

Reconstruction of sub-threshold events of cosmic-ray radio detectors using an autoencoder

**P. Bezyazeekov,^{a,*} D. Shipilov,^b D. Kostunin,^c I. Plokhikh,^d A. Mikhaylenko,^d
P. Turishcheva,^e S. Golovachev,^f V. Sotnikov^f and E. Sotnikova^g**
for the Tunka-Rex Collaboration

(a complete list of Tunka-Rex authors can be found at the end of the proceedings)

^a*Applied Physics Institute, Irkutsk State University, 664020 Irkutsk, Russia*

^b*X5 Retail Group, Moscow, 119049 Russia*

^c*DESY, 15738 Zeuthen, Germany*

^d*Novosibirsk State University, 630090 Novosibirsk, Russia*

^e*Innopolis University, 420500 Innopolis, Russia*

^f*JetBrains Research, 194100 St. Petersburg, Russia*

^g*Sobolev Institute of Mathematics, 630090 Novosibirsk, Russia*

E-mail: contact@tunkarex.info

Radio detection of air showers produced by ultra-high energy cosmic rays is a cost-effective technique for the next generation of sparse arrays. The performance of this technique strongly depends on the environmental background, which has different constituents, namely anthropogenic radio frequency interference, synchrotron galactic radiation and others. These components have recognizable features, which can help for background suppression. A powerful method for handling this is the application of convolution neural networks with a specific architecture called autoencoder. By suppressing unwanted signatures, the autoencoder keeps the signal-like ones. We have successfully developed and trained an autoencoder, which is now applied to the data from Tunka-Rex. We show the procedures of the training and optimization of the network including benchmarks of different architectures. Using the autoencoder, we improved the standard analysis of Tunka-Rex in order to lower the threshold of the detection. This enables the reconstructing of sub-threshold events with energies lower than 0.1 EeV with satisfactory angular and energy resolutions.

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

*Presenter

1. Introduction

Digital radio arrays became mature technique for the air-shower detection in last years [1]. The engineering arrays, e.g. Tunka-Rex, has shown the ability of energy spectrum [2] and mass composition reconstruction [3] as well as was used for the independent cross-check of energy scale of Tunka-133 and KASCADE-Grande [4]. After decommission of Tunka-Rex in 2019 we focus on the open data [5] and development of techniques for the future arrays, e.g. aperture estimation and self-triggering for radio [6, 7]. In this work we continue developing sub-threshold reconstruction using deep learning started in Ref. [8].

The efficiency of the method of reconstruction the primary particle parameters by measured radio emission, initiated with related air-shower, is sensitive to background characteristics. High-level background may distorts the air-shower pulses up to impossibility of reconstruction in case of low amplitude pulses. Standard methods of radio reconstruction uses the absolute amplitudes of air-shower pulses as main low-level characteristic of measured data. Reconstruction of primary particle parameters is performed by analysis of lateral distribution of measured amplitudes. For reconstruction the absolute amplitude of measured pulse in background conditions one can use analytical parametrization of mean background amplitude for reconstruction the amplitude of air-shower pulse, which works well in case of white noise and can be useful in some cases. However, radio background in real conditions is non-white and have specific features. Convolutional neural networks are effective tools for solving such type of tasks.

We show our approach to solve the background problem with autoencoder based on convolutional neural networks. Autoencoder is tuned to extract background-related characteristics from the measured radio data, and remove them from the signal traces leaving only air-shower pulses. We use the peak times of denoised signals and reconstruct the lateral distribution for further reconstruction of energy of primary particle.

2. Autoencoder architecture

Autoencoder is an architecture of artificial neural network, which codes the input data to abstract representation and after decodes it to as similar to input as possible. It can be used for problems of dimensionality reduction, denoising, anomaly detection and others. Structurally autoencoder consists of two parts: encoder and decoder. The encoder consists of a set of layers with decreasing number of neurons at each next layer and produces compressed representation of original input or the code. After encoding decoder starts to work. In typical case, the decoder has the same structure as encoder but in opposite orientation. The goal of decoder is to reconstruct original input from the code as close as possible. As we have small-size layer in the center of the network, some part of the information about input data is inevitably lost.

Denoising autoencoder is specific architecture which is proposed to removing noise from the data. During training of this type of autoencoder, it takes the noised data at input, and after processing compares output with the same data, but without noise. This way it tunes to remove noise-related information and returns only useful data.

As we have data in one-dimensional format (signal trace of measured amplitude of electric field), for this study we use one-dimensional architecture of autoencoder. For defining the optimal

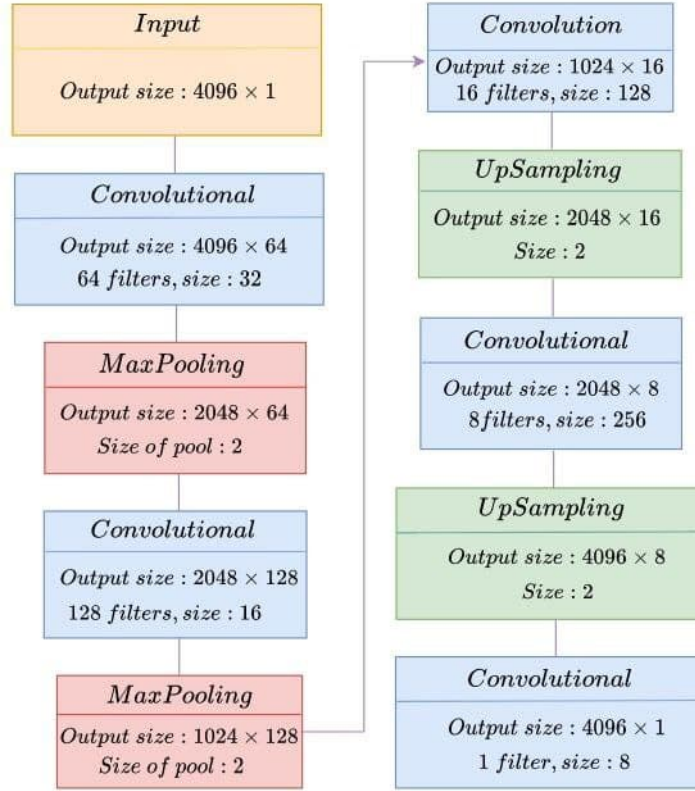


Figure 1: Autoencoder architecture

configuration for reconstruction the air-shower pulse we tested a number of autoencoder configurations with various number of layers and filters, size of filter, loss function etc. For preliminary estimation of the efficiency of autoencoder we use binary cross entropy and visual analysis of performance results. Detailed view of chosen architecture is shown in Fig 1. Loss function is ReLU (rectified linear unit). Training of AE was performed with Tensorflow+Keras on GPU server using uDocker virtualisation.

3. Denoising of radio data with autoencoder

For training the autoencoder we prepared the dataset consisting of 25000 CoREAS [9] simulations based on CORSIKA software [10] and library of background measured at Tunka-Rex consisting of 650000 samples. Training dataset is formed by adding simulated air-shower pulses with background traces. The air-shower pulse is located around the fixed position with random shift with range of ± 50 ns. Each sample contains one reconstructed polarization of measured/simulated electric field. Before processing with autoencoder traces are upsampled with factor 16 (0.3125 nanoseconds per count instead of 5 nanoseconds per count in raw data). Length of the trace is 4096 counts, which corresponds to 1.28 microseconds. For correct performance of autoencoder we normalise the amplitude of each trace to [0:1] range. Due to this we can not reconstruct the amplitude directly for very low signals (where signal-to-noise ratio close to unity) and develop a special procedure for reconstruction which is described in detail in next section.

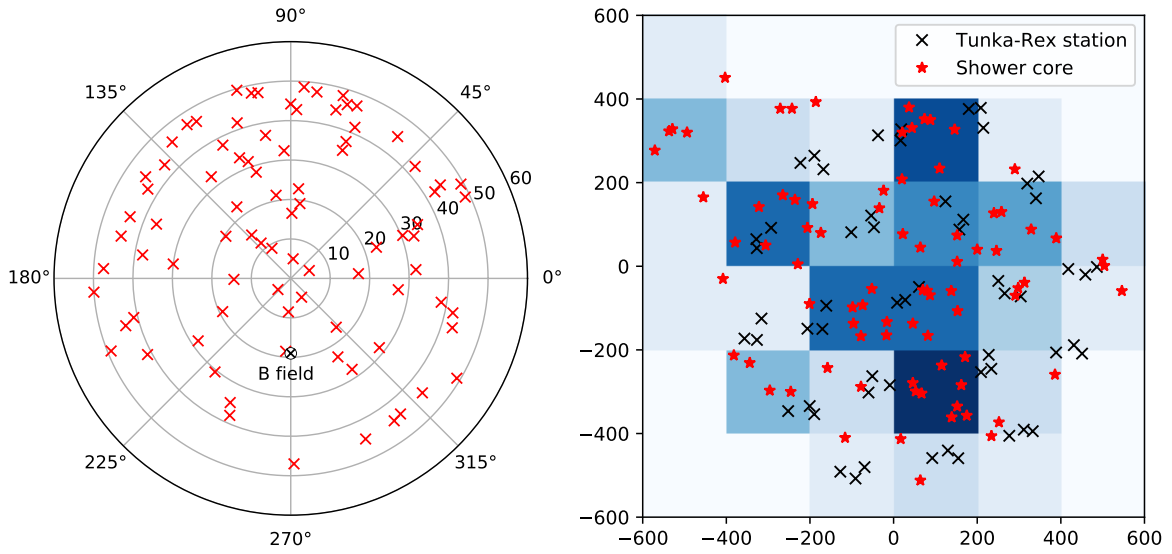


Figure 2: Distribution of events detected by autoencoder before quality cuts. *Left:* arrival directions of detected events. One can see geomagnetic asymmetry, which indicated that the significant fraction of events are true positives. *Right:* distribution of the cores of detected events taken from trigger, Tunka-133. No hotspots are obtained, which shows that autoencoder does not trigger on known local sources of the noise (e.g. central DAQ or power lines).

4. Reconstruction of real data

In the frame of this work we aim at the detection of sub-threshold events not accessible with standard reconstruction. We have selected events triggered by Tunka-133 and which have energy reconstructed by Tunka-133 in between 10^{16} and 10^{17} eV. Since the energy threshold of Tunka-Rex starts from about 10^{17} eV this set has almost no overlap with standard Tunka-Rex reconstruction. We decreased the internal threshold of autoencoder, which was set to value corresponding to 5% probability of false positive for single signal. Since we used normalization, the autoencoder returns amplitudes in the range of $[-0.5; 0.5]$ and we reduce threshold from 0.395 to 0.2.

Fig. 2 shows detected events passed standard quality cuts (difference between Tunka-133 and Tunka-Rex arrival directions less than 5° , antennas are clustered, etc.). The distributions of arrival directions and core positions of detected events indicate significant fraction of true positives, and we introduce additional quality cuts and algorithms, which allows us not only to reject false positives, but also reconstruct primary energy.

We developed so-called *adaptive LDF*¹ algorithm exploiting the idea of coherent production of radio emission during air-shower development (see Fig. 3) This algorithm sums radio traces using peak position reconstructed by autoencoder and synthesises effective amplitude at effective distance from the shower core. Assuming uncorrelated background the signal-to-noise ratio (SNR) of synthesised signal should have increase by factor of \sqrt{N} , where N is a number of original signals. This feature is used for detection of the false positives and exclusion them from the analysis, we add additional cut $\text{SNR} > 16$ for the synthesised signals.

¹Lateral distribution function (LDF) is a distribution of amplitudes projected on shower axis.

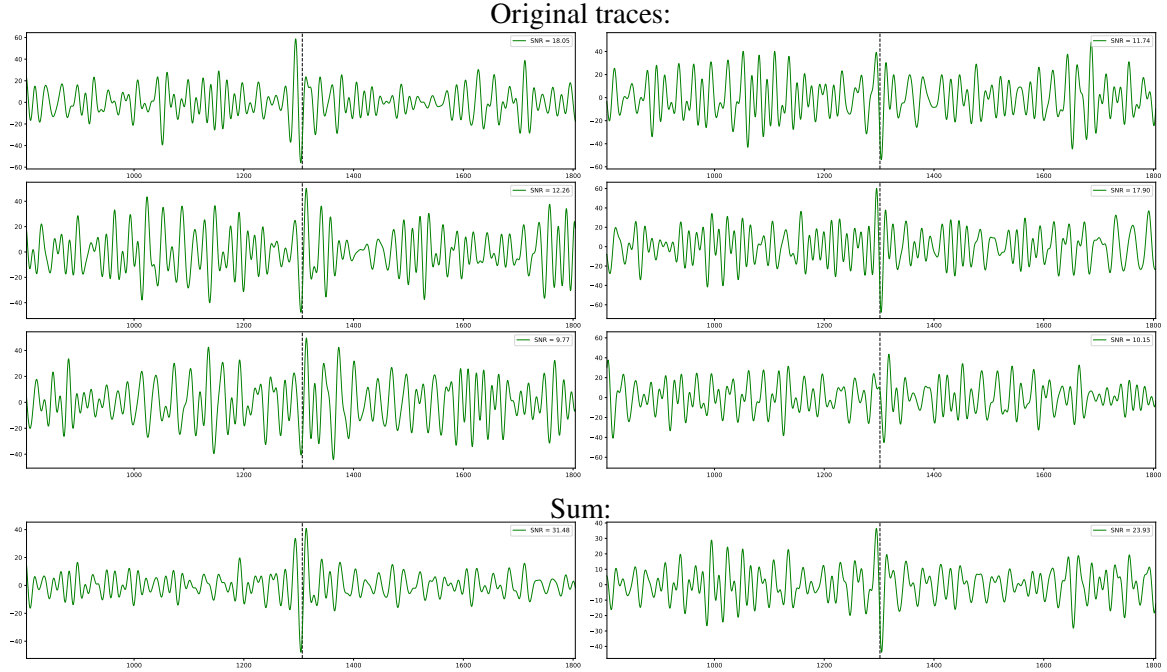


Figure 3: Example of two events with $E = 30$ PeV detected with autoencoder. Dashed lines indicate the reconstructed peak position.

Adaptive LDF does not depend on the time resolution of the detector and requires only position of the radio pulse in the trace. Coherent sum features its maximal performance for the signals spatially close to each other, since they are produced in the same part of air-shower and their shapes are almost identical.

Left panel of Fig. 4 shows the synthesised signals after processing with adaptive LDF algorithm. One can note that adaptive LDF remove significant fraction of detected events, since synthesised signals do not reach sufficient level of signal-to-noise ratio (SNR). The latest step is the energy reconstruction. We have already developed algorithms for single antenna energy reconstruction [11], which uses parametrization based on average LDF. We readjusted its parameters using LDF slope fitted from reconstructed events (see right panel of Fig. 4) and normalization from Tunka-Rex software corresponding to the actual calibration. This way, primary energy E_{pr} is reconstructed using following:

$$E_{pr} = \kappa E_{ALDF} \exp [(d_{ALDF} - d_0)\eta] / \sin \alpha, \quad (1)$$

where $\kappa = 868$ EeV/(V/m), $d_0 = 180$ m, fitted $\eta = 0.00228$ and α is geomagnetic angle (angle between shower axis and magnetic field of the Earth).

Fig. 5 shows the performance of the method for the energy reconstruction. One can see that we have successfully reconstructed energy and arrival directions for two events with energies of 30 PeV which is the record value for the Tunka-Rex.

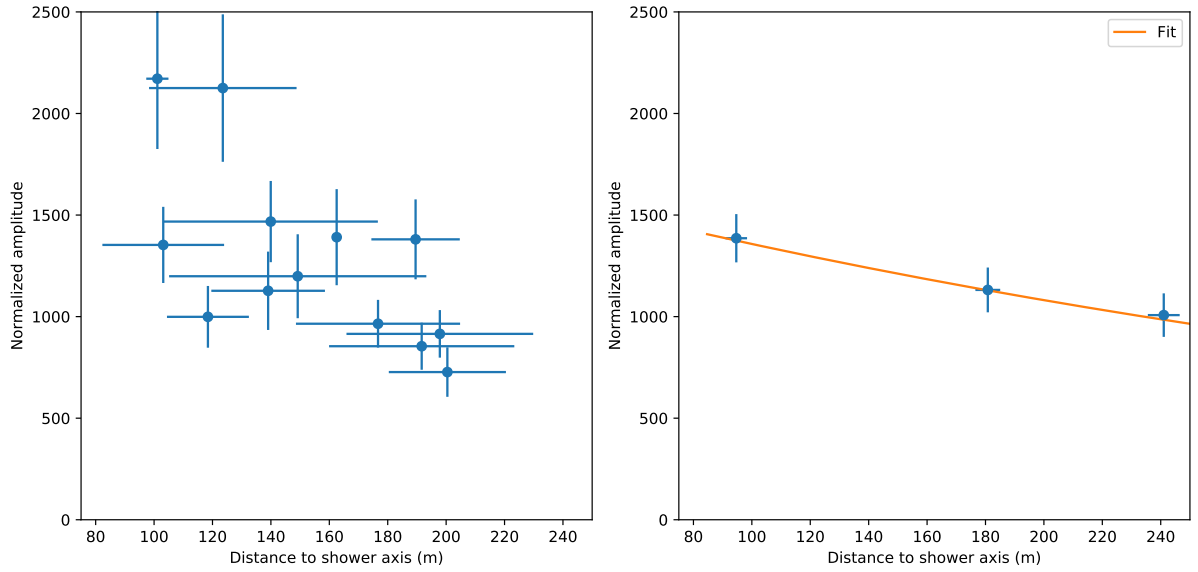


Figure 4: *Left:* The distribution of synthesised signals, which passed adaptive LDF processing. Error bars include uncertainties introduced by synthesis procedure. *Right:* Average LDF for all detected events. Orange curve indicated fit with exponential function, slope of which is used for the energy reconstruction.

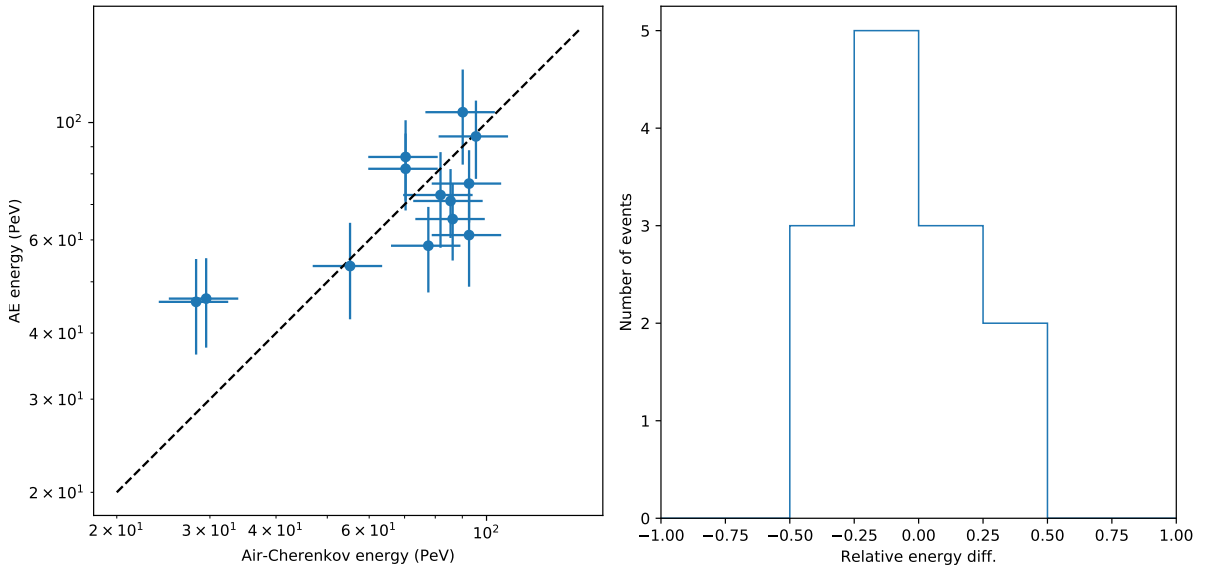


Figure 5: Performance of energy reconstruction obtained for sub-threshold events. *Left:* cross-check with Tunka-133 reconstruction for the same events. *Right:* deviation between Tunka-133 and Tunka-Rex reconstruction showing the Tunka-Rex resolution of 26%.

5. Conclusion

We have reconstructed real data measured by the cosmic-ray radio detector using a deep neural network for a first time. The algorithms developed in this works allow us to reconstruct events not detectable with standard Tunka-Rex methods obtained high angular and energy resolution. The chosen architecture (autoencoder) and training pipeline (COREAS traces mixed with real

background) show the efficiency of our approach. The developed methods show high performance for the detection of the signal, which drives us to propose the following application of this technique:

- Sub-threshold detection and reconstruction of the signals at the tails of the LDF with low SNR is accessible with combination of autoencoder and adaptive LDF. As shown in earlier studies [12, 13], the slope of the LDF characterized by its tails is sensitive to the position of shower maximum, which is a key for the identification of primary particle.
- Trigger based on the compressed autoencoder integrated on FPGA [14] is a perspective technique for the implementation of the self-triggered detector for the next generation of digital radio arrays [15, 16]. We plan to test this approach in the nearest future.

The third-party code based on our software [17] has shown its efficiency for the broader frequency range, what indicates the universality of the approach. Our software is published under free license² and we plan its further improvement. One can find the example of usage of this software in Jupyter-Hub at IAP KIT³. Further read can be also found in Refs. [8, 18, 19]. The autoencoder was also included in educational project at the first workshop of Mathematical center in Akademgorodok⁴.

Acknowledgements

The authors would like to express gratitude to the colleagues from KCDC team. The development and testing of the software was supported by the state contract with Institute of Thermophysics SB RAS.

References

- [1] Schröder, Frank G., *Radio detection of Cosmic-Ray Air Showers and High-Energy Neutrinos*, *Prog. Part. Nucl. Phys.* **93** (2017) 1 [1607.08781].
- [2] TUNKA-REX collaboration, *Seven years of Tunka-Rex operation*, *PoS ICRC2019* (2020) 319 [1908.10305].
- [3] TUNKA-REX collaboration, *Reconstruction of cosmic ray air showers with Tunka-Rex data using template fitting of radio pulses*, *Phys. Rev.* **D97** (2018) 122004 [1803.06862].
- [4] TUNKA-REX, LOPES collaboration, *A comparison of the cosmic-ray energy scales of Tunka-133 and KASCADE-Grande via their radio extensions Tunka-Rex and LOPES*, *Phys. Lett.* **B763** (2016) 179.
- [5] TUNKA-REX collaboration, *Tunka-Rex Virtual Observatory*, *PoS ICRC2021* (2021) 421.
- [6] TUNKA-REX collaboration, *Current Status and New Challenges of the Tunka Radio Extension*, *J. Phys. Conf. Ser.* **1181** (2019) 012027 [1812.06893].

²<https://gitlab.ikp.kit.edu/tunkarex/denoiser>

³tutorials/jbr/trvo/trvo_icrc.ipynb at <https://jupyter.iap.kit.edu>

⁴<https://english.nsu.ru/mca/media/news/2973965/>

- [7] TUNKA-REX collaboration, *Estimation of aperture of the Tunka-Rex radio array for cosmic-ray air-shower measurements*, *PoS ICRC2021* (2021) 210.
- [8] TUNKA-REX collaboration, *Signal recognition and background suppression by matched filters and neural networks for Tunka-Rex*, *ARENA 2018 proceedings* (2018) [1812.03347].
- [9] T. Huege, M. Ludwig and C. James, *Simulating radio emission from air showers with CoREAS*, *AIP Conf.Proc.* **1535** (2013) 128 [1301.2132].
- [10] D. Heck et al., *CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers*, FZKA Report 6019, Forschungszentrum Karlsruhe (1998).
- [11] TUNKA-REX collaboration, *Tunka-Rex: energy reconstruction with a single antenna station*, *EPJ Web Conf.* **135** (2017) 01004 [1611.09614].
- [12] D. Kostunin, P.A. Bezyazeev, R. Hiller, F.G. Schröder, V. Lenok and E. Levinson, *Reconstruction of air-shower parameters for large-scale radio detectors using the lateral distribution*, *Astropart. Phys.* **74** (2016) 79 [1504.05083].
- [13] TUNKA-REX collaboration, *Towards a cosmic-ray mass-composition study at Tunka Radio Extension*, *EPJ Web Conf.* **135** (2017) 01005 [1611.09127].
- [14] J. Duarte et al., *Fast inference of deep neural networks in FPGAs for particle physics*, *JINST* **13** (2018) P07027 [1804.06913].
- [15] C. Xie et al., *Improving Radio Frequency Detectors using High Performance Programmable Logic*, *PoS ICRC2021* (2021) 1028.
- [16] P. Bezyazeev et al., *Efficiency estimation of self-triggered antenna clusters for air-shower detection*, *PoS ICRC2021* (2021) 273.
- [17] A. Rehman et al., *Classification and Denoising of Cosmic-Ray Radio Signals using Deep Learning*, *PoS ICRC2021* (2021) 417.
- [18] P. Bezyazeev et al., *Advanced Signal Reconstruction in Tunka-Rex with Matched Filtering and Deep Learning*, in *3rd International Workshop on Data Life Cycle in Physics*, 6, 2019 [1906.10947].
- [19] P. Bezyazeev et al., *Reconstruction of radio signals from air-showers with autoencoder*, *MDPI Proc. L* **2679** (2020) 43 [2101.02943].

Full Authors List: Tunka-Rex Collaboration

P. Bezyazeev¹, N. Budnev¹, O. Fedorov¹, O. Gress¹, O. Grishin¹, A. Haungs², T. Huege^{2,3}, Y. Kazarina¹, M. Kleifges⁴, E. Korosteleva⁵, D. Kostunin⁶, L. Kuzmichev⁵, V. Lenok², N. Lubsandorzhev⁵, S. Malakhov¹, T. Marshalkina¹, R. Monkhoev¹, E. Osipova⁵, A. Pakhorukov¹, L. Pankov¹, V. Prosin⁵, F. G. Schröder^{2,7}, D. Shipilov⁸ and A. Zagorodnikov¹

¹Applied Physics Institute ISU, Irkutsk, 664020 Russia

²Karlsruhe Institute of Technology, Institute for Astroparticle Physics, D-76021 Karlsruhe, Germany

³Astrophysical Institute, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

⁴Institut für Prozessdatenverarbeitung und Elektronik, Karlsruhe Institute of Technology (KIT), Karlsruhe, 76021 Germany

⁵Skobel'syn Institute of Nuclear Physics MSU, Moscow, 119991 Russia

⁶DESY, Zeuthen, 15738 Germany

⁷Bartol Research Institute, Department of Physics and Astronomy, University of Delaware, Newark, DE, 19716, USA

⁸X5 Retail Group, Moscow, 119049 Russia