PROCEEDINGS OF SCIENCE

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

Lauri Laatu[a,*] **on behalf of the ATLAS Liquid Argon Calorimeter group**

[a] *CPPM, Aix-Marseille Université, CNRS/IN2P3 (FR),*
163 Avenue de Luminy 13009, Marseille, France

E-mail: laatu@cppm.in2p3.fr

Within the Phase-II upgrade of the LHC, the readout electronics of the ATLAS Liquid Argon (LAr) Calorimeters are prepared for high luminosity operation expecting a pile-up of up to 200 simultaneous proton-proton interactions. Moreover, the calorimeter signals of up to 25 subsequent collisions are overlapping, which increases the difficulty of energy reconstruction. Real-time processing of digitized pulses sampled at 40 MHz is performed using FPGAs. To cope with the signal pile-up, new machine learning approaches are explored: convolutional and recurrent neural networks outperform the optimal signal filter currently used, both in assignment of the reconstructed energy to the correct bunch crossing and in energy resolution. Very good agreement between neural network implementations in FPGA and software based calculations is observed. The FPGA resource usage, the latency and the operation frequency are analyzed. Latest performance results and experience with prototype implementations are analyzed and are found to fit the requirements for the Phase-II upgrade.

The European Physical Society Conference on High Energy Physics (EPS-HEP2021)
26-30 July 2021
Online conference jointly organized by Universität Hamburg and the research center DESY

[*] Speaker

PoS(EPS-HEP2021)752

1. Introduction

The ATLAS detector [1] is one of the general purpose detectors at the Large Hadron Collider [2] (LHC) measuring the properties of the particles produced in high-energy proton-proton collisions that happen every 25 ns (40 MHz). In the Run-4 of the LHC starting in 2027, also known as the high-luminosity phase of LHC (HL-LHC), the machine is expected to produce instantaneous luminosities of 5–7 × $10^{34}\,\text{cm}^{-2}\text{s}^{-1}$ corresponding to 140–200 simultaneous proton-proton interactions. The liquid-argon (LAr) calorimeters of ATLAS mainly measure the energy of electromagnetic showers of photons, electrons and positrons using their ionization signal that causes an electronic pulse. This pulse is shaped to a bi-polar shape that takes up to 25 bunch crossings (BCs) shown in Figure 1 which can lead to out-of-time pileup by overlapping with previous pulses in the HL-LHC conditions.

A new energy reconstruction method capable of continuous energy measurement and selection of subsequent collision events is required for the LAr calorimeter for the HL-LHC era [3]. The energy reconstruction is done by field-programmable gate arrays (FPGAs) for the 182 000 calorimeter cells with 384 or 512 LAr calorimeter cells per one Intel Stratix-10 FPGA [4] with a latency requirement of about 150 ns [3, 5]. For this small Artificial Neural Networks (ANNs) based on Convolutional [6] and Recurrent Neural Networks [7] (CNNs, RNNs) are developed to replace the current optimal filtering [8] (OF) algorithm.

2. Energy Reconstruction

The current energy reconstruction method is based on the optimal filter method. A peak finder is used to assign energies to the correct BCs. The two shortcomings of this method for the use in the HL-LHC are that the optimal filter assumes a perfect pulse shape which leads to degraded performance when the pulse is distorted by previous events as well as peak finder failing to assign the energy to the correct BC. ANNs can use information from previous collisions to mitigate the effect of distorted pulse shape and to assign the energy to the correct BC.

The ANNs are trained using simulated data of HL-LHC conditions created with AREUS [9] containing electronics noise and low-energy deposits from particles produced in inelastic proton-proton collisions of up to 1 GeV combined with higher energy deposits up to 5 GeV injected randomly with a mean interval of 30BCs. The simulation is of a LAr calorimeter cell in the middle layer of the barrel (labelled EMB middle) ($\eta = 0.5125$, $\phi = 0.0125$) with an average pileup $\langle \mu \rangle = 140$.

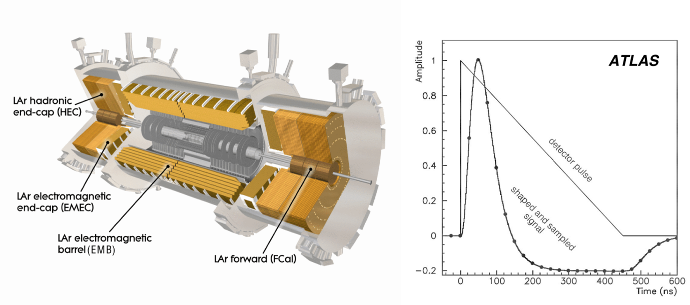
LAr hadronic end-cap (HEC)
LAr electromagnetic end-cap (EMEC)
LAr electromagnetic barrel (EMB)
LAr forward (FCal)
ATLAS
Amplitude
1
0.8
0.6
0.4
0.2
0
−0.2
0
100
200
300
400
500
600
Time (ns)
detector pulse
shaped and sampled signal

Figure 1: Left - cutout of the LAr calorimeter, right - shaped and digitized LAr calorimeter pulse [1].

2.1 Convolutional Neural Networks

The developed Convolutional Neural Networks for energy reconstruction use 1-dimensional filters to process time-series data and they consist of a two-staged architecture shown in Figure 2 where the first two layers are trained to detect deposit above the noise threshold and the later layers reconstruct the energy. Their input is a sliding-window of the detector signal with a window size of 28 for the network labeled Conv-3 and window size of 13 for Conv-4.

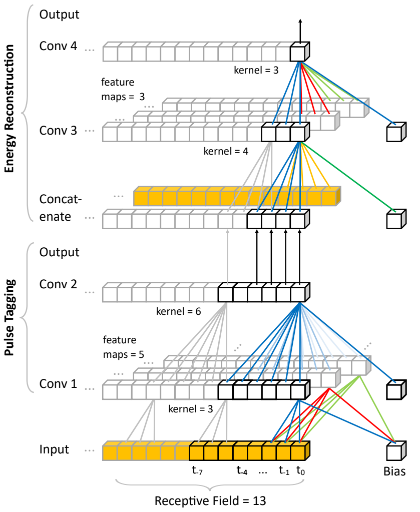

Figure 2: The CNN architecture [13].

2.2 Recurrent Neural Networks

Recurrent neural networks are a family of neural networks for processing sequential data. Long Short Term Memory (LSTM) [10] contains complex internal structure to gate the flow of information to the next timestep using neural network layers with sigmoid and tanh activation functions. This allows LSTM to process longer sequences and as such it can be applied in both sliding window in which the digitized signal from the calorimeter is split into overlapping sub-sequences of length of 5 where each sub-sequence has a single reconstructed energy shown in Figure 3 as well as single cell method in which the full signal is processed in a stream. Vanilla RNN is a simpler network structure for which a ReLU activation was chosen has significantly fewer parameters. Only sliding window architectures are used to train the Vanilla RNN due to its simplicity which makes it unstable to longer sequences.

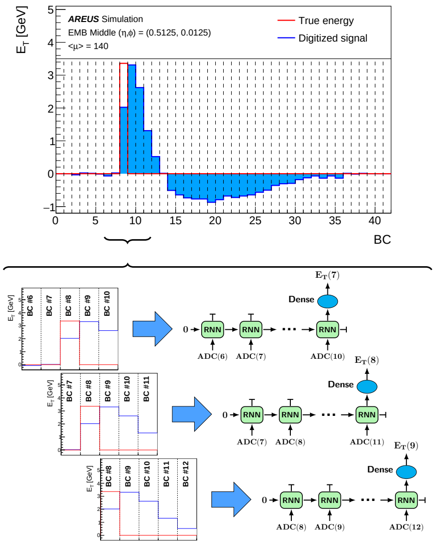

Figure 3: Sliding window application of RNNs [13].

2.3 Performance

The transverse energy resolution given by various ANNs for energy deposits 3σ above the noise level ($E_T > 240$ MeV) is shown in Figure 4. The ANNs perform better than the OF with MaxFinder by having both smaller bias for the mean and a better resolution. The main improvement is with overlapping events as shown in Figure 5. Networks using more samples from past events yields a better correction for overlapping events.

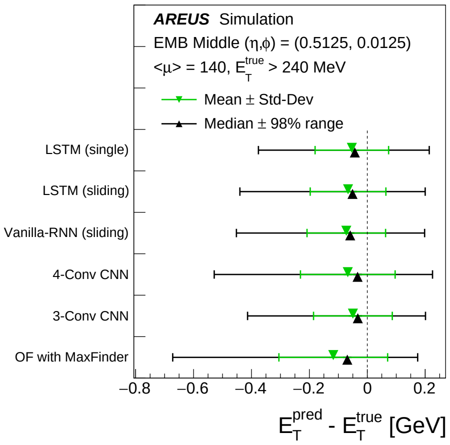

Figure 4: Resolution for energy deposits 3σ above over noise level [13].

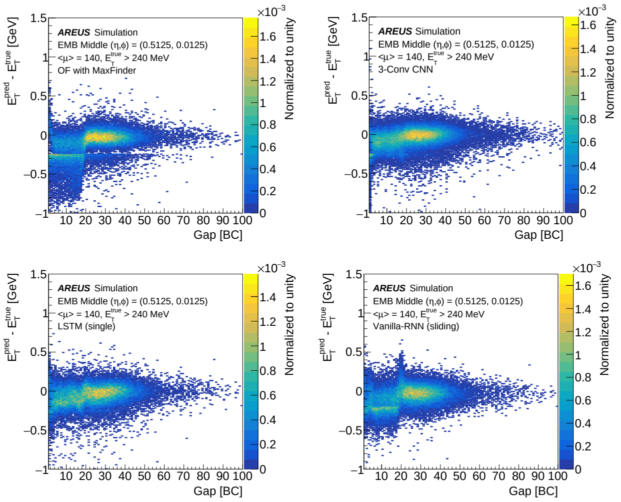

Figure 5: Resolution as a function of the distance to previous high energy deposit (gap) [13].

3. FPGA Implementation

The CNNs are implemented using Very High-speed integrated circuit hardware Description Language (VHDL). RNNs are implemented using High Level Synthesis (HLS). The fixed point bitwidths are chosen to ensure a resolution of the order of 1%. The results shown in Figure 6 compare the Keras output to the FPGA implementations simulated with Quartus 20.4 [11] and Questa Sim 10.7c [12]. The small differences observed are caused by quantization and by the LookUp Table (LUT) based realization of the activation function. To be able to reconstruct the energies of up to 384-512 calorimeter cells with a single board, it is necessary to compute the energies of several cells with a single instance of the NNs between the inputs coming in every 25 ns. This is referred to as multiplexing. The required cell count with one FPGA for the Phase-II upgrade can be achieved by the 3-Conv CNN and the Vanilla RNN implementations as shown in table 1. Further performance comparison is available in [13].

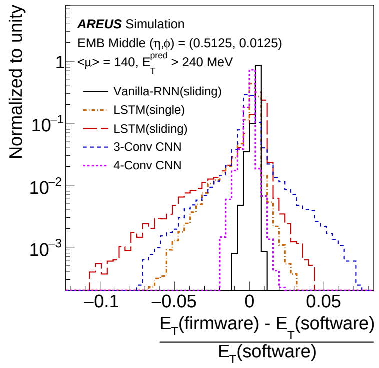

Figure 6: Relative deviation of the firmware and software results [13].

	Multiplexing	Freq	Latency	LAr	Resource Usage
		F_{max} MHz	clk_{core} cycles	Channels	DSP/ALM
3-Conv	6	344	81	390	0.8% / 1.5%
Vanilla	15	640	120	576	2.6% / 0.6%

Table 1: Occupancy, latency, maximum achievable clock frequency and the maximum number of LAr channels that can fit in one Stratix-10 FPGA assuming all FPGA resources are dedicated to energy reconstruction algorithms [13].

4. Conclusion

CNNs and RNNs have shown to outperform the optimal filter algorithm in reconstructing the energies deposited in the LAr calorimeter in HL-LHC conditions. The performance increase is greatest in the case of overlapping events. Latencies of about 200 ns and maximum execution frequencies of 344–640 MHz of the multiplexed networks partially fill the real-time processing requirements of the LAr.

The deployment of ANNs on FPGAs has a great potential to improve the energy reconstruction by the ATLAS LAr calorimeters at high luminosities, which will allow more sensitive physics analyses and a more efficient event selection by the ATLAS trigger system.

5. Acknowledgements

This work was in part supported by the German Federal Ministry of Education and Research within the research infrastructure project 05H19ODCA9. The project leading to this publication has received funding from Excellence Initiative of Aix-Marseille Université - A*MIDEX, a French "Investissements d'Avenir" programme, AMX-18-INT-006.

References

[1] ATLAS Collaboration (2008) The ATLAS Experiment at the CERN Large Hadron Collider, JINST 3:S08003. https://doi.org/10.1088/1748-0221/3/08/S08003

[2] Evans L, Bryant Ph, eds. (2008) LHC machine. JINST 3:S08001. https://doi.org/10.1088/1748-0221/3/08/S08001

[3] ATLAS Collaboration (2017) Technical Design Report for the Phase-II Upgrade of the ATLAS TDAQ System, CERN-LHCC-2017-020, ATLAS-TDR-029. https://cds.cern.ch/record/2285584

[4] Intel Corporation (2020) Intel Stratix-10 Device Datasheet, Version 2020.12.24

[5] ATLAS Collaboration (2017) Technical Design Report for the Phase-II Upgrade of the ATLAS LAr Calorimeter, CERN-LHCC-2017-018, ATLAS-TDR-027. https://cds.cern.ch/record/2285582

[6] LeCun Y, et al. (1989) Backpropagation applied to handwritten zip code recognition, Neural Computation 1(4):541–551. https://doi.org/10.1162/neco.1989.1.4.541

[7] Sherstinky A (2020) Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network, Physica D: Nonlinear Phenomena 404:132306. https://doi.org/10.1016/j.physd.2019.132306

[8] Cleland W E, Stern E G (1994) Signal processing considerations for liquid ionization calorimeters in a high rate environment, NIM A Volume 338:467-497. https://doi.org/10.1016/0168-9002(94)91332-3

[9] Madysa N (2019) AREUS: A Software Framework for ATLAS Readout Electronics Upgrade Simulation, EPJ Web Conf. 214:02006. https://doi.org/10.1051/epjconf/201921402006

[10] Hochreiter S, Schmidhuber J (1997) Long Short-Term Memory Neural Computation 9:1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[11] Quartus, ModelSim and HLS tools available from https://www.intel.com; Accessed: 2021-02-18

[12] Questa Sim available from https://eda.sw.siemens.com/; Accessed: 2021-06-20

[13] Aad, G., Berthold, AS., Calvet, T. et al. (2021) Artificial Neural Networks on FPGAs for Real-Time Energy Reconstruction of the ATLAS LAr Calorimeters. Comput Softw Big Sci 5, 19 https://doi.org/10.1007/s41781-021-00066-y