

Device and Memory array Models for Flash EEPROM Technology

Hassen AZIZA¹Bertrand DELSUC²

¹ IM2NP, Institut Materiaux Microelectronique Nanosciences de Provence, UMR 6242 CNRS,
University of Marseille

² ST-Microelectronics, ZI de Rousset BP 2, F-13106 Rousset
^{1&2} FRANCE

hassen.aziza@polytech.univ-mrs.fr

Abstract: - In this paper, three EEPROM memory cell models are presented. The first model is a compact model based on Mos Model 11 (MM11) and fully validated on silicon. From this first model we propose two alternative models (level 1 & level 2). These last models allow a reduction of simulation time and memory space overheads, with respect of accuracy, compared to the original compact model. The technique used to build the level 1 and level 2 models is based on the complexity reduction of the original EEPROM model. We also present simulation time results using the different models within memory arrays.

Key-Words: - EEPROM, Modeling, Memory arrays, Simulation time.

1 Introduction

Modeling and simulation of memory circuits remains an outstanding problem [1] especially when dealing with complex memories like EEPROMs.

In the Non Volatile Memory field, many EEPROM memory cell models have been presented for DC [2][3] and Transient [4][5] analysis. In [2], a charge sheet and a depletion approximation are used to describe the charge distribution in the semiconductor. In [3], the authors present a new compact SPICE model of floating gate non volatile memory cells. This model, simple and easy to implement since it uses SPICE circuit elements is capable to reproduce effectively the complete DC electrical behaviour in every bias conditions. In [4] and [5], theoretical and experimental analyses of Write/Erase characteristics of EEPROM cells are presented. Experimental investigation of the Fowler-Nordheim programming current is performed and results are included in a detailed cell model which greatly improves the understanding of the device physics.

Other models, often macro models are suitable for DC and transient simulations [6].

All these studies were done at a device level in order to understand and describe the EEPROM

memory behavior. However, no information was given about the behavior of a whole memory array composed of interconnected elementary cell models, particularly in terms of simulation time. Moreover, the model description never treat the way memory cells are interfaced to the peripheral circuits to meet designer's requirements. A. Pirola [7] presents a solution based on a VHDL synthesizable description, for modeling an analog Flash EEPROM array. This approach relies on available static RAM and ROM models and results in a finite state machine which describes the functionality of the memory array.

The proposed approach is fully analogical and targets electrical simulators. This paper presents, in a first time a new EEPROM memory model which is based on a physical description of the memory cell behavior. This model offers the capability to correctly simulate the behavior of the memory cell during programming operations at a circuit level. The validation of this model is done by comparing results of the simulation with experimental data obtained with EEPROM cells.

From this first model, two other models are developed for the I-V characteristics of the EEPROM cell under given operating conditions. These two models are obtained by varying the complexity of the initial model.

Section 2 gives an overview of the EEPROM technology. Section 3 presents the compact EEPROM model based on the Philips MM11. In section 4, two alternative models (level 1 and level 2) are presented. Section 5 is dedicated to simulation results. Finally, Section 6 gives some concluding remarks.

2 EEPROM Technology overview

The EEPROM architecture is globally similar to conventional RAM and consists in an array of memory cells identified by their row and column position. This array is surrounded by the following elements:

- A control logic and registers,
- A row decoder to select the active row,
- A column decoder to select the active column,
- Column high voltage latches and Row latches,
- A sense amplifier for the read operations.

However, the EEPROM architecture exhibits also some specificity directly due to the programming process. To perform data storing operations, the EEPROM includes an on-chip high voltage generator to provide high voltage levels during programming operations [8][9][10]. High voltage generators are a key component in flash EEPROM devices. Indeed, flash EEPROM devices require a high voltage pulse to erase any existing data in a particular memory cell before it can be written with a new value. The memory write and erase time depends on the maximum internally generated high voltage value which in turn depends on the technology used.

The global architecture of an EEPROM is depicted Fig.1. Typically, the memory array can be organized as a NOR or NAND array. In our case, the NOR organization is considered (Fig.2.a). The memory cell is composed of two transistors, a select transistor (ST) and a sense transistor, as illustrated Fig.2.b. The core element of the EEPROM memory cell is the sense transistor, which is a Floating Gate Transistor (FGT).

To have a memory cell that can commute from one state to the other and that can store the information independently of external conditions, the storing element needs to be a device whose conductivity can be changed in a non destructive way. One solution is to have a transistor with a threshold voltage that can change repetitively from a high to a low state, corresponding to the two

states of the memory cell, i.e., the binary values ('1' and '0') of the stored bit [11].

To do that, the threshold voltage V_T of a FGT can be shifted alternatively from a high to a low state corresponding to the two logical values of the memory cell. When a negative charge is stored in the floating gate, the V_T increases, the cell is then said erased and its logical value equals '0'. Respectively, when the charge in the floating gate is positive, the V_T decreases, the cell is then said written and stores a logical '1'.

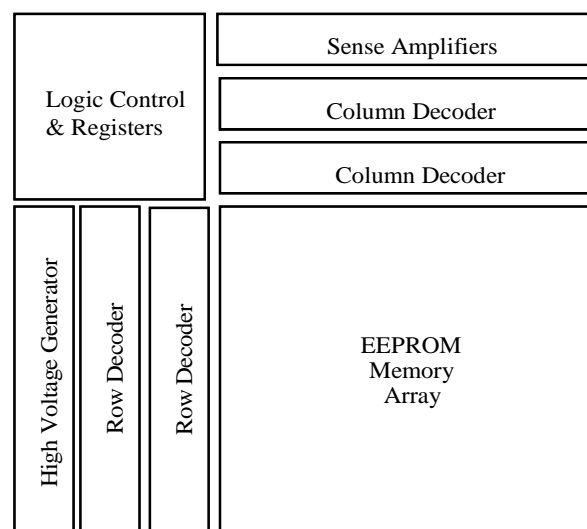


Fig.1 EEPROM architecture overview

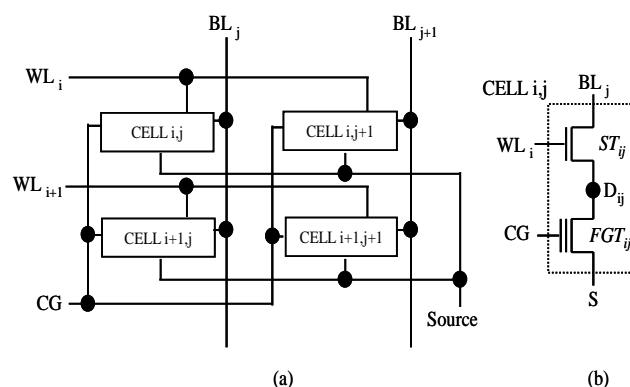


Fig.2 Memory array structure (a) and EEPROM cell structure (b)

In order to write a cell, a high voltage level called V_{pp} is applied on the bit line node (BL_i) of the selected cell. During an erase operation, the high voltage level is applied on control gate node (CG) of the selected cell. V_{pp} is provided by the high voltage generator. The schematic cross section of a generic floating-gate memory device is shown in Fig. 3: the upper gate is the control gate, while the lower one, completely surrounded by the

dielectric, is the floating gate. The basic concepts and the functionality of an FGT can be easily understood by determining the relationship between the floating gate potential, which physically controls the channel conductivity, the drain node and the control gate potential, both controlled by an external circuitry.

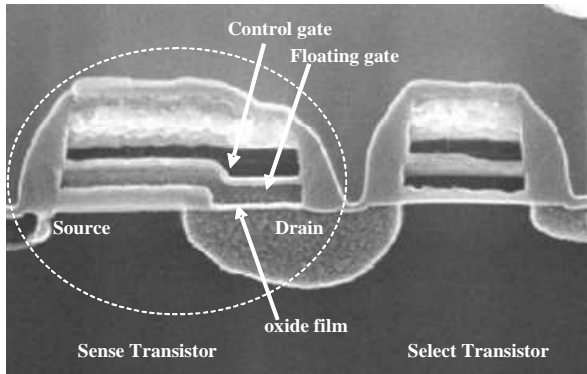


Fig.3 EEPROM cell cross section

There are many solutions used to transfer electric charge from and into the FG. For both erase and program, the problem is making the charge pass through a layer of insulating material (oxide film). In the EEPROM industry-standard “Flotox”, the charge transfer from the drain node to the floating gate node is due to Fowler-Nordheim tunneling [12]. This tunnelling mechanism is completely dependant of the geometry of the cell, leading to critical problems of process control. Indeed, any small variations of any geometric parameters (i.e. oxide film thickness, drain-floating gate overlap) among the cells in a memory array leads directly to a spread threshold voltage distribution, limiting the overall performance of the memory. The Fowler-Nordheim current is given by equation (1) where Q_{fgo} represents the initial charges trapped in the floating gate, α , β the Fowler-Nordheim parameters, $S_{tun} = T_{tun} \cdot W$ the surface of the oxide film and E_{tun} the electric field throw the oxide film.

$$I_{fn} = \alpha S_{tun} \cdot E_{tun}^2 \left[\exp\left(\frac{-\beta}{E_{tun}}\right) \right] \quad (1)$$

The tunnelling-injection mechanism is widely used in NVM, particularly in EEPROM. There are two main reasons for this choice: first, tunnelling is a pure electrical mechanism; second, the involved current level is quite low and thus allows the internal generation of supply voltages needed for

all operations; third it allows one to obtain the time to program ($< 1\text{ms}$) 12 orders of magnitude shorter than retention time ($> 10\text{years}$) which is a fundamental request for all NVM technologies [11].

3 EEPROM compact model

The compact EEPROM model is developed for DC and transient simulations. This model is surface potential dependant and takes advantage of the MM11 formulation. This model can be used during all the EEPROM basic operations: erase, write and read operations. The erase and write operation require a high voltage pulse to be applied to the source and drain of the memory cell in order to activate the FN tunnelling.

A simplified equivalent circuit of the EEPROM model is depicted Fig.4 where C_{pp} is the interpoly capacitance, V_{fg} the potential on the floating gate, V_{cg} the voltage applied on the control gate, C_{tun} the tunnel oxide capacitance and I_{fn} the Fowler-Nordheim current. The central part of this model is based on MM11 formulation. To build the EEPROM model, all the FGT transistor characteristics are added to the MM11 formulation. MM11 not only gives an accurate description of charges and currents and their first-order derivatives (transconductance, conductance, capacitances), but also of their higher order derivatives. It includes an accurate description of all physical effects important for modern and future technologies, such as gate tunneling current, influence of pocket implants, poly-depletion, quantum-mechanical effects and bias-dependent overlap capacitances [13]. These features make MM11 the ideal candidate to build the EEPROM model.

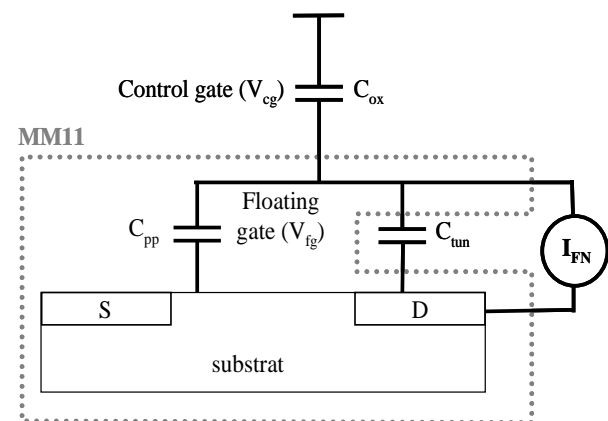


Fig.4 EEPROM equivalent circuit

Using this approach, the floating gate potential V_{fg} is computed by resolving an implicit equation.

3.1 Static Model (DC)

The Gauss law applied to the surface including the floating gate (Fig. 4) allows determining the charges Q_{fg} trapped on the floating gate by using the following equation:

$$Q_{fg} = C_{pp} \cdot (V_{fg} - V_{cg}) + C_{tun} \cdot (V_{fg} - V_d) + Q_g + Q_{fgo} \quad (2)$$

The charge Q_{fgo} represents, when necessary, the initial charges trapped on the floating gate. The charge Q_g is the MM11 gate charge computed by using the Mos Model 11 charge model. Q_{fgo} and Q_g are computed knowing a set of variables and more specifically the variable V_{fg} which is the floating gate potential.

Equation (2) is an implicit equation of the variable V_{fg} . V_{fg} is obtained for all possible potentials applied on the EEPROM cell nodes by solving equation (2). Then, knowing the potential V_{fg} , all the variables of the model are given by changing in the MM11 formulation the potential of the gate with the V_{fg} potential.

3.2 Dynamic Model (Transient)

The non-quasi-static approximation is used to develop the intrinsic dynamic model. This approach is based on the charges' neutrality in the structure [14][15]. The floating gate potential is computed implicitly using the following equation (in which the Fowler-Nordheim current I_{fn} given by (1))

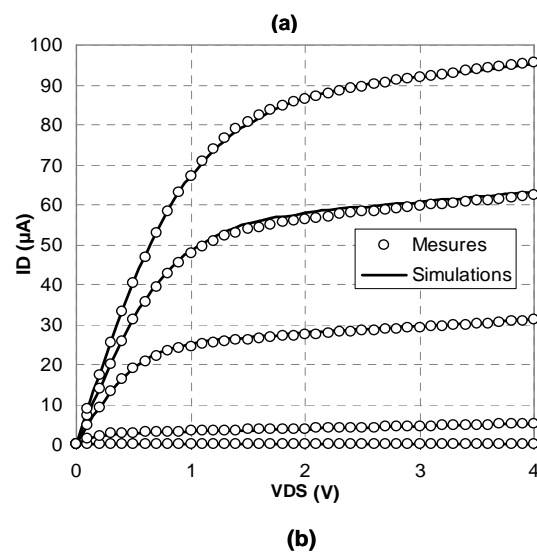
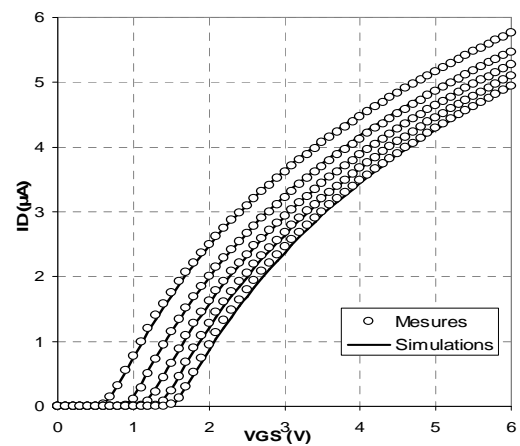
$$0 = Q_{fg} - Q_{fgo} - \int_0^t I_{fn} dt \quad (3)$$

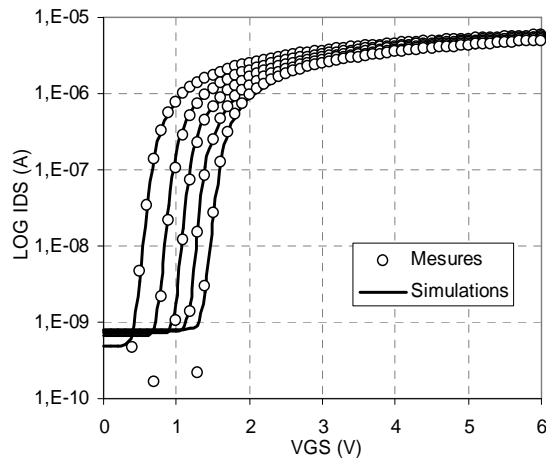
3.3 EEPROM Model

To obtain a full EEPROM cell model, a select transistor is associated to the sense transistor model. The select transistor is modeled with a classical MM11 high voltage transistor (provided by the design kit). It is important to notice that the sense transistor floating gate potential computation leads to an increase of simulation time. Model parameters of the sense transistor have been extracted on a 0,18 μ m EEPROM technology. Fig.5a.b.c.d.e presents I-V silicon results. The silicon validation of the model shows that EEPROM cell measurements fit the model behavior for $I_{Ds}(V_{Gs})$, $I_{Ds}(V_{DS})$, $\log I_{Ds}(V_{Gs})$ the

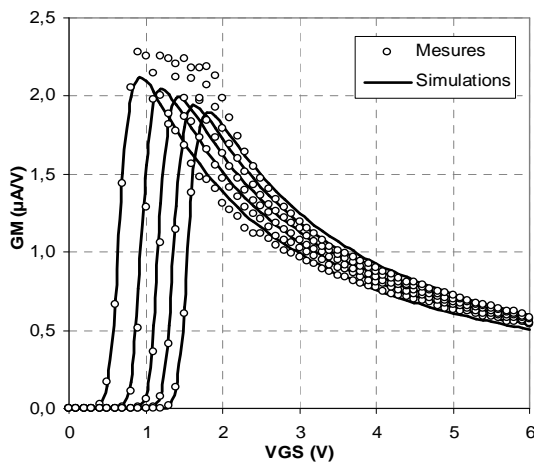
transconductance and the conductance characteristics. Concerning the extraction of the Fowler-Nordheim parameters (α, β), the method based on the $C(V)$ and $I(V)$ measurements presented in [16] is used.

Advantages of such a compact model are obvious at a device level. The memory model characteristics include V_T changes, drain current, tunnelling current and take advantage of the robustness of the MM11 equations. Indeed, this model is used as a predictive model when scaling down the cell dimensions or changing the operating conditions. Drawbacks appear at a circuit level when using the compact EEPROM model to build memory arrays: it leads to huge simulation time. This limitation avoids the use of this kind of model to build memory arrays and studies like memory fault simulation can not be considered. For instance, impact of a defective isolated cell on its neighbours, as well as the impact of a defective cell on the peripheral circuitry or the impact of a resistive Bit Line on the cell characteristics can not be considered.

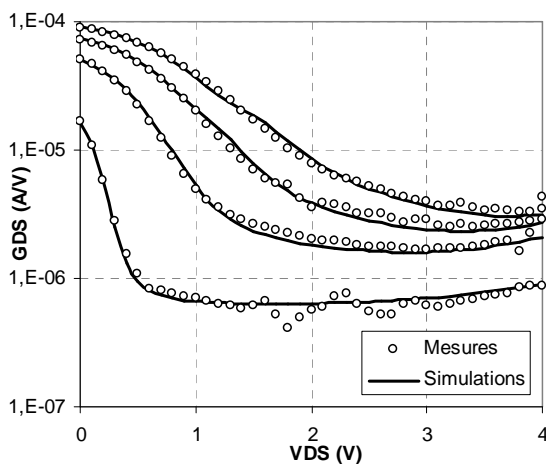




(c)



(d)



(e)

Fig.5.a.b.c.d.e EEPROM silicon results

4 EEPROM alternative models

4.1 Level 1 model

At a device level, when interfacing an EEPROM cell with peripheral circuits, it is necessary to know the memory cell I-V characteristics. As shown in Fig.6, during an EEPROM cell typical read operation, the control gate voltage V_{cgREAD} is set to a value close to the Virgin threshold voltage V_{TV} . In these conditions:

- If the memory cell is in the write state, the cell conducts the reading current I_{READ} (logical '1'),
- If the memory cell is erased, no current flows through the memory transistor (logical '0').

Then, the reading current is converted in a logical value at the output of the Sense Amplifier [17]. During this step, a high current resolution is mandatory in order to assure reliability. At a circuit level the bit line voltage is regulated ($V_{BL} \sim 0.8V$) to guarantee a sufficient reading current. This current, which depends mainly on the V_{cgREAD} value in a first approximation, is chosen so as to turn on only written cells.

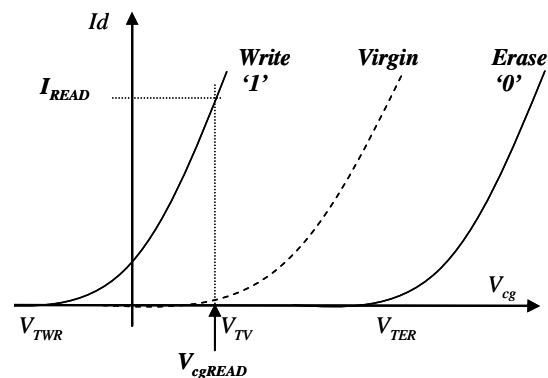


Fig.6 EEPROM read operation

Fig.7 shows that the EEPROM programming cycle is made up of two parts: an erase operation followed by a write operation. During a programming cycle, the signal BUSY is high and the memory state is evaluated according to the programming levels. The third basic operation is the read operation. During this step, the signal EN_SA is high and the suitable EEPROM cell current equation is chosen according to the state of the memory.

To use the level 1 model at a circuit level, additional nodes have to be considered by the model in order to respond to a program and read operation ordered by the control logic. Thus, an external signal called BUSY (this signal is set high at the beginning of a program operation and is reset at its end) and an external signal called EN_SA (this signal is set high at the beginning of a read operation and is reset at its end) are added as external PINs of the model. These signals are

presented Fig.7. Moreover, the memory state (erase or write) is an internal parameter of the model which means that when the model is evaluated, the state of the memory depends on the last programming cycle. External nodes of the level 1 EEPROM model are the bit line (BL), the word line (WL), the control gate (CG), the source (S), the bulk (B), the BUSY and the EN_SA PINs (see Fig.8).

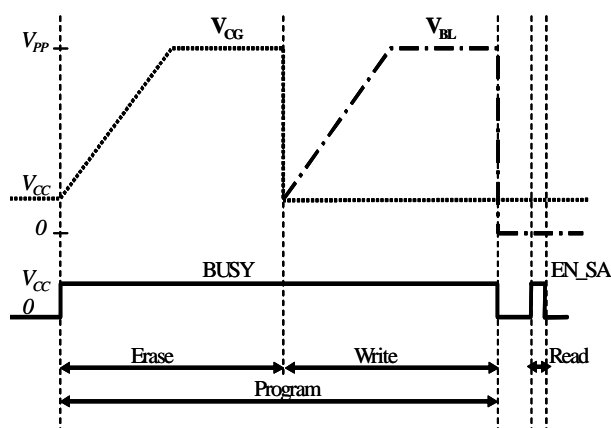


Fig.7 EEPROM basic operations (at a circuit level)

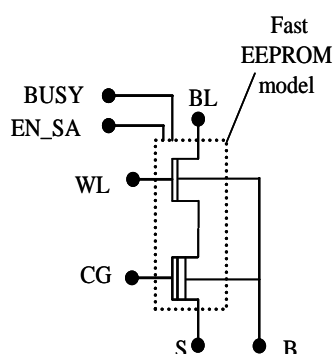


Fig.8 Level 1 EEPROM model

Based on these observations, the level 1 EEPROM model is built in order to determine first, the state of the memory after a programming cycle, and then to approximate as close as possible the conducting current as a function of the control gate voltage V_{cg} . More concretely, two sixth order polynomial equations of the bit line current I_d are used during the reading step according to the programming voltage levels V_{BL} and V_{CG} . In fact, the level 1 EEPROM model works as follow:

- If the CG node voltage reaches a given high voltage value V_{pp} , then the cell is erased (the internal memory state variable is set to '0') and the

equation of the erase characteristic is used during the read operation.

- If the BL node voltage reaches a given high voltage value V_{pp} , then the cell is written (the internal memory state is set to '1') and the equation of the write characteristic is used.

Advantages of such model are obvious: this basic model allows evaluating quantitatively the behavior of large memory arrays during the programming and read cycles. Nevertheless, this kind of model can not detect the high programming voltage variation and its impact on V_T and, therefore, on $I_d(V_{cg})$ curves. Besides, to detect Sense Amplifier marginalities during a read operation, the current equation definition have to consider the sub threshold voltage slop.

4.2 Level 2 model

The level 1 model assumes that the reading current I_d equals I_{off} and is close to 0 for an erased cell during a read operation. This assumption is only true for well-erased memory cells (V_{TER} value far from the V_{TV} one) and without taking into accounts the sub threshold voltage slope of the $I_d(V_{cg})$ memory characteristic. In addition, at a circuit level the high voltage V_{pp} has to be monitored as well as possible because it is one of the main causes of EEPROM memory issues (it leads to a V_T shift). Thus, impact of the high programming voltage variation has to be considered. Knowing that, a second EEPROM model is developed. This model includes two important features: the impact of the high voltage levels variation (i.e., V_{CG} and V_{BL} variations) on the V_T during the program operation and a more accurate expression of the current I_d during the read operation. The structure of this second model is the same as the one used for the previous one (see Fig.7). However, a new internal model parameter is introduced. This parameter is the threshold voltage variation called ΔV_T which takes into account the impact of V_{pp} variation on the V_T .

In this new model, the equation complexity increases. The memory drain current is defined by three equations (in agreement with the compact model behavior): the linear and quadratic current equation, the sub threshold voltage slop equation and the I_{off} current equation. The transition between each equation of the drain current is obtained by using smoothing functions.

The level 2 EEPROM model works as follow: after a programming operation, the drain current

characteristic (in its virgin state) is dynamically shifted according to the programming levels. Experimental results showing the programming levels impact on V_T are presented Fig.9. The evolution of the virgin curve as a function of V_{PP} is clearly shown for the $\log I_d(V_{CG})$ characteristic.

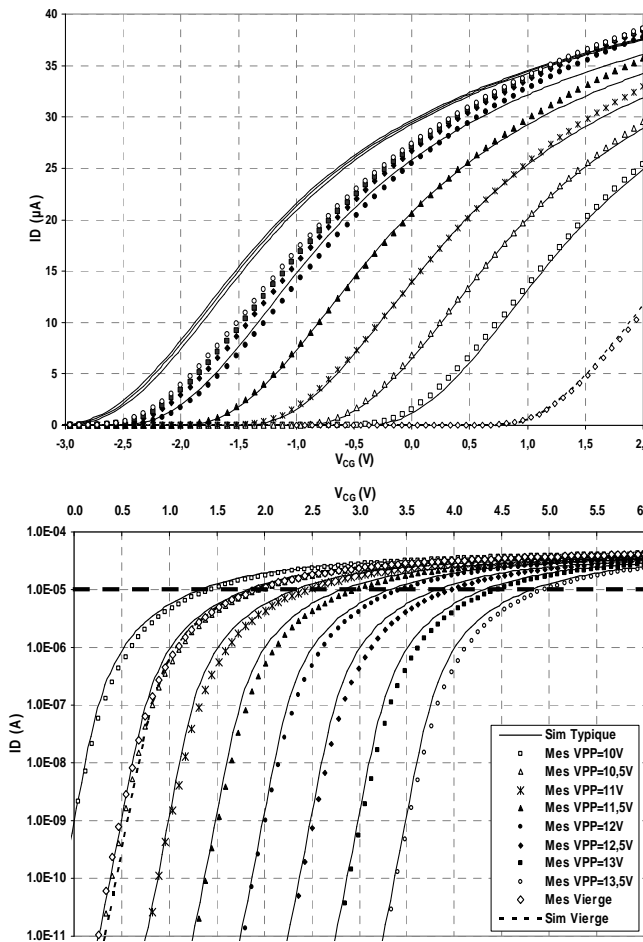


Fig.9 V_{PP} impact on I-V characteristics

The virgin curve evolution exhibits a horizontal translation and follows a linear law; this property has been validated on silicon. Thus, two equations of the V_T variation (ΔV_T) as a function of the write high voltage level (V_{PPWR} , equation 4) and the erase high voltage level (V_{PPER} , equation 5) are implemented in the Level 2 model. According to the V_T variation value, the memory drain current (defined in three parts) is dynamically shifted.

$$\Delta V_T = -1.0 * (1.035 * V_{PPWR} - 9.18) \quad (4)$$

$$\Delta V_T = +1.006 * (V_{PPER} - 10.35) \quad (5)$$

The new features of this second level model are essential if we want to evaluate more finely interactions between the memory array and the

peripheral circuitry.

5 Simulation results

To validate this approach, elementary memory arrays are extracted from an existing memory circuit: a 512Kbits EEPROM product. Generally, memory products are simulated without the memory array to check if the signals probed around the memory block match the specifications. This is done in order to save simulation time and because no memory models are available at product level.

To evaluate the EEPROM memory arrays, we build arrays with different densities: 64, 128, 256, 512 and 1k bits. For each array, elementary EEPROM cells (bits) are represented by the different EEPROM models: the compact model, the level 1 model and the level 2 model. The structure of a 128 bits memory array is presented Fig.11.

The elementary memory blocks are placed in actual operating conditions:

- The signals applied to the elementary arrays are obtained after performing global simulations using 512Kbits EEPROM circuit,
- The sense Amplifier, row and Columns latches are added to obtain realistic simulation conditions. The aim is to use these elementary memory arrays in a realistic simulation context.

Array Size	Compact	Level 1	Level 2
1cell	16s 100ms	12s 42ms	12s 61ms
64 cells	8mn 44s	1mn 27s	2mn 06s
128 cells	14mn 5s	2mn 13s	3mn 56s
256 cells	34mn 27s	4mn 25s	6mn 12s
512 cells	4h 18mn 1s	8mn 28s	11mn 50s
1k cells	23h 26mn	18mn 57s	24mn 06s

Table 1: Simulation time results

Simulation time results, obtained after a programming cycle (in global mode) followed by a read cycle are reported in Table 1. It appears clearly that the compact model is unsuitable when used to build memory arrays. For a 1k bits memory array with minimum peripheral circuitry, we obtain a 1 day simulation time. The important simulation time is due to the memory programming operation. When comparing the level 1 and level 2 models, the difference in simulation time is not important. The reason is simple: the current models equations are used during the reading operation and this last operation is very short comparing to the programming time.

Concerning the models accuracy, simulations have been performed using the different models. After a programming cycle (Fig.10), logical values (Fig.11) and the cell drain current I_d (Fig.12) are measured. Logical values are the same for each model. Concerning the cell current I_d , the maximum difference between the compact model and the level 1 model is around $0.8\mu\text{A}$. This current measurement is made in DMA (Direct Memory Access) mode. In this test mode, peripheral circuitry is bypassed, making the bit lines accessible; thus, the current sunk by each cell can be directly measured.

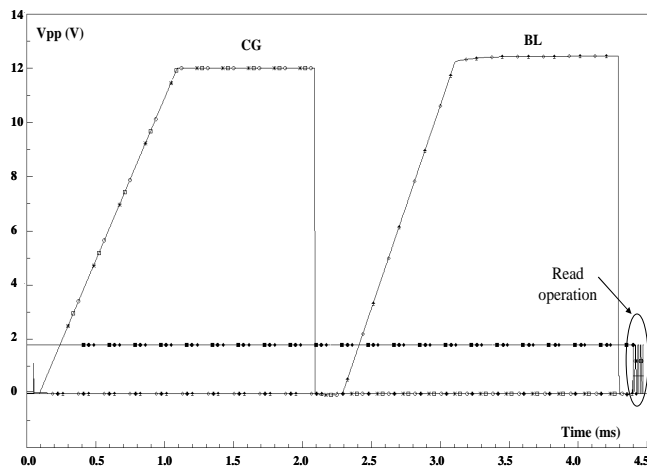


Fig.10 V_{CG} , V_{BL} and DATA signals for each model

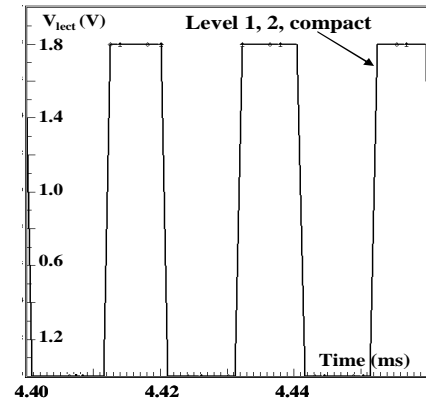


Fig.11 Logical values

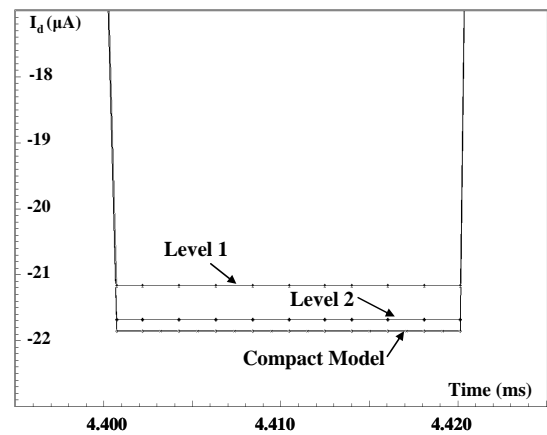


Fig.12 I_d cell current

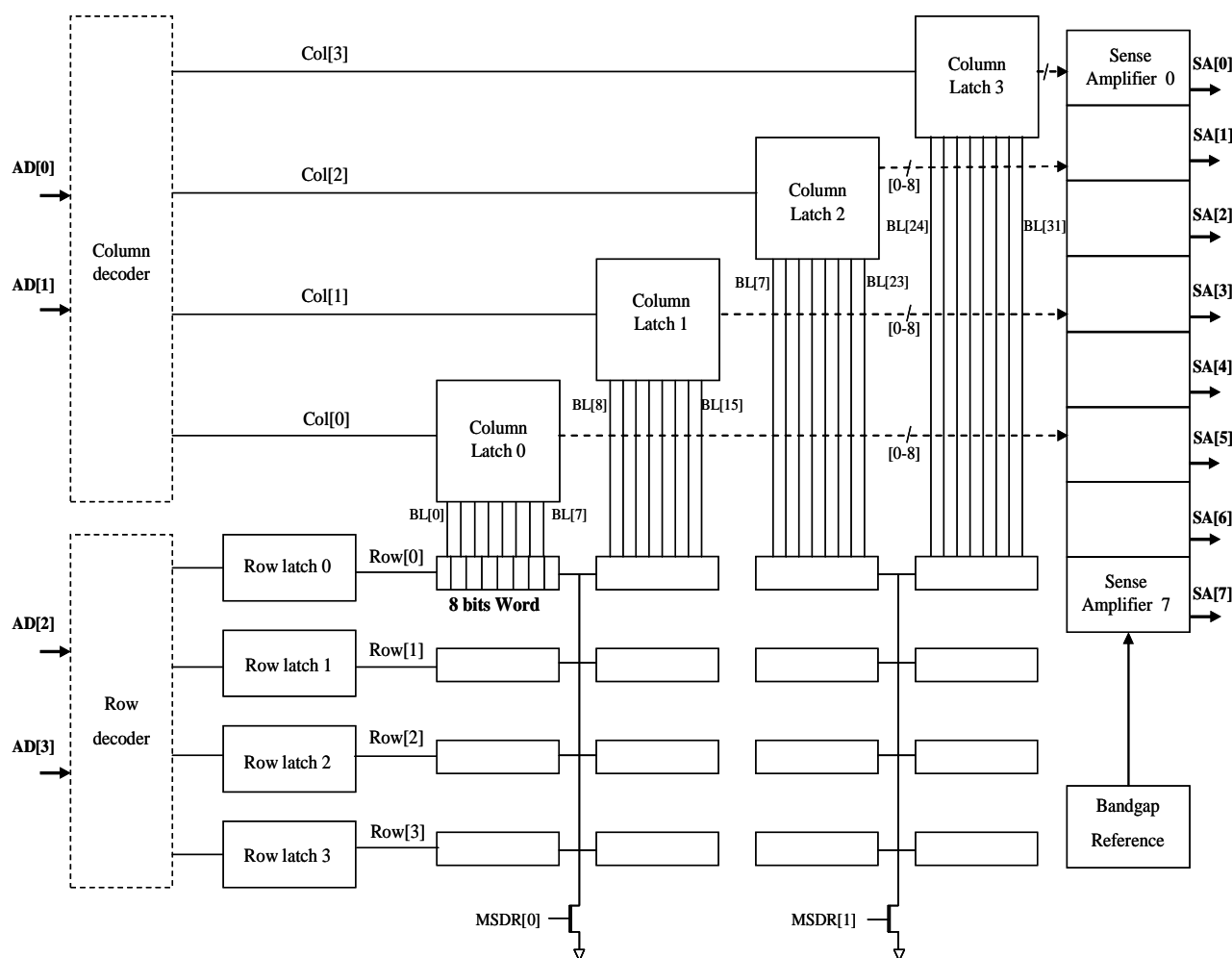


Fig.13 128 bits EEPROM memory array

6 Conclusion

In circuit designs, EEPROM memory blocks are critical in terms of simulation time. Fast EEPROM models (level 1 and level 2) presented in this paper are a solution to speed up simulation time with respect of accuracy when compared to compact models. The compact EEPROM model based on MM11 equations is first introduced and validated on silicon. Then, 2 alternative EEPROM models based on analogical equations are derived from this initial model. To check the models performances, the models are evaluated within elementary memory arrays. For a 1k memory array, level 1 model is seventeen times faster and level 2 models is sixteen times faster comparing to the compact model with respect of accuracy (0.5% of accuracy lack for level 1 and 0.1% for level 2 concerning the reading current). Thanks to this new models memory fault

simulation (impact of a defective isolated on its neighbours or on the array) can be considered.

References:

- [1] S. Bose and A. Nandi, "Extraction of Schematic Array Models for Memory Circuits", *Proceedings of the conference on Design, automation and test in Europe (DATE'04)*, Vol. I, p. 10570, 2004.
- [2] L. C. Liong and Po-C. Liu, "A Theoretical model for the Current-Voltage Characteristics of a Floating-Gate EEPROM cell", *IEEE Transactions on Electron Devices*, Vol. 40, No. 1, pp.146-151, 1993.
- [3] L. Larcher, P. Pavan, F. Gattel, L. Albani and A. Marmiroli, "A New Compact DC Model of Floating Gate Memory Cells Without Capacitive Coupling Coefficients", *IEEE*

- Transactions on Electron Devices*, Vol. 49, No. 2, 2002.
- [4] A. Kolodny, S. T. K. Nieh, B. Eitan and J. Shappir, "Analysis and modeling of floating-gate EEPROM cells", *IEEE Transaction on Electron Devices*, Vol. ED-33, pp. 835-844, 1986.
 - [5] Roberto Bez, Daniele Cantarelli and Paolo Cappelletti, "Experimental Transient Analysis of the tunnel Current in EEPROM Cells," *IEEE Transactions on Electron Devices*, Vol. 37, No. 4, pp.1081-1086, 1990.
 - [6] M. Meng and K.V Noren, "A Macromodel for FLOTOX EEPROM", *IEEE 39th Midwest symposium on Circuits and Systems*, Vol. I, pp.39-42, 1996.
 - [7] Alessandro Pirola, "A Solution for Hardware Emulation of Non Volatile Memory Macrocells", *Proceedings of the conference on Design, automation and test in Europe, DATE'03*, 2003.
 - [8] E. Vittoz, "The design of high performance analog circuits on digital CMOS chips", *IEEE Journal of Solid State Circuits*, Vol. SC-20, No 3, pp. 657-665, 1985.
 - [9] J.F. Dickson, "On-chip high-voltage generation in MNOS integrated circuits using an improved voltage multiplier technique", *IEEE Journal of Solid-State Circuits*, Vol. 11, Issue 3, pp. 374-378, 1976.
 - [10] M. Zhang and N. Llasar, "Dynamic analysis of Dickson charge pump circuits with a resistive load", *Proceedings of the 2003 10th IEEE International Conference on Electronics, Circuits and Systems*, Vol. 2, pp. 431-434, 2003.
 - [11] P. Pavan, R. Bez, P. Olivo, E. Zanoni, Flash memory cells-an overview, *Proceedings of the IEEE*, Vol. 85, Issue 8, pp. 1248-1271, 1997.
 - [12] M. Lenzlinger and E.H. Snow, "Fowler-Nordheim tunneling into thermally grown SiO₂", *Journal of Applied Physics*, Vol. 40, pp.278-283, 1969.
 - [13] R. Van Langevelde, "MOS Model 11", *Unclassified Report NL-UR 2001/813*, Philips Electronics N.V., 2004.
 - [14] Y. Tsvetidis, "Operation and Modeling of the MOS transistor", Mc Graw Hill international Edition, 1999.
 - [15] R. Bouchakour, H. Recoules, W. Benzerti and H. Petit, "An analytical damaged submicron MOSFET model for CAD application", *Proceedings of the 40th Midwest Symposium on Circuits and Systems*, Vol. 2, pp. 1473-1476, 1997.
 - [16] N. Harabech, R. Bouchakour, P. Canet, P. Pannier, J.P. Sorbier, "Extraction of Fowler-Nordheim parameters of thin SiO₂ oxide film including polysilicon gate depletion: validation with anEEPROM memory cell", *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, pp.441-444, 2000.
 - [17] D. Baderna, A. Cabrini, G. De Sandre, F. De Santis, M. Pasotti, A. Rossini and G. Torelli, "A 1.2V Sense Amplifier for High-Performance Embeddable NOR Flash Memories", *IEEE International Symposium on Circuits and System*, Vol. 2, pp.1266-1269, 2005.