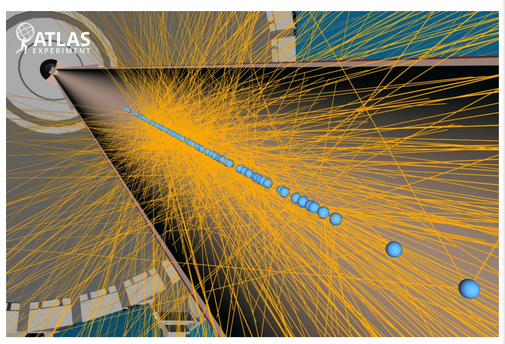
Analog-Domain Implementation of Neural Networks for Energy-Efficient High Energy Physics Applications

Author: Dewen Zhong[1],Mark Neubauer[1]
Institutions: [1]*University of Illinois at Urbana-Champaign*

Computing Challenge @ HL-LHC

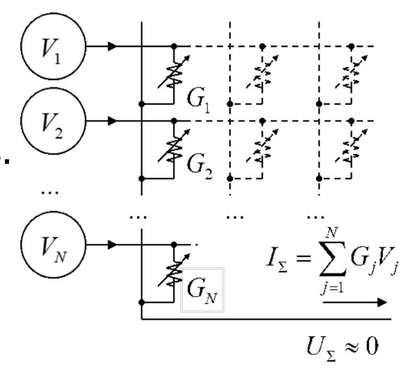
Efficient computational strategies are paramount for devices in resource-limited settings, particularly within high-energy physics experiments.
At HL-LHC, 10x more data per second than Run 1 & 2.



Analog AI

Application of analog computing techniques to perform artificial intelligence (AI) tasks.

- Analog computing operates on continuous physical quantities, such as voltages or currents.
- Inherent parallelism and efficiency of analog computations.
- More energy-efficient, Better Latency to accelerate AI algorithm.

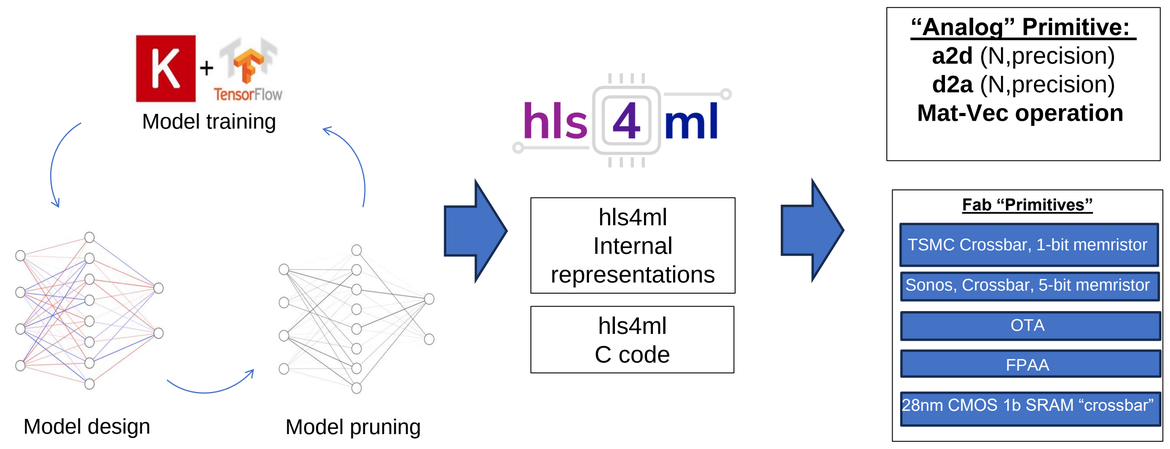


HLS4ML

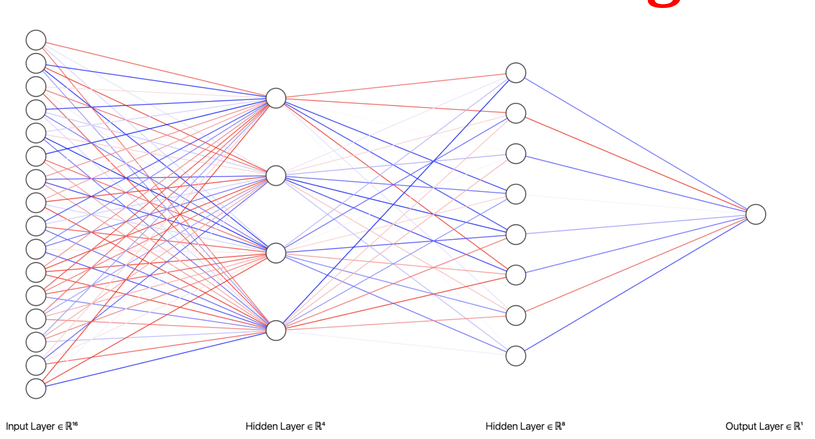
hls4ml[1] is a firmware implementations of machine learning algorithms using high level synthesis language (HLS) in FPGAs with ultra low latency.

- Simplify the hardware implementation process.
- Support for popular machine learning libraries.
- Compatibility with diverse hardware platforms such as FPGAs and ASICs.
- Deploy models in real-time and embedded applications.

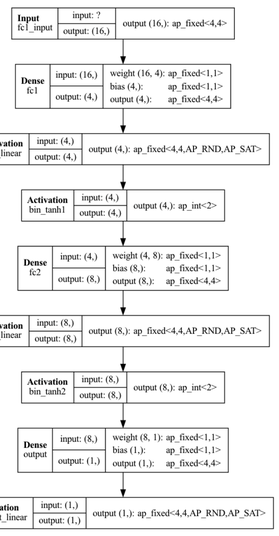
Analog Path



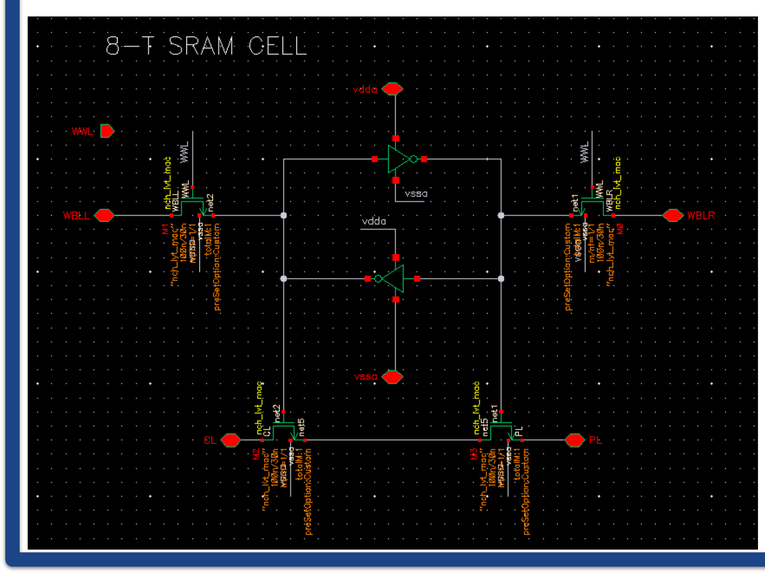
Machine Learning architectures



Custom ANN model



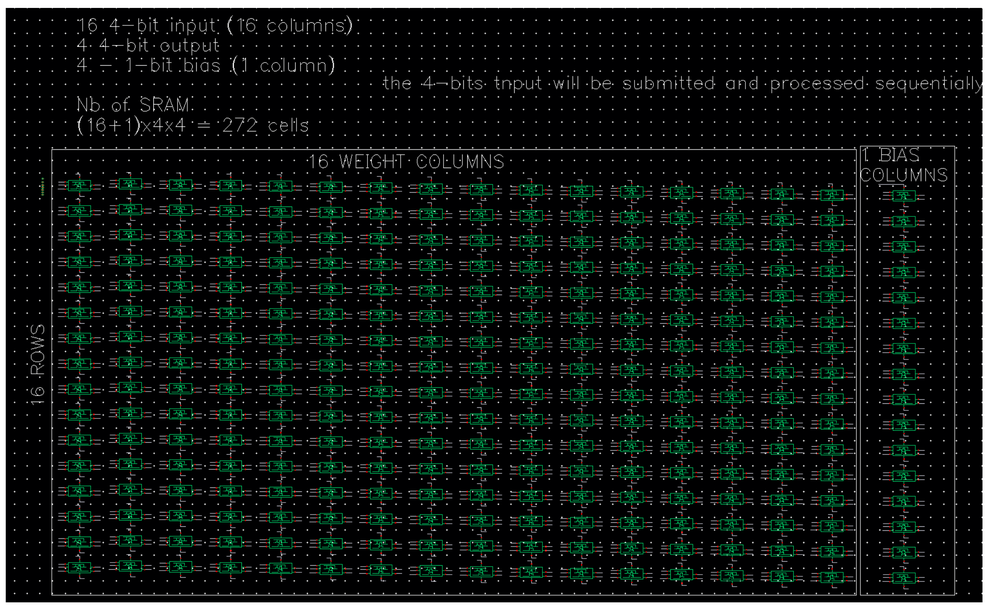
SRAM-UNIT Cell



Based off design

- Mf-net: Compute-in-memory sram for multibit precision inference using memory-immersed data conversion and multiplication-free operators
- MC-CIM: Compute-in-Memory With Monte-Carlo Dropouts for Bayesian Edge Intelligence

First Layer Analog Representation



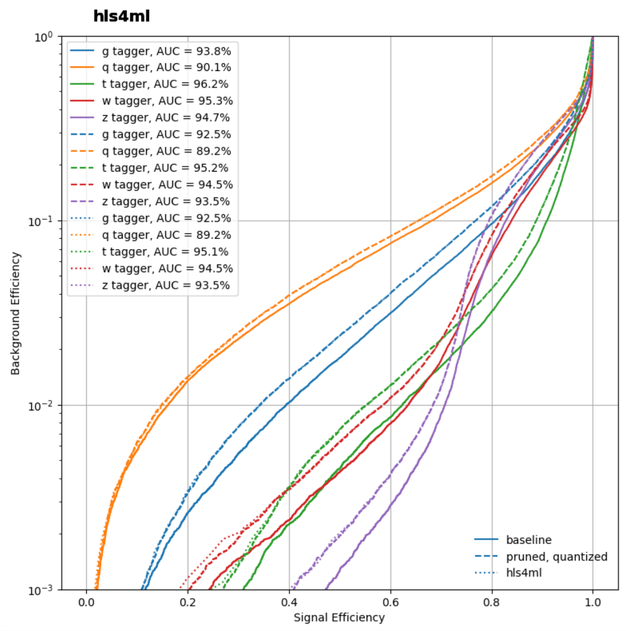
Basic Matrix for fc1 layer:

- 16 4-bits input (16columns)
- 4 4-bits output
- 4 1-bit bias (1 bias column)

All production lines charged are accumulated and summed horizontally bit by bits.

Each bit needs to be binary weight.

Expectation and Goal



- Using hls4ml to generate analog AI models and implement analog models to ML-based jet Tagger.
- Provide similar performances as digital hls4ml ML models.

Conclusion

- Investigate the analog path to produce analog AI model from custom ANN model via hls4ml
- Simulate the first layer analog representation for ANN model.
  Further Work
- Complete the full analog representation for entire ANN model structure.
- Tunning and Optimize the analog AI model performance for High Energy Physic models and beyond .

References & Acknowledges

This project is funded by NSF OAC-2117997

[1] hls4ml: An Open-Source Codesign Workflow to Empower Scientific Low-Power Machine Learning Devices F. Fahim, B. Hawks, +27 authors Zhenbin Wu ,Computer Science, arXiv:2103.05579 2021

National Science Foundation