Evaluation of Soft Error Resilience in Floating-Point Units

Thesis advisor: David DEFOUR, Professor in Computer Science

Location: University of Perpignan Via Domitia, France

Keywords: soft error, gpu, floating-point computation

---

1. Introduction

In recent years, the increasing demand for high-performance computing (HPC) and artificial intelligence (AI) workloads has driven the adoption of Graphics Processing Units (GPUs) across various industries. Among the critical components of a GPU, the Floating-Point Unit (FPU) is responsible for executing complex numerical computations. As GPUs scale in power and complexity, they are more prone to **soft errors**, which are transient faults caused by high-energy particles, electromagnetic interference, or thermal fluctuations.

Soft errors can compromise the accuracy of computations, leading to incorrect results, system crashes, or compromised data integrity. Given the widespread usage of NVIDIA GPUs in mission-critical applications like autonomous vehicles, healthcare simulations, and financial systems, it is crucial to evaluate and improve the soft error resilience of their floating-point units. This master thesis aims to **evaluate the soft error resilience of floating-point units targeting specifically NVIDIA GPUs** and explore methodologies for improving their robustness.

2. Research Objectives

The main objectives of this thesis are:

1. **Investigate the vulnerability of floating-point units (FPUs) to soft errors**
    - Quantify the rate and types of soft errors in different various GPU architectures and extend it to CPU as well.
    - Examine how floating-point precision (single, double, mixed) influences error resilience.
2. **Evaluate existing error mitigation techniques**
    - Analyze NVIDIA's error correction mechanisms (ECC) and other built-in error detection methods to mitigate soft errors.
3. **Develop a framework for benchmarking the soft error resilience of FPUs**
    - Design a testbed to evaluate error resilience in controlled conditions using synthetic and real workloads.
    - Establish a set of metrics for benchmarking error resilience.

3. Literature Review

Soft errors have been widely studied in the context of main memory and traditional Central Processing Units (CPUs). However, GPUs, with their highly parallel architecture, introduce unique challenges when it comes to error resilience, particularly in floating-point computations. Recent studies have explored soft error propagation in memory, but limited research focuses specifically on FPUs.

1. **Soft Error Mechanisms**
    - Overview of soft error causes (cosmic rays, thermal variations).
    - The difference between permanent faults (hardware defects) and transient soft errors.
2. **Error Resilience in FPUs**
    - Soft error resilience studies and how these approaches can be extended to GPUs.
    - Previous work on hardware redundancy, software-based error detection, and correction codes.
3. **NVIDIA GPU Architecture**
    - Analysis of NVIDIA's GPU architecture, focusing on its FPUs, memory hierarchy, and parallelism.
    - Review of existing fault tolerance mechanisms in NVIDIA GPUs, such as ECC for memory and instruction replay for execution units.

A curated list of recent research papers, technical documentation from NVIDIA, and prior studies on fault tolerance in computing systems:

- How SHAKTI-F (RISC-V based) processor tackled soft and hard errors (SEEs)?: https://www.linkedin.com/pulse/how-shakti-f-risc-v-based-processor-tackled-soft-hard-abhishek-jadhav/
- Transient Error Resilient Hessenberg Reduction on GPU-based Hybrid Architectures https://www.netlib.org/lapack/lawnspdf/lawn279.pdf
- Transient Error Analysis, https://apps.dtic.mil/sti/tr/pdf/ADA155395.pdf
- R. Baumann. Radiation-induced soft errors in advanced semiconductor technologies. Device and Materials Reliability, IEEE Transactions on, 5(3):305–316, 2005
- S. Michalak, K. Harris, N. Hengartner, B. Takala, and S. Wender. Predicting the number of fatal soft errors in los alamos national laboratory's asc q supercomputer. Device and Materials Reliability, IEEE Transactions on, 5(3):329–335, 2005.
- D. Defour and E. Petit, GPUburn: A system to test and mitigate GPU hardware failures," 2013 International Conference on Embedded Computer Systems: Architectures, Modeling, and Simulation (SAMOS), Agios Konstantinos,

4. Methodology

1. **Experimental Setup**

- Selection of target NVIDIA GPUs (from architectures such as Pascal, Volta, Ampere) and CPUs for comparison purposes
- Creation of synthetic benchmarks and real-world workloads (AI, scientific computing) to stress the various FPUs (adder, multiplier, SFU, tensor unit)
- Introduction of simulated soft errors using fault injection techniques.

2. **Error Injection and Detection**
    - Inject soft errors in floating-point registers and arithmetic units to observe their effects on computation.
    - Utilize existing tools (e.g., NVBitFI, GPGPU-Sim) for fault injection at the hardware level.
3. **Analysis and Metrics**
    - Measurement of error rates, fault propagation, and system-level impacts (e.g., performance degradation, crashes).
    - Evaluation of how different precision levels (e.g., FP32, FP64) respond to soft errors.
4. **Proposed Enhancements**
    - Develop and implement new error correction schemes ()
    - Measure and compare the performance, power consumption, and resilience of the new methodologies against existing ones.

---

5. Expected Contributions

1. A detailed analysis of the vulnerability of FPUs to soft errors.
2. A comparative evaluation of existing error resilience techniques.
3. Novel techniques to enhance the resilience of FPUs to soft errors, validated through simulation and testing.
4. A framework for benchmarking soft error resilience in FPUs that can be used by future researchers and developers.