A Two-Layer SPICE Model of the ATMEL TSTAC eFlash Memory Technology for Defect Injection and Faulty Behavior Prediction

Pierre-Didier Mauroux, Arnaud Virazel, Alberto Bosio, Luigi Dilillo, Patrick Girard, Serge Pravossoudovitch, Benoît Godard, Gilles Festes, Laurent Vachez

A Two-Layer SPICE Model of the ATMEL TSTAC™ eFlash Memory Technology for Defect Injection and Faulty Behavior Prediction

P.-D. Mauroux A. Virazel A. Bosio L. Dilillo
P. Girard S. Pravossoudovitch
LIRMM – University of Montpellier / CNRS
Montpellier, France
{mauroux, virazel, bosio, dilillo, girard, pravo}@lirmm.fr

B. Godard G. Festes L. Vachez
ATMEL
Rousset, France
{benoit.godard, gilles.festes, laurent.vachez}@atmel.com

***Abstract*—Flash memories are based on the floating gate technology allowing the write and erase data electronically. Such a technology can be prone to complex defects leading to faulty behaviors. In this paper, we introduce an electrical model of the ATMEL TSTACTM eFlash memory technology. The model is composed of two layers: a functional layer representing the floating gate and a programming layer able to determine the channel voltage level controlling the Fowler-Nordheim tunneling effect. The proposed model has been validated by means of simulations and comparisons with ATMEL silicon data. We apply this model for the analysis of defect-induced failures. As a case study, a resistive defect injection is considered.**

Keywords-embedded Flash; electrical model; coupling effects; Fowler-Nordheim; defects; fault modeling.

I. INTRODUCTION

System-on-Chip technologies allow embedding in a single chip a large number of memories from different types, sizes, access protocols and timing. The increased use of portable electronic devices such as mobile phones and digital camera produces a high demand for non-volatile memories. Flash memories are non-volatile memories based on the floating-gate concept that allows to electronically write and erase memory data [1, 2]. Moreover, their low-power consumption and high density make them popular for portable applications.

The high integration density of eFlash memories and their particular manufacturing process steps make them prone to defects. Moreover, as high electric field is required for its memorization principle, eFlash may be subject to complex disturbance phenomena [3, 4, 5].

Two types of array can be used to realize a Flash memory; NOR and NAND-based arrays [6]. Up to now, the NOR-based array was preferred to build embedded Flash memories for performance reason, especially, the short read access time. Meaningful studies related to electrical simulation models and defect injection for NOR-based array can be found in the literature [7, 8, 9].

Nowadays, the demand on embedded non-volatile memories makes that designers develop new concepts allowing to build eFlash with high performances and high integration density. In this context, we are using the ATMEL TSTAC™ eFlash memory technology.

Like NOR based eFlash, TSTAC™ eFlash memories can also be prone to defects as they still use floating gate transistors and high voltages for charge injection and removal during programming modes. In [10], we have performed a qualitative analysis of actual hard (open and short) defects that may occur in the TSTAC™ eFlash memory array. This analysis has shown that observed faulty behaviors could be modeled by classical fault models such as stuck-at faults (SAF0 and SAF1) and state coupling fault (CFst).

In order to develop an efficient tests for TSTAC™ eFlash memories, other type of defects have to be considered such as resistive and coupling defects. For this set of defects, electrical simulations become mandatory to analyze the possible resulting faulty behaviors of TSTAC™ eFlash under read, write and erase operations. During read operations, the sense amplifier measures the current provided by the core-cell according to its logic state. In this case, electrical simulations can be easily done by means of SPICE-like description with an appropriate core-cell set-up and voltage levels on eFlash array nodes. The case of write and erase operations is more difficult to simulate than the read operation. Here, many physical phenomena to program the core-cell have to be considered (*i.e.* Fowler-Nordheim and coupling effects), since the two operations change the core-cell state (from a logic '0' to a logic '1' and vice versa).

In this paper, we propose an electrical model of the TSTAC™ eFlash core-cell. The proposed model is a two-layer model: a functional layer and a programming layer Compared to the existing Flash model given in [11], we have developed a SPICE model of the TSTAC™ eFlash core-cell, which is fully dynamic, *i.e.* it allows read, write and erase operations. Next, we describe in detail how we build this model and we compare simulation results with silicon data measured on an actual ATMEL technology. Then, we present a first analysis of a resistive defect impact in a TSTAC™ eFlash array based on the proposed model. This analysis highlights the interest of the proposed model to provide a realistic set of fault models that has to be tested, thus enhancing existing solutions for eFlash testing.

978-1-4244-5833-2/10/$26.00 ©2010 IEEE

The rest of the paper is organized as follows. Section 2 describes the TSTAC™ eFlash and gives the basic operations we can perform. Section 3 presents the proposed electrical TSTAC™ core-cell model. Section 4 gives a first application of the model for resistive defect analysis in the TSTAC™ eFlash array. Finally, in Section 5 conclusion and perspectives are given.

II. TSTAC™ eFLASH DESCRIPTION

The functional scheme of an eFlash memory is presented in Figure 1. Like other memories, the eFlash is composed of a core-cell array (CORE), data latches (DLATCH), bit-line / word-line decoders and sense amplifiers. In addition, eFlash memories need two particular building blocks to perform dedicated functions: a Charge Pump device for the High Voltage Generation (HVG) allowing the write and erase operations, and a Sense Reference Voltage used during the read operation.

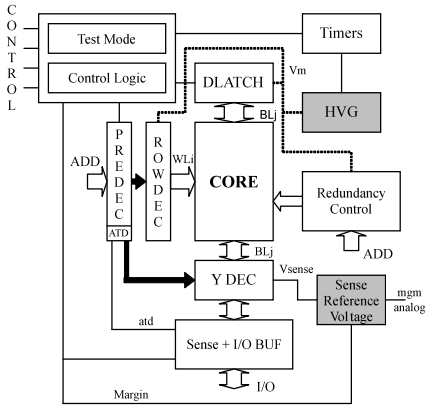

Figure 1. Functional scheme of eFlash memories

The TSTAC™ Flash array is composed of strings as shown in Figure 2. Each string contains some core-cells that are selected through the word-line signal (WL) and the activation of the select gate transistors (SG_1 and SG_2). For the sake of simplicity, the bulk of each transistor is not represented in Figure 2 and we consider only one core-cell in the string.

The TSTAC™ eFlash core-cell is based on the floating gate (FG) concept. There are two typical mechanisms to transfer electric charges from and into the FG: hot carrier injection (HCI) [2, 6] and the Fowler-Nordheim (FN) tunneling effect [2, 6]. The FN tunneling effect is used for charge injection or removal in the TSTAC™ eFlash FG.

Three different operations can be performed on a TSTAC™ eFlash: erase, write and read.

The erase operation consists in removing charges from the FG. A high voltage is applied to the bulk of the FG-transistor while its gate must be maintained at ground. All the other nodes are set to HZ (high impedance). The erase operation is performed simultaneously on all core-cells within the same string and the same page and not on a single core-cell for structural reason. At the end of the erase operation, the Vt of the FG-transistor is a low Vt denoted VTL. From a functional point of view, VTL corresponds to logic '1'.

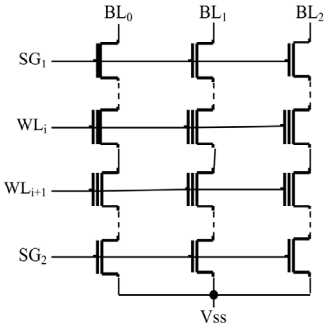

Figure 2. TSTAC™ eFlash structure

The write operation consists in injecting electrons to the FG by putting the gate of the FG-transistor at high voltage while maintaining its bulk at ground. Specific voltage levels are applied on the other nodes to allow the write on a unique core-cell without affecting other core-cells of the same word-line. More details will be given in Section 3. Under this operation, charges injected into the FG change the Vt from a low to a high Vt (VTH) that corresponds to logic '0'.

For the read operation, the core-cell is selected and a sense amplifier, working in current measurement mode, performs the data sensing. During the read operation, if the FG-transistor has a VTL, a current passes through the bit-line (between 5µA and 30µA) and the sense amplifier provides a logic '1' on its output. Otherwise, if the FG-transistor has a VTH, there is no current through the bit-line and the sense amplifier returns a logic '0'.

III. THE TWO-LAYER TSTAC™ eFLASH MODEL

In our previous work [9], we have developed a SPICE-like simulation model targeting NOR eFlash memories. This model is a dynamic model, i.e. it allows representing the erase and write operations. Here after, we introduce a novel model for the TSTAC™ eFlash. The main difference with our previous work is that the TSTAC™ eFlash model is composed of two layers: a functional layer and a programming layer. The functional layer is similar to the NOR eFlash model and it allows representing the FG-transistor. The programming layer allows evaluations of the channel voltage level that controls the electric field of the Fowler-Nordheim tunneling effect.

In the following sub-sections, we first present the principles of the model. Then, we provide electrical simulation results that are compared with silicon data measurements in order to validate the pertinence of the proposed electrical model.

A. Model description

To develop our model, we consider one FG-transistor (*i.e.* one word-line) per bit-line. Generalization of the model to the case of multiple FG-transistors per bit-line (*i.e.* the real case of multiple word-lines) will be discussed in Section 3.3.

A specific device from ATMEL technology was used to model the two select transistors. Concerning the FG-transistor, the description was more complex due to particular coupling effects and the Fowler-Nordheim tunneling effect phenomenon to be considered.

To have a good understanding of coupling effects, Figure 3 represents the different capacitances of the FG-transistor as proposed in [12]. C_{ONO} and C_{OX} are the ONO (Oxide-Nitrite-Oxide) capacitance and the tunnel oxide capacitance respectively. C_{CHAN} is the total channel capacitance, which is the sum of the junction capacitance between the inversion layer / p-well and that between the source / drain diffusion regions and p-well.

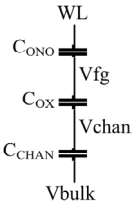

Figure 3. Coupling effects in the FG-transistor

Now, we describe the charge injection or removal mechanisms in the floating gate. The erase and write operations are regulated by the Fowler-Nordheim tunneling effect that modulates the charge quantity in the floating gate and thus the threshold voltage value of the FG-transistor. As in [9], we represent the floating gate by a capacitive charge that varies with the Fowler-Nordheim tunneling effect. Following the law Q = C * U (which is the law defining the electrical charge 'Q' contained in a capacitor 'C' under a potential 'U'), the voltage of this capacitance is proportional to the charges injected or removed. From a physical point of view, the equivalent capacitance value (C_{TOT}) represents the total capacitance of the FG-transistor.

We have seen in the previous section that the charge stored in the floating gate impacts the Vt value of the FG-transistor. When the core-cell is erased, Vt is low (VTL), and when the core-cell is written, Vt is high (VTH). Thus, to build the model, we need a transistor whose Vt value changes according to the quantity of charges stored in C_{TOT}. This principle can easily be implemented in an electrical SPICE-like model.

The Fowler-Nordheim tunneling effect is represented by the following equation [13]:

$$I_{FN} = A \times \alpha \times E_{ox}{}^{2} \times \exp\left(\frac{-\beta}{E_{ox}}\right) \qquad (1)$$

with:

A = Tunnel area
α = Fowler Nordheim constant
β = Fowler Nordheim constant
E_{ox} = Oxide electric field

The tunneling effect is equivalent to a current source controlled by an electric field (E_{ox}). This electric field is due to the voltage between the floating gate (Vfg) and the channel (Vchan). In a SPICE simulator, we can describe such a current source if we know all the parameters of Equation 1. From ATMEL silicon measurements, we have extracted all these parameters; A, α, β and the oxide thickness of the tunnel window used to calculate E_{ox}.

With all these elements, the proposed SPICE model has been implemented and its functional layer is presented in Figure 4. Note that, this layer represents one TSTAC™ eFlash BL with only one FG-transistor. The blocks Ki represent the coupling factors due to the different capacitances with:

$$K_G = \frac{C_{ONO}}{C_{TOT}} \text{ and } K_C = \frac{C_{OX}}{C_{TOT}} \qquad (2)$$

The other coupling factors do not contribute significantly to the floating gate voltage due to their low values.

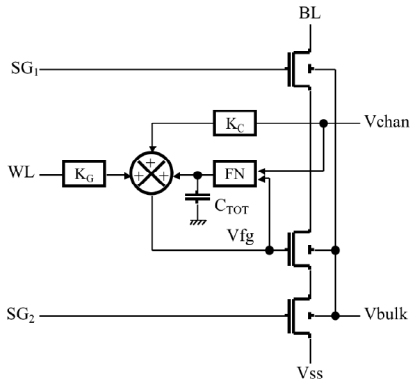

Figure 4. Functional layer

The capacitance C_{TOT} is used to store charges provided by the block *FN* representing the Fowler-Nordheim tunneling effect. This block is implemented as a current source in SPICE. We sum these three effects (block *Sum*) to control the gate of a nmos transistor representing the FG-transistor. The voltage Vfg represents the equivalent floating gate voltage (*i.e.* the threshold voltage) of a core-cell under the Fowler-Nordheim effect and capacitive coupling effects.

Now, we have to determine the voltage channel level (Vchan) in order to control the electric field for the erase and write operations. For this purpose, we use the capacitive view of the FG-transistor shown in Figure 3 and we add the two select gates to modulate the Vchan level. Figure 5 presents what we call the programming layer, which is able to produce the required Vchan level for the erase and write operations. As for the functional layer, this programming layer represents only one TSTAC™ eFlash BL with only one FG-transistor.

Now, let us analyze the function of the programming layer. As already mentioned, the erase operation consists in applying a voltage to the bulk of the FG-transistor and the ground to the WL node. As the C_{CHAN} capacitance is low compared to the others, the high voltage is also present on the Vchan node. Consequently, the electric field is high enough to remove charges from the FG.

During the write operation, the two select transistors are helpful to allow the write on a unique FG-transistor without disturbances on the other FG-transistors of the same word-line.

Let us assume that we act a write operation on the FG-transistor of the first bit-line. Specific voltage levels applied on SG_1 and SG_2 make T_1 'ON' and T_2 'OFF'. In addition, a high voltage is applied on WL, while the ground is applied to the bulk and BL. Consequently, Vchan is at ground while Vfg, due to the ONO coupling effect, is high enough to create the required electric field for charge injection from the channel to the FG.

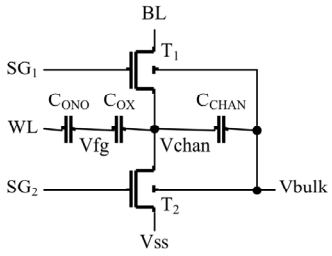

Figure 5. Programming layer

As in the same row all FG-transistors share the same WL signal, the high voltage is also applied on FG-transistors of the other bit-lines where the write operation must not be acted. Consequently, BL levels on non-selected bit-lines have to set all T1 transistors in the 'OFF' state. Then, as Vchan levels increase due to the capacitive bridge, the electric field is not enough important to inject charges into the FG. Consequently, there is no write disturb on the other FG-transistors connected to the selected word-line.

B. Experimental validation of the model

To validate the proposed two-layer model of TSTAC™ eFlash memories, we have performed simulations on the hypothetical 1x2 (1 word-line, 2 bit-lines) memory array shown in Figure 6. At SPICE level, both bit-lines are built with their own functional and programming layers.

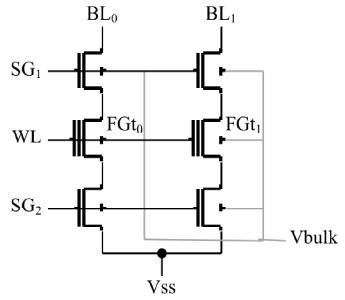

Figure 6. A 1x2 TSTAC™ eFlash memory array

Simulation results of erase, write and read operations are presented in Figure 7. Note that, all voltage levels and times have been removed for confidential reasons.

Simulations start with an erase operation on both FGt0 and FGt1. The high voltage is applied to the bulk. Both Vchan0 and Vchan1 are increasing and the resulting electric fields remove charges from the FG of FGt0 and FGt1. At the end of the erase operation, both FGt0 and FGt1 have a low Vt (VTL).

Then both FGt0 and FGt1 are read. During each read operation, a current is passing through the bit-line that corresponds to a logic '1.

The next operation consists in writing the FGt0 only. From waveforms of Figure 7, we can observe that Vchan0 (the channel voltage level of the FGt0) is 0v. Consequently, charges are injected into the FG and the Vt of FGt0 has changed from VTL to VTH. On the other hand, Vchan1 increases and is important enough to do not disturb FGt1 that remains at a low Vt (VTL).

Finally, both FGt0 and FGt1 are read again. The simulation has confirmed that no current passes through FGt0 when it is written.

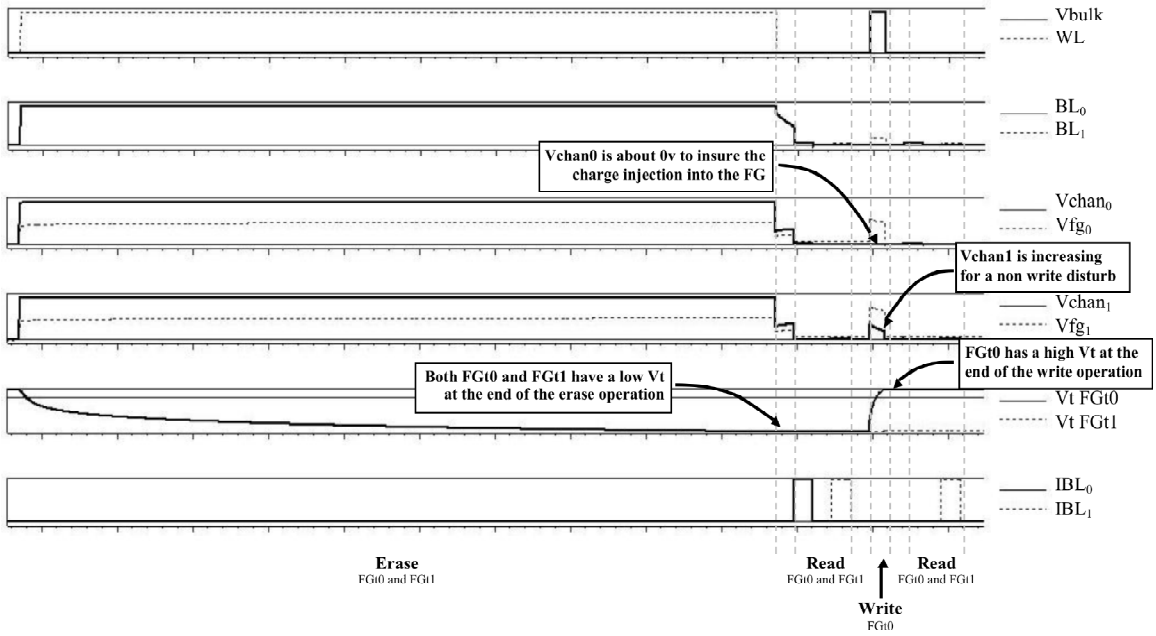

Figure 7. Electrical simulations of Erase, Write and Read operations

After these simulations, we now present some comparisons with silicon data obtained on the ATMEL technology. We analyze the VTL and VTH values after the erase and write operations for different programming times. Table 1 and Table 2 give a representative set of these comparisons between ATMEL Silicon data and those obtained with the proposed model.

In Table 1, we have reported the Vt window (VTW) comparisons for different values of the programming times (erase and write operations). The first column gives the different programming time conditions. The two next columns present the VTW of the proposed model (VTWm) and the VTW measured on silicon (VTWd). Finally, the last column indicates the deviation of the model. Note that, the grey line corresponds to the typical programming time. This first set of comparisons demonstrates the accuracy of the proposed SPICE model since the average error is about 5.2%.

TABLE I. VT WINDOW COMPARISONS

	Vt window (V)		**Error**
Timing Condition	SPICE model VTWm	Silicon Data VTWd	$\frac{VTWm - VTWd}{VTWd}$ (%)
-300%	0.811	0.886	8.5%
-100%	1.80	1.70	5.8%
T_T	**2.78**	**2.68**	**3.5%**
+60%	3.805	3,930	3.2%
+80%	4.550	4.795	5.1%

A second set of comparisons has been done to verify the accuracy of Vt values (VTL and VTH). Table 2 shows the differences of the observed Vt on silicon data and the achieved Vt with the model. The first column gives the different programming time conditions. The next column presents the delta (in %) between the VTL achieved with the model (VTLm) and the VTL observed on silicon data (VTLd). The last column gives the VTH comparisons.

TABLE II. VTL AND VTH COMPARISONS

Timing Condition	**Δ (VTLm, VTLd)**	**Δ (VTHm, VTHd)**
-300%	3.1%	2.9%
-100%	5.5%	1.3%
T_T	**2.9%**	**6.6%**
60%	0.2%	10.3%
80%	0.7%	12.5%

As for VTW comparisons, Vt comparisons demonstrate the effectiveness of the proposed model, as the average error is about 6.7%.

C. Generalization of the model

The proposed model has been developed for a one core-cell per string TSTAC™ eFlash. Generally, there are many core-cell in a string (2, 4, 8, etc.). Consequently, our model has to be scalable to match with actual TSTAC™ eFlash memories.

The scalability of the functional layer is easily done by implementing as many equivalent FG-transistors (*i.e.* K_G, K_C, FN current generator, etc.) as needed. The scalability of the programming layer is done by adding, in parallel, coupling capacitances of FG-transistors and adjusting the C_{CHAN} value (eqC_{CHAN}) as shown in Figure 8.

Experimental validations of the generalization of our model will be part of our future work.

IV. CASE STUDY: RESISTIVE DEFECTS IN THE ARRAY

To underline the interest of the proposed electrical model, we have simulated the impact of a resistive short defect between two adjacent bit-lines. Consider again the TSTAC™ eFlash array shown in Figure 6 where the defect is injected between BL_0 and BL_1 nodes. The impact of the resistive defect has been analyzed during erase, write and read operations on both FG-transistors (FGt0 and FGt1).

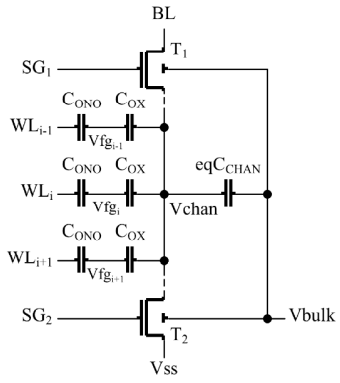

Figure 8. Scalability of the programming layer

During the erase operation, both bit-lines are at HZ. Consequently, the defect does not impact the erase operation. The proposed model has delivered a stable Vt value for FGt0 and FGt1 (VTH) whatever the defect size.

The defect impacts the write operation since bit-lines are not set to the same voltage level. Waveforms presented in Figure 9 are obtained with the proposed electrical model. Figure 9.a shows the Vt variation on both FG-transistors when a write operation is acted on FGt0. Note that, when the defect size decreases the write is not correctly acted on FGt0 since the voltage level on BL_0 increases. On the other hand, FGt1 has a Vt that increases as the defect make that the bad write is acted on both FG-transistors.

Finally, Figure 9.b presents the bit-line currents measure after a write operation on FGt0. It appears that, for a defect size higher than 100kΩ, the current that passes through BL_0 is lower than the threshold, and thus the sense amplifier will provide a logic '0'. When the size of the defect decreases, the current increases and is interpreted as a logic '1'. The current measured through BL_1 is always higher than the read current threshold, so FGt1 is always read as an erased core-cell (*i.e.* containing a logic '1'), and thus is considered as fault-free.

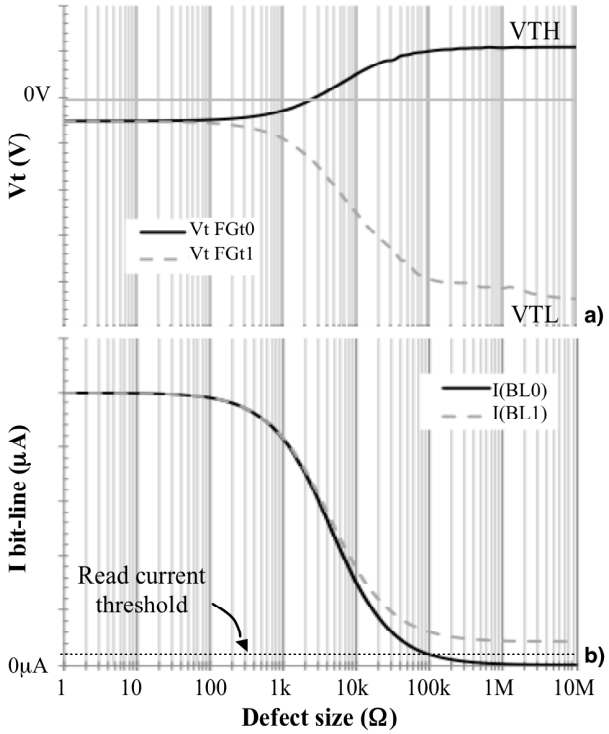

Figure 9. Simulation results of the defect injection
a) Vt values and b) bit-line currents

With the help of the electrical TSTAC™ eFlash model, we are able to determine the resulting faulty behavior of defect injection. The case study of the resistive short defect between two adjacent bit-lines can be modeled by a stuck-at-1 fault (SAF1) for a defect size lower than 100kΩ.

V. CONCLUSION AND FUTURE WORKS

In this paper, we have explored possibilities to create a simulation model of the TSTAC™ eFlash memory. We have detailed the different steps required to create this model and its basic scheme. It is made by two layers; a functioning layer and a programming layer. From measurements on the ATMEL technology, we have validated the accuracy of the proposed model in terms of Vt and current values during erase, write and read operations. Then, a typical resistive defect injection has been experienced and the resulting faulty behavior has been modeled from knowledge of the CMOS memory test literature.

In a future work, further investigations on any other kind of defects (resistive and coupling) will be done with the help of this simulation model. For each defect, we will analyze the faulty behavior to finally give a comprehensive set of fault models. Finally, with a complete eFlash electrical defect analysis, we will be able to enhance the existing eFlash test solutions.

REFERENCES

[1] S. Aritome, R. Shirota, G. Hemink, T. Endoh and F. Masuoka, "Reliability Issues of Flash Memory Cells", Proc. of the IEEE, vol. 81, N° 5, May 1993, pp. 776-788.

[2] P. Pavan, R. Bez, P. Olivo and E. Zanoni, "Flash Memory Cells – An Overview", Proc. of the IEEE, vol. 85, N° 8, August 1997, pp. 1248-1271.

[3] "IEEE Standard Definitions and Characterization of Floating-gate Semiconductor Arrays", IEEE 1005-1998, Revision of the IEEE std. 1005-1991.

[4] M. Mohammad and K. Saluja, "Flash Memory Disturbances: Modeling and Test", Proc. of the IEEE VLSI Test Symposium, pp. 218-224, 2001.

[5] O. Ginez, J.-M. Daga, P. Girard, C. Landrault, S. Pravossoudovitch and A. Virazel, "Retention and Reliability Problems in Embedded Flash Memories: Analysis and Test of Defective 2T-FLOTOX Tunnel Window", Proc of the IEEE VLSI Test Symposium, pp. 47-52, 2007.

[6] A.-K. Sharma, "Semiconductor Memories: Technology, Testing and Reliability", IEEE Press, Piscataway, 1997.

[7] M. Mohammad and K. Saluja, "Simulating Disturb Faults in Flash Memories Using SPICE Compatible Electrical Model", IEEE Trans. on Electron Devices, vol. 50, N° 11, November 2003, pp. 2286-2291.

[8] O. Ginez, J.-M. Daga, M. Combe, P. Girard, C. Landrault, S. Pravossoudovitch and A. Virazel, "An Overview of Failure Mechanisms in Embedded Flash Memories", Proc. of IEEE VLSI Test Symposium, pp. 108-113, 2006.

[9] O. Ginez, J.-M. Daga, P. Girard, C. Landrault, S. Pravossoudovitch and A. Virazel, "A SPICE-Like 2T-FLOTOX Core-Cell Model for Defect Injection and Faulty Behavior Prediction in eFlash", Journal of Electronic Testing - Theory and Applications, vol. 25, N° 2-3, Jun 2009, pp 127-144.

[10] P.-D. Mauroux, A. Virazel, A. Bosio, L. Dilillo, P. Girard, S. Pravossoudovitch and B. Godard, "NAND Flash Testing: A Preliminary Study on Actual Defects", Proc. of IEEE International Test Conference, poster 13, 2009.

[11] L. Larcher, A. Padovani, P. Pavan, I. Rimmaudo, A. Calderoni, G. Molteni, F. Gattel and P. Fantini, "Modeling NAND Flash Memories for Circuit Simulations", Simulation of Semiconductor Processes and Devices, vol. 12, September 2007, pp. 293-296.

[12] K. Takeuchi, S. Satoh, K. Imamiya and K. Sakui, "A Source-Line Programming Scheme for Low-Voltage Operation NAND Flash Memories for Circuit Simulations", IEEE Journal of Solid State Circuits, vol. 35, N° 5, May 2000, pp. 672-681.

[13] M. Lenzlinger and E. H. Snow, "Fowler-Nordheim Tunneling into Thermally Grown SiO2", Journal of Applied Physics, vol. 40, January 1969, pp. 278-283.