

Measuring the Radiation Reliability of SRAM Structures in GPUs Designed for HPC

Paolo Rech¹, Luigi Carro¹, Nicholas Wang², Timothy Tsai², Siva Kumar Sastry Hari², and Stephen W. Keckler²

¹Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil
{prech, carro}@inf.ufrgs.br

²NVIDIA Corporation, Santa Clara, CA, USA
{niwang, timothyt, shari, skeckler}@nvidia.com

Abstract—Graphics Processing Units specifically designed for High Performance Computing applications require a higher reliability than GPUs used for graphic rendering or gaming. Particular attention should be given to GPU memory structures because these components have been shown to be the most vulnerable for various codes. This paper describes a test framework to assess neutron sensitivity of GPU caches and register files. It also presents results from an extensive radiation test campaign that was performed at LANSCE in Los Alamos, New Mexico. Results show that the neutron sensitivity of the latest GPUs designed for HPC is significantly lower than a previous generation device. This paper also discusses the occurrences of Multiple Bits and Cells Upset and efficacy of the available ECC mechanisms.

Keywords—GPU, neutron sensitivity, caches, multiple errors, ECC

I. INTRODUCTION

Graphics processing units (GPUs) are used not only for graphic rendering or gaming, but also as accelerators in high performance computing (HPC) applications. The high computational power of a GPU combined with low cost, reduced power consumption, and flexible development platforms are pushing their adoption in supercomputers, like TITAN at Oak Ridge National Laboratory [1].

As the newest GPUs are built with cutting-edge technologies, offer a great amount of resources, and operate at high frequencies, they may be particularly susceptible to experience radiation-induced errors, including those originating from the terrestrial neutron radiation environment [2]. When GPUs are employed in gaming or video editing applications, radiation reliability may not be a primary issue as a certain number of failures can be tolerated [3]. In contrast, GPUs designed for HPC applications have reliability as a major concern. The large number of devices that compose a supercomputer (TITAN is composed of 18,688 NVIDIA GPUs [1]) exacerbates the probability of having at least one GPU corrupted by neutrons.

Caches and internal register files, due to their large size, high density, and impact in increasing the performance of the

GPU, represent a major target of radiation effects. As reported in [4] and [5], most of the radiation-induced failures in GPUs are caused by corruption of memory resources. Evaluating precisely the terrestrial radiation-induced error rate of GPU memory structures is extremely important.

GPUs designed for HPC include some architectural solutions specifically introduced to increase the reliability of memory resources. As an example, Tesla C20XX series accelerators feature SECDED ECC on register files, L1 caches, L2 caches, and DRAM. The more advanced Tesla K20 adds a parity-plus-retry mechanism to the read-only texture cache.

This paper presents results of an extensive radiation test campaign performed at LANSCE in Los Alamos, New Mexico. Devices from NVIDIA's K20 and C2050 families were tested to compare the radiation reliability of SRAM structures of GPUs designated for HPC systems, spanning two process and architecture generations, highlighting the reliability improvement of modern devices. In June 2011, 3 of the top 5 supercomputers used an NVIDIA 2050 accelerator. Since 2013, the more advanced K20 has replaced the 2050 and acts as an accelerator in 2 of the top 10 supercomputers [6].

The proposed experimental analysis gives particular attention to the occurrence of multiple bit and multiple cell upsets (MBU and MCU). An MBU occurs when a neutron corrupts more than one bit in the same memory word. If the corrupted bits do not belong to the same word, the neutron induces a so-called MCU.

Recently, manufacturers have adopted error correction code (ECC) mechanisms against soft errors affecting all GPU memory modules. The ECC mechanism included in NVIDIA devices is able to correct single-bit errors and detect double-bit errors. As shown in this paper, the occurrence of multi-bit errors with more than two bits corrupted in the same memory row in the K20 is negligible, justifying the chosen ECC's correcting capabilities. Including an ECC mechanism with better correction capabilities would then introduce useless overhead. In these devices, the user can enable or disable

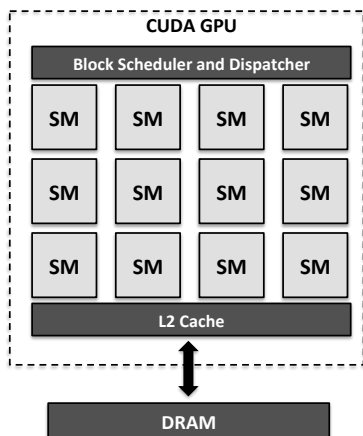


Figure 1: Simplified structure of NVIDIA Kepler and Fermi GPUs composed of an array of streaming multiprocessors that share L2 cache and an external DRAM. The block scheduler and dispatcher are in charge of assigning one or more blocks of threads to idle streaming multiprocessors [7][8].

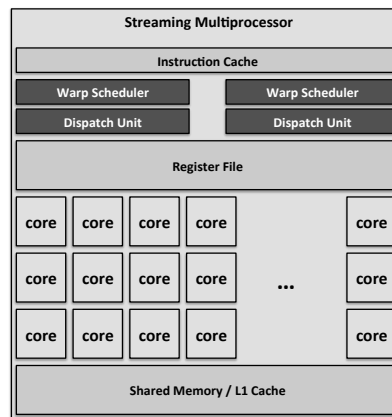


Figure 2: Simplified architecture of a streaming multiprocessor. A block of threads assigned to a SM is divided into warps of 32 threads each. Each thread has a set of dedicated registers and can access a shared memory and L1 cache [7][8].

ECC. During our tests, we turned off ECC to access the raw radiation sensitivity of GPU SRAM structures.

The rest of the paper is organized as follows. Section II describes the GPU internal structure, giving particular attention to memory organization and hierarchy, Section III introduces the experimental setup, Section IV summarizes and discusses the obtained results, and Section V concludes the paper.

II. GPU INTERNAL STRUCTURE

As depicted in Figure 1, GPUs are divided into computing units, named *streaming multiprocessors (SMs)*, each of which has the ability of executing a block of threads in parallel, which are physically executed by *CUDA cores* (i.e., the basic GPU computing unit). The GPU physical design may be viewed as composed of several isolated elementary computing units such that a radiation-induced failure in one CUDA core will affect only the thread assigned to it. Threads that follow the affected one or threads assigned to CUDA cores next to the one that is struck will not be affected. However, the corruption of shared resources like the cache or the scheduler may generate multiple output errors [4][5].

Figures 1 and 2 depict a simplified architecture for Kepler and Fermi GPUs, highlighting the major memory structures. An L2 cache is shared among all SMs (Figure 1). A thread executing on an SM has exclusive access to dedicated internal registers and shares L1 cache or shared memory with all threads in the same SM (Figure 2). A thread can access the L1 cache or shared memory of only the SM in which it is being executed. Caches and DRAM are particularly critical to resilience, as a radiation corruption in a bit may affect the executions of many threads, leading to multiple output errors [4].

III. GPU RADIATION TEST SETUP

A) Tested Device

We tested two commercially available NVIDIA GPUs: the Fermi-based C2050 and the Kepler-based K20. These GPUs were designed in 40nm and 28nm technology nodes, respectively and were released approximately two years apart.

The C2050 features an 1150MHz SM core clock, a 768kB L2 cache, a total of 896kB in L1 caches, and a total of 1.75MB of register file storage. The K20 features a 706MHz SM core clock, a 1.25MB L2 cache, a total of 832kB in L1 cache, and a total of 3.25MB of register file storage.

B) Experimental Methodology

Experiments were performed at Los Alamos National Laboratory's (LANL) Los Alamos Neutron Science Center (LANSCE) Irradiation of Chips and Electronics House II, called ICE House II, in August 2013. The ICE House II beam line provides a white neutron source that emulates the energy spectrum of the atmospheric neutron flux. The available neutron flux was approximately $5 \times 10^5 \text{ n}/(\text{cm}^2 \cdot \text{s})$ for energies above 10MeV. The beam was focused on a spot with a diameter of 2 inches plus 1 inch of penumbra, which provided uniform irradiation of the GPU chip without directly affecting nearby board power control circuitry and DDR memories (Figure 3).

A desktop PC running Ubuntu 10.04 with the CUDA 5.5 driver controlled the board under test. The GPU was connected through a 14-inch PCI-Express bus extension to prevent neutron effect to the host PC. Moreover, the extension was provided with fuses to prevent radiation-induced latch-ups in the GPU from propagating to the host PC motherboard. No latch-up was observed during the overall radiation campaign.

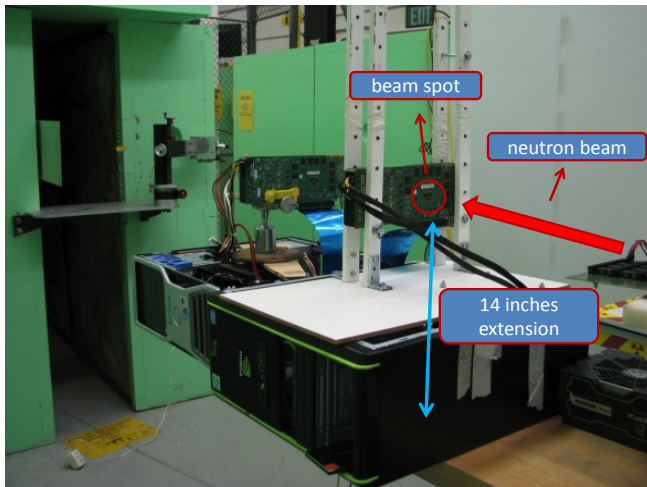


Figure 3: Radiation test setup inside the ICE House II, LANSCE, LANL, Los Alamos, New Mexico.

C) Caches and Register File Test Frameworks

To measure cache and register file neutron sensitivity, it is first necessary to (1) force the memories to a given value, (2) expose the device to a given radiation fluence (i.e., the number of high-energy neutrons hitting the device during the exposure time per unit area) and allowing errors to accumulate, and (3) check whether all the bits in the memory are still holding the initial value. The first and last steps are particularly critical, since specific cells in a cache are not directly accessible to the user.

The tests were written in the CUDA programming language and run natively on the device under test. As is typical in GPU programming, each kernel's work was partitioned among many different threads executing in parallel. For the SM register file test, each thread was responsible for testing its privately accessible registers. For the L2 cache test, each thread was responsible for testing a pre-assigned portion of the L2 cache. To precisely measure the duration of testing, each thread recorded a timestamp into system memory at the beginning and end of its program execution. After the completion of each kernel, the CPU program sums the elapsed execution time of each thread to determine the aggregate runtime of all threads. In this way, we were able to precisely calculate the amount of testing time to feed into cross section calculations.

The exposure time of register files and caches was parameterized and carefully engineered to make it unlikely for more than one neutron to generate a failure. This approach is essential to evaluate the occurrences of MBUs and MCUs. If more than one error is detected in the same test, those errors are likely to be produced by the same impinging neutron. If the errors belong to the same word, an MBU is detected. Otherwise, an MCU is detected.

The pattern to be written and checked in the memory was also parameterized. Some memory cell design is not symmetric, meaning that the pull-up and pull-down transistor

capacitances may differ. Under radiation, this turns into a different cross section of bits set to 0 or to 1. Both patterns were checked to measure the probability of corrupting bits set to 0 or to 1 respectively. Finally, we tested a checkerboard pattern with alternating 1s and 0s. Such a pattern represents the worst case for MBUs or MCUs, as the sensitive nodes of cells are close to each other, and more easily affected by the same impinging neutron [9].

During the overall test campaign, we disabled the available ECC mechanism to measure the raw sensitivity of the SRAM structures of the GPUs.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A) L2 Cache Neutron Sensitivity

To evaluate the reliability of GPU memory structures we first experimentally measured the cross section for the L2 caches of K20 and C2050 boards. The cross section was calculated by dividing the number of observed errors by the received neutron fluence and the number of exposed bits. The higher the cross section, the more sensitive to radiation a bit is. In this calculation, multiple bit upset (MBU) and multiple cell upset (MCU) were considered as events generated by a single neutron. A discussion on MBU and MCU is included in the following subsection.

Experiments were performed by writing all zeros (00 pattern), all ones (FF pattern), and alternating ones and zeros (AA pattern) in the storage elements. For the C2050, the AA pattern did not provide a statistically significant number of errors, and thus is not listed in Table I. Reported values (in arbitrary units or a.u.) were normalized to the cross section of K20 L2 cache obtained with an FF pattern, which was the lowest measured cross section. This table also shows the 95% confidence intervals for our results, which is a combination of neutron counts uncertainty and statistical error.

As depicted in Figure 4, the L2 cross section for the K20 depends on the written pattern. For the 00 pattern, the L2 cross section is approximately 40% higher than that of the FF pattern. This means that L2 bits set to 0 are more likely to be corrupted by high-energy neutrons than bits set to 1. As the AA pattern is a combination of 00 and FF patterns, one expects and sees a cross section in between those of 00 and FF. The same trend applies to C2050 L2 cache sensitivity (last row of Table I). The observed dependence on test pattern is due to the asymmetries intrinsic in the cache cell design. This specific result can be achieved only through radiation experiments, and is fundamental to precisely evaluating the resilience of GPUs.

Fault injection simulators may particularly benefit from the reported results. Table I provides experimentally obtained error probability functions that can be applied to the fault simulators. In particular, taking the reported 0-to-1 and 1-to-0 error probabilities into account will result in more precise simulation outcomes than considering the two radiation-induced transitions as equally probable.

Figure 4 shows the sensitivity of K20 and C2050 L2 cache cells graphically. Note that these results are normalized to the number of bits, thus the reported relative sensitivities do

TABLE I
K20 AND C2050 L2 CACHE NORMALIZED CROSS SECTIONS
(NORMALIZED TO THE K20 L2 CROSS SECTION FOR FF PATTERN)

	00	FF	AA
K20 L2 [a.u.]	1.44±10%	1.00±10%	1.36±10%
C2050 L2 [a.u.]	3.09±12%	2.29±12%	N/A

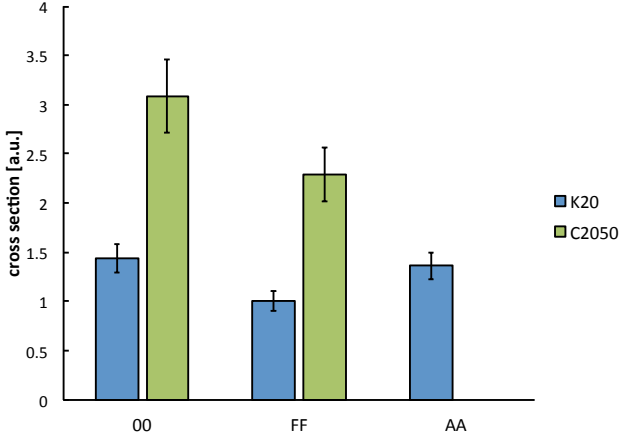


Figure 4: Normalized cross sections for the K20 and C2050 L2 cache cells. The K20 L2 cache cells are approximately 3 times less sensitive to high energy neutrons with respect to the C2050 cells.

not account for the increase in the size of L2 cache from C2050 to K20. Figure 4 can be used to compare the sensitivity to neutrons of a bit in the L2 of K20 and C2050 caches. It is easy to note that K20 cache cells have an increased reliability with respect to those of the C2050.

We believe that the increased reliability is caused mainly by two factors: process node (40nm for the C2050 vs. 28nm for the K20) and a potential different resulting bit cell design. A reduced transistor dimension lowers the device cross section, as the exposed area becomes smaller. Nevertheless, reducing the feature size typically reduces the device node capacitance. It is possible that a neutron hitting a bit cell in K20 may have a higher probability of generating a failure with respect to a neutron hitting a bit cell in C2050. The combination of reduced sensitive area and capacitance is expected to bring an overall benefit in reducing the radiation sensitivity for future technologies [10].

Supply voltage plays a key role in the radiation sensitivity of memory cells. Reducing the memory cells supply voltage has the countermeasure of increasing linearly the device radiation sensitivity [11]. For current and future devices, the voltage reduction per generation is limited to 5-10% [12]. Hence, no significant radiation sensitivity increase is expected in the K20 with respect to the C2050 due to a reduced supply voltage. As supply voltage reduction in future technology generations is expected to be small, we anticipate that increases in circuit sensitivity will be relatively small.

B) Register File Neutron Sensitivity

Table II reports the experimentally obtained cross sections for the register file cells of both K20 and C2050 boards, with

TABLE II
K20 AND C2050 REGISTER FILE NORMALIZED CROSS SECTIONS
(NORMALIZED TO THE K20 L2 CROSS SECTION FOR FF PATTERN)

	00	FF	AA
K20 Reg [a.u.]	1.17±10%	1.13±10%	1.11±10%
C2050 Reg [a.u.]	2.86±12%	2.64±12%	N/A

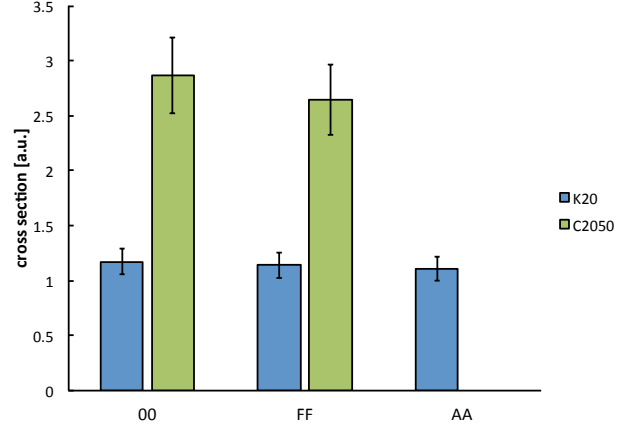


Figure 5: Normalized cross sections for the K20 and C2050 register file cells. The K20 register file cells are approximately 3 times less sensitive to high energy neutrons with respect to the C2050 cells.

95% confidence intervals. The cross sections were obtained by dividing the number of observed errors by the received neutron fluence and the number of exposed bits. Reported values were normalized with the K20 L2 cache cross section obtained with the FF pattern to ease the comparison of register and cache cell sensitivity.

Our results indicate that the sensitivity of register file cells do not vary significantly with the data pattern used during the test (Figure 5). Further investigation is needed to understand the reason behind such an observation, which is a part of our ongoing effort. Figure 5 also compares the cross section of K20 and C2050 register file cells. Again, the K20 shows a higher reliability than the C2050.

Data in both Tables I and II are normalized with the FF pattern for K20 L2 cache cross section, allowing a direct comparison between L2 and register file sensitivity. This data also permits building a fault injection framework that precisely emulates radiation-induced failure in both L2 and register file.

C) Multiple Bit and Multiple Cell Upset

Even if the K20 device design increases the reliability of both cache and internal register cells significantly, when a large number of GPUs operate simultaneously, as in supercomputers, the overall system neutron-induced error rate can become unacceptable. Both the C2050 and K20 are equipped with an ECC mechanism that is able to detect double bit errors and correct single bit errors. Such an ECC is called Single Error Correction Double Error Detection (SECDED). To evaluate the efficacy of the ECC mechanism, one must measure the probability of occurrence of a multiple bit upset (MBU) in the GPU memory resources. An MBU occurs when

a single impinging neutron interacts with more than one transistor in such a way to flip more than one memory bit in the same row of a RAM. If the bits flipped do not belong to the same row, the neutron generates a so-called multiple cells upset (MCU). Both MBU and MCU are particularly critical for electronic device reliability. MBUs, in particular, can defeat ECC mechanisms if sufficient word interleaving is not employed.

The test framework described in the previous section enables us to measure of the number of bits corrupted per word. Table III lists the percentage of errors detected in register files that were found to have two or more bits corrupted. Our results show that the register file row is somewhat prone to MBUs. As shown in Figure 6, the K20 register files are more likely to be affected by MBUs than C2050s. As the K20 is implemented in a 28nm technology node, its transistors are smaller than those in the 40nm C2050. Thus it is easier for a neutron to interact with more than one transistor in the K20 than in the C2050. No MBU with more than 2 bits corrupted per row was observed during the experiments. As a result, the employed SECDED ECC mechanism is guaranteed to detect all register file faults observed during our testing.

Table IV lists the percentage of neutron-induced failures in the register files that were MCUs. Unfortunately, the tests did not provide a statistically significant number of errors for the C2050. The K20 seems very prone to experience MCUs, as expected for a 28nm SRAM with high density and capacity. An MCU will not affect the ECC mechanism efficacy, as all the instantiated words in the cache and in the internal registers are protected independently. However, like single-bit upsets, the occurrence of an MCU is a major concern for GPU reliability without error detection and correction. From Table III and IV it is clear that MBUs are less common than MCUs in K20 internal registers. We suspect that this is because most of the multiple bit errors occur along the bit-lines as shown by [13] and are correctable with ECC.

V. CONCLUSIONS

Radiation induced failures represent a major issue for all devices designed for the HPC market, including GPUs. The high number of devices that work in parallel in a supercomputer makes it very likely for at least one GPU to be corrupted by neutrons. Memory structures including caches and register files are responsible for the majority of failures on modern GPUs and should be carefully designed and protected.

In this paper, we developed a test framework to evaluate the sensitivity of both L2 caches and registers of GPUs. As experimental results demonstrate, current GPUs devoted to the HPC market are more reliable than previous generation devices. The efforts made in increasing GPU radiation reliability succeeded in lowering the SRAM structures' neutron sensitivity. If the device error rate is still found to be too high for the mission requirements, an ECC mechanism to correct single errors and detect double errors can be activated. As experimentally demonstrated, such an ECC is sufficient to detect and correct all observed errors.

TABLE III
K20 AND C2050 MBU PERCENTAGE

	00	FF	AA
K20 Reg	4%	6%	1%
C2050 Reg	1%	1%	N/A

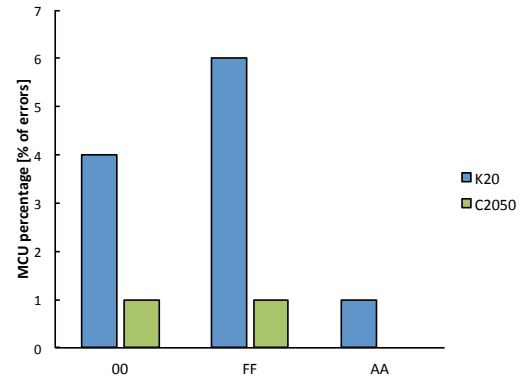


Figure 6: Percentage of errors that were found to be double Multiple Cells Upset for the K20 and C2050 register files.

TABLE IV
K20 AND C2050 MCU PERCENTAGE

	00	FF	AA
K20 Reg	19%	18%	17%
C2050 Reg	N/A	N/A	N/A

REFERENCES

- [1] Introducing Titan, Available online: <http://www.olcf.ornl.gov/titan/>
- [2] A. Dixit and A. Wood, "The Impact of New Technology on Soft Error Rates", IEEE International Reliability Physics Symposium (IRPS), 2011, pp. 5B.4.1-5B.4.7.
- [3] M. A. Breuer, S. K. Gupta, and T. M. Mak, "Defect and Error Tolerance in the Presence of Massive Number of Defects", IEEE Design and Test of Computers, Vol. 21, No. 3, pp. 216-227, 2004.
- [4] P. Rech, C. Aguiar, C. Frost, and L. Carro, "An Efficient and Experimentally Tuned Software-Based Hardening Strategy for Matrix Multiplication on GPUs", IEEE Transactions on Nuclear Science, Vol. 60, No. 4, pp. 2797-2804, 2013.
- [5] P. Rech, L. L. Pilla, F. Silvestri, C. Frost, P. O. A. Navaux, M. Sonza Reorda, and L. Carro, "Neutron Sensitivity and Hardening Strategies for Fast Fourier Transform on GPUs", IEEE Radiation Effects on Components and Systems (RADECS), 2013.
- [6] Top 500 Supercomputer Sites, Available online: <http://www.top500.org/lists>
- [7] Whitepaper: "NVIDIA's Next Generation CUDA Compute Architecture: Fermi", 2009.
- [8] Whitepaper: "NVIDIA's Next Generation CUDA Compute Architecture: Kepler GK110", 2012.
- [9] A. D. Tipton, J. A. Pellish, P. R. Fleming, R. D. Schrimpf, R. A. Reed, R. A. Weller, M. H. Mendenhall, and L. W. Massengill, "High Energy Neutron Multiple-Bit Upset", International Conference on IC Design and Technology (ICICDT), 2007.
- [10] R. Baumann, "Radiation-Induced Soft Errors in Advanced Semiconductor Technologies", IEEE Transactions on Devices and Materials Reliability, Vol. 5, No. 3, pp. 305-316, 2005.
- [11] J. F. Zigler and H. Pucher, "SER - History, Trends and Challenges", Cypress press, 2010.

- [12] M. White, "Scaled CMOS Technology Reliability User Guide", JPL Publication 09-33, 2010.
- [13] D. F. Heidel, P. W. Marshall, J. A. Pellish, K. P. Rodbell, K. A. LaBel, J. R. Schwank, S. E. Rauch, M. C. Hakey, M. D. Berg, C. M. Castaneda, P. E. Dodd, M. R. Friendlich, A. D. Phan, C. M. Seidleck, M. R.

Shaneyfelt, and M. A. Xapsos, "Single-Event Upsets and Multiple-Bit Upsets on a 45 nm SOI SRAM," IEEE Transactions on Nuclear Science, Vol. 56, No. 6, pp. 3499-3504, 2009.