We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,300

Open access books available

172,000

International authors and editors

190M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Computational Modeling and Monte Carlo Simulation of Soft Errors in Flash Memories

Jean-Luc Autran, Daniela Munteanu,
Gilles Gasiot and Philippe Roche

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/57220>

1. Introduction

As CMOS technologies are scaling down, the susceptibility of integrated circuits (IC's) to radiation coming from space or present in the terrestrial environment has been found to seriously increase [1]. Until now, radiation effects in IC's have mainly been an issue for space or avionics applications. At ground level and in nowadays ultra-scaled devices, natural atmospheric radiation principally induces Single event effects (SEE), which has been identified to induce one of the highest failure rates of all reliability concerns for devices and circuits entering in the area of nano-electronics [2]. SEE's are the result of the interaction of highly energetic particles, such as protons, neutrons, alpha particles, or heavy ions, with the sensitive region(s) of a microelectronic device or circuit. A single event may perturb the device/circuit operation (e.g., reverse or flip the data state of a memory cell, latch, flip-flop, etc.) or definitively damage the circuit (e.g. gate oxide rupture, destructive latch-up events) [3].

Among all integrated circuits used in many application areas for which a high reliability level is required (medical, space, automotive, networking, nuclear), non-volatile memories are known for their relative robustness to single events, even if the different components of a flash memory circuit (on one hand the memory cell array, on the other hand the peripheral control circuitry) exhibit distinct levels of radiation sensitivity. In addition, the specific question of their sensitivity to the terrestrial radiation environment has been little studied until now. Cellere et al. [4-5] and Gerardin et al. [6-7] have been the first to clearly state, using accelerated tests, that atmospheric neutron induced soft error occurrence is possible in flash memories, although with extremely low probabilities at ground level. A very recent study by Just et al. [8], based for the first time on real-time tests performed in a mountain altitude natural environment, has concluded in a similar way: natural atmospheric radiation at ground level can induce

soft-errors in flash memories, typically several decades below the soft-error rate (SER) of static RAM (SRAM) of comparable technological nodes.

In this context, we recently developed a numerical simulation code capable of computing the SER of floating-gate flash memories. Our simulation platform named TIARA-G4 and described in Ref. [9], has been adapted to flash memory architectures (TIARA-G4 NVM release for "Non-Volatile Memories") by modifying the device/circuit 3D geometries and by implementing a model for charge loss from the floating gates induced by ionizing particles. This chapter presents in detail our modeling and simulation approach as well as the code validation by comparison of numerical results with experimental data reported in [8].

The chapter is organized as follows: in Section II, we briefly introduce some basic knowledge about the architecture and electrical operation of floating gate flash memories. Section III also briefly reviews the current comprehension of radiation effects in floating-gate memories. The objective of these two first sections is to introduce for a non-specialist reader the technical background necessary for a good understanding of the second part of this chapter more specifically dedicated to computational modeling and Monte Carlo simulation issues. In Section IV, we detail our modeling and simulation approach based on the adaptation of our TIARA-G4 simulation platform [9] to flash memory architectures. Finally, in Section V, we expose the simulation results and compare them to experimental results obtained on a large collection of memories exposed to natural radiation.

2. Flash memory architectures and electrical operation

This paragraph provides a brief introduction to the architecture and operating principles of floating gate flash memories at both device and circuit levels. Flash memory is an electronic non-volatile storage device that can be electrically erased and reprogrammed; it offers fast read access times, as fast as dynamic RAM, although not as fast as static RAM or ROM. Flash memories are used in a wide variety of electronic devices for general storage, configuration data storage or data transfer. Modern flash memories store logical information in an array of memory cells built from floating-gate transistors. In traditional single-level cell devices, each cell stores only one bit of information. Some newer flash memory, known as multi-level cell devices can store more than one bit per cell by choosing between multiple levels of electrical charge to apply to the floating gates of its cells.

As illustrated in Fig. 1, the memory cell consists of a single n-channel transistor with a control-gate (CG) and an electrically isolated polysilicon floating gate (FG). The two gates are separated by an oxide-nitride-oxide dielectric stack (ONO), often called "inter-poly oxide". Data can be stored in the cell by adding or removing electrons in the FG, which induces changes of the threshold voltage of the cell transistor. Charge injection into the floating-gate through the tunnel oxide (TO) is governed by the electrical signals applied on the control-gate owing to the electrostatic coupling existing between the two gates. Indeed, the electrostatic potential of the FG (V_{FG} , see Fig. 2 left) is directly determined by the potential of the CG and the amount of electrical charge stored in the FG. These operations require high voltage signals produced

Figure 1. Schematic cross-sections of a floating-gate transistor, acting as the elementary memory element in a flash memory circuit, along its length (left) and its width (right).

on-chip using special DC-to-DC converters (charge pumps) that uses capacitors as energy storage elements to create higher voltages from the circuit external supply voltage. Two threshold voltage levels (V_T^0 and V_T^1 , see Fig. 2 right) are considered to store one bit of information in the cell. The difference between the two levels, ΔV_T , is directly linked to the variation of the charge amount in the FG and to the coupling capacitance between CG and FG electrodes. A reference voltage value V_T^{REF} , intermediate between V_T^0 and V_T^1 is considered as a demarcation level between the two logical states “0” and “1” (Fig. 2 right).

Figure 2. Left: Equivalent capacitance network of the floating-gate transistor with four terminals and definition of the main voltage and capacitance values. Right: Electrical characteristics $I_{DS}(V_{CG})$ of the floating-gate transistor with two different values of the floating-gate charge corresponding to erased and programmed states.

To form dense circuits with storage capacities up to several millions or billions of bits, elementary memory cells are arranged in matrix, i.e. in rows and columns. In addition, cells are generally grouped to form a hierarchical organization of increasing size: groups, blocks, pages, etc. Lines, called “wordlines” are connected to the control gates and columns, called “bit-lines”, are connected to the drain terminals. Around the matrix memory is the peripheral control circuitry composed of additional circuits for decoding cell addresses, generating high voltage signals (charge pumps), reading cells (sense amplifiers) and managing circuit information. There are two main types of circuit architectures at the memory plan level, called NOR and NAND gate flash memories, each corresponds to a certain manner to associate several cells. The construction and the operation of NOR and NAND flash memories are briefly described in the following.

- NOR architecture:** The organization of the NOR gate flash, shown in Fig. 3(a) is the following: several cells are connected to a bit line; each cell has the source terminal connected directly to ground, and the drain terminal is connected directly to a bit line. The drain contacts of individual transistors connected to the bitline are shared between two adjacent cells. This setting of elementary devices is called "NOR flash" because it operates as a NOR gate. The default state of a single-level NOR flash cell is logically equivalent to a binary "1" value: when a suitable voltage is applied on the control gate the current flows through the channel and the bitline voltage is pulled down. The programming operation of a NOR flash cell (i.e. setting to a binary "0" value) is done by injection of hot carriers from the channel. The high current, required by this mechanism, limits the parallelism of the operation (only some cells can be programmed in same time) [10]. The programming procedure is the following: a voltage increase (typically > 5 V) is applied to the control gate which turns on the channel and the electrons can flow from source to drain (for an n-channel MOS transistor). The source-drain current is sufficiently high so that a certain number of electrons of high energy are able to pass through the insulating layer on the floating gate by hot electrons injection mechanism. The erasing operation of a NOR flash cell (resetting it to the "1" state) is done by Fowler-Nordheim (FN) mechanism. For this purpose, a large voltage of the opposite polarity is applied between the control gate and the source terminal, pulling out the electrons from the floating gate by FN tunneling. This organization of the NOR flash allows a fast random access (approximately 100 ns). The programming operation is carried out at block level, and is much slower (approximately 5 μ s). The erasing operation is carried out on the level of block and is even slower, typically 200 ms [10]. Taking into account these characteristics, the NOR flash is used principally as a read-only memory mainly for code storage, for which the random access time is important, but where the programming/erasing operations are rarely carried out [10]. In the NOR architecture, the manufacturer guarantees that all individual bits are functional and meet retention and endurance specifications, as explained in [10]; no implementation of Error Correction Code (ECC) is needed from the user side. In some cases (e.g., multi-level architecture), an internal ECC, totally transparent to the user, may be present. NOR devices typically have separate buses for addresses and data [10].
- NAND architecture:** The organization of the NAND gate flash is shown in Fig. 3(b). In this configuration, several groups of floating-gate transistors are connected in series. These groups are then connected via some additional transistors to a NOR-style bit line array in the same way that single transistors are linked in NOR flash. Due to this arrangement, the bit line is pulled low only if all word lines are pulled high (above the threshold voltage of the transistors). This organization of elementary transistors is called NAND flash because transistors are connected in a way which is similar to a NAND gate. Compared to NOR flash, replacing single transistors with serial-linked groups adds an extra level of addressing. As explained in [10], the series arrangement and the great level of parallelism, which is achieved with this organization thanks to the low program/erase currents, give rise to a poor random access time but a very good serial access. Programming of the NAND flash is performed by FN tunneling at the page level (which is typically a few kBytes) and is carried out in about 0.2 ms [10]. The erasing operation is performed at the block level (typically a

few MBytes) and takes about 2 ms [10]. Block erasure is carried out also by FN tunneling, but by using opposite polarity. Thanks to these characteristics flash NAND is adapted better for the data storage, where the problems of latency are minor and the random access time is not very important. In this configuration, the use of external ECC is mandatory (which increases the latency), because the manufacturer does not guarantee each single bit and the commercial devices may contain a few defective blocks [10].

Figure 3. Schematic representations of NOR and NAND architectures.

3. Radiation effects in floating-gate memories

Floating-gate memories are sensitive to ionizing radiation, both to total ionizing dose (TID) and single event effects (SEEs). Very schematically, ionizing radiation induces charge loss in the floating-gate and charge trapping in the different dielectric layers of the transistor stack; it can also generate interface states. The induced current transients and such parasitic charges and defects cause degradation of circuit functionality and/or loss of logical information stored in the FG array in addition to possible global circuit performance degradation. Detailed results of both TID and SEEs in flash memories are available in recent papers or review presentations [4-7,10-13].

In this study, we will exclusively focus on soft-errors induced by atmospheric neutrons in the FG array of flash memories. SEEs in FG memories are due to highly energetic particles that directly (heavy ions) or indirectly (neutrons) induce charge loss from the FG. Other effects may be possible, such as Single Event Functional Interruptions (SEFI) or destructive events (Single Event Gate Rupture – SEGR) at the level of the FG array or in the peripheral circuitry. Note that SEEs only affect FG cells impacted by at least one particle whereas TID uniformly impacts the programmed FG cells.

Figure 4. Schematic illustration of the effects of highly ionizing particle irradiation on the threshold voltage distributions of a floating gate array. Adapted from Paccagnella et al. [12].

Neutrons are not ionizing (they not directly create e^-/h^+ pairs in the matter) and specifically due to their neutral character, they can penetrate deeply into the chip atomic structure. Only the resulting products of the neutron-silicon (or other atoms of the circuit, O, W, Al, etc.) collisions are ionizing and, by consequence, only the impact of such secondary products on the FG can result in charge loss. This is the reason why, in the following, we will focus on the underlying physical mechanisms of charge loss induced by ionizing particles, but the link with atmospheric neutrons remains evident.

Figure 4 illustrates the effects of ionizing-particle irradiation on the threshold voltage distributions of a large array of FG devices. Before irradiation, threshold voltages of individual cells are distributed following a typical Gaussian distribution, sharply centered on the programmed value V_T^1 . A secondary peak and a tail appear after irradiation: these structures correspond to all cells that have been hit by incident ionizing particles. The position (V_T shift) of the peak with respect to the initial distribution gives the average threshold voltage shift: it is directly linked to the ion Linear Energy Transfer (LET) and to the electric field in the tunnel oxide. The height of the peak is related to the irradiation fluence. Finally, the tail is related to all memory cells for which the V_T values are intermediate between the secondary peak and the initial distribution. Of course, all cells, initially programmed at V_T^1 and having their post-irradiation V_T value below V_T^{REF} , have been upset.

In their IRPS 2008 paper [14], Butt and Alam reviewed several models of charge loss due to a radiation particle strike. Different physical mechanisms have been proposed in the literature for the modeling of the charge loss from floating gates after single radiation particles strikes.

The authors summarized these earlier models and underlined their strengths and limitations; such classical models include the trap assisted tunneling (TAT), the conductive pipe model, the generation-recombination-transport in oxide and the electron emission. The most important limitation of these approaches is their failure to quantitatively predict the charge loss on the basis of a set of physics-based equations without any fitting parameter and/or phenomenological assumption. The authors have then proposed a new model, called Transient Carrier Flux (TCF) model, which quantitatively explains the observed charge loss in FG memories irradiated with heavy ions. Figure 5 illustrates all these different physical mechanisms and models, which are shortly detailed in the following.

1. **Generation/Recombination/Transport in Oxide:** The ion strike produces hot holes in the tunnel oxide or inter-poly dielectric and a certain fraction of these hot holes are not recombined in the prompt recombination phase. This model considers that the not-recombined holes [15] may drift into the floating gate. This may be possible since the negative electron charge stored on the floating gate itself produces an electrical field across the oxide which attracts holes. These holes that drift in FG are recombined with electrons stored in FG and cause a reduction in the negative charge on the floating gate. At the same time, the electrons produced by the ionizing particle are quickly transported to the silicon bulk or to the control gate due to their high mobility. However, this model lacks sufficient experimental validation because it does not agree quantitatively with data loss measured in FG Flash memory cells [16]. Indeed, the number of holes that survive the prompt recombination after a heavy ion strike in a 10 nm tunnel oxide is less than 100, while the data show that the charge loss is a few thousand electrons [14].
2. **Electron Emission:** This phenomenon was originally proposed by Snyder et al. as one of the main mechanisms of charge loss in FG EEPROM cells under gamma ray irradiation [17]. The charge loss is explained by the fact that electrons stored in the floating gate can gain energy from ionizing radiation and can be emitted over the oxide barrier in the control gate or in the silicon substrate. This mechanism is also called photoemission. The emission over the oxide has been empirically modeled. However, this mechanism has not been physically modeled or extended heavy ions or to other particles [14]. Moreover, in this model, the photoemission is limited only to electrons stored in the floating gate which has no physical justification [14]. In fact, an ionizing particle strike can generate a number of electrons much larger than the net number of electrons stored in the floating gate. Some of the electrons generated by the particle strike may have enough energy to be emitted over the oxide barriers [14].
3. **Trap Assisted Tunneling (TAT):** This mechanism is one of the most important causes of oxide wear out due to electrical stress of program/erased cycles of a FG cell. When an ionizing particle strikes the cell, defects are created in the tunnel oxide. These defects may provide a percolation path for the electrons which can thus pass by tunneling effect through the tunnel oxide; therefore, this mechanism is called trap assisted tunneling (TAT). It has been shown that TAT is responsible for retention problems in at least a certain percentage of irradiated devices [14]. However, a very long time is required to discharge a FG cell by TAT mechanism (a few hours to a few weeks) [16]. Therefore, the TAT

mechanism cannot be responsible for SEU in FG cell taking into account that SEU data are taken immediately after the cell irradiation and then do not change with time [14]. TAT may nevertheless result in hard errors that cause retention problems of FG cell.

4. **Conductive Pipe Model:** This model has been proposed by Cellere et al. to explain the charge loss due to heavy ions strikes [15], [18]. This model assumes that the dense plasma of e-h pairs generated by the ion strike creates a temporary very thin (~ 10 nm) conductive path in the tunnel oxide during a short time (sub picosecond) after the strike. This is accompanied by the local lowering of the oxide energy barrier, which allows the electrons stored in the floating gate to pass through this conducting pipe. This phenomenological model reproduces well the experimental data of charge loss. However, there is a lack of physical explanation of the mechanisms governing both the resistance of the conductive path and the oxide barrier lowering [14].

Figure 5. Schematic illustration of different models of charge loss due to a radiation particle strike. Adapted from Butt et Alam [14].

5. **Transient-Carrier-Flux (TCF) model:** This model was proposed by Butt and Alam [14] to explain the charge loss due to a SEU in FG memory cells. In this model it is assumed that the dominant physical mechanism that causes the FG charge loss due to a particle strike is the net flux of hot carriers flowing within a short time ($\sim$ ps) over the oxide barrier at the FG/oxide interfaces. After a particle strike, a dense cluster of hot electron-hole pairs are generated with carriers having broad energy distributions which return to thermal equilibrium in a time $\sim$ 1 ps [14]. The tail of the high energy distribution induces a transient carrier flow in and out of the floating gate over the tunnel and inter-poly oxides. In case of a zero electric field in the oxide, the incoming and outgoing carrier flow balances each other at both oxide/FG interfaces and therefore the net flux is zero. On the contrary, in the programmed state, the electron negative charge stored in the floating gate induces a relatively high electric field in the oxide. Due to this electric field the electrons flux leaving the floating gate is greater than the electron flux entering the floating gate. In addition, the incoming holes flux is greater than the holes flux exiting the floating gate. The net flux therefore causes a reduction of the number of electrons stored in the floating gate. A small imbalance between the incoming and outgoing fluxes may be sufficient to disturb the state of the memory cell for which the tolerance of charge loss can be 100 electrons or less [14]. Butt and Alam validated their model by numerical simulations using a high-energy particle physics based toolkit - Geant4 for the generation and initial energy distributions in the high energy range ($\sim$ 10eV - $\sim$ keVs). The hydrodynamic model coupled with Monte Carlo simulations was used for carrier relaxation in low energy ($<$ 10eV) range, in order to accurately take into account the energy relaxation due to phonon scattering and impact ionization [14]. The transient fluxes of hot carriers flowing in and out the floating gate over the barrier oxides are calculated by solving self-consistently a system of equations including the transmission probability through the oxides and the Poisson equation, until carriers relax and reach the thermal equilibrium. These fluxes are then used to obtain the charge loss in flash memory cells due to alpha particles and cosmic neutron strikes. Butt and Alam finally demonstrated that the TCF model is in very good agreement with experimental data from Ref. [16], as will be shown later in Section 4.2.

4. Modeling and simulation of non-volatile memories using TIARA-G4 platform

In this section, we describe in details our modeling and numerical simulation approaches to compute the SER related to the floating-gate array of a flash memory circuit.

4.1. Description of TIARA-G4 NVM platform

The Tool Suite for Radiation Reliability Assessment (TIARA) platform has been developed these last years conjointly at Aix-Marseille University (IM2NP laboratory) and at STMicroelectronics (Central R&D, Crolles). The last version of the code has been called TIARA-G4 in reference to the fact that it is totally rewritten in C++ using Geant4 classes

and libraries and compiled as a full Geant4 application [19]. This major evolution of TIARA allows us to consider now all the complexity of a given integrated circuit in terms of materials, doping and 3D geometry, using the Virtual Geometry Model (VGM [20]) factory and interface with both Geant4 for calculation and Root [21] for visualization. Up to now, TIARA-G4 has been used to simulate the interaction of Geant4 particles (including high energy and thermal neutrons, protons, muons, alpha-particles and heavy ions) with various SRAM and Flip-Flop architectures [9].

Figure 6 shows a schematic of the new TIARA-G4 NVM simulation flow structured into several independent modules and integrating new dedicated modules/subroutines to floating-gate NVM devices. In particular, we wrote a new cell/circuit construction model to reproduce the flash chip geometry (floating-gate array) with high fidelity. A second dedicated module implementing a physical model for radiation-induced charge loss from the floating-gate has been also developed, as detailed in paragraph 4.2.

Figure 6. Schematics of the TIARA-G4 NVM simulation flow showing the different code inputs and outputs and the links with Geant4 classes, libraries, models or external modules and visualization tools.

To test the capability of the code to consider a real geometry, we based our developments on a NOR floating-gate flash memory architecture designed and fabricated by STMicroelectronics using a 90 nm CMOS process. This process is based on a Boro-Phospho-Silicate Glass (BPSG)-free Back-End Of Line (BEOL) which eliminates the major source of ^{10}B in the circuits and drastically reduces the possible interaction between ^{10}B and low – thermal energy neutrons

[22]. Figure 7(a) shows a TEM cross-section of the floating-gate devices along the transistor channel and Figure 7(b) shows a portion of the cell array layout at metal1/metal2 level. The elementary memory cell has an area of $0.18 \mu\text{m}^2$. In TIARA-G4 NVM, the different transistor domains have been modeled as simple axis-aligned box volumes (Geant4 elements) of different materials (silicon, silicon dioxide, ONO and back-end-of-line stack), as illustrated in Figure 7(c) for a portion of the memory array. Figure 7(d) also shows a larger view of the array with different particle tracks interacting with certain floating-gate stacks.

4.2. Physical model considered

In complement to geometrical aspects, we also implemented in TIARA-G4 NVM a new module describing the charge loss from floating gates after single radiation particles strikes. From the review of the different available models in literature presented in Section 3, our initial choice was to adopt the full physical model of the Transient Carrier Flux (TCF) proposed by Butt and Alam [14]. The original approach of these authors is therefore based on complex simulations, in particular for the computation of carrier relaxation in the low energy ($< 10\text{eV}$) range, using coupled hydrodynamic and Monte Carlo simulations in order to correctly account for energy relaxation due to phonon scattering and impact ionization. This requires outsourcing from the main code the calculation of the charge loss from FG as a function of the incident particle properties.

Figure 7. nm NOR floating-gate flash memory architecture considered in this work. (a) TEM cross-section of the floating-gate transistor geometry along the transistor channel and (b) layout of the cell array at metal1/metal2 level. (c) and (d): ROOT screenshots of a TIARA-G4 simulation showing detailed (c) and global (d) views of the memory array and different particle tracks resulting from atmospheric neutrons interaction with circuit materials. For a better view at FG cell level, all BEOL materials (6 metal levels), silicon substrate and intra-cell silicon and dielectrics are not shown.

An example of Butt and Alam's simulations is illustrated in Figure 8. Simulated curves very well reproduce experimental data without any fitting parameter (data extracted from Fig. 12 of Ref. [14]). From a practical point-of-view and in absence of a relatively simple computational solution to implement the Butt and Alam's model, we adopted a pragmatic approach assuming that an ionizing particle of LET striking the FG produces a Number of Electron Loss given by the following analytical function:

$$NEL = A \times LET^2 + B \times LET \quad (1)$$

Figure 8 shows that Eq. (1) is able to very well reproduce data. Of course, fitting coefficients A and B must be carefully evaluated for each device considered for the simulation from experimental measurements (heavy ion irradiation) or complementary numerical simulation using the Butt and Alam's complete computational procedure. The next release of TIARA-G4 NVM will integrate such an external dedicated module to confer to the code the capability to simulate a wide variety of NVM devices.

Figure 8. Number of electron loss (NEL) as a function of the particle LET for device T3 of Ref. [16] under an oxide electric field of 3.5 MV/cm. Simulation results from Butt & Alam (Ref. [14]) are also reported. The full line corresponds to the fitting function (1) on experimental data.

In the particular case of the present study and by chance, Figure 8 is based on data from Cellere et al. who precisely worked on STMicoelectronics FG arrays. It has been found that device T3 in Ref. [16] is technologically very close to our circuit, with the same thicknesses for the different layers composing both the FEOL and BEOL stacks. In order to consider values given by Eq. (1) to our memory devices, we introduced a scaling factor coefficient to take into account the difference in the dimensions of the floating gate polysilicon electrodes between devices considered in Fig. 8 and the present memory cell architecture (simple ratio of the volumes). Without any other calibration, we use in the following Eq. (1) to directly derive the threshold voltage shift resulting from a single particle strike in the FG domain using:

$$\Delta V_T = \frac{q \times NEL}{C_{pp}} \quad (2)$$

where C_{pp} is the coupling capacitance between the FG and CG electrodes (see Fig. 2 left).

5. Results and discussion

This last section presents the numerical simulations performed with TIARA-G4 NVM for the 90 nm NOR floating-gate flash memory architecture previously described. In a second part, we report experimental measurements obtained from the direct exposition of a large number of circuits to natural radiation. These two sets of data are finally compared and discussed in the last part of this section. It is important to notice that, in the following, all numerical results concerning the characterization and the simulation of the 90 nm flash circuit have been normalized by a common arbitrary scaling factor for confidentiality reasons imposed by the semiconductor manufacturer.

Figure 9. ROOT screenshots of a TIARA-G4 simulation showing several hundred of memory cells and different particle tracks resulting from atmospheric neutrons interaction with circuit materials. All BEOL materials, intra-cell silicon and dielectrics are not shown, silicon substrate is represented in yellow.

5.1. Numerical simulations

Using TIARA-G4 NVM, we performed extensive Monte Carlo simulations on large arrays of memory cells (up to 10^5 cells) considering the JEDEC atmospheric neutron source for high-energy incident neutrons above 1 MeV [23]. Other simulations have been also performed using a random generation of alpha particles inside the silicon material for mimicking the presence of ^{238}U contamination at ppb-level (we considered in this case the eight alpha-particle emitters of the ^{238}U decay chain) [24].

Two simulation screenshots are shown in Figure 7(c) and (d) in the case of a reduced matrix of 30×30 cells (considered for a better view). A larger simulation view is shown in Figure 9 for several hundred cells. They illustrate the interaction of atmospheric neutrons with the circuit materials and the way in which the neutron-induced secondary particles can impact the memory cells (direct strikes on the FG electrodes). A large part of the events are induced by secondary particles generated in the proximity of the FEOL/BEOL interface and predominantly by protons and silicon recoil nuclei. The BEOL stack is found to contribute marginally (< 2%) to the total SER in spite of the presence of several layers and vias of high density materials (W, Cu, Ta).

Figure 10. Distributions of V_T values computed by TIARA-G4 NVM for a population of 100,000 memory cells before and after irradiation with atmospheric neutrons.

Figure 10 shows the simulated V_T distributions for 10^5 cells before and after irradiation. The initial distribution corresponds to a Gaussian distribution with the mean and standard deviation values calibrated on experimental data (see 5.2, Fig. 13). The final distribution is the result of 10^9 incident JEDEC neutrons on the cell matrix, which corresponds to 50×10^6 h (i.e. more than 5700 years!) under natural atmospheric radiation at New-York City (NYC), the reference location defined by a high energy neutron flux of $20 \text{ n/cm}^2/\text{h}$ (neutron energies above 1 MeV). One can observe the emergence of a typical neutron-induced tail on a large domain of V_T values below 7 V. This tail indicates that the V_T value has sufficiently decreased for a certain number of cells to appear outside the Gaussian distribution. Among them, some cells have shifted below the sense value fixed at 5.7 V: their state has thus changed from a logical point-of-view (0 $\rightarrow$ 1 transition) and their number must be taken into account for the evaluation of the neutron-SER. For the other cells of the distribution tail, ΔV_T values are not sufficient to decrease their V_T below 5.7 V but large enough to shift the cells outside the initial curve.

Figure 11. Distributions of ΔV_T values extracted from data of Fig. 10.

In complement to Figure 10, Figure 11 shows the threshold voltage shift distribution for all the cells of the simulated array. The peak at $\Delta V_T = 0$ V indicates that the great majority of the cells have not been impacted during the simulation run. For $\Delta V_T > 0$ V, the distribution is decreasing when ΔV_T increases. This directly reflects (cf. Eqs. (1) and (2)) the LET distribution of the secondary particles (i.e. neutron byproducts) striking the floating gates: the lightest particles (protons, alphas) with low LET values (typically below $1.5 \text{ MeV.cm}^2/\text{mg}$) induce a large number of events characterized by a small or moderate ΔV_T shift (< 1 V); on the contrary, particles with the highest LET values, much less numerous, induce the largest ΔV_T (> 3 V). From the number of cells verifying $V_T < 5.7$ V after irradiation, the neutron-SER at sea-level has been numerically evaluated to 7.7 (in arbitrary unit taking into account the common arbitrary scaling factor for confidentiality reasons). This value is expressed for the reference location (NYC).

For the alpha-SER, a value of 0.12 (a.u.) has been obtained considering a concentration of 0.2 ppb of ^{238}U uniformly distributed in the volume of circuit materials at both FEOL and BEOL levels. This concentration was directly deduced from experimental emissivity measurements (see below).

5.2. Experimental characterization and results

In parallel to this work of modeling and numerical simulation, previously described, we launched an experimental verification procedure to estimate the circuit SER from direct measurements. For this, we considered a large collection of NOR floating-gate flash memory circuits fabricated by STMicroelectronics using a 90 nm CMOS process. Circuits have been directly operated and characterized at wafer-level.

Figure 12. Layout (left) and die (right) of the ANNA test chip (area $9.230 \times 7.044 \text{ mm}^2$) fabricated by STMicroelectronics in CMOS 90 nm technology. The memory array is segmented into 32 blocks of 4 Mbits or 128 sectors of 1 Mbits (total capacity of 128 Mbits per chip).

The test chip, named “macrocell ANNA” and shown in Figure 12, is a 128 Mbit array of memory cells organized in 1 Mbit sectors, 4 Mbit blocks and 16 Mbit quarters without ECC. Several tens of macrocells are available per test wafer (200 mm wafers); more than 50 Gbits (~ 20 wafers) were used and fully characterized for the present experiment.

The test began by an initial wafer-level characterization at ST-Rousset (near Marseille) of all the circuits using a high performance tester (Verigy® V93000 platform). The test platform uses high precision voltage sources and parameter analyzers calibrated before each measurement campaign: the accuracy on V_T extraction is guaranteed to be less than 10 mV. Memory arrays have been written (all “0” pattern) and then read several times, allowing the compilation of a reference threshold voltage (V_T) mapping for all the test chips, cell per cell and wafer per wafer. The corresponding numerical data have been stored on a hard disk bay. During this initial characterization, all the wafers were also submitted to a 24h bake at 250°C followed by a new V_T characterization in order to identify (and thus to eliminate) all the test chips exhibiting electrical instabilities and/or abnormal FG charge loss. Figure 13 shows a typical V_T distribution, sharply centered around 7.8 V for a population of memory cells corresponding to all functional test chips for a series of five wafers (same technological lot). The reproducibility (i.e. repeatability) of such an electrical characterization has been attested by the fact that repeated measurements on the same wafer show exactly the same V_T distribution within measurement margins ($< 10 \text{ mV}$), cell per cell.

In addition to this initial electrical characterization, we also performed alpha-emissivity measurements at wafer-level using a XIA UltraLo-1800 alpha-particle counter. Figure 14 shows the results of this characterization, in terms of emissivity and measurement error (Fig. 14 left) and of energy distribution (Fig. 14 right) of the emitted alpha particles from the fully processed wafers. An emissivity level of $0.0013 \alpha/\text{cm}^2/\text{h}$ was measured, which corresponds to a concentration of 0.2 ppb of ^{238}U uniformly distributed in the volume of circuit materials at both FEOL and BEOL levels. Such a correspondence has been estimated using a reverse α -particle emissivity analytical modeling recently developed [25].

Figure 13. Initial distribution of the measured threshold voltage V_T values for all the programmed FG memory cells (all "0" pattern) related to a series of 5 wafers.

Figure 14. Alpha-particle emissivity characterization of 90 nm flash memory wafers using a XIA UltraLo-1800 alpha-particle counter. Left: emissivity and measurement error sigma as a function of measurement duration. Right: energy distribution of the detected alpha particles emitted from the surface of the wafers (fully-processed wafers).

After the initial characterization, approximately one half of the total number of wafers was stored in Rousset and the second half was delivered to an altitude test site by express mail and exposed to natural radiation. Figure 15 shows the flowchart of this test method that illustrates the sequencing of the different characterization and wafer transportation steps. Two different radiation environments have thus been considered: the first one at sea-level in Rousset for reference and the second one in altitude on the ASTEP platform [26]. The two sites are characterized by a relative atmospheric neutron flux of 1.04 and 6.02 with respect to New-York City, respectively [27-28]. After a period of exposition of several months, the wafers stored on ASTEP (see Figure 16) have been delivered to ST-Rousset for complete electrical characterization. Those remained in Rousset were also measured in the same time. The complete characterization loop Rousset → ASTEP → Rousset was repeated 3 times for the present work.

Figure 15. Flowchart of the multi-site characterization technique developed to evaluate the soft error rate of flash memories written and read at wafer-level using a Verigy® V93000 platform.

Figure 16. Global view of one of the ASTEP experimental room showing, in the foreground, six wafers of flash memories stored on the ground during their exposition to natural radiation on the ASTEP platform and, in the background, a real-time test setup based on 40nm SRAM circuits [9].

Figure 17 shows the results for the two series of wafers exposed in Rousset and on the ASTEP Platform. Three reading operations have been performed on the wafers stored in Rousset, respectively after 5, 12 and 18 months of exposition. Similarly, two reading operations have been performed on the ASTEP wafers, after 5 and 12 months of natural irradiation in altitude.

For wafers exposed at sea-level, one memory cell compared to more than several tens of Gbits has been detected with a V_T value changing at $t_0 + 12$ months and becoming inferior to the reference value ($V_T^{REF} = 5.7$ V) delimiting the "0" and "1" logical states. For this memory cell, the threshold voltage shifted from 8.0 to 4.5 V. Likewise, 2 and 3 shifted- V_T cells have been detected on the ASTEP wafers, respectively after 5 months and one year of exposition. Measured V_T values for these flipped cells are also reported in Figure 17.

Figure 17. Number of memory cells with shifted V_T below the reference value (5.7 V) delimiting the "0" and "1" logical states and detected during the first and the second wafer readings.

Figure 18. Comparison between the two distributions of V_T values measured at t_0 and at $t_0 + 12$ months for population of programmed memory cells exposed to natural radiation on the ASTEP platform.

A detailed analysis of the analogic V_T bitmaps (not shown) for all these impacted cells shown that these latters correspond to isolated cells (i.e. adjacent cells not impacted) randomly distributed in the FG array and on the exposed wafers. A more detailed investigation on the complete V_T distributions shows that several other cells have been potentially impacted during their exposition to natural radiation. Figure 18 shows such a distribution for the whole cell population exposed on ASTEP. At $t_0 + 12$ months, two groups of impacted cells can be distinguished: a first group of 3 cells, labeled A, which corresponds to the 0→1 flipped cells reported in Figure 17 (bottom graph, at $t_0 + 12$), and a second group of 6 cells, labeled B, for which the V_T have shifted but not enough to cross the limit of 5.7 V delimiting the two binary states "0" and "1".

From data of Figure 17 obtained at two different locations, the global soft error rate (SER) and its two components can be determined, as suggested in [29]. The two components are, on one hand, the n-SER taking into account the atmospheric neutrons contribution to the SER and, on the other hand, the so-called α -i-SER accounting for all the internal failure mechanisms in the chips, including the possible alpha-particle emitter contribution. Indeed, several physical intrinsic mechanisms can be invoked to explain the long-term charge loss generally observed in FG devices, in particular different leakage mechanisms through the tunnel oxide or through the ONO interpoly dielectric based on various possible trap/defect assisted tunneling [30]. These latter are not inevitably related to radiation effects but can be also linked to material properties or induced by the technological process or by an electrical stress. This is the reason why the second contribution to the SER is called here α -i-SER and not only α -SER. We thus have a system with two equations and two unknown quantities:

$$\alpha - i - SER + AF_{Rousset} n - SER = SER_{Rousset} \quad (3)$$

$$\alpha - i - SER + AF_{ASTEP} n - SER = SER_{ASTEP} \quad (4)$$

where $AF_{Rousset} = 1.04$ and $AF_{ASTEP} = 6.02$ are the neutron flux acceleration factor, as previously reported in II.B.

Figure 8 shows the results of this SER extraction, considering results of Figure 17, durations and memory capacities related to the different experiments. Global SER values of 9.7 and 28.8 a.u. are obtained for Rousset (sea-level) and ASTEP (altitude) experiments, which leads to an estimation of α -i-SER = 3.2 and n-SER = 8.7 a.u.

These results demonstrate a very limited impact of the atmospheric radiation on the total SER without ECC, typically in the range [10-100] FIT/GBit. With respect to all other internal failure mechanisms, the external natural radiation constraint is found to represent less than one third (27%) of the total SER. Note that all these SER values are found strictly equal to 0 if ECC is activated on the chips, due to the fact that only rare events always corresponding to single cell upsets have been detected.

Figure 19. Summary of the SER deduced from data of Fig. 4 for sea-level and ASTEP conditions. The two components of the SER are given for normalized New-York City conditions. SER values are in a.u. for confidentiality reasons but the order of magnitude of these values is a few hundreds FIT per GBit.

5.3. Discussion

In this last paragraph, we conclude by comparing these experimental results with predictive values obtained using the TIARA-G4 NVM simulation platform. Figure 20 summarizes this comparison for the different defined SER components.

A good agreement is found for the neutron-SER taking into account all experimental and simulation uncertainties, in the first instance, the relatively weak statistics of the experiment in terms of number of events detected. Indeed, despite the duration of the experiment (18 months) and the huge quantity of data to manipulate (the individual V_T evolution of more than 50 Gbits of memory cells has been stored and processed), the statistics of this first experiment remains relatively weak because of the extremely low rate of cell flips in this kind of memory.

For alpha-SER, the discrepancy is flagrant between the two values. This confirms our initial precaution to name the second extracted component of the SER (Fig. 19) α -i-SER instead of classically α -SER because, in the present case of FG devices, this component may be the result of other intrinsic failure mechanisms occurring in parallel inside the chips. From literature [31-33], we can invoke different intrinsic or extrinsic leakage current mechanisms through the dielectric layers present in the floating gate stack (tunnel oxide, ONO, spacers). Intrinsic mechanisms that contribute to charge loss are field-assisted electron emission, thermionic emission and electron detrapping. Extrinsic mechanisms are essentially oxide defects that can form conductive paths through a given dielectric. Whatever the mechanism or eventually the activation of several leakage paths, our results suggest that these electrical processes appear to be dominant in the observed failure rate with respect to the contribution of alpha-particle internal emission. This point will have to be carefully reevaluated in future works.

Figure 20. Comparison of the SER component values obtained by TIARA-G4 NVM simulation and from exposition to natural radiation in Rousset and on ASTEP.

Another interesting point of comparison comes from the ratio of the numbers of upset cells to the numbers of cells for which V_T have shifted but not enough to cross the limit of 5.7 V delimiting the two logical states. Although statistics are low for data of Fig. 18, the ratio (number of cells B/number of cells A) can be roughly evaluated to 50%. From simulation results with a much larger statistics, this ratio is 40.7%, which is clearly in the same order of magnitude. Beyond the fact that this point consolidates the comparison between experiment and simulation, this result shows that the number of impacted cells with a final V_T ranging between the sense voltage value and the edge of the initial Gaussian distribution is approximately two times larger than the number of cells verifying the upset criterion.

6. Conclusion

In conclusion, we developed in this work a numerical simulation code (TIARA-G4 NVM) capable of computing the soft-error rate of floating-gate flash memories induced by the two main natural radiation components at ground-level: the atmospheric high-energy neutrons and the alpha-particles emitted from ultra-traces of radioactive contaminants in circuit materials. Based on Geant4 geometry classes, elements and materials, the code is able to reproduce the circuit geometry from silicon substrate to back-end-of-line levels with fidelity. In complement to geometrical aspects, TIARA-G4 NVM also integrates a new module describing the charge loss from floating gates as a function of the properties (LET) of the incident ionizing particles. Using this code, we performed extensive Monte Carlo simulations on large arrays of memory cells (up to 10^5 cells) related to a 90 nm NOR floating-gate flash memory architecture designed by STMicroelectronics. Values of the SER for atmospheric neutrons and alpha-particle emitters have been computed and expressed, respectively, at sea-level (New-York City) and for a concentration of ^{238}U in the circuit materials separately

determined through experimental emissivity measurements. The experimental verification of these simulated results has been conducted, for the first time, following a totally new approach: the direct exposition to natural radiation of a large amount of test circuits programmed and periodically read at wafer-level with a dedicated industrial test equipment. In spite of a relatively weak statistics achieved during this experimental phase, the remarkable convergence of the experimental results and our numerical simulations (considering no fitting parameter in the complete simulation chain) for the neutron-SER indicates that this later value is more than two decades below the soft error rate usually measured in modern SRAMs. In the same way, the comparison of experimental data measured at sea-level and alpha-SER simulations clearly suggests that another mechanism than internal alpha-particle production in bulk materials may be responsible of charge loss from floating gates. This point will have to be carefully investigated in future works.

Acknowledgements

The authors would like to acknowledge G. Just, A. Regnier and J.L. Ogier from STMicroelectronics (Reliability and Electrical Characterization Laboratory, Rousset Plant, France) for their contribution to the electrical characterization of the flash memory wafers and Sébastien Sauze (ASTEP platform) for his technical and support. They also sincerely acknowledge their past post-doc students Sébastien Martinie (now at CEA-LETI, Grenoble) and Sébastien Serre (now at TRAD, Toulouse) for their different contributions to this work. The logistical support of the Institute for Radio astronomy at Millimeter Wavelengths (IRAM) is finally gratefully acknowledged.

Author details

Jean-Luc Autran¹, Daniela Munteanu¹, Gilles Gasiot² and Philippe Roche¹

¹ Aix-Marseille University & CNRS, Marseille, France

² STMicroelectronics, Crolles, France

References

- [1] Massengill L.W., Bhuvu B.L., Holman W.T., Alles M.L., Loveless T.D. Technology Scaling and Soft Error Reliability. In: Proceedings of the IEEE International Reliability Physics Symposium 2012, 3C.1
- [2] International Technology Roadmap for Semiconductors (ITRS), available online at www.itrs.net

- [3] Munteanu D., Autran J.L. Modeling of digital devices and ICs submitted to transient irradiations. *IEEE Transactions on Nuclear Science* 2008;55(4) 1854-1878.
- [4] Cellere G., Gerardin S., Bagatin M., Paccagnella A., Visconti A., Bonanomi M., Beltrami S., Harboe-Sorensen R., Virtanen A. and Roche P. Can atmospheric neutrons induce soft errors in NAND floating gate memories?. *IEEE Electron Device Letters* 2009;30(2) 178-180.
- [5] Cellere G., Gerardin S., Bagatin M., Paccagnella A., Visconti A., Bonanomi M., Beltrami S., Roche P., Gasiot G., Harboe-Sorensen R., Virtanen A., Frost C., Fuochi P., Andreani C., Gorini G., Pietropaolo A., Platt S. Neutron-induced soft errors in advanced flash memories. In: *Proceedings of the IEEE International Electron Devices Meeting* 2008, pp. 1 – 4.
- [6] Gerardin S., Bagatin M., Paccagnella A., Cellere G., Visconti A., Beltrami S., Andreani C., Gorini G., Frost C.D. Scaling Trends of Neutron Effects in MLC NAND Flash Memories. In: *Proceedings of the IEEE International Reliability Physics Symposium* 2010, pp. 400 – 406.
- [7] Gerardin S., Bagatin M., Ferrario A., Paccagnella A., Visconti A., Beltrami S., Andreani C., Gorini G., Frost C. D. Neutron-Induced Upsets in NAND Floating Gate Memories. *IEEE Transactions on Device and Materials Reliability* 2012;12(2) 437-444.
- [8] Just G., Autran J.L., Serre S., Munteanu D., Sauze S., Regnier A., Ogier J.L., Roche P., Gasiot G. Soft Errors Induced by Natural Radiation at Ground Level in Floating Gate Flash Memories. In: *Proceedings of the IEEE International Reliability Physics Symposium* 2013, 3D-4.
- [9] Autran J.L., Semikh S., Munteanu D., Serre S., Gasiot G. and Roche P., “Soft-Error Rate of Advanced SRAM Memories : Modeling and Monte Carlo Simulation, Numerical Simulation - From Theory to Industry”, edited by Mykhaylo Andriychuk, ISBN : 978-953-51-0749-1, InTech, pp. 309-336, Sept. 2012. Available online at <http://dx.doi.org/10.5772/50111>
- [10] Gerardin S., Bagatin M., Paccagnella A., Grünmann K., Gliem F., Oldham T.R., Irom F., and Nguyen D.N. Radiation Effects in Flash Memories. *IEEE Transactions on Nuclear Science* 2013;60(3) 1953-1969.
- [11] Gerardin S. and Paccagnella A. Present and future non-volatile memories for space. *IEEE Transactions on Nuclear Science* 2010;57 3016-3039.
- [12] Paccagnella A., Bagatin M., Cellere G., Gerardin S. Flash memories and soft errors at ground level. In: *RADSOL workshop*, June 2009.
- [13] Gerardin S., Bagatin M., Ferrario A., Paccagnella A., Visconti A., Beltrami S., Andreani C., Gorini G., Frost C. D. Neutron-Induced Upsets in NAND Floating Gate Memories. *IEEE Transactions on Device and Materials Reliability* 2012;12(2) 437-444.

- [14] Butt N.Z., Alam M. Modeling Single Event Upsets In Floating Gate Memory Cells. In: Proceedings of the IEEE International Reliability Physics Symposium 2008, pp. 547 – 555.
- [15] Ma T.M. and Dressendorfer P.V. Ionization Radiation Effects in MOS Device and Circuit. Wiley, New York, 1989.
- [16] Cellere G., Paccagnella A., Visconti A., and Bonanomi M. Subpicosecond conduction through thin SiO₂ layers triggered by heavy ions. Journal of Applied Physics 2006;99 074101.
- [17] Snyder E.S., McWhorter P.J., Dellin T.A., and Sweetman J.D. Radiation response of floating gate EEPROM memory cells. IEEE Transactions on Nuclear Science 1989;36 2131-2139.
- [18] Cellere G., Paccagnella A., Visconti A., Bonanomi M., and Candelori A. Transient conductive path induced by a Single ion in 10 nm SiO₂ Layers. IEEE Transactions on Nuclear Science 2004;51 3304-3311.
- [19] Agostinelli S. et al. Geant4, a simulation toolkit. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 2003;506 250-303. See also <http://geant4.cern.ch>
- [20] Virtual Geometry Model (VGM), available online: <http://ivana.home.cern.ch/ivana/VGM.html>
- [21] ROOT, an object oriented framework for large scale data analysis. <http://root.cern.ch>.
- [22] Baumann R. and Smith E. Neutron-Induced ¹⁰B Fission as a Major Source of Soft Errors in High Density SRAMs. Microelectronics Reliability 2001;41 211-218.
- [23] JEDEC Standard “Measurement and Reporting of Alpha Particles and Terrestrial Cosmic Ray-Induced Soft Errors in Semiconductor Devices”, JESD89 Arlington, VA: JEDEC Solid State Technology Association, October 2006.
- [24] Gedion M., Wrobel F., Saigné F., Schrimpf R.D. Uranium and Thorium Contribution to Soft Error Rate in Advanced Technologies. IEEE Transactions on Nuclear Science 2011;58(3) 1098-1103.
- [25] Martinie S., Autran J.L., Munteanu D., Wrobel F., Gédion M., Saigné F. Analytical Modeling of Alpha-Particle Emission Rate at Wafer-Level. IEEE Transactions on Nuclear Science 2011;58(6) 2798-2803 2011.
- [26] Altitude Single Event Effects Test European Platform (ASTEP), www.astep.eu.
- [27] Lei F., “Quotid Atmospheric Radiation Model” – QARM (previously Qinetiq Atmospheric Radiation Model). Available online on request to the author.

- [28] Autran J.L., Serre S., Munteanu D., Martinie S., Semikh S., Sauze S., Uznanski S., Gasiot G., Roche P. Real-time Soft-Error testing of 40nm SRAMs. In: Proceedings of the IEEE International Reliability Physics Symposium 2012, pp. 3C.5.1 - 3C.5.9.
- [29] Puchner H. Correlation of Life Testing to Accelerated Soft Error Testing. In: Third Annual IEEE-SCV Soft Error Rate (SER) 2011, Workshop, San Jose, USA.
- [30] Kim J.H., Choi J.B. Long-term electron leakage mechanisms through ONO interpoly dielectric in stacked-gate EEPROM cells. IEEE Transactions on Electron Devices 2004;51(12) 2048-2053.
- [31] Pavan P., Larcher L., Marmiroli A. Floating Gate Devices: Operation and Compact Modeling. Kluwer Academic Press, 2004, Chapter 2.
- [32] Golla C., Zanoni E., Cappelletti P. Flash Memories. Kluwer Academic Press, 2011.
- [33] Van Houdt J., Degraeve R., Groeseneken G., Maes H.E. Physics of Flash Memories. In: "Nonvolatile Memory Technologies with Emphasis on Flash: A Comprehensive Guide to Understanding and Using NVM Devices", edited by J. E. Brewer, M. Gill, Wiley & Sons, 2007.