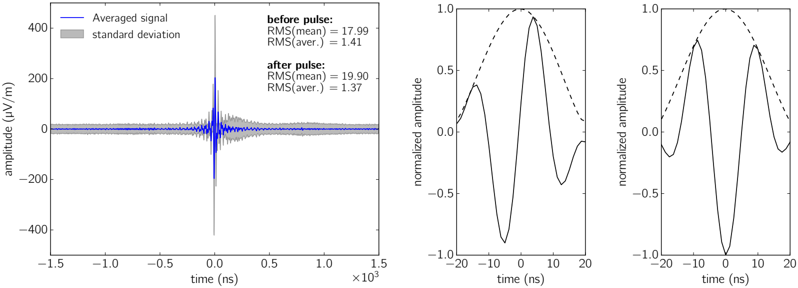
Denoising radio traces with autoencoder on Tunka-Rex experiment

Dmitriy Shipilov, Dmitriy Kostunin for the Tunka-Rex Collaboration
February 20, 2018

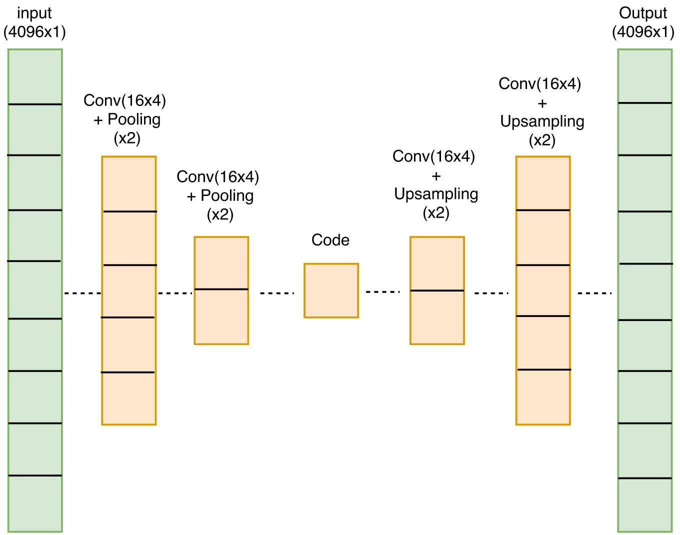
INSTITUT FÜR KERNPHYSIK

Motivation

- Real background is not absolutely white
- It is hard to manually categorize each possible RFI
- The true pulse shape is known with low accuracy



Architecture



Learning strategy

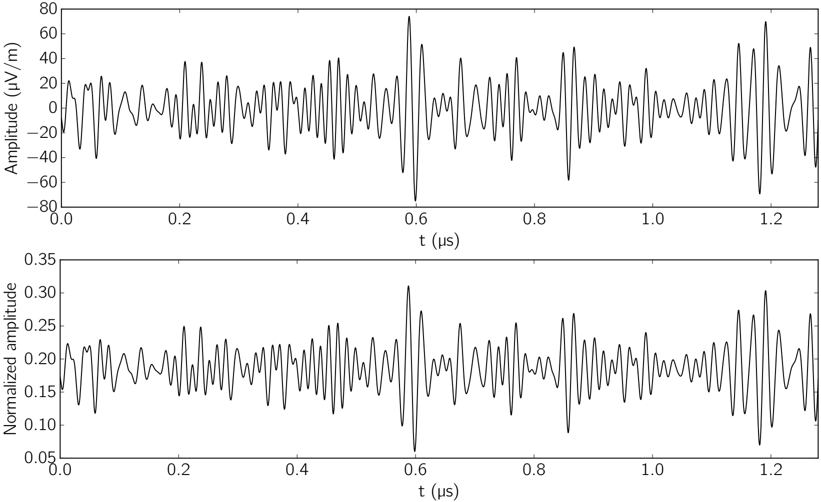
Dataset

- 650k samples of Tunka background recorded in 2014-2017
- CoREAS simulations of Tunka-Rex signals
- Pulse is randomly located inside signal window (200 ns)
- Using single polarization
- Upsampling 16, using 4096 bins per trace (1280 µs)

Three different networks

- High amplitudes $A > 200$ µV/m
- Medium amplitudes $100 < A < 200$ µV/m
- Low amplitudes $A < 100$ µV/m – present work
  - 11k samples + 1k augmented samples (signal is out of signal window)
  - 90% for training
  - 10% for control

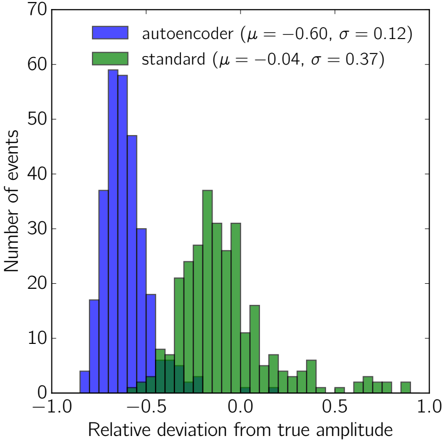
Signal normalization

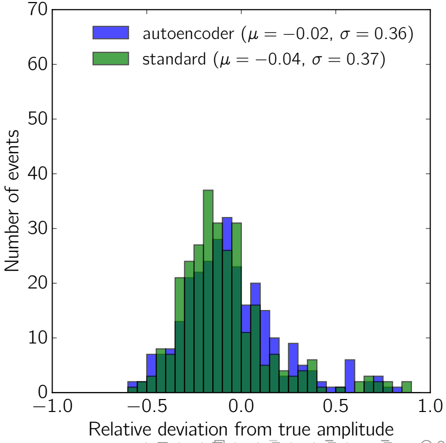


Amplitude is normalized to be bounded in [0;1]

Issue with normalization

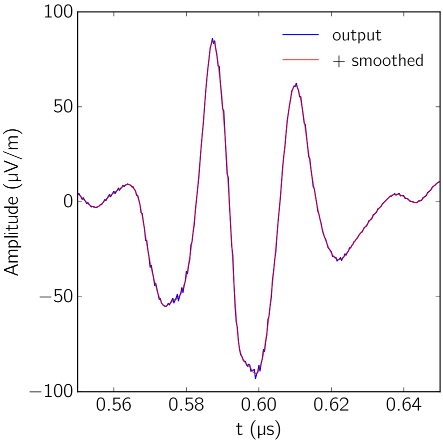
- Absolute amplitude of denoised trace is not reconstructed
- Re-normalization to scale of input trace can fix the problem

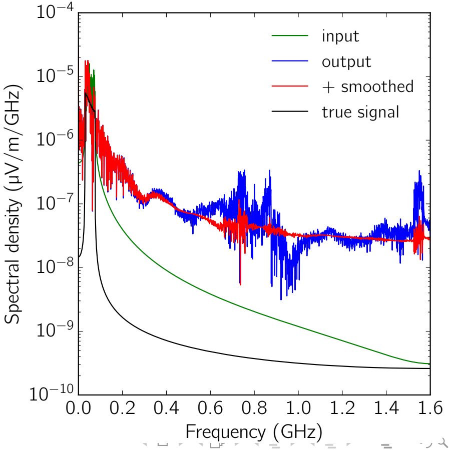




Smoothing of the signal

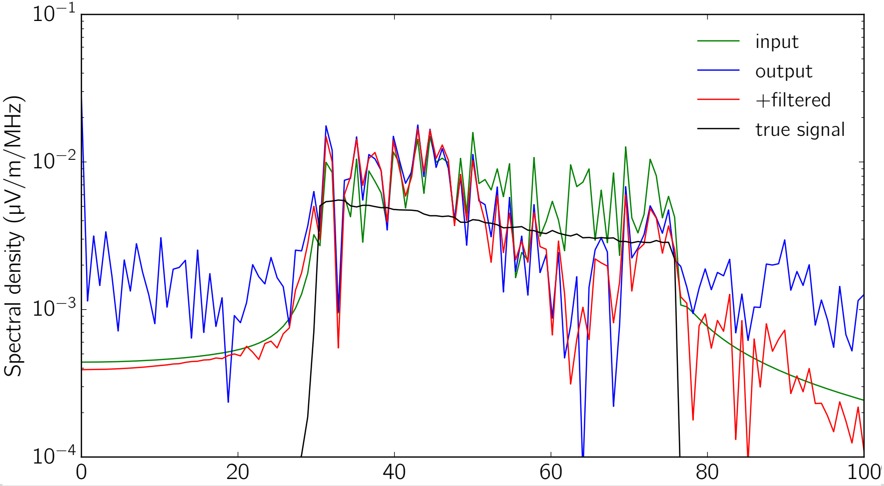
- Reconstruction adds high-frequency numerical noise
- Savitzky-Golay filter is used for smoothing



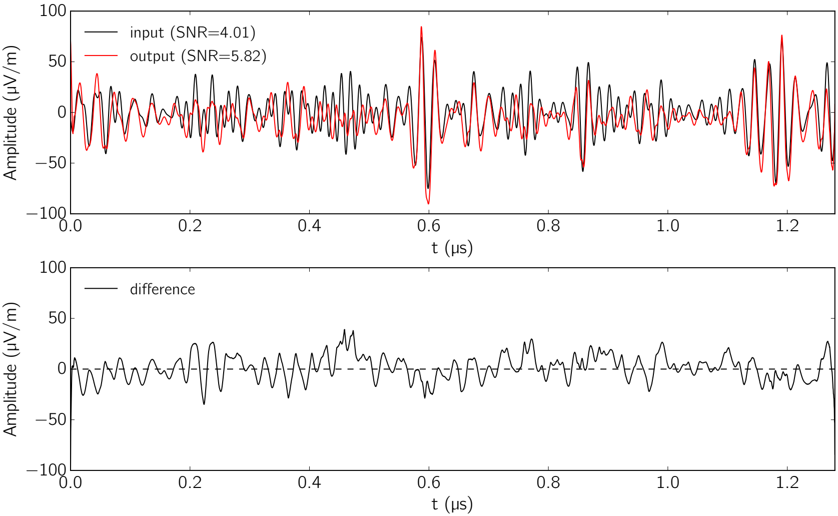


Bandpass filtering

- Signal can be filtered with Butterworth filter
- The impact is not investigated well (not used in present work)

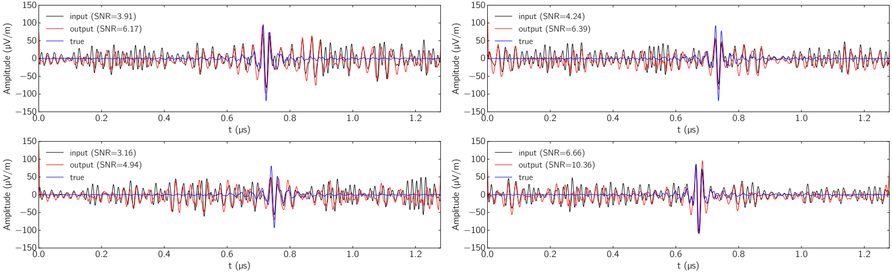


Example of denoising

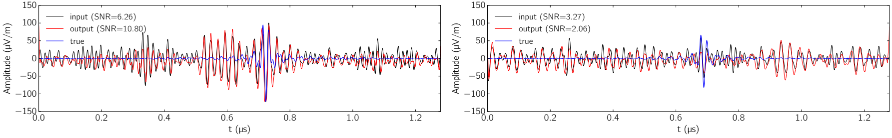


Few more examples

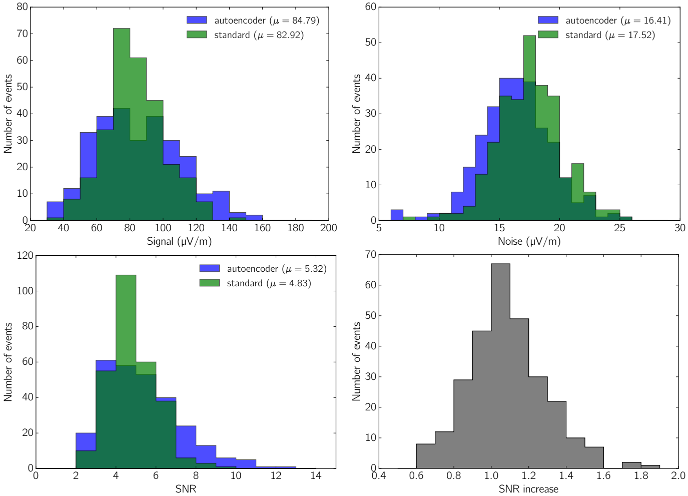
Autoencoder has learned how to remove various RFI



However, sometimes it leaves them, or even removes a signal



Improving signal-to-noise ratio



The signal-to-noise ratio is increased by ≈10%

Summary

Present status

- Autoencoder shows promising results
- The overall SNR is increased on the chosen dataset
- Prospects of investigation of spectral features

Next steps

- Fixing the issue with normalization
- More sophisticated criteria for estimation of performance is required
- Comparison of performance with matched filtering
- Investigation the different configurations: upsampling, trace length, size of dataset
- Including second polarization in analysis
- Alternative training in Fourier domain