Signal recognition and background suppression by matched filters and neural networks for Tunka-Rex

D. Shipilov[1], *P.A.* Bezyazeekov[1], *N.M.* Budnev[1], *D.* Chernykh[1], *O.* Fedorov[1], *O.A.* Gress[1], *A.* Haungs[2], *R.* Hiller[2,⋆], *T.* Huege[2,⋆⋆], *Y.* Kazarina[1], *M.* Kleifges[3], *E.E.* Korosteleva[4], *D.* Kostunin[2], *L.A.* Kuzmichev[4], *V.* Lenok[2], *N.* Lubsandorzhiev[4], *T.* Marshalkina[1], *R.* Monkhoev[1], *E.* Osipova[4], *A.* Pakhorukov[1], *L.* Pankov[1], *V.V.* Prosin[4], *F.G.* Schröder[2,5] and *A.* Zagorodnikov[1] (Tunka-Rex Collaboration)

[1] *Institute of Applied Physics ISU, Irkutsk, Russia*
[2] *Institut für Kernphysik, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany*
[3] *Institut für Prozessdatenverarbeitung und Elektronik, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany*
[4] *Skobeltsyn Institute of Nuclear Physics MSU, Moscow, Russia*
[5] *Department of Physics and Astronomy, University of Delaware, Newark, DE, USA*

Abstract. The Tunka Radio Extension (Tunka-Rex) is a digital antenna array, which measures radio emission of the cosmic-ray air-showers in the frequency band of 30-80 MHz. Tunka-Rex is co-located with the TAIGA experiment in Siberia and consists of 63 antennas, 57 of them are in a densely instrumented area of about 1 km^2. In the present work we discuss the improvements of the signal reconstruction applied for Tunka-Rex. At the first stage we implemented matched filtering using averaged signals as template. The simulation study has shown that matched filtering allows one to decrease the threshold of signal detection and increase its purity. However, the maximum performance of matched filtering is achievable only in case of white noise, while in reality the noise is not fully random due to different reasons. To recognize hidden features of the noise and treat them, we decided to use convolutional neural network with autoencoder architecture. Taking the recorded trace as an input, the autoencoder returns denoised traces, i.e. removes all signal-unrelated amplitudes. We present the comparison between the standard method of signal reconstruction, matched filtering and the autoencoder, and discuss the prospects of application of neural networks for lowering the threshold of digital antenna arrays for cosmic-ray detection.

1 Introduction

Tunka-Rex is an antenna array, which measures the radio emission of air showers produced by cosmic rays with energies above 100 PeV in the frequency band of 30-80 MHz [1]. Tunka-Rex requires an external trigger and operates jointly with the non-imaging air-Cherenkov light detector Tunka-133 [2] and the scintillators of Tunka-Grande [3].

⋆ now at the University of Zürich
⋆⋆ also at Vrije Universiteit Brussel, Brussels, Belgium

The main background at the Tunka-Rex location is the Galaxy. However there are many sources of non-white and non-stationary background in the Tunka Valley. Due to this we use two different approaches: matched filter with predefined signal template and neural network with optimized convolutional filters.

2 Matched filter and autoencoder for signal reconstruction

In the present work we use 650 000 samples of measured Tunka background and 25 000 CoREAS simulations folded with Tunka-Rex hardware response. We use single polarization ($v \times B$) and the following upsampling rates: 64 for matched filter and 16 for neural network. The true peak position in the simulated trace is defined as a position of the maximum of the Hilbert envelope of the trace before adding the background [1].

A matched filter (MF) convolutes a template with the input trace and the maximum of the convolution defines the position of the peak. Templates are obtained from averaging of many CoREAS[4] simulations. In the present work we use templates with lengths of 60 ns, upsampled to 800 MS/s rate (see Fig. (1)). The threshold is defined as 5% probability of false positive. Amplitude is estimated as the function of the square root of cross-correlation. We have implemented the MF in the Auger Offline [5] and tested it on the set of simulated events. The MF is able to reconstruct pulses with lower amplitudes and features resolution of arrival direction similar to standard method. The distribution of reconstructed events and arrival directions can be seen in Fig. (1). Here it is worth noticing, that we did not perform any corrections and whitening of our noise, by this the performance of the MF does not reach it full efficiency.

At the next step we use neural network which called autoencoder (AE). AE is based on 1D convolutional layers with rectified linear units and max pooling after convolution layers. Binary cross-entropy is used as a loss function. For minimization of the loss, all data should be normalized in [0;1] range and baseline always should be put on a 0.5 level, what helps the AE to extract features from noise.

The structure of the AE is defined by the following: depth (D) and number of filters per layer (N) are free parameters. The i-th encoding layer ($i = 1, ..., D$) is described by the following:

$$S_i = S_{\min} \times 2^{D-i} \,, \quad n_i = 2^{i+N-1} \,, \tag{1}$$

where S_i is size of the i-th filter, n_i is a number of filters per layer. $S_{\min} = 16$ is minimal size of layer (corresponding to few ns).

To assess the quality of the networks we have introduced 2 metrics: efficiency: $N_{\text{rec.}}/N_{\text{tot.}}$, namely fraction of events that passed the threshold; and the purity: $N_{\text{hit}}/N_{\text{rec.}}$, namely the fraction of events with reconstructed position of the peak $|t_{\text{rec.}} - t_{\text{true}}| < 5$ ns. Networks with 3-4 layers show similar result, only addition of the 5th layer increases of the purity. Networks with 5 layers have a large number of degrees of freedom which leads to overfitting with our limited size of the training dataset.

3 Summary

We have improved Tunka-Rex signal reconstruction by implementing matched filters and autoencoders. New methods show promising results in the lowering of the threshold. We will improve current results by extending the library of templates for matched filter and optimization of autoencoder architectures.
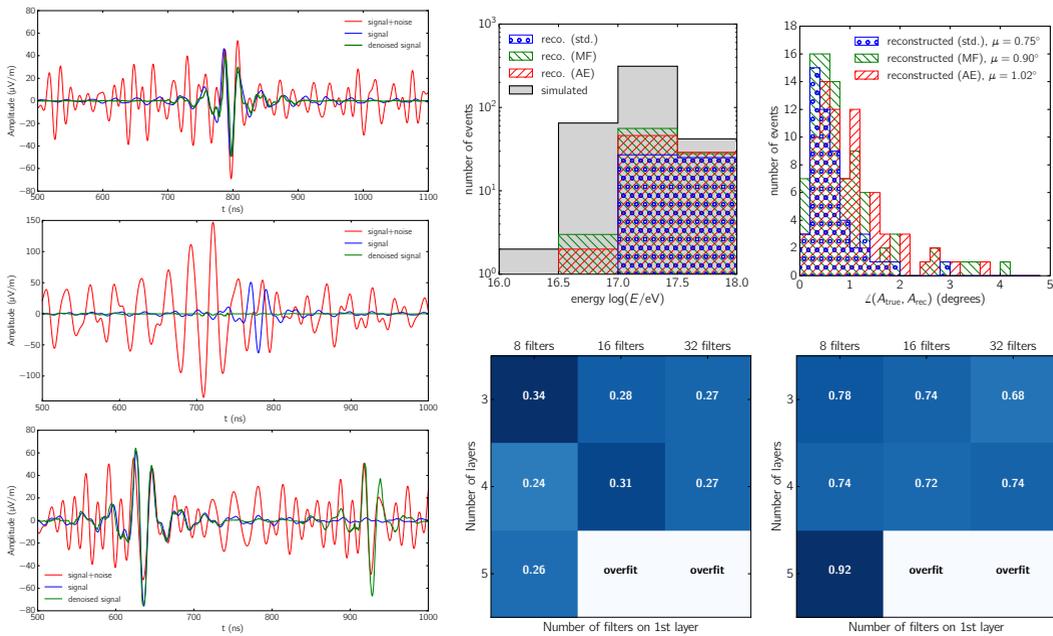
Figure 1. Left panel, examples of AE performance: correct identification (true positive), no identification (true negative), double identification (true plus false positive). Right panel, from the left upper conner to the bottom right: 1) distribution of reconstructed events as function of energy; 2) angular resolution of different methods; 3) efficiency of AE; 4) purity of AE.

Acknowledgements

References

[1] P. A. Bezyazeekov *et al.*, Nucl. Instrum. Meth. A **802**, 89 (2015)
[2] V. V. Prosin *et al.*, Nucl. Instrum. Meth. A **756**, 94 (2014)
[3] N. M. Budnev *et al.*, Bull. Russ. Acad. Sci. Phys. **79**, 395 (2015) [Izv. Ross. Akad. Nauk Ser. Fiz. **79**, no.3, 430 (2015)]
[4] T. Huege *et al.*, AIP Conf. Proc. **1535**, 128 (2013), arXiv:1301.2132 [astro-ph.HE]
[5] P. Abreu *et al.*, Nucl. Instrum. Meth. **A635**, 92 (2011), arXiv:1101.4473 [astro-ph.IM]