Design and Mitigation techniques of Radiation induced SEEs on Open-Source Embedded Static RAMs

Design and Mitigation techniques of Radiation induced SEEs on Open-Source Embedded Static RAMs

S. Azimi, C. De Sio, A. Portaluri, L. Sterpone

[1]Dipartimento di Automatica e Informatica, Politecnico di Torino, Torino, Italy
{sarah.azimi, corrado.desio, andrea.portaluri, luca.sterpone}@polito.it

Abstract. Static RAM modules are widely adopted in high performance systems. Single Event Effects (SEEs) resilient memories are required in many embedded systems applied in automotive and aerospace applications to increase their overall resiliency against SEEs. The current SEE resilient SRAM modules are obtained by applying radiation-hardened by design solutions which leads to elevated area overhead and difficulty to tune the resiliency capability with respect to the particle's radiation profile. To overcome these limitations, we propose a methodology for the analysis and mitigation of embedded SRAMs generated by the OpenRAM memory compiler. A technology-oriented radiation analysis tool is presented to support the interaction of the charged radiation particles with the SRAM layout and depict the sensitive transistors of the SRAM memory. A selective duplication of the sensitive transistors has been applied to the 6T-SRAM cell designed at the layout level. The designed cell is included in the OpenRAM compiler and used to generate a mitigated 8Kb SRAM-bank, a DMA interface is also added to the bank in order to evaluate the interface capabilities. We evaluated the SEEs sensitivity by comparative simulation-based radiation analysis observing a reduction more than 6 times with respect to the original 6T-SRAM cell for the SEE sensitivity at high energy heavy ions particles, with negligible degradation of operations margins and power consumption and area overhead of less than~ 4%. The performance of the developed OpenRAM module has been also evaluated considering its application on a neural network behavioral model that demonstrate the feasibility of the proposed solution on large scale memory block circuitry.

Keywords: Radiation Effects, Single Event Effects, SRAM memory, Transistor Layout.

1 Introduction

Embedded Static Random Access Memories are widely applied in various kinds of commercial applications, and they are today an integrated module of aerospace and automotive microprocessor systems [1] . RAMs are crucial components in System-on-chips (SoCs) and due to their wide application, SRAM modules are characterized by several memory configuration requirements and constraints especially when they

are adopted in harsh environments [2] . SRAMs are vulnerable to two main effects, on one side they are really sensitive to wear-out mechanisms such as aging, where the Bias Temperature Instability (BTI) has been discovered as the main reliability concern. On the other side, SRAM cells are extremely sensitive to radiation-induced errors such as Single Event Effects (SEEs) caused by charged particles passing through the semiconductor device and generating electron-hole pairs along the particle track. The collected charge (Q_{coll}) of electron-hole pairs may change the memory cell state in the case it is greater than the critical charge (Q_{crit}) [3] . The radiation sensitivity of embedded SRAM is more emphasized considering that the area used by SRAM memory is dominating the physical layout of CPUs or GPUs [3] . Hence, SRAM layout is typically characterized by minimum device geometry that tends to reduce the Q_{crit} and conversely increase the sensitivity to radiation-induced errors. In order to manufacture robust SRAM modules and to increase the immunity to SEE, design and mitigation strategies for SRAM apply radiation-hardened-by-design (RHBD) that are generally adopting special epitaxial or eventually SOI substrate to limit ionizing radiation particle track length and including high-density capacitors and resistors to avoid circuit response to the collected charge [4] . Since the elevated cost of RHBD, typically error detection and correction (EDAC) approaches are applied to SRAM modules such as caches and shared memory [5] . However, the inclusion of extra combinational logic, such as the one used for Error Correction Mechanism, may also increase the occurrences of Single Event Transients (SETs) since these errors are not easily protected by EDAC.

Considering the growing role of embedded SRAM in system performances, several memory compiler tools have been recently developed [6] . The need for these tools was supported by the fact that most academic ICs design approaches are limited by the effective availability of memories. Nowadays, with the advent of an open-source customizable compiler, researchers are able to design their own memory module with the proper regular structure and configuration. This represents an undoubted advantage in hardware design since the basic building blocks are provided by foundries in technology process design kits (PDKs) and they are essential for hardware and device realization [7] . Thanks to the availability of open-source PDKs for RAMs, several researchers recently started to investigate the applicability of reliability analysis and mitigation of open-source hardware designs [8] .

In this work, we propose an analysis and mitigation framework targeting the Single Event Effects (SEE) radiation-induced phenomena on Open-RAM physical design, extending the methodology introduced [8] . The OpenRAM project is an open-source memory compiler freely available under BSC license [6] . The compiler may be used for the design of new architectures in order to evaluate power, performances and area overhead on the other side, OpenRAM is usable to prototype and evaluate technological modification [10] We selected Cyclone, the cyclotron of the University Catholic the Louvain (UCL) heavy ions high-penetration cocktail as a reference for the radiation characterization. We perform a complete radiation sensitivity evaluation of the RAM block core cells in order to individuate the SEE cross-section and potential weak points of each memory component. Secondly, we apply selective mitigation solutions to the 6T cell of the SRAM by hardening the most sensitive transistor. Final-

ly, we evaluated the performances and fault tolerance capability considering the OpenRAM bank connected to a DMA module and stimulated in different data transfer conditions. Furthermore, an extended performance evaluation of the developed solution has been integrating the developed OpenRAM memory bank into the simulation model of a Neural Network with DSP-oriented neuron architecture.

The hardening insertion has been automatized by the development of a tool to manipulate the physical layout of each cell and to insert a resized duplicated transistor into the 6T cell physical layout structure capable to increase the critical charge of the 6T cell structure.

This work has two main scientific contributions. The former is characterized by the first heavy ions radiation sensitivity evaluation of open-source embedded RAMs showing promising results and a large margin of improvements. The second is the realization of a tool to manipulate physical layout on large scale and to introduce mitigation strategies, the tool may be also ported to commercial technology nodes using library technology files. In order to evaluate the developed methodology, we designed two memory blocks using the physical implementations at 45nm technology with the OpenRAM compiler adopting the FreePDK45 design kit and we performed comparative simulation-based radiation analysis. Experimental results demonstrate that the mitigated Open-RAM memory is approximately 35% more robust than the original Open-RAM design with a marginal degradation of the circuit performance and an area overhead of less than 4%.

This paper is organized as follows. Section II presents previous works related to analysis and mitigation methods for SEE effects on SRAM modules. Section III gives an overview on the OpenRAM analysis method, while the mitigation approach is described in Section IV. The experimental results are reported in Section V. Finally, Section VI drafts some conclusions and future works.

2 Related Works

Several previous works have already analyzed the impact of radiation particles on SRAM cells. With the progressive technology scaling, the number of errors within SRAM module drastically increases. This effect can be explained by both the SRAM cell junction reduction and by the reduced space between cells and lower values of critical charge [10] [11] .

Radiation tests and 3-D simulations already demonstrated that the bipolar parasitic physical mechanism of the MOS transistors is activated by radiation particle strikes and is the cause of memory upsets [12] . In the last decade, real-time radiation test explored the sensitivity of 45nm SRAM modules and identify the soft error projection with respect to the type of radiation particles and energies. The obtained results demonstrated the importance of the device manufacturing and the thick interconnect metallization and dielectric layers with respect to the effective sensitivity to charged particles [13] .

On the other hand, two main categories of mitigation techniques targeting the corrections of single and multiple cell upsets were proposed. The first category relies on

the insertion of Error Correction Code (ECC) mechanisms at the architectural level, while effective for Single Event Upsets (SEUs) these approaches are not able to cope with many Multiple Event Upsets (MEUs), usually happening in the same word and not necessarily in adjacent cells [14] . Besides, traditional error detection and correction approaches introduce critical timing on memory access making these solutions difficulty applicable in cache memories [15] .

The second category is based on radiation-hardened-by-design (RHBD) techniques that allows to apply radiation mitigation circuit solutions to the manufacturing process of commercial foundries in order to minimize the impact of radiation particles [16] . The focus of RBHD techniques have been on Single Event Upsets (SEUs) affecting the 6T-SRAM cell. The developed mitigation solutions were based on resizing of the sensitive transistors [17] or adding extra transistors to reduce the proximity of the radiation strike and distribute the charge collection. The insertion of extra transistors to the original 6T-RAM cell has been also provided in the 8T-SRAM [18] where two transistors are added to eliminate the charge sharing effect between the bit lines or in the differential-ended 10T-SRAM [19] for increasing the speed of the bit line signal or even as 12T-SRAM to reduce the noise disturbance of the bit cell interleaving structure. The performance of the memory block is strictly dependent on the Central Processing Unit (CPU) or Direct Memory Access (DMA) modules.

In particular, the DMA module allows to speed up the data transfer versus and from the memory. In System-on-a-Chip (SoC), DMA module is implemented through the allocation of a controller physically close to the memory layout in order to provide a high-bandwidth infrastructure. In order to provide error tolerant data transfer mechanisms, radiation tolerant heterogeneous multi DMA core systems have been developed to support Single Event Upsets (SEUs) tolerant high-rate connections between the multi-core module and the memory elements [20] .

In order to evaluate the performance of the developed OpenRAM memory block with respect to the typical DMA data transfer, we evaluated the Scatter-Gather (SG) algorithm, which consist on the data transfer initialization on blocks of 32-bit words through Buffer Descriptors (BD). A set of DMA configurations have been evaluated considering different direction of the data transfer, parametric data length and different test of read and write addresses (random and full burst transfer mode). We also evaluated single data request from the DMA with data transfer request of a low number of data packets in order to evaluate the performance limits of the OpenRAM block versus the DMA data transfer routines.

In this work, we investigate specifically the sensitivity and mitigation of recoverable SEEs phenomena affecting a SRAM memory bank. Typically, this type of SEE effect happens in ground and avionics applications or, for aerospace applications at Low Earth Orbit (LEO), where the eventuality of ultra-high energy radiation particle is nullified by the Earth magnetic field. Thanks to the layout-oriented radiation analysis we implement a RHBD mitigation solution at the layout level increasing the robustness of the original 6T-SRAM cell by adding two parametrizable transistors that can be tunable with respect to the radiation particle energy required

3 Radiation Analysis of VLSI technology

In order to achieve an accurate radiation sensitivity of the SRAM module, we developed the radiation particle simulation environment illustrated in Figure 1. The simulation flow is based on the 3D radiation particle propagation tool presented in [22] and available as open-source code. We extracted the physical layout description of the OpenRAM basic cell library, and we inserted the layer material, thickness, and depth to the Graphic Data Systems (GDS) description.

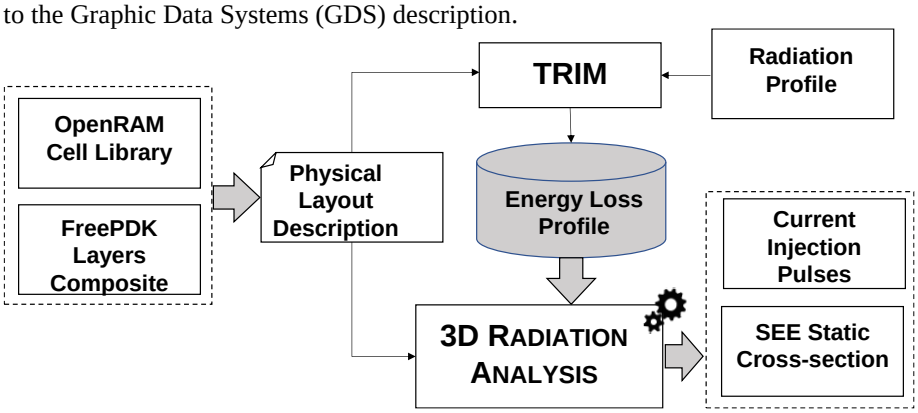

Fig. 1. The developed 3D simulation for the radiation particle analysis on the OpenRAM cell library layout geometry.

The physical description and the particle radiation profile generated by the Transport of Ions in Matter (TRIM) tool are used to calculate the energy distribution released by the ions traversing the layer section of the cell. The last step of the analysis consists on the analysis of the physical layout description and the energy loss of the radiation particle used to calculate the transient sensitivity of the logic cells in terms of SEE cross-section and the correspondent current injection pulses generated within the cell geometry.

3.1 The FreePDK45nm technology node

The OpenRAM memory module is generated by the open-source compiler using the open-source variation-aware physical design kit FreePDK [21] based on Scalable CMOS design rules. The cell library includes variation-aware tools compatible with commercial design tools based on a theoretical 45nm technology where each cell is described by a proper structure including rectangular vias, metallizations and interconnects as well as silicon active regions. We integrated the FreePDK library information building a 3D model of the cell library adding thickness and layer material adopting the modeling provided by the 45nm high performance bulk logic platform technology lithography [23] . The generated model consists of 13 layers from the Active region, the Well and implant sections up to three aluminum metallizations connected by copper vias. The data of the generated model are represented in Table 1 while a structural view of a 6T-SRAM cell of the OpenRAM module is illustrated in

Fig. 2.

The size of each layer and the position of each volumetric region of the FreePDK library has been modeled considering the thickness and the implant position in order to achieve a compliant three-dimensional model.

Table 1. 45nm FreePDK layers, thickness, and composite materials

Layer Name	Layer [#]	Thickness [nm]	Layer Material
Active	1	520	SiO_2
N-Well	2	100	n-Si
P-Well	3	110	p-Si
N-implant	4	100	n-Si
P-implant	5	110	p-Si
S-Block	6	85	SiN
Poly-Silicon	9	85	Poly-Si
Contact	10	150	Si_3N_4
M1	11	130	Al
Via1	12	120	Cu
M2	13	140	Al
Via2	14	120	Cu
M3	15	140	Al

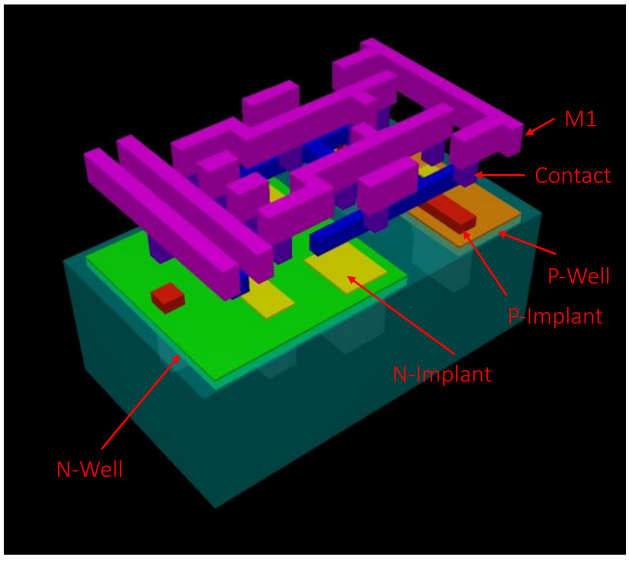

Fig. 2. The 45nm 6T-SRAM cell 3D model view from the Active layer up to the metallization M1.

3.2 SEE Radiation Analysis

The radiation analysis is performed considering four heavy ions energy profile related to the UCL facility [24] . Table 2 reports the energies, ion range and Linear Energy Transfer (LET) used. The analysis starts by extracting the geometry and size data of layer volumes, material composition as well as the radiation profile for the considered particles. The TRIM application calculates the energy loss level of the particle for each layer of the cell. Fig. 3 represents the amount of released energy in each layer of the cell considering the *Aluminium* and *Xenon* heavy ion particles at the energies defined in Table 2. Interestingly, *vias* are the volumes with the highest value of energy loss, while the metalizations, contacts and in implants and well have a low energy loss.

Table 2. Radiation Particle Characteristics

Ion	DUT Energy [MeV]	Range [µmSi]	LET [MeV/mg/cm^2]
$^{13}C^{4+}$	131	269.3	1.3
$^{27}Al^{l8+}$	250	131.2	5.7
$^{58}Ni^{18+}$	582	100.5	20.4
$^{124}Xe^{35+}$	995	73	62.5

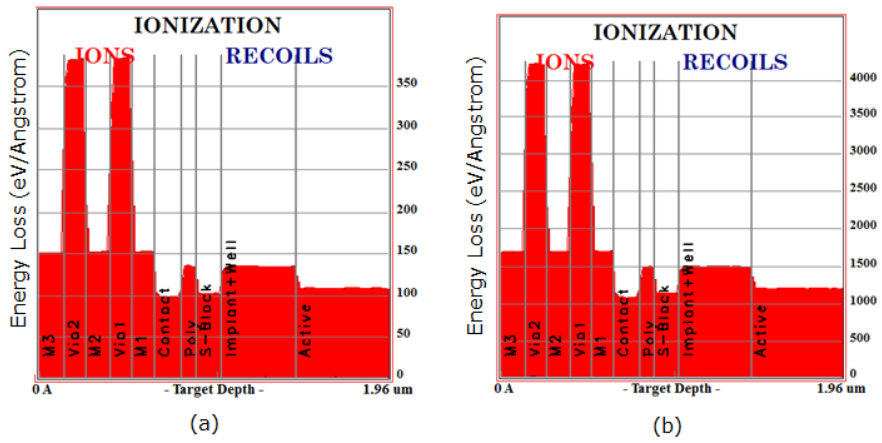

Fig. 3. Release Energy profile for the different layers of the 45nm cell for the Aluminum (a) and the Xenon (b) energy levels.

The developed tool elaborates the physical description of the cell, generating the 3D mesh structure of the layout of the logic cell. Based on the size, shape, and material of metallization and volumes of the cell with respect to the radiation profile of the mission represented in Table 2, the developed radiation analysis tool simulates the effects of highly charged particles traversing the silicon junction of the device and calculate

the generated eV transmitted to the Silicon matter by the particles and provide the current profile for each particle strike.

3.3 OpenRAM radiation sensitivity and SoC interface

The System-on-Chip (SoC) under evaluation consists on an OpenRAM module and a DMA core. They are connected through two buses: a control bus with read, write and enable signals, and a data-bus of 32 bit wise. The overall scheme is illustrated in Fig. 4, where it is also possible to distinguish the eight blocks of the OpenRAM architecture. The hierarchical blocks of the memory bank are based on six main logic cells: Data Flip-Flop (DFF), Master and Slave Flip-Flip (MS-Flop), Write Drivers, Three states buffer, Sense Amplifier and the 6 Transistors RAM cell. The hierarchical decoder and the control logic gates are outside of the memory bank; however they are mainly based on DFFs and combinational logic cells which radiation sensitivity can be determined with traditional analysis method [25] . We analyzed the memory bank cells with 10,000 heavy ions particle using the 3D simulation approach and we computed the SEE cross-section for each cell. The results are illustrated in Fig. 5.

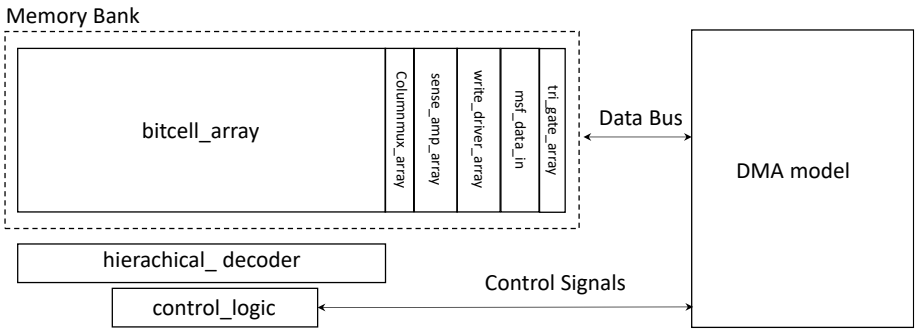

Fig. 4. The overall OpenRAM hierarchical blocks. The hierarchical decoder and control logic modules are outside from the memory bank. The DMA model directly configure the control logic and performs the data transfer through the data bus.

The SEE cross-section may vary from $5.44 \cdot 10^{-14}$ up to $5.46 \cdot 10^{-13}$. The DFFs and the MS-Flops are the most sensitive cells while the 6T-RAM cell is interestingly the cell with the lower cross-section curve. However, considering the number of cells per block, the memory bank cross-section is fully determined by the sensitivity of the 6T-SRAM cell. For example, considering a memory bank of 8Kb, the cross-section is equivalent to $1.43 \cdot 10^{-9}$ totally due by the 6T-SRAM radiation sensitivity.

We performed a Monte Carlo analysis in order to depict the vulnerability regions of the 6T-SRAM cells and individuating the parasitic thyristor resistance spectrum distribution on the cell layout considering a static and unpowered condition of the cell. In Fig. 6, it is possible to observe that the SRAM has various sensitive area, most of them correlated to the layout position of the 6 transistors. Besides, we calculated the distribution of the current pulses observing that 96.62% of the radiation particle injected over the 40,000 injections performed by the selected ions are generating current

pulse below 0.5 µA with a maximal peak of 17.4 µA.

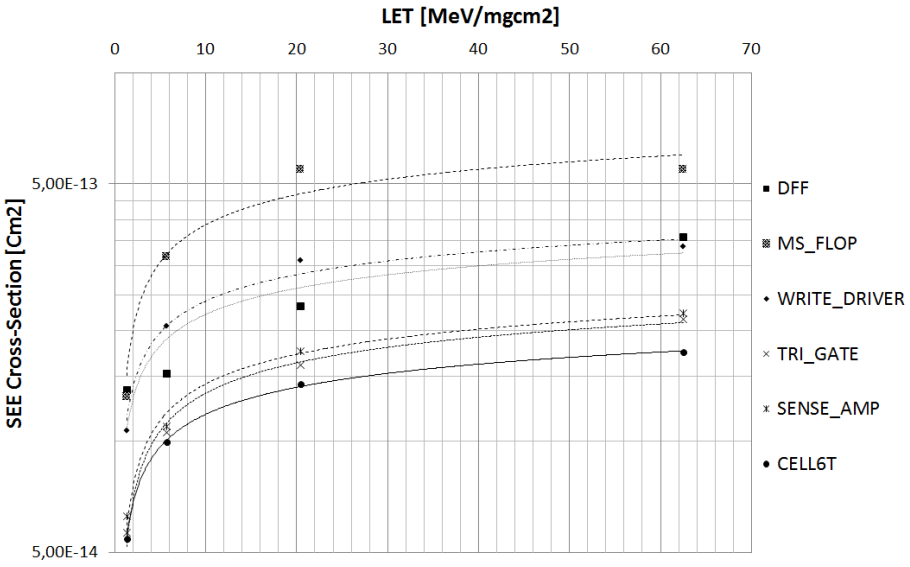

Fig. 5. The Single Event Effects (SEEs) cross-section sensitivity for the OpenRAM memory bank individual cell components.

4 8T-SRAM Mitigation Strategies

Traditional 8T-SRAM schemes are based on the insertion of 2 additional transistors to the 6T scheme depending on the target application of the memory. In case of mitigation solutions for soft-errors, extra transistors are generally added to introduce redundancy to the bit-lines or to the NMOS and PMOS used to implement the SRAM storage. The main purpose of our approach is to insert two redundant transistors in parallel to the original PMOS transistors, in order to distribute the radiation particle charge injected by those particles directly crossing MP5 and MP6 and to increase the overall Q_{crit} margins for the transient effects introduced by particles crossing other regions of the cell.

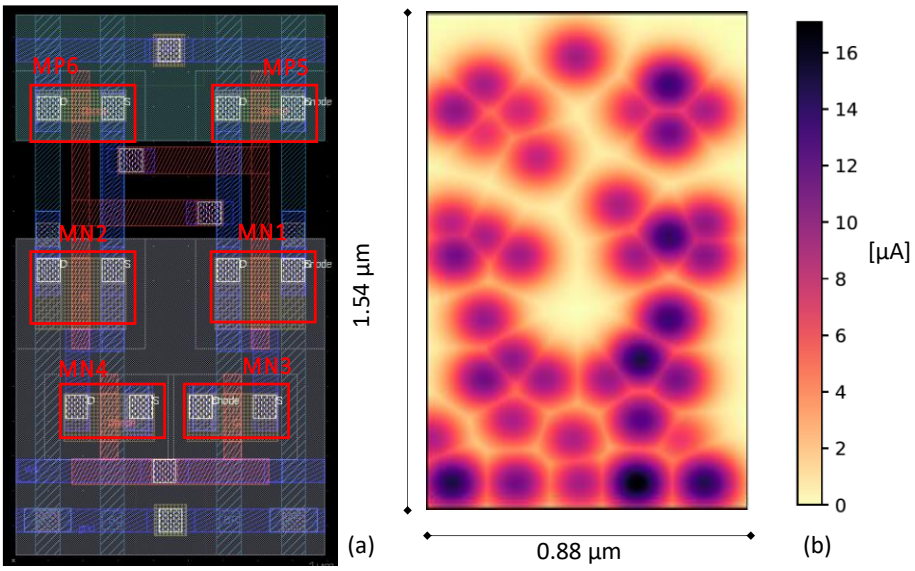

Fig. 6. The 6T-SRAM layout (a) and the vulnerability region reporting the µA current spectrum for the Xenon energy analysis on the overall area (b).

We adopted a different approach to insert the redundant transistors to the original scheme. Thanks to the availability of the OpenRAM layout, instead to start the mitigation insertion from the electrical scheme, we considered at first the 6T-SRAM original layout available regions that can be modified without introducing area overhead to the cell. We identify on the top of the MP6 and MP5 transistors enough physical space to introduce to redundant transistors without affecting the SRAM cell size.

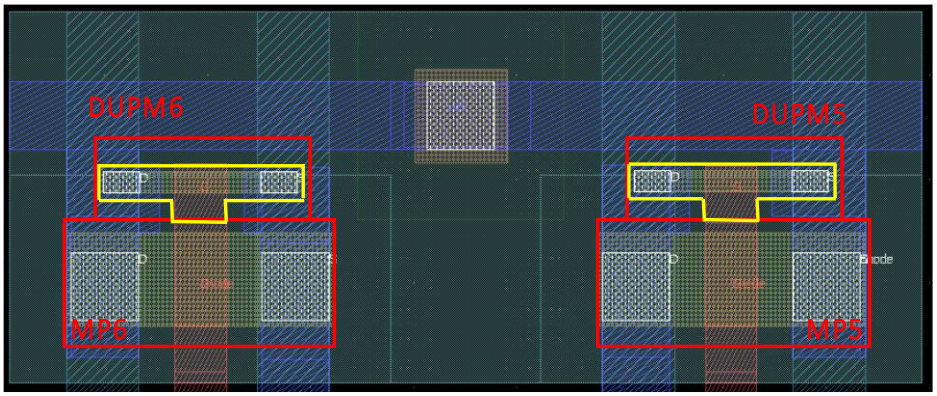

Fig. 7. The view of the top area of the SRAM cell including the duplicated transistor T structure, highlighted in yellow.

We introduced the following layers for each transistor: an active layer with *h*=0.02 µm and *w*=0.19 µm; a P-implant layer with *h*=0.02 µm and *w*=0.18 µm centered with respect to the active layer; the poly-silicon section vertical to the active layer has been extended of 0.05 µm in order to be effective with respect to the active layer; two contact regions of *h*=0.02 µm and *w*=0.035 µm, finally, we extended the metallization M1 of 0.06 µm in order to connect properly the VDD source and the drain and source junctions.

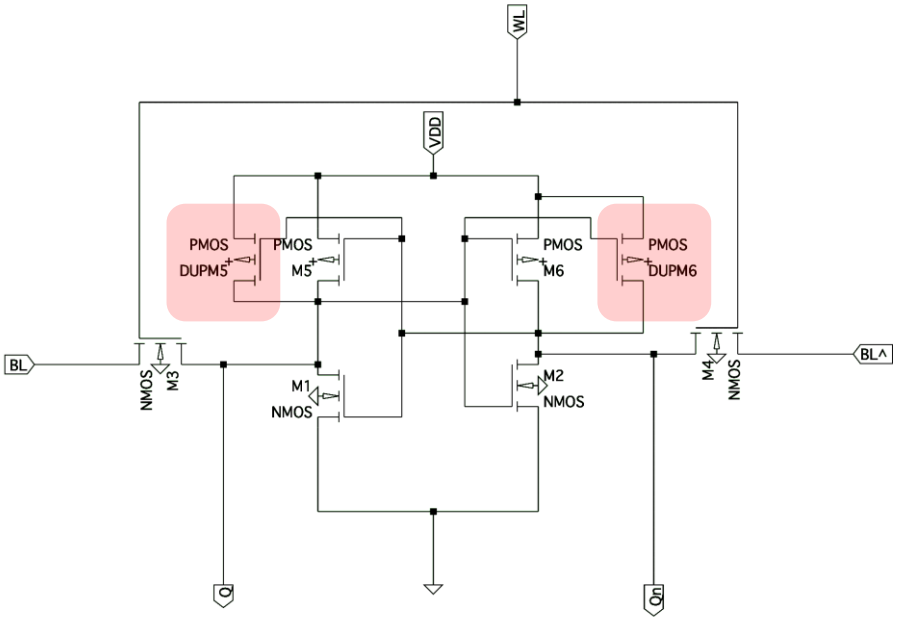

Fig. 8. The electrical scheme obtained from the layout technology extraction of the mitigated 8T-SRAM cell with the highlights on the added transistors replica.

The result of this modification is a *T-structure* added on the top of the original transistor as illustrated in Fig. 7. The layout insertions have been validated by a commercial layout editor tool configured with the FreePDK45 library design rules check. Finally, we performed the technology extraction and conversion to an H-spice model, reported in Fig. 8, using the same layout editor tool.

5 Experimental Results

We designed two 8Kb memory modules with 1 memory bank, 256 words and 32-bits configuring the OpenRAM compiler with the original 6T-SRAM and with the developed 8T-SRAM adopting the T-structure redundant transistors. We performed two experiments evaluating the static behavior of the memory bank. The former consists on a

fault injection campaign for evaluating the mitigation capability of the developed 8T-SRAM cell, the latter consists on the comparative analysis of area, leakage current and SRAM performance characteristics.

5.1 SEE Radiation Analysis

The fault injection simulation setup consists on modeling the SEE at the circuit level by inserting transient current sources at the impact nodes. The fault injection has been executed in two different campaigns. The former campaign measures the maximal current pulse threshold tolerated before to create the upset for each individual transistor. The latter campaign measures the dynamic sensitivity of the cell by injecting the current pulse extracted from the current pulse profile generated by the radiation analysis tool applied to the cell.

Table 3. SRAM cell SEE threshold current

SRAM Configuration	SEE Threshold Current Pulse [μA]	
	$Q=1\ Q_n=0$	$Q=0\ Q_n=1$
6T-Original	0.93	0.46
8T-Proposed	3.83	2.12

In general, the most sensitive part of the SRAM inverter configuration is the drain of the n-mos transistor which is in the off stage; however, we performed the fault injection in all the transistors, and we extracted the SRAM cell threshold current for creating a bit-flip within the cell. The obtained results are reported in Table 3, as it possible to notice the proposed mitigated cell increases the maximal current threshold of around 4 times at the static condition Q=1 and Q_n=0 and more than 4.6 times for the condition Q=0 and Q_n=1.

A plot of the injection of the maximal current pulse for the configuration Q=1 and Q_n=0 is illustrated in Fig. 9. The injections are performed at the static storage condition of the SRAM cell and for a duration of 2ns which is 18% longer than the maximal current pulse width measured by the 3D radiation simulation. In details, the current pulse effect has a duration of around 800ps when the Q and Q_n are simultaneously at low voltage. As consequence, the two values are upset for 2.05ns. Once the current pulse expires, the original SRAM values are restored in less than 420ps.

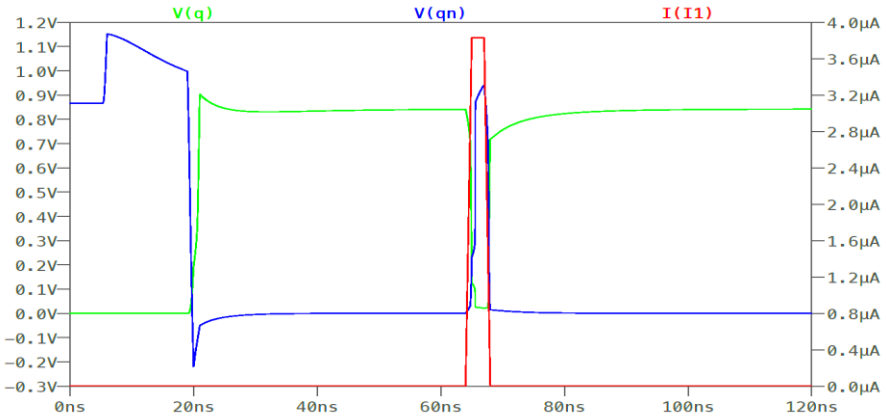

Fig. 9. An example of maximal current pulse injection on the proposed 8T-SRAM cell.

We evaluated the overall robustness of the mitigated OpenRAM memory bank considering the radiation particle spectra described within the radiation analysis section and we performed 40,000 particle injection comparing the achieved SEE cross-section. The results, illustrated in Fig. 10, show that the developed 8TSRAM is robust more than 6 times with respect to the original cell at higher energy. Please note, that the proposed 8TSRAM is drastically more robust for low energy particles, since it is resilient more than one order of magnitude at energies below 10MeV/mgcm^2.

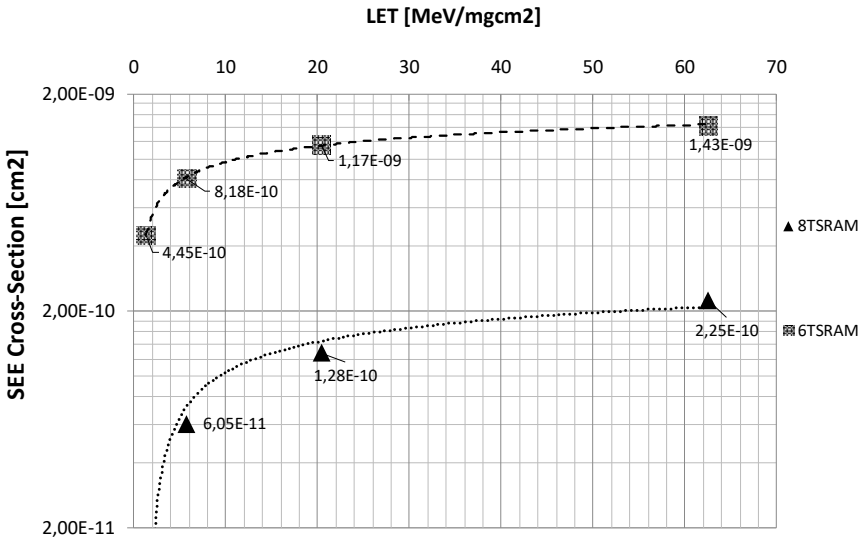

Fig. 10. The OpenRAM 8Kb memory bank SEE cross-section comparison.

5.2 DMA performance analysis

The main goal of the benchmark analysis is to mimic the hardware computing operations of Direct Memory Access (DMA) data transfer considering a single core. As benchmark DMA we selected a data transfer module capable to perform the Scatter-Gather algorithm on memory blocks of 32-bit. We hypothesized to configure the DMA with a single channel mode managing data transfer individually per data transfer direction. We settled two data transfer directions: *Memory Mapped-to-Stream* (MM2S) mainly based on reading operation from the block RAM and *Stream-to-Memory Mapped* (S2MM) which is performing continuous writing on the block RAM. The directions are addressed by different buffer descriptor allocated in a dedicated memory within the DMA module. Please note that the interrupt signals typically coming from the DMAs and connected to the CPU have been properly monitored by the simulation model in order to measure the performance of the memory block. A software routine running on the DMA is settled in order to initialize and stimulate the memory and collect the reports on the simulation environment. The achieved data are reported in Table 4, where we reported the minimal clock period that allows to perform an error immune data transfer.

Table 4. OpenRAM memory block DMA performances

32-bit words [#]	Original 6TSRAM [ns]		Proposed 8TSRAM [ns]	
	MM2S	S2MM	MM2S	S2MM
1	2.18	2.19	2.94	3.02
8	17.41	17.49	23.55	23.55
64	139.31	141.45	148.41	159.41
256	556.05	562.08	645.66	682.32

5.3 Comparative analysis

The power dissipation, area and delay are marginally affected by the insertion of the two redundant transistors. The size of the SRAM cell is not changed by the insertion of the T-structure transistors since we included them in the original 6T-SRAM layout. However, in case of further optimization, our insertion will limit the reduction of the SRAM cell area for less than 3%.

Table 5. SRAM cells characteristics Comparison

Characteristic	Original 6TSRAM	Proposed 8TSRAM
V_{DD}(V)	1.5	1.5
Leakage Current (μA)	0.81	0.88
SNM	41.53	43.20
RSNM	27.40	38.42
WSNM	129.62	141.75

Considering the power consumption and the functional characteristics, we compared the original and mitigated cells in order to compare the Static Noise Margin (SNM), the Read Static Noise Margin (RSNM) and the Write Static Noise Margin (WSNM). As expected, the results reported in Table 5 indicates that the proposed 8T-SRAM cell is slightly degrading the leakage current while maintaining almost equivalent the SNM.

We compared the reading and writing delays characteristics of a single SRAM cell. The delay is degraded due to the additional parasitic resistive capacitive load effects that increases the average response time from 68ps up to 92ps. Finally, we also compared the standby power for the traditional 6TSRAM cell with the developed 8TSRAM observing a negligible increasing of power consumption from 6.30 µW to 6.83µW.

6 OpenRAM module within Neural-Network structure

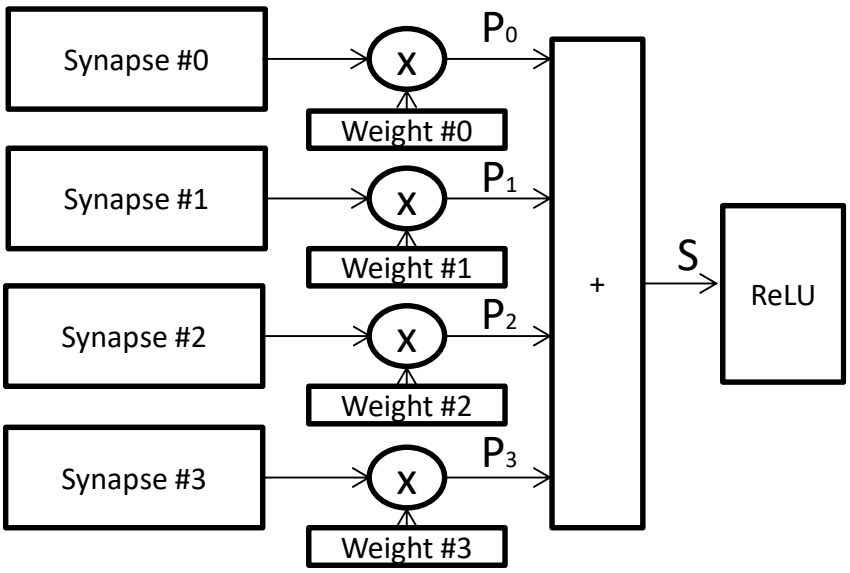

Fig. 11. The neuron structure: the basic element of CNN architecture.

The typical NN structure illustrated in Figure 11, where floating point data are used for weights, inputs and outputs. In order to implement a complete CNN, a number of parallel neurons can be instantiated in parallel. All the data flow traversing the structure from the synapse inputs up to the post rectified linear output are represented by 16 bits. Considering that the data are generally representing values in a range from 0 to 1 and that weights are signed values with a limited range of precision; the product will mandatory requires higher resolution for the multiplication and extra range for the accumulation to avoid overflow conditions of any arithmetic process. For the sake

of this work, we implemented a full neural network model consisting in a behavioral description of fully parallel neurons in order to evaluate the impact of OpenRAM mitigated module on a large scale circuit.

A feasible solution to insert the developed OpenRAM module in its original unmitigated version and with the developed 8TSRAM cell is to adopt a Digital Signal Processing (DSP) architecture for each neuron. The scheme of the developed implementation is illustrated in Figure 12.

The structure of the hardware synthesizable neuron consists of an input stream of 256 16 bits data words simultaneously read by all the neuron in the same layer. The layer of parallel neurons is reducing the limitations on the input bandwidth thanks to the essential data caching. The data inputs are multiplied with the weights; typically, each weight value is used several times by the neurons depending on the position related to the data set. The OpenRAM module is inserted in order to store the neuron weight. The architecture used is the classical convolutional network [26] with reduced size, suitable to evaluate mapping and implementation tools.

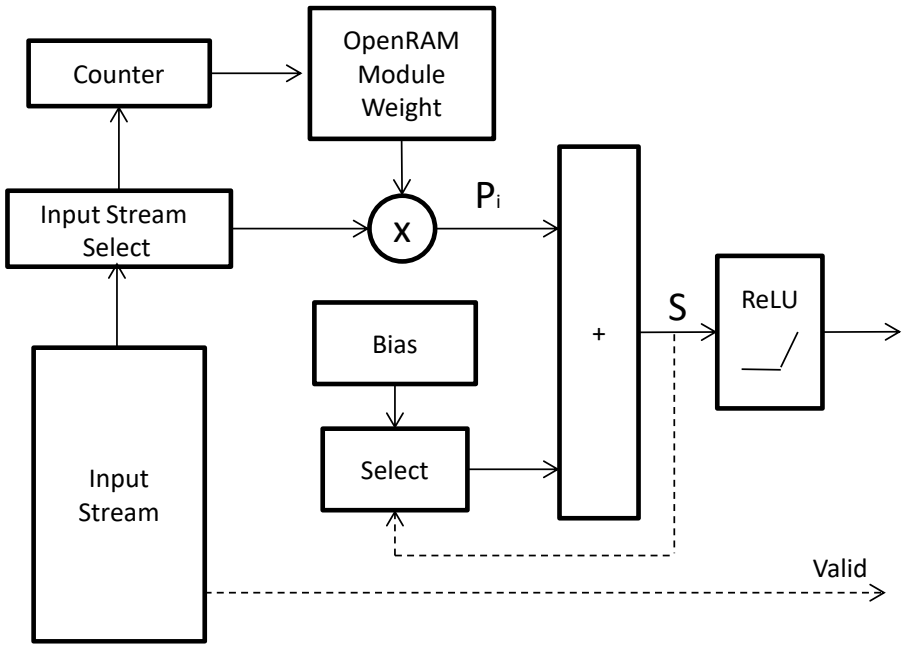

Fig. 12. The hardware neural network structure implementation.

We tested the network with an input data stream consists of a 224 by 224 image crop with 3 colors map convolved with 96 filters at the first layer, each one with a size of 7 by 7and adopting a stride of 2 on both x and y. The feature map is then passed through a rectifier linear function, max pooled with a 3x3 matrix with stride 2 and finally normalized across feature maps and generating 55 x 55 elements feature map. The intermediate layers 2 to 5 repeat the same operation, while the final two layers are fully connected and are elaborating the features from the top convolutional

layer in a vector of 9,216 dimensions. Finally, the last layer is a soft-max function with *i*-way, with *i* being the number of classifications.

The developed design flow has been applied on the NN implemented and simulated using the ModelSIM simulator instrumented with the OpenRAM structural netlist. The NN has been implemented in five different implementation:

- *Original*: implemented with timing performance optimization considering a customized memory block obtained by synthesis
- *+Open*: implemented with the original 8Kb OpenRAM module used for the storage of the neural node weights.
- *+OpenOptiRead*: implemented with the original 8Kb OpenRAM module used for the storage of the neural node weights, optimized for the reading operations
- *+OpenMit*: implemented with the developed mitigated 8Kb OpenRAM module used for the storage of the neural node weights.
- *FullOpen:* implemented with the developed mitigated 8Kb OpenRAM module used for the storage of the neural node weights and for the storage of the input and data output of the NN.

The obtained results are presented in Table 6 where we show the computational time to elaborate a classification of 10 images. The performance results show the overall computational time considering a simulation frequency of 100MHz, however, we estimated that that maximal working frequency of 220MHz can be reached without incurring in errors within the OpenRAM blocks.

The result shows that the insertion of the OpenRAM memory block provides an improvement of performances versus the original implementation of the Neural Network with a customized memory block. We believe that this optimization is due to the faster decoding circuitry of the OpenRAM block versus the customized ones. On the other hand, it is possible to notice a comparable trend on the NN with the mitigated OpenRAM, as expected we observed a marginal degradation of the computational time. However, we believe that this degradation is mainly due to the smaller size of the developed OpenRAM block with respect to the requested size which may reach up to 33Kb for the largest Neural Network layer.

Table 6. Neural Network Performances considering the OpenRAM integration

Characteristic	Overall Computational Time [ms]
Original	12.4
+Open	11.8
+OpenOptiRead	11.4
+OpenMit	12.1
FullOpen	14.9

7 Conclusions and Future works

In this paper, we propose a methodology for the analysis and the mitigation of SRAM circuit generated by the open-source OpenRAM memory compiler. The main novelty of the proposed methodology is the capability to details the interaction of the radiation particle with the SRAM memory layout, to depict the sensitive transistors and to selectively mitigate the radiation effects by layout-oriented modifications. We applied the workflow methodology to the design and mitigation of the 6TSRAM cell. Thanks to the availability of the layout description provided by the OpenRAM project, we developed a new mitigation strategy to increase the current thresholds and reduce the voltage transients. An experimental analysis performed on an 8Kb memory module generated by the OpenRAM compiler demonstrates that the developed 8TSRAM-based memory module is 6 times more resilient than the original memory block at high energy particle. The resiliency is improved up to one order of magnitude at lower energies, as specifically target for ground and low earth orbit applications. We compared the functional characteristics with the original cell, and we observed a minimal deviation in leakage current and an evident improvement versus reading and writing noise margins. The performance analysis of the developed cell has been also experimentally evaluated considering a DMA System-on-a-Chip and a Neural Network simulation. Thanks to these two analyses it has been possible to evaluate the proposed solution considering an high performance architecture and large scale memory block usage. As future works, we plan to extend the mitigation features to other memory components and to evaluate the robustness versus destructive Single Event Latch-up (SEL) effects and to evaluate the application of error detection and correction schemes.

References

[1] B. D. Sierawski et al., "Muon-Induced Single Event Upsets in Deep-Submicron Technology," in *IEEE Transactions on Nuclear Science*, vol. 57, no. 6, pp. 3273-3278, Dec. 2010.

[2] T. Karnik and P. Hazucha, “Characterization of soft errors caused by single event upsets in CMOS processes” *IEEE Trans. Depend. Secure Comput.*, vol. 1, no. 2, pp. 128–143, Apr.–Jun. 2004.

[3] S. Azimi, B. Du, L. Sterpone, "On the prediction of radiation-induced SETs in flash-based FPGAs", *Microelectronics Reliability*, pp 230-234, 2016.

[4] P. E. Dodd and L. W. Massengill, "Basic mechanisms and modeling of single-event upset in digital microelectronics" in *IEEE Transactions on Nuclear Science*, vol. 50, no. 3, pp. 583-602, June 2003.

[5] R. Lacoe et al., "Application of Hardness-By-Design Methodology to Radiation-Tolerant ASIC Technologies", *IEEE Trans. Nuc. Sci.*, vol. 47, no. 6, pp. 2334-2341, Dec. 2000.

[6] X. Yao, L. T. Clark, D. W. Patterson and K. E. Holbert, "Single event transient mitigation in cache memory using transient error checking circuits*" IEEE Custom Integrated Circuits Conference* 2010, San Jose, CA, USA, 2010, pp. 1-4.

[7] M. R. Guthaus, J. E. Stine, S. Ataei, Brian Chen, Bin Wu and M. Sarwar, "OpenRAM: An open-source memory compiler," 2016 IEEE/ACM International Conference on Computer-Aided Design (ICCAD), Austin, TX, USA, 2016, pp. 1-6.

[8] S. Azimi, C. D. Sio and L. Sterpone, "On the Evaluation of SEEs on Open-Source Embedded Static RAMs," 2021 IFIP/IEEE 29th International Conference on Very Large Scale Integration (VLSI-SoC), 2021, pp. 1-6

[9] T. Ansell and M. Saligane, "The Missing Pieces of Open Design Enablement: A Recent History of Google Efforts : lnvited Paper," 2020 IEEE/ACM International Conference On Computer Aided Design (ICCAD), San Diego, CA, USA, 2020, pp. 1-8.

[10] S. Azimi, L. Sterpone, "Digital Design Techniques for Dependable High Performance Computing," IEEE International Test Conference (ITC), 2020, pp. 1-10.

[11] S. Yoshimoto et al., "A 40-nm 0.5-V 20.1-µW/MHz 8T SRAM with low-energy disturb mitigation scheme," 2011 Symposium on VLSI Circuits - Digest of Technical Papers, Kyoto, Japan, 2011, pp. 72-73.

[12] R.C. Baumann and E. B. Smith, "Neutron-induced 10B fission as a major source of soft errors in high density SRAM", Microelectronics Reliability, Vol. 41, No. 2, pp. 211-218, February 2001.

[13] Y.-W. Chiu *et al*, "40nm bit-interleaving 12T subthreshold SRAM with data-aware write-assist", IEEE Transactions on Circuits and Systems I, Vol. 61, pp. 2578-2688, 2014.

[14] X. Yao, L. T. Clark, D. W. Patterson and K. E. Holbert, "Single event transient mitigation in cache memory using transient error checking circuits," IEEE Custom Integrated Circuits Conference 2010, pp. 1-4.

[15] N. Seifert, M. Kirsch, "Rea-Time Soft-Error Testing Results of 45nm High-K Metal Gate, Bulk CMOS SRAMs", IEEE Transactions on Nuclear Science, Vol. 59, No. 6, December 2012.

[16] J. Li, W. Chen, R. Li, G. Wang and S. Yang, "Study on Transient Ionizing Radiation Effect of 40nm SRAM", 3rd International Conference on Radiation Effects of Electronic Devices (ICREED), pp. 1-4, 2019.

[17] K. Mohr and L. Clark, "Delay and Area Efficient First-level Cache Soft Error Detection and Correction", IEEE ICCD Proceedings, pp. 88-92, October, 2006.

[18] T. Hoang, *et al*. "A Radiation Hardened 16-Mb SRAM for Space Applications", IEEE Aerospace Conference, pp. 1-6, 2006.

[19] B. S. Mohammad, H. Saleh, M. Ismail, "Design Methodologies for Yield Enhancement and Power Efficiency in SRAM-based SoCs", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Vol. 23, Issue 10, pp. 2054-2064, 2015.

[20] T. Solokhina et al., "Radiation tolerant heterogeneous Multicore "system on chip" with built-in multichannel SpaceFibre switch for onboard data management and mass storage device: Components, short paper," 2016 International SpaceWire Conference (SpaceWire), 2016, pp. 1-6

[21] L. Chang *et al*, “An 8T-SRAM for variability tolerance and low-voltage operation in high-performance caches”, IEEE Journal of Solid-State Circuits, Vol. 43, pp. 518-529, 2008.
[22] I. J. Chang *et al*. “A 32 Kb 10T sub-threshold SRAM array with bit-interleaving and differential read scheme in 90nm CMOS”, IEEE Journal of Solid-State Circuits, Vol. 44, pp. 650-658, 2009.
[23] J. E. Stine et al., "FreePDK: An Open-Source Variation-Aware Design Kit", IEEE International Conference on Microelectronic Systems Education, pp. 173-174, 2007.
[24] L. Sterpone, F. Luoni, S. Azimi, B. Du, “A 3-D Simulation-based Approach to Analyze Heavy Ions-Induced SET on Ditigal Circuits”, IEEE Transactions on Nuclear Science, Vol. 67, Issue 9, pp. 2034 – 2041, 2020.
[25] S. Azimi, C. De Sio, L. Sterponet., "A Radiation-Hardened CMOS Full-Adder Based on Layout Selective Transistor Duplication", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 29, no. 8, pp. 1596-1600, 2021.
[26] Matthew D. Zeiler and Rob Fergus, "Visualizing and Understanding Convolutional Networks," Lecture Notes in Computer Science, vol. 8689, pp. 818–833, 2014, Springer