

Designing intelligent DAQ systems for radiation instrumentation with hls4ml

Monday, 3 October 2022 13:45 (15 minutes)

As detector technologies improve, the increase in resolution, number of channels and overall size create immense bandwidth challenges for the data acquisition system, long data center compute times and growing data storage costs. Much of the raw data does not contain useful information and can be significantly reduced with veto and compression systems as well as online analysis.

The improvements in artificial intelligence, particularly the many flavours of machine learning, adds a powerful tool to data acquisition strategies by providing embedded analysis, reducing the data at the source. However large and deep neural network are still compute intensive and one of the most important challenges ML designers face is minimizing the model size without losing the precision and accuracy required for a scientific application. The combination of signal processing algorithms and compression algorithms with machine learning can improve the latency and accuracy of edge systems by reducing the width and depth of the model. Using this strategy, we develop hardware compatible signal processing and machine learning systems for various radiation detectors. We will present two current use cases:

1) The CookieBox, an attosecond angular streaking detector used for X-ray pulse shape recovery generating ~ 800 GB/s. This system requires microsecond latency to apply a veto on downstream detectors. We designed both a fully connected neural network and a convolutional neural network (CNN) each combined with a non-uniform data quantizer. These networks achieve 86 % accuracy in $8 \mu\text{s}$ for the fully connected network and 88 % accuracy in $23 \mu\text{s}$ for the CNN (including data transfer times) on a ZYNQ XC7Z02.

2) The billion pixel X-ray camera for use in synchrotrons, XFEL facilities and pulsed power facilities generating up to 15 TB/s. Data is compressed using a trained neural network to accelerate the ISTA algorithm (Learned ISTA) and combined with a DEFLATE compression to achieve 83:1 compression. The network compressed each 6×6 pixel patch in less than $2 \mu\text{s}$ when implemented on a ZYNQ XC7Z02.

For both systems complementary, additional hardware modules were designed and integrated with hls4ml to implement the complete data analysis on FPGA. Several other ongoing projects are currently benefitting from the methods developed with these systems, including medical imaging and dark matter search.

Primary author: Prof. CORBEIL THERRIEN, Audrey (Université de Sherbrooke)

Co-authors: Mr GOUIN-FERLAND, Berthié (Université de Sherbrooke); Mr GRANGER, Charles-Étienne (Université de Sherbrooke); Mr EZZAOUI RAHALI, Hamza (Université de Sherbrooke); Mr RAHIMIFAR, Mohammad Mehdi (Université de Sherbrooke)

Presenter: Prof. CORBEIL THERRIEN, Audrey (Université de Sherbrooke)

Session Classification: Contributed Talks