Latest Tunka-Rex activity towards deep learning techniques and open data

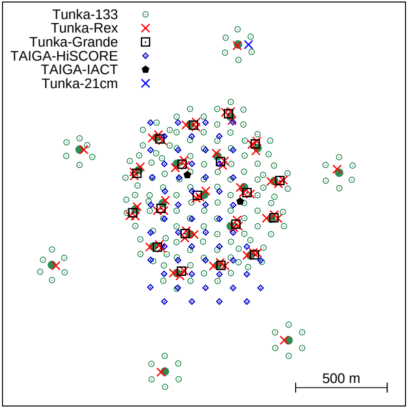
Pavel Bezyazeekov for the Tunka-Rex Collaboration

Applied Physics Institute, Irkutsk State University, Russia

February 19, 2018

Tunka-Rex

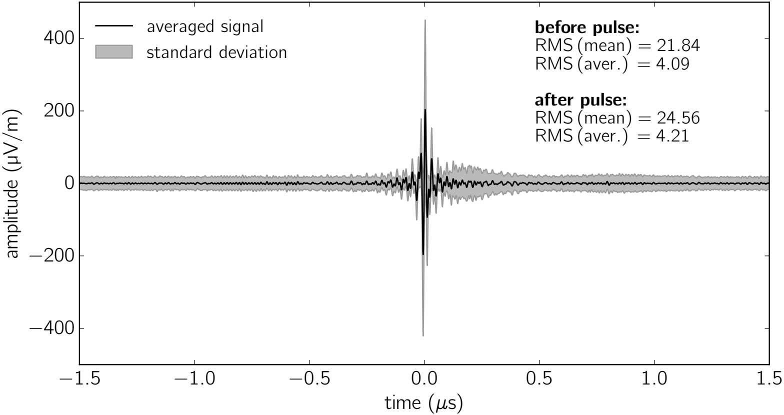
- 12 bit / 200 MHz signal digitalization
- 30-80 MHz radio band
- 10-40 m distance between antennas
- 57 antennas / 1km^2



DENOISING USING AUTOENCODER

Deep learning: motivation

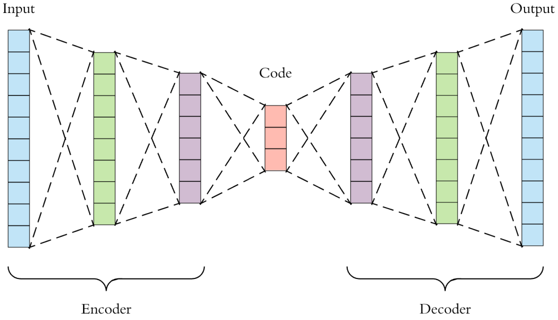
- Extract features of background
- Denoising traces



Average of 400 events; RMS should reduce by factor of 20 ⇒ Noise not white

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



Simulation set

- 650k samples of Tunka background recorded in 2014-2017
- CoREAS simulations of Tunka-Rex signals (25k samples)
- Pulse is randomly located inside signal window (200 ns)
- Using single polarization ($v \times B$)
- Folded with Tunka-Rex hardware response
- Upsampling by factor 16

Learning strategy and training pipeline

Datasets:

- 25k samples for training

Subsets grouped by amplitudes:

- 10 – 100 µV/m (used in present work)
- 100 – 200 µV/m
- 200 – 300 µV/m

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

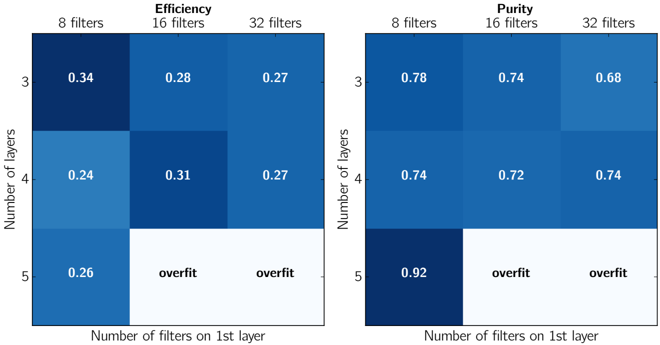
i-th encoding layer is described by the following ($i = 1, ..., D$):

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

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 (corresp. to few ns)

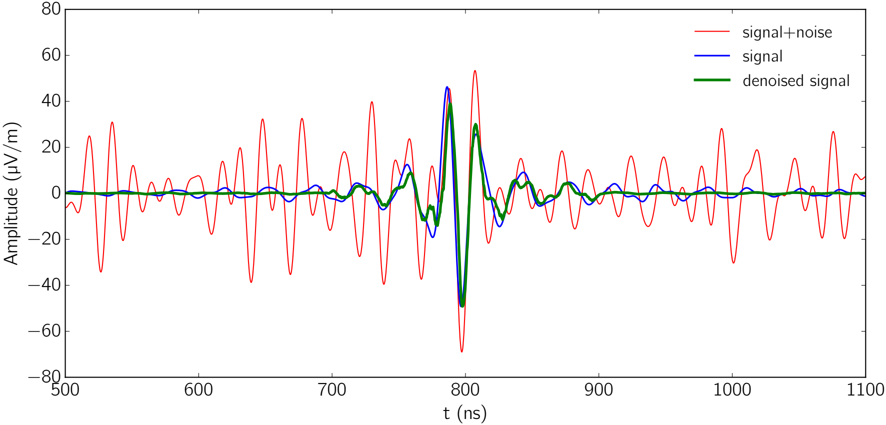
Threshold and metrics

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



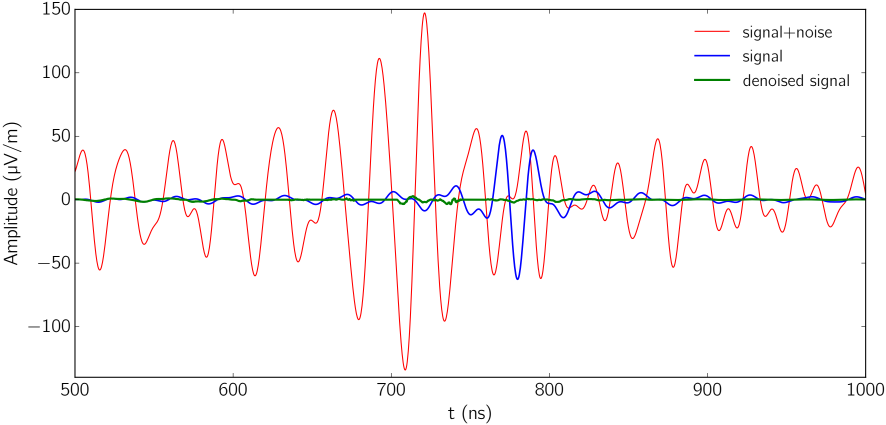
Best architecture contains $N_{\text{dof}} = 10240$

Example: correct identification



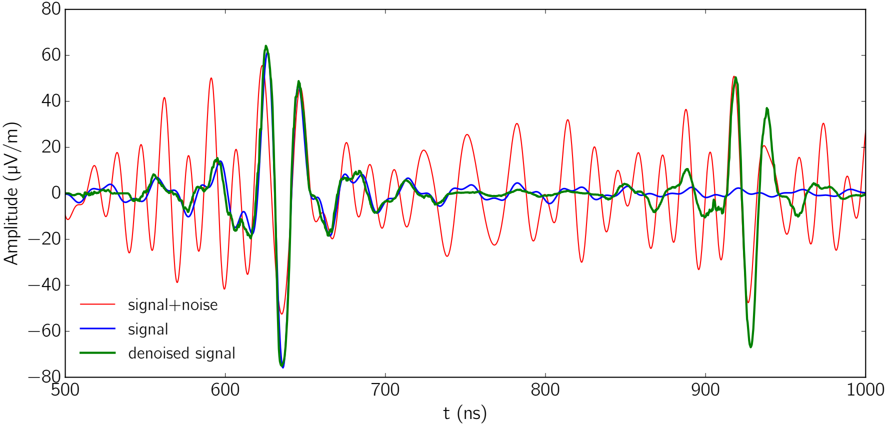
True signal and noise are identified correctly, noise is removed

Example: no identification



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

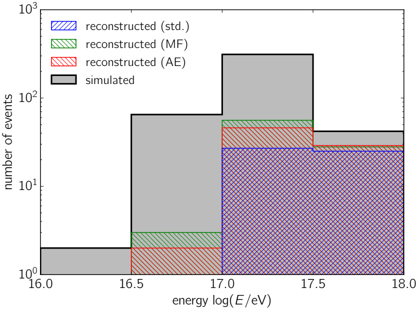
Example: double identification

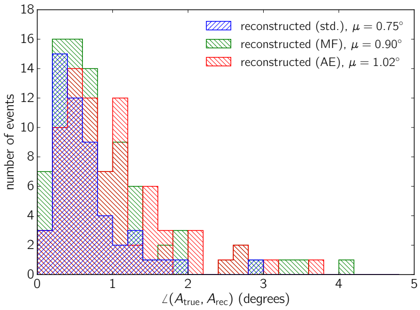


Signal-like RFI is identified as signal

Full-pipeline reconstruction with autoencoder

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





Autoencoder summary

- “Stack more layers” rule works, but requires larger training sets
- Signal properties of denoised traces are under investigation
- We plan to try different architectures of neural networks, check of signal position reconstruction by real data and investigate the properties of background.
- Blind check shows precision compared with matched filtering and classical reconstruction

DATA ACTIVITIES

Data activities

Open data in frame of virtual observatory:

- Analyse archive data
- Process and make it open
- Search for radio transient
- Test the performance of digital radio array for cosmic ray detection in application to astromonical problems

Tunka-21cm

- Side project of Tunka-Rex for astronomic purposes
- Studying 21cm hydrogen line with big red shift at MHz frequences
- Beamforming
- Estimated data flux ≈10 Gb/hour

Tunka-21cm

Summary

Autoencoder

- Blind check shows precision compared with standard reconstruction
- Plan to check of signal position reconstruction by real data

Virtual observetory

- Preparing and processing data

Tunka-21cm

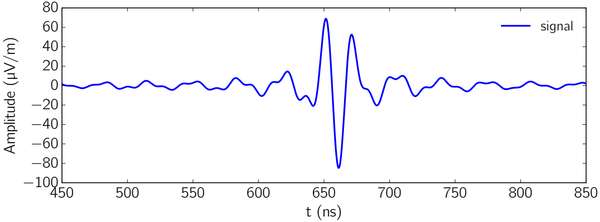
- Parallel calculations over the field of view
- How to realtime?

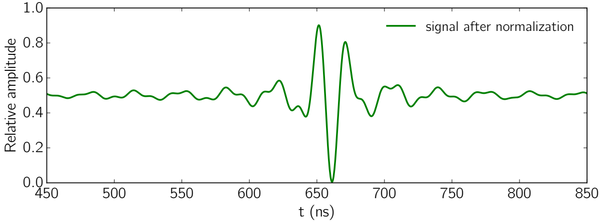
BACKUP

Traces normalization

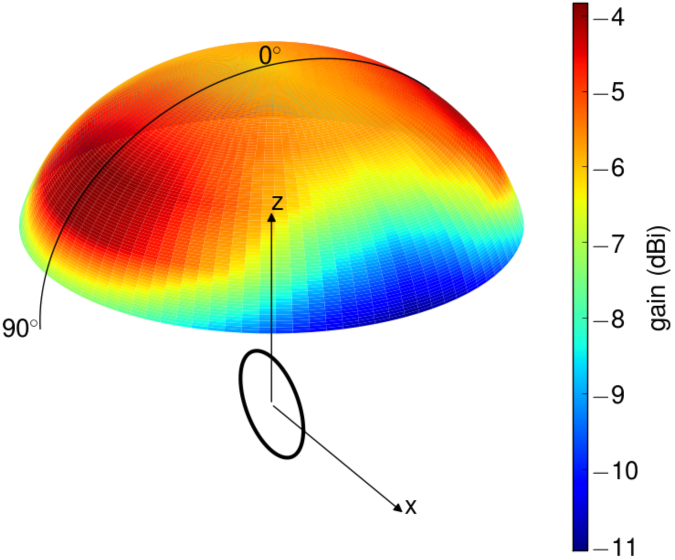
Traces should be normalized to 0–1 values, baseline should be located at 0.5 level

$$s'_{\mathrm{i}} = \frac{s_{\mathrm{i}}}{\max(u_{\mathrm{i}})} + 0.5,\ \text{where } u_{\mathrm{i}} \text{ is envelope of trace}$$

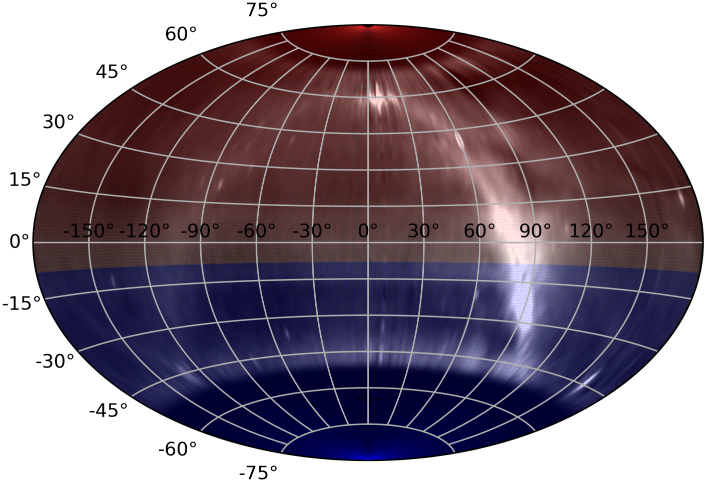




Radiation pattern



Tunka-21cm field of view



HDF5 container for Tunka-Rex

Detector config:

- Coordinates
- Layout
- Hardware response

Summary

- Spectra
- FITS maps

Events (array)

- timestamp
- coordinates (ra, dec)
- weather
- reconstruction
- traces (array)