Sources of Single Event Effects in the NVIDIA Xavier SoC Family under Proton Irradiation

Ivan Rodriguez-Ferrandez*†, Maris Tali‡, Leonidas Kosmidis*†, Marta Rovituso§, David Steenari‡

*Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
†Barcelona Supercomputing Center (BSC), Barcelona, Spain
‡European Space Agency (ESA), Noordwijk, The Netherlands
§Holland PTC, Delft, The Netherlands

***Abstract*—In this paper we characterise two embedded GPU devices from the NVIDIA Xavier family System-on-Chip (SoC) using a proton beam. We compare the NVIDIA Xavier NX and Industrial devices, that respectively target commercial and automotive applications. We evaluate the Single-Event Effect (SEE) rate of both modules and their sub-components, both the CPU and GPU, using different power modes, and we try for the first time to identify their exact sources using the on-line testing facilities included in their ARM based system. Our conclusion is that the most sensitive part of the CPU complex of the SoC is the tag array of the various cache structures, while no errors were observed in the GPU, probably because of its fast execution compared to the CPU part of the application during the radiation campaign.**

I. Introduction

The space industry is facing a dramatic increase in the payload processing performance required by future missions - upcoming spacecraft require acquiring orders of magnitude more data compared to existing ones [1]. Much higher resolutions, precision and sampling frequencies are becoming more common. The sheer amount of obtained data cannot be transmitted to ground, therefore the only viable solution is to process these data on-board or compress them before their transmission.

Moreover, Artificial Intelligence (AI) is creating a revolution in algorithms for space exploration and earth observation missions [1] [2] which require an order of magnitude more performance capabilities compared to previous missions.

Other types of upcoming space missions require increased autonomous operation and decision-making capabilities due to the lack of real-time control, communication delays or narrow communication windows. These include robotic exploration such as the Rosalind Franklin ExoMars [3] and Asagumo [4] rovers, as well as new types of space missions and concepts like the space tug [5] and active debris removal [6] or the Mars Helicopter Ingenuity [7].

These performance requirements cannot be met by existing processor technologies used in space, such as the radiation hardened LEON [8] or PowerPC [9] families of processors. Interestingly, COTS devices (i.e. devices not designed specifically for space) are already considered as enabling solutions [10].

Among several hardware options, Embedded Graphics Processing Units (GPUs) have shown a great potential in high performance processing in temperature and battery-constrained devices, following the widespread and successful use of GPUs in the high-performance and the mobile phone domain. For example, the Mars Helicopter [7] technology demonstrator is using such an embedded GPU SoC, based on the SnapDragon 801, originally targeting the mobile phone sector, for all spacecraft's operations, not only the payload ones. Moreover, the recent ESA-funded activity GPU4S (GPUs for Space) [11] has identified embedded GPUs as ideal high performance processing device for space, provided that their radiation performance is well studied and necessary mitigation is used if needed. In this work, we take a step closer towards this direction.

A similar trend is seen in other safety critical domains, which have similar high performance demands due to their need of autonomy. In particular, the automotive sector is one of the main drivers of that need, with the introduction of autonomous cars [12]. Industrial automation sees a similar growth, with the goal to revolutionise factory production for Industry 4.0 [13]. As a result, embedded GPU manufacturers have introduced embedded GPU products targeting these two regulated domains. These new products come with enhanced reliability in their hardware design, offering features which are beneficial for the increased availability and safety of the system, as required by their functional safety standards. Additionally, they include functionality to increase their radiation resilience, such as error correction (ECC) in the device caches [14].

In this paper we test two such devices, the NVIDIA Xavier NX and Industrial, both from the NVIDIA Xavier family under proton irradiation. We take advantage of the reliability features provided by these products, especially in the case of the NVIDIA Xavier Industrial, which targets the automotive and industrial sectors. This allows us to gain more insights into the causes of Single Event Effects (SEE) as opposed to only observing the functional correctness or operational status of the device during the test, as the majority of similar studies in the literature do. To our knowledge this is a unique feature

978-1-6654-7355-2/22/$31.00 ©2022

 DOI 10.1109/IOLTS56730.2022.9897236

of our work compared to prior work in this area and a major contribution of our paper.

The rest of the paper is organised as follows: In Section II we provide Background information on radiation effects as well as describe briefly related works in the area. In Section III we describe the reliability features of the devices under test. Section IV explains the experimental setup of our irradiation campaign and Section V presents our results. Finally, Section VI provides the conclusions of our work.

II. Background on Radiation Effects

Radiation effects in semiconductors are a well known effect [15] and are taken into account for reliability analysis of a specific device [16]. When an energetic particle impacts a semiconductor, it displaces charges in the different layers of the transistor causing it to behave in a non-predicted way. This phenomenon is called a *single event effect* (SEE) [17].

Depending how the SEE affects the semiconductor, it is categorised in multiple types [18]. When the effect only flips a bit in memory without affecting the functional integrity of the device, it is called *single event upset* (SEU). In the case of a bit-flip that happens in the control logic of the device and results in the lost of its functionality, it is called a *single event functional interrupt* (SEFI). Neither SEUs nor SEFIs cause any physical damage to the device, but in the case that the radiation event forces a transistor to be continuously open and this happens in a MOSFET transistor related to the power regulation of the device, this event can result in the destruction of the board. Such an event is called a *single event latch-up* (SEL) and is a destructive event. Most of the SEEs – including SELs if they are detected before they damage the device – are transient errors, which are cleared from the device with a power cycle.

The reliability of electronic devices for use in space or in other safety critical systems, including automotive and industrial computing is assessed through *radiation testing*. Radiation testing can be performed using protons, neutrons, heavy ions, two-photon absorption [19] (known also as laser testing) or gamma radiation, with the purpose of accelerating the appearance of SEEs, which are collected for further analysis. Other types of testing like *Total Ionizing Dose* (TID) is also used in order to accelerate the aging of the device through accumulation of radiation. The most common type of testing for terrestrial applications (i.e. automotive or industrial systems) is using neutrons, while in the case of space proton and heavy ions are the preferable choice. In our work we chose proton irradiation since we focus on space systems.

Since SEEs are caused by an energetic particle hitting the device, in order to characterise a radiation environment we define the number of particles that cross the area every second, the *flux* which is measured in [particles/ions]/cm2/sec. During an exposure or a radiation test, the integral of all of the particles or ions that cross the area, is called *fluence*. Knowing the number of particles which cross the area of a device during a radiation test, if we record the number of events (SEE) which are observed during the exposure, we can compute the sensitivity of the device for particle/ion radiation. The metric of sensitivity, defined by the equation in Figure 1, is called *cross-section* of the device and it is used to compare the resilience of multiples devices to different radiation types [18].

$$\sigma = \frac{\text{number of events}}{\text{fluence}} \tag{1}$$

Fig. 1: Cross-section equation

Apart from the overall cross-section, which takes into account all types of SEEs, other interesting metrics are also used. For example if we only count the SEFI events, we compute the sensitivity of the device to lose functionality. This is an interesting metric for robustness/availability of the device in a specific orbit, indicating the number of functional interrupts per day/year which can be expected to experience, using some models which simulate the number of particles which will hit the device during the space mission.

A. Related work

Since embedded GPUs are considered promising target devices for space, multiple devices from different vendors such us NVIDIA [20] [21] and AMD [22] have been exposed to proton radiation or total ionizing dose (TID) testing. As heterogeneous systems-on-chip, featuring multiple CPUs, an integrated GPU and several peripherals, as well as a complex software stack using a general purpose operating system (Linux), which is required in order to use the GPU, embedded GPU platforms are significantly more complex devices than the traditional devices which are used in space and have been tested under radiation in the past.

However, prior works in the literature use rudimentary forms to identify errors which have happened in the CPU or GPU complex, simply trying to test them in the same way that simple space processors have been tested so far. For example, Hiemstra et al. [21] use the system reboot or the unresponsiveness of the system as a way to identify the SEE count, without a way to identify the hardware structure or type of error. Similarly, in [23] Badia et al. use the watchdog timers to detect when a SEFI has happened, but again not knowing the precise point in which the error was generated. Moreover, all prior works only compare the program output with a known result to identify SEUs. However, no published work so far has been able to observe SEUs using this method in these platforms, especially on the GPU.

Our work differs from all prior GPU radiation studies, since it is the first one that uses the built in reliability features of these devices, in order to identify the hardware structure in which the radiation effects have happened, and it is the first one that has successfully observed SEUs.

Finally, Wyrwas et al. [22] is the only work that tries to identify the source of the errors, so it is the work more similar to ours. The authors are using the machine check (MC) errors reported by the x86 CPU during the test of a discrete AMD GPU e9173. However, it is important to mention that in their

radiation setup the main CPU complex and OS are outside of the beam area in order to record these errors, unlike our setup in which the entire SoC is affected by radiation.

III. Built-in Reliability Features in NVIDIA Xavier

A. ARM safety features

Since the NVIDIA Xavier family of SoCs are primarily targeting automotive and industrial sectors, their ARM cores have a feature called ARM Reliability, Availability, and Serviceability (RAS) [24]. RAS is a set of hardware and software diagnostic systems which focuses on the reliability and error reporting of ARM-based CPU systems. In addition, it includes protocols for communication of the different safety systems with the high level software, such as the operating system. In the case of our test campaign, this system provided us with the error reporting of all of the Single Event Upsets (SEU) and most of the Single Event Functional Interrupts (SEFI) that happened in these systems. Therefore, RAS allowed us to pinpoint the location and the source of the errors we observed in most of the cases. Figure 2 shows an example of the reporting of a RAS error in the NVIDIA Xavier NX platform.

```
[6707425.443660] ras_ccplex_serr_callback: Scanning CCPLEX Error Records for Uncorrectable Errors
[6707425.443708] ****************************************
[6707425.443714] RAS Error in SCF:L3_0, ERRSELR_EL1=768:
[6707425.443726]         Status = 0x74007204
[6707425.443732]         IERR = L3 Protocol Error: 0x72
[6707425.443738]         SERR = Assertion Failure: 0x4
[6707425.443741]         Uncorrectable (this is fatal)
[6707425.443834]         MISC0 = 0x0
[6707425.443837]         MISC1 = 0x0
[6707425.443858] ****************************************
```

Fig. 2: Example of a RAS error

A lot of information can be extracted from the error report. In this case this error is an uncorrected error, so it will trigger the board to reboot and therefore it will count as a SEFI. The error originates from the L3 section 0, so it happened in the fist CPU cluster of the Xavier (cores 0 and 1). Another source of information is the SERR and IERR fields, which are defined in the RAS manual [24]. SERR is a general error common for all RAS-enabled devices while the IERR is a vendor specific one, adding more information to the SERR error code. In our example the SERR code indicates an assertion error and the IERR specifies that is a L3 protocol error. This error log can be directly copied from the *dmesg* in a Linux system or from the register associated to the RAS status and error reporting.

B. NVIDIA Xavier SoC

The two systems under test have very similar characteristics, but target different markets. The NVIDIA Xavier NX targets small robots, while the Xavier Industrial is a board that targets safety critical applications. Therefore, the Industrial features safety characteristics like ECC and non-correctable error detection (SECDEC) in all memories including DRAM, while the NX only in the SoC ones. The interesting point is that since both platforms are based in the same chip, both have these benefits. This is important since the form factor of the NVIDIA Xavier industrial module is too big for cubesats

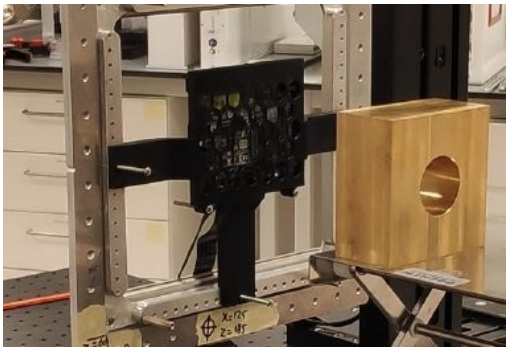

Fig. 3: Xavier NX under the Beam with proton blocker to limit the exposure to the SoC

($105\,\mathrm{mm}$ x $105\,\mathrm{mm}$) but the form factor of the NX could easily fit in cubesats ($70\,\mathrm{mm}$ x $45\,\mathrm{mm}$), making it easier to develop and launch the latter in a demonstrator mission.

Regarding the hardware characteristics, the NVIDIA Xavier industrial features 8 CPU cores and the NVIDIA Xavier NX 6 cores, and 32GB and 8GB of LPDDR4 main memory respectively. The rest of the memory hierarchy comprise a 128 KB L1 instructions cache and 64 KB L1 data cache, a shared 2MB L2 cache per complex (cluster of two cores) and a 4MB shared L3 cache which is shared among all cores [25][26].

IV. Experimental Setup

A. Holland PTC Proton Beam

Our test was performed in Holland PTC, an oncology centre focused on the treatment of cancer using proton radiation located in Delft, the Netherlands [27]. This facility includes a proton beam-line intended for research and development purposes. The beam can deliver between $1\,\mathrm{nA}$ to $800\,\mathrm{nA}$ beam nominal current with a energy range between $70\,\mathrm{MeV}$ to $250\,\mathrm{MeV}$. For our specific test we chose a beam size of $9\,\mathrm{cm}^2$, as it can be seen in Figure 3, in order to focus our testing on inducing errors only in the SoC and not in the associated peripherals, such us power delivery, storage system or main memory. We used the $200\,\mathrm{MeV}$ proton beam, which is most commonly used for proton SEE testing. This also means we did not need to remove the cooling system of the board, as the protons have sufficient penetration to reach the sensitive volumes of the board through the back of the main board, similar to other tests using the NVIDIA Xavier AGX module [21]. We followed this methodology for both tested boards.

B. Device under test: NVIDIA Xavier

As already mentioned both modules share a lot of common characteristics. They are both fabricated using TSMC $12\,\mathrm{nm}$ metal-oxide-semiconductor field-effect transistor (FinFET) with a total die area of $350\,\mathrm{mm}^2$. Both boards have custom ARM $64\,\mathrm{bit}$ CPU cores from NVIDIA codenamed Carmel. In addition to the rest of the characteristics introduced in Section III-B, both boards feature an NVIDIA Volta GPU, with 8 streaming multiprocessors (SM) in the Industrial, and 6 in the NX.

C. Setup procedure

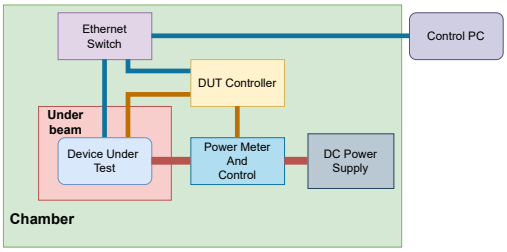

Fig. 4: Schematic of the test setup configuration

In order to be able to control these two complex devices during the testing without entering the radiation testing chamber, a series of systems were designed to be able to monitor and control the behaviour of the systems in detail during the test, as shown in the schematic diagram in Figure 4. The Devices Under Test (DUT) are powered from a standard power supply unit to the *Power meter and control* board. This module records the power consumption of the DUT in real-time and sends the data to the DUT Controller. Then the power goes through a disconnect switch connected to a microcontroller. If the power consumption reaches a limit value, i.e. due to a latch-up, the power to the DUT is automatically interrupted.

Two connections are attached to the DUT, a standard Ethernet connection and a serial UART. The DUT controller sends the commands to execute the different tests through the Ethernet port. Finally the sensors metrics and results of the test are sent to a database in the control PC and displayed with a custom-designed control panel with the critical metrics built with Grafana. Additionally, the control panel allows to force a manual reboot of the device.

D. Power Modes

During the irradiation campaign we executed each test in three different power modes. These modes target relevant power demands for the use of the modules in a thermally constrained satellite. Additionally, some radiation induced effects, such as latch-up, are highly temperature and biasing dependent. All the specification of the power modes can be seen in Table I.

Of these power modes, the 15W and 10W are provided and verified by NVIDIA, but the sub-10W mode is a custom power mode which we designed not to exceed 10W in any of the open source GPU4S benchmark suite [28]. The purpose of this power mode is to test the device under the minimum power consumption that can be achieved on the DUT.

E. Test Procedure

In order to assess the radiation response of the devices in close to real operating conditions, we tested them as follows:

- Idle: The boards were only running the operating system and the monitoring software
- CPU compute: The CPU was computing matrix multiplication from the GPU4S benchmark suite in a loop, using matrix sizes 1K×1K, while the GPU is idle.

Property	Board	15 W	10 W	SUB 10 W
Number of CPU cores	**Industrial**	4	2	2
	NX	2	2	2
CPU Frequency	**Industrial**	1.2 GHz	1.2 GHz	0.3 MHz
	NX	1.9 GHz	1.5 GHz	0.3 MHz
Number of SM	**Industrial**	8	4	4
	NX	6	6	4
GP Frequency	**Industrial**	0.7 GHz	0.5 GHz	0.1 GHz
	NX	1.1 GHz	0.8 GHz	0.1 GHz
Memory Frequency	**Industrial**	1.3 GHz	1.1 GHz	0.2 GHz
	NX	1.6 GHz	1.6 GHz	1.6 GHz

TABLE I: Power modes used for the test

- GPU compute: The CPU is idle and the GPU is running matrix multiplication from the GPU4S benchmark suite in a loop, using 1K×1K matrix sizes.

These were combined with the 3 power modes, giving us a total 9 different configurations per device. The reason for selecting the matrix multiplication as software under test is that it is a computationally intensive kernel frequently used in such tests. An external program checks and verifies the output using a golden reference.

V. Experimental Results

In this Section we discuss the effects of the proton radiation on both devices. An important initial note is that no destructive Single Event Latch-up (SEL) was observed during the radiation campaign. Figure 5 shows a full duration test without a reboot, in which there is no visible increase of the current consumption of the device that could indicate that a latch up is happening in its power circuits [29].

In addition, no errors in the output of the matrix multiplication were observed during the test, similar to all prior works in the literature, which however erroneously concluded that no SEUs had happened. In our case, despite the absence of output errors, SEUs errors were observed during the testing using the RAS reporting system through the debug serial interface and all originate from the main CPU complex. This includes both corrected and uncorrectable errors. However, no errors were reported from the GPU complex.

Table II and Table III present the cross-section of the NVIDIA Xavier Industrial and NVIDIA Xavier NX respectively, including the total SEE count consisting of the total number of Single Event Upsets (SEU) and the total number of Single Event Functional Interrupts (SEFI). In the case that several runs of the same test were performed, the upset and fluence were combined for the cross-section calculation.

The correctable errors reported by RAS were classified as SEUs. Any uncorrectable errors reported by RAS, which automatically caused device resets, were classified as SEFIs.

Figure 6 and Figure 7 present the break down of SEUs for the NVIDIA Xavier industrial and for the NX, respectively. We observe that the SEUs are roughly equally divided between the L2 and L3, for both boards. In the case of the L2 cache, we see that the SEUs are also split 50% percent between the ”L2 MLC

Fig. 5: NVIDIA Xavier Industrial power, voltage and thermals during the test in our custom Grafana-based control panel.

Power Mode	Test Type	# of Events	Fluence	Cross-section
10W	Idle	24	1.2×10^{9}	2.05×10^{-8}
	CPU	8	4.9×10^{7}	6.85×10^{-8}
	GPU	7	9.2×10^{7}	2.67×10^{-8}
15W	Idle	6	9.4×10^{7}	6.38×10^{-8}
	CPU	1	8.4×10^{7}	1.19×10^{-8}
	GPU	2	7.6×10^{7}	2.63×10^{-8}
SUB 10W	Idle	1	7.5×10^{7}	1.33×10^{-8}
	CPU	8	8.4×10^{7}	9.51×10^{-8}
	GPU	1	7.7×10^{7}	1.29×10^{-8}

TABLE II: Radiation results of the different test configurations per power modes for the NVIDIA Xavier Industrial.

Power Mode	Test Type	# of Events	Fluence	Cross-section
10W	Idle	2	8.4×10^{7}	2.38×10^{-8}
	CPU	1	8.0×10^{7}	1.25×10^{-8}
	GPU	3	5.5×10^{7}	3.2×10^{-8}
15W	Idle	7	6.7×10^{7}	1.04×10^{-7}
	CPU	2	6.9×10^{7}	2.89×10^{-8}
	GPU	1	8.6×10^{7}	1.16×10^{-8}
SUB 10W	Idle	1	7.5×10^{7}	1.32×10^{-8}
	CPU	1	7.5×10^{7}	1.33×10^{-8}
	GPU	1	6.38×10^{7}	1.56×10^{-8}

TABLE III: Radiation results of the different test configurations per power modes for the NVIDIA Xavier NX.

Correctable Error" and the "SCF to L2 Correctable ECC", were MLC stands for mid level cache. The former indicates an error in the data array of the L2 cache, while the latter seems to come from the interface that maintains the cache coherence between the L2 caches.

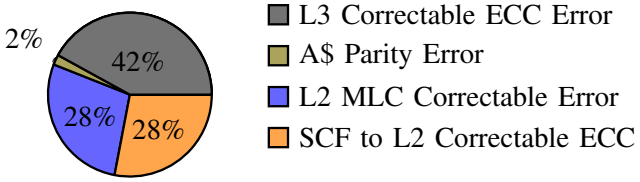

Fig. 6: SEU Distribution in Xavier Industrial

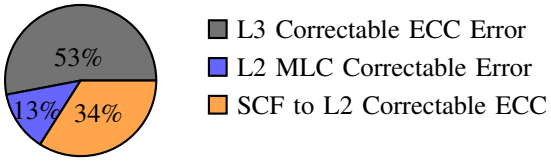

Fig. 7: SEU Distribution in Xavier NX

Table IV summarises the SEFIs reported by RAS. Most of them are related to the L3 and the cache tag protection. All the reported SEFIs are triggered by an uncorrectable error in RAS which forces a kernel panic, which in turn reboots the device. As seen in the Table, all the RAS detected errors are related to the Tag part of the cache, which makes sense because as specified in the official Xavier manual [30], the cache tags of all caches are only protected with a parity bit, so errors cannot be corrected. Regarding the effectiveness of RAS, only one SEFI error was not detected by the system, resulting in a high rate of SEFI detection.

Finally, as a summary, from all of the observed SEEs only 11% were SEFIs and 89% were SEUs.

SEFI Causes

Board	RAS Detected	Cause
IND1	YES	L3 Tag Parity Error
IND1	No	Unable to handle kernel paging request at virtual address 00010014
IND1	Yes	L3 Tag Parity Error
IND1	Yes	L3 Tag Parity Error
IND1	Yes	DVMU Interface Timeout
IND1	Yes	L2 MLT Tag Parity Error
IND1	Yes	L2 MLT Tag Parity Error
NX1	Yes	Multi-Hit Tag Error
NX1	Yes	Multi-Hit Tag Error

TABLE IV: RAS SEFI error reports

A. *Effect on Power Modes on the SEE cross-section*

During this study we also tried to identify whether the power mode and clock speed has an impact in the overall cross-section of the different elements of the board.

Figure 8 shows the cross-section per bit for the L3 cache of both boards. Both the NVIDIA Xavier Industrial and the

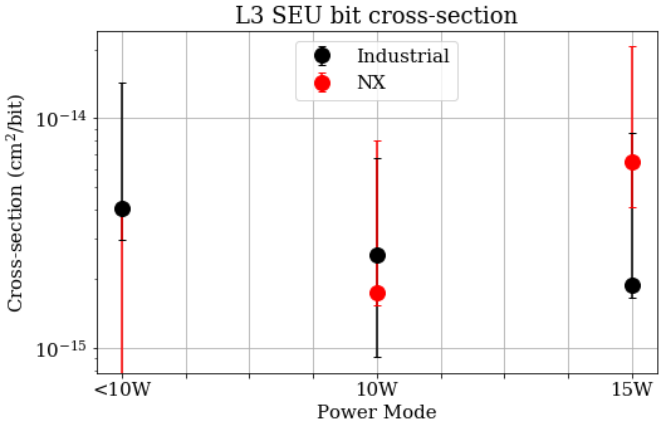

Fig. 8: L3 cache SEU cross-section dependence on the power mode for both the Xavier NX and Industrial.

		IND		NX	
	Events	**GEO**	**LEO**	**GEO**	**LEO**
SEU	**Day**	1.31×10^{-2}	4.04×10^{-2}	1.23×10^{-2}	3.77×10^{-2}
	Year	4.78	1.47×10^{1}	4.49	1.38×10^{1}
	Orbit	1.31×10^{-2}	2.83×10^{-2}	1.23×10^{-2}	2.64×10^{-2}
SEFI	**Day**	5.53×10^{-3}	1.70×10^{-2}	4.73×10^{-3}	1.46×10^{-2}
	Year	2.02	6.21	1.73	5.33
	Orbit	5.53×10^{-3}	1.19×10^{-2}	4.73×10^{-3}	1.02×10^{-2}
L2 SEFI/bit	**Day**	2.99×10^{-9}	9.19×10^{-9}	2.91×10^{-9}	8.94×10^{-9}
	Year	1.09×10^{-6}	3.35×10^{-6}	1.06×10^{-6}	3.26×10^{-6}
	Orbit	2.99×10^{-9}	6.44×10^{-9}	2.91×10^{-9}	6.26×10^{-9}
L3 SEFI/bit	**Day**	5.90×10^{-10}	1.81×10^{-9}	1.13×10^{-9}	3.48×10^{-9}
	Year	2.15×10^{-7}	6.61×10^{-7}	4.12×10^{-7}	1.27×10^{-6}
	Orbit	5.90×10^{-10}	1.27×10^{-9}	1.13×10^{-9}	2.44×10^{-9}

TABLE V: Estimation of the number of events in two different orbits for the two devices

Xavier NX have very similar cross-section. Moreover, the cross-sections measured using the different power modes are within error margins from each other and no significant effect of the modes could be observed. This could be explained by the low number of events observed during the test.

B. Orbit reliability

Based on the data extracted during the test for the different events for each of the most sensitive parts of both devices, we can predict the mean time between events in different possible scenarios in which the boards can be deployed. In order to do that we use the OMERE 5.6 simulator from TRAD [31]. In this way we can simulate the number of events that we can have in multiple orbits, using the cross-section data and the energy of the beam used to calculate the cross-section.

We simulate two typical orbits in which the use of GPUs can be beneficial for earth observation missions. The first one LEO (Low-Earth Orbit), is a 800 Km orbit with an inclination of 98º (polar orbit), and the other one GEO (geosynchronous orbit), in which the orbit speed and the rotation speed of the Earth matches, that is a 35784 Km orbit with 0º inclination.

We simulate both protons and ions, but because we only have data for proton effects, we use only the proton models. Moreover, we use the AE8 model [32] for the trapped particles, and the ESP [33] model for the solar particles. Table V shows the predicted number of radiation events that the devices will experience in the two orbits, broken down to different categories among SEUs, SEFIs, L2 SEFIs/bit and L3 SEFIs/bit in three different time frames, per day, per year and per orbit.

We notice that both devices will only experience a total of 2 SEFI errors per year in a GEO orbit and a total of 6 errors in LEO orbit, which is an acceptable reliability figure, and shows that GPUs can be employed for such missions.

VI. Conclusions

Our radiation data suggest that both NVIDIA Xavier SoCs are good candidates for high performance computation for space applications in thermally constrained satellites deployed in both low earth and geosynchronous orbits, from a reliability point of view. The measured SEU cross-sections were $3.61 \times 10^{-8}\,\mathrm{cm}^2$ for the Industrial and $2.91 \times 10^{-8}\,\mathrm{cm}^2$ for the NX. The SEFI cross-sections were $1.52 \times 10^{-8}\,\mathrm{cm}^2$ for the Industrial and $1.30 \times 10^{-8}\,\mathrm{cm}^2$ for the NX. The largest contribution to the overall SEE cross-section was the L2 and L3 SEU events, while SEFIs accounted only for 11% of SEEs.

We use a novel approach in order to identify errors using the ARM RAS system, which allows us to carefully pinpoint the source of the radiation errors. Unlike previous works in the literature on the radiation testing of GPUs, we were not only able to observe SEUs but also find the reason of each SEE using RAS, which is a unique feature of our work. Most of the errors were correctable, thanks to the ECC protection in the caches. The SEFI errors that caused almost all the restarts were associated with the L3 cache tags. Moreover, although we tested multiple power modes, they did not have a measurable effect on the cross-section, which could be explained by the low number of events we observed during our radiation testing.

Finally, thanks to our ability to find out the source of SEEs in our radiation campaign, we were able to predict the behaviour of errors in orbit using well known and used models assessing the robustness of both devices in two particularly useful orbits for the use of GPUs in earth observation applications.

Acknowledgments

The authors would like to thank Wouter van Burik from HPTC for his assistance in the test. This work was supported by ESA through the 4000136514/21/NL/GLC/my co-funded PhD activity "Mixed Software/Hardware-based Fault-tolerance Techniques for Complex COTS System-on-Chip in Radiation Environments" and the GPU4S (GPU for Space) project. Moreover, it was partially supported by the Spanish Ministry of Economy and Competitiveness under grants PID2019-107255GB-C21 and IJC2020-045931-I (Spanish State Research Agency / http://dx.doi.org/10.13039/501100011033) and the HiPEAC Network of Excellence. We also thank ST Microelectronics for the STM32F401 microcontroller used for power meter and control, which was an award in HiPEAC's CSW Autumn 2019 student challenge and NVIDIA for their assistance in securing the Xavier Industrial units.

REFERENCES

[1] G. Furano, G. Meoni, A. Dunne, D. Moloney, V. Ferlet-Cavrois, A. Tavoularis, J. Byrne, L. Buckley, M. Psarakis, K.-O. Voss *et al.*, "Towards the Use of Artificial Intelligence on the Edge in Space Systems: Challenges and Opportunities," *IEEE Aerospace and Electronic Systems Magazine*, vol. 35, no. 12, pp. 44–56, 2020.

[2] G. Giuffrida, L. Fanucci, G. Meoni, M. Batič, L. Buckley, A. Dunne, C. Van Dijk, M. Esposito, J. Hefele, N. Vercruyssen *et al.*, "The Φ-Sat-1 Mission: The First On-board Deep Neural Network Demonstrator for Satellite Earth Observation," *IEEE Transactions on Geoscience and Remote Sensing*, 2021.

[3] K. McManamon, R. Lancaster, and N. Silva, "ExoMars Rover Vehicle Perception System Architecture and Test Results," *Symposium on Advanced Space Technologies in Robotics and Automation (ASTRA)*, 2017.

[4] Spacebit. (2020) ASAGUMO Rover. [Online]. Available: https://spacebit.com/

[5] M. Mammarella, C. Paissoni, N. Viola, A. Denaro, E. Gargioli, and Massobrio, "The Lunar Space Tug: A Sustainable Bridge Between Low Earth Orbits and the Cislunar Habitat," *Acta Astronautica*, vol. 138, pp. 102 – 117, 2017.

[6] S. Kawamoto, Y. Ohkawa, H. Okamoto, K. Iki, T. Okumura, Y. Katayama, M. Hayashi, Y. Horikawa, H. Kato, and N. Murakami, "Current Status of Research and Development on Active Debris Removal at JAXA," *7th European Conference on Space Debris (SDC7)*, 2017.

[7] B. Balaram, T. Canham, C. Duncan, H. F. Grip, W. Johnson, J. Maki, A. Quon, R. Stern, and D. Zhu, "Mars Helicopter Technology Demonstrator," in *AIAA Atmospheric Flight Mechanics Conference*, 2018.

[8] J. Gaisler, "The SPARC History in Space," *Keynote at ESA Workshop on Avionics, Data, Control and Software Systems (ADCSS)*, 2017.

[9] R. W. Berger et al., "The RAD750 -A Radiation Hardened PowerPC Processor for High Performance Spaceborne Applications," in *IEEE Aerospace Conference Proceedings*, vol. 5, March 2001, pp. 2263–2272.

[10] M. Esposito, S. Conticello, M. Pastena, and B. C. Domínguez, "In-orbit Demonstration of Artificial Intelligence Applied to Hyperspectral and Thermal Sensing From Space," in *CubeSats and SmallSats for remote sensing III*, vol. 11131. International Society for Optics and Photonics, 2019, p. 111310C.

[11] L. Kosmidis, I. Rodriguez, A. Jover-Alvarez, S. Alcaide, J. Lachaize, O. Notebaert, A. Certain, and D. Steenari, "GPU4S: Major Project Outcomes, Lessons Learnt and Way Forward," in *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2021, pp. 1314–1319.

[12] S. Kato, S. Tokunaga, Y. Maruyama, S. Maeda, M. Hirabayashi, Y. Kitsukawa, A. Monrroy, T. Ando, Y. Fujii, and T. Azumi, "Autoware on Board: Enabling Autonomous Vehicles With Embedded Systems," in *2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS)*. IEEE, 2018, pp. 287–296.

[13] R. L. Silva, O. Canciglieri Junior, and M. Rudek, "A Road Map For Planning-deploying Machine Vision Artifacts in the Context of Industry 4.0," *Journal of Industrial and Production Engineering*, pp. 1–14, 2021.

[14] D. A. Oliveira, P. Rech, L. L. Pilla, P. O. Navaux, and L. Carro, "GPGPUs ECC Efficiency and Efficacy," in *2014 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT)*. IEEE, 2014, pp. 209–215.

[20] W. S. Slater, N. P. Tiwari, T. M. Lovelly, and J. K. Mee, "Total Ionizing Dose Radiation Testing of NVIDIA Jetson Nano GPUs," in *2020 IEEE High Performance Extreme Computing Conference (HPEC)*. IEEE, 2020, pp. 1–3.

[15] K. Iniewski, *Radiation Effects in Semiconductors*. CRC press, 2018.

[16] D. A. G. G. de Oliveira, L. L. Pilla, T. Santini, and P. Rech, "Evaluation and Mitigation of Radiation-induced Soft Errors in Graphics Processing Units," *IEEE Transactions on Computers*, vol. 65, no. 3, pp. 791–804, 2015.

[17] S. Duzellier and R. Ecoffet, "Recent Trends in Single-event Effect Ground Testing," *IEEE Transactions on Nuclear Science*, vol. 43, no. 2, pp. 671–677, 1996.

[18] ESCC. (2014) Specification (25100) Single Event Effects Test Method and Guidelines. [Online]. Available: https://escies.org/specification/viewapppub?id=3095

[19] S.-C. Pu, M.-J. Yang, C.-C. Hsu, C.-W. Lai, C.-C. Hsieh, S. H. Lin, Y.-M. Cheng, and P.-T. Chou, "The Empirical Correlation Between Size and Two-photon Absorption Cross Section of CdSe and CdTe Quantum Dots," *Small*, vol. 2, no. 11, pp. 1308–1313, 2006.

[21] D. Hiemstra, C. Jin, Z. Li, R. Chen, S. Shi, and L. Chen, "Single Event Effect Evaluation of the Jetson AGX Xavier Module Using Proton Irradiation," in *2020 IEEE Radiation Effects Data Workshop (in conjunction with 2020 NSREC)*. IEEE, 2020, pp. 1–4.

[22] E. J. Wyrwas, "Proton Testing of AMD e9173 GPU," NASA, Tech. Rep. GSFC-E-DAA-TN72682, 2019.

[23] J. M. Badia, G. Leon, J. A. Belloch, M. Garcia-Valderas, A. Lindoso, and L. Entrena, "Comparison of Parallel Implementation Strategies in GPU-accelerated System-on-Chip Under Proton Irradiation," *IEEE Transactions on Nuclear Science*, 2021.

[24] ARM. (2020) Arm Architecture Reference Manual Supplement Reliability, Availability, and Serviceability (RAS), for Armv8-A. [Online]. Available: https://developer.arm.com/documentation/ddi0587/latest

[25] NVIDIA. (2021) NVIDIA Xavier Industrial Specifications. [Online]. Available: https://developer.nvidia.com/embedded/jetson-xavier-ind

[26] ——. (2021) NVIDIA Xavier NX Specifications. [Online]. Available: https://developer.nvidia.com/embedded/jetson-xavier-nx

[27] HPTC. (2022) Pre-clinical and Clinical Facilities. [Online]. Available: https://www.hollandptc.nl/research/pre-clinical-and-clinical-research-facilities/?theme=researchers

[28] I. Rodriguez, L. Kosmidis, J. Lachaize, O. Notebaert, and D. Steenari, "GPU4S Bench: Design and Implementation of an Open GPU Benchmarking Suite for Space On-board Processing," Universitat Politècnica de Catalunya, Tech. Rep. UPC-DAC-RR-CAP-2019-1, https://www.ac.upc.edu/app/research-reports/public/html/research_center_index-CAP-2019,en.html.

[29] J. Schwank, M. Shaneyfelt, J. Baggio, P. Dodd, J. Felix, V. Ferlet-Cavrois, P. Paillet, D. Lambert, F. Sexton, G. Hash *et al.*, "Effects of Particle Energy on Proton-induced Single-event Latchup," *IEEE Transactions on Nuclear Science*, vol. 52, no. 6, pp. 2622–2629, 2005.

[30] NVIDIA. (2021) NVIDIA Xavier TRM. [Online]. Available: https://developer.nvidia.com/embedded/downloads#?search=trm&tx=$product,jetson_agx_xavier

[31] TRAD. (2022) TRAD Omere. [Online]. Available: https://www.trad.fr/en/space/omere-software/

[32] J. D. Gaffey Jr and D. Bilitza, "NASA/National Space Science Data Center Trapped Radiation Models," *Journal of Spacecraft and Rockets*, vol. 31, no. 2, pp. 172–176, 1994.

[33] M. A. Xapsos, J. Barth, E. Stassinopoulos, E. A. Burke, and G. Gee, "Space Environment Effects: Model for Emission of Solar Protons (ESP): Cumulative and Worst Case Event Fluences," NASA, Tech. Rep. NASA/TP-1999-209763, 1999.