

Enhancing neural network robustness: Laser fault injection resistance in 55-nm SRAM for space applications

Qing Liu(刘清), Haomiao Cheng(程浩淼), Xiang Yao(姚骧), Zhengxuan Zhang(张正选), Zhiyuan Hu(胡志远), and Dawei Bi(毕大炜)

Citation: Chin. Phys. B, 2025, 34 (**4**): 046104. DOI: 10.1088/1674-1056/adacd1
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Classification and structural characteristics of amorphous materials based on interpretable deep learning

Jiamei Cui(崔佳梅), Yunjie Li(李韵洁), Cai Zhao(赵偲), and Wen Zheng(郑文)
Chin. Phys. B, 2023, 32 (**9**): 096101. DOI: 10.1088/1674-1056/acd7d0

Mechanisms of atmospheric neutron-induced single event upsets in nanometric SOI and bulk SRAM devices based on experiment-verified simulation tool

Zhi-Feng Lei(雷志锋), Zhan-Gang Zhang(张战刚), Yun-Fei En(恩云飞), Yun Huang(黄云)
Chin. Phys. B, 2018, 27 (**6**): 066105. DOI: 10.1088/1674-1056/27/6/066105

Large energy-loss straggling of swift heavy ions in ultra-thin active silicon layers

Zhang Zhan-Gang (张战刚), Liu Jie (刘杰), Hou Ming-Dong (侯明东), Sun You-Mei (孙友梅), Zhao Fa-Zhan (赵发展), Liu Gang (刘刚), Han Zheng-Sheng (韩郑生), Geng Chao (耿超), Liu Jian-De (刘建德), Xi Kai (习凯), Duan Jing-Lai (段敬来), Yao Hui-Jun (姚会军), Mo Dan (莫丹), Luo Jie (罗捷), Gu Song (古松), Liu Tian-Qi (刘天奇)
Chin. Phys. B, 2013, 22 (**9**): 096103. DOI: 10.1088/1674-1056/22/9/096103

Angular dependence of multiple-bit upset response in static random access memories under heavy ion irradiation

Zhang Zhan-Gang (张战刚), Liu Jie (刘杰), Hou Ming-Dong (侯明东), Sun You-Mei (孙友梅), Su Hong (苏弘), Duan Jing-Lai (段敬来), Mo Dan (莫丹), Yao Hui-Jun (姚会军), Luo Jie (罗捷), Gu Song (古松), Geng Chao (耿超), Xi Kai (习凯)
Chin. Phys. B, 2013, 22 (**8**): 086102. DOI: 10.1088/1674-1056/22/8/086102

Radiation damage effects on power VDMOS devices with composite SiO_2–Si_3N_4 films

Gao Bo (高博), Liu Gang (刘刚), Wang Li-Xin (王立新), Han Zheng-Sheng (韩郑生), Song Li-Mei (宋李梅), Zhang Yan-Fei (张彦飞), Teng Rui (腾瑞), Wu Hai-Zhou (吴海舟)
Chin. Phys. B, 2013, 22 (**3**): 036103. DOI: 10.1088/1674-1056/22/3/036103

Enhancing neural network robustness: Laser fault injection resistance in 55-nm SRAM for space applications

Qing Liu(刘清)[1,2], Haomiao Cheng(程浩淼)[1,2], Xiang Yao(姚骧)[1,2]
Zhengxuan Zhang(张正选)[1,2], Zhiyuan Hu(胡志远)[1,2], and Dawei Bi(毕大炜)[1,2,†]

[1] *State Key Laboratory of Materials for Integrated Circuits, Shanghai Institute of Microsystem and Information Technology, Shanghai* 200050, *China*
[2] *University of the Chinese Academy of Sciences, Beijing* 100049, *China*

(Received 19 December 2024; revised manuscript received 20 January 2025; accepted manuscript online 22 January 2025)

The integration of artificial intelligence (AI) with satellite technology is ushering in a new era of space exploration, with small satellites playing a pivotal role in advancing this field. However, the deployment of machine learning (ML) models in space faces distinct challenges, such as single event upsets (SEUs), which are triggered by space radiation and can corrupt the outputs of neural networks. To defend against this threat, we investigate laser-based fault injection techniques on 55-nm SRAM cells, aiming to explore the impact of SEUs on neural network performance. In this paper, we propose a novel solution in the form of Bin-DNCNN, a binary neural network (BNN)-based model that significantly enhances robustness to radiation-induced faults. We conduct experiments to evaluate the denoising effectiveness of different neural network architectures, comparing their resilience to weight errors before and after fault injections. Our experimental results demonstrate that binary neural networks (BNNs) exhibit superior robustness to weight errors compared to traditional deep neural networks (DNNs), making them a promising candidate for spaceborne AI applications.

Keywords: single event effects, convolutional neural network, fault injection, SRAM
PACS: 61.82.Fk, 61.80.Hg, 07.05.Mh, 85.35.–p
DOI: 10.1088/1674-1056/adacd1 **CSTR:** 32038.14.CPB.adacd1

1. Introduction

To achieve high-performance, cost-effective solutions for new space missions, a novel paradigm has emerged. This paradigm is characterized by the innovative adaptation of established technologies and methodologies from the traditional space industry. As a result, the industry has increasingly shifted toward flexible, cost-efficient, and high-performance equipment for space missions. Among these, commercial-off-the-shelf (COTS)[1] devices, which are pre-manufactured and readily available, have gained popularity for critical mission tasks due to their ability to meet these requirements. These attributes make COTS SRAM an optimal choice for space-based applications. SRAM-based FPGAs are widely deployed in satellites and other spacecraft because of their abundant logic resources and superior reconfigurability.

Convolutional neural networks (CNNs) are increasingly utilized across a wide range of fields, from image processing to high-performance computing in space applications.[2–5] However, when deployed in spaceflight contexts, CNNs encounter significant reliability issues due to the effects of space radiation. Specifically, the memory of hardware accelerators used to run CNNs has been found to be particularly vulnerable soft errors caused by space radiation. After irradiation to a fluence of 2.6×10^7 cm^{-2}, 29% of 10^5 output neuron magnitudes had changed due to SEUs in the SRAM.[6] The memory of hardware accelerators holds a large amount of data and weight parameters for CNNs. SEUs in storage circuit can result in abnormal data interactions, particularly those affecting the weight parameters, can propagate through the networks computations, ultimately resulting in reduced accuracy in CNN performance.

To mitigate soft error rates, various traditional system-level SEU mitigation techniques have been developed, focusing on enhancing fault tolerance through redundancy methods.[7,8] Prominent techniques include the triple module redundancy (TMR) architecture and error correction coding (ECC) technology. However, due to the substantial computational demands and large parameter sets characteristic of CNNs, these methods are suboptimal for deep learning applications. On the one hand, the large number of weight parameters in CNNs consumes significant storage resources on hardware accelerators. On the other hand, error correction coding methods reduce computational speed and are ineffective in addressing multiple bit upsets (MBUs).[9] To be applied broadly across many fields, neural networks must be shown to be robust against main memory errors caused by nuclear radiation, or cosmic rays.

Recent research has primarily focused on the impact of single event effects (SEEs) on basic convolutional neural network (CNN) models, particularly in the realm of image classification tasks. For example, the study conducted by Libano

[†]Corresponding author. E-mail: davidb@mail.sim.ac.cn
© 2025 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

et al.[10] demonstrated that quantizing convolutional layer weights can reduce radiation sensitivity by 39%. Additionally, Sergio Alejandro Rosales Vargas *et al.*[11] introduced the dual-binary approach for modeling the effects of single event transients (SETs) on networks presents innovative methods to address transient errors in smaller-scale models. Xu Zhao *et al.*[12] developed a CNN system for the MNIST dataset using a 28-nm system-on-chip (SoC) and assessed its susceptibility to SEEs through alpha particle irradiation experiments and fault injection techniques. While previous studies have advanced the understanding and mitigation of SEEs' impacts on CNN weights and overall model performance, a gap persists in exploring the comprehensive effects of SEEs, particularly regarding data integrity. SEEs can lead not only to erroneous weight updates but also to the corruption of data stored in memory, including inputs to CNNs such as images in the NMIST-Fashion dataset. This dual-point of impact introduces additional complexity, further reducing the reliability of the CNN model.

To address these challenges, this work uses image denoising accuracy to assess the effects of radiation in binary neural networks and single-precision CNNs. These networks are designed to combat data corruption and ensure reliability in irradiation environments. The main contributions of this study are summarized as follows:

(i) The impact of single event upsets (SEUs) on neural network feature inputs and model weights is examined. It is found that bit flips occurring in SRAM mainly follow a Poisson distribution.

(ii) Neural networks are employed at the system level to denoise and restore image pollution caused by radiation. The results indicate that soft errors have a greater impact on the last layer of convolution in the network.

(iii) Laser fault injection in Bin DNCNN suggests that binary neural networks exhibit superior robustness to weight errors compared to traditional deep neural networks.

The rest of the paper is organized as follows. Section 2 summarizes the state-of-the-art fault injection platforms, as well as the background related to binary neural networks. Section 3 describes the experimental methodology, including the proposed solution. Experimental results and a detailed discussion are presented in Section 4. Section 5 concludes the paper.

2. Background

2.1. Fault injection

To investigate the effects of SEEs caused by various types of radiation on integrated circuits, fault injection has emerged as a fundamental and straightforward method. Several fault injection techniques have been developed,[5,13] with simulation-based and hardware-based fault injection being two commonly used approaches for injecting faults into memory. Both techniques can alter memory logic values with high precision.

Simulation-based software fault injection methods are widely utilized due to their simplicity and adaptability. Specifically, register-transfer level (RTL) fault injection simulates the circuitry of the device under test, enabling the observation of fault propagation from the gate level to the software output. Although this approach is generally cost-effective, the low-level abstraction involved in fault injection can require considerable time to complete, with the duration depending on the complexity of the hardware and the application being executed.[14]

Hardware-based fault injection can be performed using contact-based methods, which perturb the external interface of the integrated circuit,[15,16] or non-contact methods that use external energy sources. A prominent technique involves laser beam-induced faults in SRAM. By exposing the chip to a laser beam, a temporary conductive channel is formed in the dielectric, which causes the transistor to switch states in a precise and controlled manner. By adjusting laser parameters such as beam diameter and emission energy, attacks can not only affect signal timing on the chip but also precisely flip bits in the memory.[17,18]

2.2. Binary neural networks

Binary neural network (BNN) utilizes the most lightweighted quantization $\{-1, +1\}$ as its weight and activation values in each layer. The full-precision convolution $\mathcal{X}_i^{\mathrm{f}} \otimes \mathcal{W}_i^{\mathrm{f}}$ ($\mathcal{X}_i^{\mathrm{f}}$, $\mathcal{W}_i^{\mathrm{f}}$, and $\otimes$ are full-precision activations, weights, and Conv respectively) is approximated by the binary convolution $\mathcal{X}_i^{\mathrm{b}} \otimes \mathcal{W}_i^{\mathrm{b}}$. For binary convolution, both weights and activations ($x_{i,j}^{\mathrm{b}}$) are binarized to -1 and $+1$. Efficient bitwise XNOR and bit counting operations can replace computationally heavy floating-point matrix multiplication, which can be defined as[19]

$$\mathcal{X}_i^{\mathrm{b}} \otimes \mathcal{W}_i^{\mathrm{b}} = \mathrm{bitcount}\left(\mathrm{XNOR}\left(\mathcal{X}_i^{\mathrm{b}}, \mathcal{W}_i^{\mathrm{b}}\right)\right), \tag{1}$$

$$x_{i,j}^{\mathrm{b}} = \mathrm{Sign}\left(x_{i,j}^{\mathrm{f}}\right) = \begin{cases} +1, & \text{if } x_{i,j}^{\mathrm{f}} > 0, \\ -1, & \text{if } x_{i,j}^{\mathrm{f}} \le 0. \end{cases} \tag{2}$$

In contrast to full-precision networks, binarized models significantly expedite the forward propagation process and alleviate computational complexity. However, they may experience a decrease in accuracy due to the constrained representational capacity of features and the non-differentiability of binary functions. To counteract this issue, various approaches have been proposed to ameliorate the adverse effects of network binarization and enhance performance without substantially increasing computational complexity. BNN[20] pioneers the complete binarization of CNN weights and activations, employing Xnor and PopCount operations to simplify full-precision matrix multiplication. Xnor-Net[21] enhances the performance of BNN on large-scale datasets by

introducing scaling factors to both the weights and activations. ReActNet[22] further optimizes BNN performance on extensive datasets by employing reshaped output distributions. This is accomplished through the incorporation of learnable thresholds before the sign function and the integration of RPReLU activation functions following each residual connection.BiSRNet,[23] designed a new binary convolution module called binarized spectral redistribution convolution for spectral compression reconstruction tasks, which achieved results comparable to full precision CNN.

3. Methodology

In this section, we first present the practical laser fault injection setup in Subsection 3.1. Then, in Subsection 3.2, we provide a detailed introduction to the proposed binary neural network.

3.1. Laser experiments

Our setup is depicted in Fig. 1(a). The laser source of the test system was pulsed Nd:YAG laser, with a wavelength of 1064 nm and a pulse duration of 25 ps, and its penetration depth in silicon reaches 1000 μm, the depth for sensitive region in a device is less than 30 μm.[24] The pulse repetition rate ranged from 1 kHz to 5 kHz, and the focused laser spot could be adjusted between 1 m and 2 m. The device under test (DUT) was a 55-nm, 64-kbit radiation-hardened SRAM[25] as shown in Fig. 1(b), placed on a field-programmable gate array (FPGA)-based testing system. The DUT was supplied with a core voltage of 1.2 V and a peripheral I/O voltage of 3.3 V by the testing system. The package of this chip was opened so that there is direct visibility on a back-side silicon die with a laser. The board was placed on an XYZ positioning table with a step precision of 0.1 m in the XY directions. A variety of data patterns were used, including blanket zero data (0000), blanket one data (FFFF), checkerboard data (5A5A), image data (each pixel data was represented as an 8-bit unsigned integer), and pre-trained neural network weight data (represented as 32-bit single precision numbers). The data pattern was written into the DUT by the testing system before laser irradiation. Five laser energies were chosen and the calculated equivalent linear energy transfer (LET) was similar to heavy ion, as shown in Table 1. Three samples per cell were measured for laser irradiation.

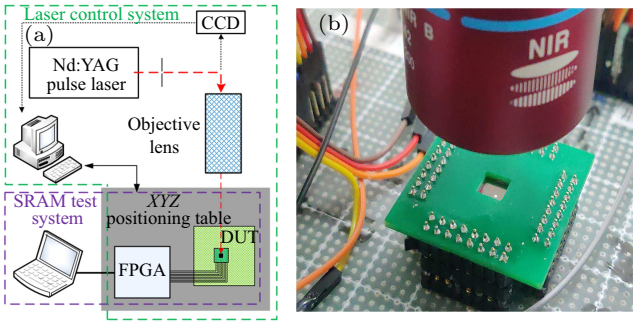


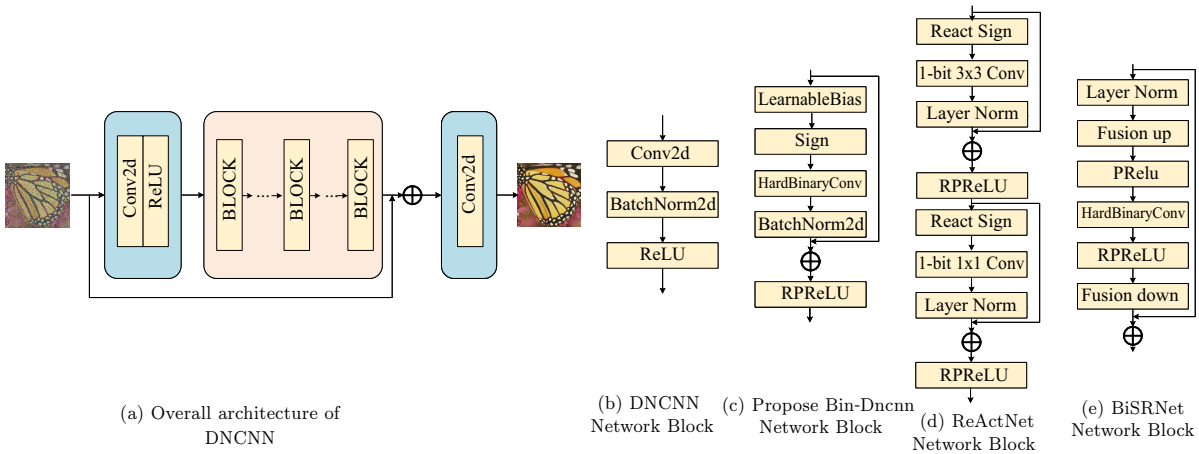
Fig. 1. (a) Setup components and (b) device under test.

Table 1. Laser's equivalent LET.

Laser energy (pJ)	150	180	350	900	1800
LET (MeV·cm^2/mg)	5±1.25	8±2.0	15±3.75	37±9.25	75±18.75

3.2. Proposed method

To address the issue of radiation-induced data corruption, such as the degradation of image quality, we propose a binary convolutional neural network-based image denoising approach, referred to as Bin-DNCNN. The proposed method builds upon the existing DnCNN,[26] as illustrated in Fig. 2(a). The core building block of DnCNN, depicted in Fig. 2(b), consists of key components, including convolutional (Conv) layers, batch normalization (BN) layers, and ReLU activation layers. Batch normalization, in particular, helps mitigate the internal covariate shift encountered during training, accelerating the learning process and preventing the vanishing gradient problem. The primary function of DnCNN is not to directly transform noisy images into noise-free ones but to identify and model noise patterns within the noisy images. This residual learning strategy improves model stability during training and enhances the effectiveness of denoising.



(a) Overall architecture of DNCNN

(b) DNCNN Network Block

(c) Propose Bin-Dncnn Network Block

(d) ReActNet Network Block

(e) BiSRNet Network Block

Fig. 2. (a) Overall architecture of DNCNN, network block for (b) DNCNN, (c) Bin-DNCNN, (d) ReactNet, (e) Bisrnet.

To improve the performance of binary neural networks, the Bin-DNCNN architecture, shown in Fig. 2(c), follows the design principles of ReAct-Net, depicted in Fig. 2(d). Residual connections are incorporated to refine activation information, enhancing the networks ability to propagate and preserve feature representations. Additionally, learnable biases are introduced in the sign and activation functions to correct gradient mismatches, optimizing the activation distribution, which is crucial for the binary network's convergence and accuracy during training.[22] The denoising performance is also compared with the BiSRNet network, shown in Fig. 2(e).

Following a description of the experimental methodology, we provide insights into the effects of SEU errors on both the feature inputs to neural networks and the weights of the models.

4. Experimental results

In this section, we detail the impact of SEU errors within a medium-sized convolutional neural network. Following a description of the experimental methodology, we provide insights into the effects of SEU errors on both the feature inputs to neural networks and the weights of the models.

4.1. The impact of SEU on data inputs to neural networks

The measured SEU cross-section under pulsed laser irradiation is shown in Fig. 3. The Weibull function, as given in Eq. (3), was used to fit the SEU cross-sections for various data patterns stored in SRAM, where σ represents the SEU cross-section, σ_{sat} denotes the saturation SEU cross-section, W is the shape factor (a dimensionless parameter), S is the exponential factor (also a dimensionless parameter), and L_{th} is the threshold LET for SEU occurrence. According to the fitted curves in Fig. 3, the L_{th} and σ_{sat} of the SRAM are 5.61 MeV·cm^2/mg and 2.36×10^{-7} cm^2/bit, respectively. As the LET increased from 37 MeV·cm^2/mg to 75 MeV·cm^2/mg, the device current began to increase exponentially, and single event latch-ups (SELs) occurred in the DUT, causing irreparable damage to the device.

$$\sigma = \begin{cases} 0, & L < L_{\text{th}}, \\ \sigma_{\text{sat}}\left(1-\exp\left[-\left(\dfrac{L-L_{\text{th}}}{W}\right)^{S}\right]\right), & L \geqslant L_{\text{th}}. \end{cases} \tag{3}$$

Figure 4 illustrates the bit flips caused by SEU in image data under various LET values of 8, 15, and 37 MeV·cm^2/mg, with each noise pixel in the image indicating a single-bit flip or multiple-bit flips. We use two metrics,[27] SSIM (structural similarity index measure) and PSNR (peak signal-to-noise ratio), to assess the impact of radiation on image data. The lower the values of these metrics, the more severe the data contamination. It is evident from the figure that higher LET lead to an increase in noise pixels and a greater degradation of image quality.

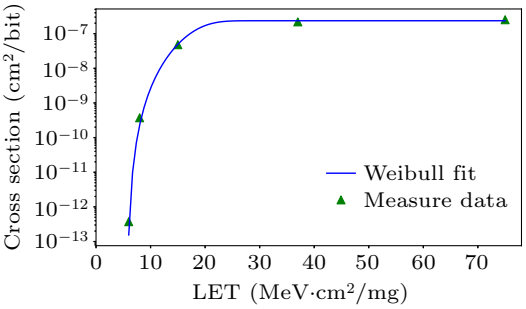


Fig. 3. The cross sections of 64-kbit SRAM.

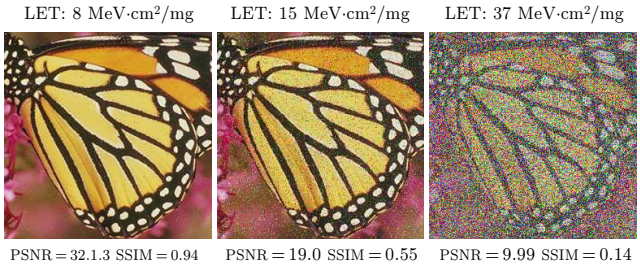


Fig. 4. The noise in the image data caused by bit flipping due to SEU under LET values of 8, 15, 37 MeV·cm^2/mg.

To evaluate the impact of irradiation on data integrity, bitwise XOR operations and bit counting were performed on the data before and after laser irradiation. The number of bit flips, denoted as $N_{\text{ff}} = \text{Bitcount}(X_{\text{pre-irradiation}} \oplus X_{\text{post-irradiation}})$, provides a statistical measure of the bit flips. As shown in Fig. 5, the data remained unchanged when the LET was below 5 MeV·cm^2/mg. At an LET of 8 MeV·cm^2/mg, fewer than 1% of the bits underwent SEU-induced single-bit flips. When the LET reached 15 MeV·cm^2/mg, multiple-bit flips began to occur. At an LET of 37 MeV·cm^2/mg, the proportion of multiple-bit flips increased to 70%. This is attributed to the charge-sharing effect in SRAM cells, where a single-particle upset in one bit can cause adjacent bits to flip, leading to consecutive bit flips.

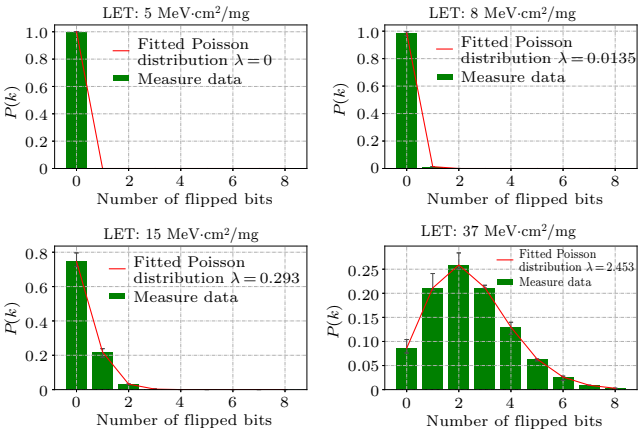


Fig. 5. Histogram of probability distribution of bit flipping under different LET values of 5, 8, 15, 37 MeV·cm^2/mg.

The probability $P(X=k)$ of bit flips occurring in SRAM during a single test cycle can be estimated using the Poisson distribution[28,29]

$$P(X=k)=\frac{\lambda^k \mathrm{e}^{-\lambda}}{k!} \quad (k=1,2,3,\ldots). \tag{4}$$

Here, λ represents the expected number of SEUs during one test cycle. Statistical analysis has shown a significant correlation between λ and the energy injected by the laser.

Images corrupted by noise can significantly affect subsequent analysis and processing in space applications. Therefore, algorithmic optimization to reduce image noise and improve resolution, without altering the system hardware, is crucial for enhancing the reliability of COTS devices. Figure 6 demonstrates the image restoration performance of the proposed binary neural network, Bin-DNCNN. When the LET is $\leq$ 15 MeV·cm^2/mg, the image data can be substantially restored to a quality comparable to the pristine image. At an LET of 37 MeV·cm^2/mg, the model effectively mitigates noise, though there is a significant loss in image resolution.

To validate the effectiveness of the model, we used the Kodak24 and CBSD100 datasets as benchmarks, as shown in Table 2. The proposed Bin-DNCNN model outperforms other binarized models but still shows a slight accuracy gap when compared to the single-precision DNCNN model. In terms of model complexity, metrics such as operations per second (OPs) and the number of trainable parameters (Params) demonstrate that the binary neural network substantially reduces storage requirements and accelerates computational speed, making it well-suited for hardware acceleration. Specifically, at a Poisson noise level of 0.293 on the CBSD100 dataset, our method achieved only a 2.68-dB reduction in PSNR compared to DNCNN. This highlights the potential of deploying binary neural networks on embedded devices.

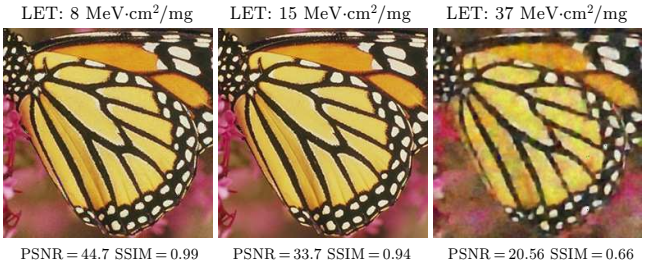


Fig. 6. The denoising performance of the Bin-DNCNN network under different LET values of 8, 15, 37 MeV·cm^2/mg.

Table 2. Quantitative comparisons (average PSNR/SSIM) of classical image denoising on a benchmark dataset.

Methods	OPs (G)	Params (K)	Kodak24			CBSD100		
			$\lambda=0.0135$	$\lambda=0.293$	$\lambda=2.453$	$\lambda=0.0135$	$\lambda=0.293$	$\lambda=2.453$
DNCNN[26]	36.717	559.363	38.42/0.979	41.03/0.990	25.52/0.700	38.30/0.983	40.39/0.992	24.52/0.679
ReActNet[22]	0.855	9.216	36.55/0.973	35.18/0.932	22.54/0.484	36.54/0.980	34.84/0.940	22.07/0.492
BiSRNet[23]	0.540	10.496	37.83/0.967	35.64/0.944	21.95/0.476	38.04/0.973	35.23/0.953	21.44/0.461
Bin-DNCNN	0.289	4.416	38.10/0.988	38.14/0.967	24.24/0.619	38.10/0.987	37.71/0.971	23.42/0.607

4.2. The impact of SEU on neural network

In this section, we investigate the impact of single-event upsets on neural network weights. Full-precision neural networks are encoded as single-precision floating-point numbers according to the IEEE 754 standard, which consists of three components: the sign bit, the exponent, and the mantissa.

$$\begin{aligned} 0.33 &= 0,01111100,10011001100110011001101 \\ 0.012561035 &= 0,01111000,10011011100110011001101 \\ 1.1229\times 10^{12} &= 0,11111000,10011011100110011001101 \end{aligned} \tag{5}$$

For instance, in Eq. (5), storing the number 0.33 in memory may result in a significant change in value due to a bit flip. Specifically, a bit flip at the red-marked position can decrease the value by more than twentyfold, while a bit flip at the green-marked position can exponentially increase the value. When a single-bit flip occurs in the mantissa, the impact on the parameter value is relatively small. However, a flip in the sign or exponent bit has a far more substantial effect. Moreover, considering that particles can induce multiple-bit flips, a flip in one position of a parameter significantly increases the likelihood of additional bit flips in the same parameter.

Figure 7(a) illustrates the variation in the noise reduction performance of the DNCNN convolutional network under different LET values. Each column of data represents the effect of error injection into the weights of the current convolutional layer on the model's noise reduction performance. For the same convolutional layer, it is evident that as the LET value increases, the model's noise reduction performance declines sharply due to an increase in single-event upset (SEU) error injection. When the LET value is low and fewer errors are injected into the model weights, the noise reduction performance varies significantly. For instance, in the first convolutional layer, the denoising effect may drop from 0.98 to 0.12. This substantial change is likely caused by bit flips in the exponent part of the weight data, which drastically alter the weight values. However, not all errors are critical to model accuracy; in most cases, the model's output can still achieve the baseline level, fully utilizing its noise reduction capabilities. This highlights the inherent robustness of convolutional neural net-

works. At an LET value of 37 MeV·cm^2/mg, the accuracy of nearly all convolutional layers falls below 0.3. In extreme cases, the errors adversely affect image noise reduction, degrading the model's output quality.

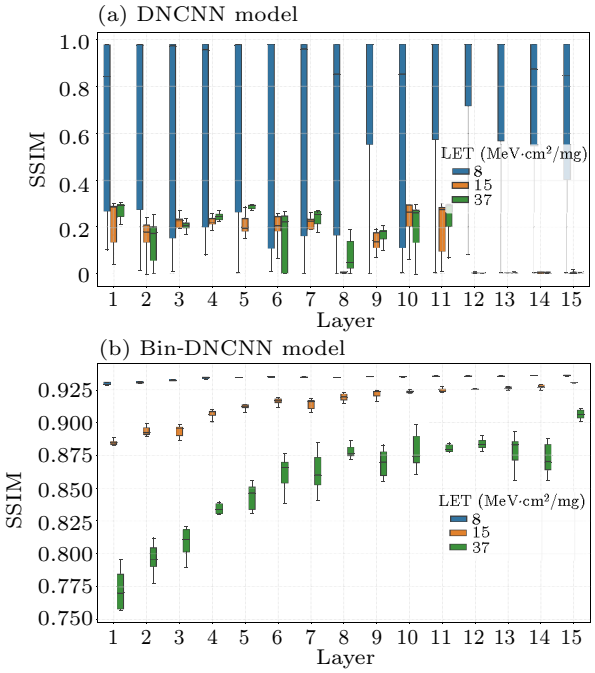


Fig. 7. SSIM accuracy degradation due to SEU affecting the weights at different layers of convolution under different LET values of 8, 15, 37 MeV·cm^2/mg.

A comparison of the layers under the same radiation conditions shows that the vulnerability of neural network layers is related to their position in the network. Specifically, the final layer is the most vulnerable, as illustrated in Fig. 7(a). Errors in SEUs occurring in later layers are more likely to affect recognition accuracy than those in earlier layers. This is because CNNs include batch normalization (BN) layers following each convolutional layer, which help pull weight data back into a normal distribution, potentially shielding errors and enhancing stability. As faults in earlier layers pass through multiple BN layers, the probability of data being shielded increases. Consequently, weight changes in later layers are more likely to affect the output, making SEU errors in these layers more detrimental to noise reduction performance.

The binary convolutional shows greater robustness to radiation, as indicated in Fig. 7(b). With the increase in LET value, the number of faults induced by laser injection also increases, and the accuracy of the denoising results is relatively reduced, following a similar trend to single-precision floating-point neural networks, but with far less decay in noise reduction performance. The significant difference in accuracy decreases depending on the affected layers. For example, When the LET reached 37 MeV·cm^2/mg, DNCNN decreased accuracy to 0.001 in the first convolutional layer, while BIN-DNCNN achieved a denoising result of 0.765 in the last convolutional layer. At an LET of 8 MeV·cm^2/mg, the denoising performance of DNCNN exhibits considerable variability, with denoising effectiveness fluctuating between 0.12 and 0.98 in the first convolutional layer. In contrast, BIN-DNCNN demonstrates strong stability with the denoising effect remaining essentially constant at 0.92.

In fault injection tests across various convolutional layers, Bin-DNCNN demonstrates a unique pattern relative to the DNCNN model, with the initial layer being the most susceptible to errors. Similar findings are reported in Ref. [30]. The weight variations in the first layer substantially affect the processing and propagation of input features to subsequent layers, which may result in increased errors or further deterioration of the model's accuracy.

5. Conclusion

This paper provides a comprehensive assessment of the reliability of neural networks under laser fault injection. Firstly, Our initial analysis focused on the impact of fault injection on data inputs and network weights, revealing that SEUs in data follow a predominantly Poisson distribution, with increased laser energy intensifying Poisson noise in the image data. Subsequently, To mitigate the effects of SEUs on image data, we employed the DNCNN model for noise reduction. However, soft errors induced extreme weight values, significantly perturbing the convolutional layers and diminishing the model's denoising effectiveness, particularly in the later layers. To further enhance neural network reliability, we introduced the Bin-DnCNN framework, a binary neural network that restricts convolutional layer weights to the binary set $\{-1, 1\}$. This approach reduces the memory footprint of the DNN and enhances the resilience of certain network parameters against soft errors. Nevertheless, soft errors have a more pronounced impact on the earlier layers, Finally, the laser experiments result indicates that binary neural networks enhance the precision and robustness of the network under irradiation environments.

Overall, this study's results significantly enhance our understanding of the effects of SEUs on neural networks designed for space applications. This research lays the foundation for developing more robust, fault-tolerant AI chips and systems, thereby ensuring the reliability and stability of neural network-based technologies in critical space missions.

References

[1] Brunetti G, Campiti G, Tagliente M and Ciminelli C 2024 *IEEE Access* **12** 76478
[2] Dario Izzo, Marcus Märtens and Pan B F 2019 *Astrodynamics* **3** 287
[3] Benoot W 2023 *European Data Handling & Data Processing Conference (EDHPC)* pp. 1–6
[4] Li J, Yang Z and Luo Y 2024 *Advances in Space Research* **75** 2184
[5] Rech P 2024 *IEEE Transactions on Nuclear Science* **71** 377
[6] Brewer R M, Moran S L, Cox J, Sierawski B D, McCurdy M W, Zhang E X, Iyer S S,Schrimpf R D, Alles M L and Reed R A 2020 *IEEE Transactions on Nuclear Science* **67** 108–115

[7] KCHAOU A, YOUSSEF W E H and TOURKI R 2017 *International Journal of Advanced Computer Science and Applications* **8**

[8] Krcma M, Kotasek Z and Lojda J 2017 *IEEE East-West Design & Test Symposium (EWDTS)* pp. 1–6

[9] Das A and Touba N A 2018 *Proceedings of the* 2018 *Great Lakes Symposium on VLSI* GLSVLSI '18 p. 219

[10] Libano F, Wilson B, Wirthlin M, Rech P and Brunhaver J 2020 *IEEE Transactions on Nuclear Science* **67** 1478

[11] Rosales Vargas S A and García-Ortiz A 2024 13*th International Conference on Modern Circuits and Systems Technologies (MOCAST)* pp. 01–05

[12] Zhao X, Du X, Xiong X, Ma C, Yang W, Zheng B and Zhou C 2024 *Chin. Phys. B* **33** 078501

[13] Eslami M, Ghavami B, Raji M and Mahani A 2020 *Integration* **71** 154

[14] Condia J E R, Guerrero-Balaguera J D, Dos Santos F F, Reorda M S and Rech P 2022 *IEEE International Test Conference (ITC)* pp. 278–287

[15] Salami B, Onural E B, Yuksel I E, Koc F, Ergin O, Cristal Kestelman A, Unsal O, Sarbazi-Azad H and Mutlu O 2020 50*th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)* pp. 138–149

[16] Liu W, Chang C H, Zhang F and Lou X 2020 57*th ACM/IEEE Design Automation Conference (DAC)* pp. 1–6

[17] Selmke B, Brummer S, Heyszl J and Sigl G 2016 *Smart Card Research and Advanced Applications*, eds. Homma N and Medwed M pp. 193–205

[18] Hou X, Breier J, Jap D, Ma L, Bhasin S and Liu Y 2020 *IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)* pp. 1–6

[19] Yuan C and Agaian S S 2023 *Artif. Intell. Rev.* **56** 12949

[20] Hubara I, Courbariaux M, Soudry D, El-Yaniv R and Bengio Y 2016 *Proceedings of the* 30*th International Conference on Neural Information Processing Systems* NIPS'16 p. 4114

[21] Rastegari M, Ordonez V, Redmon J and Farhadi A 2016 *Computer Vision – ECCV* 2016 pp 525–542

[22] Liu Z, Shen Z, Savvides M and Cheng K T 2020 *Computer Vision – ECCV* 2020 pp. 143–159

[23] Cai Y, Zheng Y, Lin J, Yuan X, Zhang Y and Wang H 2023 arXiv: 2305.10299v3 [cs.CV]

[24] Huang J and Han J 2005 *Sci. China- Phys., Mech. Astron.* **48** 113

[25] Zhang H, Liu Z, Xie Y, Jia Y and Zhang Z 2021 *IEICE Electronics Express* **18** 20210404

[26] Zhang K, Zuo W, Chen Y, Meng D and Zhang L 2017 *IEEE Transactions on Image Processing* **26** 3142

[27] Horé A and Ziou D 2010 2010 20*th International Conference on Pattern Recognition* pp. 2366–2369

[28] Schirmeier H, Borchert C and Spinczyk O 2015 45*th Annual IEEE/IFIP International Conference on Dependable Systems and Networks* pp. 319–330

[29] Deng Q, Shao C, Li H and Fang J 2019 *IEEE* 13*th International Conference on Anti-counterfeiting, Security, and Identification (ASID)* pp. 314–319

[30] Khoshavi N, Broyles C and Bi Y 2020 21*st International Symposium on Quality Electronic Design (ISQED)* pp. 99–104