Soft Faults and the Importance of Stresses in Memory Testing

Zaid Al-Ars Ad J. van de Goor
Delft University of Technology, Faculty of EE, Mathematics and CS
Laboratory of Computer Engineering, Mekelweg 4, 2628 CD Delft, The Netherlands
E-mail: z.e.al-ars@ewi.tudelft.nl

Abstract: *Memory testing in general, and DRAM testing in particular, has become greatly dependent on the modification of stresses (timing, temperature and voltages) in a way that is difficult to justify using the current understanding of memory faults. This paper introduces a new class of fault models (soft faults) based on a special classification of memory faults, that shows why it is fundamentally necessary to apply stresses. The paper calculates the relative probability of soft faults for a specific failure mechanism and compares this probability in DRAMs with that in SRAMs. In addition, the concept of soft faults is validated using defect injection and electrical simulation of a Spice DRAM model.*

Keywords: *Fault modeling, soft faults, memory testing, stress application, defect simulation.*

1 Introduction

Memory testing has seen a continued growth in complexity in the past decade, due to the need to detect many newly observed subtle types of faulty behavior. The extent of this increase in test complexity is most apparently reflected in the extensive employment of *stress (ST)* modifications, such as timing, temperature and voltages, to increase the effectiveness of test patterns. STs today are considered as an integral part of any modern industrial memory test, used either to ensure a higher fault coverage of a given test, or to target specific failure mechanisms not detected at nominal operational conditions [Falter00].

From a theoretical point of view, STs have always been considered necessary in the context of validating the specifications, where proper functionality of the memory is examined across the operational range specified in the data sheets [vdGoor98]. It has not yet been theoretically justified, however, why it would be *necessary* sometimes to apply far more stressful STs than the specifications allow, STs that bring the memory very close to total operational failure.

This paper defines a new class of memory faults, *soft faults*, based on a special classification of memory faults that identifies specific fault related voltage ranges within a faulty memory cell. The new fault class identifies the underlying fundamental reasons for using STs in memory testing, and makes it possible to analyze the types of STs needed for a specific defect.

Section 2 introduces a classification of memory faults and defines the concept of soft faults. To appreciate the importance of soft faults, Section 3 derives their relative probability in DRAMs and compares it with that in SRAMs. Section 4 discusses ways for detecting soft faults and describes the way they are related to STs. The concepts presented in the paper are then validated, using defect injection and electrical simulation of a memory model in Section 5. Finally, Section 6 ends with the conclusions.

2 Types of memory faults

In this section, basic aspects of DRAM operation are discussed first, followed by defining three different classes of DRAM faults.

2.1 Basic DRAM functionality

As shown in Figure 1(a), a DRAM cell consists of an access transistor controlled by a word line (WL), which connects a bit line (BL) with a cell capacitor. A DRAM cell stores its logic value in a leaky storage capacitor, a fact that results in the gradual loss of the stored charge in the capacitor. In order to store data in a DRAM for a long period of time, the stored data in the cell needs to be refreshed regularly to prevent the total depletion of stored information. There are many causes of leakage current, some pull the cell voltage up, while others pull the cell voltage down, so that the net voltage change within the cell is determined by the net effect of all active leakage mechanisms for a given memory cell [Keshavarzi97].

Directly after a write operation, the voltage in the capacitor should be set to a high (or a low) enough level that allows enough time before the stored voltage is completely destroyed by leakage. Therefore, it is possible to divide stored cell voltages into three different regions (see Figure 1(a)). The “1” region directly after performing a write 1 operation, the “0” region directly after performing a write 0, and the “faulty” region directly after performing a faulty write 1 or 0 operation. The logic 1 region extends from the

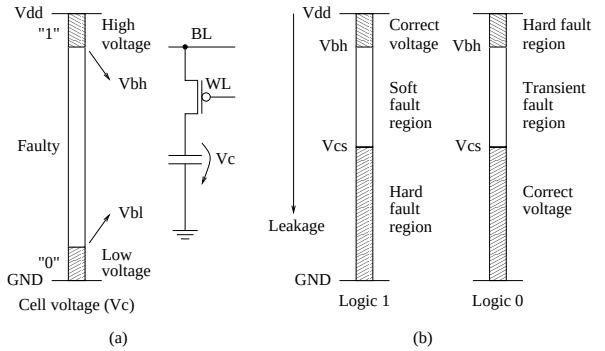

Figure 1. Definition of memory faults. (a) Basics of DRAM operation. (b) Voltage ranges in DRAM cells.

high power supply voltage (V_{dd}) to the border high voltage (V_{bh}), the logic 0 region extends from GND to the border low voltage (V_{bl}), while the faulty region takes on the voltages between V_{bh} and V_{bl}. For the remainder of this paper, we assume that in a defect free memory the following voltages are valid: $V_{dd} = 2.4$ V, $V_{bh} = 95\% \cdot V_{dd} = 2.28$ V, $V_{bl} = 5\% \cdot V_{dd} = 0.12$ V. In a defective memory, however, these values change according to the type and severity of the defect.

Assume that a defective memory cell has a net leakage current that pulls the cell voltage down, then the voltage ranges corresponding to a stored logic 0 and stored logic 1, directly after performing a write operation, are shown in Figure 1(b). The figure shows that, in addition to the region of proper operation, there are three faulty regions: the hard fault region, the soft fault region and the transient fault region. In the following each of the three types of faults is described in more detail.

2.2 Hard faults

Hard faults are memory faults that do not depend on time in any way, neither for sensitization nor for detection. Figure 1(b) shows the range of hard fault voltages directly after performing a write operation on a DRAM cell, having a net leakage current to GND. The voltage level V_{cs} is the *cell sense threshold voltage*, above which the sense amplifier detects a stored logic 1, and below which the sense amplifier detects a stored 0. Therefore, in the column representing logic 1 in the figure, any voltage below V_{cs} is considered as a hard fault voltage since it can be directly detected as 0 by a read operation. In the column representing logic 0 in the figure, the hard fault range is between V_{dd} and V_{bh}, since it takes more time than the refresh time to deplete the faulty charge in the cell to a detectable 0.

A more formal definition of hard faults can be given using the fault primitive (FP) notation used to describe faults in memory devices. Fault primitives are represented as $<S/F/R>$, where S is the sensitizing operation sequence that results in the fault, F is the logic level present in the faulty cell, and R is the read output in case S ends with a read operation [vdGoor00]. Since hard faults should not depend on time, neither S nor F should have a time parameter attached to it.

Figure 2 shows examples of memory defects that generate hard faults for DRAM and SRAM cells. For the DRAM cell, the open defect between the pass transistor and the cell capacitor restricts current flow to and from the cell and prevents write operations from changing the value stored in the cell, thereby causing an up-transition fault ($<0w1/0/->$) as well as a down-transition fault ($<1w0/1/->$) for high open resistance values [Al-Ars01a]. An up-transition fault means that a $0w1$ operation fails to flip the logic value stored in the cell from 0 to 1. The transition faults here represent hard faults.

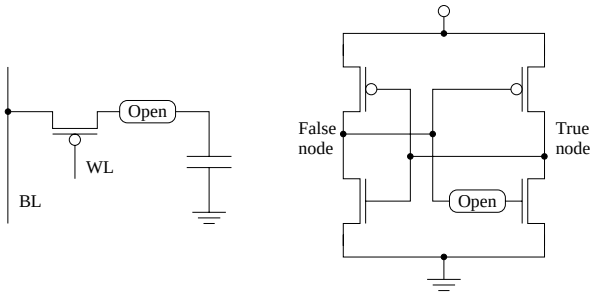

Figure 2. Opens in DRAM and SRAM cells causing hard as well as soft faults, each for a specific range of defect resistances.

For the SRAM cell, the open defect at the gate of the pull-down transistor causes a down-transition fault ($<1w0/1/->$) when the open resistance has a high value [Hamdioui02]. This type of fault is operation related (a write operation in this case) and will always be sensitized once the faulty operation is performed.

2.3 Soft faults

Soft faults are memory faults sensitized not only by the sequence of memory operations used in the test, but also by time. Figure 1(b) shows the region of soft fault voltages within a DRAM cell having a net leakage current to GND. With such leakage current, soft faults take place after a $w1$ operation that sets a cell voltage between V_{bh} and V_{cs}, since this faulty voltage sensitizes a fault when some time passes to deplete the voltage to a level below V_{cs}. In terms of the FP notation, soft faults are represented as $<S_T/F/R>$, where the sensitizing operation sequence has an added time parameter T to indicate that some time should first elapse before the fault effect is sensitized.

Figure 2 shows examples of memory opens that cause soft faults in both DRAM and SRAM cells. The DRAM

open shown in the figure is located along the data path of the cell, and for a given intermediate range of open resistances, write operations succeed but are only able to write a *weak voltage*. As time passes, a weakly written voltage is depleted gradually, thereby losing the stored information over time. Therefore, this defect causes what we may call *soft transition faults*, which can be written as $<0w1_T/0/->$ and $<1w0_T/1/->$.

The SRAM open shown in Figure 2 is located at the gate of the pull-down transistor, a defect position that (in combination with a low floating gate voltage) degrades the ability of GND to compensate for leakage currents through the pull-up transistors. Therefore, a stored 0 voltage on the defective side of the cell will be gradually degraded through leakage, until the cell flips and looses its stored voltage a while later. This faulty behavior is referred to as data retention fault, and has the FP notation $<0_T/1/->$ [Dekker90].

2.4 Transient faults

Transient faults are memory faults that do not remain sensitized indefinitely, but they tend to correct themselves after a period of time. Transient faults have already been treated in the literature [Al-Ars01b], and are tested for by performing a detecting read operation directly after sensitizing the transient fault. Figure 1(b) shows the range of transient fault voltages in a DRAM cell having a net leakage current to GND. With such leakage current, transient faults are only sensitized after a $w0$ operation that sets a cell voltage between V_{bh} and V_{cs}, since such a faulty voltage is automatically corrected by leakage after some idle time. In terms of the FP notation, transient faults are represented as $<S/F_L/R>$, where the faulty cell value F has an added time parameter L (life time) to indicate that these faults are time limited.

As an example of transient faults, consider the DRAM open shown in Figure 2, which forces write operations into setting a faulty voltage within the cell that is not strong enough to qualify as a hard fault. As time passes, a weakly written faulty voltage is depleted gradually, thereby correcting the faulty information over time. Therefore, the shown defect causes what we may call *transient transition faults*, which can be written as $<0w1/0_L/->$ and $<1w0/1_L/->$. Transient faults have not yet been validated for SRAMs.

3 Significance of soft faults

In this section, the probability of occurrence of soft faults is calculated for DRAMs and compared with that observed for SRAMs. In the case of SRAMs, the only type of soft fault known to take place is modeled as data retention faults, and therefore a calculation of the probability of data retention faults in SRAMs is equal to the probability of soft faults. A study of Intel's embedded SRAM caches analyzed the occurrence probability of common fault models for SRAMs, using spot defect injection and electrical simulation [Hamdioui02]. The study shows that, although soft faults (represented by data retention faults in SRAMs) are indeed observed in the faulty behavior, the probability that they actually take place is as low as 0.71% of all the faults observed in the faulty behavior of the SRAM.

Unfortunately, there is no study of DRAMs to identify the occurrence probability of soft faults, but it is possible to approximate this probability for a specific memory model having the DRAM cell open shown in Figure 3(a). The open resistance (R_{op}) in the figure consists of the track resistance (R_{tr}) of the conductive path within the cell, in addition to a resistive deviation (ΔR) related to the parameter distribution of the fabrication process. Figure 3(b) plots the probability density function (PDF) of ΔR, represented by a normal distribution with a mean value of $\mu = 0$ and a standard deviation of $\sigma = 5$ kΩ:

$$f_{\Delta R}(r) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-r^2}{2\sigma^2}}$$

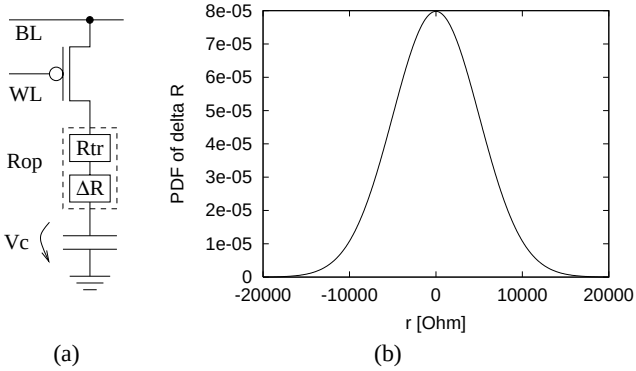

Figure 3. (a) Model of open defect. (b) PDF of defect.

Figure 4(a) shows the impact of the open defect on the cell voltage (V_c), where V_c is initialized to 0 V and then a sequence of $w1$ operations is performed. The curve $(1)w1$ plots the cell voltage after a single $w1$ operation as a function of ΔR, while the curve $(2)w1$ plots the voltage after a second $w1$. Note that with a small value of ΔR, the $w1$ operation behaves properly (same is true for negative ΔR)[1], and as the resistance increases, it becomes increasingly difficult to write a high voltage into the cell. Figure 4(a) also plots the V_{cs} curve (the cell sense threshold

[1]Negative ΔR values mean that the total resistive value of the track resistance is below what is expected in the fabrication process, in which case the cell continues to behave properly.

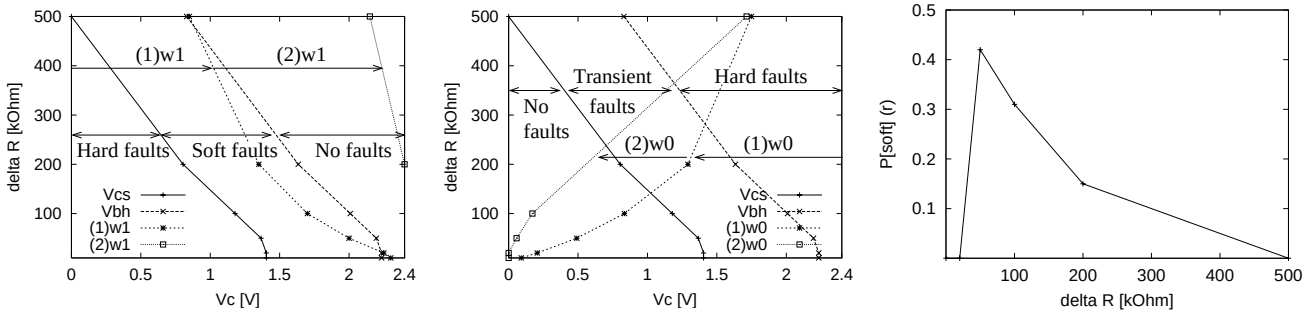

Figure 4. Impact of ΔR on V_c of (a) $w1$ and (b) $w0$. (c) Probability of soft faults as a function of ΔR.

voltage curve), which shows that as ΔR increases, it becomes increasingly difficult to sense a 0. In addition, the figure plots the V_{bh} curve (the border high voltage curve), which keeps a constant distance from the V_{cs} curve. Three voltage regions are identified in the figure, corresponding to hard faults, soft faults and no faults, where a net leakage current to GND is assumed in the cell (see Figure 1(b)). In a similar way, Figure 4(b) shows the impact of ΔR on the cell voltage, where V_c is initialized to $V_{dd} = 2.4$ V and then a sequence of $w0$ operations is performed. The figure identifies three voltage regions corresponding to hard faults, transient faults and no faults.

Using the information in Figures 4(a) and (b), it is possible to calculate the probability of hard, soft and transient faults as a function of ΔR after performing a single write operation, by assuming that V_c is uniformly distributed before performing the write operation. As an example, Figure 4(c) shows the probability of the soft faults region as a function of ΔR. As an example, for a ΔR of, say, 200 kΩ, the probability of the soft fault region is the sum of the probability of the region part to the left of $(1)w1$ where V_c = 1.4 V, plus the region part to the right of $(1)w1$. This is calculated as follows:

$$\begin{aligned} \Delta R \quad &= 200\text{k}\Omega \\ P[\text{soft}]|_{200\text{k}\Omega} &= P[V_{cs} < V_c < 1.4\text{V}] + P[1.4\text{V} < V_c < V_{bh}] \\ &= 0 \qquad\qquad + \tfrac{1.4\text{V}}{2.4\text{V}} \cdot \tfrac{V_{bh} - 1.4\text{V}}{V_{dd} - 1.4\text{V}} \\ &= 0.15 \end{aligned}$$

Figure 4(c) shows that the probability of soft faults is 0 for low values of ΔR, which makes sense since for low values of the open resistance, the memory should function properly without any faults. As the value of ΔR increases above 20 kΩ, the probability of soft faults increases rapidly to about 0.45 at ΔR = 50 kΩ, which is due to the gradual degradation of the $w1$ ability to write a proper high voltage. The probability of soft faults then decreases gradually toward 0 at ΔR = 500 kΩ and stays there for higher resistance values.

In a similar way, the probabilities of transient and hard faults can be calculated as a function of ΔR, which may then be used, in combination with the PDF of ΔR, to calculate the total probability of soft, transient and hard faults caused by an open within the cell. The calculation takes the form of compound probabilities according to the relation:

$$P[\text{fault}] = \int f_{\Delta R}(r) \cdot P[\text{fault}](r) dr$$

which results in $P[\text{soft}] \approx 1.97 \cdot 10^{-10}$, $P[\text{transient}] \approx 4.95 \cdot 10^{-150}$, and $P[\text{hard}] \approx 0$. These probabilities can finally be used to calculate the relative probability of a given type of fault, assuming that a fault does take place:

- $P[\text{soft}]_{\text{rel}} \approx 100\%$
- $P[\text{transient}]_{\text{rel}} \approx 0\%$
- $P[\text{hard}]_{\text{rel}} \approx 0\%$

These results closely represent DRAM fail count numbers acquired from industrial manufacturing data. The probabilities indicate that soft faults are by far the most probable type of faulty behavior, at least for cell opens in DRAMs with the simulated behavior of Figures 4(a) and (b). Comparing the high probability of soft faults for DRAMs (100%) with that for SRAMs (0.71%) highlights the fact that soft faults are much more important for DRAMs than they are for SRAMs.

4 Testing for soft faults

This section discusses how soft faults are usually tested for in SRAMs and in DRAMs.

4.1 DFT for SRAM testing

For some specific types of defects and with specific resistance ranges, SRAMs suffer from soft faults represented by data retention faults, which get sensitized by waiting for

some time. Traditionally, these defects have been tested for by tests that simply *wait* for some time after a 1 is written and after a 0 is written, before reading the written data. The following is one such test referred to as the IFA-9 scheme [Dekker90]:
IFA-9 = $\{\Updownarrow(w0);\ \Uparrow(r0,\ w1);\ \Uparrow(r1,\ w0);\ \Downarrow(r0,\ w1);\ \Downarrow(r1,\ w0);$ Del; $\Updownarrow(r0,\ w1);$ Del; $\Updownarrow(r1)\}$
where Del in the test stands for waiting for a specific amount of time, which is in the order of 100 milliseconds.

Since testing should take place on each and every memory component, and because there are typically hundreds of millions of memory components produced per month, test time is considered a major bottleneck for memory manufacturing. Therefore, soft faults are usually given special attention and solved by the introduction of DFT techniques. The DFT technique introduced to solve the problem of data retention faults in SRAMs is called "Weak Write Test Mode", where special "weak RAM write" circuits are used to weakly overwrite a previously written logic value, such that only defective cells are overwritten [Meixner96]. Subsequently, a read operation directly identifies the overwritten cells without the need to introduce any explicit delays into the test.

4.2 Stresses for DRAM testing

As discussed in Section 3, soft faults are much more important for DRAMs than they are for SRAMs, since they make up a large portion of the faults taking place in DRAMs. Soft faults in DRAMs are caused by almost any defect present in the memory, a fact that makes it costly to design a DFT solution for each and every possible DRAM defect. Therefore, DRAM tests today commonly resort to the modification of *stresses (STs)*, which generally include cycle time t_{cyc}, temperature T, and V_{dd}, to eliminate the soft fault problem [Vollrath00].

These STs can be used to bring a faulty memory closer to failure in order to induce a (directly detectable) hard or transient fault in a defective memory that exhibits soft faulty behavior. Figure 5(a) shows how the maximum voltage achievable by a $w1$ operation performed on a cell containing GND (V_{w1}) decreases gradually by decreasing the access time to the memory cell. When timing is significantly shortened by aggressively driving it below the boundaries set in the specifications, all cells (functional and defective) are brought closer to failure, and are actually forced to fail when $V_{w1} \leq V_{cs}$. Figure 5(b) shows a similar dependence on timing for $w0$ operations performed on cells containing V_{dd}, with the exception that here transient faults are forced and not hard faults.

When an ST is selected to eliminate the soft fault problem of a specific defect, it should satisfy two conditions.

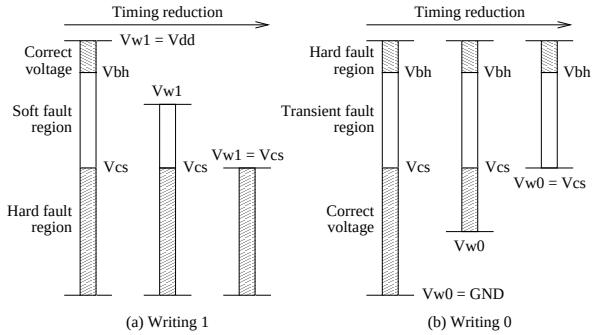

Figure 5. Eliminating soft faults by forcing a directly detectable fault using timing for (a) a $w1$ and (b) a $w0$ operation.

1) It is necessary that the ST is able to force a directly detectable fault in *any* defective cell, no matter how minor the defect is, which can only be achieved by bringing even functional cells to the verge of failure. STs able to achieve this for a given defect are called *decisive* STs for the considered defect, while STs that only influence, but do not force detectable faults are called *indecisive* STs for that defect.

2) The degradation in functionality induced by the decisive ST should take place gradually in such a way that badly defective cells are killed first and functional cells are killed last. These STs are referred to as *continuously* decisive STs.

It follows directly from this discussion that:

> It is sufficient to identify only *one* continuously decisive stress for a given defect in order to eliminate soft faults from the faulty behavior of that defect.

There are practical difficulties, however, in applying STs to eliminate soft faults in DRAMs. The most serious of which is related to the amount of degradation required to eliminate the soft fault problem. Since the fabrication process of integrated circuits is not a perfect one, the characteristics of functional memories represent a statistical distribution around an ideal norm, which makes it practically impossible to identify an exact border that separates functional components from faulty ones. Therefore, great care should be taken when defining the values for decisive STs in order to prevent cutting into the range of functional components. In the industry, it is commonly accepted to eliminate some functional (but weak) components in order ensure the high quality of the memories delivered to the customer [Vollrath00].

5 Validation by simulation

In this section, electrical Spice-based simulations are performed on a number of memory cell defects to validate the concepts presented above.

5.1 Simulation methodology

The used electrical simulation model is a reduced design-validation model of a real DRAM manufactured in 0.35 μm technology, with one folded bit line pair (2×2 memory cells, 2 reference cells, precharge devices and a sense amplifier), one write driver and one data output buffer. The used simulation tool is a Siemens/Infineon in-house electrical Spice-based simulator.

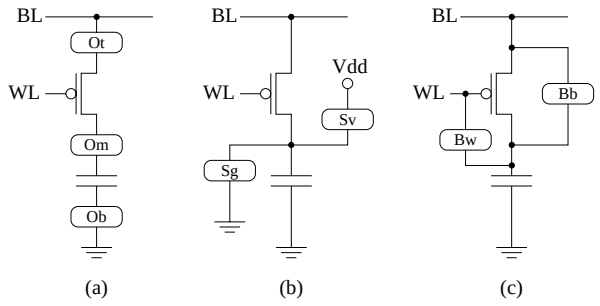

Figure 6. Simulated cell defects: (a) opens, (b) shorts and (c) bridges.

Figure 6 shows the 7 analyzed defects: 3 opens, 2 shorts and 2 bridges. For opens within the cell (Ot, Om, Ob) and the short to GND (Sg), the net effect of leakage is considered to pull V_c down toward GND. The short to V_{dd} (Sv) is considered to cause a net V_c leakage toward V_{dd}. Bridges are considered to cause a net V_c leakage toward the idle voltage of the aggressor node, which means that both word line bridges (Bw) and bit line bridges (Bb) pull V_c high.

5.2 Simulation results

Table 1 summarizes the simulation results. The first column lists the analyzed defects as shown in Figure 6, the second column shows whether the defects cause soft faults or not, while the next three columns indicate whether each ST can be used as decisive STs for the analyzed defects.

Table 1. Effectiveness of different STs on defects shown in Figure 6.

Defect	Soft faults	Timing	Temperature	Voltage
Ot,m,b	yes	decisive	indecisive	indecisive
Sg	yes	decisive	indecisive	decisive
Sv	yes	decisive	indecisive	decisive
Bw	yes	decisive	indecisive	decisive
Bb	yes	decisive	indecisive	decisive

Table 1 shows that all simulated defects result in soft faults. The table also shows that timing can be used as a decisive ST for all analyzed defects. This is due to the strong control timing has on the maximum voltage written in the cell. This is also due to the limited impact timing has on V_{cs}, which means that, to a large extent, timing follows the behavior symbolized in Figure 5.

Temperature, on the other hand, is considered as an indecisive ST for all simulated defects. This can be attributed to the limited control temperature has on the short-term behavior of the memory in general, and on the behavior of write operations in particular. Temperature has a huge impact, though, on the long-term behavior by controlling the amount of leakage current into the cell. Therefore, temperature should mainly be used in data retention testing to validate proper functionality according to the specifications.

Voltage represents a decisive ST for some defects, but it is indecisive for others. This can be explained by the fact that voltage has large influence on both the maximum voltage of write operations and on the V_{cs} curve. This means that voltage has a complex impact on the internal behavior of the memory, thereby making it a rather tricky ST to use and apply in testing.

6 Conclusions

This paper presented a new classification of DRAM faults, that identifies the new class of soft faults. It is shown that soft faults are the reason that makes stress application a necessary part of memory testing. The importance of soft faults is validated by calculating their probability for a specific defect. In addition, ways to detect soft faults in DRAMs have been presented and evaluated using a simulation-based analysis of a number of DRAM defects.

References

[Al-Ars01a] Z. Al-Ars and A.J. van de Goor, "Static and Dynamic Behavior of Memory Cell Array Opens and Shorts in Embedded DRAMs," *in Proc. Design, Automation and Test in Europe*, 2001, pp. 496–503.

[Al-Ars01b] Z. Al-Ars and A.J. van de Goor, "Transient Faults in DRAMs: Concept, Analysis and Impact on Tests," *in Proc. IEEE Int'l Workshop on Memory Technology, Design and Testing*, 2001, pp. 59–64.

[Dekker90] R. Dekker *et al.*, "A Realistic Fault Model and Test Algorithms for Static Random Access Memories," *in IEEE Trans. on Computer-Aided Design*, vol. C-9, no. 6, 1990, pp. 567–572.

[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.

[Hamdioui02] S. Hamdioui and A.J. van de Goor, "Efficient Tests for Realistic Faults in Dual-Port Memories," *in IEEE Trans. on Computers*, vol. 51, no. 5, 2002, pp. 460–473.

[Keshavarzi97] A. Keshavarzi, K. Roy and C.F. Hawkins, "Intrinsic Leakage in Low Power Deep Submicron CMOS ICs," *in Proc. IEEE Int'l Test Conf.*, 1997, pp. 146–155.

[Meixner96] A. Meixner and J. Banik, "Weak Write Test Mode: An SRAM Cell Stability Design for Test Technique," *in Proc. Int'l Test Conf.*, 1996, pp. 309–318.

[vdGoor98] A.J. van de Goor, *Testing Semiconductor Memories, Theory and Practice*, ComTex Publishing, Gouda, The Netherlands, 1998, http://ce.et.tudelft.nl/-vdgoor/

[vdGoor00] A.J. van de Goor and Z. Al-Ars, "Functional Memory Faults: A Formal Notation and a Taxonomy," *in Proc. IEEE VLSI Test Symp.*, 2000, pp. 281–289.

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