GLOBAL NEUTRAL HYDROGEN DATA ANALYSIS PIPELINE FOR A LUNAR-BASED SATELLITE.

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: DAPPER (Dark Ages Polarimeter PathfindER) is a NASA-funded mission concept to use a low radio frequency instrument for sky-averaged 21-cm observations above the farside of the Moon, avoiding Earth's ionospheric effects and terrestrial radio frequency interference.

Challenges faced: The study of neutral hydrogen (HI) astrophysics and cosmology in the early Universe requires the extraction of a HI signal that is about 10^{4-6} times smaller than the foreground. In addition, this is currently to be performed with unmeasured, large solid-angle, chromatic beams that introduce spectral distortions, complicating the data analysis.

Novel pattern recognition as proposed solution: We recently completed a data analysis pipeline to self-consistently separate the global 21-cm signal from the beam-weighted foreground. In the first paper of the pipeline series [1], we showed how to employ training sets based on HI theory, beam simulations and foreground maps to optimal basis vectors for modelling the signal and beam-weighted foreground. The pipeline then utilizes these models to linearly fit both components simultaneously with the minimal number of terms that best extracts the signal given its overlap with the foreground.

Robust constraints: In the second paper of this series [2], 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.

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. The latter are more computationally expensive to account for, so this strategy significantly augments the efficiency of the MCMC exploration. In turn, this allows the complexity
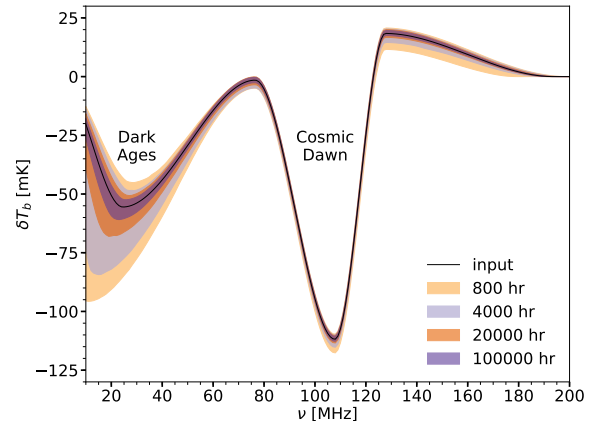

Figure 1: Plot from the second paper of the pipeline series [2]. Full (statistical plus systematic) uncertainties of a turning-point model case in frequency space for four evenly increasing integration times up from our reference value of 800 hours. The same random seed is used for noise generation in all the cases for comparison purposes.

of the foreground model to be greatly increased with negligible computational time costs.

Cases study: Using two nonlinear signal models, one based on EDGES observations [3] (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. Figure 1 shows the increases in constraining power on the spectral shape of the signal for one of these cases, when evenly increasing the integration time. These illustrate what could be increasingly achieved by utilizing many antennas individually to integrate more time.

References: [1] Tauscher K., Rapetti D., Burns J., Switzer E. (2018) ApJ, 853, 187. [2] Rapetti D., Tauscher T., Mirocha J., Burns J., arXiv: 1912.02205. [3] Bowman J., Rogers A., Monsalve R., Mozdzen T., Mahesh N. (2018) Nature, 555, 67-70.