed the analysis method in detail, starting from the stage of defining an electrical model used in the simulation, through the stage of fault analysis, till the stage of test stress optimization. The paper also considered the effect of BL coupling on the faulty behavior and showed that stress did not modify the worst case coupling background needed to get the maximum stress on the defective cell.

References

[Adler95] E. Adler *et al.*, "The Evolution of IBM CMOS DRAM Technology," *in IBM J. of Research and Development*, vol. 39, no. 1–2, 1995, pp. 167–188.

[Al-Ars05] Z. Al-Ars, *DRAM Fault Analysis and Test Generation*, PhD thesis, Delft Univ. of Technology, Delft, the Netherlands, 2005, http://ce.et.tudelft.nl/~zaid/books/phd.html

[Al-Ars08] Z. Al-Ars, S. Hamdioui, A.J. van de Goor and Georg Mueller, "Defect Oriented Testing of the Strap Problem Under Process Variations in DRAMs," *in Proc. Int'l Test Conf.*, 2008.

[Azimane04] M. Azimane and A.K. Majhi, "New Test Methodology for Resistive Open Defect Detection in Memory Address Decoders," *in Proc. IEEE VLSI Test Symp.*, 2004, pp. 123–128.

[Falter00] T. Falter and D. Richter, "Overview of Status and Challenges of System Testing on Chip with Embedded DRAMs," *in Solid-State Electronics*, no. 44, 2000, pp. 761–766.

[Lin00] K.-J. Lin and C.-W. Wu, "Testing Content-Addressable Memories Using Functional Fault Models and March-Like Algorithms," *in IEEE Trans. Computer-Aided Design*, vol. 19, no. 5, 2000, pp. 577–588.

[Mohammad03] M.G. Mohammad and K.K. Saluja, "Simulating Program Disturb Faults in Flash Memories Using SPICE Compatible Electrical Model," *in IEEE Trans. Electron Devices*, vol. 50, no. 11, 2003, pp. 2286–2291.

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