

Pulse shape Discrimination for Online Data Acquisition in Water Cherenkov Detectors Based on FPGA/SoC

Luis G. García Ordóñez,^{a,b,*} Romina Soledad Molina,^{a,b,d} Iván René Morales Argueta,^a Maria Liz Crespo,^a Andres Cicuttin,^a Sergio Carrato,^b Giovanni Ramponi,^b Héctor Eduardo Pérez Figueroa^c and Maynor Giovanni Ballina Escobar^c

^a(ICTP-MLAB) *The Abdus Salam International Centre for Theoretical Physics
Strada Costiera, 11, 34151 Trieste Italy*

^b(UNITS) *Università degli studi di Trieste, Department of Engineering and Architecture
Piazzale Europa, 1, 34127 Trieste, Italy*

^c(USAC-ICFM) *Instituto de Investigación en Ciencias Físicas y Matemáticas, Universidad de San Carlos de Guatemala
Ciudad Universitaria, Zona 12, Guatemala, Guatemala*

^d(UNSL-LEIS) *Departamento de electrónica, Universidad Nacional de San Luis
Av. Ejército de los Andes 950, D5700, San Luis, Argentina*

E-mail: lgarcia1@ictp.it

Discrimination of secondary particles produced in extensive air showers is needed to study the composition of primary cosmic rays. High speed data acquisition and the increase in resources in modern FPGAs with the addition of a microprocessor in System-on-Chip (SoC) technologies allow to implement complex algorithms for digital signal analysis. Pulse shape Discrimination (PSD) can be carried out in real-time on the digital front-end of the detector; indeed online data analysis permits to save computational resources in post-processing and transmission bandwidth. We describe two methods for PSD, the first one based on artificial neural network (ANN) using the novel hls4ml package, and the other based on a correlation approach using finite impulse response (FIR) filters. Both methods were implemented and tested on Xilinx FPGA SoC devices ZU9EG Zynq Ultrascale+ and XC7Z020 Zynq. Data from a Water Cherenkov Detector (WCD) were acquired with a 500 Mhz, 8-bit high speed analog-to-digital converter acquisition system. Experimental results obtained with both methods are presented along with timing, accuracy and resources utilization analysis.

*37th International Cosmic Ray Conference (ICRC 2021)
July 12th – 23rd, 2021
Online – Berlin, Germany*

*Presenter

1. Overview

Pulse shape Discrimination (PSD) is a method used for separating particles with different specific ionization by analyzing the features of the signal [1]. The ease to identify this features usually depends of the physics of the detector medium used to generate the signal, which have lead to the study of new scintillator materials with better characteristics to do PSD [2]. However equally important is the acquisition system and its capacity of recognizing said features, as PSD relies on the difference in the pulse shapes produced. For instance, pulses generated by neutrons will have a longer tail than the pulses generated by gamma rays [3]. Different type of particles have dissimilarities in rise-times and amplitudes [4]. Furthermore, higher number of samples per pulse (pulse-rate) allows us to make a better comparison among different type of signals produced by cosmic rays.

Traditional PSD techniques utilize charge-integration because it is a robust, low computational demanding method; as improvements of data acquisition hardware like faster Field Programmable Gate Arrays (FPGAs) and System on Chip (SoC) arise, other more demanding methods have been implemented. Some examples are frequency gradient analysis (FGA) which studies variation of the frequencies created by the pulses, Pulse Gradient Analysis (PGA) which takes advantage of the differences between the peak and the sample amplitude of the pulses, and Charge Comparison Method (CCM) which is the study of the entire pulse and its tail to discriminate the event in small time windows [5].

Neural networks have also been used due to their potential as classifiers because of the use of experimental data for training [6]. We presented a PSD study with Artificial Neural Networks (ANN) and FIR-Based Correlation Digital Signal Processing (DSP) [7] capable of discriminating between two different type of particles. Thus, following this study, an upgraded version of the acquisition system is presented heading towards higher speed data sampling and processing capabilities. As in the previous design, two real-time digital PSD methods are presented: a traditional FIR-based DSP, as well as a Multi-Layer Perceptron (MLP) Artificial Neural Network (ANN). Moreover, the latter has been updated by creating a shrunk version, not only improving its latency, but also a reducing the area is intended for smaller FPGA fabric implementations.

2. Pulse selection criteria for shape discrimination

2.1 Centroid-based correlation distance clustering

An objective way of classifying the pulses is by grouping them by their similarities. Such classification was done with a centroid-based clustering algorithm. Due to the fact that measurement of similarity is based on shape, the computation of the distance was performed using Pearson correlation function, instead of Euclidean distance (such as in traditional K-means algorithm) [8] allowing the system to distinguish among pulse waveforms [7]. A set of ~3 million triggered pulses have been acquired and classified using this method; a subset of raw pulses within each cluster of interest is shown in Fig. 1.

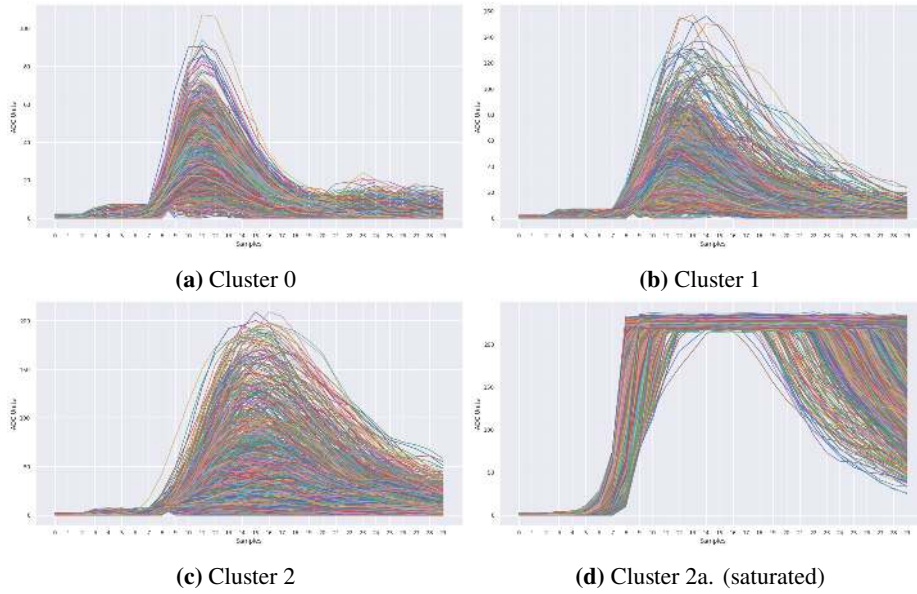


Figure 1: Sample of raw pulse traces of each cluster

2.2 Figure of Merit

To establish the capabilities of our system in isolating different particles, a standard figure of merit (FOM) was used [3]. The FOM was established by computing a charge histogram of the acquired data and selecting the bins with higher charge counts. In Fig. 2a it is possible to distinguish two different types of pulses, the first located between the 200 and 400 units of charge, and the second one between 500 and 700 units. With this information we selected the clusters that contribute the most counts to the charge in the two established regions as shown Fig. 2b. These clusters will become the ground truth for the ANN and for calculating the FIR coefficients.

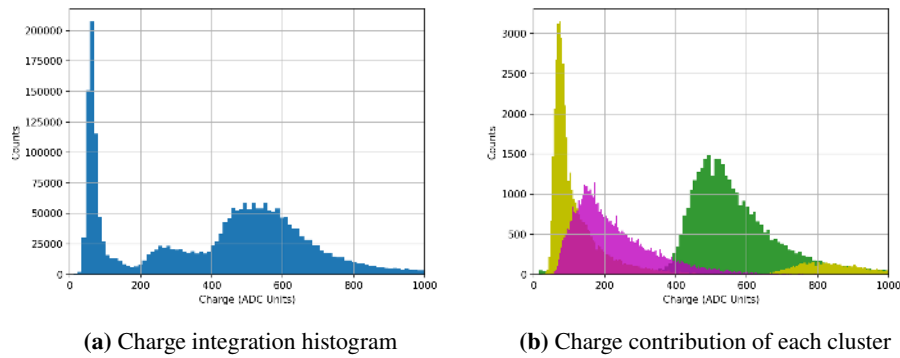


Figure 2: Charge contributions expressed in integrated ADC channel units

Further analyses are needed to do a precise determination of the particles that corresponds to each cluster. However, based on the work of Salazar [9] as reference we estimate that the area tagged in yellow may be mostly formed by electrons. The area tagged in magenta may correspond to inclined muons, whilst the green area may correspond with vertical muons. A precise determination requires more studies that are out of the scope of this paper.

Other differences among the pulses are shown in Fig. 3 where it can be distinguished the relation between the rise time and amplitude of the selected clusters. The first thing to notice is the divergence of rise time among the green cluster (type 2) and both, yellow and magenta groups (type 0 and 1, respectively); the last two separate each other by a change in the amplitude. By further contrasting the slew rate and the amplitude is noticeable the dissimilarity in slopes among the three types of pulses, where the green cluster clearly shows different inclination than the other two clusters.

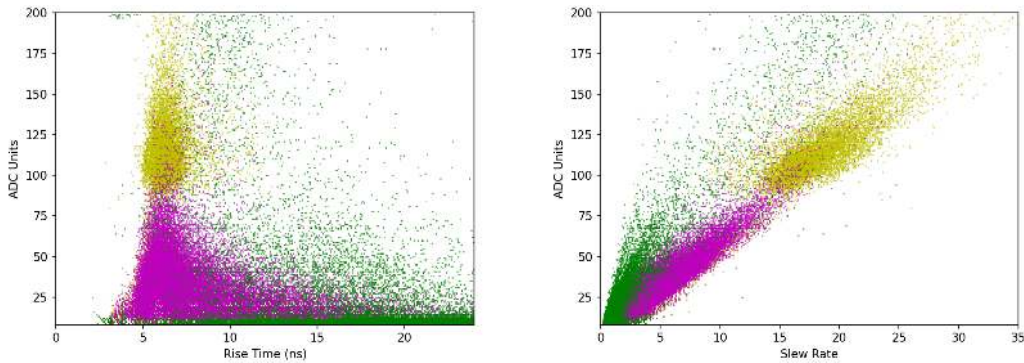


Figure 3: Rise Time vs Amplitude and Slew rate vs Amplitude plots for the three selected clusters.

3. Online PSD methodology

Two different methodologies for online PSD are presented. One based on ANN and the other based on correlation analysis using FIR. The details are shown below.

3.1 Neural network architecture

In this approach, the main objective is to classify 4 types of signals through an ANN based on MLP. As FPGA/SoC is the final platform for the implementation of PSD based on MLP, we must compress the model in order to make efficient use of this technology, considering resource utilization and latency, without compromising the final performance of the overall system in terms of accuracy. In this direction, model compression was performed using a combination of different techniques: (1) knowledge distillation, (2) pruning and (3) quantization.

The first strategy used for compression was knowledge distillation[10]: a larger model (teacher) was trained in order to distill the knowledge to a smaller model, called the student. Both networks were defined through fully-connected layers, with rectifier linear unit (ReLU) as activation function. For the output layer of each architecture, Softmax was chosen to perform the final classification. Based on this, the first step is to train the network for the teacher. Then, the knowledge is transferred to the student network, which is composed by one input layer, 2 hidden layers and the output layer.

QKeras [11] allows to perform the training with quantized layers defined by this package. This technique is useful to reduce the memory footprint in resource-constrained devices. Regarding pruning, it was applied to remove redundant parameters with a target sparsity of 0.7.

Once the final model (the student) was obtained after training and compression stages, the hls4ml package [12] was used to translate the neural network inference to a Vivado High-Level Synthesis project, for a subsequent implementation in the FPGA fabric as an IP-core.

3.2 FIR-based correlation analysis

Pulse selection based on Pearson correlation analysis was implemented taking advantage of a Tapped delay FIR line from a Xilinx FIR v7.2 core [13], configured on Systolic Multiply-Accumulate architecture. Three FIR filters, one for each cluster, are set in parallel in free-running mode connected to a voting system where the pulse will be identified and tagged. The FIR coefficients are calculated based on a representative pulse of each cluster.

An ideal representative pulse is the one which has in average the same correlation distance with the rest of the pulses. By applying standard normalization, it is possible to obtain a representative pulse with average correlation distance to the rest of the cluster [8].

In Fig. 4 it is shown a heat map of all the pulses contained on each of the selected clusters (scaled to 1 and aligned to the maximum value). Overlaid to each map (also scaled to 1) is the typical pulse obtained by averaging the standard score.

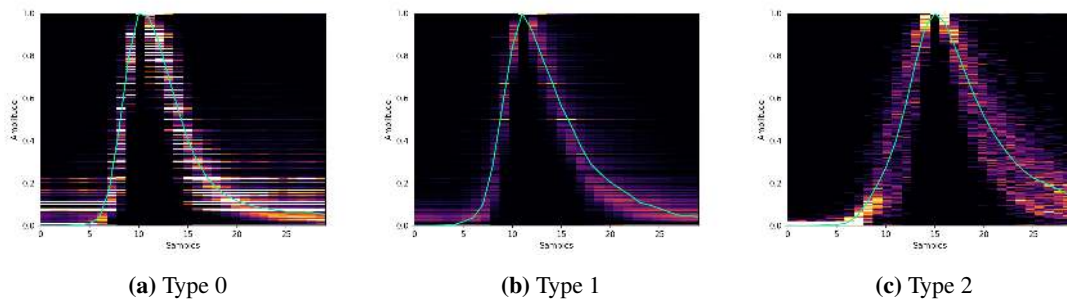


Figure 4: Heat map of pulses (scaled to 1) found in the cluster and typical pulse obtained by standard score for each cluster.

The pulses correlation matrix is shown in the table 1. To compensate for the high correlation between types 0 and 1, the voting system will also extract the amplitude of the pulse for the final decision.

	Type 0	Type 1	Type 2
Type 0	1.0000	0.9600	0.7744
Type 1	0.9600	1.0000	0.8360
Type 2	0.7743	0.8360	1.0000

Table 1: Pearson Correlation between the three selected pulses

4. Experimental setup

A light-tight 220 liters reservoir tank filled with demineralized water is used as the main scintillator, in which a Photonis XP1802 Photomultiplier Tube (PMT) is submerged at half of its

height. The PMT is polarized with a variable 2 kV high-voltage (HV) power supply (EMCO C-20), set to a fixed value of ~ 1.3 kV for this experiment. A custom board accommodates a 500 MSPS 8-bit analog-to-digital converter (ADC) where the PMT signal is fed into, which lays over a Zedboard development board housing a Zynq XC7Z020 SoC [14]. Communication between the DAQ and the computer is done via TCP/IP over ethernet, enabling remote control and sampling through an existing network infrastructure.

A detailed insight of the test framework is shown in Fig. 5. The design was implemented in a ZCU102 development board housing a ZU9EG Ultrascale+ to measure the accuracy and timing of the design. For measurements in the Zynq XC7Z020 the designs were implemented individually to obtain the occupancy.

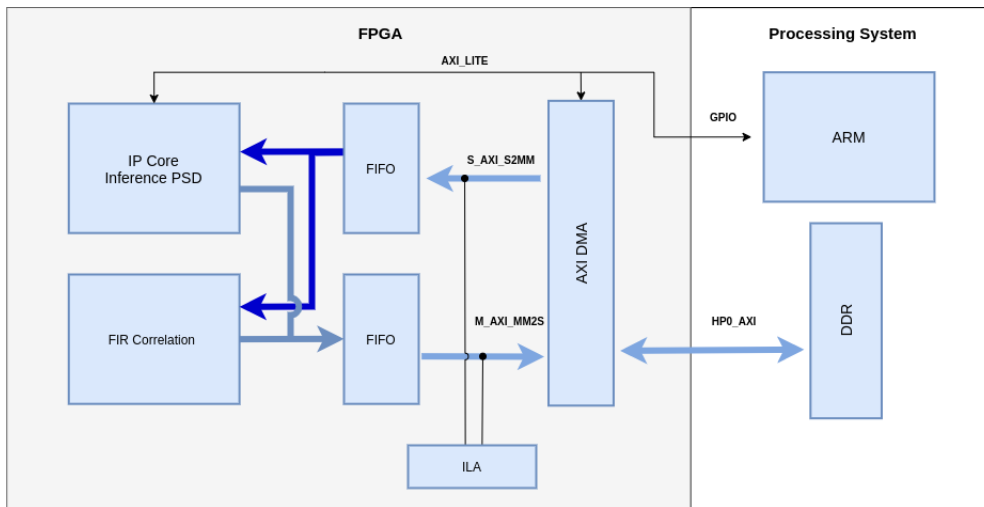


Figure 5: DAQ hardware implementation tested in ZU9EG as in [7]

5. Results

A comparison of the resources utilization for the MLP and FIR PSD blocks for the two platforms under test is shown in Table 2. Both designs were synthesised and implemented using the default options. The difference of available resources between the XC7Z020 (Series 7) and ZU9EG (Ultrascale+) families may have lead to a different optimal distribution chosen by the synthesis and implementation tools.

Platform	PSD	LUT	Registers	BRAM	DSP Blocks
XC7Z020	FIR	1.02%	2.00%	2.00%	47.7%
	MLP	68.1%	29.5%	0.00%	70.5%
ZU9EG	FIR	0.20%	0.39%	0.39%	0.00%
	MLP	11.2%	2.29%	2.29%	57.5%

Table 2: Resource utilization comparison between XC7Z020 and ZU9EG

The latency and dead-time of each block was measured using Vivado's Integrated Logic Analyzer (ILA). After the 30 sample window was filled, the MLP block had a dead time of 15

clock-cycles in the ZU9EG and 68 in the XC7Z020, whereas the FIR implementation presented a free-running latency of 26 clock-cycles. Using a 250 MHz system clock, the total processing time for each pulse for the MLP is of 180 ns in the ZU9EG Ultrascale+ and 392 ns in the XC7Z020 Zynq. The FIR block timed 224 ns in both platforms.

The classification accuracy for both methods is shown in Fig. 6. The MLP was able to predict with accuracy above 90% for each type of pulse with an overall accuracy of 94.2%, while the FIR-based correlation was able to predict each value with accuracy above 80% and an overall accuracy of 84.3%.

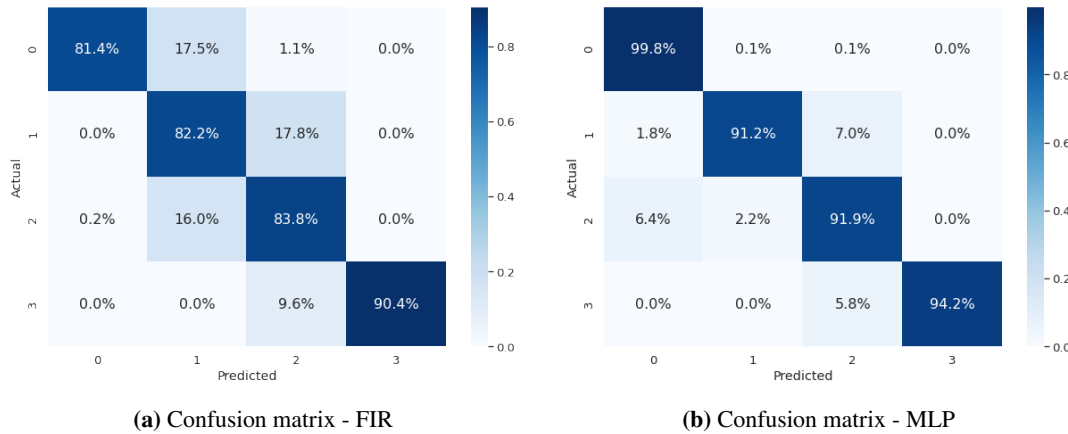


Figure 6: Classification accuracy representation in confusion matrices for each method

6. Conclusions and further work

With overall accuracy of 84.3% for the FIR and 94.2% for the MLP, we can conclude that both techniques present high detection rate for pulse shape analysis. The MLP presented higher overall accuracy than the FIR correlation system at expense of higher use of resources plus a small dead-time between pulses which is easily resolved by adding a FIFO before the input as buffering for the trigger signal.

The presented framework will be used as a base for subsequent works related to cosmic rays studies. Further analysis of the acquired data will be done to identify the type of particle corresponding to each cluster. Moreover, adding a system of tracking detectors based on silicon photomultipliers (SiPM) and plastic scintillators may be used for pulse coincidence, allowing for a more precise pulse isolation. It is also planned to do, a hybrid design using the FIR as a pre-classification system before the neural network to increase the scope of particle identification.

7. Acknowledgments

The author Iván René Morales acknowledges the receipt of a fellowship from the ICTP Programme for Training and Research in Italian Laboratories, Trieste, Italy. And to the LAGO Collaboration for the training workshops and seminars.

References

- [1] M. Roush, M. Wilson and W. Hornyak, *Pulse shape discrimination*, *Nuclear Instruments and Methods* **31** (1964) 112.
- [2] M. Grodzicka-Kobylka, T. Szczesniak, M. Moszyński, K. Brylew, L. Swiderski, J. Valiente-Dobón et al., *Fast neutron and gamma ray pulse shape discrimination in EJ-276 and EJ-276g plastic scintillators*, *Journal of Instrumentation* **15** (2020) P03030.
- [3] A.T. Lintereur, J.H. Ely, J.A. Stave and B.S. McDonald, *Neutron and gamma ray pulse shape discrimination with polyvinyltoluene*, May, 2016.
- [4] H. Salazar and L. Villasenor, *Ground detectors for the study of cosmic ray showers*, *Journal of Physics: Conference Series* **116** (2008) 012008.
- [5] M.J. Balmer, K.A. Gamage and G.C. Taylor, *Comparative analysis of pulse shape discrimination methods in a 6Li loaded plastic scintillator*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **788** (2015) 146 .
- [6] J. Griffiths, S. Kleinegesse, D. Saunders, R. Taylor and A. Vacheret, *Pulse shape discrimination and exploration of scintillation signals using convolutional neural networks*, *Machine Learning: Science and Technology* **1** (2020) 045022.
- [7] L.G. Garcia, R.S. Molina, M.L. Crespo, S. Carrato, G. Ramponi, A. Cicuttin et al., *Muon–electron pulse shape discrimination for water cherenkov detectors based on fpga/soc*, *Electronics* **10** (2021) .
- [8] M. Berthold and F. Höppner, *On clustering time series using euclidean distance and pearson correlation*. *arxiv 2016*, *arXiv preprint arXiv:1601.02213* .
- [9] H. Salazar and L. Villasenor, *Separation of cosmic-ray components in a single water cherenkov detector*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **553** (2005) 295.
- [10] G.E. Hinton, O. Vinyals and J. Dean, *Distilling the knowledge in a neural network*, *CoRR* **abs/1503.02531** (2015) .
- [11] C.N.C.J. au2, A. Kuusela, S. Li, H. Zhuang, T. Aarrestad, V. Loncar et al., *Automatic heterogeneous quantization of deep neural networks for low-latency inference on the edge for particle detectors*, 2021.
- [12] J. Duarte, S. Han, P. Harris, Jindariani and et al, *Fast inference of deep neural networks in fpgas for particle physics*, *Journal of Instrumentation* **13** (2018) P07027–P07027.
- [13] Xilinx, *Fir compiler v7.2 logicore ip product guide*, Tech. Rep. Xilinx (jan, 2021).
- [14] L.G. Garcia Ordóñez, I.R. Morales Argueta, M.L. Crespo, S. Carrato, A. Cicuttin, H.D.L.T. Perez et al., *DAQ platform based on SoC-FPGA for high resolution time stamping in cosmic ray detection*, *PoS ICRC2019* (2019) 266.