

Machine Learning for Real-Time Processing of ATLAS Liquid Argon Calorimeter Signals with FPGAs

Thursday 12 January 2023 17:40 (20 minutes)

The ATLAS Liquid Argon (LAr) calorimeter measures the energy of particles produced in LHC collisions. This calorimeter has also trigger capabilities to identify interesting events. In order to enhance the ATLAS detector physics discovery potential, in the blurred environment created by the pileup, an excellent resolution of the deposited energy and an accurate detection of the deposited time is crucial.

The computation of the deposited energy is performed in real-time using dedicated data acquisition electronic boards based on FPGAs. FPGAs are chosen for their capacity to treat large amount of data with very low latency. The computation of the deposited energy is currently done using optimal filtering algorithms that assume a nominal pulse shape of the electronic signal. These filter algorithms are adapted to the ideal situation with very limited pileup and no overlap of the electronic pulses in the detector. However, with the increased luminosity and pileup, the performance of the optimal filter algorithms decreases significantly and no further extension nor tuning of these algorithms could recover the lost performance.

The back-end electronic boards for the Phase-II upgrade of the LAr calorimeter will use the next high-end generation of INTEL FPGAs with increased processing power and memory. This is a unique opportunity to develop the necessary tools, enabling the use of more complex algorithms on these boards. We developed several neural networks (NNs) with significant performance improvements with respect to the optimal filtering algorithms. The main challenge is to efficiently implement these NNs into the dedicated data acquisition electronics. Special effort was dedicated to minimising the needed computational power while optimising the NNs architectures.

Five NN algorithms based on CNN, RNN, and LSTM architectures will be presented. The improvement of the energy resolution and the accuracy on the deposited time compared to the legacy filter algorithms, especially for overlapping pulses, will be discussed. The implementation of these networks in firmware will be shown. Two implementation categories in VHDL and Quartus HLS code are considered. The implementation results on Stratix 10 INTEL FPGAs, including the resource usage, the latency, and operation frequency will be reported. Approximations in the firmware implementations, including the use of fixed-point precision arithmetic and lookup tables for activation functions, will be discussed. Implementations including time multiplexing to reduce resource usage will be presented. We will show that two of these NNs implementations are viable solutions that fit the stringent data processing requirements on the latency ($O(100\text{ns})$) and bandwidth ($O(1\text{Tb/s})$ per FPGA) needed for the ATLAS detector operation.

Presenter: LAATU, Lauri Antti Olavi (Centre National de la Recherche Scientifique (FR))

Session Classification: Parallel Session H