RFI Removal with Neural Networks

Benjamin Saliwanchik

Brookhaven National Laboratory

21cm Cosmology Workshop
University of Wisconsin
August, 31st 2022

Computational Problem of RFI

- Radio Frequency Interference (RFI) is significant issue for radio frequency telescopes.
- Especially for 21cm intensity mapping experiments, which need exceptionally high mapping sensitivities to detect BAO.
- Current generation intensity mapping experiments such as HIRAX, HERA, and CHIME are of order 1000 dishes. Use correlator to synthesize beams, take advantage of intensity and phase information from full array.
- Next generation arrays such as PUMA will require O(10,000) dishes. So being able to effectively process all dish data (in real-time) is computationally difficult.

HERA

HIRAX (rendering)

Computational Problem of RFI

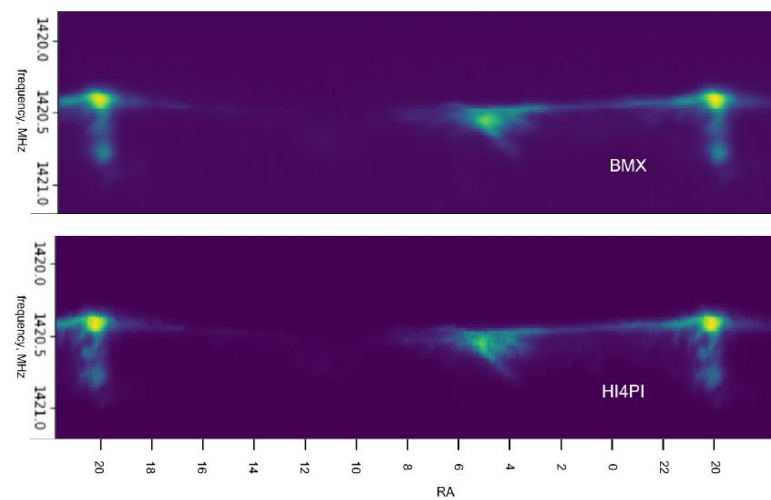
- RFI is often narrow-band and transitory, these sources can be filtered by timestream or pixel thresholding methods.
- Can also filter at the level of the spectrometer output:
  - ➢ Data has typically been massively reduced at this point into measurements of N spectral bins integrated over some time (for BMX N = 2048, t~0.1s)
  - ➢ Easy on resources, but necessarily lossy
- Or filter at the level of input waveform:
  - ➢ Working with raw digitized signal after the front-end amplifiers and before spectrometers – minimizes data loss
  - ➢ Massive data rate: typically ~2GB/s per channel => 40TB/s for next-gen array! Roughly half the current world internet traffic.
  - ➢ Must be processed in real time because storage of this volume is impossible!

Computational Problem of RFI

- Current spectral methods include linear algebraic methods (SVD, PCA) and Machine Learning (ML)
- Supervised ML methods include K-nearest neighbors, random forest classifiers, long short-term memory (LSTM), You Only Look Once (YOLO), and Convolutional Neural Networks (CNN).
- Unsupervised ML more desirable. Methods include Convolutional Auto-Encoders (CAE) and Generative Adversarial Networks (GAN).
- We propose a novel Deep Convolutional Neural Network method for unsupervised detection and removal of RFI at the raw timestream level.

Baryon Mapping Experiment (BMX)

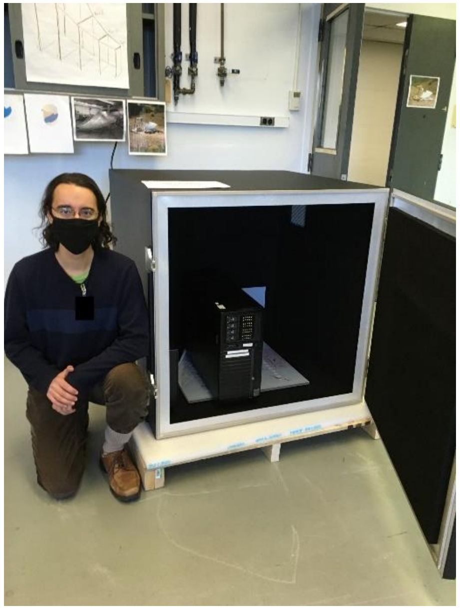
- BMX telescope is 21-cm prototype at BNL, technology demonstrator for PUMA (arxiv: 1907.12559, puma.bnl.gov)
- BMX design and calibration observations described in O'Connor et al. (arxiv: 2011.08695)
- Four zenith pointing off-axis parabolic dishes, frequency range 1.1-1.55 GHz. Objective is to develop next-gen hardware and software, observe local HI (Milky Way and galaxies at z < 0.3)



BMX and HI4PI images of HI in Milky Way

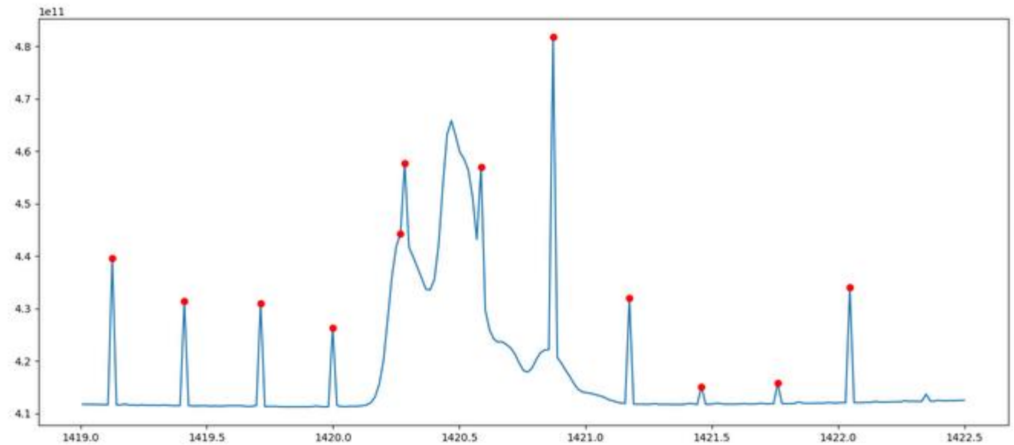
BMX – Hardware RFI Mitigation

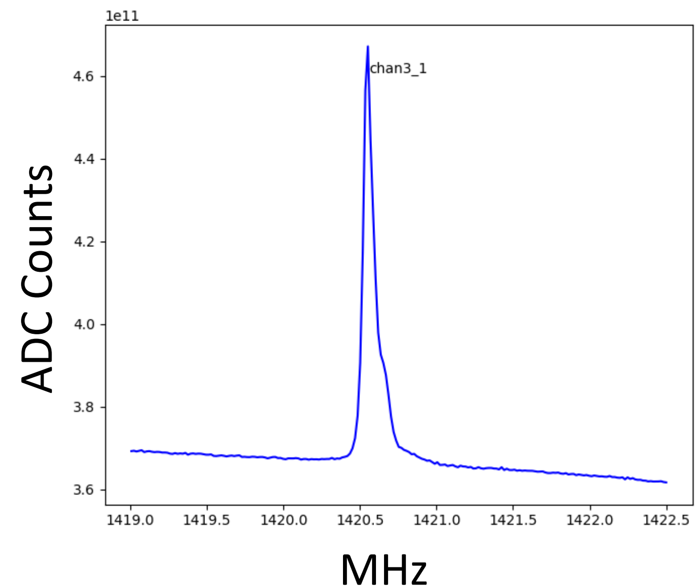
- Important to minimize RFI that enters system to simplify post-processing. Use site features, shielding.
- BMX has double-walled enclosure with shielded bulkhead connections and waveguide air vent to suppress RFI from backend electronics.
- Contained in weather enclosure made of Al T-slot frame and Al-PE composite panels for UV resistance.



BMX – Hardware RFI Mitigation

- RFI enclosure provides 75-90dB of RFI suppression across band in lab tests
- In field, local RFI features in observations completely suppressed to below noise floor.
- RFI marked in red, 21cm line (1420.4 MHz in rest frame) unmarked
- Are still RFI non-local sources (including GPS, TV, cell phones, and air traffic control) and foregrounds that are not suppressed by shielding





Non-Gaussian Signal Recognition

- NN potentially very good at recognizing RFI
- Signal separation typically requires knowledge of all potential RFI waveforms to train NN
- We propose an alternative:
  - Astrophysical signals of interest are from thermal sources, so they are Gaussian random fields (amplitude and phase Gaussian random distributed)
    => **statistically completely defined by their power spectrum**
  - Majority of RFI sources are attempts at communication
    => **they are maximally non-Gaussian**
  - Can use non-Gaussianity of RFI to distinguish it
  - (This will also detect non-stationary sky signals, such as FRBs. Further processing will be necessary to distinguish such sources.)

Non-Gaussian Signal Recognition

- Let g be the gaussian signal (including noise) and ng be the contaminating non-Gaussian signal

$$s = g + ng$$

- If you know the ng, you can train a neural network so that

$$s - \mathrm{NN}(s) = g$$

$$\Rightarrow \mathrm{NN}(s) = ng$$

- The problem is that we don't know either g nor ng, just s

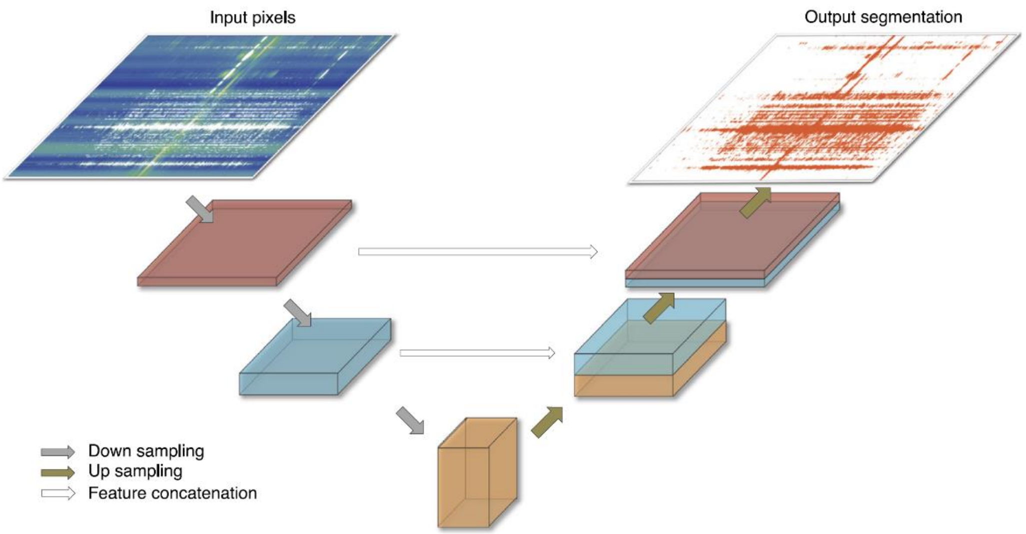
Non-Gaussian Signal Recognition

- Say you can generate a new (known) Gaussian signal with the same spectrum as *g*, call it *g'*.
- Create a new signal:
  - $s' = s + g' = g+ng + g'$
  - *(Or more properly:* $s' = (1-\lambda^2)^{1/2}s + \lambda g' = (1-\lambda^2)^{1/2}(g + ng) + \lambda g'$)
- Train neural network so that
  - s' - NN$(s') = \lambda g'$
- If NN*(s')* picks up non-Gaussian part, it will improve the loss function
- If NN(*s*') could pick up non-Gaussian + intrinsic gaussian, it could do even better -- but it cannot possibly do this, because it is theoretically impossible to distinguish *g* and *g'*
- So: network will find the compressible pieces of information that are best at lowering the variance – which will be RFI since *g* is (nearly) incompressible
- There is the curious case of λ=0, in which we are solving for s - $NN(s) = 0$ and no additional known Gaussian signal is necessary for training.

“U-Net” Neural Network

- We use a neural network that consists of an “encoder” and “decoder” network.
- Each step is a linear operation that reduces/increases the size of the array, potentially with non-linear activation functions between, and a dropout function before the final linear down sampling step.
- Forcing data through a “choke point” of minimal dimension forces the NN to learn a few parameters that explain the most features of the full data, which is precisely what we need to distinguish non-Gaussian signals.

J. Akeret et al. / Astronomy and Computing 18 (2017) 35–39

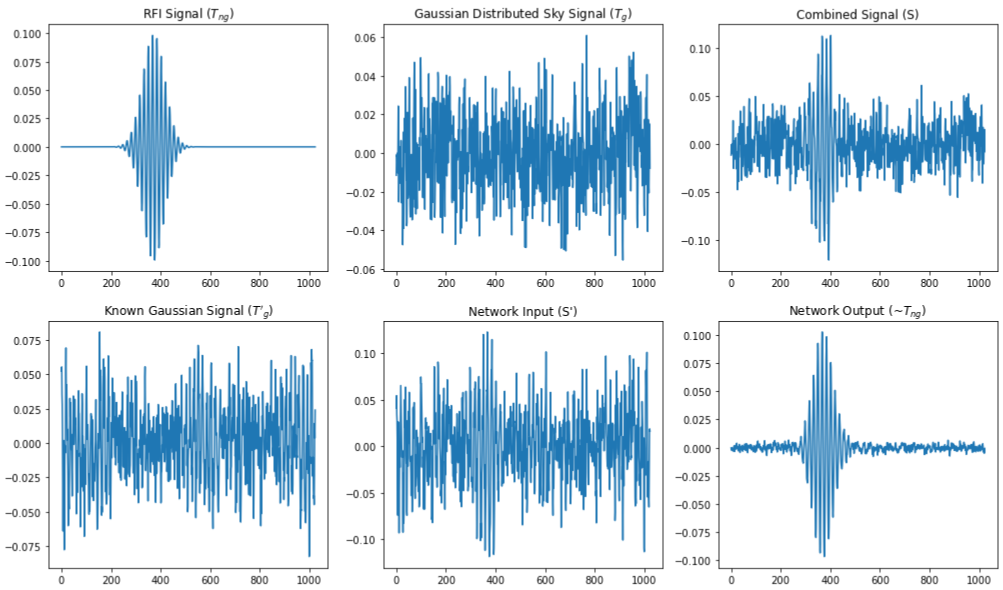


It Works!

- Run for 30 epochs with 100k training datasets, 10 test datasets
- Baseline case where input to network is

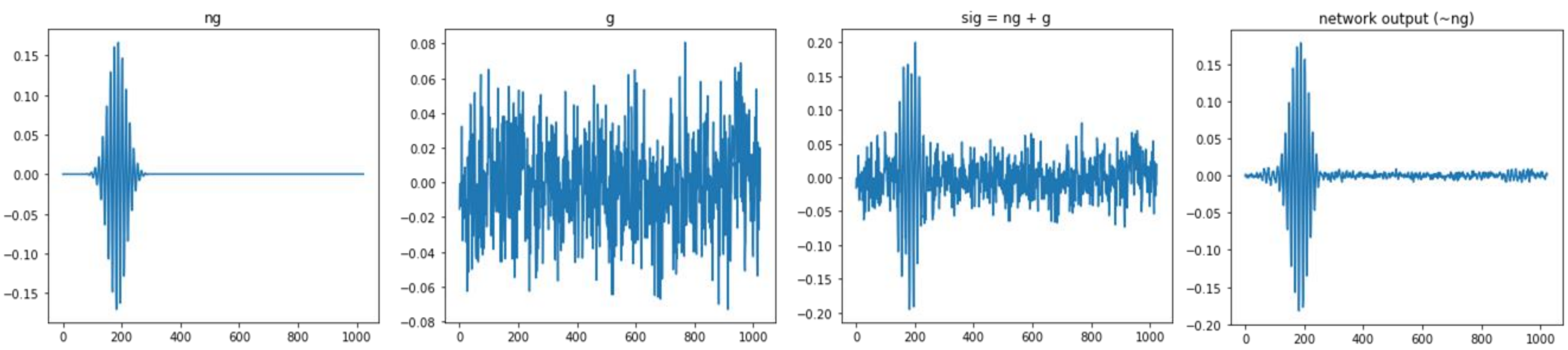
  $s' = g + ng + g'$
- Parameterize goodness of fit by RMS error:

  $E_{RMS} = sqrt((S\_recov - s')^2 / Np)$
- For baseline case average RMS error across test cases is: $E_{RMS} = 0.022$



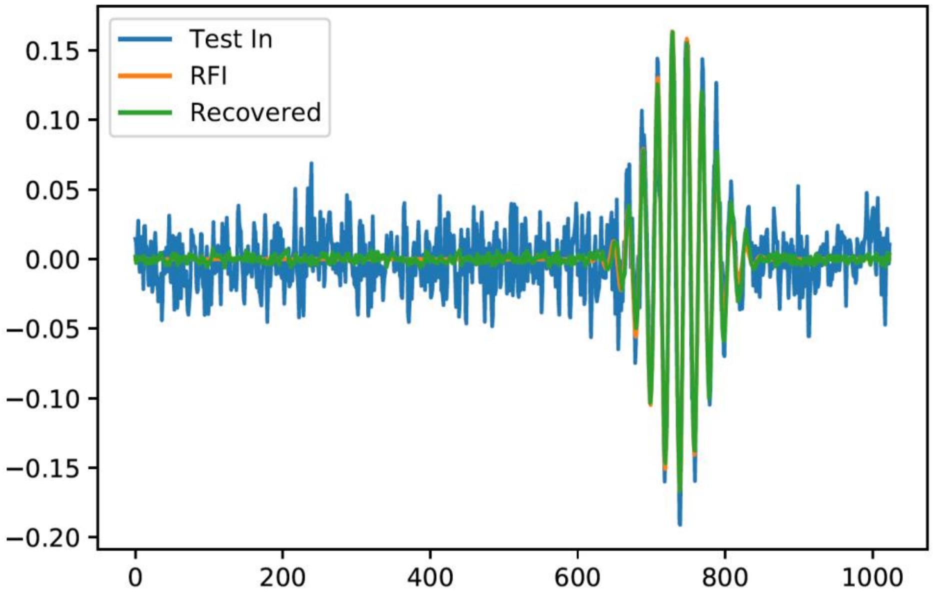
Optimizing NN

- Fit improved with normalization if λ is small.
- λ=0 case works very well, and we no longer need *g'* which saves processing time.
- RMS error reduced to $E_{RMS} = 0.004$
- Recovered RFI fits input envelope shape better, amplitude of recovered signal away from input RFI region smaller.



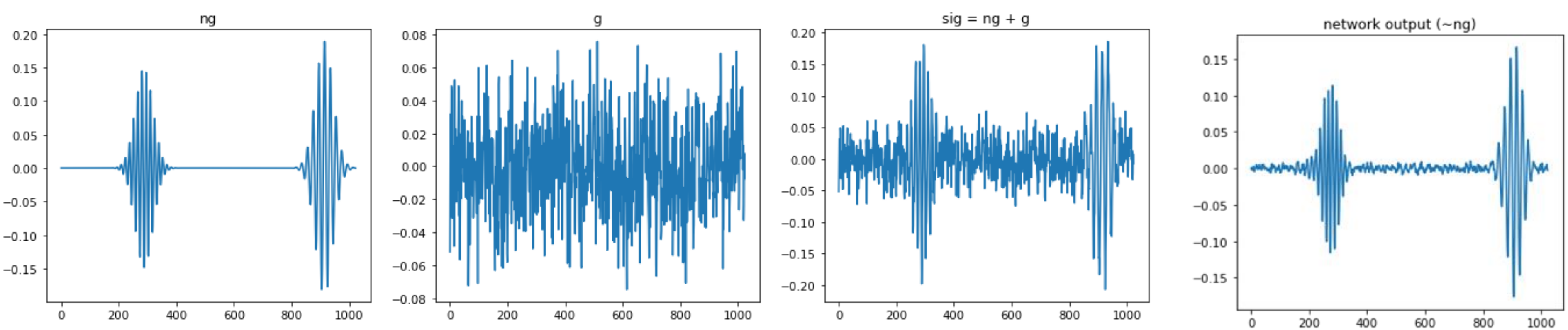
Optimizing NN

- Error in recovered timestream can be further improved by optimizing the number and size of steps in encoder/decoder network.
- Optimal architecture found used 3 hidden dimension steps, the first two with 1024 elements, and third with 16.
- RMS error reduced to $E_{RMS} = 0.0037$



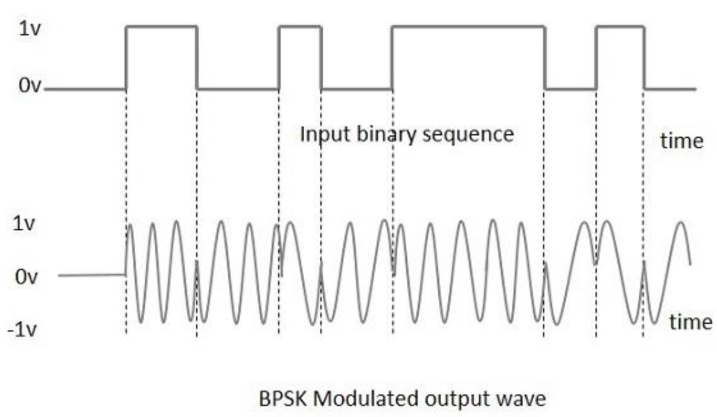
Multiple RFI Events

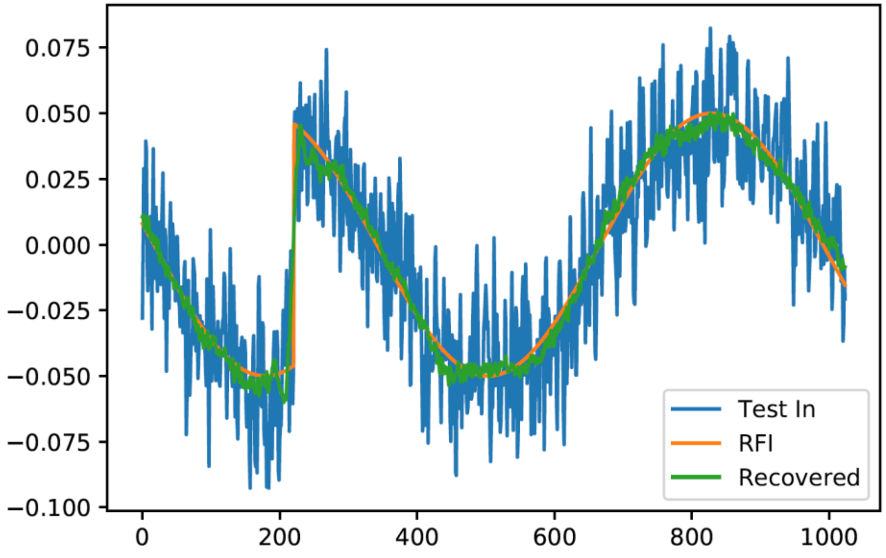
- We can fit for more than one RFI signal in the same timestream, but recovery is degraded.
- Fit improved by increasing size of final phase of encoder to give more degrees of freedom. Using zdim=32 instead of zdim=16, $E_{RMS} = 0.012$
- Can avoid fitting multiple RFI events in one timestream in most cases by judicious choice of length.



RFI Morphology

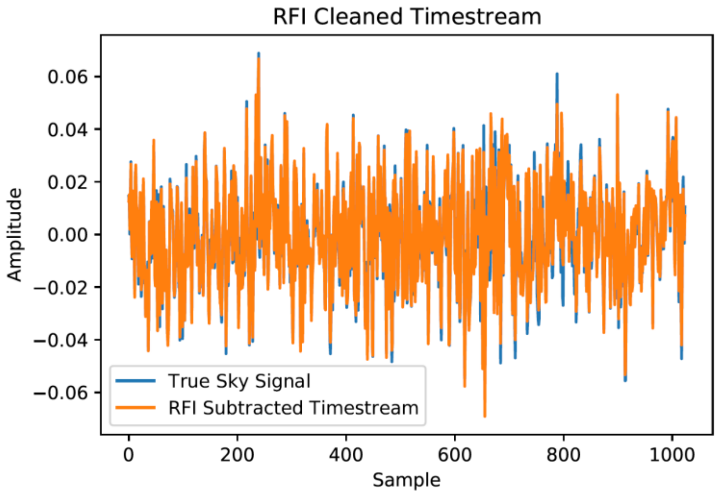
- Also tested long period sinusoidal RFI morphology, potentially with sign-flip at random location.
- This emulates a bit-flip in a binary phase-shift keying (PSK) digital signal.
- With no sign-flips we find E_{RMS} = 0.0032, with sign-flips: 0.0048.
- NN is good at fitting sinusoidal variation, even in presence of discontinuous features like sign-flips

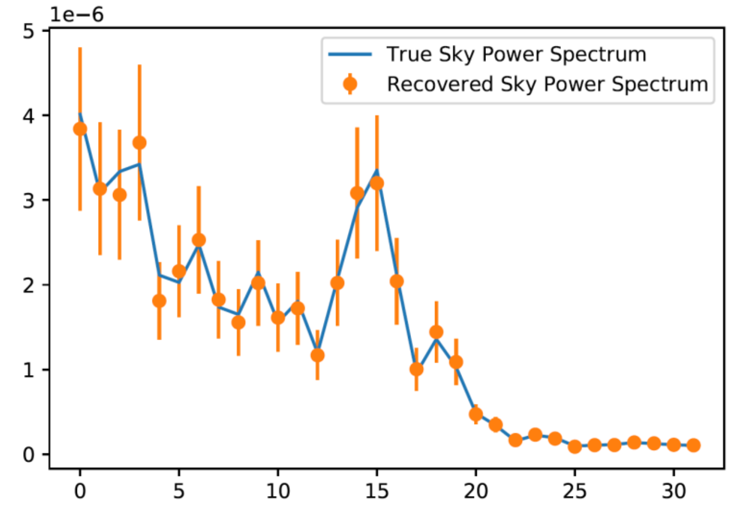




Power Spectrum Recovery

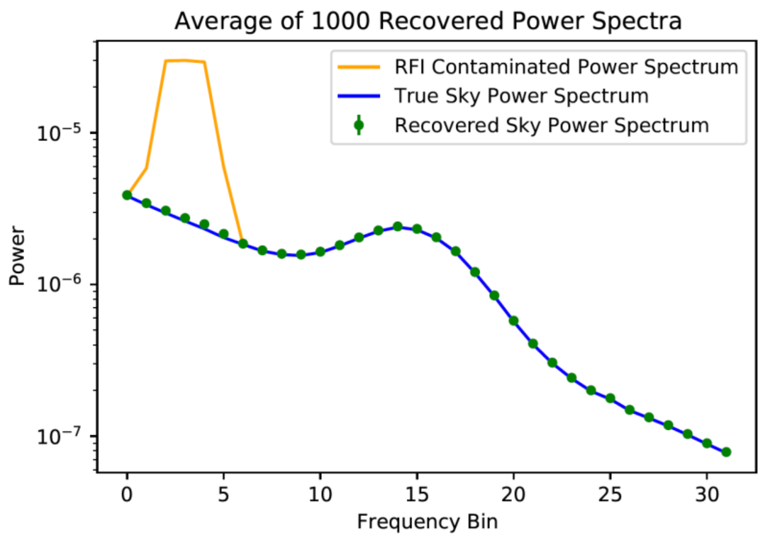
- Want to verify that RFI removal does not alter sky signal.
- After subtracting RFI event with amplitude 10x sky signal, input and recov timestreams vary by 0.56%.
- Power spectrum of recovered timestream consistent with spectrum of input timestream.
- Chi^2 = 0.0496, with 31 degrees of freedom, for a p-value of 1 to within 3x10^-38.

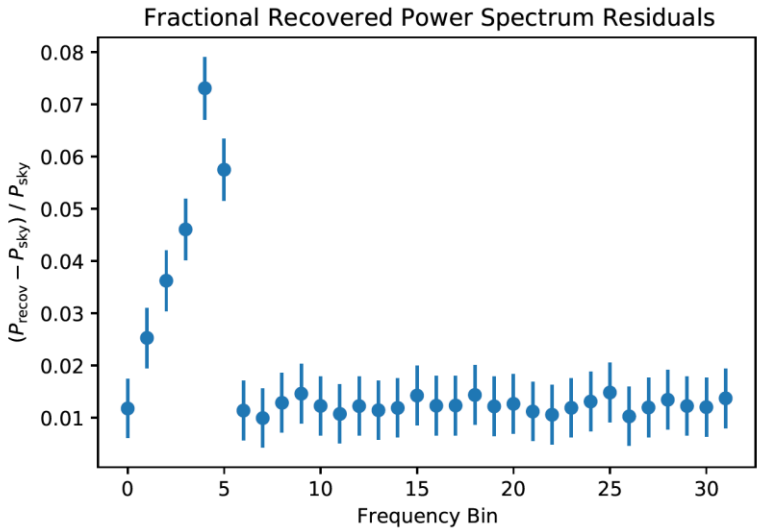




Power Spectrum Recovery

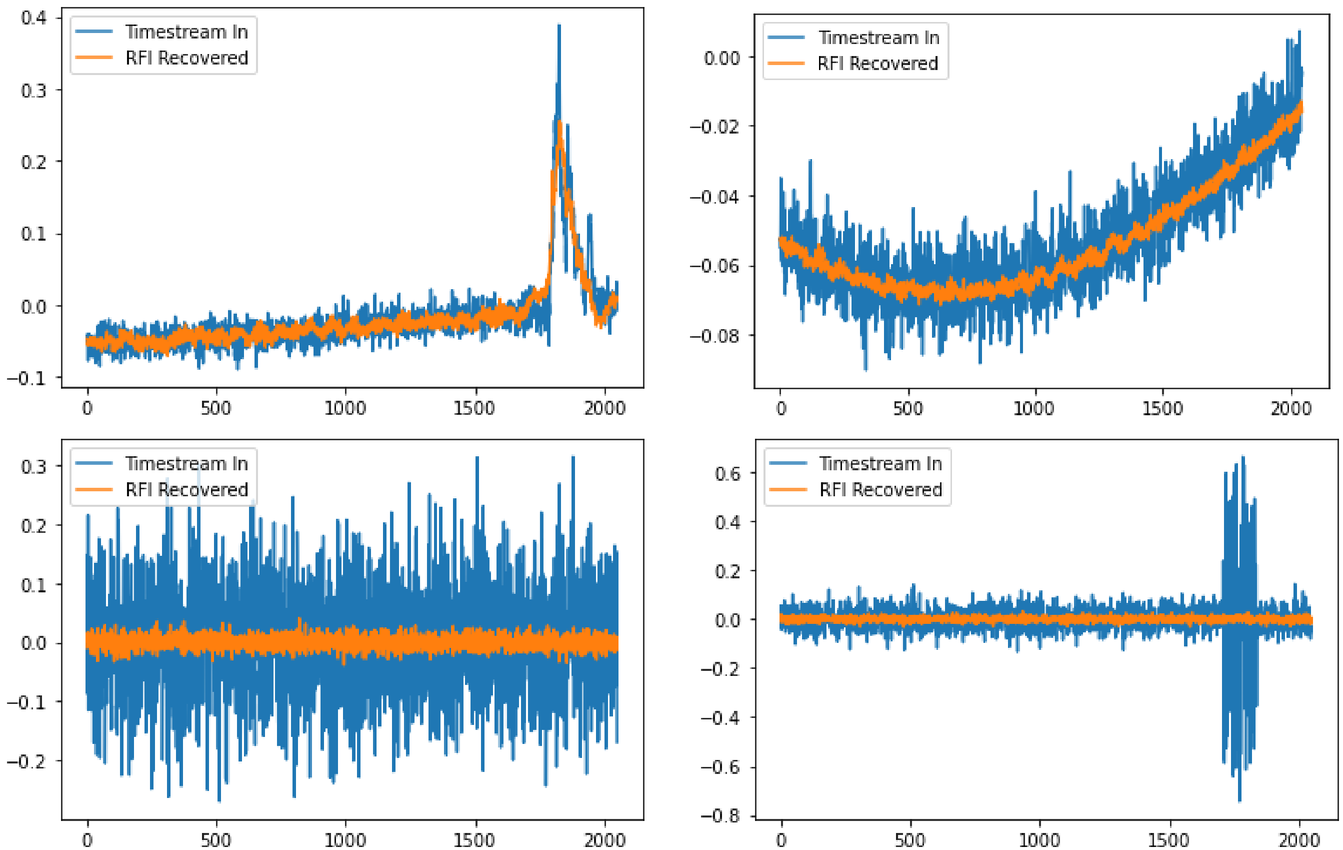
- Input spectrum is recovered even in PS bins where RFI power is 10x higher! 1σ error bars smaller than data points.
- Residuals show there is detectable bias when averaging over many timestreams: 1% increase in noise across band, and up to 7% bias in RFI contaminated bins.
- Most likely a noise bias that could in principle be calibrated. (Future work.) Even without correction, significantly better than throwing away data from RFI contaminated frames.
- This is also maximally pessimistic case in which every frame is RFI contaminated. In reality, only 0.1% to 1% of frames are typically contaminated by RFI, taking BMX data as our reference.





Sky Data from BMX

- Ran on observed sky data from BMX!
- Able to detect several RFI morphologies.
- Does not detect RFI that looks like white noise of increased amp. Would have to flag and cut timestream using standard amp thresholding method.



Summary

- We have developed a NN for detecting and removing RFI
- Optimized NN, and tested on different RFI morphologies
- Run on real sky data from BMX observations
- Use in real-time for next-gen telescope arrays will likely require hardware implementation of NN, such as in a Radio Frequency System-on-Chip (RFSoC). Will require significant work!
- Paper on method on arxiv, updated version with PS results soon. https://arxiv.org/pdf/2203.16607.pdf