

The use of adaptive predictor filter as a trigger mechanism in simulated cosmic rays radio signals corrupted with Gaussian noise

Watanabe, C.K.O.,^{a,b,c,*} Diniz, P.S.R.^b and De Mello Neto, J.R.T.^a

^aFederal University of Rio de Janeiro (UFRJ), Physics Institute, Rio de Janeiro, Brazil

^bFederal University of Rio de Janeiro (UFRJ), The Alberto Luiz Coimbra Institute for Graduate Studies and Research in Engineering (COPPE), Signal, Multimedia and Telecommunications Laboratory (SMT), Rio de Janeiro, Brazil

^cInstitute for Astroparticle Physics (IAP), Karlsruhe Institute of Technology (KIT), Germany

E-mail: cla.watanabe@pos.if.ufrj.br, diniz@smt.ufrj.br, jtmn@if.ufrj.br

Adaptive filtering belongs to the realm of learning algorithms, widely used in our daily life in the context of machine learning, artificial intelligence, pattern recognition, etc. [1]. It is formally defined as a self-designing device with time-varying parameters that are adjusted recursively in accordance with the input data [2].

The trigger mechanism is a central task in experiments using antennas to detect cosmic rays as it selects a cosmic-ray induced signal among all the voltages traces events that reach the antennas. This work presents the efficiency of a trigger mechanism developed using the adaptive predictor filter technique, whose capability is well known for time series prediction usage. This technique is independent of an external detector, using only the online temporal field recorded by the antennas in a simulated data set and Gaussian noise.

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

*Presenter

1. Introduction

Radio-detection of a cosmic ray is a modern, well established, and low-cost technique that uses antennas to detect the electromagnetic component of the air shower [4]. Nowadays, several experiments are running worldwide and new experiments are planned. In both cases, the community expects improvements on the accuracy of the primary particle energy, composition, and arrival direction. The main challenge for this type of detection is the background present at the experimental site due to human-made radio noise, and the Galactic Gaussian noise [7]. The former can be drastically reduced if site is chosen in deserted and remote areas, as proposed by the GRAND experiment [6], while the second one is irreducible.

Using thresholding methods on the signal read by the antennas, a self-triggered mechanism leads to a high number of false-positive or false-negative events. This scenario pushed radio experiments to develop a hybrid trigger mechanism: the timestamp of the signal is detected in coincidence with particle and/or fluorescence detectors indicating when an event happened in temporal series. Then, one may search for this timestamp in the radio signal given by the antennas. Ideally, the trigger would have to work independently of another detector, especially in large-scale experiments, i.e., a higher number of available antennas covering a large area at which an external detector for each antenna is not available. Thus, the presence of a sophisticated self-trigger mechanism becomes necessary to improve detection efficiency.

An adaptive filter is an online learning algorithm, that among several applications, can be used for denoising, system recognition, and time series prediction. The latter one we use to test its efficiency as a self-trigger mechanism. The simulated data used in this analysis were generated using the COREAS [5] radio emission simulation tool from CORSIKA [3] software with a background composed of a Gaussian distribution of zero mean and a variable width.

2. The simulated dataset

The electric field was generated using the CORSIKA software using the CoREAS radio tool package. This simulation was performed in the GRAND experiment scenario, i.e., using the site's magnetic field and the experiment's antenna configuration. No impulsive noise is added to the background. We consider only the Galactic background described by a Gaussian spectrum. At the GRAND site, for example, the background is mainly of galactic origin: a Gaussian distribution with roughly $15 \mu\text{V}$ standard deviation [6].

The antenna response is specific to each type of antenna, considering its frequency band and geometry. The output read by the electronics will be a voltage value. Given the electric field traces provided by COREAS, we also simulated the GRAND collaboration' antenna response. We added the noise to the signal in order to mimic the realistic scenario, i.e., the voltage trace plus the Gaussian distribution. The dataset consists of a single voltage trace template with 5000 Gaussian background samples of different widths. These ranged from $1 \mu\text{V}$ to $50 \mu\text{V}$ in $1 \mu\text{V}$ steps, and for each width, 100 distributions were generated. The result for this simulation procedure was performed and shown in figure 1. This procedure is an attempt to span all possible signal-to-noise ratio scenarios of signals. The SNR in dB scale is defined as

$$\text{SNR} = 20 \log \left(\frac{\text{RMS}_{\text{signal}}}{\text{RMS}_{\text{noise}}} \right). \quad (1)$$

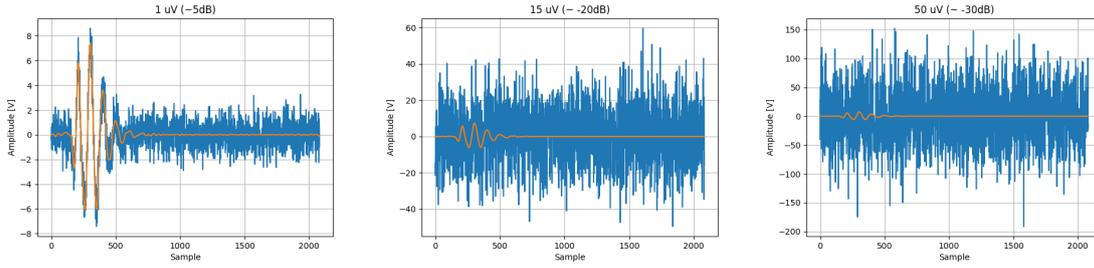


Figure 1: Starting with a standard radio signal template for each polarization, the Gaussian noise sample was ranged from $1\mu\text{V}$ to $50\mu\text{V}$ standard deviation, which gives a range from 5dB to -30dB. In the first plot, a good SNR, then the real SNR for GRAND site, and a very poor SNR. For each standard deviation, a hundred background samples were generated.

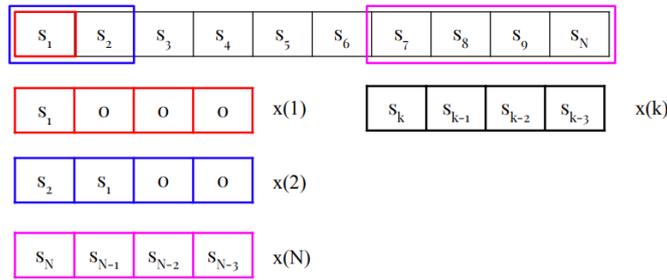


Figure 2: The FIFO (*First In First Out*) memory buffer design.

Both $\text{RMS}_{\text{signal}}$ and $\text{RMS}_{\text{noise}}$ were calculated considering the size of the trace, i.e., N equals to 2082 samples.

3. Adaptive filters

The signal read by the antennas can be defined as $\mathbf{s} = (s_1, \dots, s_i, \dots) \in \mathbb{R}^N$, where $s_i = 0$ for all $i \in \mathbb{N}$, except for a finite number of indexes. The adaptive filter does not read the entire signal \mathbf{s} at once. The input $\mathbf{x}(k)$ is a finite buffer of length M obtained from \mathbf{s} and defined as $\mathbf{x}(k) = (s_k, s_{k-1}, \dots, s_{k-i}, \dots, s_{k-(M-1)})$, for $k = 0, 1, 2, \dots$, where $s_j = 0$ if $j < 1$ (see fig. 2). Given an input signal $\mathbf{x}(k)$, the output signal y_k is defined as the inner product of a coefficients weight vector $\mathbf{w}(k)$ with the input-data vector,

$$y_k = \mathbf{w}(k)^T \cdot \mathbf{x}(k). \tag{2}$$

The goal is to minimize a function of the error between an ideal desired signal d_k and its output signal y_k , i.e., the error squared

$$e_k = (d_k - y_k)^2, \tag{3}$$

with a constant adjustment of the coefficients weight vector, based on the cost function minimization algorithm [1]

$$\mathbf{w}(k + 1) = \mathbf{w}(k) + \mu \mathbf{x}(k) e_k. \tag{4}$$

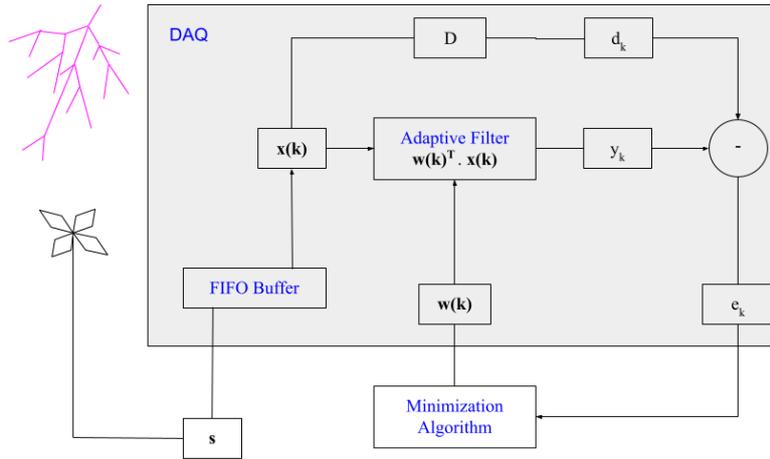


Figure 3: Full chain of the self-trigger mechanism of the radio signal induced by a cosmic ray, considering real data and the predictor adaptive filter schematic. The D parameter is the time delay in the signal. Adapted from [1].

The adaptive predictor filter uses a delayed version of the signal itself, defined as $d_k = s_{k-D}$ as the desired signal. The input parameter D is a fixed constant chosen to not overlap the input signal and the desired signal, i.e, it is larger than a typical time window of a radio signal. The μ is also an input parameter set dependent on the input signal power, which controls the amount of variation that the weight vector will increase, and k is the loop control variable that will run over the signal s read by the antennas [1].

The behavior of the weight vector can be used as a trigger mechanism because it will vary during a possible transient since the weight vector will vary as it tries to minimize the error. Minimizing the error means attempting to copy the desired signal with the linear combination of the input signal $\mathbf{x}(k)$ and the vector of weights $\mathbf{w}(k)$. In the transient regime, d_k is expected to be a higher value, and $\mathbf{x}(k)$ is expected to be a fluctuation of the background. Therefore, the weight vector will have to change in order to copy the desired signal abruptly. The main advantage of this method is that the behavior of the weight vector spectrum can be more sensitive than the signal power spectrum itself [8].

Note that this mathematical basis is similar to the one used in neural networks, with the main difference of the non-linear component related to the behavior of the neuron. Adaptive filters are usually easily implemented, have a fast computational response, do not need a massive amount of data to be trained, and have similar functionality. The input data is processed in real-time if the data is recorded using an FPGA board. In this sense, an adaptive filter can be implemented online, directly programmed inside the FPGA, and when a radio signal induced by a cosmic ray arrives, it can be triggered almost instantaneously.

4. Analysis

Since the $\mathbf{w}(k) = (w_1, \dots, w_i, \dots, w_M)$ vector shows variational behavior during a transient, we used its local average as an estimator defined as

$$\Delta w_k = \sum_{i=1}^M w_i / M. \quad (5)$$

The global average is defined as

$$\mu_W = \sum_{k=1}^{k_{max}} \Delta w_k / N, \quad (6)$$

and the global width is defined as

$$\sigma_W = \sum_{k=1}^{k_{max}} (\Delta w_k - \mu)^2 / N, \quad (7)$$

As a generic cut, if the local average is larger than $2\sigma_W$: $\Delta w_k > 2\sigma_W$, we consider it as a positive candidate. As the GRAND antenna model has three polarizations, we require a positive candidate in each polarization for an induced radio signal positive detection, although it is a strictly cut. In this sense, using more simulated data, we estimate its efficiency and false positive rate. Efficiency is defined as

$$\text{eff} = n_p / n_{\text{sim}}, \quad (8)$$

where n_p is the number of positive detection and $n_{\text{sim}}=5000$, which is the number of simulated events. The false-positive rate is defined as

$$\text{FPR} = n'_p / n'_{\text{sim}}, \quad (9)$$

where n'_p is the number of positive detection in case the simulated data set does not include any trace inside, i.e., only composed of Gaussian noise over the three polarizations and $n'_{\text{sim}}=100$ for each SNR. In other words, the FPR indicates the number of times, in a completely random distribution set, the trigger was activated although there was not a signal present.

5. Results

The data set used in this analysis is composed of 5000 traces and a single template. We have chosen a single template since in principle, this analysis is template-independent. The efficiency was estimated using the three antenna polarizations (x, y, z), and the detection is given by the combined detection in all three of them. It is assumed that the signal in each polarization should have its peak temporally near to each other. Thus, if three peaks are found in the sample (one for each polarization) and if they are temporally close to each other (in the same time window), then the trigger is activated, i.e., a positive detection occurs. It is important to mention that a signal is not necessary available in all polarizations; e.g. when a shower arrives from the west, the signal will only be north-south (vertical) [4]. This means this method is valid for a subset of arrival directions

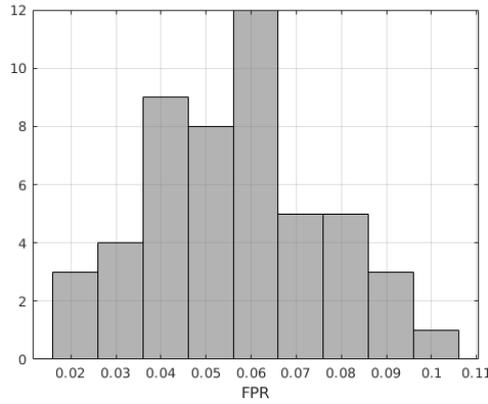


Figure 4: The False Positive Rate (FPR) using the predictor adaptive filter in a complete random distribution data. It is possible to note the very low rate $\approx 6\%$ of most probable value.

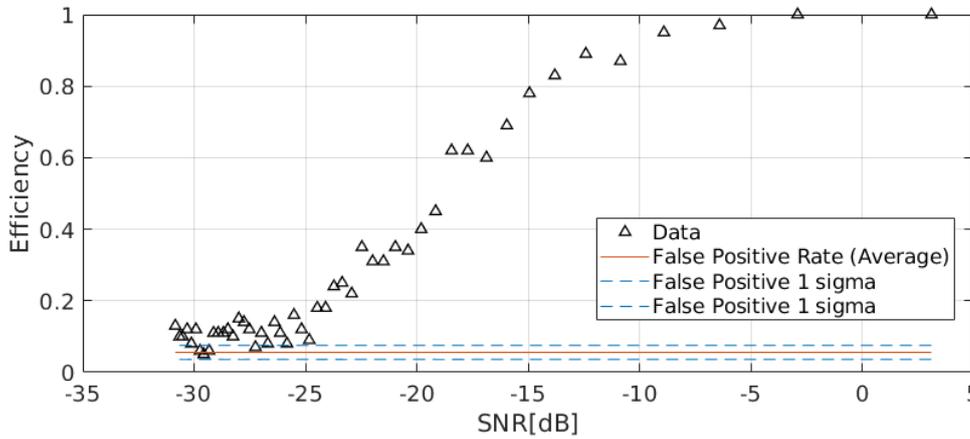


Figure 5: For highly corrupted signal (very low SNR) the efficiency is very low almost compatible with false positive rate. As the SNR increases, also the efficiency increases, and beyond -15dB we expect more that 90% of trigger efficiency.

not quantified in this analysis. Further investigations must be performed concerning the estimation of these limits.

To estimate the amount of the FPR the same analysis was used, however the dataset was only composed of random Gaussian traces also ranging from $1\mu\text{V}$ (5dB) to $50\mu\text{V}$ (-30dB). The FPR distribution is shown in the histogram in figure 4. The detection efficiency of the predictor adaptive filter as a trigger mechanism is presented in figure 5. We can assume that above -15 dB, the efficiency is almost 90%, with a very low and constant FPR.

The FPR was estimated using traces of size N (sample), where N equals to 2082. It may be also expressed in counts per second if we know the sampling rate of the experiment. The sampling rate of the GRAND experiment is 500 MHz [6], which gives a FPR ≈ 15 kHz.

6. Conclusion

The adaptive predictor filter has shown to be a potential and innovative tool for a self-independent trigger mechanism applied in air shower radio signal detection. Using this tool on measured massive data we expect to validate it in a real scenario, also considering different types of background and templates.

7. Acknowledgments

We thank the GRAND Collaboration for allowing us to use the GRAND detector response simulation and the fruitful discussions about this work. The present work was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Brazil, process no. 88887.357727/2019-00 and the Deutscher Akademischer Austauschdienst (DAAD).

References

- [1] Paulo Diniz, *Adaptive Filtering: Algorithms and Practical Implementations*, Springer, 2020..
- [2] Simon Haykin, *Adaptive Filter Theory*, Pearson, 2014.
- [3] Heck, D. and Knapp, J. and Capdevielle, J. N. and Schatz, G. and Thouw, T., CORSIKA: A Monte Carlo code to simulate extensive air showers, 1998, FZKA-6019.
- [4] Tim Huege, Radio detection of cosmic ray air showers in the digital era, *Physics Reports* 620 (2016) 1-52.
- [5] T. Huege, M. Ludwig and C.W. James, *AIP Conf. Proc.* 1535 (2012) 128, arxiv:1301.2132.
- [6] Álvarez-Muñiz, Jaime and Alves Batista, Rafael and others, *The Giant Radio Array for Neutrino Detection (GRAND): Science and design*, *Science China Physics, Mechanics & Astronomy*, 2019.
- [7] International Telecommunication Union, *Radio noise, Recommendation ITU-R P.372-13*, 2016.
- [8] Sobron, I., Diniz, P. S. R., Martins, W. A., and Velez, M., *Energy Detection Technique for Adaptive Spectrum Sensing*, *IEEE Transactions on Communications*, 2015.