RFI Detection Across Six Orders of Magnitude in Intensity: A Unifying Framework with Weakly Supervised Machine Learning

Stephen Bach [1] Jonathan Pober [2] Bryna Hazelton [3] Miguel Morales [4] Alexander Ratner [5]

Shawn Dubey [2] Jade Ducharme [2*] Charles Duong [1] Aidan LaBella [1] Yishan Liu [1] Christopher Tripp [1] Eli Lilleskov [4]

[1]C.S. Dept., Brown [2]Dept. of Physics, Brown [3]eScience Institute, U. of Washington [4]Dept. of Physics, U. of Washington [5]C.S. Dept., U. of Washington

RFI and 21 cm Cosmology

- 1 cm cosmology experiments must integrate > 1000 hours of data to detect the faint spectral signature of neutral hydrogen in the early Universe
- < 1 mJy (1 Jy $= 10^{-26}$ W m^{-2} Hz^{-1}) of residual RFI in the final integration can preclude a successful detection (Wilensky et al. 2020)
- Experiments are built at radio quiet locations (e.g. the MWA in Western Australia), but faint sources include reflections of RFI from below the horizon
- Need techniques that can detect RFI below the noise level of a single observation, but also need a way to combine the multiple techniques (including those which find bright RFI) in an internally-consistent way

Non-ML RFI Detection Methods

- `AOFlagger`: Most popular flagger, automatically runs on all MWA observations. Good at detecting edges but overall low performance
- `SSINS`: Good at detecting RFI bursts, but poor performance when RFI spans the entire observation
- χ^2: Good at detecting RFI that spans the entire observation

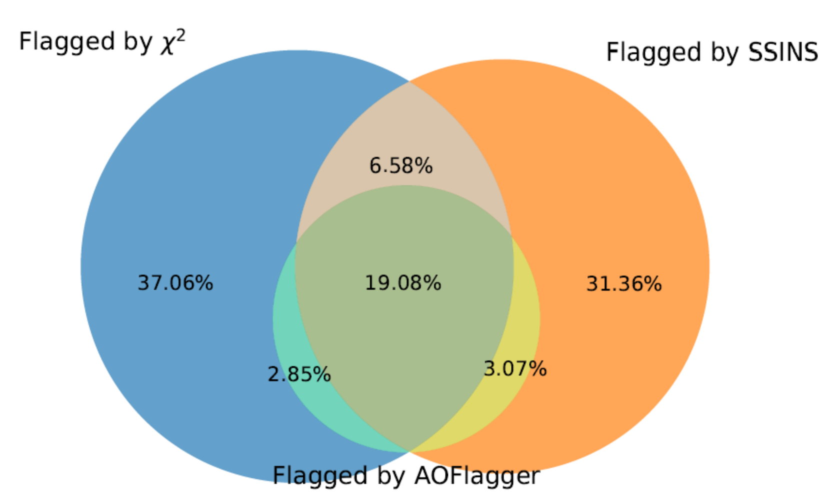

Figure 1. Percentage of observations with RFI in the DTV Channel 7 band flagged by each method, classified by detection or non-detection (Kunicki & Pober, in review).

Near-Field Interferometry

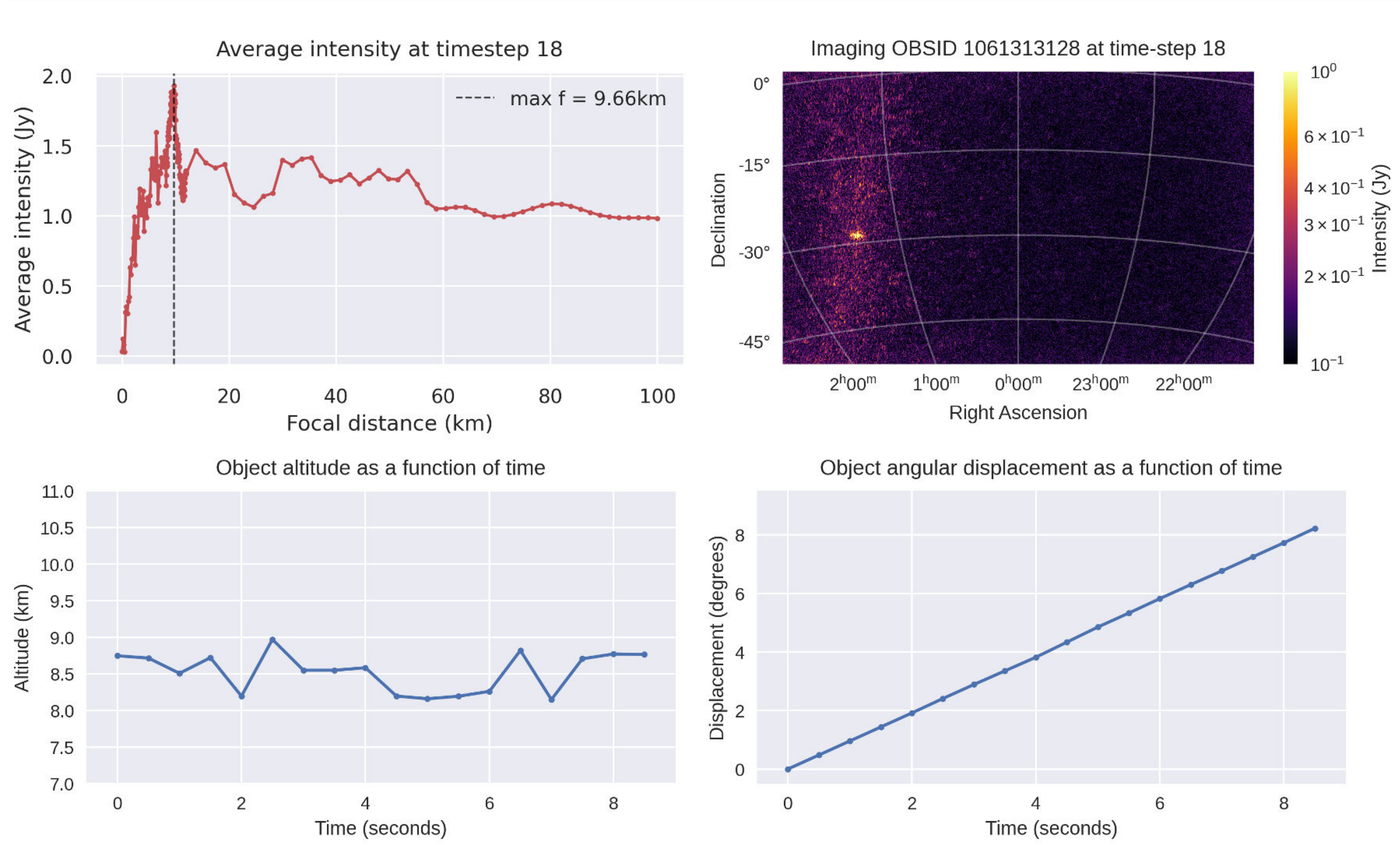

Figure 2. By applying near-field corrections, we can image and locate RFI emitting objects in MWA observations. (Top-left) The beamformed visibility is maximal at the focal distance corresponding to the location of the RFI emitter. (Top-right) Imaging the observation. (Bottom-left) The object's altitude (calculated using maximum of top-left plot) as a function of time. (Bottom-left) Object's radial displacement as a function of time.

Research Progress

- Developed a framework to reliably train U-Net models on MWA data using non-ML detection method ground truths.
- Developed a method to evaluate trained U-Net model performance by comparison to non-ML method, using metrics such as the intersection-over-union.
- Began development of new methods to identify RFI sources in image space.

Flagging Comparison

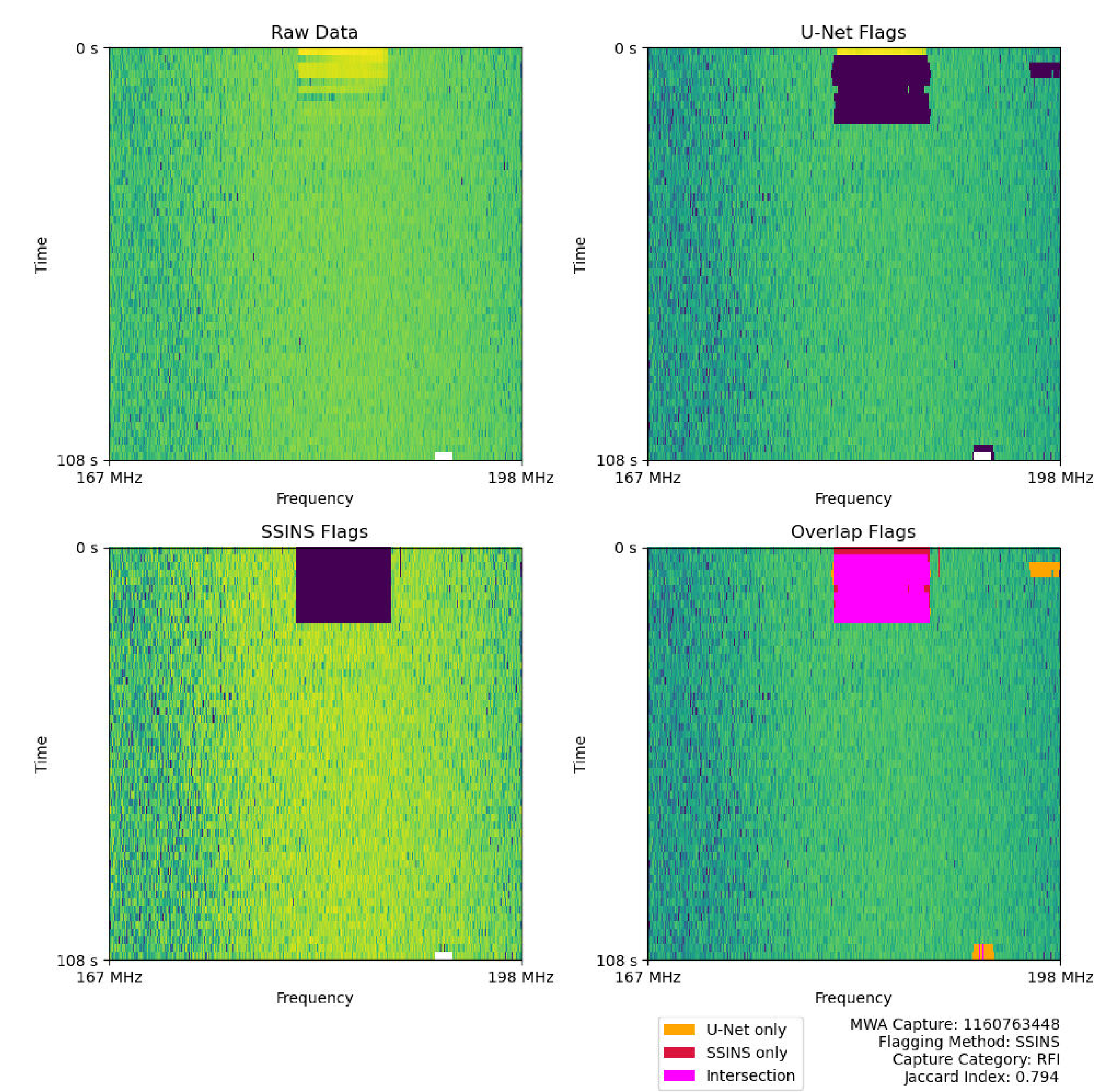

Figure 3. Comparison of traditional flagging methods and a trained U-Net

Future Work

- Further develop image space RFI identification, in order to analyze large amounts of data for RFI which may not be easily detectable in Fourier space using existing methods.

These are the guiding questions that we're attempting to answer in the next iterations of our research:

- Develop a method that intelligently and efficiently selects antenna pairs and polarizations in a manner that (1) eliminates redundant examples of RFI and (2) provides a representative sample of all types of RFI that exist within a capture.
- Determine what data representation provides the model with necessary information to make an inference that finds even the faintest RFI.
- Develop weak supervision machine learning methods to further improve U-Net performance.

Using Image Space to Identify Interference

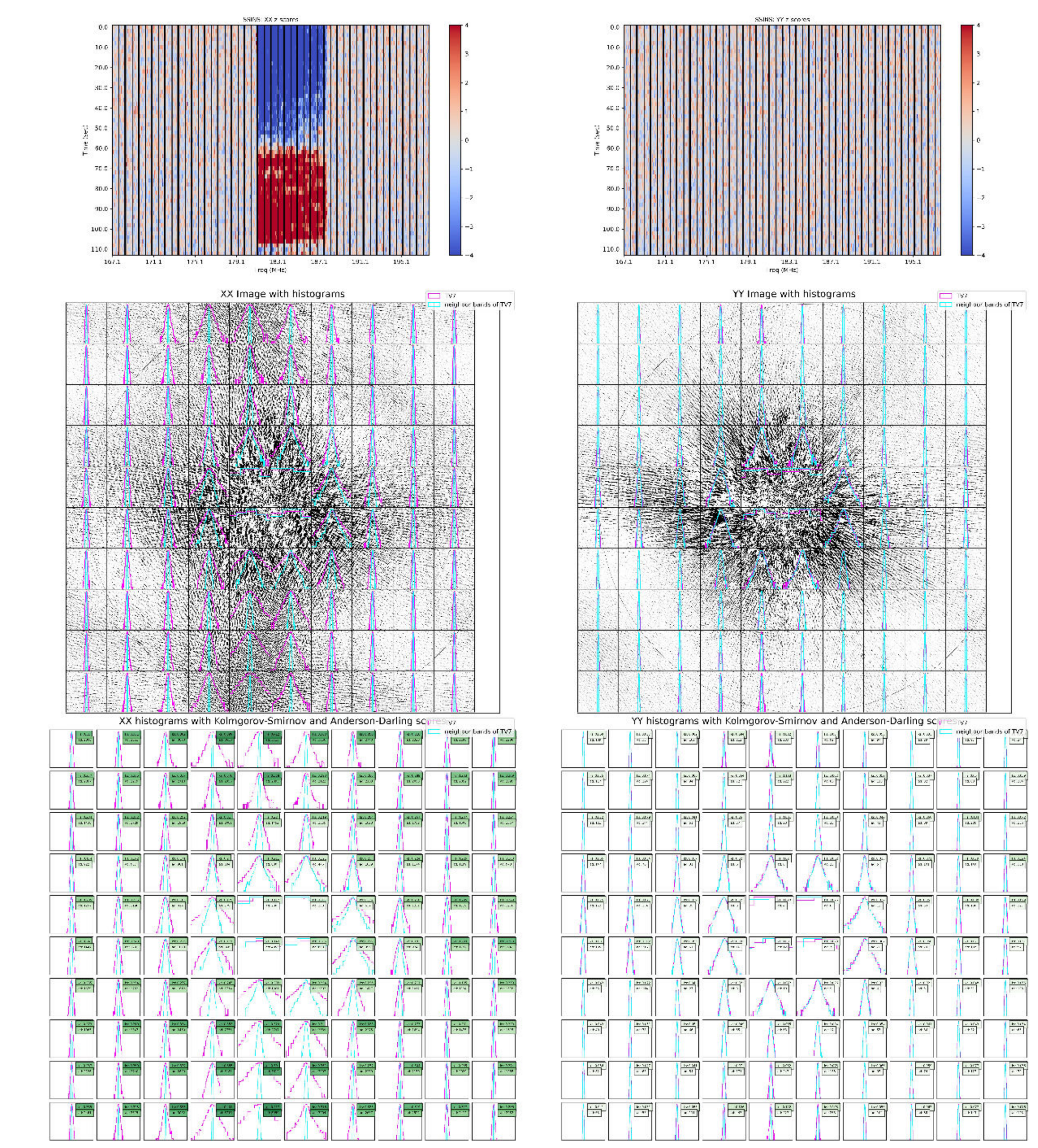

Figure 4. An example of polarized interference detected by the MWA is identified using SSINS. The top panel show the SSINS z-scores as a function of frequency and time, in both XX and YY polarizations. We can see the RFI event is highly concentrated in the TV7 band. In the next panel down, composite images are shown in both polarizations. These composites are summed in frequency and time, giving a snapshot of the whole observation. Overlayed are histograms which show the different brightness distributions for the TV7 band and the immediately neighboring bands, such that the neighboring bands combined have the same number of frequency channels as the TV7 band. It's clear that in the RFI-heavy XX polarization, there are significant differences in the histograms, especially in the area where RFI appears to be present. This can be further seen in the bottom panels, where we highlight results from the Kolmogorov-Smirnov and Anderson-Darling tests of distribution similarity, respectively. These preliminary results show that there may be ways of effectively detecting RFI in the image space directly. New methods using image-space detection could be further used to train ML algorithms to detect RFI.

References

[1] P. Arrubarrena, M. Lemercier, B. Nikolic, T. Lyons, and T. Cass. Novelty detection on radio astronomy data using signatures. *arXiv preprint arXiv:2402.14892*, 2024.

[2] J. Cai. *humidity: Calculate Water Vapor Measures from Temperature and Dew Point*, 2019. R package version 0.1.5.

[3] E. Dong, H. Du, and L. Gardner. An interactive web-based dashboard to track covid-19 in real time. *The Lancet infectious diseases*, 20(5):533–534, 2020.

[4] C. D. Du Toit, T. L. Grobler, and D. J. Ludick. A comparison framework for deep learning rfi detection algorithms. *Monthly Notices of the Royal Astronomical Society*, page stae892, 2024.

[5] F. Gu, L. Hao, B. Liang, S. Feng, S. Wei, W. Dai, Y. Xu, Z. Li, and Y. Dao. Radio frequency interference detection using efficient multi-scale convolutional attention unet. *Monthly Notices of the Royal Astronomical Society*, page stae868, 2024.

[6] J. Kerrigan, P. L. Plante, S. Kohn, J. C. Pober, J. Aguirre, Z. Abdurashidova, P. Alexander, Z. S. Ali, Y. Balfour, A. P. Beardsley, et al. Optimizing sparse rfi prediction using deep learning. *Monthly Notices of the Royal Astronomical Society*, 488(2), 2019.

[7] D. A. Mitchell, L. J. Greenhill, S. M. Ord, G. Bernardi, R. B. Wayth, R. G. Edgar, M. A. Clark, K. Dal, H. Pfister, S. J. Gleadow, et al. The murchison widefield array. *arXiv preprint arXiv:1008.2551*, 2010.

[8] National Centers for Environmental Information. U.s. local climatological data, 2021.

[9] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

NSF awards #2228989 and #2228990

sbach@cs.brown.edu