Floating Gate Devices: Operation and Compact Modeling

Paolo Pavan*, Luca Larcher** and Andrea Marmiroli***

*Dipartimento di Ingegneria dell'Informazione, Università di Modena e Reggio Emilia,
Via Vignolese 905, 41100 Modena, Italy, paolo.pavan@unimore.it
**Dipartimento di Scienze e Metodi dell'Ingegneria, Università di Modena e Reggio Emilia,
Via Fogliani 1, 42100 Reggio Emilia, Italy, larcher.luca@unimore.it
***STMicroelectronics, Central R&D, Via C. Olivetti 2, 20041 Agrate Brianza (MI), Italy,
andrea.marmiroli@st.com

ABSTRACT

This paper describes a possible approach to Compact Modeling of Floating Gate devices. Floating Gate devices are the basic building blocks of Semiconductor Nonvolatile Memories (EPROM, EEPROM, Flash). Among these, Flash are the most innovative and complex devices. The strategy followed developing this new model allows to cover a wide range of simulation conditions, making it very appealing for device physicists and circuit designers.

Keywords: compact model, nonvolatile memory, floating gate, reliability, circuit design.

1 INTRODUCTION

Flash Memories are one of the most innovative and complex types of high-tech, nonvolatile memories in use today, see for example [1]. Since their introduction in the early 1990s, these products have experienced a continuous evolution from the simple first ones to emulate EPROM memories, to the extreme flexibility of design application in today products. In the memory arena, Flash memory is the demonstration of the pervasive use of new electronic applications in our lives, exploiting this flexible and powerful memory technology either as a stand-alone component or embedded in a product. Flash are not just memories, they are "complex systems on silicon": they are challenging to design, because a wide range of knowledge in electronics is required (both digital and analog), and they are difficult to manufacture. Physics, chemistry, and other fields must be integrated; and conditions must be carefully monitored and controlled in the manufacturing process.

Compact Models (CMs) of Floating Gate (FG) devices are therefore needed and they have the same purpose of all compact models: *to be used within a program for circuit simulation*. The Floating Gate transistor is the building block of a full array of memory cells and a memory chip. In a first approximation, the reading operation of a FG device can be considered a single-cell operation. Nevertheless, CMs are fundamental to simulate the effects of the cells not directly involved in the operation under investigation and the effects of the parasitic elements. Furthermore, they allow the simulation of the interaction with the rest of the device, and hence they can be used to check the design of the circuitry around the memory array: algorithms for cell addressing, charge pump sizing taking into account current consumption and voltage drops, etc…

In this scenario, despite of the wide diffusion of FG-based Non-Volatile Memories, no complete CMs of FG devices were proposed and used in the industry until few years ago. Usually, MOSFET transistors whose threshold voltage was *manually* changed to model programmed and erased state of the FG memory cell were used in circuit simulations to reproduce (with poor accuracy) the FG memory behavior.

2 FLOATING GATE DEVICE MODEL

The FG device is the building block of a nonvolatile memory cell. The device is a MOS transistor with a conductive layer "floating" between gate and channel, Fig 1 [2]. The basic concepts and the functionality of this kind of device are easily understood if it is possible to determine the FG potential. The schematic cross section of a generic FG device is shown in Fig. 1. The FG acts as a potential well. If a charge is forced into the well, it cannot move from there without applying an external force: the FG stores charge. The presence of charge in this floating layer alters the threshold voltage of the MOS transistor (low threshold and high threshold, corresponding to "1" and "0").

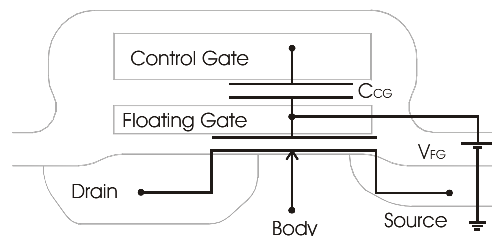

Figure 1. Cross section of a FG device and basic schematic of the CM subcircuit.

The FG device CM is the basic building block to model a single memory cell, full array, a memory chip. The simple idea underneath is to model the FG device as a circuit with a MOS transistor and a capacitor between the control gate and the FG node (which is the gate of the MOS transistor)

Fig 1. This CM exploits the MOS transistor model. Many MOS models have been developed (Philips MM11 [3], BSIM4 [4], EKV [5], SP [6], HiSIM [7]).

The approach for the FG device modeling followed here is independent of the specific MOS model adopted, thus exploiting all the improvements carried out for the basic MOS transistor models and the definition of their parameter extraction algorithms. The capacitance value is used, together with the charge injected in the floating gate, to calculate the FG node potential that is applied through a voltage controlled voltage source (V_{FG} in Fig. 1). This voltage generator is indispensable in DC conditions, as in circuit simulators there is no general solution to the calculation of the potential of a floating gate node in a DC conditions.

This model has also the advantage to allow the modeling of programming and erasing operations by simply adding a set of suitable current generators between the various electrodes. This modular approach enables the modeling of read/program disturbs, retention, leakage currents, in a rather simple way.

Two main limitations of this model can be foreseen. First, usually MOS compact models target thin gate oxide transistors with Lightly– or Medium– Doped Drain (LDD / MDD) diffusions. The oxide thickness of all FG devices is above the 7 nm, while the source and drain junctions are usually abrupt. It might become necessary to adapt the existing transistors models to this kind of devices. Second, there are a few coupling capacitances which are neglected: the coupling between the control gate node and the source, drain, and body nodes. Furthermore, as memory cells are getting smaller and closer one to the other, the coupling capacitance between the electrodes of two neighbor cells (which are not included in the model) may become more important.

2.1 DC Operation: Read

There are very few works in the literature to address the task of simulating the DC behavior of the FG memory cells, see for example [8]. With this specific CM, the FG node is biased to its correct value by an external source: the voltage-controlled voltage source, V_{FG}, which constitutes the core of the model in DC conditions.

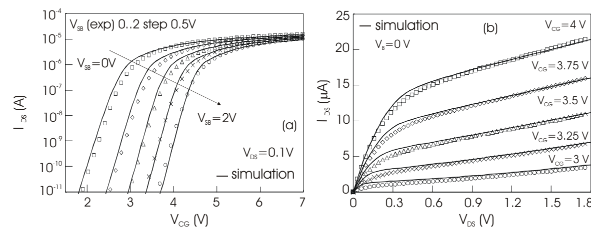

Figure 2. DC characteristics: experimental (symbols) and model (solid lines) for a 0.25μm Flash memory cell (W=0.25μm, L=0.375μm, C_{CG}=0.8fF).

The new approach gives many advantages compared to standard models [1]:

1. *scalability*: scaling rules are already included in the compact MOS model adopted and they do not affect directly the V_{FG} calculation routine;
2. *implementation*: it uses standard circuit elements whose parameters can be determined by applying the MOS parameter extraction procedure to the dummy cell, and the few additional parameters can be easily estimated from cell layout and cross section;
3. *accuracy*: it depends mainly on the compact MOS model adopted, taking advantage of the many efforts to improve and scale MOS CMs;
4. *computation time*: comparable to a MOS transistor;
5. *modularity*: it can be easily extended to simulate transient behaviors of FG memories by adding a suitable set of voltage controlled current sources to its basic structure.

2.2 Parameter extraction procedure

The procedure to extract the parameters of device CMs is not a "push-button" task. For FG devices, this task is even more complex than for standard MOS transistors. Reasonable results are obtained paying attention to the slightly different physics of the *dummy cell* (which is the cell where FG and CG are short-circuited) compared to a standard MOS transistor: narrow and short geometry, the lack of LDD and Pocket Implant determine a less ideal behavior such as larger DIBL effect and higher multiplication current. Care has to be devoted to extract the overlap capacitance values (overlap capacitances are very small: their evaluation is particularly critical).

Except the CG-FG capacitance (C_{CG}), additional parameters depend on the kind of FG memory considered. In EEPROM memory cell they are: 1) the area of the tunneling region; 2) the tunnel oxide thickness; 3) the doping levels of the drain well and the FG. In Flash memory cells they are: 1) the areas of S-FG, D-FG, and channel-FG overlap regions; 2) the doping levels of S and D wells, channel and FG. Generally, these parameters are either directly evaluated from the layout of the cell (C_{CG}, tunnel and overlap region areas), or straightly derived from the process recipe (doping). Sometimes, dedicated measurements performed on MOS capacitor test structures.

2.3 Transient Operations: Program and Erase

To simulate the program/erase operations of FG devices we have extended the model by adding a suitable set of voltage controlled current sources to the basic framework to implement compact formulae of program/erase currents. The number and position of current generators depend on the FG memory considered (Flash or EEPROM cells) and the writing mechanisms used to transfer charge to and from the FG. For example, to extend the DC model of Flash memory devices to account for writing operations, three

voltage controlled current sources have to be added to reproduce program/erase currents [9]:

1. a voltage controlled current source between FG and S, $\mathbf{I_{W1}}$, which models the Fowler Nordheim (FN) current flowing at the source side (needed when modeling Flash memories erased by FN tunnel at the source side);
2. a voltage controlled current source connected between FG and B, $\mathbf{I_{W2}}$, which models the FN tunnel current flowing toward the substrate;
3. a voltage controlled current source connected between FG and D, $\mathbf{I_{W3}}$, which models Channel Hot Electron (CHE) and Channel Initiated Secondary Electron (CHISEL) injection currents, via suitable compact formulae.

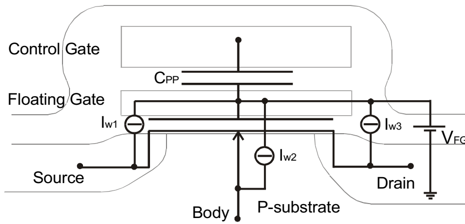

Figure 3. The complete CM of a Flash memory cell: basic framework plus three voltage controlled current sources, I_{W1}, I_{W2} and I_{W3}, that model P/E currents.

When using this model, the simulation accuracy of FG device program/erase operations depends strictly on the precision of CMs developed to describe FN, CHE and CHISEL currents. Therefore, great attention has to be devoted to develop effective CMs of these currents mechanisms.

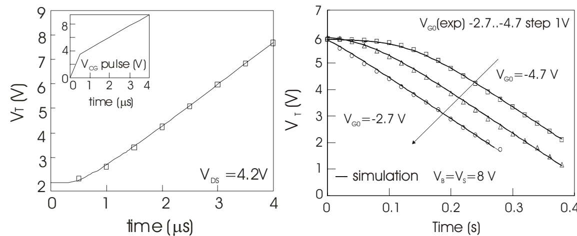

Figure 4. Threshold voltage (V_T) shifts measured (symbols) and simulated (solid lines) during program and erase of a Flash memory cell.

3 CIRCUIT SIMULATIONS

The schematic of the sense amplifier circuit simulated is shown in Fig. 5 [10]. It is a classic scheme where active load p-channel transistors are biased to provide the wanted constant current, thus allowing a controlled trip point voltage and temperature compensation. The structure is fully differential to have good noise immunity. Mn1 and Mn3, Mn2 and Mn4 provide the current/voltage conversion to bias the reference cell and the cell to be read in the matrix. V_{CELL} and V_{REF} are voltages deriving from the I-V conversion of currents driven by the cell in the memory array and the reference cell, that are compared to generate the V_{SENSE_OUT} digital level.

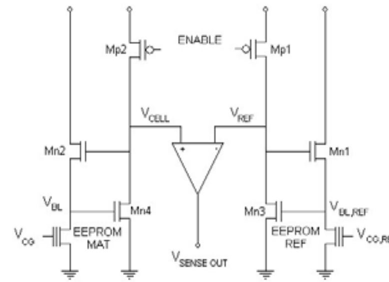

Figure 5. Schematic of the sense amplifier and the direct I-V conversion circuits of an EEPROM memory.

Simulation results in Figures 6 (a)-(b) [13] show that the output signal of the sense amplifier switches correctly according to the programmed/erased state of the EEPROM memory cell. This model is effective to simulate FG-based memory cells also in complex circuits, and therefore it can be used to simulate any circuit including a FG memory cell: read paths, non-volatile latches, X and Y decoders, voltage pumps.

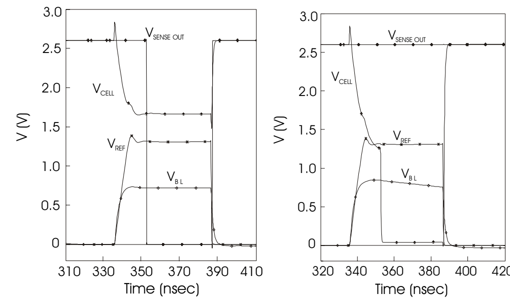

Figure 6. Control signals, and sense amplifier output obtained from read-path circuit simulations in the two cases of a programmed (a) and erased (b) EEPROM memory cell.

4 RELIABILITY SIMULATIONS

Usually, the reliability of FG memory devices is investigated through experimental techniques and the use of suitable ad-hoc models to describe leakage currents through their oxide layers. In fact, leakage currents through gate and interpoly oxides are the most serious concern for the reliability of FG memory devices, since they can strongly degrade data retention properties and increase program and read disturbs. In this scenario, we will show that the CM of FG devices (extended to include leakage current effects) can be a versatile and powerful tool for reliability predictions. CMs allow also to bridge the gap between the oxide quality characterization activity performed traditionally on MOS transistors and capacitors, and the actual impact of Stress Induced Leakage Current (SILC) on

FG memory reliability. This CM is an effective tool to predict FG memory reliability degradation, the influence on data retention of P/E cycles, P/E bias conditions, thickness and quality of tunnel oxide, and storage field.

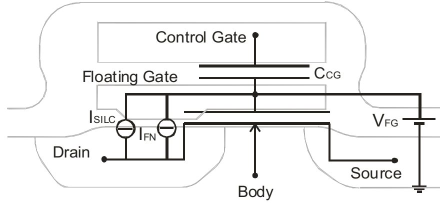

Figure 7. The CM of an EEPROM memory cell extended to simulate SILC-induced EEPROM reliability degradation by including the current generator, I_{SILC}, implementing an analytical SILC formula.

To this purposes, the CM of EEPROM memory cell can be extended by including a voltage-controlled current source implementing the empirical SILC expression proposed in [11]. As shown in Fig. 7, the SILC current generator is connected between the drain and the floating gate since in this region SILC is much larger due to the thinner thickness (~7nm) of tunnel oxide compared to the gate oxide one (~20nm).

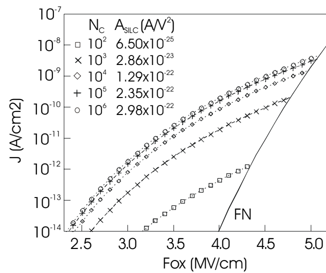

Figure 8. SILC curves (symbols) simulated through (1) in a 7 nm thick oxide on increasing the number of P/E cycles (N_C). The A_{SILC} corresponding to N_C is also indicated. The "classic" FN current is shown by a solid line.

To correlate the tunnel oxide degradation induced by high-field stress to the P/E cycles, N_C, and the P/E bias conditions, the charge exchanged during a P/E cycle and the P/E current density flowing through the tunnel oxide have been evaluated using the model. In Fig. 8, SILC curves simulated by considering typical P/E conditions (*Program*: D is ramped from 0 to V_R=12 V with a ramp rise time T_R=0.5 ms, CG and B are grounded, and S is left floating; *Erase*: CG is ramped from 0 to V_R=12 V with a ramp rise time T_R=0.5 ms, S, D, and B are grounded) are depicted for different P/E cycles.

Read disturb simulations have demonstrated that SILC is not a concern for EEPROM cells considered, since the oxide field and the time involved in read operations are too low to induce significant FG charge variations, i.e. V_T modification regardless the SILC magnitude. On the contrary, data retention losses are strongly affected by Stress Induced Leakage Current, as predicted by reliability simulations and also confirmed by experimental data.

Fig. 9 shows V_T shifts simulated in an erased EEPROM cell left unbiased for ten years at room temperature. There are two aspects worth stressing. First, the threshold voltage reduction occurring after ten years increases with the number of P/E cycles (see dashed lines). This V_T trend is due to the SILC rise on increasing N_C (see Fig. 8), whereas the overlap of V_T-time curves for $N_C \leq 10$ is determined by the fact that in these storage field conditions the tunnel current is dominated by the FN component. Second, the ten year threshold voltage after 10^5 cycles does not depend on the initial V_T, i.e. on the initial storage oxide field, $F_{OX,S}$: as shown in Fig. 9, V_T-time curves simulated for EEPROM cells after 10^5 P/E cycles assuming different initial V_T (solid lines) converge to a similar value after 2-3 years, which depends on SILC magnitude, but it is independent on the initial V_T, i.e. $F_{OX,S}$.

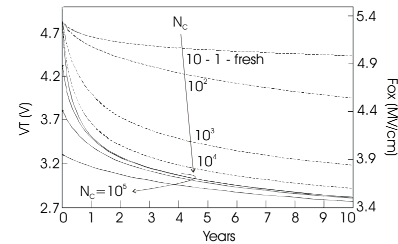

Figure 9. V_T decay for a single EEPROM cell left unbiased at room temperature on increasing the cycle number N_C (dashed lines). The oxide field is also indicated. All dashed lines start from the same initial storage field. For N_C=10^5, different initial storage fields have been assumed, and the decay curves are shown by solid lines.

REFERENCES

[1] VV.AA., Proceedings of the IEEE, Special Issue on: *Flash Memory Technology*, April 2003

[2] .L. Larcher, et al., IEEE Trans. on Electron Devices, Vol.49, N.2, pp. 301-307, 2002.

[3]http://www.semiconductors.philips.com/Philips_Models/

[4] W. Liu, X. Jin, K. M. Cao, an C. Hu, "BSIM4.0.0 MOSFET Model User's Manual," 2000.

[5] http://legwww.epfl.ch/ekv/

[6] G. Gildenblat, et al, Proc. 1997 International Semiconductor Device Research Symposium, p. 33, 1997.

[7] .http://home.hiroshima-u.ac.jp/usdl/HiSIM.html

[8] A. Kolodny, et al. IEEE Trans. Electron Dev., Vol. ED-33(6), pp.835-844, 1986.

[9] L. Larcher, P. Pavan, Proc. MSM2002, Puerto Rico, April 21-25, 2002, pp. 738-741

[10] P. Pavan, et al., IEEE Trans. on CAD, Vol. 22, N. 8, August 2003, pp. 1072 –1079

[11] L. Larcher, et al., IEEE Transactions on Device and Materials Reliability, Vol. 2, N. 1, March 2002, pp. 13 - 18