DATA ANALYSIS PIPELINE FOR GLOBAL NEUTRAL HYDROGEN OBSERVATIONS WITH THE LUNAR-BASED FARSIDE ARRAY. D. Rapetti[1,2,3], K. Tauscher[3,4], J. Mirocha[5], J. O. Burns[3], [1]NASA Ames Research Center, [2]Universities Space Research Association, [3]Center for Astrophysics and Space Astronomy, Department of Astrophysical and Planetary Science, University of Colorado Boulder, David.Rapetti@colorado.edu, [4]Department of Physics, University of Colorado Boulder, [5]Department of Physics & McGill Space Institute, McGill University.

Introduction: FARSIDE (Farside Array for Radio Science Investigations of the Dark ages and Exoplanets) is a Probe-class concept to place a low radio frequency interferometric array on the farside of the Moon [1] (see also a companion abstract from J. O. Burns for technical details on the FARSIDE array).

As primary goals, FARSIDE would enable near-continuous monitoring of the closest stellar systems in the search for the radio signatures of coronal mass ejections and energetic particle events, and would also detect the magnetospheres for the nearest candidate habitable exoplanets. At the same time, FARSIDE would be used to characterize similar activity in our own solar system, from the Sun to the outer planets, including the hypothetical Planet Nine.

Through precision calibration via an orbiting beacon, and exquisite foreground characterization, FARSIDE would also measure the Dark Ages global 21-cm signal at high redshifts z~50–100. We are developing a full pipeline to analyze such sky-averaged 21-cm observations.

Robust constraints: Our data analysis pipeline self-consistently separates global 21-cm signals from large systematics, such as the beam-weighted foreground (which is ~$10^{5\text{-}6}$ times larger than the signal), using a novel pattern recognition methodology. In the first paper of the pipeline series [2], we showed how to obtain optimal basis vectors from signal and foreground training sets to linearly fit both components at the same time with the minimal number of terms that best extracts the signal given its overlap with the foreground. In the second paper of this series [3], we utilize the spectral constraints derived in the first paper to calculate the full posterior probability distribution of any nonlinear signal model of interest. This spectral fit provides the starting point for a Markov Chain Monte Carlo (MCMC) algorithm to sample the signal without traversing the foreground parameter space, as described next.

Efficient analysis: At each step of the nonlinear MCMC calculation, we marginalize over the weights of all linear foreground modes and suppress those with unimportant variations by applying priors gleaned from the foreground training set. Conveniently, the application of foreground priors circumvents the need for selecting a minimal number of foreground modes. Also, the analytical integration (marginalization) over the linear foreground terms (which are typically many in our analyses) drastically reduces the number of required MCMC parameters (which are more computationally expensive to account for), and therefore augments the efficiency of the MCMC exploration. In turn, this allows the complexity of the foreground model to be greatly increased with negligible computational time costs.

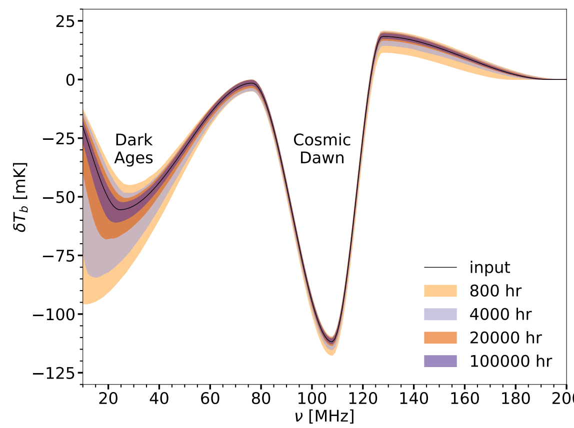

Figure 1: Plot from the second paper of the pipeline series [3]. Full (statistical plus systematic) uncertainties of a turning-point model case in frequency space for four evenly increasing integration times up from our reference value of 800 hours, with the same random seed for noise generation to ensure comparability.

Our method simultaneously describes the selected nonlinear signal model using MCMC parameters, and the foreground linearly via its four Stokes parameters plus multiple observed spectra for different rotation angles or drift-scan time bins, without requiring extra MCMC parameters. Since contrary to the signal, the foreground is polarized and changes in the sky over time, our technique can robustly employ this information to separate the signal from the foreground.

Cases study: Using two nonlinear signal models, one based on EDGES observations [4] (flattened Gaussian model) and the other on phenomenological frequencies and temperatures of theoretically expected extrema (turning point model), we demonstrate the success of the pipeline by recovering the input parameters from several randomly simulated signals at low radio frequencies (10-200 MHz), while rigorously accounting for realistically modeled beam-weighted foregrounds.

For one of the cases studied, Figure 1 shows the increases in constraining power on the spectral shape of the signal, including both the Dark Ages (at lower frequencies) and Cosmic Dawn troughs, when evenly increasing the integration time by factors of 5 from our 800 reference value to 4000, 20000, and 100000 hours. These illustrate what could be increasingly achieved by utilizing many FARSIDE antennas individually to integrate more time. For these particular simulations, these results show that up to the highest integration time used here the constraints are not limited by systematics, which in this case is the overlap between signal and foreground.

Acknowledgments: This work was directly supported by the NASA Solar System Exploration Virtual Institute Cooperative Agreement 80ARCC017M0006. J.M. acknowledges support from a CITA National Fellowship.

References: [1] Burns J. O., Hallinan G., Lux J., Teitelbaum L., Kocz J., MacDowall R., Bradley R., Rapetti D., Wu W., Furlanetto S., Austin A., Romero-Wolf A., Chang T.-C., Bowman J., Kasper J., Anderson M., Zhen Z., Pober J., Mirocha J., *arXiv: 1911.08649*. [2] Tauscher K., Rapetti D., Burns J. O., Switzer E. (2018) *ApJ, 853, 187*. [3] Rapetti D., Tauscher T., Mirocha J., Burns J. O., *arXiv: 1912.02205*. [4] Bowman J. D., Rogers A. E. E., Monsalve R. A., Mozdzen T. J., Mahesh N. (2018) *Nature*, 555, 67-70.