

## Анализ программных методов повышения надёжности аппаратных SNN-ускорителей

Зиновичев Е. С. (Университет ИТМО)

Научный руководитель – к.т.н., Быковский С. В.

(Университет ИТМО)

**Введение.** SNN-ускорители обеспечивают энергоэффективное выполнение задач искусственного интеллекта [1]. Однако их аппаратные реализации подвержены различным видам отказов – «мягким» ошибкам (битовые перевороты из-за космических частиц) и «жёстким» дефектам (дефекты производства, старение, перегрев) [2], [3]. В условиях высоких требований к надёжности актуален поиск эффективных методов обеспечения отказоустойчивости SNN-платформ на всех уровнях – от аппаратного обеспечения до прошивки, драйверов ОС и инструментального ПО. Цель исследования – анализ существующих подходов к повышению надёжности SNN-ускорителей и выявление их ограничений в части системной и платформенной поддержки.

**Основная часть.** В рамках проведённого анализа были рассмотрены основные подходы к обеспечению надёжности ускорителей SNN, с акцентом на платформенные компоненты: прошивку, драйверы ОС, инструменты диагностики и поддержки отказоустойчивости. Проанализированы методы инъекции ошибок (Fault Injection) [1], механизмы мониторинга состояния, RAS-архитектуры (Reliability, Availability, Serviceability) [5], а также программные средства восстановления и реагирования на отказы.

Особое внимание уделено тому, как данные решения реализуются в существующих ускорителях, и какие из них доступны на уровне платформенного стека. Установлено, что большинство инструментов Fault Injection, таких как SpikeFI, функционируют на уровне симуляторов и не интегрированы в реальные системы или SDK ускорителей [1], [4]. Средства диагностики, как правило, ограничиваются тестированием памяти и регистров при инициализации, без возможности динамического наблюдения за отказами в процессе работы [5].

В работе также проанализированы типовые архитектурные сценарии: поддержка ECC в памяти, регистрация ошибок в аппаратных логах (например, MCE), а также реакция прошивок и драйверов ОС на сбои [5], [6]. Несмотря на то, что такие механизмы активно применяются в серверных системах, в случае SNN-ускорителей они реализованы частично либо отсутствуют вовсе.

По результатам анализа выделены ключевые пробелы в области: отсутствие стандартных API для доступа к телеметрии отказов, слабая связность между слоями (аппаратным и программным), а также недостаток средств для диагностики в ходе выполнения и адаптации вычислений в условиях отказов. Это подчёркивает необходимость разработки комплексного подхода, объединяющего аппаратные средства защиты с доступными платформенными механизмами управления отказами [2], [3], [7].

**Выводы.** Проведённый анализ показывает, что надёжность SNN-ускорителей зависит от совокупности аппаратных и программных мер: от встроенных RAS-модулей и ECC до методов тестирования и адаптивных алгоритмов. Сочетание активного мониторинга с динамической адаптацией SNN-модели к моделям отказов и программной логикой управления позволит добиться высокой устойчивости. Полученные результаты подчёркивают важность объединения независимых техник повышения надёжности в единую платформенную архитектуру. Это открывает пути для дальнейших исследований – например, внедрения рассматриваемых методов в реальные программные платформы ускорителей и оценки их эффективности на прототипах SNN-чипов, а также применения результатов при проектировании новых нейроморфных систем.

**Список использованных источников:**

1. T. Spyrou, S. Hamdioui, H.-G. Stratigopoulos. SpikeFI: A Fault Injection Framework for Spiking Neural Networks. arXiv:2412.06795 (2024).
2. E.-I. Vatajelu et al. Reliability of Hardware-Implemented Spiking Neural Networks (SNN). 2019 IEEE VLSI Test Symposium (VTS).
3. C. Zhang et al. Thermal-Aware Layout Optimization and Mapping Methods for Resistive Neuromorphic Engines. ASP-DAC, 2022.
4. Y. Kim et al. Functional Fault Analysis in Spiking Neural Networks. IOLTS 2023.
5. J. E. Oster. Reliability, Availability, and Serviceability (RAS). Intel Developer Zone, 2020.
6. M. Sabbar et al. Input-Aware Resilience Optimization of AI Accelerators. IEEE Transactions on Circuits and Systems, 2022.
7. R. V. W. Putra et al. RescueSNN: Enabling Reliable Executions on Spiking Neural Network Accelerators Under Permanent Faults. IEEE Transactions on Computers, 2023.