A CMOS OxRAM-Based Neuron Circuit Hardened with Enclosed Layout Transistors for Aerospace Applications
Pablo Ilha Vaz, Patrick Girard, Arnaud Virazel, Hassen Aziza

To cite this version:

HAL Id: lirmm-03035780
https://hal-lirmm.ccsd.cnrs.fr/lirmm-03035780
Submitted on 2 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A CMOS OxRAM-Based Neuron Circuit Hardened with Enclosed Layout Transistors for Aerospace Applications

Pablo Ilha Vaz*, Patrick Girard*, Arnaud Virazel† and Hassen Aziza†

*Laboratory of Informatics, Robotics and Microelectronics of Montpellier (LIRMM)
University of Montpellier / CNRS, Montpellier, France
Emails: pablo.ilha-vaz@lirmm.fr, girard@lirmm.fr, virazel@lirmm.fr
†IM2NP laboratory, Aix-Marseille University / CNRS, Marseille, France
Email: hassen.aziza@univ-amu.fr

Abstract—Brain-inspired computing architectures, brought by Artificial Neural Networks (ANNs), are an attractive solution to reduce the energy consumption of the conventional von Neumann’s computation, with an excellent parallel processing capability. Therefore, critical applications, such as Space & Satellite, which impose severe constraints in terms of power consumption and computing efficiency, are excellent candidates to embed such networks. Nonetheless, integrated circuits operating during long-term and cumulative exposure to incidence levels of ionizing radiation may have their physical components degraded, thus drastically reducing their reliability and expected lifetime. A possible solution to enhance the radiation hardening characteristics of a conventional bulk CMOS device is to use the non-standard gate geometry referred to as Enclosed Layout Transistor (ELT). In this work, we propose to harden the design of an existing OxRAM-based neuron circuit [1] through the inclusion of ELTs, i.e., to improve the radiation hardening characteristics of a preexistent convenient neuron circuit topology by using the enclosed gate geometry for the n,pMOS devices. Electrical simulations, considering a standard commercial bulk CMOS fabrication process, in a 180 nm technology, have been carried out to validate our proposed design. The simulation results, supported by the analysis of former works regarding the incidence of ionizing radiation in OxRAM and ELTs, indicate that the proposed hardened neuron circuit is a feasible solution to embed neuromorphic computing in aerospace applications.

Index Terms—Artificial Neural Networks, bulk CMOS, Enclosed Layout Transistor, OxRAM, Neuron circuit, Radiation Hardening, Fault-tolerance.

I. INTRODUCTION

The emergence of Artificial Neural Networks (ANNs) has brought a paradigm shift in computing architectures with brain-inspired systems. In contrast with the conventional von Neumann’s computation, ANNs have the ability to learn and adapt through complex nonlinear relationships [2]. Moreover, this biological-inspired approach has an excellent parallel computing capability with a significantly lower power consumption when compared to Central Processing Units (CPUs) and Graphic Processing Units (GPUs) [3]. In this context, critical applications, such as Space & Satellite (S&S), which impose severe constraints in terms of power consumption and computing efficiency, are excellent candidates to embed ANNs. Nonetheless, applications exposed to incidence levels of ionizing radiation, as in the case of S&S, may have their physical components degraded. Depending on the energy of the incident particle, and time exposure to such ions, this ionizing radiation gives rise to transient upsets or even permanent damage to the device’s materials [4]–[6]. Thus, this vulnerability becomes a significant issue to design a circuit capable of properly operate between acceptable error margins, i.e., failure tolerance, for a baseline application.

In this context, to embed a state-of-the-art neuromorphic circuit operating in harsh environments, it is necessary to face the challenge of enhancing its radiation hardening characteristics, thus satisfying its reliability constraints.

There are several techniques capable of hardening a device to mitigate the effects associated with the incidence of ionizing radiation. The methods range from the higher abstraction level (i.e., the system-level) to the lower abstraction level (i.e., the layout-level). In fact, by combining complementary approaches, applicable at various abstraction levels, it is possible to achieve the highest level of fault-tolerance [7], [8]. Indeed, this composition can complement one to each other, allowing to exploit the ultimate benefits brought by the emerging brain-inspired computing paradigm, enhanced through well-established fault-tolerance techniques, necessary to properly operate in harsh environments.

Thus, in this work, taking into account the reliability challenges associated with the possibility to consolidate a state-of-the-art ANN in aerospace applications, we present the design of the basic building block of brain-inspired networks, i.e., a neuron cell, hardened through the inclusion of ELTs. To this purpose, we use the neuron circuit topology introduced in [1] that includes Oxide-based Resistive Random Access Memory (OxRAM) device as the memory element, which is inherently tolerant against ionizing radiation [9], and, hence, better suited for a fully hardened circuit.

Previous related works [5], [10]–[12], which are supported based on measurement results, have demonstrated the intrinsic hardening characteristics of ELTs when compared to standard gate geometry. These studies confirm that the proposed neuron
circuit may have even higher radiation tolerance than the original topology based only on OxRAM devices and standard gate geometry.

To the best of our knowledge, this is the first work proposing to harden the design of an existing neuron circuit topology integrating these two hardened solutions (i.e., ELT and OxRAM). Therefore, such design has relevant value to explore further artificial neural units, e.g., synapses, and to solve complex tasks in aerospace applications using neuromorphic computing.

The remaining of this work is organized as follows. Section II introduces the basic concepts related to Ionizing Radiation, its effects on circuits and systems, and the ELT paradigm. Section III introduces the fundamentals of brain-inspired computing and OxRAM technology. Section IV presents the neuron circuit topology adopted in this work. Section V describes the simulation setup, and simulation results. Finally, Section VI provides the conclusion of the work.

II. IONIZING RADIATION AND ENCLOSED LAYOUT TRANSISTOR THEORY BACKGROUND

Circuit reliability constitutes an imperative role for applications operating in environments exposed to ionizing radiation, for instance, those in aerospace, aeronautics, and, even at the ground level, as high-energy physics. Therefore, to maintain an integrated circuit proper operation inside acceptable error margins throughout an extended lifetime, it is mandatory to enhance its radiation hardening characteristics [13].

When radiation particles travel through solid materials, they may transfer kinetic energy to atomic electrons [4]. Depending on the energy of the incident particle, this interaction produces a path composed of electron-hole pairs ($\bar{e}/h$) [14], [15]. However, these $\bar{e}/h$ produced during the ion incidence may migrate when exposed to an external electric field, for instance, as generated when biasing a MOS device. The change resulted from this charge migration, i.e., the flow rate of electric charge, results in unwanted behavior in MOS devices, giving rise to transient upsets or permanent damages both for analog and digital circuit operations [14], [15].

Regarding spurious currents, which may lead to functional failures, a proper architecture-level circuit design, based on redundancies and voting schemes, can be adopted to mitigate unwanted characteristics, e.g., such as bit-flips [6].

On the other hand, to achieve high levels of tolerance to permanent damage caused by ionizing radiation, a robust design should be embedded at the layout-level to prevent the susceptible regions to store parasitic charges [5], [16]. One well-established solution to minimize the accumulation of long-term radiation-induced parasitic charges is to change the standard commercial gate disposal (i.e., straight polysilicon stripe) to the annular geometry referred to as Enclosed Layout Transistors (ELTs) [5], [7], [11]–[13], [16]–[18]. Instead of standard two-edged devices, depicted in Fig. 1 (I) and (III), the use of ELTs, depicted in Fig. 1 (II), eliminates the transition region where the polysilicon layer extends over the well-to-substrate boundary, hence, significantly reducing the radiation-induced leakage current which may exist between source and drain (S/D) contacts [5], [16].

Fig. 1. a) Minimum-sized standard two-edged device, b) Minimum-sized ELT, and c) Standard two-edged device equivalent to the minimum-sized ELT.

In the field of radiation tolerance, in our recent work [17], based on experimental results, we have performed an extensive analysis of ELTs and deeply investigated the effects of the incidence of cumulative radiation, thus corroborating the findings presented in former works [5], [7], [10], [11], [19].

From this study [17], we have extracted Fig. 2, which provides a sample of the transfer function behavior, respectively for minimum-sized (a) standard (STD), and (b) ELT when exposed to different irradiation doses. An oxide thickness of 12.5 nm was used in both cases.

Fig. 2. ($I_D$ vs. $V_{GS}$) nMOS characteristics, respectively for: STD devices (a) and ELTs (b).

Based on referred experimental irradiation measurements (i.e., from Fig. 2), we observed increments of approximately 3 orders of magnitude in the leakage current ($I_{leak}$) of nSTD devices when irradiated up to 500 krad(Si). On the other hand, in the same irradiation experiments, for ELTs, we verified deviations of less than 1% in the $I_{leak}$. These previous experimental irradiation results clearly demonstrate that the proposal of hardening an OxRAM-based neuron circuit through the addition of ELTs will improve its radiation hardening characteristics.

Nonetheless, the determination of the aspect ratio ($W/L$) in ELTs is not a straightforward task. In ELTs, conversely to standard-devices, the electric field under the gate corners is not uniform and, thus, a methodology is needed to estimate the equivalent aspect ratio of such devices. In our previous
work [17], we have deeply investigated a wide range of ELT’s \((W/L)_{eff}\) predictions. In this work, we adopted the equations first presented in [16].

III. OxRAM-based Neuromorphic Systems Theory Background

The basic building block of ANNs are neurons (or nodes), and the transmission of their information occurs through electrical signals, in the form of spikes (conveyed by synapses).

In an ideal ANN, each neuron should be able to receive an input signal, i.e., stimuli from a previous synapse, and change its internal structure accordingly to better process it.

Through this brain-inspired paradigm, it is viable to accomplish matrix-vector-based tasks, such as image pattern recognition, with remarkable energy efficiency and matching performance when compared to central processing units (CPUs) and graphic processing units (GPUs) [3], [20].

Together with ANNs, recent advances in non-volatile memories (NVMs) have emphasized the opportunity to perform in-memory computing, a pivotal character to emulate an \textit{in situ} learning mechanism. In this direction, during the last few years, the scientific community has proposed the first neuron circuits and synapses integrating NVMs [21].

In this context, one of the promising device to act as the memory element is the Oxide-based Resistive Random Access memory (OxRAM) cell. The OxRAM cells are excellent candidates to be designed within ANNs, especially regarding their compatibility with CMOS back-end-of-line (high integration density), exceptional switching speed (fast programming), and low energy operation [22], [23].

An OxRAM device is a metal-insulator-metal (MIM) based element. Its metal oxide can be composed of a range of different materials, such as \(\text{HfO}_2\), \(\text{NiO}\), \(\text{TiO}_2\), and \(\text{TaO}_2\), all of them compatible with the CMOS processes. In this work, as depicted in Fig. 3, we considered a \(\text{HfO}_2\)-based cell.

From the designer’s perspective, i.e., behavioral view, the stack of \(\text{TiN/Ti/HfO}_2/\text{TiN}\), depicted in Fig. 3, can be seen as a resistive switching device or, in other words, as a two-terminal device adjustable resistor.

The OxRAM operation relies on the formation/dissolution of an internal Conductive Filament (CF), as depicted in Fig. 3. While the CF does not exist (or is weakly formed), the device can be considered in the OFF state (with very High Resistance State - HRS). Then, it is possible to gradually switch their conductance by increasing the magnitude of the voltage potential between its top \((V_T)\) and bottom terminals \((V_B)\). With the occurrence of this voltage potential increment, the OxRAM passes throughout a plethora of intermediate states, until the total formation of the CF, where it reaches the ON state (with very Low Resistance State - LRS).

The evolution of the CF occurs based on the generation and migration of oxygen ions \((O^{2-})\) and vacancies \((V_o)\) inside the OxRAM cell. Thus, to accurately simulate such devices, it is necessary to have a proper physical model prediction. In this sense, we invite the reader to consult the related work [24], which presents an extensive comparison among a series of distinct OxRAM models.

In this work, we adopted the model introduced by Stanford [25]. This referred model captures both DC and AC electrical characteristics of OxRAM devices with physics-based compact model descriptions.

IV. Bulk CMOS Neuron Architecture

One of the first conceptual and purest implementations of a neuron model is known as the Leaky Integrate and Fire (LIF) model [26]. The global mixed-signal concept of a LIF neuron suggests two main procedures: the integration of input signals, and an additional trigger mechanism, which is activated when the integration process reaches a certain threshold value.

The first procedure, i.e., the integration of input synaptic stimuli, plays the role of the brain’s ability to store synapsis history. The electronic equivalent to emulate such brain-inspired memory behavior can be designed, for instance, through OxRAM devices – regarding their advantages when compared to CMOS based-components, as introduced in Sec. III.

The second procedure, i.e., trigger activation, receives the signal of integration threshold and, when reaches the limit value, generates a digital output spike to be sent to post-neurons and to reset the current one.

Taking into account the previous conceptual implementation of a LIF neuron circuit, the block diagram presented in Fig. 4 introduces the complete neuron topology analyzed in this work.

The original circuit architecture has been introduced in [1]. In this work, we propose reducing the long term degradation caused by the incidence of ionizing radiation of the neuron circuit by changing the gate geometry of the standard two-edged devices (which are identified by \(M_{ref}\), \(M_1\) to \(M_7\), and the switches \(S_1\) to \(S_4\)) using the ELT paradigm.

The neuron circuit can be understood as the composition of three distinct blocks: the \textit{Current comparator}, the trigger (the combination of the \textit{RS Latch} and \textit{Pulse generator}), and the \textit{Feedback path}.
The Current comparator block is responsible for integrating the input signal coming from the VPreSpike pin (i.e., \( V_T \) in the current block). The core of this current integration block is the OxRAM element, in which each excitatory input signal slightly changes its internal conductance. The maximum current which flows through the OxRAM \( I_{OxRAM} \) is regulated by the constant current input source pin \( (I_{thr}) \) – due to \( M_{1,2} \) current mirror. Hence, the current difference between \( I_{thr} \) and \( I_{OxRAM} \) is “calculated” and amplified due to the current mirror \( M_{3,4} \). Finally, when the \( I_{OxRAM} \) exceeds \( I_{thr} \), the analog output \( V_{out} \) reaches the digital equivalent high level, i.e., the voltage necessary to activate the next block.

The trigger mechanism, represented in Fig. 4 by the combination of the RS Latch and Pulse generator sub-blocks, receives (and monitors) the analog output of the preceding block \( (V_{out}) \). When \( V_{out} \) reaches the threshold voltage necessary to active (or reset) the RS Latch, the intermediate signal \( Q \) falls to the digital equivalent low level. Thus, the Pulse generator produces the VPostSpike signal – a single output pulse with both voltage and period predetermined.

The last block, the Feedback path, is formed by a set of switches \((S_1 \text{ to } S_4)\). The role of this block is to allow a proper input/output signal decoupling between VPreSpike (neuron input) and VPostSpike (neuron output and internal reset signal \( V_B \)).

### V. EXPERIMENTAL RESULTS

#### A. Simulation Setup

To evaluate the behavior of the proposed neuron circuit hardened with ELTs, we propose to exploit a sub-set of simulations based on three different n,pMOS device variations:

- \( STD_{min} \): Minimum-sized standard two-edged devices.
- \( ELT_{min} \): Minimum-sized ELT rules.
- \( STD_{equiv} \): Standard two-edged devices with the \((W/L)_{eff}\) equivalent to the \( ELT_{min} \).

For simulating the OxRAM switching behavior, we adopted the SPICE-compatible Stanford model V2 [25]. Besides, the electrical level simulations are performed at the SPICE level, using a commercial CAD tool in a 180\( nm \) bulk CMOS technology.

The neuron cell circuit behavior is analyzed in the topology presented in Fig. 4. As in the previous OxRAM configuration setup, the simulation Test Bench (TB) is proposed to compare the figures of merit of ELTs fare against those of standard two-edged devices. Thus, the same n,pMOS arrangements are considered for \( STD_{min} \), \( ELT_{min} \), and \( STD_{equiv} \).

In each configuration, the input/output signals \( V_{PreSpike} \), \( V_{PostSpike} \), \( V_{out} \), and, \( V_Q \) are monitored during simulations. Initially, the input value for the threshold current takes into account the detection precision of the difference between \( I_{OxRAM} \) and \( I_{thr} \). In this case, related studies about OxRAM programming current have reported an accuracy of less than 0.1% for 100\( \mu \)A embedded in the same built-in current comparator and similar oxide thickness [1]. Thus, such nominal value is chosen.

Regarding devices sizing, Table I presents both width and length of n,pMOS devices. The value 1 designates the minimum device size; higher values correspond to multiples of their minimum dimensions.

![Fig. 4. Schematic view of the analyzed neuron circuit.](image)

<table>
<thead>
<tr>
<th>Variable</th>
<th>( I_{thr} )</th>
<th>( t_{period} )</th>
<th>( t_{fall, rise} )</th>
<th>( V_{DD} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit</td>
<td>( 100\mu )</td>
<td>( 100n )</td>
<td>( 100p )</td>
<td>1.8</td>
</tr>
<tr>
<td>Value</td>
<td>A</td>
<td>s</td>
<td>s</td>
<td>V</td>
</tr>
</tbody>
</table>

### B. Simulation Results

Figure 5 shows the simulation results for the OxRAM-based neuron circuit. In the referred figure, (a) illustrates the behavior of the main input/output signals monitored during one cycle of \( ELT_{min} \) topology functioning (i.e., \( V_{PreSpike}, V_{out}, V_Q \), and, \( V_{PostSpike} \)). Additionally, for the three neuron simulated configurations, Fig. 5 (b) presents the current which flows both through the OxRAM and the \( M_3 \) device (specified in Fig. 4).

For each analyzed configuration, the quantitative values for the calculation of the average current through OxRAM \((I_{cell,avg})\) and the energy per cycle (or spike) \((E/Cycle)\) are presented in Table III.

Regarding the neuron circuit, this work adopted the same simulation methodology and calculations procedure, as reported in the reference work [1]. This conformity allows a direct behavior and performance comparison against the original one. In this context, the reference work, evaluated in 130\( nm \)
Moreover, the increment in the $E/Cycle$ of the $ELT_{min}$ when compared against $STD_{min}$ represents $\approx 1.4\times$. This increase is due to the area increment, which in practice, establishes the increment in the effective $(W/L)$ aspect ratio, i.e., the effective $(W/L)$ of a minimum-sized ELT is about $9.8$ times higher than the $(W/L)$ of the $STD_{min}$.

It is noteworthy that the inclusion of ELTs does not affect the functional behavior of the neuron circuit. In this case, the functional correspondence can be confirmed not only by $I_{cell(Avg)}$ and $E/Cycle$, presented in Table III, but also substantiated with Fig. 5 (b) in which the neuron circuit fires in the same cycle (i.e., when comparing $ELT_{min}$ vs. $STD_{equ}$). The asymmetry between the source/drain areas of the $ELT_{min}$ when compared to $STD_{equ}$ configuration diminishes the $E/Cycle$ about $\approx 0.2\%$ and the $I_{cell(Avg)}$ about $\approx 0.13\%$. This almost negligible difference is in absolute accordance with what was expected since the $STD_{equ}$ has sized according to the same effective $(W/L)$ of the $ELT_{min}$, and, thus, the only remaining difference is related to the source and drain different areas.

Regarding the physical neuron circuit area estimation, we calculated the relation among the three MOS different configurations. The quantitative values for the area analysis are given in Table IV. In this table, the $Device relation$ ($STD_{min}$ vs. #) represents the area difference calculated between MOS type and the reference value $STD_{min}$. Moreover, the $High$ vs. $Length$ represents the relation between the cell’s high and length, and, finally, $Area$ refers to the total neuron cell area for each MOS type. It is noteworthy that all area estimations only take into account the layout of the MOS devices, as depicted in Fig. 1.

From Table IV, it is possible to calculate the inherent area penalty necessary to implement ELTs, as shown in Fig. 1. When comparing the area overhead between a minimum-sized single two-edged device to a minimum-sized ELT, the area increases by $4.51$ times. Nonetheless, when estimating the total neuron circuit area, i.e., taking into account the OxRAM device alongside ELTs, we presume that the silicon overhead should be negligible. It occurs because the OxRAM devices are placed on top of the CMOS subsystem during the back-end phase, contrary to conventional capacitors, which demand an extensive area. These values for the area estimation are in accordance with our previous work [18], in the same technology node, based on practical results.
fore, shorter switching times are required to charge/discharge such capacitances [17], [18].

Finally, the electrical simulation results presented in this work, together with the support of previous practical irradiation measurements, presented in Fig. 2, clearly demonstrate that the proposal of hardening an OxRAM-based neuron circuit through the addition of ELT paradigm improves its radiation hardening characteristics.

VI. Conclusion

In this work, to face the reliability challenges related to the incidence of ionizing radiation in ICs and to consolidate a state-of-the-art ANN in aerospace applications, we presented the design of an OxRAM-based neuron circuit, hardened through the inclusion of ELTs.

To the best of our knowledge, this is the first work that addressed the design of a preexistent convenient OxRAM-based neuron circuit integrating ELT paradigm.

We presented the neuron circuit simulation results in a subset of three different n,pMOS arrangements: the standard two-edged device, the minimum-sized ELT, and the standard two-edged equivalent to the minimum sized ELT. The electrical level simulation results presented in this work are in accordance with practical measurements reported in former works. Therefore, this coherence supports a proper OxRAM model usage and a correct neuron circuit functional behavior.

After all, when introducing ELTs, the simulation proposals and analysis presented in this work have proven that, at the cost of the area, it is possible to harden a preexistent neuron circuit topology. Moreover, when taking into account the OxRAM devices, laid-out into back-end-of-line (BEOL), we expect a negligible area overhead, thus, reiterating the relevance of this work in the field of reliability of neural networks embedded in aerospace applications.

VII. Acknowledgment

The authors would like to thank LRI/IEAv/DCTA for the test facilities used in our previous work [17].

References