Neural network algorithm for HL-LHC ATLAS hardware τ trigger

Yuval Frid, Maayan Tamari, Boping Chen, Liron Barak
Tel Aviv University
boping.chen@cern.ch

1. Introduction

- The High-Luminosity Large Hadron Collider (HL-LHC) is planned to start to collect data in 2029 and achieve 4000 fb^{-1} at ATLAS detector, with an upgraded new global trigger subsystem to be added into the hardware trigger.
- Due to the much higher pile-up condition, the performance of the current hardware trigger algorithm will be largely degraded, especially for the low energy region.

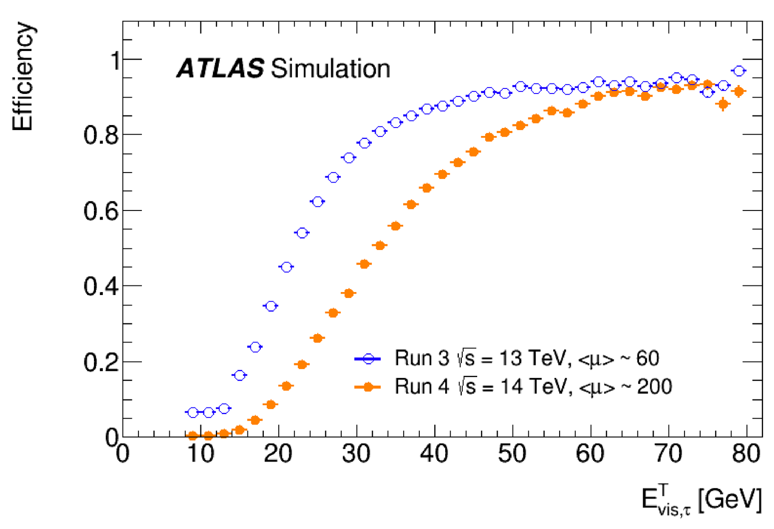

Figure 1: Current τ trigger performance with Run 3 sample (blue) and Run 4 (HL-LHC) sample (orange)

- Since more advanced Field Programmable Gate Arrays (FPGAs) will be implemented, it is possible to run machine learning and even neural networks (NN) at the hardware trigger level.
- Several NN structures, including CNN/DNN, are tested, and the software high level synthesis for machine learning (HLS4ML) is used to create firmware implementation of the NN.
- Compared with the current ATLAS default τ trigger algorithm, the NN can significantly improve the performance of the low energy regime.

2. NN input and structure

- Signal: Standard Model $Z \to \tau\tau$. Background: QCD dijet events.
- Using cell information from EM and Had Calorimeter.
- Seeding with local maximum from upstream, using all the cells within 0.3 x 0.3 in $\eta \times \phi$ around the seed to build the input.
- Different calorimeter layers have different granularity. Rebin or split the cell to get the same granularity for all the layers.

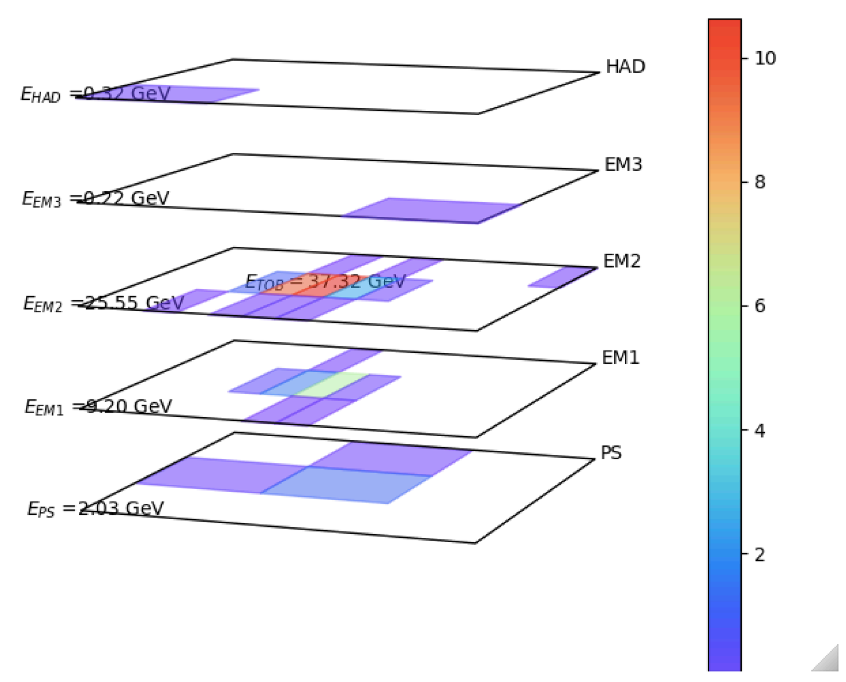

Figure 2: Example of signal input

- Convolutional neural network (CNN) is used for the NN study as a common image identification method.
- The input is treated as a 3x12 image with 5 channels.
- CNN consists of three 2D convolutional layers with kernel size 3x3, two 2D MaxPool layers, and one linear layer to flatten the tensor to a single score.
- Output is one single score between 1 (as signal) and 0 (as background).
- The CNN consists of 7 layers with 12821 parameters
- The threshold is calculated to achieve the background fake rate for the di-τ trigger.

3. NN performance

- Figure below shows signal efficiency vs transverse energy of visible τ.
- Orange is for the current ATLAS default τ trigger (Run 3) algorithm. Blue is the using the CNN score. Green is the using the CNN score multiplied by the total transverse energy of the input image.
- The raw CNN performance (blue) is much better than the current Run 3 algorithm in the low energy region, but worse in the high energy region. After reweighting with the transverse energy (green), the reweighted CNN performance is better than the Run 3 algorithm for the whole range.

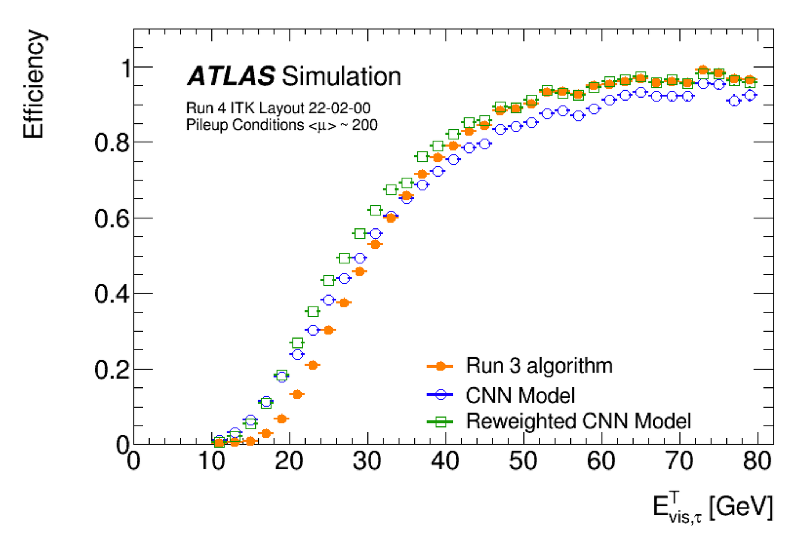

Figure 3: Example of signal input

4. HLS4ML and performance

- HLS4ML is a Python package for machine learning inference in FPGAs.
- It can be used to create firmware implementations of machine learning algorithms using high level synthesis language (HLS), by translating traditional open-source machine learning package models into HLS for microsecond-scale latency on predictions.
- It supports different models, including DNN, RNN and CNN.

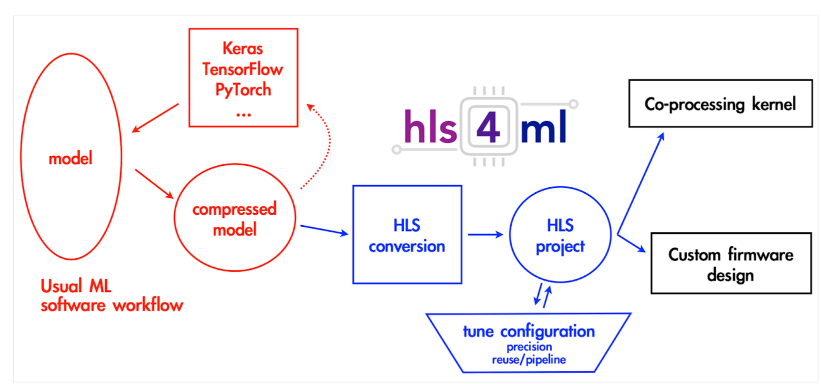

Figure 4: Example of signal input

- Implementation is done on the xcvu9P FPGA.
- For the latency strategy: 1.05% FF, 4.95% LUT, and 32.82% DSP slices, with an estimated latency of 104 clock cycles (520 ns).
- For the resources strategy: 0.52% FF, 1.61% LUT, and 20.82% DSP slices, with an estimated latency of 548-561 clock cycles (2.7 ms).

5. Summary

- The ALTAS trigger system will be upgraded for the HL-LHC, with a new trigger system and more advanced FPGAs.
- Current τ trigger algorithm will be largely degraded due to the much higher pile-up condition for the HL-LHC.
- It is possible to implement the NN to have a better trigger performance.

References

[1] FastML Team. fastmachinelearning/hls4ml, 2023.
[2] Yuval Frid. Neural Network Algorithms For HL-LHC Tau Trigger In ATLAS, 2021.