FPGA performance for signal reconstruction of the ATLAS Tile Calorimeter in the HL - LHC environment

Sonakshi Ahuja[1*], Fernando Carrió[1], Antonio Cervelló[1], Francesco Curcio[1], Luca Fiorini[1], Antonio Gómez[1], David Hernandez[1], Francisco Hervás[1], Arantxa Ruiz[1], Alberto Valero[1]
[1] Instituto de Física Corpuscular, Spain
(*) sonakshi.ahuja@cern.ch

***Abstract*—The ATLAS Tile Calorimeter (TileCal) will face considerable challenges from increasing radiation and high data throughput in the High Luminosity Large Hadron Collider (HL - LHC) era. This study aims to find a balance between FPGA resource usage and latency with efficient signal reconstruction algorithms of the calorimeter pulses. The precise reconstruction of signal pulse from calorimeter pulses is crucial as it corresponds to the energy of the particle deposited in the calorimeter. Optimal Filtering (OF) is currently used for reconstructing the pulse amplitude from digitised samples. While efficient, the performance degrades considerably with increasing pile up. A solution to combat this challenge is to work with Neural Networks. This study shows the performance of OF contrasted with a Single Layer Perceptron (SLP) with data trained on simulated ATLAS TileCal pulses. The results show that the SLP outperforms the OF algorithm in the pile-up case, which is expected in the HL - LHC era.**

***Keywords* — HL - LHC, TileCal, Signal Reconstruction**

I. Introduction

THE Tile Calorimeter (TileCal) [1] is the central hadronic calorimeter of the ATLAS [2] detector at the Large Hadron Collider (LHC). It is a sampling calorimeter that uses steel plates as the absorber material and plastic scintillators as the active medium. Fig 1. shows the TileCal barrels and electronics drawer unit.

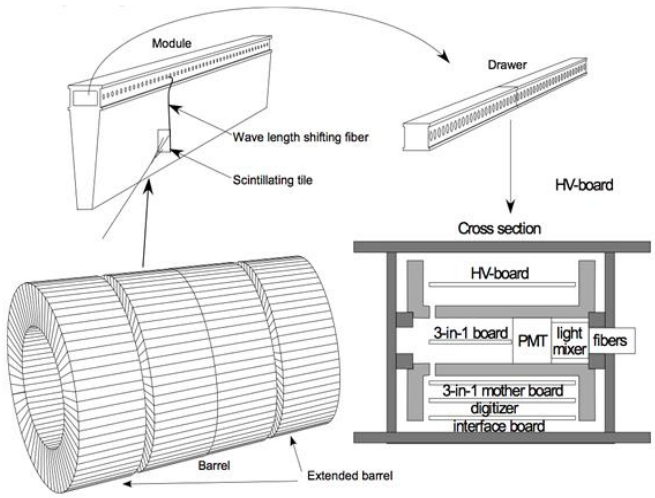

Fig. 1 The Tile Calorimeter extended and long barrel sections with the drawers containing the electronics [3]

The calorimeter is partitioned in three main sections: the central barrel consists of the the Long Barrel ($\eta < 1.0$), followed by two extended barrels covering a pseudorapidity (η) range of 0.8 - 1.7. Each barrel is segmented into 64 azimuthal modules covering 2π in ϕ. Each cell in the calorimeter corresponds to a combination of radial and azimuthal segmentation and is read out by two PMT's for redundancy.

When particles traverse the detector, they deposit energy in the scintillators, producing light that is collected by wavelength-shifting fibres and read out by PMTs. These analog signals are shaped, amplified, and digitised by the front-end electronics.

Accurate reconstruction of the particle energy from these digitised samples is critical, especially under the high pile-up conditions anticipated at the HL - LHC [4]. Currently, TileCal employs the OF algorithm for signal reconstruction, a method used extensively during Run 2 and Run 3. OF estimates the signal amplitude as a weighted sum of the digitised samples, using precomputed weights optimised for the nominal pulse shape and expected noise. While this algorithm is robust and effective under stable conditions, its performance degrades when the signal is distorted by pile-up or is misaligned with the sampling window — conditions that are expected to become more common at the HL - LHC.

To address these limitations, this study investigates the use of neural networks for signal reconstruction in TileCal. Neural Networks can learn complex, non-linear correlations from data and adapt to varying pulse conditions.

This work focuses on a SLP, a simple feed-forward model that is both efficient and suitable for implementation in firmware on FPGAs. This makes it a promising candidate for real-time processing in the upgraded TileCal readout system. The SLP, trained on pseudo TileCal pulses, is evaluated in pile-up conditions and compared with the performance of the OF algorithm under the same conditions.

II. Readout Architecture and Signal Reconstruction

This section begins with the setup and implementation of the two signal reconstruction algorithms, Optimal Filtering and a Neural Network based Single Layer Perceptron. We also detail the data preparation, implementation on FPGA firmware and reconstruction algorithm design.

The upgraded TileCal [4] readout system introduced in the Phase-II upgrade is designed to operate under the demanding conditions of the HL-LHC, including higher luminosity, increased radiation, and substantial pile-up. A key feature of this architecture is the full digitisation of signals from each PMT individually, ensuring that the granularity of the detector

is fully preserved. Importantly, the data from every bunch crossing is transmitted off-detector, for the reconstruction of the signals.

The analog signals generated by the PMTs are processed by the front-end electronics, which perform signal shaping, amplification, and digitisation. These include the upgraded 3-in-1 cards, responsible for shaping the signal and providing both high-gain (HG) and low-gain (LG) amplification paths

Digitisation is performed with a 10-bit ADC operating at 40 MHz, corresponding to one sample per LHC bunch crossing, The upgraded front-end also changes from the 3-in-1 to new boards such as FENICS, while the digitisation uses a 12-bit width which manages digitised signal routing to the off-detector electronics [5].

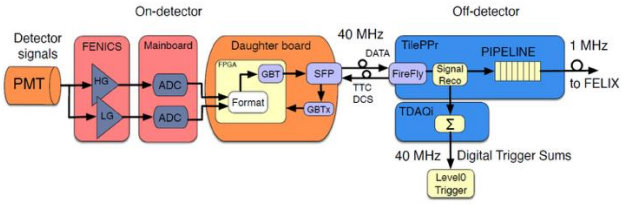

Fig. 2 The Tile Calorimeter upgraded readout architecture for the HL - LHC era.

Once off-detector, the digitised samples are passed to the PreProcessor module where energy reconstruction algorithms such as Optimal Filtering or Neural Network models are implemented in either software or firmware. The consistent digitisation and full granularity provided by the upgraded readout system are crucial for the performance of these reconstruction methods, especially under the high pile-up conditions expected at HL-LHC.

III. Signal Reconstruction Algorithms

To extract the signal amplitude from digitised samples, we explored different reconstruction methods. This study compares the well-established Optimal OF algorithm with a neural network-based approach, namely the SLP.

A. Optimal Filtering Algorithm

The OF algorithm is a lightweight and robust technique that has been used extensively in the ATLAS Tile Calorimeter readout during Run 2 and Run 3. Its primary objective is to reconstruct the amplitude of a pulse — which is directly related to the energy deposited in the calorimeter — by computing a weighted sum of the input samples, as shown in figure 2. The reconstructed amplitude A is given by:

$$A = \sum_{i=0}^{n-1} a_i \cdot y_i,$$

where y_i are the digitised samples and a_i are optimal weights precomputed based on expected conditions.

These weights are calculated using the known pulse shape, the autocorrelation matrix of the electronic noise, and assumptions on the signal timing and phase. While the OF method is efficient and works well under ideal conditions, it is known to be sensitive to pulse shape distortions and out-of-time signals (pile-up), which can lead to degraded performance in high-luminosity environments. In the TileCal readout, OF typically uses 7 input samples, centred around the expected signal peak. The signal itself is amplified in two gain stages — high gain and low gain — with a ratio of 1:40, to cover a dynamic range from approximately 200 fC to 1000 pC

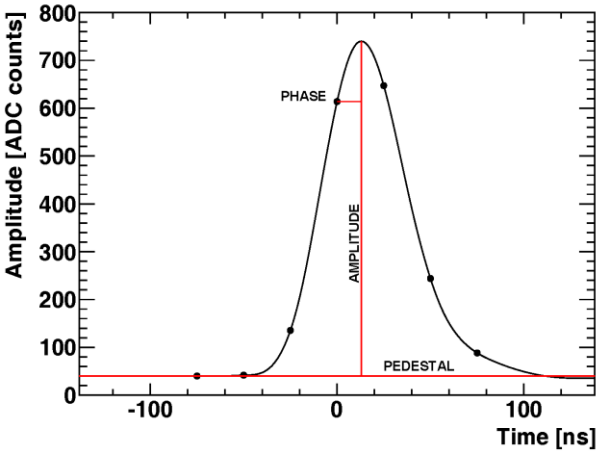

Fig. 3 Optimal Filtering *signal* reconstruction in the Tile Calorimeter using 7 input samples, and a non - centred pulse on the reconstructed amplitude.

B. Neural Network Model

The study and implementation of a neural network, specifically a SLP, was motivated by the challenges faced by the OF algorithm in conditions of high pile-up and increased data throughput. Unlike OF, which relies on fixed, pre-determined weights, the SLP offers a more flexible, data-driven approach that can learn complex, non-linear relationships from the data.

The model takes as input a sliding window of 9 samples, which correspond to the digitised signal samples from the detector. This input is processed through a single weight layer followed by a hyperbolic tangent (tanh) activation function.

The goal is to reconstruct the true amplitude of the signal for each event. A simple Mean Squared Error (MSE) loss is used to evaluate the difference between the predicted and actual amplitudes during training.

To maintain a fair comparison, the SLP is trained and tested on the same dataset used for OF. Although training is performed only on High Gain (HG) samples, a straightforward conversion factor allows switching to Low Gain (LG) whenever HG saturates. One of the main strengths of using a neural network lies in its ability to adapt to variations in the pulse shape — including those that are distorted — through learning from realistic or simulated data.

Finally, the model output, which is a floating-point amplitude, is converted into a fixed-point format suitable for efficient implementation on the FPGA, keeping in mind the constraints on the resources.

IV. Testbench Development

The main goal of developing the firmware testbench is to evaluate the performance and efficiency of signal reconstruction algorithms before deploying them in the ATLAS cavern for the HL-LHC era. The testbench environment replicates the actual hardware configuration, using the same version of the FPGA that will be installed in the detector. This enables a realistic validation of the firmware's behaviour and

timing in a controlled setting.

The SLP model is implemented in VHDL, structured into components representing the input, output, and activation logic. The activation function, in this case a piecewise-linear approximation of the tanh function, is implemented separately and integrated into the overall architecture. Computations within the model are performed using fixed-point arithmetic, and normalisation and de-normalisation stages are included to properly scale input and output values to match the training regime and restore physical meaning to the reconstructed amplitude.

Since the model processes a sliding window of 9 samples, a first-in-first-out (FIFO) buffer is used to maintain the sliding sample window. The design is tested for different fixed-point precisions (bit widths) to evaluate the trade-off between resource usage and reconstruction accuracy. Latency is also studied and adjusted to meet clock timing constraints and ensure deterministic behaviour.

The full firmware compilation proceeds in three stages: synthesis, placement and routing, and finally bitstream generation. The bitstream corresponds to the binary configuration loaded onto the FPGA.

Firmware simulation is carried out using a cocotb-based testbench[6], which enables co-simulation of the VHDL implementation and Python control code. The simulation flow is tightly synchronised with the firmware clock, where the testbench waits for valid signals from the firmware before supplying new data or proceeding to the next step. Inputs to the device under test (DUT) include the 9-sample high-gain window and the true amplitude, simulating conditions like those expected in real detector data.

To ensure reproducibility and avoid version mismatches between tools such as Vivado (a hardware design suite by Xilinx for synthesis and implementation of FPGA designs), cocotb (a coroutine-based Python framework for testing VHDL), Python, and ROOT, the entire setup is encapsulated within a Docker container. This provides a consistent and isolated environment for firmware development and testing.

The DUT outputs the reconstructed amplitude, which is written to a ROOT ntuple file. PyROOT is used for further analysis and visualisation, including histogramming and performance comparison. The reconstructed output from the FPGA is compared against the reference software model, allowing a direct evaluation of how closely the firmware reproduces the trained behaviour of the SLP.

V. Performance Evaluation and Results

The performance evaluation focuses on the firmware implementation in the high-gain range. A total of 1 million events were processed through the testbench to assess and compare the OF algorithm and the SLP under realistic conditions. We begin by contrasting their performance in scenarios with and without pile-up. In the absence of pile-up, the OF algorithm demonstrates excellent linearity in its reconstruction, clearly reflecting its efficiency in low-noise environments. However, under pile-up conditions, its reconstruction quality degrades significantly, failing to maintain a reliable amplitude response.

Given the expected increase in pile-up and data throughout the HL-LHC era, the limitations of OF in such scenarios highlight the necessity for more adaptable models.

The SLP, trained on pseudo TileCal pulses, outperforms OF in the presence of pile-up, showing a stronger linear correlation between the predicted and true amplitudes as shown in figure 4. When comparing the FPGA output of the SLP model with its software counterpart, a small discrepancy is observed — the software model offers slightly higher precision.

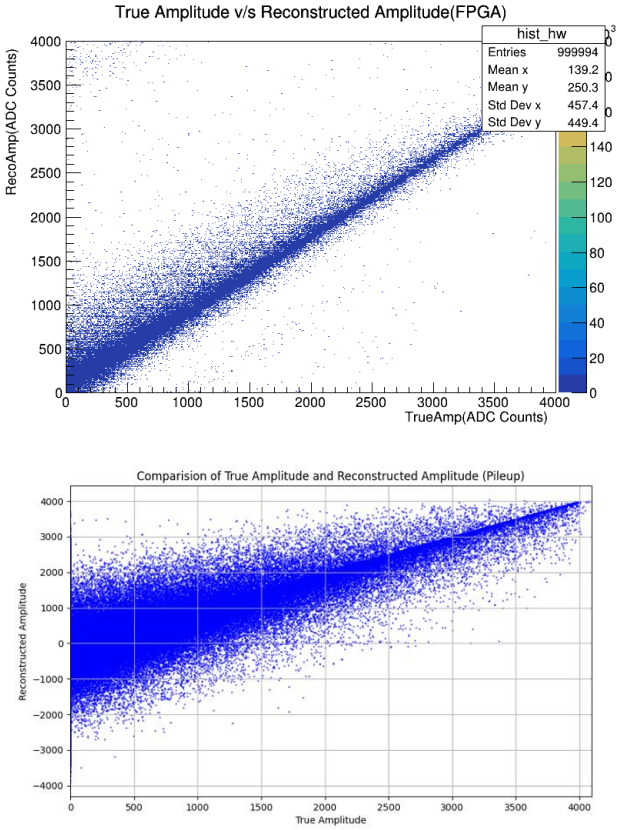

Fig. 4. Comparison between the NN (top) and OF (below) algorithm for True Amplitude and Reconstructed Amplitude in the FPGA.

This difference arises from the use of fixed-point arithmetic in the FPGA, as opposed to floating-point operations in software. Nevertheless, this minor trade-off is acceptable when considering the resource constraints and latency requirements of the firmware environment.

The firmware operates using two distinct clock domains: a 40 MHz clock aligned with the LHC bunch crossing, and a faster 400 MHz processing clock used internally for algorithm execution. In terms of latency, the OF implementation requires approximately 7 bunch crossings to produce an output, while the SLP introduces a slightly higher latency of about 12 bunch crossings, still well within acceptable bounds for real-time reconstruction.

VI. Conclusions and Results

The aim of this work was to assess the performance of signal reconstruction algorithms implemented in firmware, under the conditions expected at the HL-LHC—namely, increased pile-up, radiation, and data rates. We focused on two approaches: the Optimal Filtering algorithm, known for its simplicity and

reliability, and a more flexible Single Layer Perceptron neural network model. Our results show that while OF performs well under low pile-up conditions, the neural network method provides improved reconstruction accuracy in high pile-up scenarios.

This study implemented both reconstruction algorithms directly in the FPGA firmware. Their performance was evaluated using a custom testbench that simulates realistic TileCal conditions, allowing us to validate the firmware before deployment. The study demonstrates that both OF and neural networks can be deployed and evaluated in a controlled environment before integration in the ATLAS detector.

The work also underscores the importance of the upgraded readout architecture, which—by providing full-granularity digitised data at 40 MHz—enables robust real-time reconstruction and opens the door to more sophisticated algorithmic strategies like machine learning. Overall, this study serves as a proof of concept for implementing neural network-based reconstruction in firmware and contributes to the broader effort of preparing the TileCal readout for HL - LHC operation.The document must be in this two-column format.

VII. Future Work

The next steps involve exploring more complex neural network architectures such as Multi-Layer Perceptrons (MLPs) and Convolutional Neural Networks (CNNs) to improve reconstruction accuracy and robustness. We also plan to investigate different configurations of these models, including the use of various activation functions like ReLU and PReLU, optimisers such as ADAM, and alternative loss functions. Additionally, we will scan over different window sizes to understand how temporal context affects reconstruction. These studies aim to assess the trade-offs between reconstruction performance and FPGA resource usage.

Acknowledgment

This research was funded by Ministerio de Ciencia e Innovación with project reference PID2021-125069OB-I00.

References

[1] ATLAS Collaboration, “Operation and performance of the ATLAS tile calorimeter in LHC Run 2”, Eur. Phys. J. C 84, 1313 (2024), https://doi.org/10.1140/epjc/s10052-024-13151-4
[2] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider”, JINST 3 (2008) S08003
[3] High-Luminosity Large Hadron Collider (HL-LHC) : Preliminary design 304 report, CERN (2015), http://dx.doi.org/10.5170/CERN-2015-005
[4] ATLAS Collaboration, “Technical Design Report for the Phase-II Upgrade of the ATLAS Tile Calorimeter”, CERN-LHCC-2017-019, https://cds.cern.ch/record/2285583
[5] E. Valdes, et al., Readiness of the ATLAS Tile Calorimeter link 309 daughterboard for the High-Luminosity LHC era, PoS(TWEPP2019)087 310 (2019). http://refhub.elsevier.com/S0168-9002(22)01152-4/sb6
[6] T. Martin and F. Schmidt, “Cocotb: Coroutine-based co-simulation testbench for verifying RTL using Python,” GitHub Repository, https://github.com/cocotb/cocotb.