Politecnico di Torino

Politecnico di Torino

Master of Science Course in Materials Engineering for Industry 4.0
A.a. 2023/2024
Graduation Session October 2024

TCAD simulations of single transistor attacks through X-rays

Tutors:
Eng. ANCEAU Stephanie (CEA)
Eng. MAINGAULT Laurent (CEA)
Prof. SALVO Luc (G-INP)
Prof. RICCIARDI Carlo (PoliTo)

Candidate:
SOLAZZI Alessio

Summary

The continuous improvement of cybersecurity measures designed to protect devices from external attacks requires the exploration of new strategies capable of bypassing these protections. The "fault injection" methodology, which involves the systematic introduction of errors at the software or hardware level to extract sensitive information, has recently gained popularity due to its non-invasive approach. Among the various techniques available, hardware X-ray attacks have demonstrated promising results, particularly because of the possibility to focus the beam at the nanoscale enabling precise targeting of individual transistors.

This type of attack is currently under investigation at the Cybersecurity Department of the Commissariat à l'Énergie Atomique et aux énergies alternatives (CEA) in Grenoble, where this Thesis was conducted. Specifically, the Thesis focuses on the use of X-rays to alter specific transistors within Flash and SRAM memories and bypass their security measures. The types of X-rays considered are synchrotron radiation and radiation produced by a laboratory tomograph, both of which have demonstrated to be effective in previous studies. Synchrotron radiation can be precisely focused on individual transistors but is challenging to access due to its high cost. In contrast, tomograph radiation is naturally divergent, but by employing specialized masks it is possible to isolate the targeted transistors, enabling attacks using a more economical source.

The objective of this Thesis is to provide a detailed description of the electronic transport phenomena involved in transistor attacks using the ECORCE simulation software. A deeper understanding of the underlying mechanisms at the semiconductor level will enable the optimization of the parameters governing X-ray emission, making the attacks more effective.

To accurately model the transistors within the software, cross-sections were performed using a plasma FIB. This approach allowed to deeply understand the memory cells architecture and enabled precise measurement of the thicknesses of the material layers that constitute the transistors.

To determine the X-ray exposure time required to alter the logic state of transistors, experimental campaigns were conducted using both synchrotron and tomograph radiation. The synchrotron attacks were performed at the European Synchrotron Radiation Facility (ESRF) along the ID16B beamline, which provides a highly focused X-ray beam with a diameter of 60 nm. The second campaign took place at the Science et Ingénierie des Matériaux et des Procédés (SIMAP) laboratory, where a tomograph is available. For the tomograph-based attacks, a tungsten mask with a hole fabricated using a plasma FIB was placed over the transistors of interest. Additionally, the efficiency of an innovative system involving tungsten masks deposited directly on the memory regions of interest was evaluated.

A dedicated Python code was developed to calculate the dose absorbed by the memories during the experiments, accounting for the absorption by the various material layers observed in the cross-sections. The data obtained through this code were used as input parameters in the simulations to study the interaction of the two types of memory, and the associated transistors, with X-rays. The Flash memory simulations focused on the irradiation of a single "floating gate" MOS transistor, which features a suspended gate capable of accumulating electrons. For the SRAM memory, simulations were conducted by irradiating NMOS and PMOS transistors both individually and simultaneously.

Simulations of the electrical behavior of the floating gate transistor successfully modeled the process of electron injection into the suspended gate. By this way, it was possible to estimate a threshold voltage shift of approximately 1 V induced by carrier injection and to determine the density of accumulated electrons that cause the transition from logic state 1 (empty gate) to logic state 0 (full gate).

Tomograph radiation simulations highlighted that the primary mechanism altering the transistors state is the accumulation of positive charges trapped in the oxide layers. This is supported by the evolution of the electronic density in the gate which, as shown in **Figure I (left)**, remains relatively stable during the attack, allowing the exclusion of photoemission as the dominant mechanism. The results of these simulations were validated by heat treatments at 200 °C, which successfully restored the transistors to their initial state, confirming that electrons are not emitted during the attack but only shielded from the trapped charges.

In contrast, simulations with synchrotron radiation suggested that the emission of holes from the oxide layers toward the suspended gate is the main phenomenon responsible for altering the logic state. As depicted in **Figure I (right)**, during the attack the density of accumulated electrons significantly decreases due to recombination with the injected holes. However, it was not possible to experimentally confirm this result, as the memory's behavior is heavily influenced by the accumulation of positive charges in the oxide spacers separating the memory cells. These layers, observed through the FIB analysis, present a complex three-dimensional structure that ECORCE simulations cannot model.

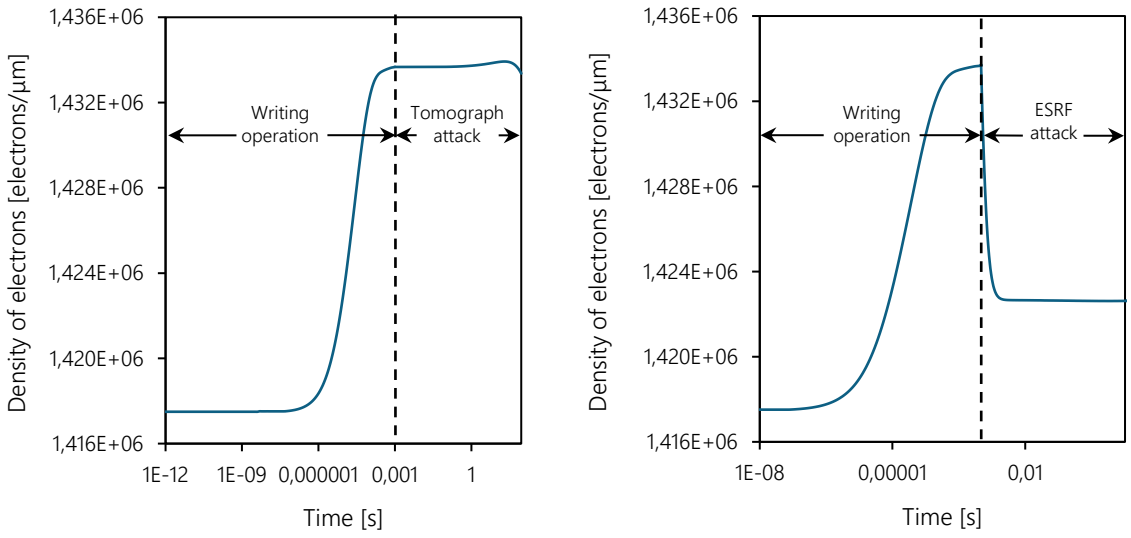

Figure I: density of electrons within the floating gate as a function of irradiation time in the case of tomograph radiation (left) and synchrotron radiation (right).

The simulations conducted on the NMOS and PMOS transistor pair allowed the determination of the optimal operating voltage for a single memory cell, which was estimated to be 2.5 V. This voltage guarantees the quickest transition from the high logic state (bit 1) to the low logic state (bit 0) in the transfer curve, which is the function that correlates the input potential with the output potential of the cell.

Irradiation simulations, using both tomograph and synchrotron radiation, identified positive charge accumulation in the gates as the primary mechanism responsible for altering the logic state in this type of memory. Both types of irradiation returned similar results: a downward shift in the maximum of the transfer curve when the device is in the ON state and an upward shift when it is in the OFF state. This result is shown in **Figure II**.

Finally, the simulations confirmed a prior experimental observation made by CEA researchers, indicating that a PMOS transistor is easier to fault than a NMOS transistor. The simulations showed that the cause of this difference relies on the different intensity of the electric field within the gates of the two transistors. Specifically, in NMOS transistors the higher electric field enhances the re-emission of trapped holes, making it more resistant to state alteration.

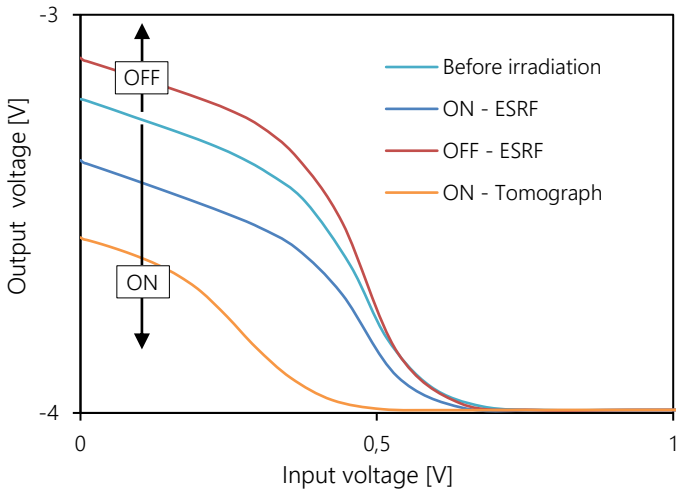

Figure II: shift of the transfer curve as a function of radiation nature and device state.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

In recent years, the security of integrated circuits (ICs) has significantly improved, making it increasingly necessary to explore new methods of attack. Fault injection techniques, which consists in the systematic introduction of errors through physical or software attacks to extract sensible information, are becoming increasingly popular. This rise in popularity is mainly due to their non-invasive nature, enabling attackers to exploit circuit vulnerabilities without physically damaging the device, thus minimizing the risk of detection.

In 2016, a team of researchers from CEA-Grenoble launched the MITIX (Modification non-Invasive de circuiTs Intégrés par rayons X) project to explore the feasibility of fault injection attacks on semiconductor memories using X-rays [1]. The primary goal is to exploit the well-known Total Ionizing Dose (TID) effects to alter the logical state of individual transistors, thereby enabling the retrieval of sensitive information stored in these memories. Their proximity to the European Synchrotron Radiation Facility (ESRF) allowed them to demonstrate the efficiency of nanofocused X-rays in altering the logical state of transistors inside a SRAM and a Flash memory [2]. Additionally, the collaboration with the SIMAP (Science et Ingénierie des MAtériaux et des Procédés) laboratory at INP-Grenoble provided access to a tomograph, allowing the researchers to demonstrate the feasibility of such attacks using a more economical and accessible laboratory source [3].

To conduct more systematic attacks, the MITIX team is willing to gain a deeper understanding of how X-rays interact with memory components at the semiconductor level. While existing research has mainly focused on the effects of space radiation on memories, there is a significant gap in knowledge about the interaction of focused X-rays with individual transistors, especially at very high doses.

To address this, the TCAD (Technology Computer Aided Desing) software ECORCE (Étude du COmportement sous Radiation des Composants Electroniques) was chosen to simulate the attacks on the SRAM and Flash memory performed using both synchrotron and tomograph radiation. This Thesis work aims to present the preliminary experiments conducted to obtain the input parameters for the simulations and explain the underlying mechanisms at play during X-ray attacks based on the simulation results. Additionally, the MITIX team is aiming to explore the effectiveness of X-ray attacks on DRAMs. Achieving this requires a deep understanding of the hardware architecture to design effective attacks. To support this, the Thesis work includes a detailed reverse engineering of a DRAM, providing insights about the architecture of the transistors within the memory.

The realization of this work was made possible through several advanced tools and collaborations. The beamline ID16B at the ESRF was used to generate the nanofocused X-ray beam, while the *EasyTom XL Ultra 230-160 micro/nano-CT scanner* from *RX Solutions* at SIMAP laboratory was exploited as the economical X-ray source. The reverse engineering experiments were conducted using the *Helios 5 DualBeam Plasma-FIB* from *ThermoFisher Scientific* at the PNFC (Plate-Forme de Nano-Caractérisation) of CEA-Leti.

1.1 Thesis structure

Following this introductory chapter, the Thesis is structured as:

- **Chapter 2** provides an overview of the current technologies used in semiconductor memories, describing the transistor architectures and storage mechanisms. **Section 2.1** focuses on MOS transistors and their use in digital memories, whereas **Section 2.2** discusses floating-gate transistors and their applications.
- **Chapter 3** gives an in-depth description of the interaction mechanisms between X-rays and transistors, as well as an overview of the X-ray sources used by the MITIX team for fault injection attacks. **Section 3.1** presents the equations that describe X-ray absorption in materials. **Section 3.2** addresses the interaction mechanisms of X-rays with single transistors, while **Section 3.3** describes the TID faulting mechanisms in memories. **Sections 3.4**, **3.5**, and **3.6** provide brief explanations of the working principles of the adopted X-ray sources, while **Section 3.7** reviews the state of the art in X-ray fault injection.
- **Chapter 4** describes the preliminary experiments required to perform simulations and the results obtained using the ECORCE software. **Section 4.1** describes the experimental procedures used to retrieve the input parameters for the simulations, while **Section 4.2** gives a brief description of ECORCE's working principles. **Section 4.3** presents the results of Flash memory simulations, while **Section 4.4** the SRAM results.
- **Chapter 5** focusses on the reverse engineering steps followed to study the hardware architecture of the DRAM. **Section 5.1** presents the chemical etching steps and their results, along with optical analysis, while in **Section 5.2** the images obtained using the PFIB are shown.
- **Chapter 6** provides an overview of the results presented in the previous chapters, presenting the conclusions of this work and offering suggestions for future work and potential software improvements.

Chapter 2

Semiconductor memories

Semiconductor memories are electronic components used to store digital information either in a permanent or volatile way. Silicon is commonly adopted to build them due to its well-established planar technology, which allows the lithography of a large number of transistors at the same time. Memories are structured as transistor matrices, allowing the storage of data in the form of bits. Depending on the adopted technology, two main types of transistors are used in memory cells: metal-oxide-semiconductor (MOS) transistors and floating-gate (FG) transistors.

2.1 MOS transistors and their applications

Metal-oxide-semiconductor field-effect transistors (MOSFETs) are employed in random-access memories (RAMs), volatile storage devices that lose the information when no power is supplied. This feature, along with the possibility of randomly addressing each cell, makes RAMs suitable for storing temporary data that are rapidly exchanged with the device's processor.

These unipolar devices owe their conductivity to one single type of charge carrier, which can be either electrons (n-carriers) or holes (p-carriers). In the former case they are called NMOS transistors, while in the latter PMOS transistors. A NMOS device (**Figure 1**) is built from a p-type silicon substrate where two n-doped regions, representing the source and drain contacts, are implanted. Between them, a conductive layer and an underlying oxide constitute the gate contact, while the opposite side of the substrate is called bulk or substrate contact. Highly doped polysilicon is generally employed as the conductive layer whereas thermal SiO_2 is the most common oxide. However, to limit leakages and improve the transistors' performances, high-κ dielectrics are nowadays adopted instead of thermal silica. The thickness of the oxide layer was about 100 nm in the first transistors, but it has decreased to a few nanometers nowadays [4]. PMOS transistors differ only for the inverted sign of the doping regions: the substrate is n-doped while the contacts are p-doped.

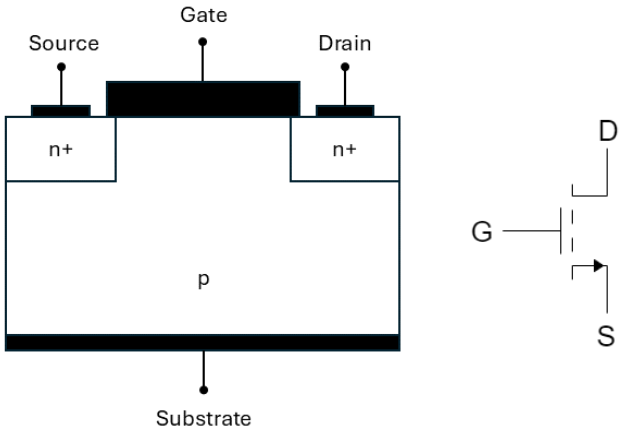

Figure 1: physical [5] (left) and symbolical (right) representation of a NMOS.

If source, drain, and bulk of a NMOS are grounded while a positive bias V_{GS} is applied to the gate, a depletion region beneath the oxide layer arises as the holes are carried away by the electric field. However, in this configuration no current can flow between the source and the drain because no conductive path connects them. When V_{GS} overcomes a characteristic threshold voltage V_{th}, a conductive path of electrons builds up for a few nanometers below the oxide layer. This phenomenon is called "population inversion" as the electrons density locally overcomes the p-doping concentration, inverting the sign of the main charge carrier.

In this configuration, if a positive bias V_{DS} is applied on the drain, electrons can migrate through the channel creating a current I_{DS} as shown in **Figure 2 (left)**. If V_{DS} is lower than $V_{GS} - V_{th}$, the current I_{DS} is directly proportional to V_{DS}, and this region is referred to as "ohmic region". For larger values instead, I_{DS} stabilizes to a constant value called "saturation current" as shown in **Figure 2 (right)**. Similar considerations apply for PMOS transistors with the only difference being the negative sign of the biases, thus the characteristic curves appear inverted.

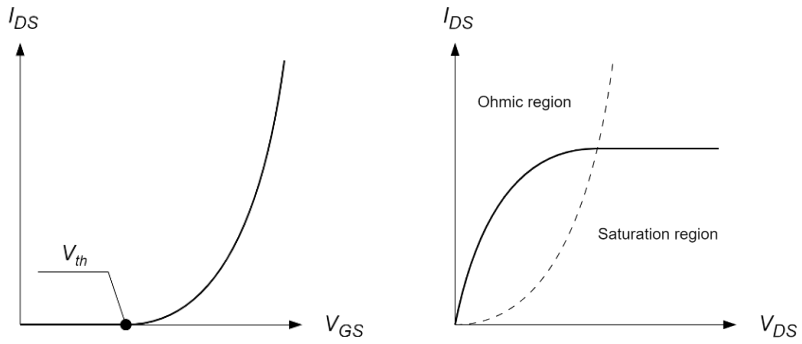

Figure 2: $I_{DS}(V_{GS})$ curve (left) and $I_{DS}(V_{DS})$ curve (right) [5].

The sequence of a PMOS and a NMOS with a common contact is called CMOS (Complementary Metal-Oxide-Semiconductor) technology. It forms the elementary unit used to build static random-access memories (SRAMs), storing devices able to hold the information without refreshing as long as the power is supplied. The circuit representation of a CMOS and its Voltage Transfer Characteristic (VTC) are reported in **Figure 3**, where LIN refers to the ohmic region and SAT to the saturation region. V_{DD} is the positive power supply voltage applied to the source of the PMOS whereas v_i and v_o are the input and output biases, respectively. The transistor connected to the power supply is a PMOS called pull-up (PU) transistor, while the other is a NMOS called pull-down (PD) transistor.

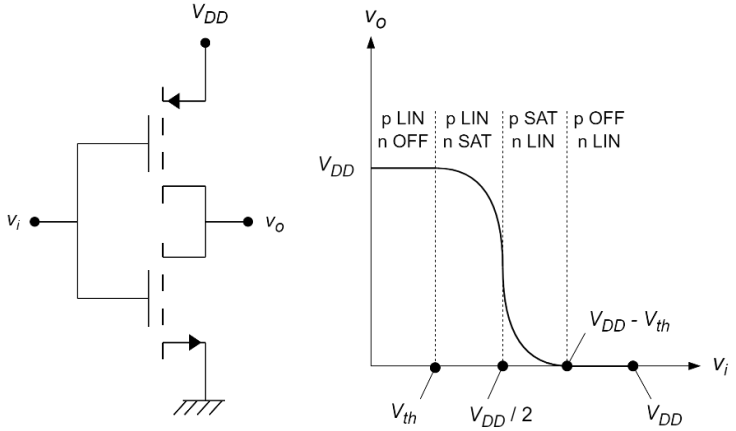

Figure 3: circuit representation of a CMOS (left) and its VTC (right) [6].

It is clearly visible that this system can operate in a high-state ($v_i < V_{DD}/2$) and a low-state ($v_i > V_{DD}/2$), therefore making it possible to represent a binary variable. An output voltage equal to V_{DD} enables the representation of the bit 1, while a 0 V output is used to represent the bit 0. The logical operator NOT embodies this behavior by correlating a low input value with a high output value and vice versa. For this reason, it is also referred to as "inverter".

The elementary cell of a SRAM features a two cross-coupled inverters architecture called "latch". The output nodes of the two inverters are called N1 and N2. They represent the storage nodes, as the logic value of the cell is directly related to their voltage. Two NMOS access transistors called pass-gate (PG) transistors are required to read and write the latch, thus making the number of transistors per cell equal to six. The access transistors are connected to two different bit-line, called BL and $\overline{BL}$, whereas their gates communicate with the same word-line WL. The architecture of a single SRAM cell is reported in **Figure 4**.

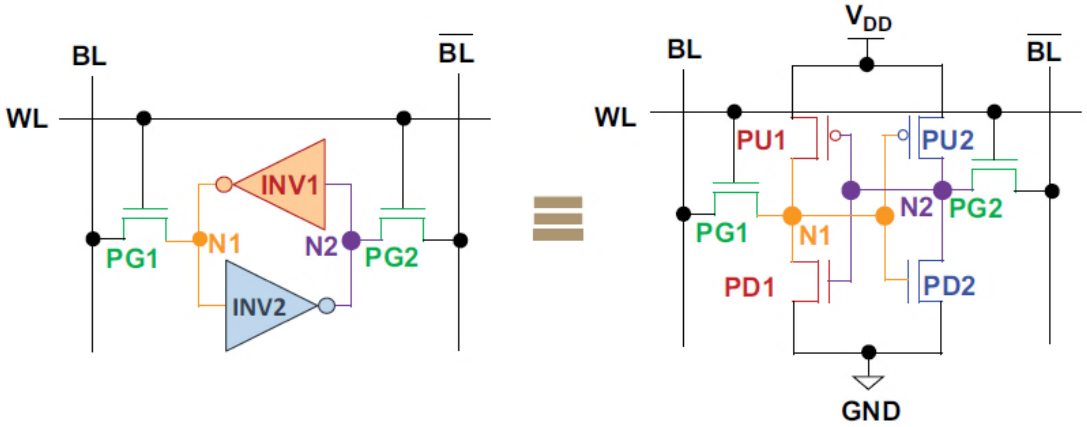

Figure 4: logic representation of a latch (left) and architecture of a SRAM cell (right) [7].

It is possible to perform three main operations in a single memory cell:

- **Holding**: the data is maintained unvaried between two reading/writing operations. To do so, the WL is grounded to turn off the control transistors, therefore isolating the nodes N1 and N2.
- **Reading**: the content of the storage nodes N1 and N2 is read. To do so, the BL and $\overline{BL}$ are pre-charged to V_{DD} while the WL is activated through a voltage pulse of amplitude V_{DD}. As previously shown in **Figure 3**, depending on the value of the voltages at the nodes N1 and N2 (hence, the digital value stored in the cell), a pull-up and a pull-down transistors result in the OFF state. A reading current is able to flow from one bit-line through the non-deactivated PD transistor to the ground, causing a small voltage drop along the interested bit-line. The difference in voltage between the two bit-lines is amplified by a sense amplifier and translated into the stored logic value.
- **Writing**: the content of the storage nodes N1 and N2 is flipped. To protect the memory cell from bit-flips, during write operations the first node to be flipped is the one storing the value 1. To do so, the corresponding bit-line is grounded and the other one is pre-charged to V_{DD}, while the WL is turned on through a voltage pulse of amplitude V_{DD}. This deactivates a pull-up and a pull-down transistor, allowing a current to flow from the WL through the node where 1 is stored up to the associated bit-line. The node experiences a voltage drop from V_{DD} to 0 V, resulting in a bit-flip from 1 to 0. Consequently, the output of the latch at the other node is inverted from 0 to 1.

MOS transistors are also used in dynamic random-access memories (DRAMs), volatile storing devices that require constant refreshing of the saved data. A single memory cell features an access transistor, controlled by a bit-line and a word-line, connected to a capacitor. The shared node between the transistor and the capacitor is called storage node. The charging state of the capacitor allows the representation of a binary variable, which corresponds to 1 when the capacitor is fully charged and 0 when it is depleted. When the stored bit is 1, positive charges arise at the interface between the external plate and the dielectric, while negative charges accumulate at the opposite interface. However, a periodic recharging of the capacitor is required due to the existence of leakage currents that depletes its charge content. According to industrial standards, cells must be refreshed every 64 ms [8]. A schematic representation of a DRAM cell is reported in **Figure 5 (left)**.

As visible from the schematic cross-section of a DRAM reported in **Figure 5 (right)**, the capacitor is generally processed vertically to increase the exposed surface and therefore its capacity. For this reason, DRAMs generally feature higher cell densities compared to SRAMs, making them suitable to be used as main volatile memories in digital devices. The conductive plates of the capacitor are usually made of polysilicon, while the dielectric can either be an oxide-nitride or a high-κ oxide.

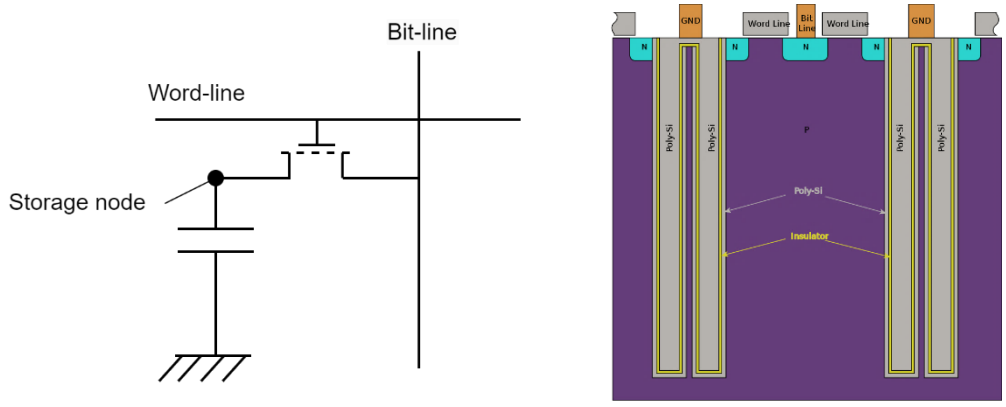

Figure 5: circuit scheme of a single DRAM cell (left) and cross-section of a DRAM cell (right) [7].

The same aforementioned operations for SRAMs can be performed also in DRAM cells:

- **Holding**: the charging state of the capacitor is maintained by deactivating the WL.
- **Reading**: the state of the capacitor is probed by measuring the potential at the storage node, which can be either V_{DD} if the capacitor is charged or 0 V if it is depleted.
- **Writing**: the WL is turned on with a voltage $V_{WL} > V_{DD}$ to allow the flow of a current through the transistor. The BL is either grounded or brought to V_{DD} when the capacitor is discharged or recharged, respectively.

2.2 FG transistors and their applications

Floating-gate transistors are a particular type of MOS transistors which feature an additional gate buried between two insulating layers, as shown in **Figure 6**. The upper gate is called "control gate" while the buried gate is referred to as "floating gate" due to its suspended position. These gates are commonly made of highly doped polysilicon. The oxide layer close to the silicon substrate (called "tunnel oxide") is made of thermal SiO_2 while the upper insulator (called "inter-poly dielectric") can be made of either SiO_2 or $SiO_2/Si_3N_4/SiO_2$ in modern devices.

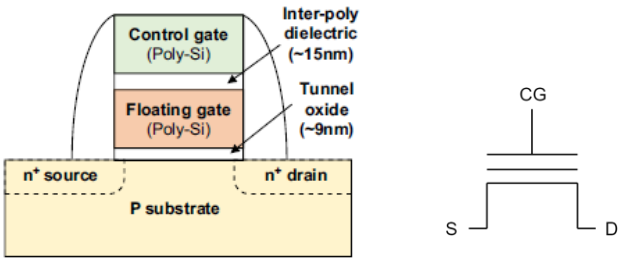

Figure 6: schematic representation of an FG transistor (left) and its circuit symbol (right) [7].

Floating-gate transistors have the unique ability to store electric charges, generally electrons, injected from the substrate within the embedded layer. Moreover, if no voltage is applied, the density of injected charges remains constant over time, as the dielectric layers prevent carriers from leaking. These charges partially screen the bias applied on the control gate, resulting in a shift to the right of the $I_{DS}(V_{GS})$ curve as shown in **Figure 7**. Thus, two threshold voltages can be defined depending on the charging state of the transistor.

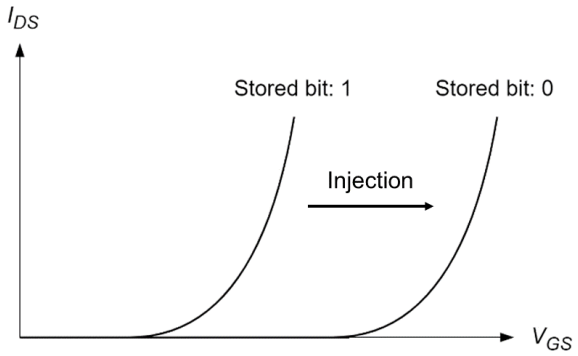

Figure 7: shift of the $I_{DS}(V_{GS})$ curve when the floating gate is charged.

Therefore, it is possible to identify two separate states: the programmed state where the floating gate is permanently charged, and the erased state where the electrons have been removed. This feature makes it possible to represent a non-volatile binary variable that can be used to store information in the form of bits. If a reading voltage V_R produces a relevant current it means that the gate is empty, and the logic value will be 1. In this condition, the transistor is said to be erased. On the other hand, if no current flows between the contacts, it means that $V_R < V_{th}$ and the gate is charged; thus, the logic value 0 will be returned and the transistor is said to be programmed.

Two main mechanisms can be exploited to program a floating gate transistor:

- **Channel hot electrons (CHE) mechanism**: a high voltage up to 12 V is applied on the control gate, a lower voltage from 4 to 6 V is applied on the drain and the source is grounded. In this configuration a portion of the accelerated electrons have enough kinetic energy to overcome the tunnel oxide barrier, therefore getting trapped within the floating gate.
- **Fowler-Nordheim mechanism**: a high voltage ranging from 10 to 20 V is applied on the control gate, and all the other contacts are grounded. The intense electric field perpendicular to the oxide attracts the electrons that tunnel through the SiO_2 layer, thus accumulating in the floating gate. This mechanism differs from direct tunneling in the shape of the energetic barrier, which is triangular in this case due to the high voltage and trapezoidal in the latter case.

These peculiarities make floating-gate transistors suitable for Non-Volatile Memories (NVMs) such as EEPROMs and Flash memories, digital storage devices able to maintain the data even if the power supply is removed. EEPROMs (Electrically Erasable Programmable Read Only Memories) have been the first NVMs enabling the user to perform erasing operations through electric signals. Flash memories are the successors of EEPROMs, as they feature higher storage densities and faster erasing operations. However, EEPROMs are still widely adopted in microcontrollers, electronic devices used as microprocessors in embedded systems. Since in this Thesis work a microcontroller was taken under analysis, a special focus will be given to the description of writing and erasing operations in EEPROMs.

The schematic of an EEPROM single cell is reported in **Figure 8 (left)**. It features a floating-gate transistor that is used to store the bit value, and an access transistor that enables its selection, thus allowing single-bit modifications. One side of the access transistor is connected to the BL, its gate to a select-line and the control gate to a WL. The floating-gate transistors employed in EEPROMs have a peculiar structure called FLOTOX (FLOating gate Thin OXide), showed in **Figure 8 (right)**. These systems feature a tunnel oxide with a reduced thickness region (8-10 nm) near the drain to facilitate the injection and removal of electrons from the floating gate. Usually, the Fowler-Nordheim mechanism is exploited to change the state of FLOTOX devices.

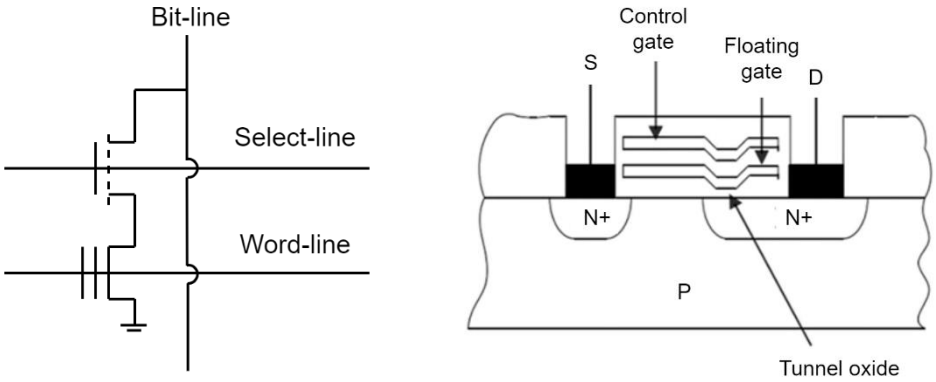

Figure 8: schematic of a EEPROM cell (left) and cross-section of a FLOTOX transistor (right) [9].

The operations that can be performed in FLOTOX EEPROMs are:

- **Writing**: the chosen WL and select-line are raised to a high voltage (e.g. 10 V) while the BL is grounded. In this configuration, the access transistor allows an intense current to flow through the thin oxide to charge the floating gate. Fowler-Nordheim writing time typically ranges from 100 μs to 1 ms.
- **Erasing**: the WL is grounded while the select-line and the BL are brought to a high voltage up to 20 V. The stored electrons are removed through an intense Fowler-Nordheim current; typical duration ranges from 100 μs to 1 ms.
- **Reading**: the BL is raised to a low voltage (e.g 2 V), the access transistor is activated through a low voltage and the WL is brought to a reading voltage. Depending on the charging state of the transistor, if a current is measured at the drain the associated logic value will be 1, otherwise 0. Reading operations are usually the fastest, commonly taking some tens of ns.

Chapter 3

X-rays/matter interaction

X-rays belong to the portion of the electromagnetic spectrum ranging from 100 eV to 100 keV, resulting in characteristic wavelengths between 10 nm and 10 pm. X-rays are ionizing radiations because the collision of X-ray photons with electrons eject them from their orbitals, therefore ionizing the atoms. The most common applications of X-rays rely on this property, ranging from medical radiographies to materials characterization. However, a new application emerging is the physical manipulation of integrated circuits and memories to alter their content using X-rays.

3.1 X-rays absorption in materials

Depending on the energy of the incident photons, three main interaction modes of matter with photons can be identified [10]:

- **Photoelectric effect**: this interaction happens when a photon possesses enough energy to overcome the binding energy of an inner shell electron, ejecting it from its orbital. This phenomenon is predominant for high-Z materials and occurs at low incident energies, generally below 0.5 MeV.
- **Compton scattering**: it represents the elastic collision of a high-energy photon, generally between 0.5 and 1 MeV, with an outer shell electron. This interaction results in the ejection of the electron and a decrease in the photon's frequency.
- **Pairs production**: this interaction is the main mechanism for higher energies and consists in the conversion of an incident photon into an electron and a positron.

The linear attenuation coefficient μ groups the effects of these mechanisms and is dependent on both the material and the X-ray wavelength. It allows the determination of the intensity of an X-ray beam travelling through a medium using the Lambert-Beer law:

$$I(x, \lambda) = I_0(\lambda) \cdot e^{-\mu(\lambda) \cdot x} \qquad \textbf{Eq. 1}$$

Here, λ is the wavelength of the radiation, I_0 is the incident beam's intensity expressed as the number of photons per unit of time and space, μ is the linear attenuation coefficient, x is the distance crossed by the beam within the medium and $I(x)$ is the intensity of the beam at that distance. If the incident beam is polychromatic, this law applies to each wavelength of the spectrum.

The linear attenuation coefficient is not adapted to calculate the absorbed energy from the material because it does not account for radiative losses. The coefficient that considers these secondary phenomena is the linear energy absorption coefficient μ_{en}. Radiative losses carry out energy from the material, therefore reducing the actual amount of absorbed energy.

Radiative losses include:

- **Photons scattered by Compton interactions**: a fraction of these photons escape the material without further interaction.
- **Bremsstrahlung radiation**: this radiation is produced by the deceleration of electrons as they pass near the nuclei of atoms.
- **Radiative energy release**: this occurs when electrons fill inner shells following photoelectric interactions.

The amount of absorbed energy per unit mass following an X-ray interaction is called absorbed dose or total ionizing dose (TID) and the SI unit is the gray (Gy). However, the historical unit rad (radiation absorbed dose) is still widely used in the scientific community. The relationship between the gray and rad is:

$$1\ \text{Gy} = 1\ \text{J/kg} = 100\ \text{rad} \qquad \textbf{Eq. 2}$$

The rate at which the material absorbs energy from the incident X-ray photons is called dose rate, typically expressed in rad/s. It is possible to evaluate the dose rate absorbed by a thin slab, such as a MOSFET's gate oxide, when a monochromatic beam impinges on it through the following equation [11]:

$$\dot{D} = I_0 \frac{\mu_{en}}{\rho} h\nu \qquad \textbf{Eq. 3}$$

Where ρ is the medium's density, h the Planck constant and ν the photons frequency. The absorbed dose D can be calculated by multiplying the dose rate with the exposure time.

3.2 X-rays interaction with MOS transistors

X-rays have been a central topic of study within the satellites and spacecrafts community, as they are part of the so-called space radiation. This term is used to group all the ionizing sources permeating the universe: mainly electrons, protons, heavy ions and various radiations. These energic particles interact with electronic components used in satellites, degrading their performances and potentially causing loss of functionality. Much of the research has focused on the detrimental effects of ionizing particles on CMOS devices, as it is the most adopted technology in integrated circuits [12], [13].

The TID effect of X-rays on CMOS devices is mostly associated with the buildup of trapped charges within the insulating oxide layers, especially in the gate oxides transistors when a positive (NMOS) or a negative (PMOS) bias V_{GS} is applied. The damaging mechanism can be divided into four steps, each occurring on different time scales: generation of electron-hole pairs, holes diffusion towards the Si/SiO_2 interface, near-interface deep trapping and formation of interface traps [4], [14]. These steps are summarized in **Figure 9**, where they are located with respect to the MOS energy band diagram.

Generation of electron-hole pairs: following X-ray exposure, electron-hole pairs are generated within the SiO_2 thin layer. However, depending on the electric field and the photons' energy, only a fraction of them known as "charge yield" survive to recombination. The remaining electrons are quickly extracted from the oxide, while the holes tend to remain near their generation points due to their significantly lower mobility (on average, μ_n = 20 $cm^2V^{-1}s^{-1}$ and μ_p < 1 $cm^2V^{-1}s^{-1}$ [15]).

Therefore, a positive charge builds up in the oxide, leading to a negative shift in the threshold voltage: NMOS transistors become more conductive (**Figure 10 (left)**), whereas PMOS become less conductive (**Figure 10 (right)**). This shift occurs within the first picoseconds after the interaction.

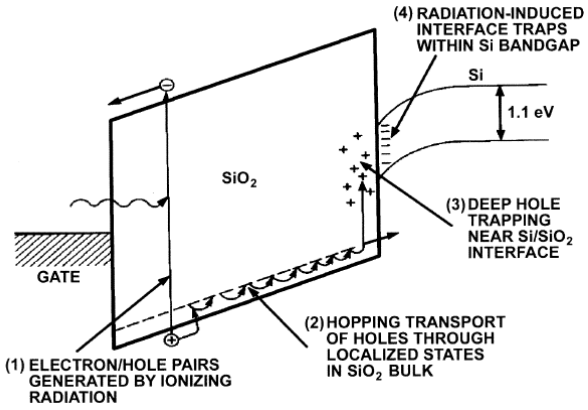

Figure 9: four-step interaction of X-rays with a MOS transistor [4].

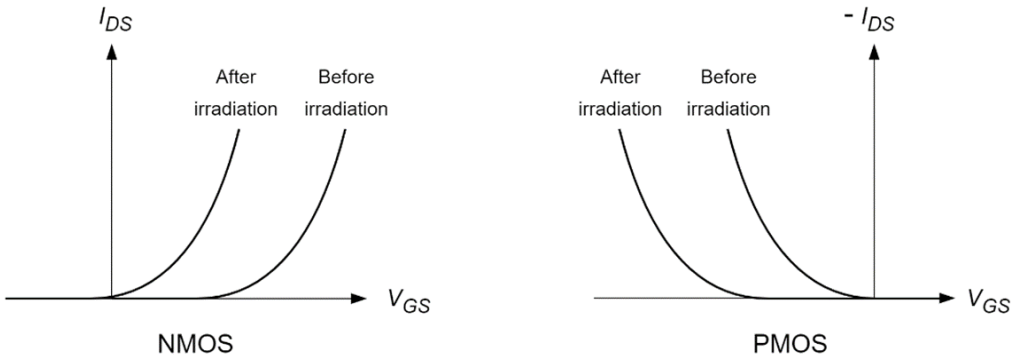

Figure 10: negative threshold voltage shift in a NMOS (left) and a PMOS (right).

Holes diffusion towards the Si/SiO_2 interface: the limited mobility of holes is tightly linked to their diffusion mechanism, which is explained by the Multiple Trapping Detrapping (MTD) theory [16], [17]. According to this model, the amorphous structure of SiO_2 give rise to a decreasing distribution of localized states within the band gap [18]. When an external electric field is applied, the holes moving in the valence band can randomly get trapped by these localized states, thus reducing the overall mobility (mechanism 1-2). Holes can escape from such traps through thermal excitation or tunneling, with their frequency of emission that decreases with the traps' energetic depths (mechanism 3) [19]. Therefore, detrapping can be enhanced either by increasing the temperature or the intensity of the electric field (Poole-Frenkel effect), making this phenomenon "semi-permanent".

Additionally, some holes recombine with electrons or are collected by the substrate, thereby reducing the amount of positive charge accumulated in the oxide (mechanism 4). The probabilities of trapping and recombination are encompassed respectively by the trapping cross-section σ_{tp} and the recombination cross-section σ_{rp}, which are inversely proportional to the intensity of the electric field [16], [19]. **Figure 11** provides an overview of the described mechanism (1 to 4) following X-ray exposure, where E_{pi} and E_{nj} denote the i and j energy levels where holes and electrons can get trapped, respectively. The recombination of holes and their collection by the substrate reduce the magnitude of the threshold voltage shift, which is proportional to the accumulated positive charge. This recovery happens in less than one second at room temperature and for this reason is called "short-term" recovery.

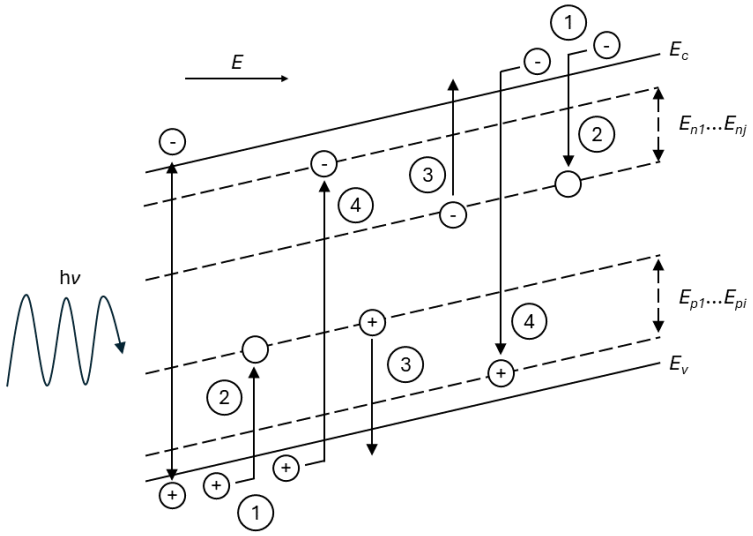

Figure 11: overview of the main phenomena at play following X-ray exposure [16].

Near-interface deep trapping: the transition region from SiO_2 to Si is rich of oxygen vacancies originated from the incomplete oxidation of the substrate, and they represent the main source of hole trapping in thermal oxides. These vacancies are most concentrated at the SiO_2/Si interface and decays exponentially with the distance from this region. Holes trapping in such sites happens follows the mechanism presented in **Figure 12 (a)** and **(b)**, which permits the relaxation of the strained Si-Si bonds. One silicon atom acquires a positive charge and relaxes into a planar configuration, while the other one exposes a dangling bond but retains the tetrahedral configuration. These traps feature deep energetic levels (> 3 eV [15]), therefore detrapping can happen after hours or years and the recovery is said to be "long-term". The proposed mechanism shown in **Figure 12 (c)** explains detrapping through the tunneling of an electron from the substrate to the tetrahedral silicon. The resulting dipole structure can either transiently return to the asymmetric configuration by losing an electron or reform the Si-Si bond, therefore ejecting the hole [20]. The probability of a tunneling electron from the substrate reaching a broken Si-Si bond decreases with distance from the SiO_2/Si surface.

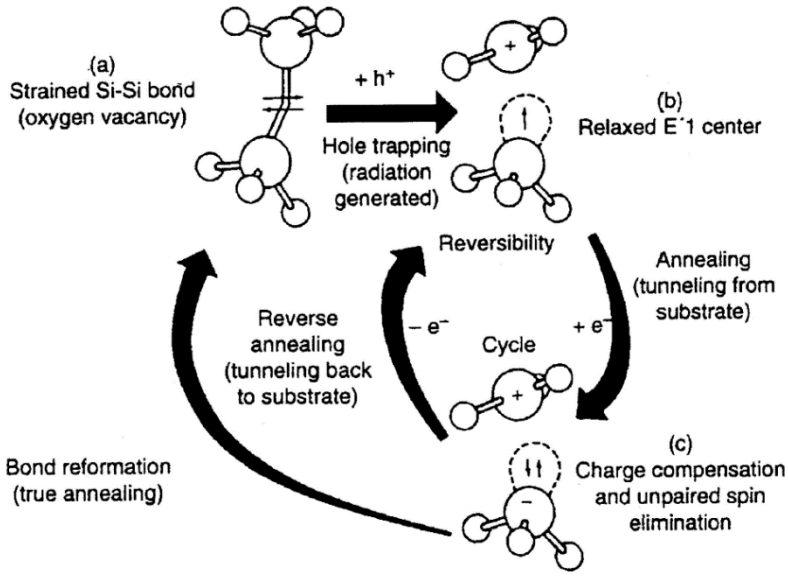

Figure 12: mechanism for hole trapping and detrapping in oxygen vacancies [4].

Buildup of interface traps: following thermal oxidation, more than 10^{13} Si atoms per cm^3 feature a Si-H bond [4]; these sites are considered responsible for the buildup of interface traps after X-ray exposure. Holes can break such bonds, leaving a dangling bond on Si atoms and releasing H^+ ions which can diffuse to the SiO_2/Si interface under the influence of an external positive bias. Protons react with other Si-H bonds at the interface, releasing H_2 and leaving positively charged Si atoms [21]. This defect has an amphoteric behavior, meaning that it can be neutral, positively charged or negatively charged depending on the bias conditions and the doping type of the MOSFET. These phenomena occur on a timescale ranging from hours to years. When $V_{GS} > 0$, in NMOS transistors such states acquire electrons from the substrate becoming negatively charged, vice versa for PMOS transistors. Therefore, NMOS transistors experience a further reduction in the threshold voltage shift, which can eventually become positive. On the other hand, the presence of additional positive charges in PMOS transistors increase the threshold voltage shift.

It is worth noting that the continuous scaling down of transistor dimensions has reduced the effect of charge trapping in gate oxides. In fact, the threshold voltage shift is proportional to the density of trapped charges, which decreases with the thickness of the gate oxide. On the other hand, STIs (Shallow Trench Isolations) have progressively become the main source of hole trapping. The positive charges trapped in field oxides can create an intense electric field, inducing a population inversion at the interface with the substrate. As shown in **Figure 13**, in NMOS transistors two new conductive paths connecting the source to the drain can arise, therefore creating leakage currents in the device. These parasitic currents can increase the static power consumption and eventually lead to a loss of functionality, as they can flow even when the device is turned off. On the other hand, in PMOS transistors the accumulation of electrons at the STI/substrate interface inhibits the creation of a conductive path for holes, thereby not affecting their functioning [22].

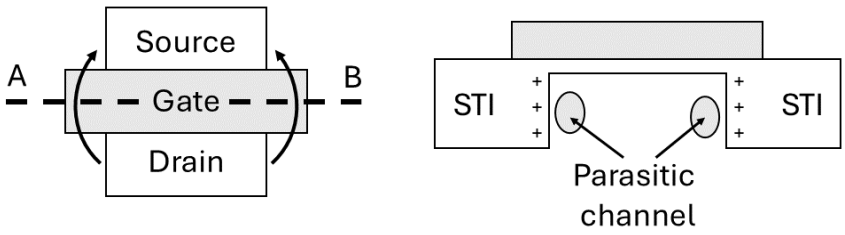

Figure 13: direction of the leakage currents seen from a top view (left) and cross-section of the transistor along the dashed line (right) [23].

3.3 X-rays interaction with memories

3.3.1 SRAM fault

The threshold voltage shift that MOS transistors experience when irradiated by X-rays can have a detrimental effect on memories' content. This interaction changes the electrical behavior of these devices, making them more conductive or more resistive, therefore altering the stored bits. For a NMOS, the accumulation of positive charges in the oxides can switch its state from OFF ($V_{GS} < V_{th}$ and $I_{DS} \sim 0$ A) to ON ($V_{GS} > V_{th}$ and $I_{DS} > 0$ A), making it semi-permanently conductive. On the other hand, a PMOS can switch from the ON state to the OFF state, becoming semi-permanently blocked.

When a SRAM memory cell is irradiated by X-rays, the TID effect can alter either a PD-NMOS in the OFF state or a PU-PMOS in the ON state. As shown in **Figure 14**, when a PD-NMOS is switched from OFF to ON following irradiation, the latch mechanism flips the logical values stored in the N1 and N2 nodes. The new configuration of the cell remains unchanged until the trapped holes in the PD gate are completely annealed, therefore making it impossible to restore the original values in the cell through a writing operation [24].

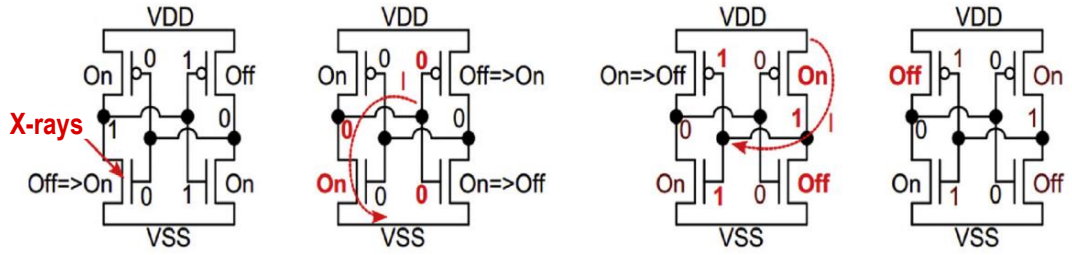

Figure 14: PD transistor faulting mechanism in a SRAM cell [24].

3.3.2 DRAM fault

Charge leakages following X-rays exposure is the main faulting mechanism for DRAMs. Discharge currents can arise if the TID effect produces a sufficient threshold voltage shift of the NMOS access transistor. The transistor becomes semi-permanently conductive even if no bias is applied to the WL, allowing current to flow beneath the gate. Additionally, the ionization of the capacitor's oxide can increase the density of interfacial traps. It has been demonstrated that such sites can enhance trap-assisted tunneling (TAT) phenomena [25], a tunneling mechanism that exploits the capture and emission of an electron in a trap within the energy barrier (**Figure 15 (left)**). This allows electrons stored at the inner interface of the capacitor to leak out, reducing the charge content and eventually flipping the bit value from 1 to 0. The capacitor will continue to leak until the interface traps are not annealed.

3.3.3 EEPROM and Flash fault

Finally, X-rays can interact with FG transistors through three main mechanisms depicted in **Figure 15 (right)** [26]. Electron-hole pairs are generated both in the tunnel and the inter-poly oxides, and they are separated thanks to the electric field generated by the stored charges. The holes move towards the floating gate, and they can either be injected within the floating gate (1) or get trapped in the oxide (2). In the former case, the holes recombine with the stored electrons, while in the latter case the positive charges trapped within the oxide mask the stored negative charges. Moreover, X-rays can transfer enough energy to the stored electron to expel them from the floating gate (3). All these mechanisms concur to the reduction of the net negative charge stored in the floating gate, shifting the $I_{DS}(V_{GS})$ curve to the left. If the effects (2) and (3) prevails, the cell experiences a reduction of the stored charge as if it was electrically erased, and it can be reprogrammed. If mechanism (1) is dominant, the shift of the characteristic curve is semi-permanent, and the logic state of the cell gets stuck until the trapped charges are not completely annealed.

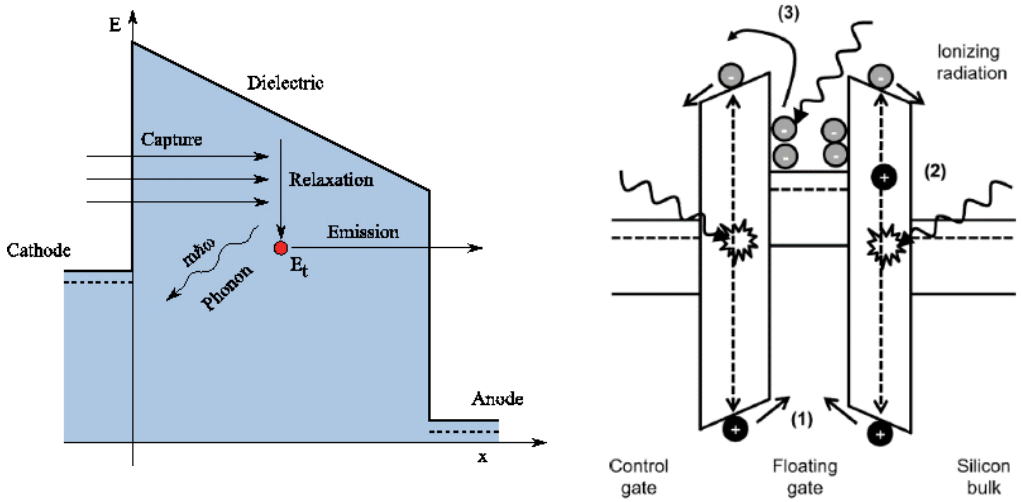

Figure 15: trap-assisted tunneling mechanism (left) and X-ray interaction mechanisms with FG transistors (right).

3.4 X-rays exploitation for cyberattacks

The spacecrafts and satellites community have extensively studied the presented effects to improve the resilience of ICs and memories during space missions. However, since the discovery that laser radiation can reproduce TID effects in ICs [27], the cybersecurity community progressively started to exploit such mechanisms to design new means of attack. Lasers, and subsequently X-rays, have been adopted as a semi-invasive alternative to traditional fault injection techniques. This family of attacks, which can be performed at either a software or hardware level, aims to exploit external stimuli to overcome countermeasures and retrieve sensible data.

Laser radiation was initially adopted due to the possibility of focusing the beam at a micrometric scale, even with commercial laser pointers [28]. This feature was exploited to demonstrate for the first time the feasibility of single-bit attacks, enabling the alteration of single SRAM cells and the retrieval of their logical address. It is important to note that the faulting mechanism caused by laser irradiation is entirely different from the mechanisms presented in the previous sections. In fact, the ionization of the oxide can only be achieved with energetic radiation such as X-rays. The energy of the photons produced by a laser (typically in the IR/visible region) is generally lower than the band gap of silica (E_g of SiO_2 = 9 eV), therefore absorption happens only in the silicon substrate. Transistor faulting is caused by an induced photocurrent generated at the drain/substrate interface, which is a reverse biased pn junction [29].

Lasers feature some drawbacks that make them unsuitable for systematic attacks [2], especially for new technologies:

- **Transient faults**: due to the intrinsic faulting mechanism previously explained, faults last as long as the transistors are irradiated and disappear once the attack is finished. Therefore, the state of the transistors cannot be permanently altered.
- **Low penetration depth**: laser radiation is absorbed by the external package of the IC, therefore depackaging is mandatory to perform an effective attack. Moreover, due to this limitation it is not possible to obtain indirect information about the internal architecture of the circuit, necessitating FIB (Focused Ion Beam) deprocessing for this purpose.

- **Resolution limit**: due to the Abbe diffraction limit [30], which states that the minimum resolvable distance is proportional to the radiation wavelength, it is impossible to focus laser radiation below the micrometric scale. This poses a limitation to the minimum technology node of the transistors that can be attacked, therefore making this technique unsuitable for new technologies.

X-rays allow to overcome these disadvantages, as they can semi-permanently fault transistors, avoid depackaging due to their high penetration depths and enable focusing down to the single transistor scale. For these reasons, X-rays represent nowadays a valid alternative to lasers as a mean of circuit perturbation, enabling single transistor faulting. In the presented work, two types of X-ray sources have been exploited to perform cyberattacks. A brief description of their working principles is given in the following section.

3.5 Synchrotron radiation

Synchrotrons are particle accelerators used to produce collimated X-rays, known as synchrotron radiation. Synchrotrons exploit the emission of electromagnetic radiation by accelerated charged particles, generally electrons. These particles are accelerated up to relativistic speeds inside a toroid-like structure called storage ring, where ultra-high vacuum is maintained to facilitate this process. Electrons' trajectories are kept circular through bending magnets, which are dipole magnets producing a magnetic field perpendicular to the particle's velocity. The Lorentz force acting on the particles results in a centrifugal acceleration, leading to the emission of radiation known as bending magnet radiation (**Figure 16 (left)**) [31]. As shown in **Figure 16 (right)**, as the ratio v/c (velocity of the particle over the speed of light) becomes closer to 1, the angular distribution of the emitted radiation aligns with the particle's velocity. Therefore, bending magnet radiation is emitted tangentially to the storage ring [32].

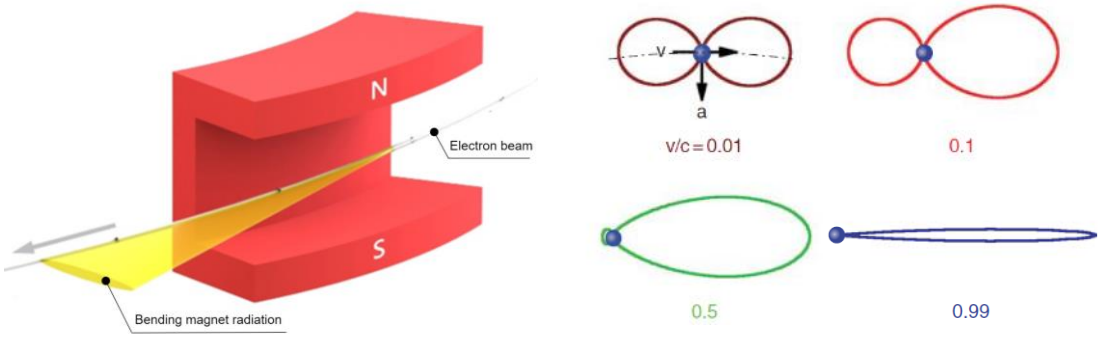

Figure 16: bending magnet radiation (left) [33] and angular distribution of the emitted radiation from an accelerated particle as a function of the v/c ratio (right) [32].

The radiation spectrum produced by electrons passing through bending magnets is broad. To optimize it, magnets with alternating polarity are placed along the storage ring to produce rapid changes in the particle's acceleration, resulting in the emission of additional radiation. These parts are called insertion devices (IDs), and they are classified as either wigglers or undulators based on the magnet arrangement. Wigglers intensify the flux and increase the beam's energy, while undulators focus the beam and make it monochromatic. Each ID has an operational station called a beamline, which runs tangentially to the storage ring [32].

In the presented work, the European Synchrotron Radiation Facility (ESRF) in Grenoble, France, was utilized to produce nanofocused X-rays for the purpose of semi-permanently faulting single transistors. To achieve this, a beamline along an undulator was necessary to produce a coherent nanometric beam. The ID16B beamline was chosen, as it can produce a monochromatic beam with a diameter of 60 nm and a flux up to 10^{11} ph/s. The nanofocused beam is produced through a complex optics system which demagnifies the beam coming from the storage ring [34].

3.6 Laboratory tomograph radiation

X-ray tomography is an imaging technique widely adopted in medicine as well as in material science to visualize the inner details of bodies and objects. This instrument can also be exploited as a practical X-ray source, as its spectrum can be easily tuned by controlling a restricted number of parameters. The radiation is produced through an X-ray tube, an apparatus that accelerates thermionic electrons towards a target, which emits X-rays afterwards. X-rays are produced either as braking radiation or following the neutralization of inner vacancies from outer shell electrons [35].

Reflective and transmission targets are generally adopted in laboratory tomography and their schematic representation is reported in **Figure 17**. Only a little fraction of the electrons interacts to produce X-rays, whereas 99% of the energy is dissipated as heat. For this reason, targets need to be in contact with a conductive material such as copper. Diamond can be used as a substrate for the target in order to enhance heat conductivity.

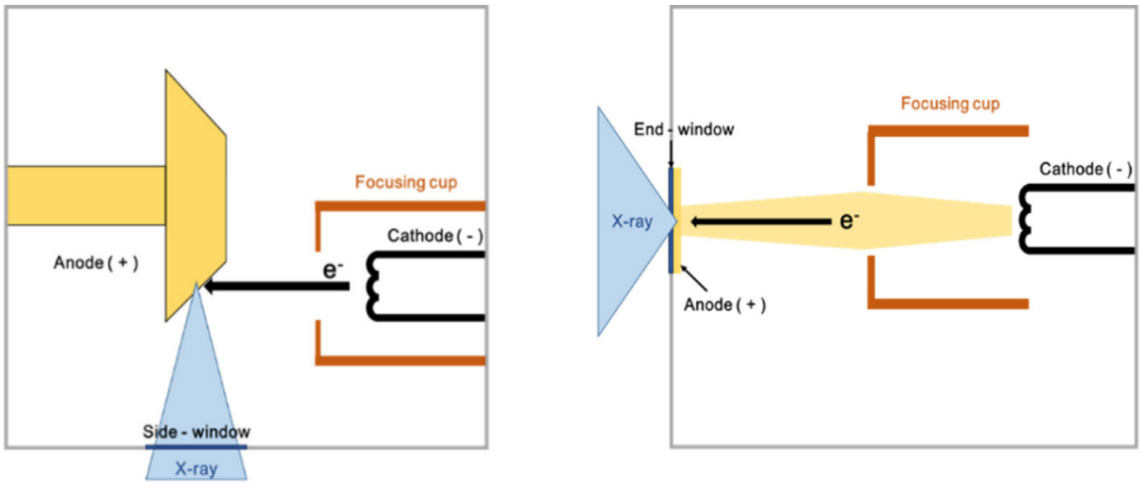

Figure 17: schematic representation of an X-ray tube with a reflective target (left) and a transmission target (right) [36].

The shape of the final spectrum can be adjusted by tuning the voltage used to accelerate the thermionic electrons, selecting different target materials, and placing a filter in front of the emission spot. Moreover, the intensity of the spectrum can be regulated by varying the current. These aspects, together with the possibility of using movable masks to focalize the beam in a precise spot, make laboratory tomographs an economic alternative to synchrotrons to perform cyberattacks on single transistors.

3.7 State of the art

In 2016, the MITIX group of researchers from CEA-Grenoble successfully demonstrated for the first time the feasibility of single-bit alteration using nanofocused X-rays [2]. The attack was performed through the ID16B beamline on the ATmega1284P, a microcontroller with a 350 nm technology node, 128 kB of Flash memory, 4 kB of EEPROM and 4 kB of SRAM. This experiment proved the efficiency of localized X-rays attack as a new fault injection technique to bypass PIN authentication programs. Since then, the researchers have shifted their focus on using a laboratory tomograph as a more accessible X-ray source and exploring the feasibility of these attacks on lower technology nodes.

3.7.1 ESRF attack

The X-ray beam available at the ID16B beamline enables the production of a sufficient number of electron-hole pairs in silica to fault single transistors in a few seconds. The beamline is equipped with an X-ray Fluorescence (XRF) spectroscope. This apparatus provides a 2D map of the distribution of elements detected in the sample at a chosen depth. Therefore, it can be used to localize tungsten, which is the most adopted material for vias at the transistor level. If the GDSII (Graphic Desing System) file of the IC is known, it is possible to precisely localize the transistors with respect to the vias. Additionally, due to the high penetration depth of X-rays, this analysis can be conducted without removing the packaging, making it suitable for attacking circuits without visible traces.

Since the GDSII file of ATmega1284P is covered by the industrial secret, the CEA researchers performed a preliminary deprocessing to obtain a precise localization of the SRAM and Flash memory transistors. Through wet etching, the plastic package, the metals and the oxide layers were removed revealing the polysilicon grids. This approach allowed the researchers to superimpose the fluorescence map of tungsten with SEM pictures of the etched memories, enabling the precise localization of the transistor contacts.

As an example, in **Figure 18** are shown from the left the SEM picture of the etched SRAM, the tungsten map obtained through XRF spectroscopy and their superimposition. These information were exploited to precisely localize the nanofocused X-ray beam above single transistors of the two memories. The researchers successfully managed to corrupt the content of single SRAM cells through the mechanism explained in **Section 3.3.1**. Additionally, by exploiting the phenomena presented in **Section 3.3.3**, the erasure of single FG transistors was also achieved. Through this procedure, it was possible to retrieve the relation between the physical coordinates of the attacked transistors and the logical address.

The knowledge of the memory map is fundamental to perform systematic cyberattacks. The CEA researchers managed to corrupt an authentication code stored in the Flash memory by irradiating a specific floating gate transistor. This cell was identified through the memory map, and its alteration allowed to bypass a four-digit PIN requirement, making the program accept 9999 wrong PINs and reject the initial one.

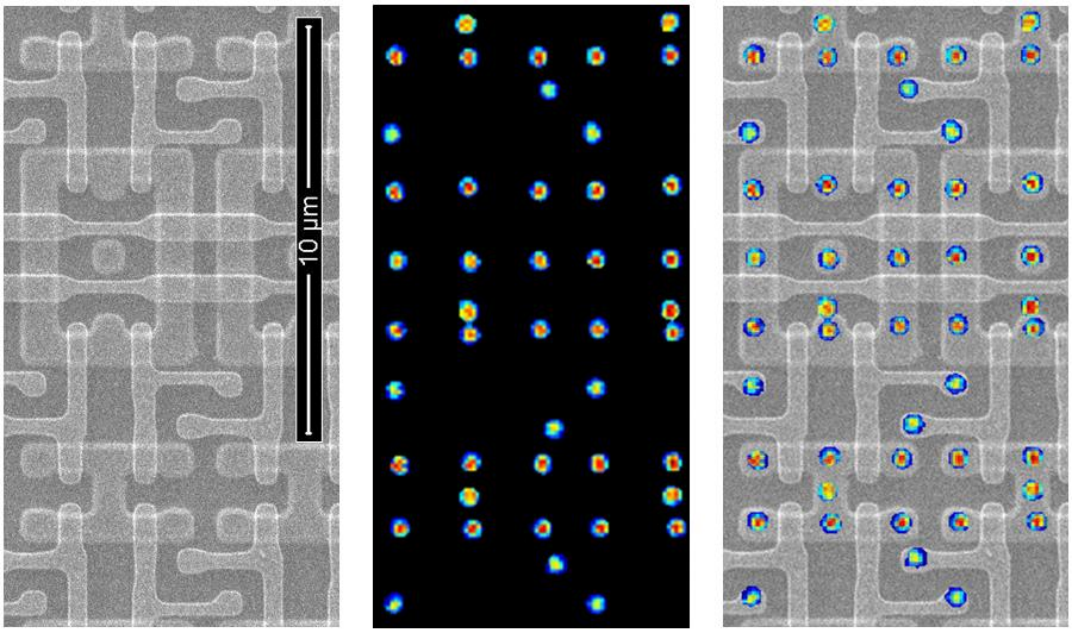

Figure 18: SEM picture of the etched SRAM (left), tungsten XRF map (center) and superimposition of the SEM-XRF pictures (right) [2].

3.7.2 Laboratory tomograph attack

The demonstration of the feasibility of single transistor alteration through X-rays led the CEA researchers to explore more accessible and economical X-ray sources, such as a laboratory tomograph [3]. For this purpose, an *EasyTom XL Ultra 230-160 micro/nano-CT scanner* from *RX Solutions* with a Mo target was employed. Unlike synchrotron X-rays, tomograph radiation is divergent, making it impossible to focus down to the nanometric scale. Furthermore, the absence of in-situ element maps complicates the localization of transistors.

To address the former issue, various mask layouts were tested to isolate the targeted transistor while protecting the surrounding ones from radiation. All these masks featured some holes drilled through a PFIB to target a single transistor or a specific region of the circuit. These attempts included: thin lead films glued over the circuit and movable lead masks placed in front of the circuit and controlled through an arm moved by piezo motors. The second layout emerged as the most adaptable and systematic, as the mask can be milled once and placed directly in front of the targeted transistor.

The latter issue was addressed by retrieving the physical position of the transistors with respect to specific features visible from the optical microscope and the bonds visible from the radiographic image returned by the tomograph. For the Flash memory, the position of the transistors was measured with respect to the periodic apertures that separate groups of eight transistors. These apertures are outlined with red rectangles in **Figure 19 (left)**, where the SEM picture of the etched memory is reported. These apertures were localized in the optical image, where they appear as black boxes as visible in **Figure 19 (center)**. Finally, the position of the black dots was correlated to the location of the bonds, which are visible in the radio image as shown in **Figure 19 (right)**.

The possibility of precisely identifying the transistors and focusing the tomograph radiation enabled the researchers to produce single bit faults in the Flash memory of the ATmega1284P microcontroller. This result served as a proof of concept, validating the feasibility of single bit attacks through a laboratory source.

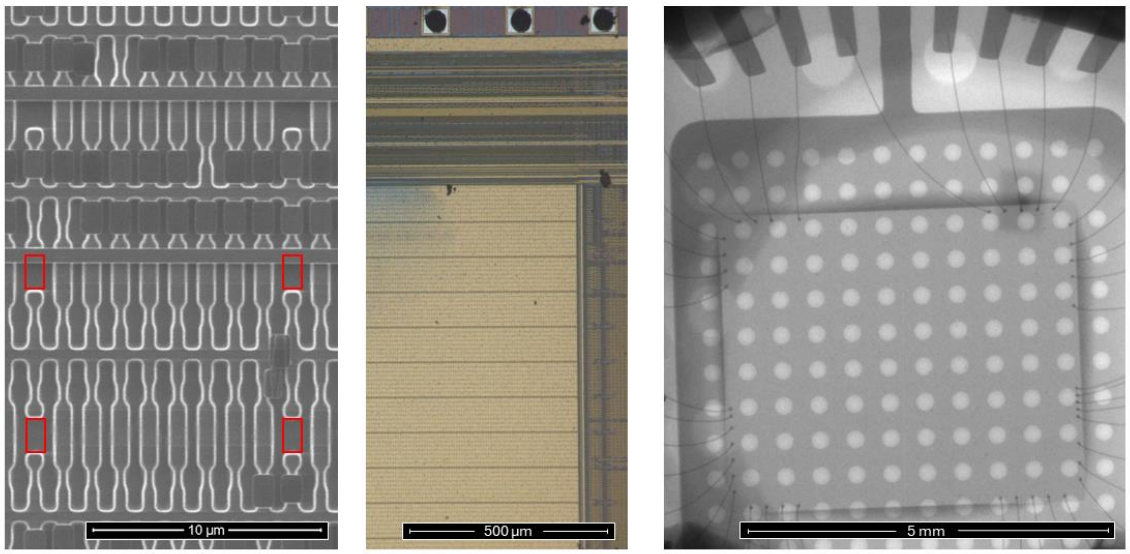

Figure 19: SEM picture of the etched Flash memory (left), optical image of the Flash memory (center) and radio image of the whole circuit (right).

3.7.3 Unsolved problems

The presented experiments successfully demonstrated that both synchrotron and tomograph radiation, when properly focused and positioned, can be used to alter the logical state of single transistors. However, since most research on X-rays/ICs interaction has focused on space radiation, there is a lack of information on the exact mechanisms at play during single transistor irradiation, especially at very high doses. To better understand the phenomena occurring at the semiconductor level, the MITIX team decided to evaluate the accuracy of TCAD simulations in reproducing the performed attacks. To perform helpful simulations, cross-sections of the ATmega1284P memories are needed to reconstruct single transistors and calculate the absorbed dose rate. Additionally, new experiments are required to determine the exact irradiation time needed to fault a transistor, which is a crucial input parameter for irradiation simulations. The preliminary experiments conducted to gather this information for the simulations, as well as the results obtained, are the main focus of this Thesis work.

Furthermore, the success of the presented experiments led the CEA team to explore the feasibility of such attacks on new types of memories. For this purpose, the DRAM of the Zybo-Z7 development board was chosen. Defining the attack procedure for a new type of memory requires a deep understanding of the memory architecture, which is often protected as an industrial secret. Therefore, reverse engineering is necessary to comprehend the transistor architecture and to establish the relationship between physical and logical addresses. The preliminary deprocessing of such memory is presented in this Thesis work.

Chapter 4

TCAD simulations

In order to fully understand the phenomena at play during single bit erasure following X-rays irradiation, the TCAD (Technology Computer Aided Desing) software ECORCE (Étude du COmportement sous Radiation des Composants Electroniques) was chosen. The software was originally developed to study the detrimental effects of space radiation on electronic components, introduced in **Section 3.2**. However, the possibility of simulating TID effects on single transistor lead the MITIX team to identify ECORCE as a powerful instrument to improve their understanding of TID effects on memories and perform more accurate attacks.

The main objective of this Thesis work has been to validate the usefulness of this software in modeling single-bit attacks on MOS and FG transistors, and to provide helpful insights for further attacks. The chapter will be divided into two sections, one related to the Flash memory simulations and the other to the SRAM simulations. For each memory, a description of the preliminary retro-engineering experiments and the codes used to retrieve the input values for ECORCE is given. Afterwards, the simulations results are presented and discussed.

4.1 Preliminary experiments

4.1.1 Flash memory cross-section

A cross-section of the ATmega1284P Flash memory was performed to understand its architecture and measure the dimensions of the floating gate transistors. A *Strata DualBeam FIB* from *FEI company* was employed for this purpose. The instrument is a "dual beam" type, featuring both a gallium ion beam for sample milling and an electron beam for SEM imaging.

The preparation of the sample for FIB manipulations involves the chemical or mechanical removal of the plastic packaging surrounding the ICs. For the ATmega1284P sample, the external package was chemically etched with nitric acid at 60 °C using a *Jet Etch CU Protect* from *Nisene Technology Group*. The device was covered with a mask to expose the packaging above the memory while protecting the rest of the sample. **Figure 20 (left)** shows the device before depackaging while **Figure 20 (right)** shows the device after the localized etching above the memory.

To obtain a clear cross-section with a FIB, it is necessary to deposit a mask above the target region. The mask helps mitigate the effects of the surface irregularities and variations in material hardness, which can divert the ions trajectories creating a jagged surface. The mask exposes a smooth surface and enables the achievement of a planar cross-section. For this purpose, a PtC mask of 5 µm thickness was deposited and the obtained cross-section is shown in **Figure 21**.

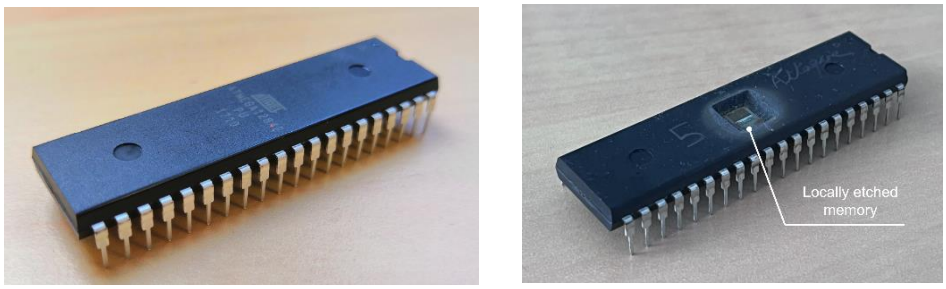

Figure 20: ATmega1284P before depackaging (left) and after localized depackaging above the memory (right).

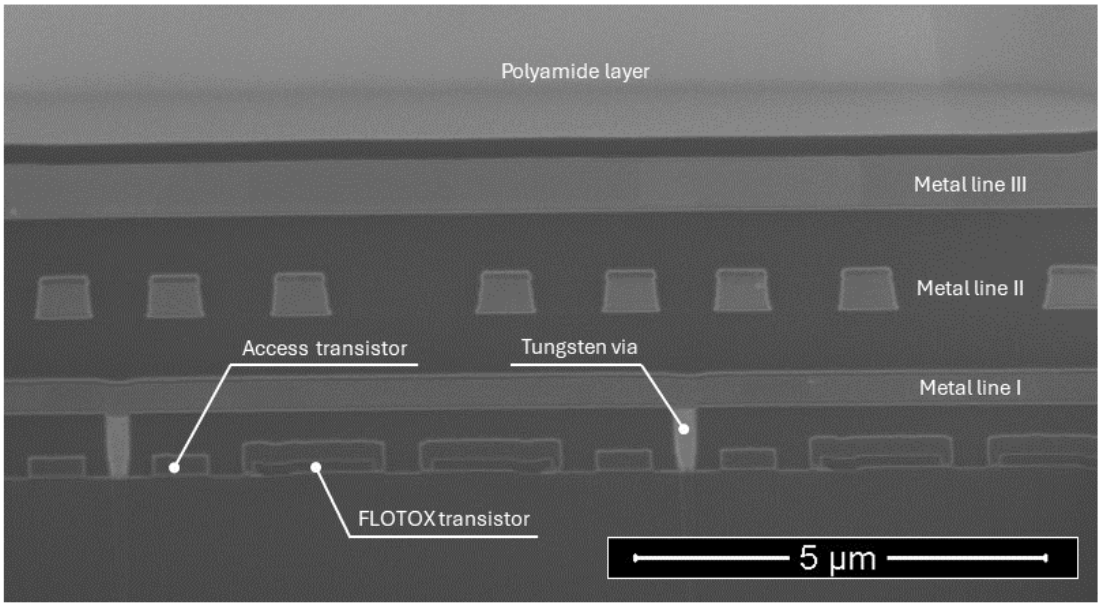

Figure 21: cross-section of the ATmega1284P Flash memory.

From the above picture it is noticeable that the Flash memory is covered by a protective polyamide layer that is not removed during the chemical etching. The electric signal is transmitted through three metal lines, and four transistors are situated between each two vertical vias. The distance between two transistors is approximately 350 nm, corresponding to the nominal technology node of the device. Two types of transistors can be identified: a double gate transistor with a region of reduced oxide thickness and a standard MOS transistor. Therefore, it can be concluded that each memory cell features the two-transistor architecture presented in **Figure 8** with a FLOTOX transistor used to store the bit value and an access transistor which enables writing and erasing operations.

Despite the FLOTOX structure being originally presented for EEPROMs, it is reasonable to think that the same structure was adopted in this Flash memory as the ATmega1284P is an old device. More advanced architectures, such as NAND or NOR memories, were introduced later for lower technology node devices. **Figure 22** shows a zoom on the FLOTOX transistor, while in **Table 1** are reported the transistor's dimensions used in ECORCE to perform simulations.

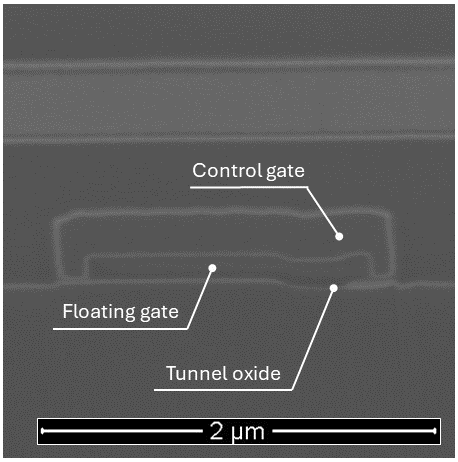

Figure 22: zoom on the FLOTOX transistor.

Feature	Dimension [µm]
Control gate width	1.67
Control gate central thickness	0.19
Floating gate width	1.39
Floating gate central thickness	0.1
Tunnel oxide thickness	0.02

Table 1: characteristic dimensions of the FLOTOX transistor used in ECORCE.

4.1.2 X-ray attacks on Flash memory

As introduced in **Section 3.1**, it is necessary to know the exposure time before faulting to calculate the absorbed dose by a single transistor during tomograph and synchrotron irradiation. For this purpose, several experiments were performed both with a laboratory tomograph and at the ESRF.

For both setups, the experiments were conducted by connecting the device to a dedicated board capable of communicating with any computer through a USB port. During each irradiation experiment, a specific program was uploaded into the memory. This program returns the physical map of the memory, which was retrieved through previous ESRF experiments (**Section 3.7.1**). After deleting the content of the entire memory, the program repeats the following cycle:

- **Programming of the FLOTOX transistors**: all the FLOTOX transistors are programmed through electrons injection in the floating gate.
- **Memory reading and monitoring**: the memory is periodically read, and a copy of its content is saved. Except for the first iteration, its content is compared with the saved one and for each bit that has changed a red dot appears on the physical map. At each cycle a dump file is generated, which contains the logical map of the memory and allows the user to locate the address(es) of the faulted bit(s).

The tomograph experiments were performed with an *EasyTom XL Ultra 230-160 micro/nano-CT scanner* from *RX Solutions* using an external voltage of 60 kV, a current of 50 mA and a pinhole of 100 µm. With respect to the experiments described in **Section 3.7.2**, the target was changed from molybdenum to tungsten on diamond. This change aimed to increase the photon flux and thereby reduce the exposure time required to fault the transistors. Two mask layouts were tested to understand the optimal attacks conditions:

- **Fixed tungsten masks**: they were deposited directly on the Flash memory external surface using the PFIB and holes were milled with the PFIB above the targeted transistors.
- **Movable tungsten mask**: this external mask was previously milled through the PFIB to create a big hole, which was aligned with the targeted transistors to expose them during the attacks.

As shown in **Figure 23 (left)**, three tungsten masks were deposited through an *Helios 5 DualBeam Plasma-FIB* from *ThermoFisher Scientific*, a new generation FIB that uses a Xe^+ ions plasma to mill materials instead of Ga^+ ions. These masks were deposited on the Flash memory with a voltage of 12 kV, a current of 70 nA and a deposition time of 20 minutes to reach a nominal thickness of 10 µm. Afterwards, two perpendicular lines were milled over each mask following the black boxes, which were previously described in **Section 3.7.2** and are visible from a close-up view in **Figure 23 (center)**. Their aim is to help the localization of the targeted transistor, given that there are always eight transistors between two black boxes that lie on the same horizontal line.

For each mask, a hole was milled above the targeted transistor using a voltage of 30 kV, a current of 1 nA and a milling time of 60 seconds to reach a 7 to 8 µm depth. **Figure 23 (right)** shows the mask following lines and hole milling. Under these conditions the obtained diameters were close to 1.5 µm, which corresponds approximately to the gate's width of the FG transistors. The hole depth was intentionally kept less than the mask thickness to protect the transistors from ion damage.

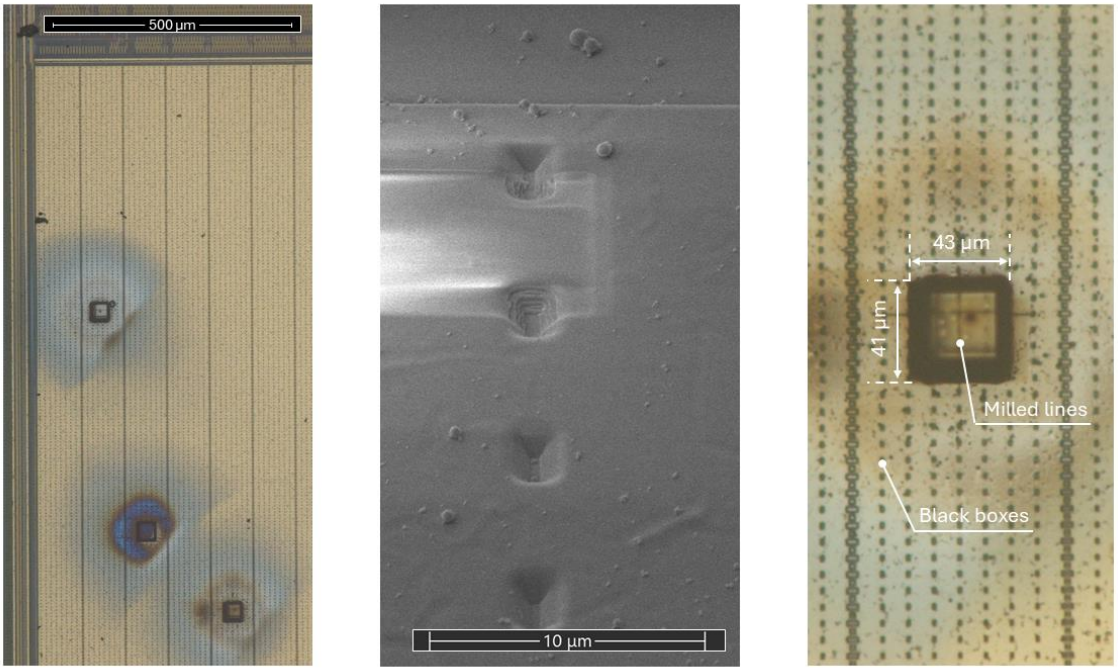

Figure 23: optical view of the three tungsten masks (left), close up view of the black boxes separating the transistors (center) and optical view of a single mask with the milled lines (right).

The sample was then shielded with lead sheets of 1 mm thickness to protect the areas surrounding the memory. Since this mask layout had never been tested, no additional masks were added to protect the remaining parts of the memory, and the attack was performed only to understand whether the transistors beneath the holes could be faulted or not. This mask layout did not prove to be efficient, as 30 minutes of irradiation did not produce any faults in the targeted areas. This result may be explained by two main reasons:

- **Misaligned holes**: the milled holes may be shifted with respect to the aimed position, resulting in a partial or incomplete exposure of the transistors, therefore increasing the exposure time required to reach a sufficient TID to fault them.
- **Insufficient hole diameter**: the diameter of the milled holes may be too small with respect to the gates' width, resulting in a lower absorbed dose and thus increasing the faulting time.

Therefore, the movable mask layout was chosen as it previously demonstrated to be most rapid and systematic method to fault transistors using a laboratory source, as explained in **Section 3.7.2**. A 25 mm x 25 mm tungsten mask of 25 µm thickness with a PFIB-drilled hole of 10 µm x 20 µm was employed. This mask is controlled by piezomotors which enable its precise location above the region of interest and is easily recognizable from the radio image as it appears as a white rectangle. The whole setup is shown in **Figure 24**. It has to be noted that the hole in the mask can cover a wide area containing more than one single transistor. However, the aim of the experiment was not to fault a precise single transistor but to evaluate the average time required to detect the first fault.

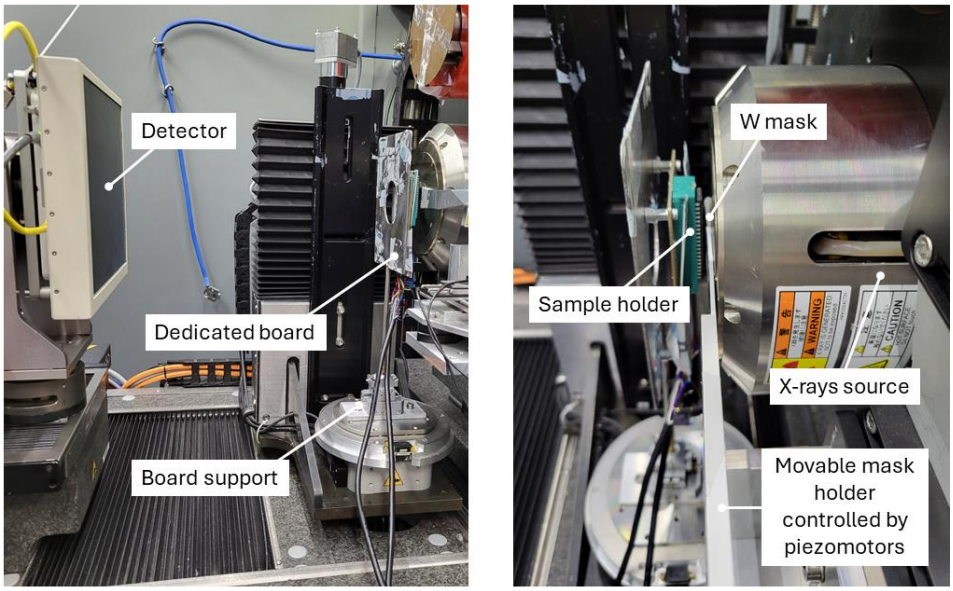

Figure 24: experimental setup for tomograph attack.

In **Figure 25 (left)**, the typical evolution of the faults distribution displayed in the physical map during a tomograph X-ray attack is shown. **Figure 25 (right)** plots the number of faulted bits as a function of irradiation time.

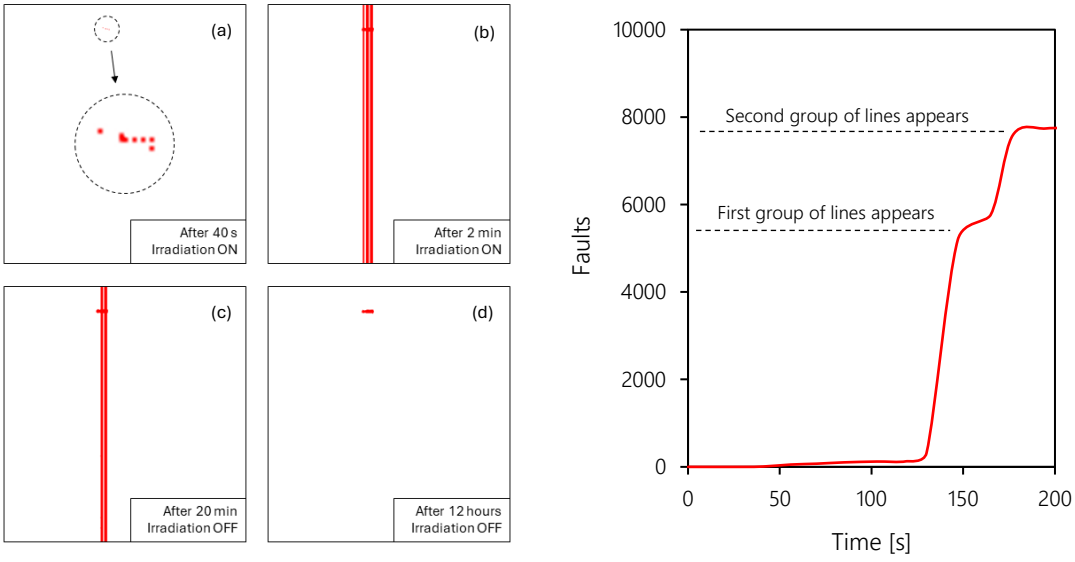

Figure 25: evolution of the faults distribution in the physical map (left) and number of faults as a function of the time (right) during a tomograph attack.

The first detected faults appear after 40 seconds (**Figure 25 (left, a)**) and correspond to the FLOTOX transistors that have changed their logical state due to the effects described in **Section 3.3.3**. After two minutes of irradiation the total number of faults increases drastically, reaching more than 7000 faults after three minutes (**Figure 25 (left, b)**). This is connected to the appearance of vertical line faults, which arise even outside of the area exposed through the mask.

While the exact mechanism is still not entirely clear, this type of fault may arise from one or more access transistors faulting due to the ionization of their gate oxides, through the mechanism explained in **Section 3.2**. This hypothesis, first presented in [3], is supported by the observation that twenty minutes after the end of the irradiation, some of the line faults start to disappear (**Figure 25 (left, c)**). Additionally, twelve hours after the end of the irradiation all the line faults have disappeared and only the faulted floating gate transistors are detected (**Figure 25 (left, d)**). This suggests that the vertical faults might be caused by the alteration of a limited number of access transistors exposed to the irradiation that may rapidly anneal, thereby restoring the initial logical value of the line, except for the faulted FLOTOX transistors.

Different attacks were performed under the same irradiation conditions in various regions of the memory, and it was not possible to determine a consistent value of the time required to detect the first faults. Consequently, an average value of 90 seconds was chosen for the simulations as the irradiation time required to consider a FLOTOX transistor faulted following tomograph irradiation.

Irradiation attacks were performed at the ESRF ID16B beamline to evaluate the time required to fault a single floating gate transistor with a nanofocused beam. The setup for the attack is shown in **Figure 26**. The sample is connected to the powerline through the same dedicated electronic board used during the tomograph attack. Above the sample holder there is the optical microscope used for pre-positioning, whereas the XRF detector is placed at the side of the sample.

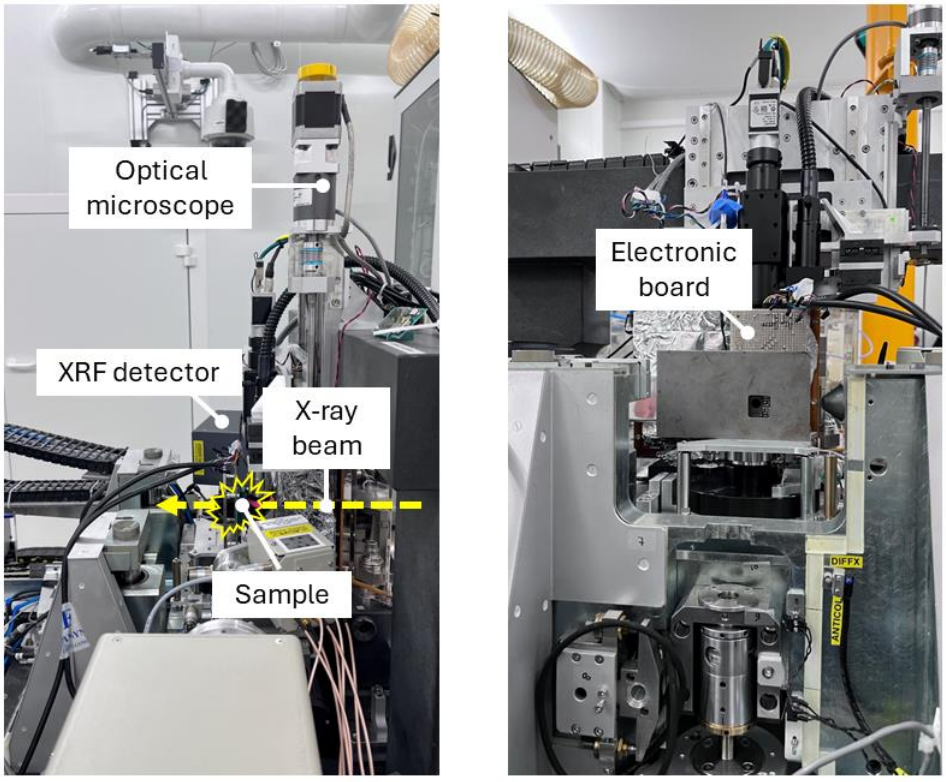

Figure 26: ID16B beamline setup for single bit attack.

Since the architecture of the ATmega1284P Flash memory was already well-known from previous experiments, the fluorescence mapping was not utilized. The attack was performed using the optical image, as its focal plane is aligned with that of the X-ray optics. Knowing that there are eight transistors between each pair of black boxes and their size thanks to **Figure 19 (left)**, it was possible to accurately position the beam on top of the targeted transistors. The evolution of the faults distribution in the physical map during a single bit attack is shown in **Figure 27 (left)**, while the number of faults as a function of the time is shown in **Figure 27 (right)**.

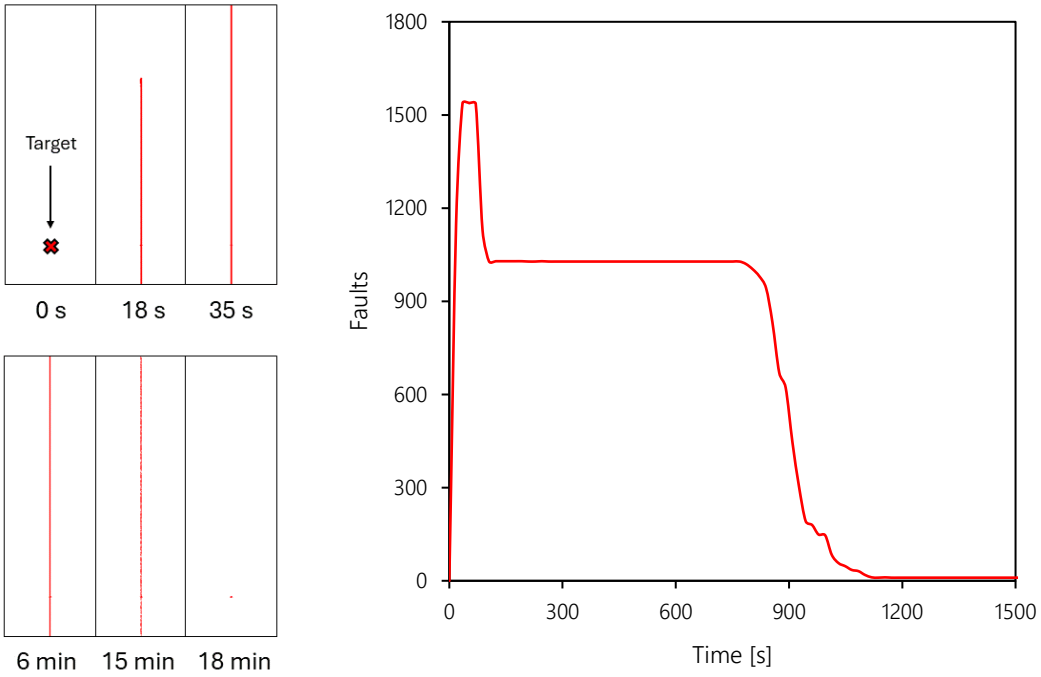

Figure 27: evolution of the faults distribution in the physical map (left) and number of faults as a function of the time (right) during the ESRF attack.

It is clearly visible that the evolution of the faults during the ESRF attack exhibits the same behavior previously described for the tomograph attack. If the attack is prolonged too much following single transistor faulting, vertical line faults appear. Due to the lower exposed surface compared to tomograph irradiation, the maximum number of faulted transistor is smaller, close to 1500. Consequently, the annealing time of the access transistors is shorter, with the single bit error being the only one remaining after all the line faults have disappeared within 18 minutes.

Several attack were performed on different transistors, and the times required to fault single transistors were consistent. On average, two seconds are required to fault one single FLOTOX transistor at the ID16B beamline. Therefore, this value was chosen for the simulations as the irradiation time to consider one single transistor faulted. However, the times required to anneal the vertical line faults varied greatly, with some faults not restoring their original value even after several days.

4.1.3 Dose rate calculation for Flash memory

To calculate the absorbed dose rate during X-ray irradiation with the tomograph and at the ESRF, a dedicated Python code was developed. The principal inputs of the program are the X-ray sources spectra, shown in **Figure 28**, and the faulting time calculated in the previous sections.

The tomograph spectrum was measured with a CdTe spectrometer from *Amptek* at 150 mm from the source with a pinhole of 100 µm to contain the flux. The acquisition surface of the spectrometer was 25 mm^2 and the thickness of the CdTe layer was 1000 µm. The result was used to estimate the tomograph intensity at 5 mm, which corresponds to the employed distance between the IC and the X-ray source during experiments. At 5 mm, the tomograph spectrum has a maximum intensity of $2.22 \cdot 10^9$ $ph \cdot s^{-1} \cdot cm^{-2}$ at 7 keV, whereas the ESRF spectrum is a single line of intensity $7.07 \cdot 10^{19}$ $ph \cdot s^{-1} \cdot cm^{-2}$ around 18 keV since the ID16B beam is monochromatic.

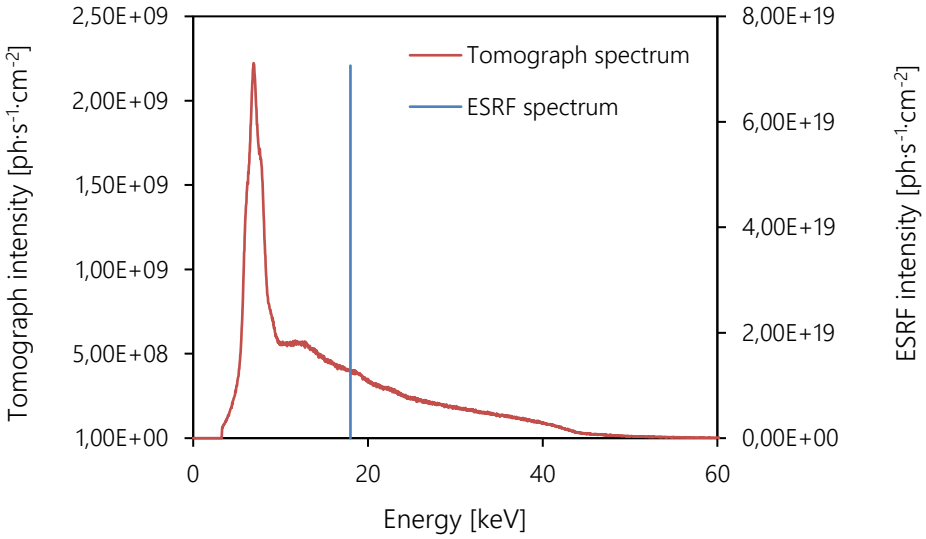

Figure 28: comparison between the tomograph and ESRF spectrum.

The Python module *xraydb* is used as it provides numerous tabulated data related to the X-rays/matter interaction. The function `material_mu (material, energy)` is used to retrieve the linear attenuation coefficient μ as a function of the energy for the main materials used in ICs. The inputs of this function are the label of the material and the energy of the radiation, expressed in eV. Instead, the linear energy attenuation coefficient μ_{en} can be derived from the NIST website [37]. In **Figure 29** the mass attenuation coefficient μ/ρ and the mass energy attenuation coefficient μ_{en}/ρ of silicon are plotted.

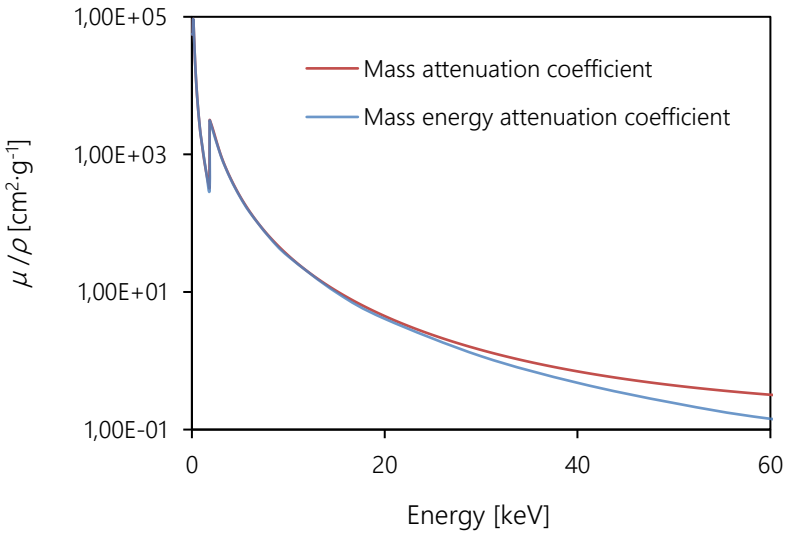

Figure 29: comparison between the mass attenuation coefficient μ/ρ and mass energy attenuation coefficient μ_{en}/ρ of Si.

It is clearly noticeable that for energies lower than 20 keV the two plots overlap, therefore the coefficient provided by `material_mu (material, energy)` can be used to simulate tomograph and ESRF irradiation. Similar considerations are valid for other common IC materials such as SiO_2 and Al.

To evaluate the absorbed dose from a single FLOTOX transistor during tomograph irradiation, the first step is to calculate the reduction of the beam's intensity due to absorption in the layers above. This is done through the function `abs_ratio (material, thickness, energy)` defined as:

```
def abs_ratio (material, thickness, energy):
        mu = xraydb.material_mu (material, energy*1e3)
        trans = np.exp(-thickness*1e-4*mu)
        return trans
```

This function takes as input the label of the material that the radiation is crossing, its thickness expressed in μm, and the radiation energy expressed in keV. These parameters are used to evaluate μ expressed in cm^{-1} as a function of the radiation energy and calculate the absorption ratio I/I_0 through **Eq. 1**. The total absorption ratio is calculated as follows:

```
abs_SiO2_1 = abs_ratio ('SiO2', 0.226, energy_tomo)
abs_Al_1 = abs_ratio ('Al', 0.685, energy_tomo)
abs_SiO2_2 = abs_ratio ('SiO2', 0.678, energy_tomo)
abs_Al_2 = abs_ratio ('Al', 0.53, energy_tomo)
abs_SiO2_3 = abs_ratio ('SiO2', 0.692, energy_tomo)
abs_Al_3 = abs_ratio ('Al', 0.459, energy_tomo)
abs_SiO2_4 = abs_ratio ('SiO2', 0.33, energy_tomo)
abs_total = abs_SiO2_1*abs_Al_1*abs_SiO2_2*abs_Al_2*abs_SiO2_3*abs_Al_3*abs_SiO2_4
```

Where `energy_tomo` is the vector collecting the energies of the tomograph spectrum, ranging from 0 to 60 keV. The three aluminum layers correspond to the metal lines shown in **Figure 19**, whereas silica is the material that separates them. The total absorption ratio corresponds to the product of each absorption ratio and the intensity of the tomograph beam impinging on the FLOTOX gate is:

```
intensity_SiO2_tomo = intensity_tomo*abs_total
```

Where `intensity_tomo` is the vector containing the values of the tomograph intensity expressed in $ph \cdot s^{-1} \cdot cm^{-2}$ for each energy value. The dose rate absorbed by the transistor gate expressed in rad/s can be evaluated using **Eq. 3** for each wavelength and summing all the values as follows:

```
dose_rate_SiO2_tomo = np.sum (intensity_SiO2_tomo*(xraydb.material_mu('SiO2',
energy_tomo*1e3)/xraydb.get_material("SiO2")[1])*energy_tomo*1.6*10**(-11))
```

Where `get_material (material) [1]` is a function that returns the density of the chosen material expressed in g/cm^3. Finally, after defining the variable `fault_time_tomo` with the time calculated in the previous section, it is possible to calculate the absorbed dose by multiplying this value by the absorbed dose rate.

```
fault_time_tomo = 90
absorbed_dose_tomo = dose_rate_SiO2_tomo*fault_time_tomo
```

This computations return a dose rate of $1.5 \cdot 10^3$ rad/s and a total absorbed dose of 0.14 Mrad. Using a similar approach, it is possible to evaluate the same quantities for the ESRF attack. By defining the variable `fault_time_ESRF` equals to 2 s, the computations return a dose rate of $7 \cdot 10^{10}$ rad/s and a total absorbed dose of 140 Grad.

4.1.4 SRAM memory cross-section

A non-functional sample prepared with the same procedure described in **Section 4.1.1** was used to perform a cross-section of the SRAM and retrieve the dimensions of the CMOS transistors. To do so, a *Helios 5 DualBeam Plasma-FIB* from *ThermoFisher Scientific* was employed. A PtC mask of 5 µm thickness was deposited over the memory to improve the quality of the cross-section. The smoothing effect of the mask can be appreciated in **Figure 30 (left)**, where the surface irregularities of the SRAM surface are clearly visible. In contrast, in **Figure 30 (right)** is shown how the milled surface appears during the initial etching steps, where the ions mill the curved edge of the mask which is highly irregular.

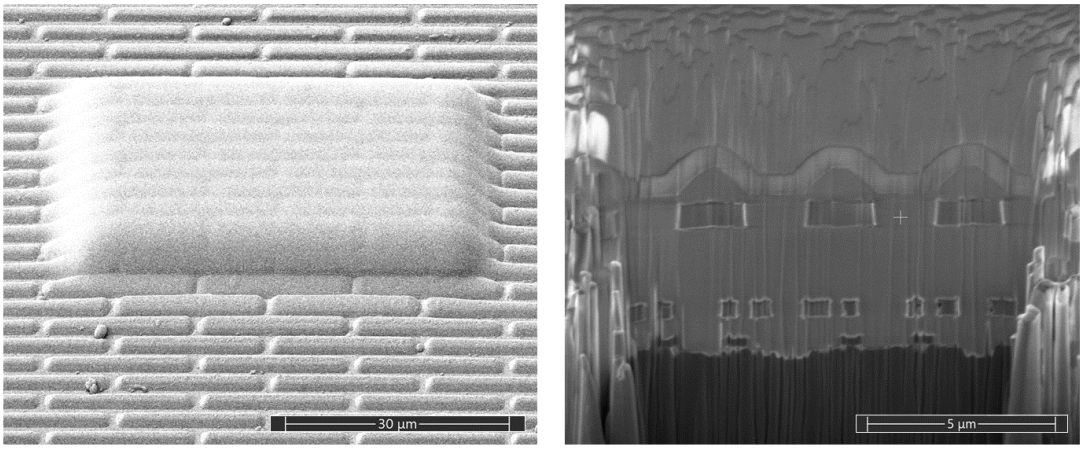

Figure 30: SEM view of the deposited mask (left) and SEM view of the milled surface during the initial steps (right).

The result of the cross-section is shown in **Figure 31 (left)**. It is evident that between each pair of STIs there are two MOS transistors; however, from this image, it is not possible to define whether they are PD1-PD2 or PU1-PU2 pairs of transistors. A better understanding of the memory architecture can be achieved through the SEM image of the etched memory shown in **Figure 31 (right)**.

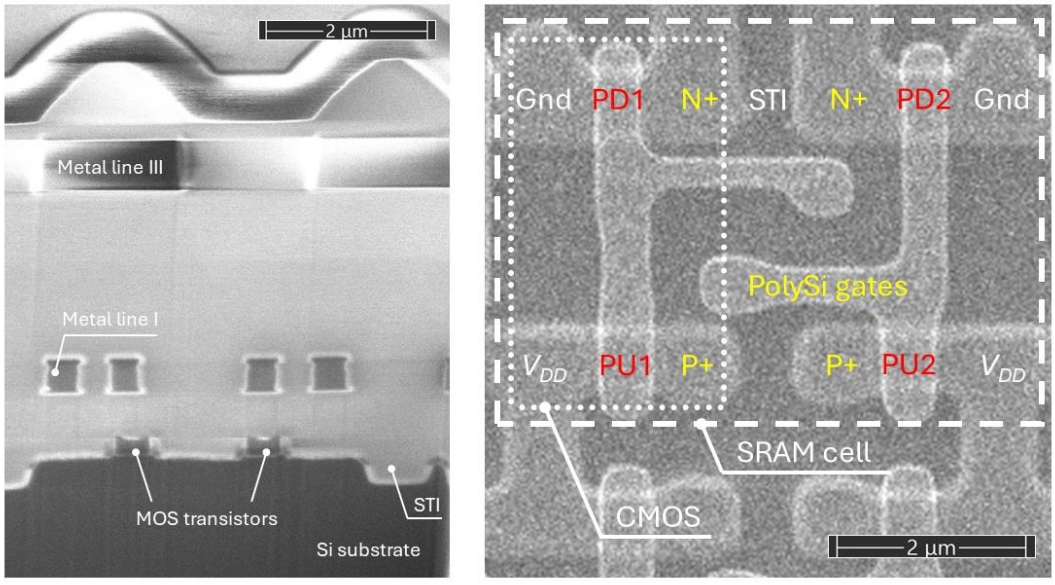

Figure 31: cross-section of the SRAM (left) and top view of the etched SRAM (right).

Since the SRAM cell develops on a plane, in ECORCE is possible to represent one CMOS only. For this purpose, the left CMOS was taken as a reference. In **Figure 32** is shown a close-up view of a MOS transistor and in **Table 2** are reported the dimensions used to build the CMOS in ECORCE.

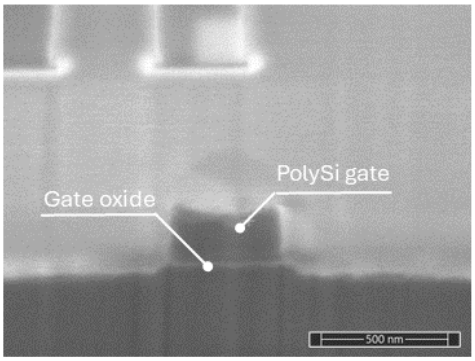

Figure 32: close-up view of a MOS transistor.

Feature	Dimension [µm]
PMOS gate width	0.57
NMOS gate width	0.47
STI width	1.81
Gate oxide thickness	0.03

Table 2: characteristic dimensions of the PMOS and NMOS transistors used in ECORCE.

4.1.5 X-ray attacks on SRAM

To perform X-ray attacks on the SRAM, the same program described in **Section 4.1.2** was uploaded into the samples. The only difference lies in the writing operations, which for the SRAM follow the mechanism explained in **Section 2.1**. However, it is important to note that a cell is detected as faulted even if only one of the four transistors constituting the latch is altered, since the program only reads the logic output of each cell. Therefore, if the program detects N faulted cells, the number of faulted transistors may range from $N/4$, in the case only one transistor per cell is faulted, to N, if all the transistors in the faulted cells are affected.

The tomograph attacks were performed under the same operative conditions adopted for the Flash memory attacks, and the same two masks layout were tested. The fixed tungsten masks were deposited using the same parameters previously described. Two types of holes were milled:

- **Rectangular hole**: dimensions of 1.5 µm x 4.5 µm to cover an entire SRAM cell, shown in **Figure 33 (left)**. The milling conditions were a voltage of 30 kV and a current of 0.3 nA.
- **Circular hole**: diameter of 1.5 µm to expose a single transistor, shown in **Figure 33 (right)**. The milling conditions were a voltage of 30 kV and a current of 1 nA.

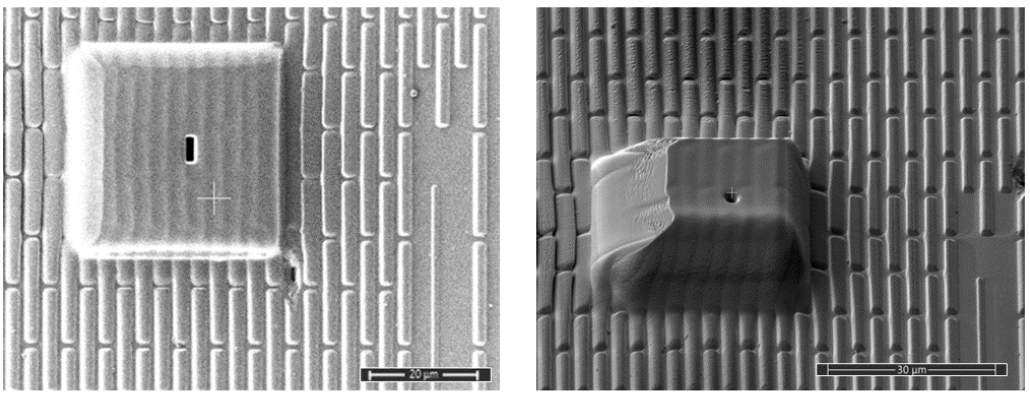

Figure 33: fixed tungsten mask with a rectangular hole (left) and a circular hole (right).

The milling time was set at 60 s, as previous experiments demonstrated it to be the optimal time to reach a depth around 7 to 8 µm. However, this mask layout did not prove to be efficient, as the studied samples lost their functionality following mask deposition. This result could be explained by two main reasons:

- **Damage during hole milling**: the targeted transistors may have been damaged during the hole milling step. To investigate this, a measurement of the masks' thickness was performed using an optical microscope. This involved the manipulation of the focal length to precisely focus on the top surface of the mask and on the circuit external surface, as shown in **Figure 34 (left)**. Through this method, the height of the mask was estimated to be around 7 µm, therefore 3 µm lower than the nominal thickness of 10 µm. Since the milling time was set to produce a trench of 7 µm, the ions may have reached the ICs beneath the mask, causing the loss of functionality of the targeted transistors.
- **Damage during mask deposition**: the memory was damaged during the deposition of the masks as visible in **Figure 34 (right)**, where the damaged spots are highlighted with yellow circles. These damages were probably caused by the presence of superficial static electricity, and the diverted ions may have reached the ICs level, determining the loss of device functionality.

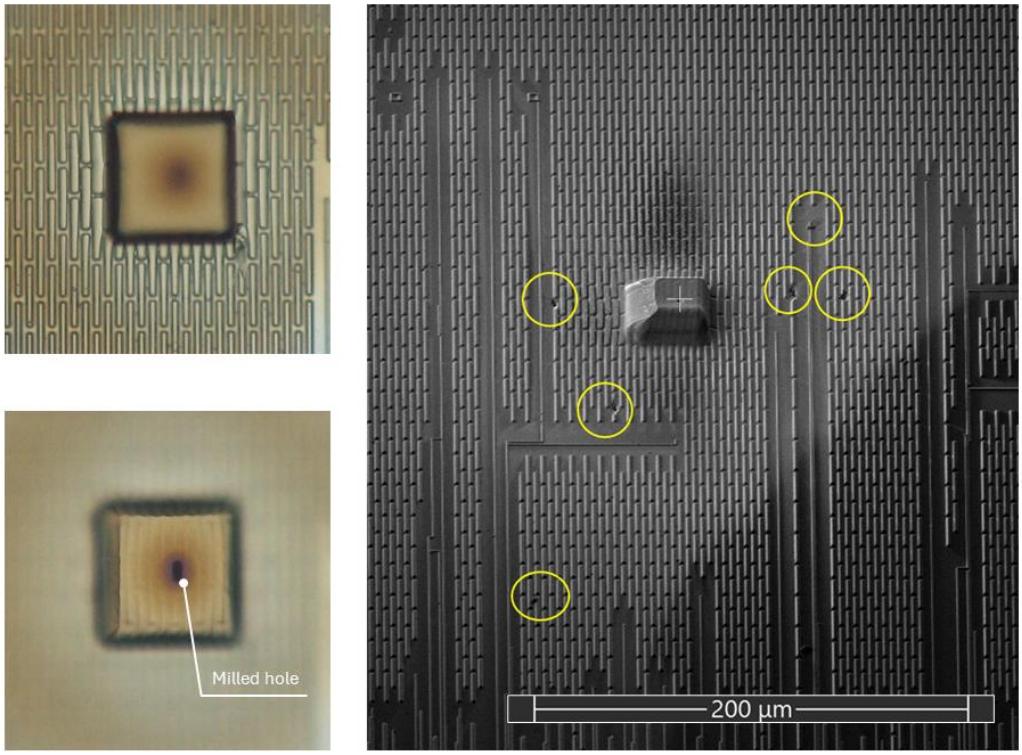

Figure 34: optical method used to determine the thickness of the mask (left) and superficial damages due to the presence of static electricity (right).

To obtain the time required to fault an SRAM cell with a tomograph source, an attack was performed with the movable mask layout. The tungsten mask with a 10 µm x 20 µm PFIB-drilled hole was placed in front of the SRAM, and 300 seconds were required to detect faults in the memory. This time was chosen for the simulations as the time required to fault a CMOS circuit. For higher irradiation times, the number of faulted cells remained constant at 140, indicating that the number of faulted transistors is between 35 and 140.

The faulting time of single SRAM transistors irradiated at the ESRF were already known from previous experiments. The chosen irradiation times are 50 seconds for faulting a PMOS in the ON state, 190 seconds for a PMOS in the OFF state, 200 seconds for a NMOS in the ON state, and 60 seconds for a NMOS in the OFF state.

4.1.6 Dose rate calculation for SRAM

The code presented in **Section 4.1.3** was exploited for the SRAM attacks to determine the absorbed dose rate from the transistors. The only difference lies in the absorption ratios, which were calculated using the layers' thicknesses measured from the cross section of **Figure 31 (left)**. The computation for the tomograph attack returns an absorbed dose rate of $1.5 \cdot 10^3$ rad/s and a total absorbed dose of 0.45 Mrad. On the other hand, the computation for the ESRF attacks returns an adsorbed dose rate of $7 \cdot 10^{10}$ rad/s and total absorbed doses in the order of thousands of Grad.

A summary of the calculated parameters used in the simulations, together with the characteristic dimensions of the transistors, is reported in **Table 3**.

<table>
<tr><th>Summary table</th><th>Gate width [µm]</th><th>Faulting time tomo [s]</th><th>Faulting time ESRF [s]</th><th>Dose rate tomo [rad/s]</th><th>Dose rate ESRF [rad/s]</th><th>Absorbed dose tomo [Mrad]</th><th>Absorbed dose ESRF [Grad]</th></tr>
<tr><td>FG transistor</td><td>1.67</td><td>90</td><td>2</td><td rowspan="3">$1.5 \cdot 10^3$</td><td rowspan="3">$7 \cdot 10^{10}$</td><td>0.14</td><td>140</td></tr>
<tr><td>NMOS</td><td>0.47</td><td rowspan="2">300</td><td>200 (ON)
60 (OFF)</td><td rowspan="2">0.45</td><td>13900 (ON)
4180 (OFF)</td></tr>
<tr><td>PMOS</td><td>0.57</td><td>50 (ON)
190 (OFF)</td><td>3480 (ON)
13200 (OFF)</td></tr>
</table>

Table 3: summary table containing the main parameters adopted for the simulations.

4.2 ECORCE software

ECORCE discretely solves the Poisson's equation to derive the electric potential distribution within semiconductors. This result is coupled with the diffusion equations to describe the distribution of charge carriers within the materials as a function of time. The main peculiarity of this software is its ability to simulate charge trapping and detrapping following X-ray irradiation through the MTD model, introduced in **Section 3.2**. Additionally, ECORCE uses a dynamic mesh, which permits a significant reduction in calculation times compared to traditional mesh generation [38].

ECORCE allows the user to manually draw the cross-section of the device under study and define the materials it is made of. The contacts and doping profile must be defined as well, and trap densities and energetic levels have to be included if irradiation simulations are performed. Originally, the software allowed to simulate the irradiation of the whole device only. However, the possibility of focusing the X-ray beam down to the nanometric scale was added specifically for this Thesis work, in order to improve the understanding of ESRF experiments.

4.3 Flash memory simulations

4.3.1 The model

The model of a single FLOTOX transistor was drawn in ECORCE using the dimensions obtained through the cross section of **Figure 21**, and its visualization is shown in **Figure 35**. The transistor is built on a silicon substrate with a thickness of 0.4 µm, and its lower edge serves as the substrate contact. The floating gate, made of polysilicon, is surrounded by an SiO_2 layer with a thickness of 0.03 µm, while the tunnel region has a thickness of 0.02 µm. The control gate, also made of polysilicon, has its upper surface used as the gate contact. The exposed lateral surfaces of the substrate are utilized as the source and drain contacts. All the contacts are set as ohmic contacts, and the background temperature is set at 300 K.

The doping concentration of the substrate is set at 10^{16} cm^{-3}, with aluminum as the acceptor atom. The source and drain contacts are doped with two n^+ regions using arsenic atoms as donors. These regions have a depth of 0.1 µm and a doping concentration of 10^{18} cm^{-3}. The drain implantation region extends below the floating gate until the region of reduced oxide thickness to enhance electron injection. The floating gate and the control gate are doped with arsenic with a concentration of 10^{19} cm^{-3}.

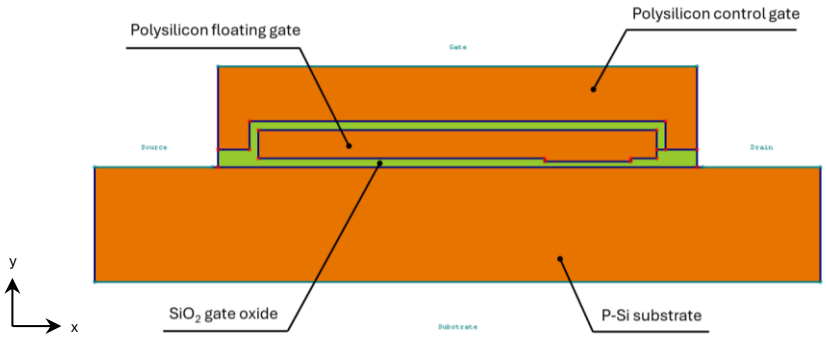

Figure 35: ECORCE model for the FLOTOX transistor.

A uniform concentration of electron and hole traps is defined within the oxide layer, with concentrations of 10^{17} cm^{-3} and 10^{18} cm^{-3} respectively. To reduce the computational complexity, only one energetic trap level is defined for each type of trapping site: 0.8 eV for the electron traps and 1.4 eV for the hole traps. Additionally, no interface traps are defined to simplify the simulations and the analysis of their results.

The distribution of the electric field when all the contacts are grounded is reported in **Figure 36 (left)**. It is clearly visible that the intrinsic electric field, arising from the coupling of different semiconductors, reaches its maximum intensity in the tunnel oxide, with a peak in the region of higher thickness. The shape of the electric field has a direct influence on the trap activation energies, as it lowers the energetic barrier that the charge carriers have to overcome to escape from the trapping sites (Poole-Frenkel effect). This influence is visible in **Figure 36 (right)**, where the activation energies of electron traps are plotted.

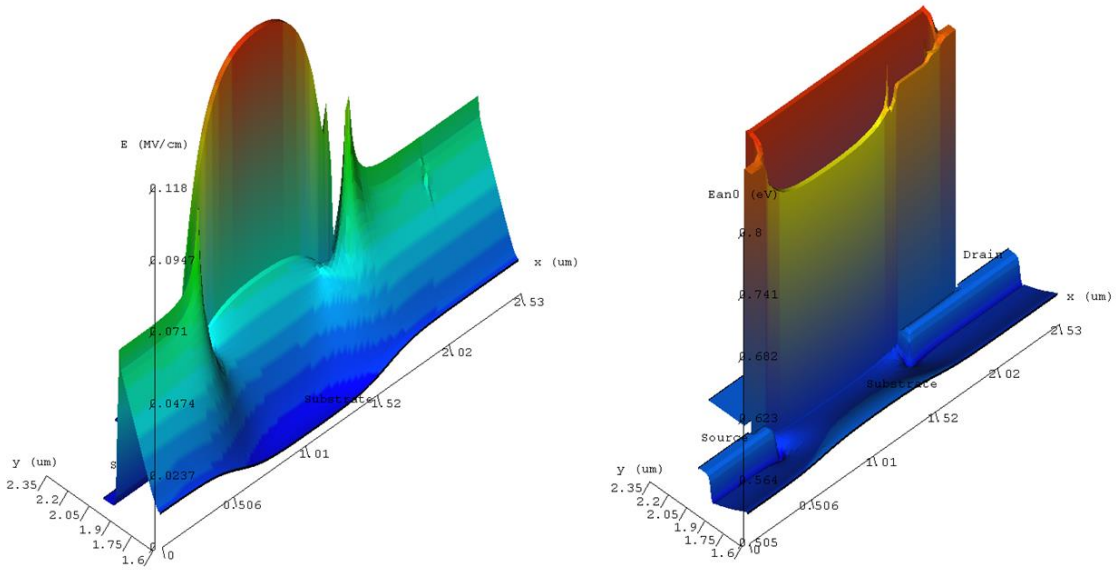

Figure 36: shape of the electric field (left) and of the electron trap activation energies (right).

In the region where the electric field is maximized, the electron trap energy decreases to 0.73 eV. A similar consideration can be made for the hole trap energy, which decreases to 1.33 eV. This phenomenon influences the residence time of trapped charges during and after irradiation, therefore affecting the extent of the TID effect on the electrical behavior of the transistor.

4.3.2 Fowler-Nordheim electron injection

The injection of charges within the floating gate is performed by increasing the voltage on the gate V_{GS} from 0 V to 10 V in 50 ns and maintaining it constant for 1 ms. This creates a strong electric field within the oxide region of reduced thickness, resulting in an intense interfacial current of electrons tunneling from the substrate to the floating gate. This is also confirmed by the change in the conduction band profile, which is plotted in **Figure 37 (left)** before (t = 0 s, V_{GS} = 0 V), during (t = 1 ms, V_{GS} = 10 V) and after (t > 1 ms, V_{GS} = 0 V) injection. The application of an external bias tilts the band within the oxide layer, while the injection of electrons increases the level of the band within the polysilicon of approximately 1 eV.

The density of electrons stored in the floating gate (expressed as electrons/µm due to the two-dimensional nature of the device) rapidly increases during the injection, as shown in **Figure 37 (right)**. Initially, the electron density in the floating gate is approximately $1.417 \cdot 10^6$ electrons/µm and rises to $1.434 \cdot 10^6$ electrons/µm. Therefore, the total number of injected electrons corresponds to $1.618 \cdot 10^4$ electrons/µm.

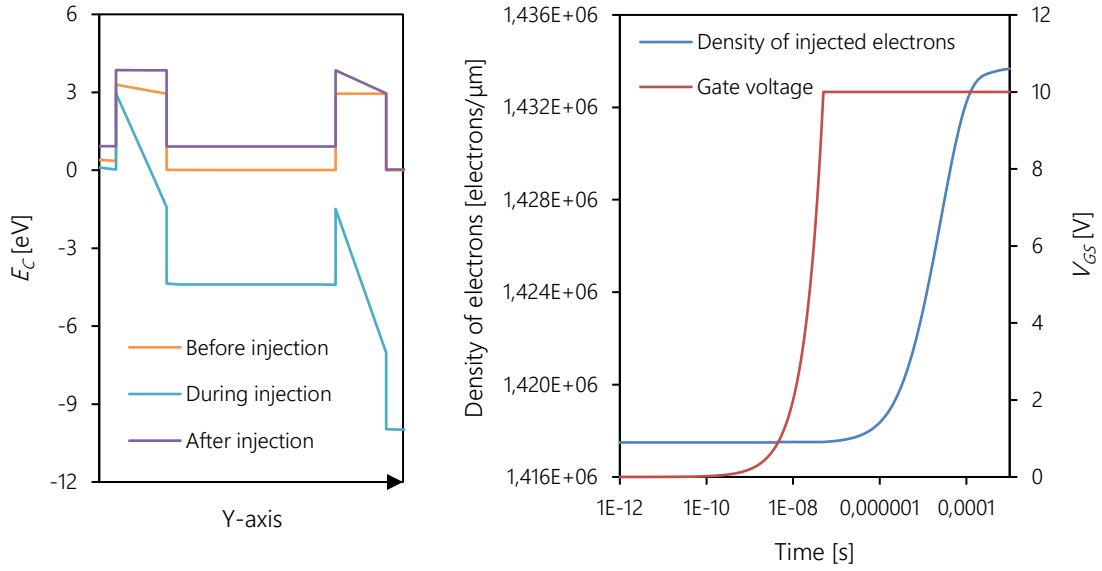

Figure 37: plot of the conduction band before and after injection (left); gate voltage and density of injected electrons as a function of the time during the writing operation (right).

After 0.1 ms, the density of injected electrons reaches a plateau, and the tunneling current begins to decrease, indicating that a physical limit has been reached. Simulations performed with longer injection times do not produce a significant change in the final density of stored electrons (only a 0.005% increase for an injection time of 10 ms), indicating that the floating gate cannot accommodate additional electrons.

To obtain the curves shown in **Figure 7**, the same simulation is performed for the device without electrons in the floating gate and the programmed transistor: V_{DS} is brought from 0 to 2 V and V_{GS} from 0 to 5 V. The results, plotted in **Figure 38 (left)**, show that the I_{DS} (V_{GS}) curve of the programmed device has shifted of approximately 0.7 V to the right compared to the curve of the erased transistor.

The shift confirms that the electrons are now stored in the floating gate, thereby altering the electric field and increasing the voltage required to open the conductive channel beneath the gate oxide. Additionally, the electric field that builds up following charge injection contributes to further reduce the trap activation energies, especially in the upper oxide layer and in the reduced thickness region where it is more intense (**Figure 38 (right)**).

An additional confirmation of the model's robustness can be found by examining the change in electron density within the floating gate when biases are applied or removed. For example, the reduction of V_{GS} from 10 V to 0 V following electron injection or the application of voltages to obtain the I_{DS} (V_{GS}) curve has a negligible effect on the electron density, which changes by less than 0.5%. This means that the change affects less than 100 electrons/µm, and the logical state of the transistor does not get altered by these operations. Thereby, this stability demonstrates the validity of the model in simulating reading and writing operations of the examined FLOTOX transistor.

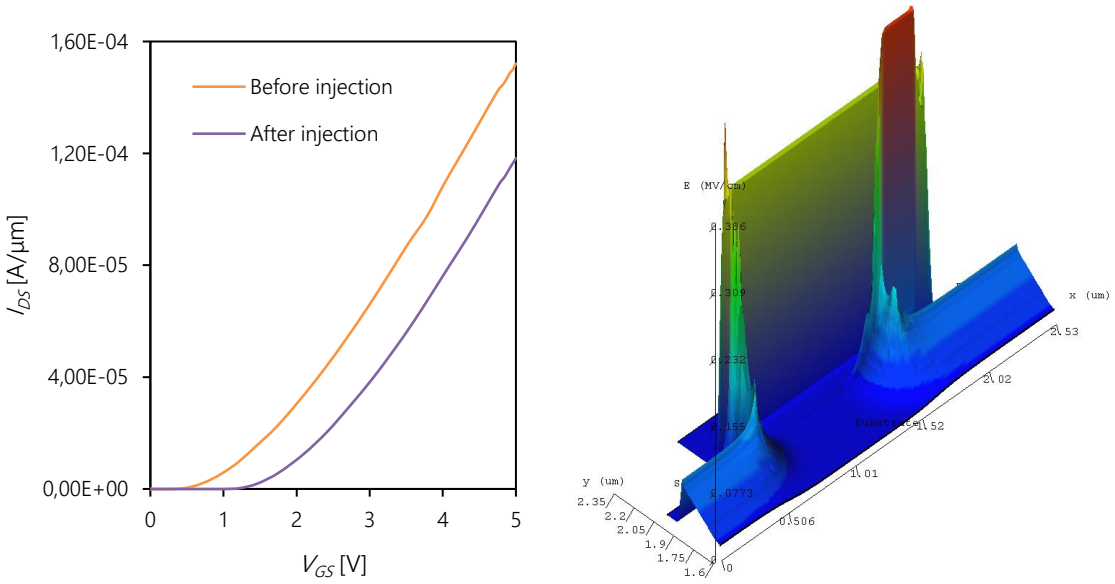

Figure 38: shift of the I_{DS} (V_{GS}) curves before and after programming (left); shape of the electric field within the transistor after programming (right).

4.3.3 Irradiation simulations with laboratory source

During irradiation experiments, the transistors are cyclically read by applying a reading voltage V_R on the gate and the activation of the bit-line. From **Figure 38 (left)**, it is possible to deduce that the value of V_R may reasonably be close to 1 V, therefore this value was used for the simulations. However, since simulating periodic applications and removals of V_R would increase the computational weight of the simulations, a simplification was made: during irradiation simulations, V_{GS} and V_{DS} are kept constant at 1 V.

Furthermore, the adopted version of ECORCE is not able to simulate photoemission during irradiation. However, since the threshold voltage shift caused by the TID effect is quadratically dependent on the oxide thickness [4], and the studied device is old with a relatively thick oxide layer, it is reasonable to assume that photoemission may play a negligible role [39]. A new version of the software which includes the photoemission effect is currently under development and some trials have been conducted to test its efficiency. The model adopted to calculate the emission rate of photoelectrons J_{pe} from the floating gate is:

$$J_{pe} = q\, n_{pe}\, \mu_n\, E \qquad \textbf{Eq. 4}$$

Where q is the elementary charge, μ_n is the electron mobility, E is the electric field and n_{pe} is the density of photoelectrons, which can be calculated as:

$$n_{pe} = g_{eh}\, t\, f \qquad \textbf{Eq. 5}$$

In this formula, g_{eh} is the generation rate of electron-hole pairs, t is the irradiation time and f is the fraction of hot electrons that manage to overcome the oxide energetic barrier, thereby becoming photoelectrons. Unfortunately, the inability to experimentally determine the value of f introduces a degree of uncertainty, making it impossible to derive meaningful results from the simulations. Consequently, the results obtained with the new version of ECORCE will not be presented.

The simulation of tomograph irradiation is performed by exposing the transistor for 90 seconds to a dose rate of $1.5 \cdot 10^3$ rad/s, which was previously calculated through the code. The negative slope of the conduction and valence bands within the oxide when an external bias is applied (**Figure 39 (left)**) influences the behavior of the electrons and holes generated by ionization in the oxide. Depending on the oxide layer where the electron-hole pairs are generated, the phenomena at play are different:

- **Upper oxide**: electrons tend to move towards the gate while holes diffuse towards the floating gate and recombine with the stored electrons at the interface.
- **Lower oxide**: electrons diffuse towards the floating gate, thereby getting injected, while holes tend to move towards the substrate.

The interaction of the carriers generated by the TID effect with the electrons stored in the floating gate controls the stored electron density during irradiation, which is plotted in **Figure 39 (right)** as the extension of **Figure 37 (right)**.

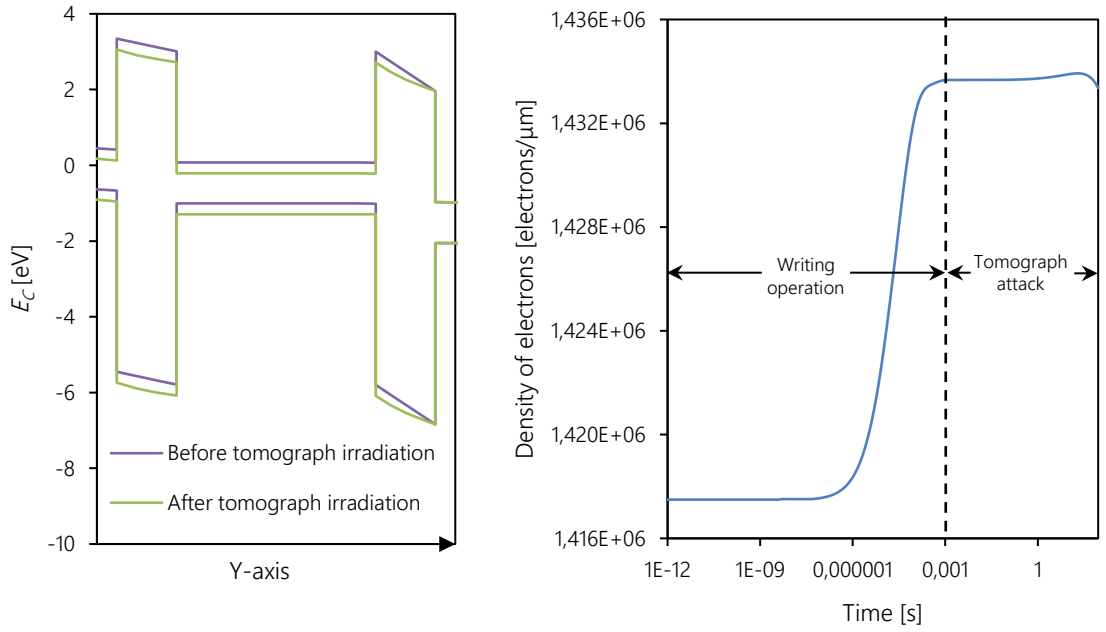

Figure 39: band structure before and after tomograph irradiation (left); density of stored electrons before and after tomograph irradiation.

It is clearly visible that the density of stored electrons does not change significantly during the tomograph attack. Therefore, it can be concluded that the recombination of holes with stored electrons at the floating gate/oxide interface is not the primary mechanism at play during an attack with a laboratory source. This is confirmed by the magnitude of the interfacial currents, which are approximately 10^{-10} A·cm^{-2} for both electrons and holes.

The density of trapped holes in the oxide surrounding the floating gate as a function of the irradiation time is reported in **Figure 40 (left)**. Its magnitude, which reaches its peak of ~ 8000 charges/µm at the end of the irradiation, confirms that the main mechanism affecting the logical state of the FLOTOX transistor is positive charge trapping within the oxide.

On the other hand, negative charge trapping is completely negligible, as its density at the end of the irradiation is only 0.03% of the trapped positive charge density. This result is the consequence of the high mobility of electrons combined with reduced activation energies due to the presence of the electric field, which drastically decreases the time spent by electrons in the oxide.

Positive charges screen the electrons stored in the floating gate, thereby shifting the I_{DS} (V_{GS}) curve to the left, as shown in **Figure 40 (right)**. This result also explains why it is not possible to remove faults by overwriting the memory. The presence of the trapped positive charges masks the electrons stored in the floating gate, regardless of its logical state. Consequently, until these charges are not completely annealed, the transistor is detected as faulted.

Moreover, the relevancy of this phenomenon justifies the low interfacial currents previously described. Due to their low mobility and the high thickness of the oxide, holes tend to get semi-permanently trapped in the trapping sites of the oxide. Only a small fraction of holes generated close to the surface manage to diffuse towards the floating gate and neutralize the stored electrons.

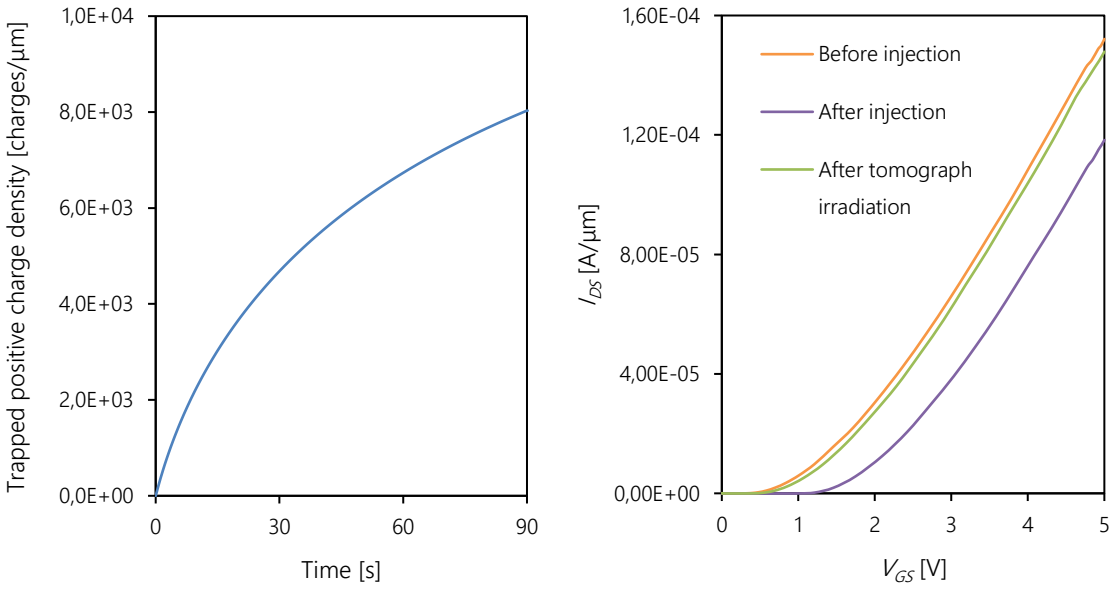

Figure 40: evolution of the trapped positive charge density during tomograph irradiation (left); comparison of the I_{DS} (V_{GS}) curves before programming, after programming and following tomograph irradiation (right).

To verify the coherence of the results provided by ECORCE with experimental observations, several thermal annealing experiments were conducted on the ATmega1284P components following tomograph irradiation. The experimental setup is depicted in **Figure 41**. The samples were locked within the pliers of a sample holder and the tip of a thermocouple was carefully placed in front of the flash memory to measure its temperature. The samples were heated through a heating gun whose temperature was tuned to reach 200 °C on the memory's surface. This temperature was identified as a tradeoff between the instrument's limits, the time required to reach an effective annealing and the minimizing the risk of debonding, which can cause the loss of functionality.

Two distinct behaviors were noted during the thermal annealing of the irradiated samples. Vertical faults were annealed more quickly compared to localized faults, with an average annealing rate of 40 bits per minute and 1 bit per minute respectively. The evolution of the faults in the memory map is shown in **Figure 42 (left, a to c)** for the line faults and in **Figure 42 (left, d and e)** for the localized faults. Additionally, in **Figure 42 (right)** is reported the evolution with time of the number of line faults, where it is visible that their removal follows a linear decreasing rule.

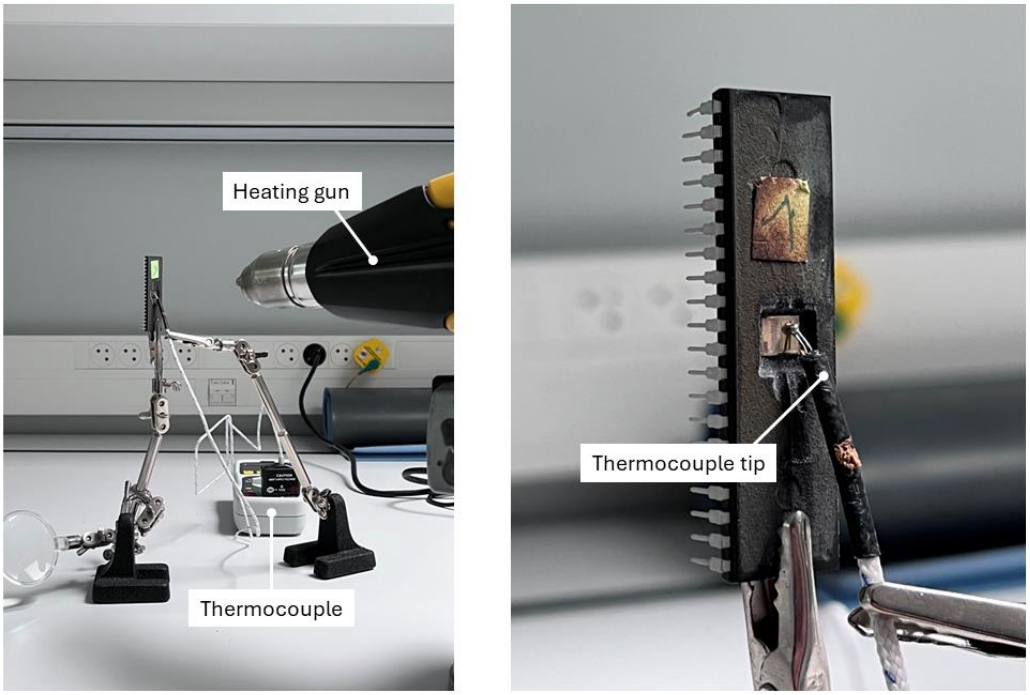

Figure 41: experimental setup for thermal annealing.

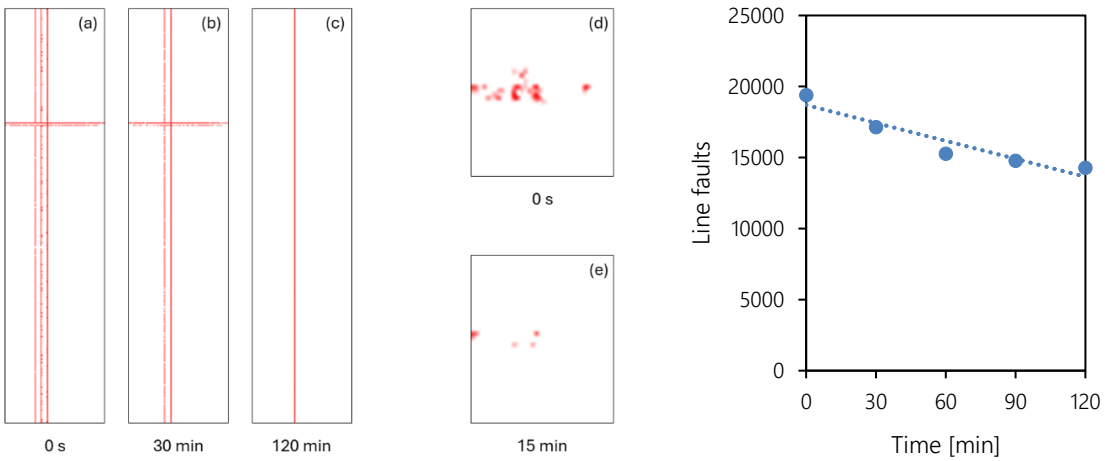

Figure 42: evolution of the faults in the memory maps during the thermal annealing (left) and number of line faults as a function of the annealing time (right).

This behavior can be explained by the fact that the gate width of floating gate transistors is nearly three times larger than that of access MOS transistors (**Table 3**). Consequently, the density of trapped positive charges is significantly lower, reducing the time required to anneal them.

The results from ECORCE simulations were confirmed to be realistic. The approximation made regarding photoemission is reasonable, as it did not appear to be the main mechanism affecting the transistor's electrical behavior. If photoemission was the dominant mechanism, the injected electrons within the floating gate would be expelled during irradiation, allowing the cell to be reprogrammed and the fault to disappear.

However, reprogrammed cells remain detected as faulted, suggesting that even though some electrons may be expelled through photoemission, it is not the primary contributor to the TID effect. For the same reason, if holes injection was the main mechanism depleting electrons in the cell, reprogramming the transistor should eliminate the fault. However, faults only disappear with thermal annealing, indicating that holes trapping in the oxide layers is the primary cause of transistor faulting during tomograph irradiation.

Overall, ECORCE effectively predicted the behavior of the floating gate transistor during tomograph irradiation. The software accurately modeled the homogeneous accumulation of positive charges in the oxide, enabling the exclusion of photoemission and hole injection as the predominant mechanisms. Thermal annealing experiments further validated the software's insights, demonstrating that the transistor's logic state is maintained after trap annealing and confirming that electrons are not expelled during irradiation.

4.3.4 Irradiation simulations with synchrotron radiation

The simulation of synchrotron irradiation involves exposing the transistor to a high dose rate of $7 \cdot 10^{10}$ rad/s for 2 seconds. The beam's radius is 30 nm and is focused on the center of the transistor's gate. As visible from **Figure 43 (left)**, after 1 µs of irradiation the transistor experiences a significant reduction in the stored charge losing $1.11 \cdot 10^{4}$ electrons/µm, which corresponds to 70% of the injected charge. This reduction continues until 1 ms of irradiation, after which the stored charge stabilizes to a constant value.

The primary mechanism responsible for this depletion is the injection of trapped holes from the oxide layer. This can be understood by analyzing the interfacial currents that arise at the lower and upper interface of the floating gate with the surrounding oxide. A scheme is depicted in **Figure 43 (right)**, where the blue arrows symbolize the electrons current and the violet arrows the holes current.

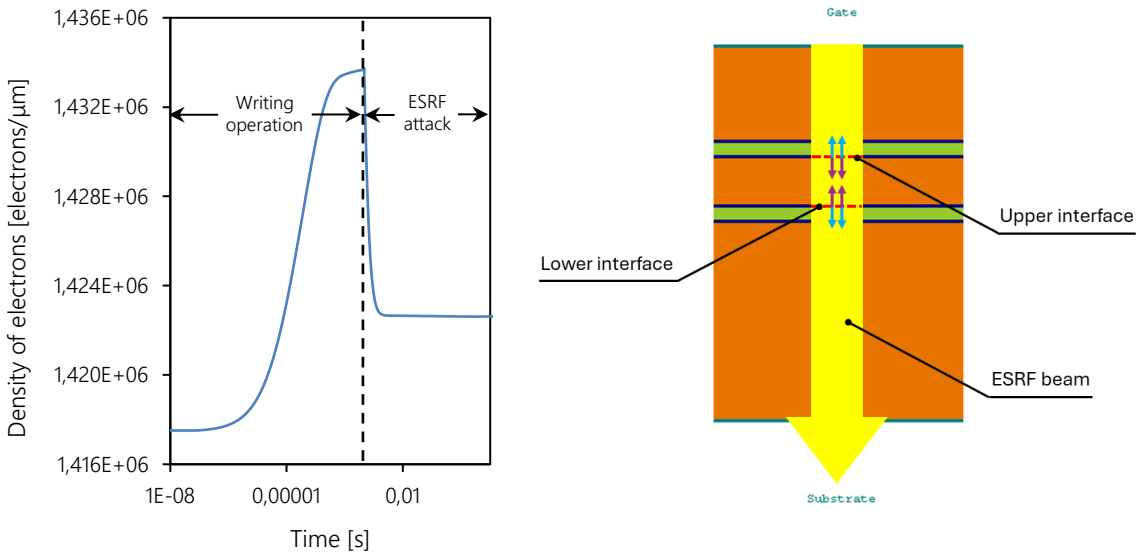

Figure 43: density of electrons within the floating gate during ESRF irradiation (left) and scheme of the interfacial currents (right).

The simulation indicates that, due to the electric field distribution, both trapped electrons and holes at the interfaces are injected into the floating gate. The injected electrons increase the total stored charge, while the holes reduce the charge content by recombining with the stored electrons. The main contribution comes from the upper interfacial currents because the electric field at the beginning of the irradiation is significantly higher in the upper oxide layer compared to the lower oxide layer. Specifically, the electric fields are 0.33 MV/cm in the upper oxide layer and 0.14 MV/cm in the lower oxide layer. This difference affects the activation energies of the traps, which are lower in the upper layer for both electrons and holes, resulting in a higher emission rate in the upper oxide.

Although electrons are injected into the floating gate, the higher rate of hole injection results in a net negative total electron current J_n^{tot}, as illustrated in **Figure 44 (left)**. This quantity is calculated as shown in **Eq. 6**, where J_n^{up} and J_p^{up} are the electrons and holes currents flowing within the beam's diameter through the upper interface, and J_n^{low} and J_p^{low} are the same currents flowing through the lower interface.

$$J_n^{tot} = (J_n^{up} - J_p^{up}) + (J_n^{low} - J_p^{low}) \qquad \textbf{Eq. 6}$$

A negative total electron current indicates that the injected holes recombine not only with all the injected electrons but also with the previously stored charges, therefore determining a net reduction of the negative stored charge. This induces a change in the logical state and a shift of the I_{DS} (V_{GS}) curve as shown in **Figure 44 (right)**, causing the transistor to be detected as faulted. The new V_{th} of the transistor is lower than that of the non-programmed transistor. This can be explained by the combined effects of stored charge reduction and positive charge trapping in the oxide within the beam's diameter. Both mechanisms contribute to the reduction of V_{th}, resulting in a value lower than the initial threshold voltage.

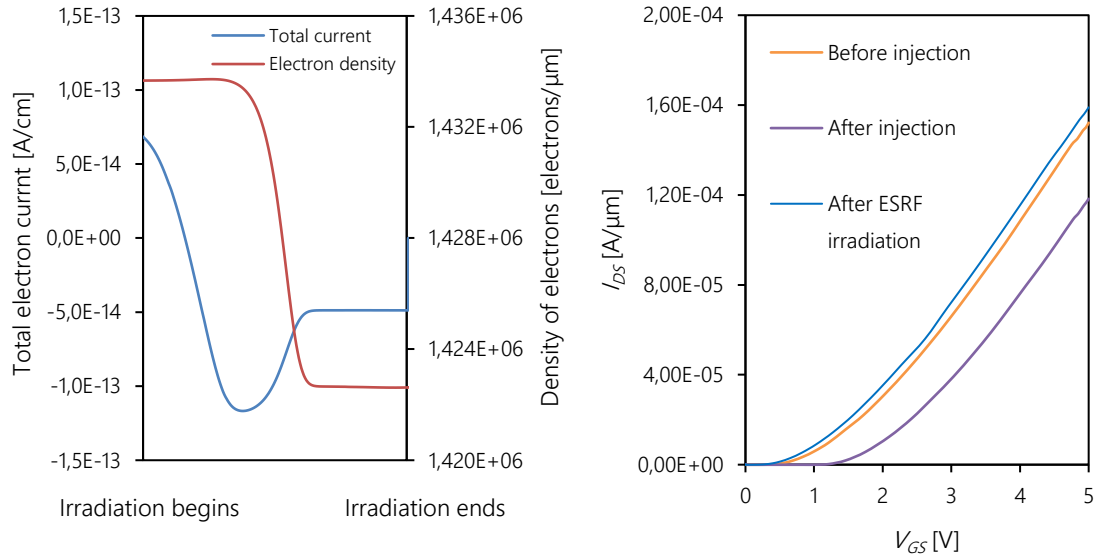

Figure 44: total electrons current as a function of the irradiation time (left) and curves before programming, after programming and after ESRF irradiation.

These results suggest that the transistor should not be detected as faulted anymore following a new writing operation. By performing a Fowler-Nordheim injection simulation using the same parameters as before, the threshold voltage returns to the programmed value. However, real experiments showed that re-writing the memory has no effect on the faults caused by synchrotron radiation, with some faults still being detected even after several days.

Only prolonged thermal treatment can restore the transistors to their initial state, implying that the persistent faulted state is connected to trapped charges in the oxide. Simulations incorporating a higher density of oxide traps and deeper trap levels yield similar outcomes, both in terms of the faulting mechanism and the restoration of the initial state via Fowler-Nordheim injection. This suggests that the semi-permanent faulted state of the transistor is probably related to the STI-related effect discussed in **Section 3.2**.

As visible in the cross-section of **Figure 45**, which was obtained by milling the memory perpendicularly to the cross-section of **Figure 21**, the memory cells are separated by 0.1 µm wide oxide spacers that partially extend beneath the floating-gate transistors. The trapped charges within these STIs can affect the logical state of the transistors, leading to leakage currents that cause the transistor to be detected as faulted until the traps are annealed through a thermal treatment.

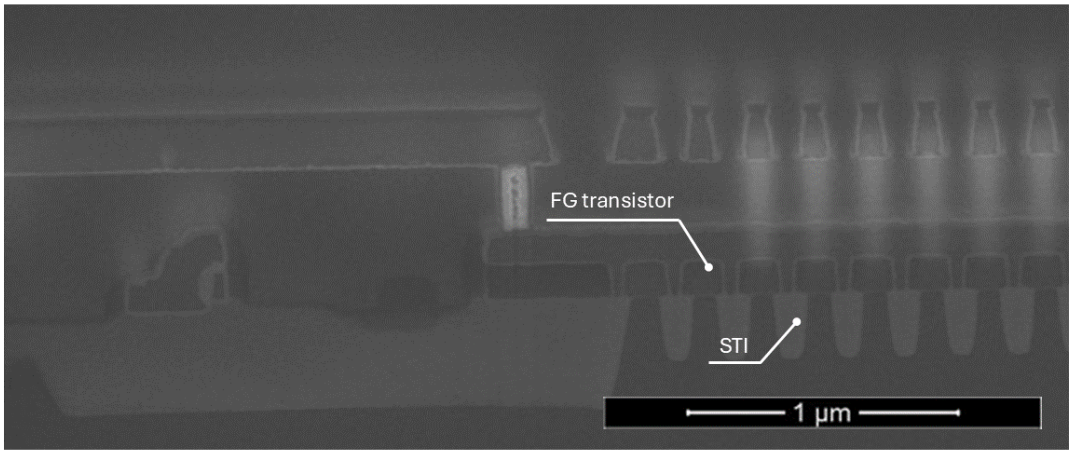

Figure 45: cross-section of the ATmega1284P where the STIs are visible.

Finally, it is interesting to analyze the evolution of the spatial distribution of trapped charges within the oxide layers during synchrotron irradiation. In the case of focalized irradiation, all trapping phenomena are localized within the X-ray beam, leading to a characteristic pattern of charge distribution. Before irradiation, only a small number of electrons are trapped in the oxide as they were captured during the injection phase, while the number of trapped holes remains negligible.

As illustrated in **Figure 46 (left)**, where the total number of positive and negative trapped charges within the beam's diameter is plotted against irradiation time, the total number of trapped electrons before irradiation is 0.7 electrons/µm, with almost no holes trapped. Additionally, the distribution of the electric potential within the lower oxide layer before irradiation is shown in **Figure 46 (right)**.

After just 1 ns of irradiation, both positive and negative trapped charges begin to increase. However, as explained in **Section 3.2**, holes have lower mobility compared to electrons, making them more likely to get trapped. As a result, the number of trapped holes becomes significantly larger than that of trapped electrons after just 1 ns, therefore altering the local electric field and potential distribution within the oxide layer.

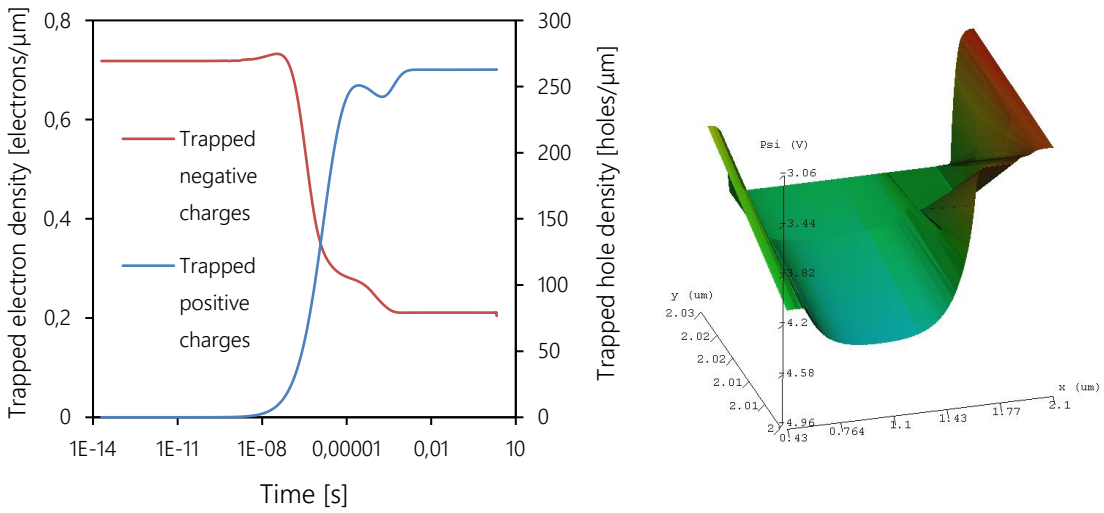

Figure 46: positive and negative trapped charges densities as a function of irradiation time (left) and distribution of the electric potential within the lower oxide layer before irradiation (right).

As shown in **Figure 47 (left)**, which illustrates the electric potential within the lower oxide layer after irradiation, the potential rises locally within the beam region due to positive charge trapping. Since holes move along positive potential gradients, this effect enhances positive trapping within the beam's diameter: previously trapped holes tend to remain in place, while newly generated holes migrate toward the center of the beam. In contrast, electrons move along negative potential gradients, therefore their trapping density experiences a decrease within this region. When previously trapped electrons are re-emitted, they tend to diffuse out of the region following the saddle-shape of the potential, as do the electrons generated from new electron-hole pairs.

This is reflected in the total number of trapped charges plotted in **Figure 46 (left)**, where it is evident that after 10 ns of irradiation, the number of trapped electrons within the beam decreases, while the trapped holes quickly increase to 2500 holes/µm by the end of the irradiation. To better visualize this phenomena, in **Figure 47 (right)** are reported the spatial distribution of positive (above) and negative (below) trapped charges within the upper and lower oxide layers of the transistor.

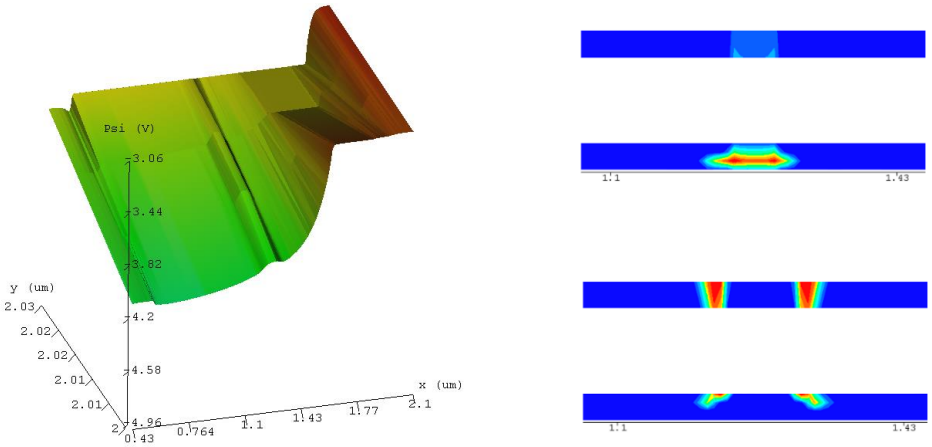

Figure 47: spatial distribution of the electric potential within the lower oxide layer following irradiation (left); distribution of the positive (top right) and negative (bottom right) trapped charges in the upper and lower oxide layers.

ECORCE provided valuable insights about the mechanisms at play during exposure to very high doses of radiation. The simulations revealed that during synchrotron irradiation, charge trapping in the oxide plays a less significant role in altering the transistor's threshold voltage compared to tomograph irradiation. Instead, ECORCE identified hole injection as the primary mechanism altering the transistor state through the recombination of stored electrons. Additionally, the possibility of analyzing the evolution of potential distribution and trapped charge distribution allowed to improve the understanding of the transport phenomena occurring during this type of irradiation. The simulations revealed a characteristic trapped charge distribution that emerges when a transistor is irradiated with an highly focused beam, which is relevantly different from the uniform distribution induced by tomograph radiation.

However, the software's inability to model photoemission limits the completeness of its analysis, especially since photoemission may significantly impact the electron density at very high doses, such as those produced by synchrotron radiation. This limitation prevented the confirmation of simulation results through real annealing experiments. Moreover, the restriction to 2D device modeling did not allow the accurate reproduction of the complex structure of the STIs, which can influence the transistor's electrical behavior by creating leakage currents. Despite these limitations, ECORCE has shown strong potential as a tool for analyzing phenomena during synchrotron radiation. However, the development of new features is necessary to achieve a more detailed description of the underlying mechanisms.

4.4 SRAM simulations

4.4.1 The models

To better understand the CMOS electrical behavior, the NMOS and PMOS transistors were first modeled separately using dimensions obtained through the cross section of **Figure 31**. As visible from **Figure 48**, both transistors are built on a silicon substrate of 0.5 µm thickness, with the lower edge serving as the substrate contact. The gates are made of a SiO_2 layer of thickness 0.03 µm, while the exposed lateral surfaces are used as the source and drain contacts. Specifically, the right surface of the NMOS serves as the ground contact, while V_{DD} is applied on the left surface of the PMOS. All contacts are modeled as ohmic, except for the gate, which exhibits Schottky behavior. The background temperature is set at 300 K.

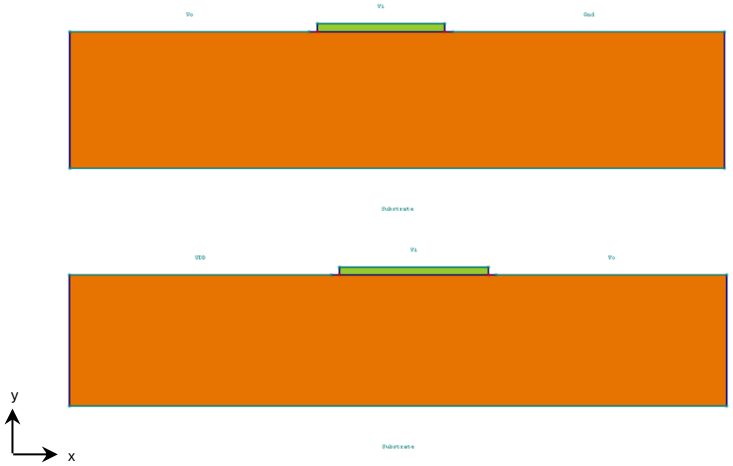

Figure 48: NMOS (above) and PMOS (below) models in ECORCE.

The doping concentration of the substrate for both transistors is set at 10^{16} cm^{-3}, with aluminum as the acceptor atom. The NMOS source and drain contacts are doped with two n^+ regions using arsenic atoms as donors, while the PMOS has two p^+ regions doped with aluminum. All regions have a depth of 0.1 µm and a doping concentration of 10^{19} cm^{-3}. Additionally, the PMOS includes an n-well that extends across its entire width, with a depth of 0.2 µm and an n^+ doping of 10^{17} cm^{-3} using arsenic atoms. A uniform concentration of electron and hole traps is defined within the oxide layer, with concentrations of 10^{17} cm^{-3} and 10^{18} cm^{-3} respectively. 0.8 eV is used as the only energetic level for the electron traps and 1.4 eV for the hole traps.

The model of the complete CMOS structure is shown in **Figure 49**. The two transistors are separated by a spacer with a width of 1.8 µm and a depth of 0.35 µm, and their V_O contacts are connected by a polysilicon line doped with arsenic at a concentration of 10^{19} cm^{-3}. The PMOS also includes an additional contact to ground the n-well.

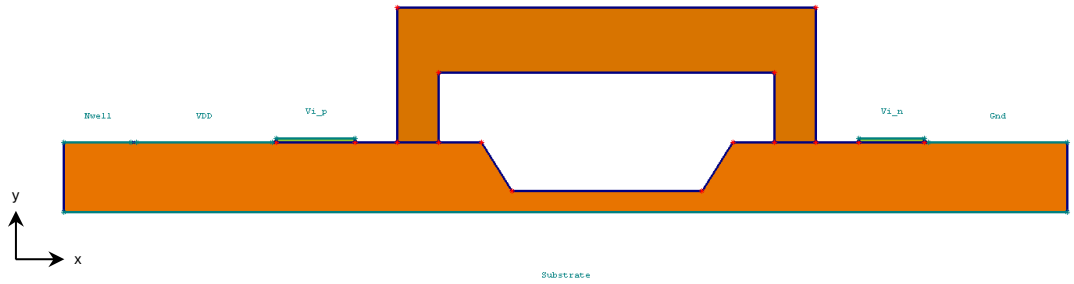

Figure 49: CMOS model in ECORCE.

4.4.2 Electrical characterizations and irradiation simulations

The I_{DS} (V_{GS}) curve of the PMOS transistor is obtained by applying a drain voltage of -1 V and a gate voltage ramp from 0 to -5 V. Similarly, the NMOS transistor is tested by applying a drain voltage of 1 V and ramping the gate voltage from 0 to 5 V. The resulting characteristic curves of both transistors are shown in **Figure 50**, where it is evident that the threshold voltage of the NMOS is close to 0.8 V, while the threshold voltage of the PMOS is around -1.8 V.

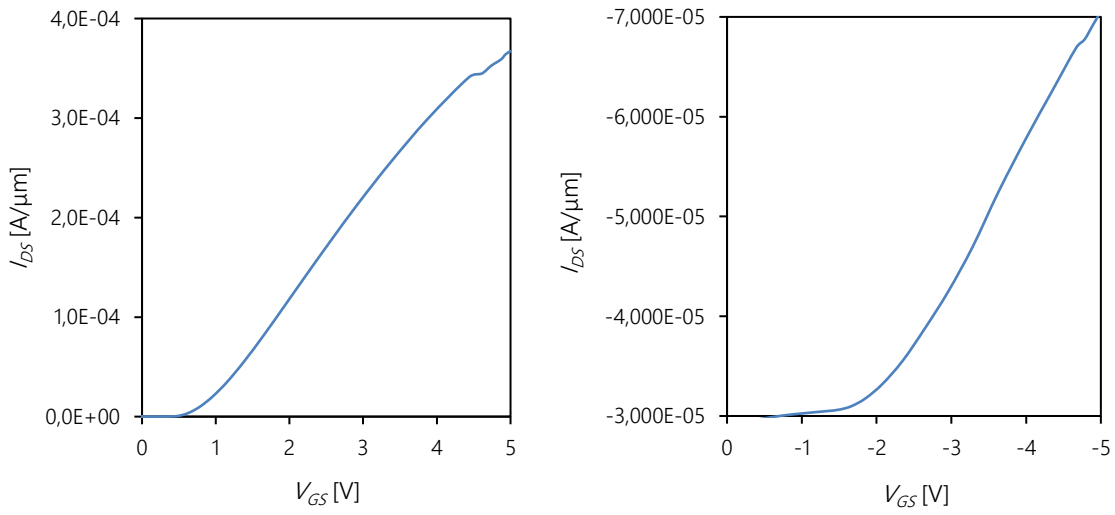

Figure 50: I_{DS} (V_{GS}) curves of the NMOS (left) and PMOS (right).

To obtain the VTC shown in **Figure 3 (right)**, numerous simulations are required to determine the correct value of V_{DD} that produces the step shape of the curve. In each simulation, V_{DD} is fixed at a specific value and the input voltage v_i is swept from 0 to V_{DD} in 50 ns, whereas the substrate and the n-well are grounded. These simulations are performed with V_{DD} values ranging from 2 to 4 V, which are common for this type of transistors. The VTC curve is obtained by probing the electrical potential at the midpoint of the polysilicon line, representing the output voltage v_o, as a function of the input voltage v_i applied to the gates.

The variation in the shape of the VTC with respect to V_{DD} is shown in **Figure 51 (left)**. It is clearly visible that the value of V_{DD} that guarantees a VTC with the steepest transition between the high-state and the low-state is 2.5 V. For higher values, the transition region becomes broader and less well-defined, while for values lower than 2.5 V but higher than 1 V, the transition is too rapid. Additionally, for values lower than 1 V, the simulations do not converge.

4.4.3 Irradiation simulations with laboratory source

The simulation of tomograph irradiation is performed by exposing the CMOS to the calculated dose rate of $1.5 \cdot 10^3$ rad/s for 300 seconds. As with the floating gate transistor, to simulate continuous reading conditions and reduce the computational load, a constant bias of 1 V is applied to the gates and as V_{DD}. Following tomograph irradiation, the VTC shifts to the left, as shown in **Figure 51 (right)**.

This shift causes the semi-permanent alteration of the CMOS output state, which becomes stuck at a fixed value. Assuming that under normal operating conditions the threshold voltage dividing the high state from the low state is at the midpoint of the step (around v_i = 0.5 V and v_o = -3.5 V), it is evident that regardless of the value of v_i, the output voltage v_o remains lower than the v_o value necessary to detect the high state, thereby fixing the CMOS' state at the low state.

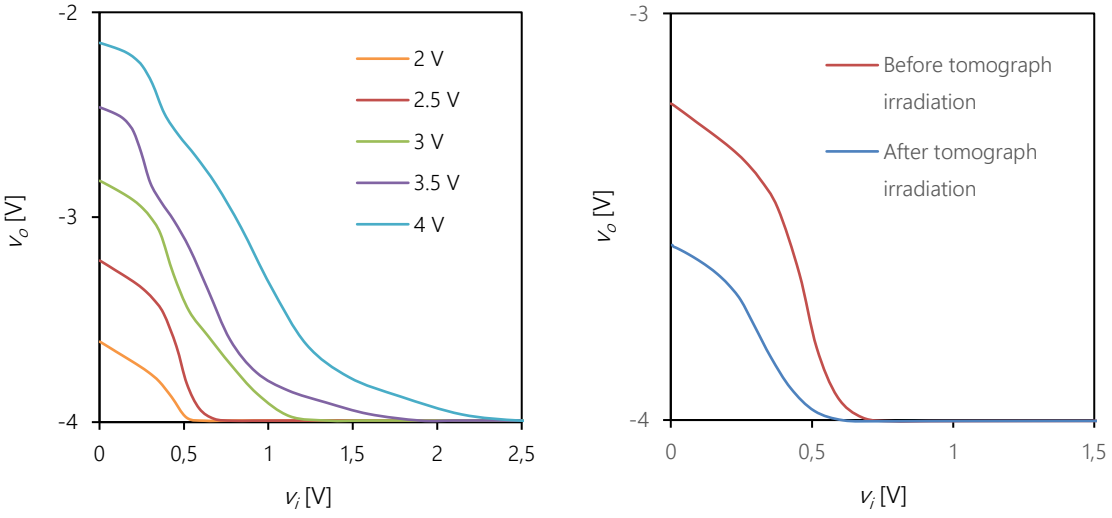

Figure 51: shape of the VTCs as a function of V_{DD} (left) and shape of the VTC before and after tomograph irradiation (right).

The VTC shift results from the combined effect of positive charge trapping in both gate oxide layers. When the same simulation is performed on individual transistors, it shows a threshold voltage shift of -0.3 V for the NMOS and -3.8 V for the PMOS. Given the same dose rate, and therefore the same amount of positive trapped charges, the PMOS experiences a more significant shift in V_{th}.

This outcome reflects the experimental observations made by the MITIX team [40], confirming that PMOS transistors are more sensitive to radiation than NMOS transistors, therefore being easier to fault. This observation can be explained by comparing the magnitude of the electric field inside the two gates. In the PMOS gate, the electric field reaches a maximum of 0.2 MV/cm, while in the NMOS gate the maximum electric field magnitude is 0.3 MV/cm. This difference affects the re-emission probability, which is higher in the NMOS gate due to the Poole-Frenkel effect, leading to fewer trapped holes. As a result, the NMOS transistor result more resistant to faulting.

4.4.4 Irradiation simulations with synchrotron radiation

The simulations of ESRF irradiation are performed by exposing the PMOS and NMOS transistors separately to the calculated dose rate of $7 \cdot 10^{10}$ rad/s for the exposure times detailed in **Section 4.1.5**. The resulting VTCs of the CMOS after NMOS and PMOS irradiation with the ESRF beam are plotted in **Figure 52 (left)**, considering both the ON and OFF states. It is visible that irradiations performed when the device is in the OFF state cause the maximum of the curves to shift upward compared to the non-irradiated device, while irradiations in the ON state result in a downward shift of the maximum.

From this group of simulations, it emerges again that PMOS transistors are generally easier to fault than NMOS transistors when in the ON state. Specifically, the PMOS transistor requires only 25% of the time needed to fault the NMOS transistor to produce a nearly identical downward shift in the VTC. During this time, the density of positive charges accumulating in the PMOS gate is comparable to the density of positive charges that arise in the NMOS after four times the irradiation duration, reaching approximately 200 charges/μm. This result can be attributed to the previously mentioned argument: even though the electric field locally changes during irradiation, a consistent difference of almost 0.1 MV/cm remains between the maximum electric fields of the two transistors.

Conversely, PMOS transistors are more difficult to fault than NMOS transistors when the device is in the OFF state. This again can be explained by the difference in electric fields: when no biases are applied, the electric field in the PMOS is higher than in the NMOS, making the PMOS less susceptible to faults in the OFF state.

Finally, it is interesting to compare the transfer curves of the device in the ON state following ESRF and tomograph irradiation. As shown in **Figure 52 (right)**, tomograph radiation induces a more pronounced downward shift in the maximum of the transfer curve. This effect can be attributed to the higher density of trapped holes, which results from the longer irradiation time - 300 seconds compared to 200 and 50 seconds - and the broader irradiated area. Unlike the focused ESRF beam, which targets only a portion of the gates, tomograph radiation affects the entire gate region, leading to a more significant accumulation of trapped charges and consequently a greater shift in the transfer curve.

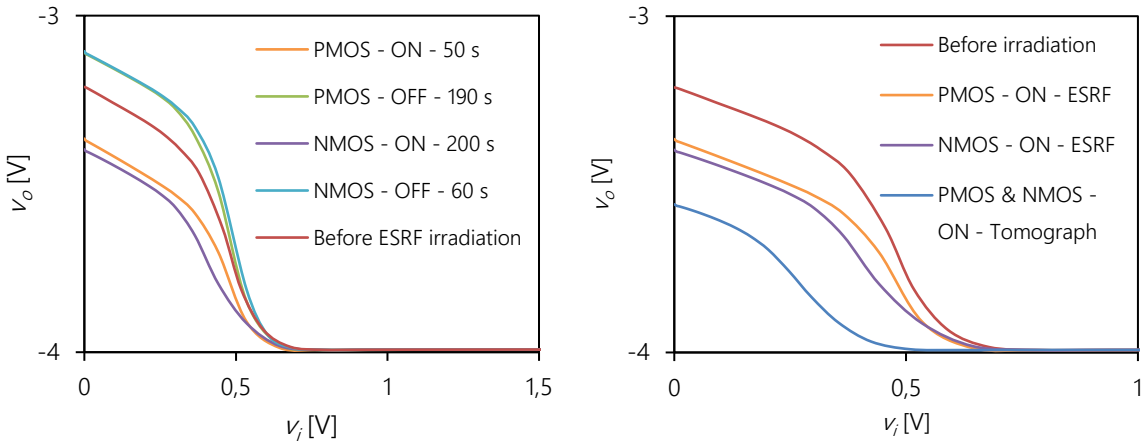

Figure 52: shift of the VTC during ESRF irradiation (left) and comparison of the VTC shift following ESRF and tomograph irradiation (right).

Overall, ECORCE provided valuable insights about the response of CMOS devices to tomograph and ESRF irradiations. The simulations using the laboratory source identified positive charge trapping as the primary mechanism responsible for memory cell faults. Furthermore, the observed shifts in threshold voltages following irradiation confirmed the findings of the MITIX team, which noted a higher sensitivity of PMOS transistors compared to NMOS transistors when in the ON state. ECORCE helped clarify the reason behind this difference by identifying the lower electric field in the PMOS gate as the cause for a higher density of trapped holes.

Additionally, the simulations with localized X-rays allowed to obtain a clearer visualization of the shifts in the VTCs, showing a downward shift in the ON state and an upward shift in the OFF state. This analysis helped explain why NMOS transistors are more susceptible to radiation with respect to PMOS when in the OFF state, attributing this vulnerability to the higher electric field within the gate when no external biases are applied.

Chapter 5

DRAM deprocessing

The DRAM of the Zybo-Z7 development board was chosen as a new type of memory to study its sensitivity to X-ray single bit attacks. To perform meaningful attacks and corrupt a new circuit, it is necessary to have a deep understanding of its hardware architecture, both at the macroscopic and microscopic scale. Knowing the relationship between the physical and logical addresses enables precise targeting of a memory cell and determination of whether its logical state has been successfully altered or not. Therefore, a preliminary deprocessing of the memory was necessary to lay the foundations for future studies that will aim to retrieve the physical-logical correlation of the memory addresses.

5.1 Chemical etching and optical analysis

Differently from the ATmega1284P microcontroller, where the memories are encapsulated in a large plastic package that communicates through the external pins, the studied DRAM is entirely covered by a thin plastic package and is directly soldered to the board. **Figure 53 (left)** shows a picture of the Zybo-Z7 board with the DRAM location highlighted by a circle. To separate it from the board, a heating torch was used for three minutes to locally raise the temperature and detach the memory. **Figure 53 (right)** shows the board without the memory, with the soldering pads used to keep the memory attached now visible.

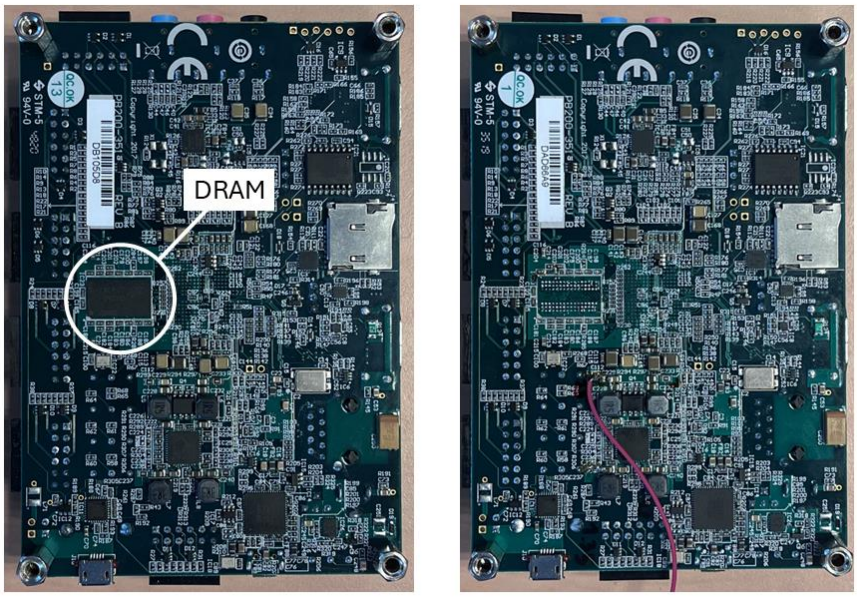

Figure 53: Zybo-Z7 board before and after DRAM removal.

A bath in fuming nitric acid at 80°C was employed to remove the plastic package, and the optical image of the depackaged memory is shown in **Figure 54 (left)**. The memory was then submerged in diluted hydrofluoric acid (50% by volume) at room temperature to etch the oxide layers and, through a lift-off mechanism, the metal layers above as well as any residual plastic. Finally, the memory was placed in sulfuric acid, which was heated to 50°C. Once the temperature was reached, hydrogen peroxide was added to effectively remove the remaining metal layers and expose the transistors. The optical image of the memory following the three etching steps is shown in **Figure 54 (right)**.

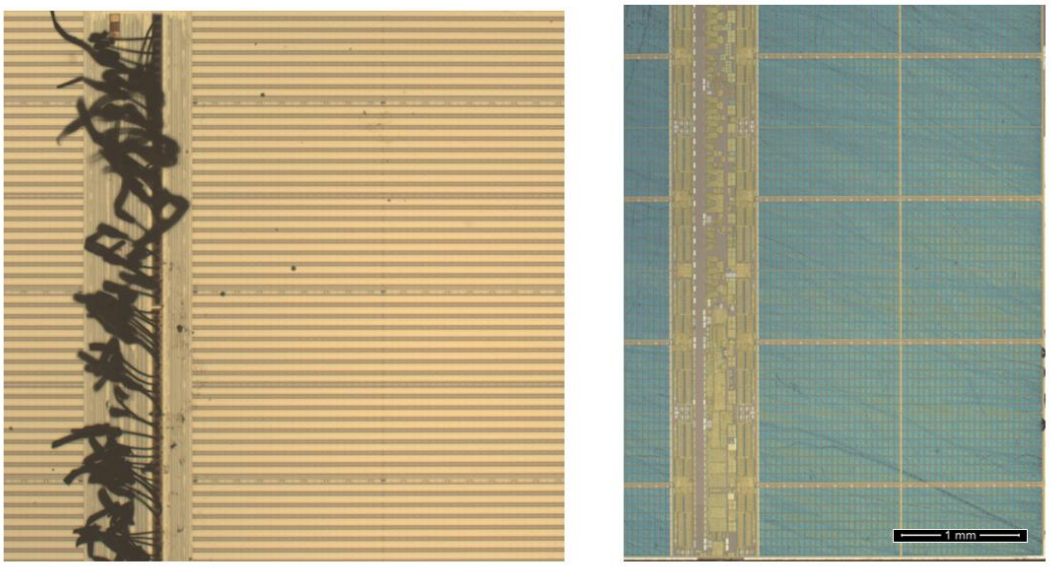

Figure 54: optical image of the DRAM after package removal (left) and after oxide and metal layers removal (right).

It is clearly noticeable that the memory cells are grouped into two halves, separated by a central vertical line that collects the peripheral circuits. Each half comprises six square regions, each with an edge length of 1.4 mm, and four rectangular regions with a minor edge length of 0.7 mm. Within each region, there are subsections measuring 80 µm x 45 µm; each rectangular region contains 200 subsections, while each square region contains 400 subsections.

5.2 PFIB analysis

To achieve a better understanding of the memory architecture at the nanometric scale, a PFIB analysis was employed, using the *Helios 5 DualBeam Plasma-FIB* from *ThermoFisher Scientific*. Firstly, a sample that underwent the three chemical etching steps was studied, and its etched surface appearance is shown in the electronic microscope image in **Figure 55 (left)**. The rectangular region at the bottom-left corner belongs to one of the subsections previously described, while the remaining circuitry corresponds to the peripheral circuits that separate each region.

From **Figure 55 (right)**, it is visible that each memory cell consists of three polysilicon squares tilted at an angle of 60°. Each cell has a total length of 125 nm and a width of 50 nm, while the distance between two adjacent cells is 25 nm. The side squares have a length of 25 nm, while the central region has a length of 30 nm.

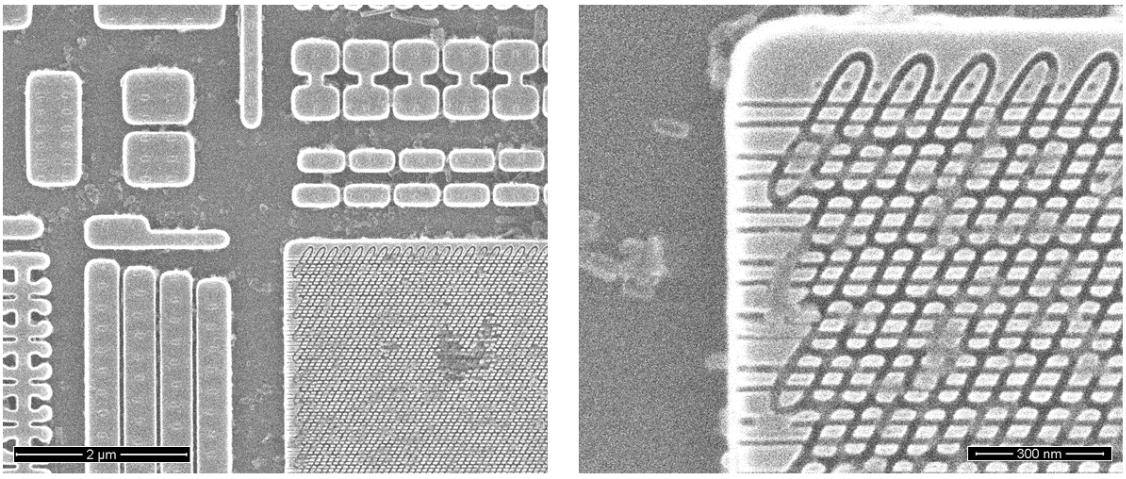

Figure 55: SEM view of the peripheral circuits and a part of the memory block (left); close up view of the memory block (right).

Following the same procedure described in the previous sections, a PtC mask of 2 µm thickness was deposited, and the cross-section was milled. As visible from **Figure 56 (left)**, each cell is composed of three polysilicon arms of height 80 nm that lie on a common polysilicon base, for a total height of 220 nm. Comparing this structure with the scheme in **Figure 5 (right)**, it is possible to assume that the void regions may contain the insulator that is removed through the chemical etching, together with the polysilicon that fills the trench through a lift-off mechanism. To confirm this hypothesis, an additional cross-section was performed on a sample treated only with fuming nitric acid. However, as visible from **Figure 56 (right)**, the oblique microstructure as well as the high aspect ratio of the trenches did not allow to obtain a clear milling surface.

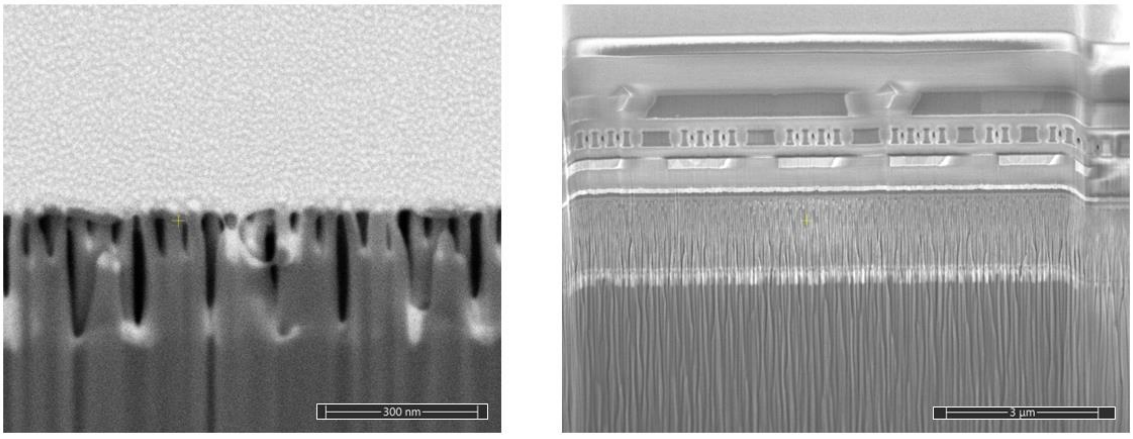

Figure 56: cross-section of the DRAM following the three etching steps (left) and cross-section following the removal of the plastic package (right).

Chapter 6

Conclusions

The increasing levels of cybersecurity require an intense research for new attack methodologies capable of bypassing existing protections, both at the software and hardware levels. Among these, single-bit attacks using X-rays have shown significant potential as a new mean of perturbation, demonstrating the capability to fault a PIN authentication program by corrupting only a single transistor. To enhance the understanding of the microscopic mechanisms involved in these attacks, and thereby increase the security levels against them, TCAD simulations represent a valuable tool.

To perform insightful simulations, especially on unknown circuits such as the ATmega1284P where the GDSII file is protected by the industrial secret, preliminary experiments are crucial for retrieving key parameters necessary for the simulations. This Thesis work provided a comprehensive summary for future attacks, outlining all the essential experiments needed to perform irradiation simulations on single transistors.

The presented work highlighted the importance of conducting cross-sections using a Focused Ion Beam (FIB) to accurately determine the dimensions of transistors that have to be modeled. Additionally, the experiments conducted in this Thesis work pointed out the necessity of employing a thick mask over the milled region to achieve clean cross-sections and avoid surface irregularities that could compromise the accuracy of the measurements, and thereby of the simulations.

Moreover, the Thesis emphasized the importance of an appropriate mask layout when performing irradiation experiments with a laboratory source to obtain precise exposure parameters. Specifically, it was found that masks deposited directly on the circuit surface are inefficient, as they fail to effectively isolate and expose a single transistor. In contrast, a movable mask layout proved to be the most effective, enabling systematic attacks and precise determination of the faulting time.

Finally, for the first time, a dedicated Python code was developed to evaluate the dose rate absorbed by the gate oxides, accounting for the device stratification observed through the cross-sections. All in all, this Thesis work emphasized that the experimental measurement of the faulting times, combined with the dose rates calculated through the dedicated code and the accurate dimensions of the transistors, represent the key input parameters for conducting useful simulations.

ECORCE was chosen as the TCAD simulation tool because of its capability to model trapped charges transport phenomena through the Multiple Trapping Detrapping (MTD) theory. Despite its original purpose of simulating space radiation interactions, the software proved to have almost all the features to model single-bit attacks, enabling to obtain valuable insights about the microscopic mechanisms at play during irradiation. This Thesis highlighted both the strengths and limitations of the ECORCE software, offering insights for its future development.

ECORCE proved to be effective in simulating tomograph attacks on floating gate transistors, clearly identifying positive charge trapping in the gate oxide as the primary mechanism that alters the transistor's logical state. Although the software currently lacks the ability to simulate photoemission, the combination of thermal annealing experiments and the measurement of the oxide's significant thickness brought to the conclusion that photoemission is not the dominant mechanism at low radiation doses.

ESRF simulations suggested that hole injection from the oxide layers may play a significant role in reducing the charge content of floating gate transistors, thereby altering their logical state. However, the absence of the photoemission model, coupled with the software limitations in modelling 2D devices that prevent the representation of STI structures, did not allow to completely exclude photoemission as a contributing mechanism. As a result, the conclusions drawn from these simulations are necessarily partial and indicate the need for further software development to provide a more comprehensive analysis.

Furthermore, ECORCE proved to be efficient in simulating the response of CMOS devices to both tomograph and ESRF radiation, enabling the visualization of the VTC shifts as a function of the radiation source and the presence or absence of applied voltages. Tomograph simulations identified charge trapping in the gate oxides as the primary mechanism affecting the transistors. Moreover, they provided an explanation for the observation made by the MITIX team regarding the higher sensitivity of PMOS transistors compared to NMOS transistors when in the ON state. On the other hand, ESRF simulations revealed that the VTCs shift downward when irradiation occurs during the ON state and upward when the device is in the OFF state. Finally, synchrotron radiation simulations highlighted the higher sensitivity of NMOS transistors compared to PMOS transistors when in the OFF state.

The distinct behaviors of these two transistors, following both tomograph and ESRF irradiation, were explained by analyzing the electric field distribution within their gate oxides. The different intensity of the electric field, and consequently the differences in re-emission rates, was identified as the main factor influencing the density of trapped positive charges within the oxides.

In conclusion, ECORCE proved to be a valid software for simulating scenarios where positive charge trapping is the only or the main mechanism at play during irradiation. On the other hand, the absence of a solid photoemission model did not allow to exclude this mechanism in the interaction of localized X-rays at high doses, making the current version of ECORCE unsuitable to model floating gate transistors.

To improve the efficiency of the simulations, future research should shift focus from studying a single transistor within a memory array as done in this Thesis work, to studying an isolated transistor. This would allow the direct measurement of the real I_{DS} (V_{GS}) curves, which could then be compared with those simulated by ECORCE. Moreover, conducting irradiation experiments on an isolated transistor would help to better estimate the roles of the interaction mechanisms identified in this work. Additionally, a study of the doping concentrations and the distribution of trap levels would reduce the number of approximations necessary to build the model. Another important improvement would be the ability to simulate 3D structures, which would enable the modeling of leakage currents caused by the STIs, a factor not currently considered in ECORCE.

Overall, the ability to study a single transistor and compare its real response to external stimuli with simulation results would make ECORCE an ideal tool for simulating single-bit attacks. This would enhance the efficiency of the attacks, thereby helping the researchers to study new countermeasures against them.

Bibliography

[1] "https://mitix.cea.fr/."

[2] S. Anceau, P. Bleuet, J. Clédière, L. Maingault, J. Rainard, and R. Tucoulou, "Nanofocused X-Ray Beam to Reprogram Secure Circuits," 2017, pp. 175–188. doi: 10.1007/978-3-319-66787-4_9.

[3] L. Maingault *et al.*, "Laboratory X-rays Operando Single Bit Attacks on Flash Memory Cells," 2022, pp. 139–150. doi: 10.1007/978-3-030-97348-3_8.

[4] T. R. Oldham and F. B. McLean, "Total ionizing dose effects in MOS oxides and devices," *IEEE Trans Nucl Sci*, vol. 50, no. 3, pp. 483–499, Jun. 2003, doi: 10.1109/TNS.2003.812927.

[5] G. Conte, F. Cerri, and D. Tomassini, *Nuovo corso di elettrotecnica ed elettronica*. Hoepli, 2023.

[6] Pieraccini Massimiliano, *Fondamenti di elettronica*. Pearson, 2014.

[7] S. Yu, *Semiconductor Memory Devices and Circuits*. Boca Raton: CRC Press, 2022. doi: 10.1201/9781003138747.

[8] "JEDEC, https://www.jedec.org/."

[9] L. Zhao, "Structural Design of an Electrically Erasable EEPROM Memory Cell," *World Journal of Engineering and Technology*, vol. 08, no. 02, pp. 179–187, 2020, doi: 10.4236/wjet.2020.82015.

[10] J. E. Martin, *Physics for Radiation Protection*. Wiley, 2013. doi: 10.1002/9783527667062.

[11] J. E. Turner, *Atoms, Radiation, and Radiation Protection*. Wiley, 2007. doi: 10.1002/9783527616978.

[12] R. Micheloni, L. Crippa, and A. Marelli, *Inside NAND Flash Memories*. Dordrecht: Springer Netherlands, 2010. doi: 10.1007/978-90-481-9431-5.

[13] U. Gatti and C. Calligaro, *Rad-hard Semiconductor Memories*. New York: River Publishers, 2022. doi: 10.1201/9781003339182.

[14] J. R. Schwank *et al.*, "Radiation Effects in MOS Oxides," *IEEE Trans Nucl Sci*, vol. 55, no. 4, pp. 1833–1853, Aug. 2008, doi: 10.1109/TNS.2008.2001040.

[15] Borghello G., "Ionizing radiation effects in nanoscale CMOS technologies exposed to ultra-high doses," University of Udine, 2018.

[16] Escoffier R., "Simulation numérique de l'effet des charges induites par irradiation dans les oxydes de structures MOS.," Université de Montpellier II, 1995.

[17] O. L. Curtis and J. R. Srour, "The multiple-trapping model and hole transport in SiO2," *J Appl Phys*, vol. 48, no. 9, pp. 3819–3828, Sep. 1977, doi: 10.1063/1.324248.

[18] M. H. Cohen, H. Fritzsche, and S. R. Ovshinsky, "Simple Band Model for Amorphous Semiconducting Alloys," *Phys Rev Lett*, vol. 22, no. 20, pp. 1065–1068, May 1969, doi: 10.1103/PhysRevLett.22.1065.

[19] Cirba C., " Simulation numerique du piegeage et du depiegeage dans les oxydes de composants MOS.," Université de Montpellier II, 1996.

[20] T. R. Oldham, *Ionizing Radiation Effects in MOS Oxides*. WORLD SCIENTIFIC, 2000. doi: 10.1142/3655.

[21] S. N. Rashkeev, D. M. Fleetwood, R. D. Schrimpf, and S. T. Pantelides, "Proton-induced defect generation at the Si-SiO/sub 2/ interface," *IEEE Trans Nucl Sci*, vol. 48, no. 6, pp. 2086–2092, Dec. 2001, doi: 10.1109/23.983177.

[22] C.-M. Zhang *et al.*, "Characterization and Modeling of Gigarad-TID-Induced Drain Leakage Current of 28-nm Bulk MOSFETs," *IEEE Trans Nucl Sci*, vol. 66, no. 1, pp. 38–47, Jan. 2019, doi: 10.1109/TNS.2018.2878105.

[23] F. Faccio and G. Cervelli, "Radiation-induced edge effects in deep submicron CMOS transistors," *IEEE Trans Nucl Sci*, vol. 52, no. 6, pp. 2413–2420, Dec. 2005, doi: 10.1109/TNS.2005.860698.

[24] T. S. Nidhin, A. Bhattacharyya, R. P. Behera, T. Jayanthi, and K. Velusamy, "Understanding radiation effects in SRAM-based field programmable gate arrays for implementing instrumentation and control systems of nuclear power plants," *Nuclear Engineering and Technology*, vol. 49, no. 8, pp. 1589–1599, Dec. 2017, doi: 10.1016/j.net.2017.09.002.

[25] G. Lee, M. Suh, M. Ryu, Y. Lee, J.-W. Han, and J. Kim, "Investigation Into the Degradation of DDR4 DRAM Owing to Total Ionizing Dose Effects," *IEEE Access*, vol. 11, pp. 97456–97465, 2023, doi: 10.1109/ACCESS.2023.3312926.

[26] S. Gerardin and A. Paccagnella, "Present and Future Non-Volatile Memories for Space," *IEEE Trans Nucl Sci*, Dec. 2010, doi: 10.1109/TNS.2010.2084101.

[27] D. H. Habing, "The Use of Lasers to Simulate Radiation-Induced Transients in Semiconductor Devices and Circuits," *IEEE Trans Nucl Sci*, vol. 12, no. 5, pp. 91–100, 1965, doi: 10.1109/TNS.1965.4323904.

[28] S. P. Skorobogatov and R. J. Anderson, "Optical Fault Induction Attacks," 2003, pp. 2–12. doi: 10.1007/3-540-36400-5_2.

[29] Viera A. C., "Simulating and Modeling the Effects of Laser Fault Injection on Integrated Circuits," Université de Montpellier, 2018.

[30] Abbe, "The Relation of Aperture and Power in the Microscope," *Journal of the Royal Microscopical Society*, vol. 2, no. 4, pp. 460–473, Aug. 1882, doi: 10.1111/j.1365-2818.1882.tb04805.x.

[31] Z. Zhao and C. Fan, *Synchrotron Radiation in Materials Science*. Wiley, 2018. doi: 10.1002/9783527697106.

[32] P. Willmott, *An Introduction to Synchrotron Radiation*. Wiley, 2011. doi: 10.1002/9781119970958.

[33] "https://www.nsrrc.org.tw/English/lightsource.aspx."

[34] G. Martínez-Criado *et al.*, "ID16B: a hard X-ray nanoprobe beamline at the ESRF for nano-analysis," *J Synchrotron Radiat*, vol. 23, no. 1, pp. 344–352, Jan. 2016, doi: 10.1107/S1600577515019839.

[35] S. Carmignato, W. Dewulf, and R. Leach, Eds., *Industrial X-Ray Computed Tomography*. Cham: Springer International Publishing, 2018. doi: 10.1007/978-3-319-59573-3.

[36] S. F. Wang *et al.*, "Respective radiation output characteristics of transmission-target and reflection-target X-ray tubes with the same beam quality," *Radiation Physics and Chemistry*, vol. 158, pp. 188–193, May 2019, doi: 10.1016/j.radphyschem.2019.02.005.

[37] "https://physics.nist.gov/PhysRefData/FFast/html/form.html."

[38] A. Michez, S. Dhombres, and J. Boch, "ECORCE: A TCAD Tool for Total Ionizing Dose and Single Event Effect Modeling," *IEEE Trans Nucl Sci*, vol. 62, no. 4, pp. 1516–1527, Aug. 2015, doi: 10.1109/TNS.2015.2449281.

[39] M. J. Marinella, "Radiation Effects in Advanced and Emerging Nonvolatile Memories," *IEEE Trans Nucl Sci*, vol. 68, no. 5, pp. 546–572, May 2021, doi: 10.1109/TNS.2021.3074139.

[40] S. Bouat, S. Anceau, L. Maingault, J. Clédière, L. Salvo, and R. Tucoulou, "X ray nanoprobe for fault attacks and circuit edits on 28-nm integrated circuits," in *2023 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT)*, IEEE, Oct. 2023, pp. 1–6. doi: 10.1109/DFT59622.2023.10313553.