Legacy of Tunka-Rex software and data

P. Bezyazeekov for the Tunka-Rex collaboration

API ISU, Russia

28 June 2021

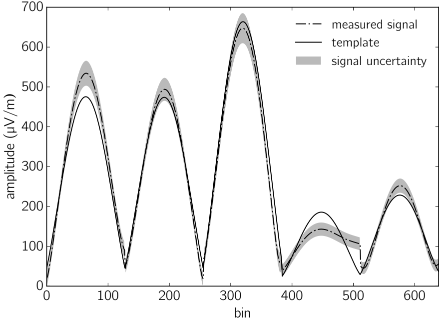
Tunka Radio Extension

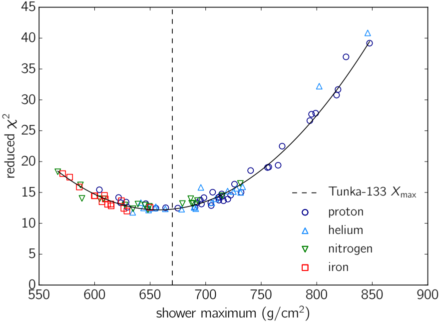
- Frequency band 30-80 MHz
- Duplex and triplex measurements: $\gamma_{\text{ch}}/\mu/e$ + radio
- World-uniqe radio and air-Cherenkov cross-calibration
- First direct measurement of shower maximum with radio
- Cost-effective
- **Proof-of-feasibility**

- Template fitting reconstruction
- Autoencoder denoising
- Efficiency model
- Self-trigger study
- Tunka-Rex Virtual Observatory

TEMPLATE FIT

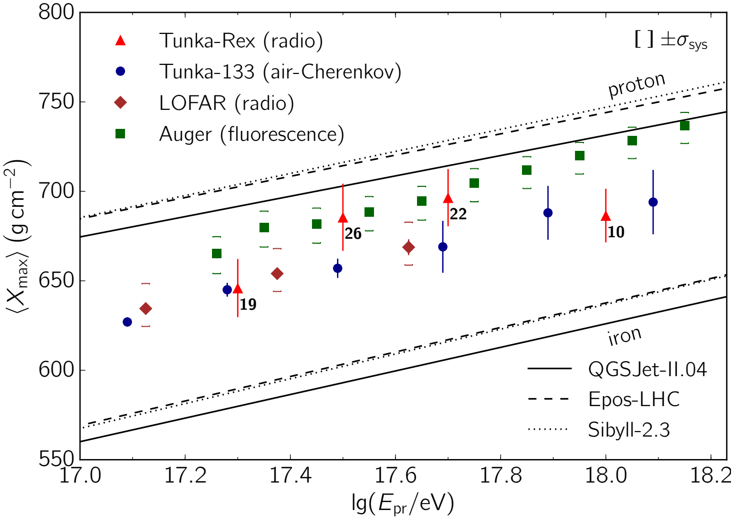
Chi-square fit of optimally-clipped envelopes concatenated to single trace





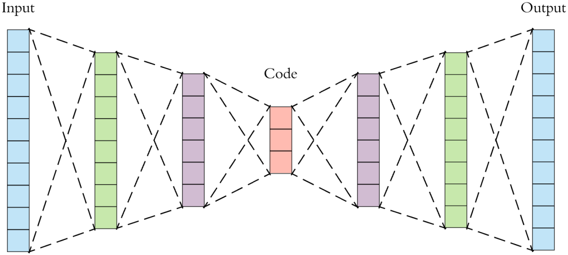
$$U(t) = \bigoplus_{i=1}^{N_a} u_i(t')\Pi((t_i - t')/t_w)\,; \quad \chi^2_{\text{red.}} = \sum_{j=1}^{N_b} \left(\frac{U_j - AV_j}{\sigma_j} \right)^2 \cdot \frac{f_{\text{ups.}}}{N_{\text{bins}}}$$

Template fit: results [Phys.Rev. D 97, 122004]

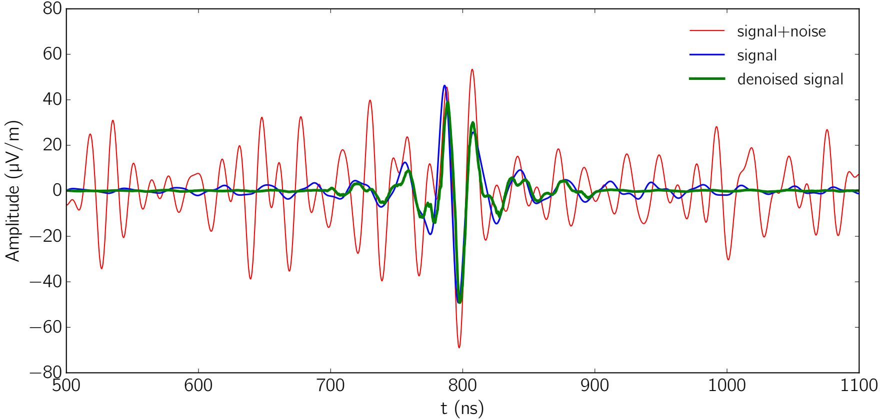


AUTOENCODER

- Unsupervised neural network with compressed representation
- Use Keras and Tensorflow with GPU support
- Based of 1D convolution layers
- ReLu ($\max(0, x)$) activation function
- Max pooling (and upsampling) after convolutional layers
- Binary crossentopy loss function and RMSprop optimizer
- Train networks via uDocker on SCC ForHLR II cluster



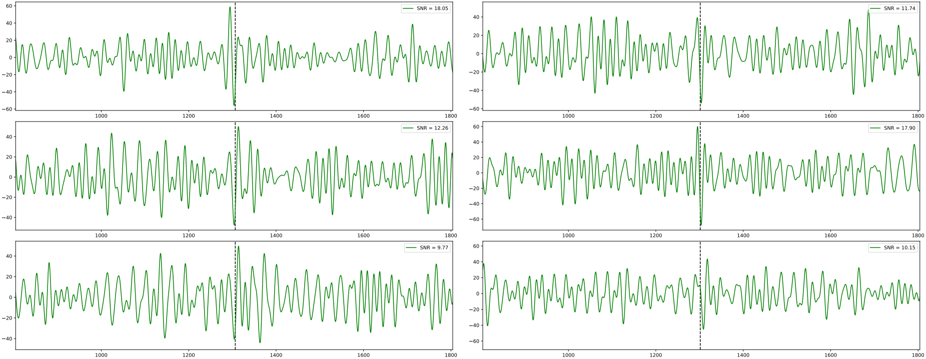
Example of correct identification



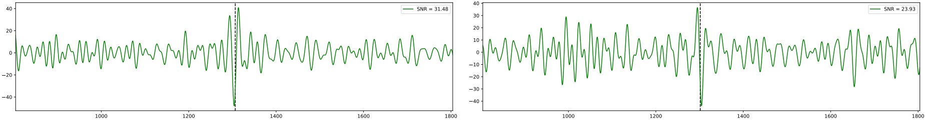
True signal and noise are identified correctly, noise is removed

Energy reconstruction by denoised data: method

Two example events with $E = 30$ PeV

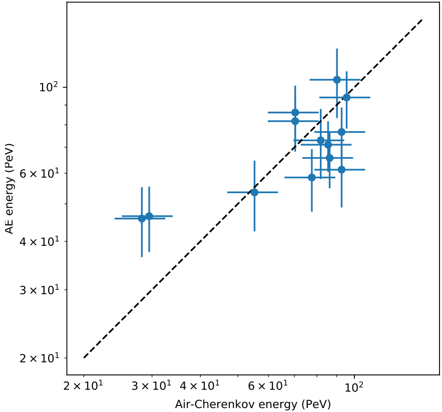


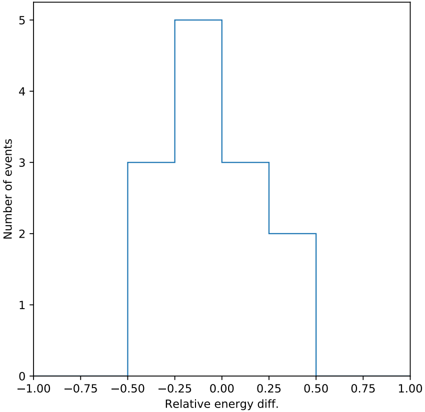
Coherent sum:



Energy reconstruction (after cuts)

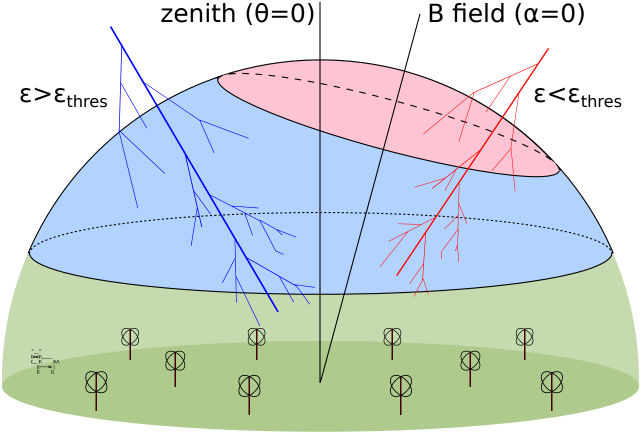
Reconstruction based on single antenna method, $E = \kappa A_d e^{-\eta(d-d_0)}$
Normalization factor from standard reconstruction; $\mu = 0\%$, $\sigma = 26\%$





EFFECTIVE APERTURE

Effective aperture model



$$\mathcal{A}(E) = \int_\Omega \int_S \varepsilon(E,\theta,\alpha,x,y)\cos\theta\,\mathrm{d}S\,\mathrm{d}\Omega$$

$$\varepsilon(E,\theta,\alpha,x,y) = \varepsilon_R(E,\theta,\alpha,x,y)\,\varepsilon_a(E,\theta,\alpha)$$

Comparison with experimental data

Gen.	Year	Antennas number	Expected events	Detected events	Efficiency
1a	2012/13	18	23	20	$0.85^{+0.05}_{-0.09}$
1b	2013/14	25	28	27	$0.96^{+0.02}_{-0.05}$
2	2015/16	44	14	14	$1.00^{+0.00}_{-0.07}$
3	2016/17	63	17	16	$0.94^{+0.04}_{-0.08}$
		Total	82	77	$0.94^{+0.02}_{-0.03}$

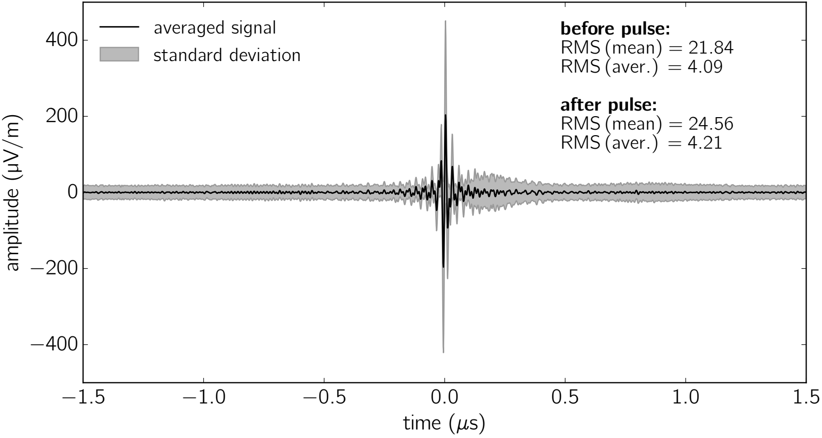
TRVO

- Over 1 million of measured and simulated events
- Modular interface
- Preparing to open access

- Developed techiques may be used in various radio experiments
- Energy resolution of 10%
- Shower maximum resolution of 25-35 g/cm^2
- Efficiency model is in agreement with experimental data
- Studies of autoencoder and self-trigger continues
- Preparing TRVO for open access

THANK YOU FOR ATTENTION

Autoencoder: motivation



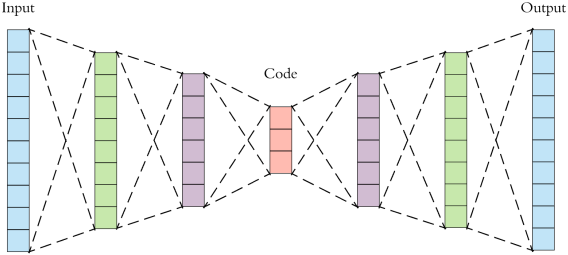
Average of 400 events, expected noise reduction with factor $\sqrt{400} = 20$

$\Rightarrow$ Noise is not white/contain features

$\Rightarrow$ Train autoencoder to learn these features

Chosen architecture (autoencoder)

- Unsupervised neural network with compressed representation
- Use Keras and Tensorflow with GPU support
- Based of 1D convolution layers
- ReLu ($\max(0, x)$) activation function
- Max pooling (and upsampling) after convolutional layers
- Binary crossentopy loss function and RMSprop optimizer
- Train networks via uDocker on SCC ForHLR II cluster



Learning strategy and training pipeline

Datasets:

- 25k upsampled ($\times 16$) traces with real background + low-amplitude simulations ($< 100\,\mu$V/m) with randomly located pulse

Training and evaluation:

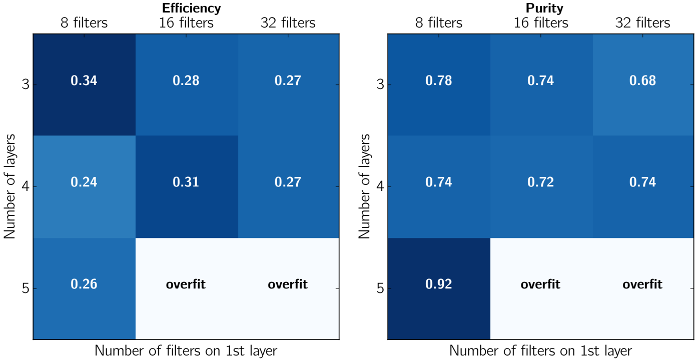
- Depth (D) and number of filters per layer as free parameters
- Primary evaluate by loss metrics
- Blind test with full-pipeline Offline reconstruction

i-th encoding layer is described by the following ($i = 1, ..., D$):

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

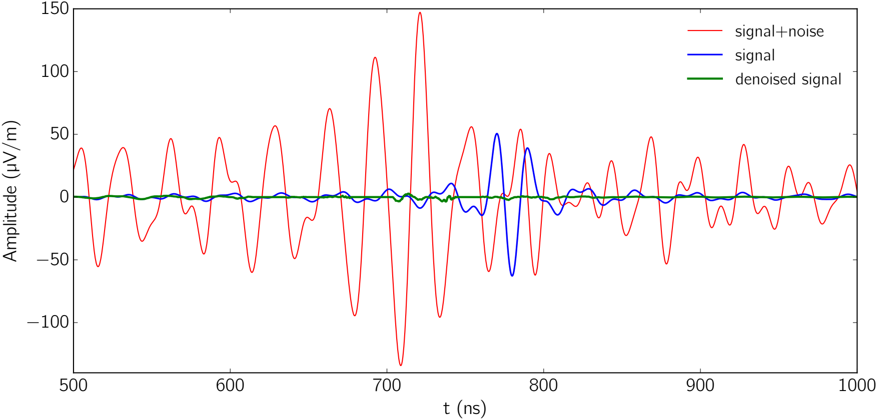
where S_i is a size of the i-th filter, n_i is a number of filters per layer
D and N are free parameters; $S_{\min} = 16$ is minimal size of layer
Size of input/output array: 4096 (1280 ns) – 25% of original trace

- Threshold amplitude $\Leftrightarrow$ 5% tolerance to false positives
- Efficiency: $N_{\mathrm{rec.}}/N_{\mathrm{tot.}}$, fraction of events passed the threshold
- Purity: $N_{\mathrm{hit}}/N_{\mathrm{rec.}}$, fraction of events with reconstructed position of the peak: $|t_{\mathrm{rec.}} - t_{\mathrm{true}}| < 5$ ns



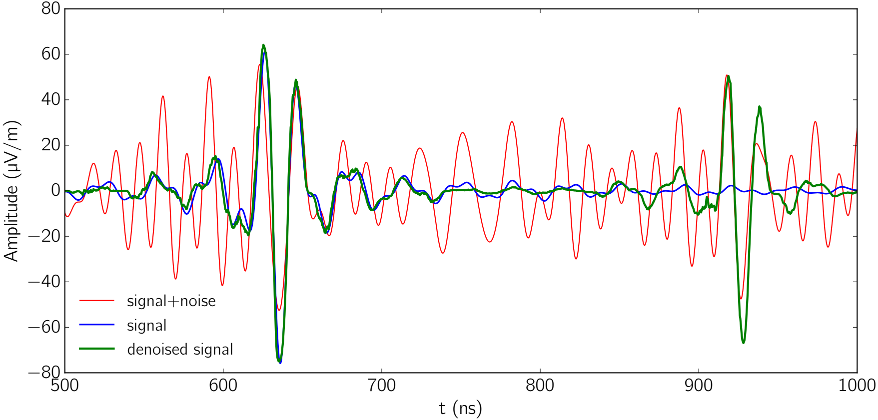
Best architecture contains $N_{\mathrm{dof}} = 10240$

Example: no identification



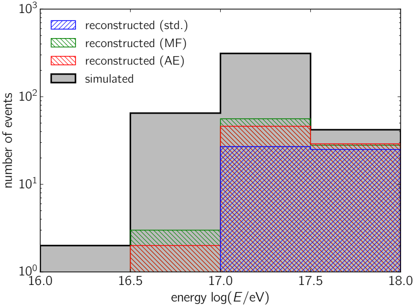
True signal is heavily distorted by noise, and removed as background

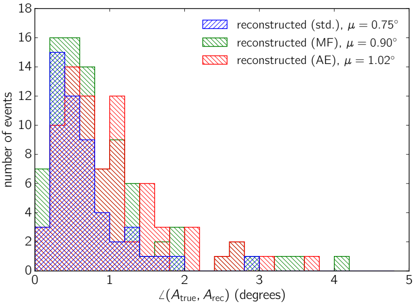
Example: double identification



Signal-like RFI is identified as signal

Autoencoder is binded with Tunka-Rex fork of Auger Offline
Reconstruction of CoREAS simulations (reproduction of 2012-2014 events)





- Tunka-133/Tunka-Rex events with $E \in [10^{16} - 10^{17}]$ eV
- Almost zero events in this energy band by standard method
- Decreasing autoencoder threshold $0.395/0.500 \rightarrow 0.200/0.500$
- Cross-check cuts: direction reconstruction $\Delta\Omega < 5°$, clustering events