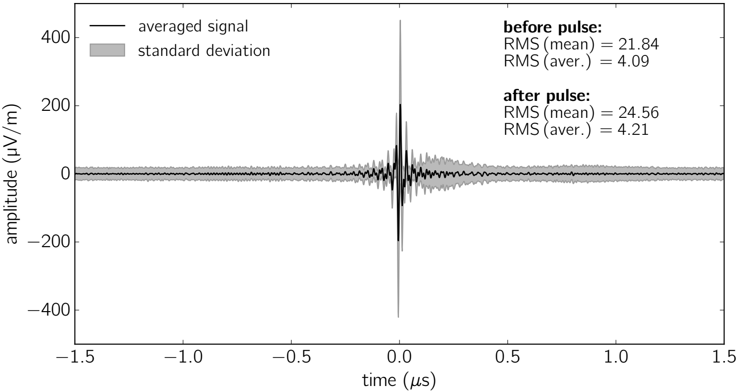
# Reconstruction of sub-threshold events of cosmic-ray radio detectors using an autoencoder

P. Bezyazeekov

International Cosmic Ray Conference, Berlin 2021

July 6, 2021

# Deep learning: motivation



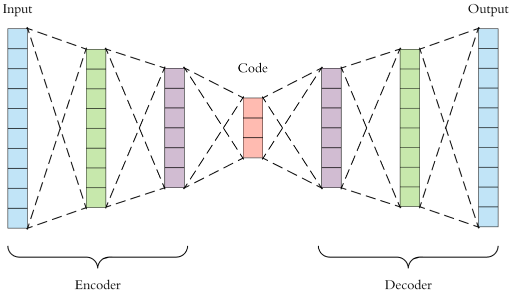

Average of 400 events, expected noise reduction with factor $\sqrt{400} = 20$

$\Rightarrow$ 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 $(< 100\,\mu\text{V/m})$ 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 $(i = 1, ..., D)$:

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

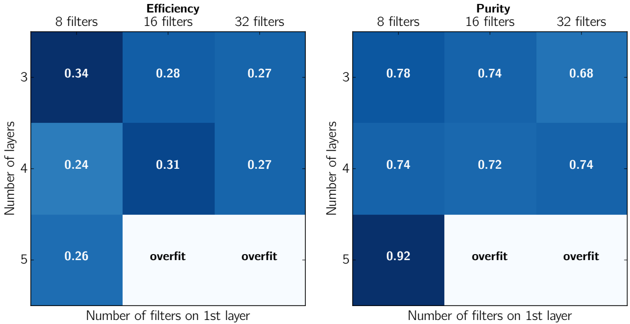
where $S_i$ is a size of the $i$-th filter, $n_i$ is a number of filters per layer

$D$ 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

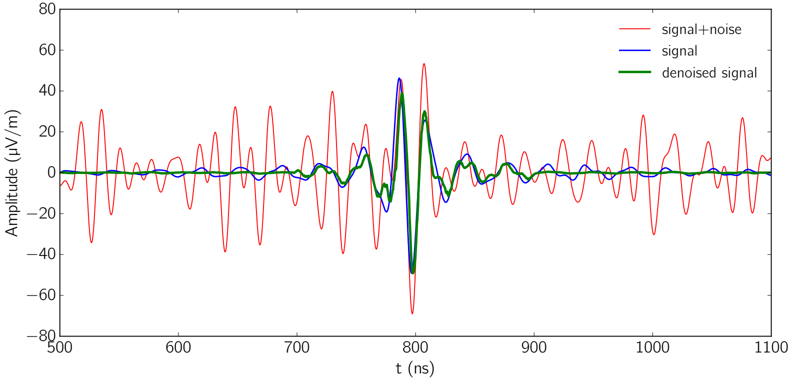
# Threshold and metrics

- 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



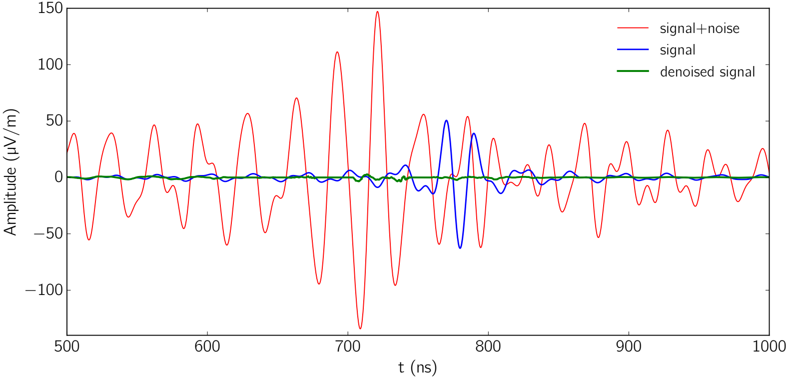

Best architecture contains $N_{\mathrm{dof}} = 10240$

# Example: correct identification



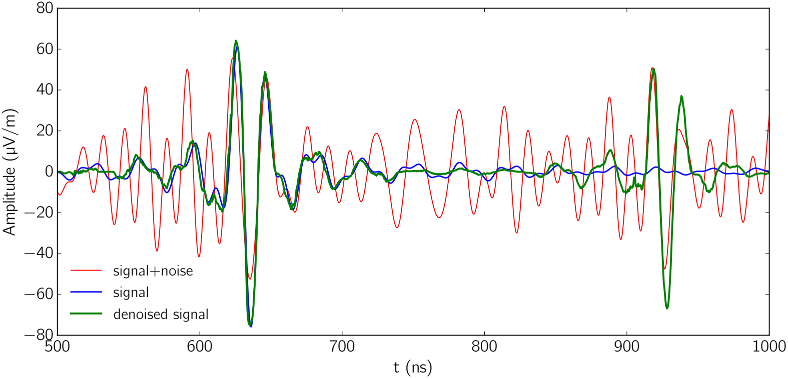

True signal and noise are identified correctly, noise is removed

# Example: no identification




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

# Example: double identification

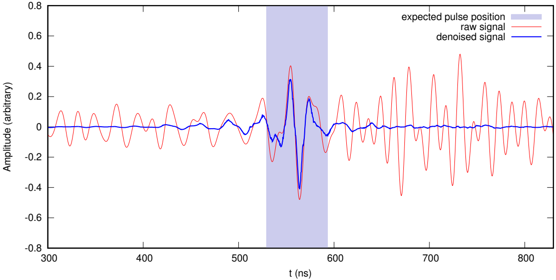



Signal-like RFI is identified as signal

# Preliminary conclusion

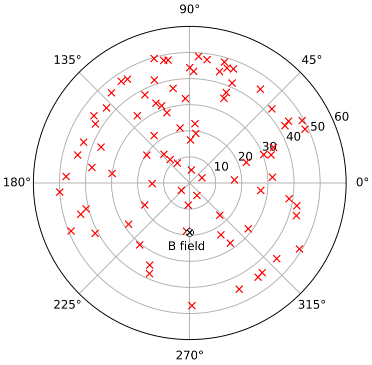
- Monte-Carlo tests show performance comparable to standard method and matched filtering
- "Stack more layers" works, but requires larger training sets
- Amplitude reconstruction degenerates when $\mathrm{SNR} < 1$ trace is normalized to $[0;1] \Rightarrow$ peak is hidden in noise

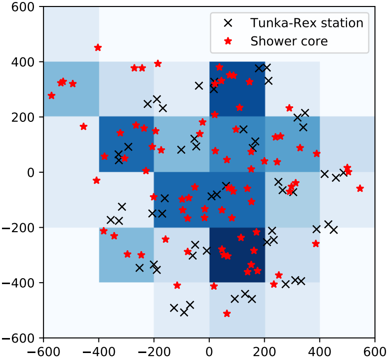
**How to convince ourselves that the reconstruction is valid when the signal is not visible?**



# Data-driven benchmark

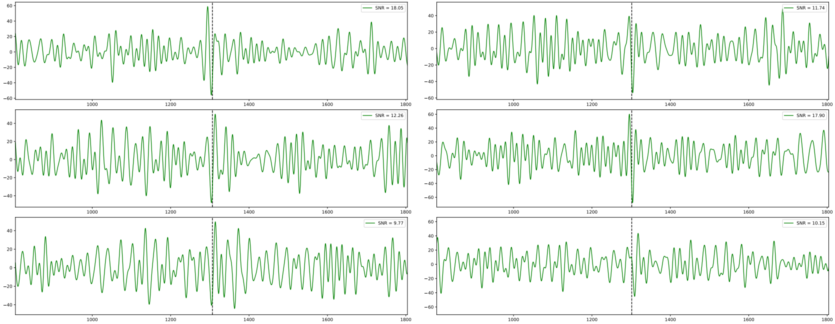
- Tunka-133/Tunka-Rex events with $E \in [10^{16} - 10^{17}]$ eV
- Almost zero events in this energy band by standard method
- Decreasing autoencoder threshold $0.395/0.500 \to 0.200/0.500$
- Cross-check cuts: direction reconstruction $\Delta\Omega < 5°$, clustering events



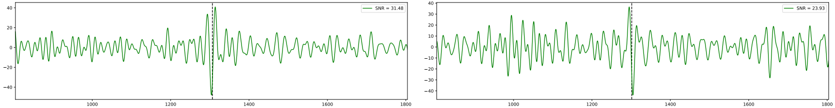



# Example reconstruction

Two example events with $E = 30$ PeV



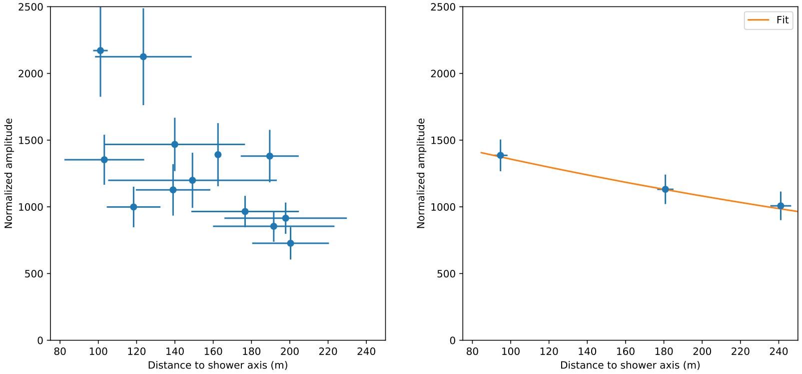

Coherent sum:



# Adaptive LDF (after cuts)

Few antennas are synthesized into single one in order to increase SNR

The slope of averaged LDF is used for energy reconstruction

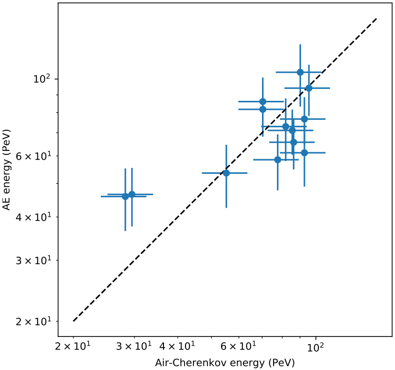


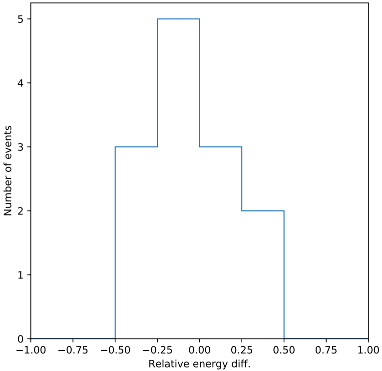
# Energy reconstruction (after cuts)

Reconstruction based on single antenna method,

$$E = \kappa A_d e^{-\eta(d-d_0)}$$

Normalization factor from standard reconstruction; $\mu = 0\%$, $\sigma = 26\%$



# Conclusion

- The performance of Tunka-Rex autoencoder has been tested on real data
- Numbers of both true and false positives are increased when loosing cuts
- We can reconstruct arrival direction but struggling with energy reconstruction

---

- Radio autoencoder can be used as self-trigger technique
- Need more sophisticated cuts to lower the threshold
- Need better training