Fast and Accurate Prediction of the Impact of Approximate Operators on a Complex Computation

Sepide SAEEDI

PhD Candidate: Email: sepide.saeedi@polito.it
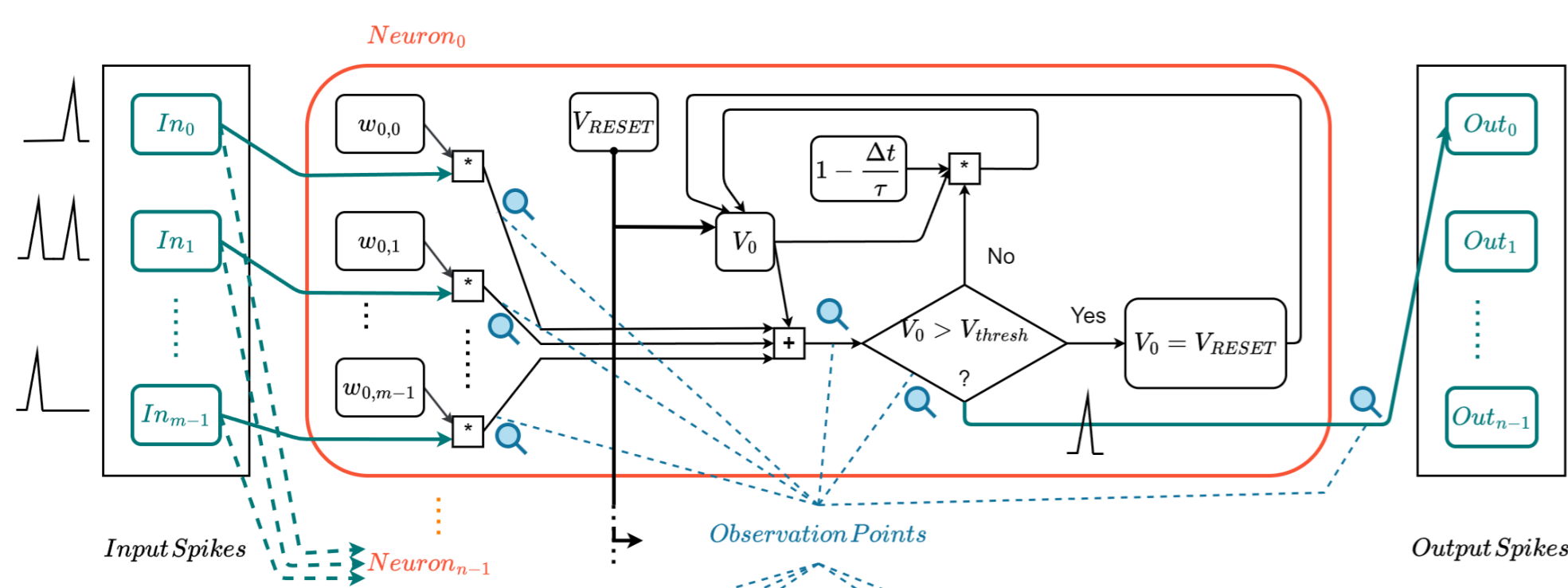
1. Introduction

Approximate Computing (AxC) techniques have become increasingly popular in trading off accuracy for gains in power, area, execution time, etc. Finding the most suitable AxC techniques to reach the best trade-off between accuracy degradation and power consumption/computation time is challenging.
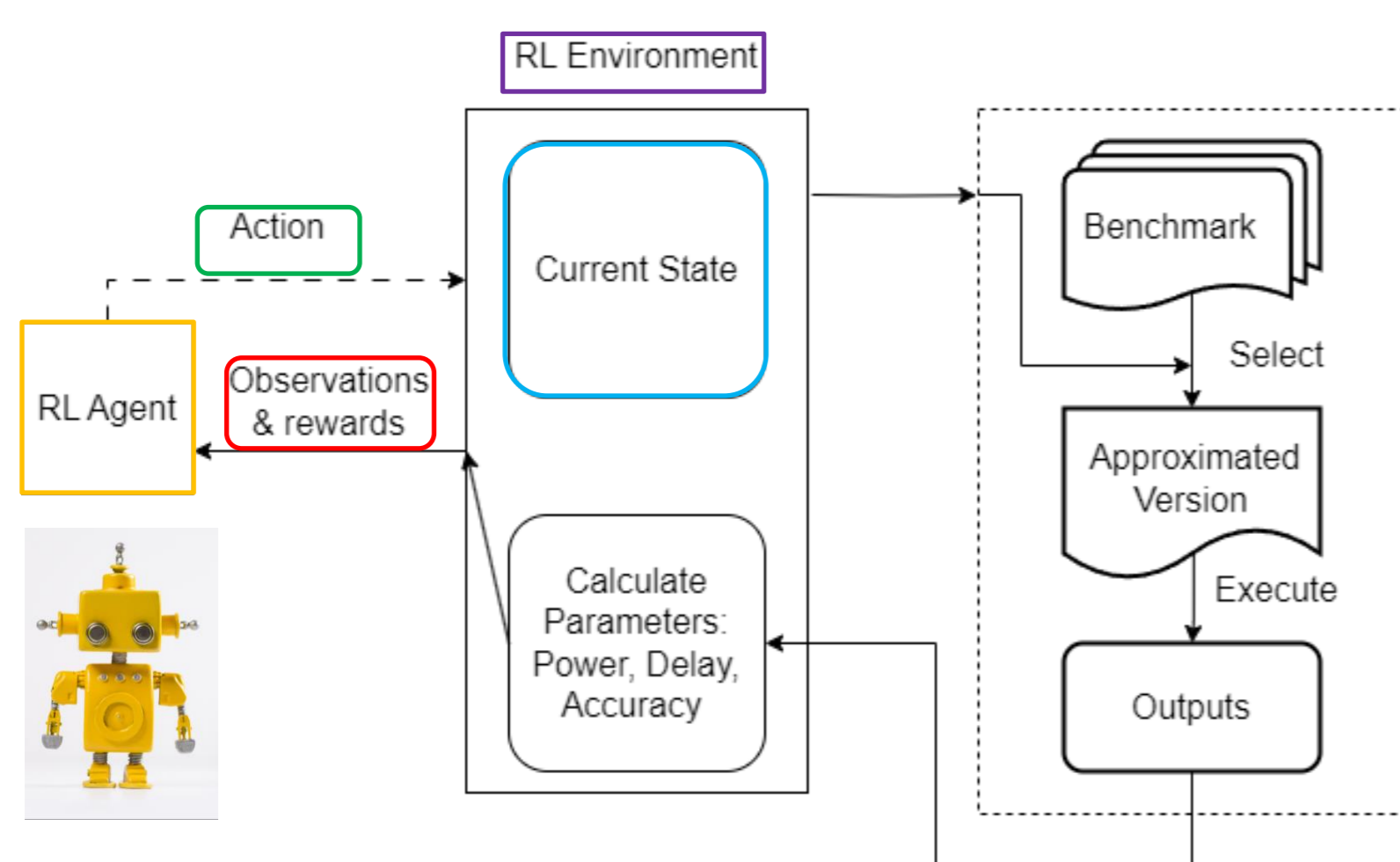
2. Goal / Objectives

A Design Space Exploration (DSE) approach can help systematically evaluate all approximate versions of an application. Performing the DSE in a reasonable time can become infeasible. This research is focused on making this DSE automatic and fast.

3. Methodology
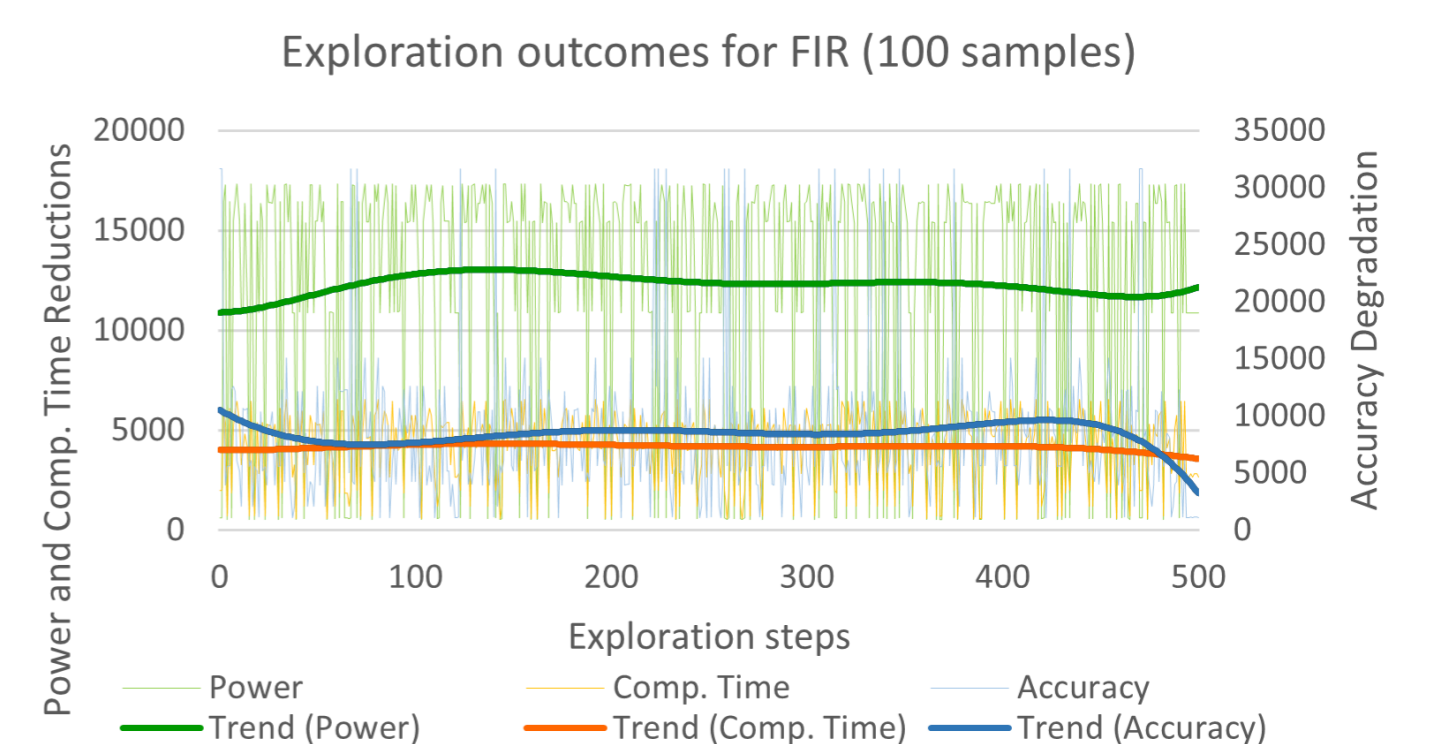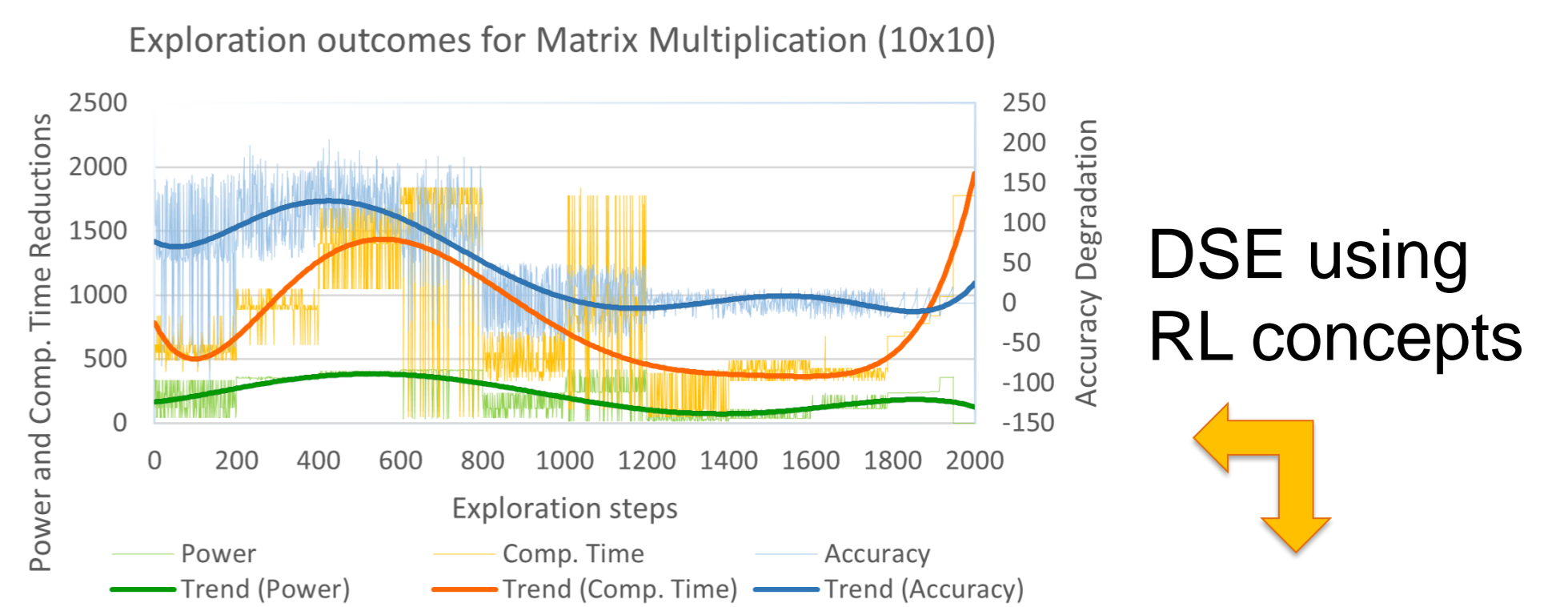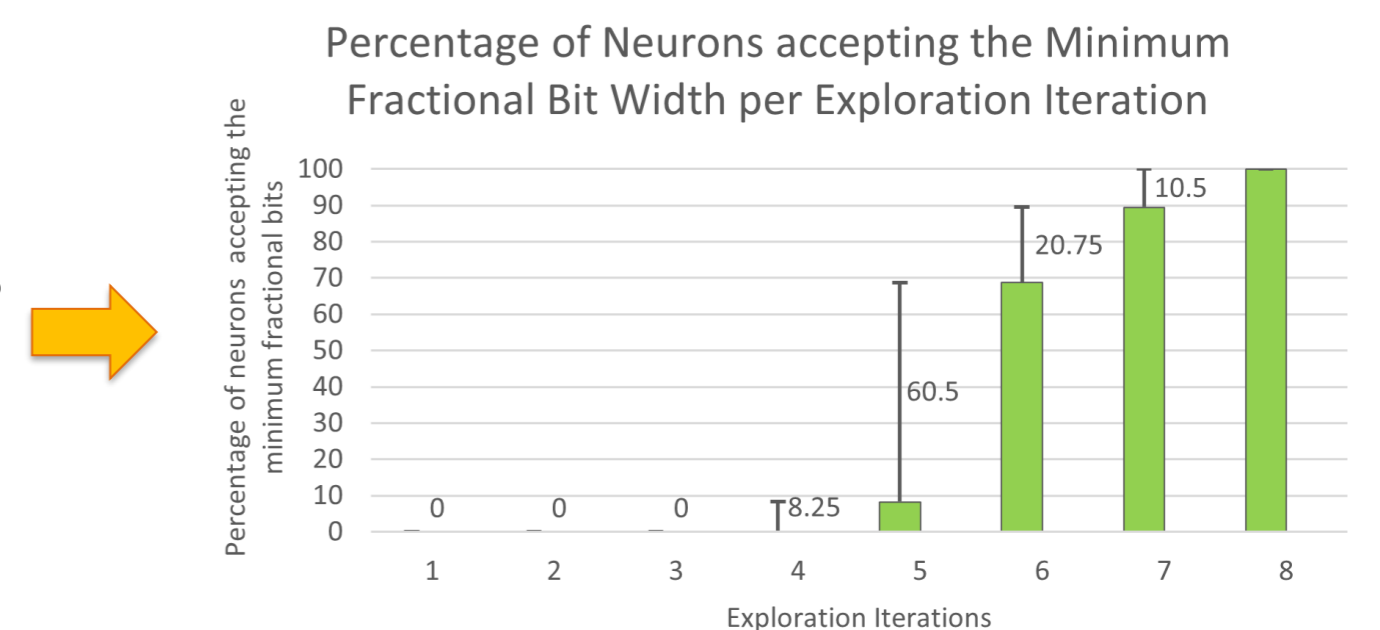
- DSE using Interval Arithmetic (IA) concepts:

- DSE using Reinforcement Learning (RL) concepts:

4. Results

DSE using IA concepts →

Percentage of Neurons accepting the Minimum Fractional Bit Width per Exploration Iteration

Exploration outcomes for Matrix Multiplication (10x10)

DSE using RL concepts

Exploration outcomes for FIR (100 samples)

5. Conclusions

- Two approaches were proposed to reduce the time of a DSE to find the most suitable AxC techniques for an application.

- Both approaches were proven effective by experimental results in most cases.

- Future Works:
  - Using the IA-based approach to compare other parameters in addition to accuracy.
  - Modifying the learning algorithm of the RL-based approach for each application.

6. References

1. S. Saeedi, A. Carpegna, A. Savino and S. Di Carlo, "Prediction of the Impact of Approximate Computing on Spiking Neural Networks via Interval Arithmetic," 2022 IEEE 23rd Latin American Test Symposium (LATS), 2022.

2. S. Saeedi, A. Carpegna, A. Savino, and S. Di Carlo, "Fast Exploration of the Impact of Precision Reduction on Spiking Neural Networks," arXiv preprint arXiv:2212.11782, 2022.

3. S. Saeedi, A. Savino and S. Di Carlo, "Design Space Exploration of Approximate Computing Techniques with a Reinforcement Learning Approach," 2023 53rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W), 2023.