Exploring the Effects of Foreground Removal Techniques and Instrumental Systematics on Observations of the 21 cm Neutral Hydrogen Signal

Benjamin L. Hoscheit
Department of Physics, University of Wisconsin Madison
Madison, WI, U.S.

Abstract

There is a substantial and growing interest in cosmology to study the 21 cm signal emitted or absorbed by large abundances of neutral hydrogen (HI) in the vast cosmic web of the universe. Such study has the potential to allow scientists to map the matter distribution of the universe over nearly its entire history. However, in order to unlock this rich potential, one must develop the ability to distinguish and remove large contaminating signals from ionized gas in the foregrounds of the 21 cm signal in addition to having high precision control of instrumental systematics. As such, we first understand and construct the associated ability to remove four common astrophysical foregrounds to 21 cm experiments. We then assess the degree to which three common instrumental systematics introduce systematic un-smoothing effects into the analysis of these astrophysical foregrounds. This study works within the framework of simulated Green Bank Telescope radio beam data.

1. Introduction

One of the primary goals of the field of cosmology is to map the evolution of structure in the universe over the approximately 14 billion years of its existence. This may seem like a wildly unreasonable task for any sane science researcher. Although, with the advent of what has come to be known as 21 cm cosmology, this process may be more seamless than previously thought. 21 cm cosmology specifically refers to the mapping of redshifted 21cm wavelength hyperfine atomic transition radiation emitted or absorbed by large abundances of neutral hydrogen (HI) in the vast cosmic web of the universe. NASA has even recommended a lunar mission referred to as the "Cosmic Dawn Mapper" in their most recent 'Astrophysics Roadmap', further stating that the radio maps produced by such a mission "will trace an intricate structure of ribbons, tunnels, and bubbles in the 200-million-year-old universe as the light of the first stars burns through the fog of absorbing hydrogen that fills intergalactic space" [9]. Thus, if we can map the three-dimensional distribution of HI in the universe at arbitrary redshifts, then we can accurately probe many important events throughout the universe's evolution. Where, redshift simply refers to a measure of how the expanding universe affects, or 'redshifts', the light that propagates through it and is typically denoted by "z".

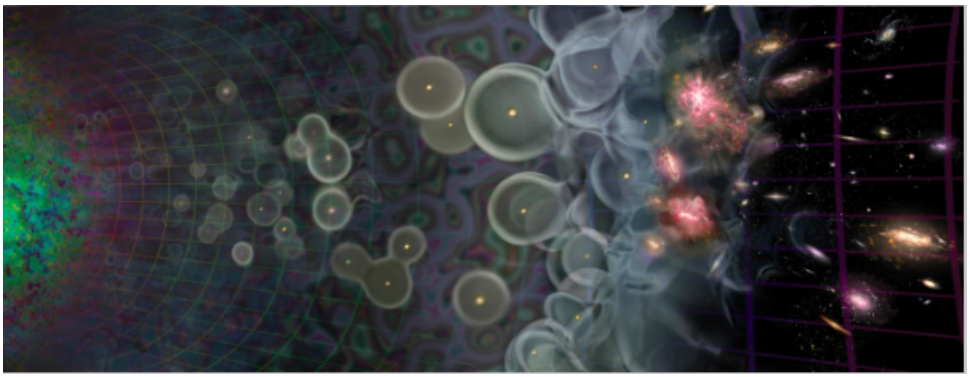

Figure 1: Artist's depiction of the evolution of neutral hydrogen (HI) in the universe. Measurements of HI via its characteristic 21 cm signal have the potential to give us extremely valuable information about important evolutionary events as well as the large scale matter distribution throughout the universe's evolution.

However, nature almost never gives out such large science returns for free. In order to justify measurements of HI at very high redshifts, one must first verify the science techniques at lower redshifts, namely, through cross-correlation intensity mapping. This technique builds off of the more well-developed, traditional galaxy survey technique of pin-pointing individual galaxies and constructing a map of the large scale structure from these counting number densities. This 21 cm intensity mapping cross-correlation method has recently shown to be quite fruitful with Tzu-Ching Chang, Ue-Li Pen, Kevin Bandura, and Jeffrey Peterson publishing the first high redshift detection ($z \sim 0.8$) of large-scale HI via observations of the 21 cm radio signal using the Green Bank Telescope (GBT) [4]. With these promising 21 cm intensity mapping cross-correlation results, many collaborations have simultaneously begun pushing onward to redshifts beyond this scale of galactic verification through autocorrelation detection methods. To name a few of these high redshift experiments with their respective operating redshift ranges: The Hydrogen Epoch of Reionization Array (HERA) [7] from z = 6 to z = 12, The Murchison Widefield Array (MWA) [2] from z = 6 to z = 12, and the Dark Ages Radio Explorer (DARE) [3] from z = 11 to z = 35. In particular, being a space-based observatory, the DARE mission will orbit the Moon for three years to escape the radio frequency interference (RFI) that plagues many of the terrestrial 21 cm intensity mapping experiments.

2. Background

Professor Peter Timbie of the University of Wisconsin-Madison (UW) Department of Physics has been an active member of the 21 cm cosmology research community through his involvement in several major related projects. First, and potentially most directly relevant to this proposed research project, Professor Timbie, and his recently graduated graduate student Christopher Anderson, are part of a larger collaboration of cosmology researchers using the Green Bank Telescope (shown in Fig. 2) in Green Bank, WV and the Parkes Observatory in Parkes, New South Wales, Australia to carry out 21 cm intensity mapping observations. In particular, Professor Timbie and Christopher Anderson have been heavily involved in the first high redshift detections of HI via the 21 cm signal using the GBT, in both cross-correlation [4], [10] and autocorrelation [13] detection methods.

Figure 2: The Green Bank Telescope located in Green Bank, WV. Professor Peter Timbie of the UW Department of Physics, and his recently graduated graduate student Christopher Anderson, are part of a collaboration of cosmology researchers using the Green Bank Telescope to measure large scale HI via intensity mapping techniques. Image courtesy of NRAO/AUI.

Furthermore, Professor Timbie is also deeply involved in the development and design of several present and future science instruments to perform such HI observations. First, Professor Timbie and Christopher Anderson are also involved in a project to design and upgrade the GBT primary focus receiver to an array of seven hexagonally packed receivers for the purpose of faster data collection with the radio telescope. Second, Professor Timbie and Christopher Anderson are also heavily involved in the development of the Tianlai Cylinder Array interferometer telescope in China

[5]. Professor Timbie, Christopher Anderson, and a former UW undergraduate, Aleksander Cianciara, have been working towards the simulation and testing of feed antennas for the Tianlai Cylinder Array experiment [6]. Third, Professor Timbie is involved in the design and the development of The Hydrogen Intensity and Real-time Analysis eXperiment (HIRAX) to be completed in full development in South Africa [11].

Finally, Professor Timbie and a former UW postdoctoral researcher, Le Zhang, have previously worked to publish a statistical foreground removal technique used for interferometric 21 cm observations [15]. In this research paper, Le Zhang, Professor Timbie, and their collaboration develop the effectiveness of their foreground removal technique, referred to as HIEMICA (HI Expectation-Maximization Independent Component Analysis), using simulated 21 cm signal and astrophysical foreground data to better understand quantitatively how each foreground effects the science results of the 21 cm signal.

3. Simulating 21 cm Astrophysical Foregrounds

First, an astrophysical foreground is any secondary astrophysical object whose radiation is cosmologically redshifted to be in the frequency range and located within the angular coverage of the desired detector, i.e. between 680 and 920 MHz for the GBT. The main difficulty arises from the fact that the astrophysical foregrounds typically dwarf the 21 cm radiation in signal amplitude. The most prominent foreground is synchrotron emission foregrounds, caused by electrons spiraling in galactic magnetic fields. This foreground is of order one kelvin (K) whereas the 21 cm signal is of order one millikelvin (mK). The effect of a successful foreground cleaning technique can be seen in Fig. 3 from the work done in [10] with the GBT.

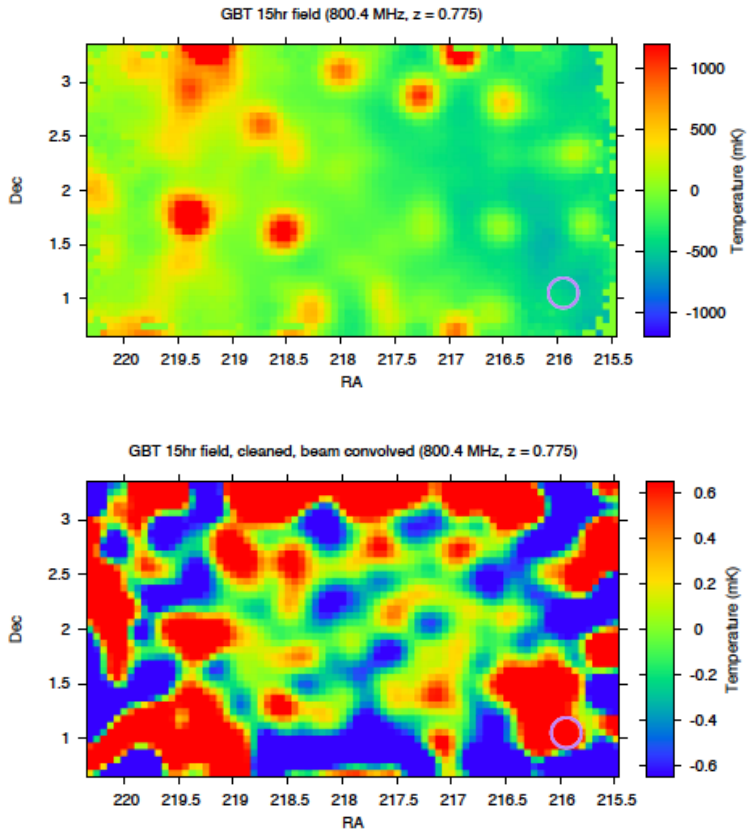

Figure 3: Maps of the GBT 15hr field at the band-center. Plot axes are Declination (Dec), Right Ascension (RA), and Temperature (mK) respectively. (a) Combined signal, including astrophysical foregrounds, instrumental systematics, and the 21 cm signal. (b) 21 cm signal plus residual foreground signal, after applying foreground removal techniques and accounting for instrumental systematics. The Singular Value Decomposition (SVD) cleaning method used here was able to reduce the overall beam signal down to the 1 mK level of the 21 cm signal with marginal 21 cm signal lose. Images courtesy of [10].

3.1. Foreground Models We have analyzed the effects of four main astrophysical foregrounds, namely, galactic synchrotron emission, extragalactic point sources, galactic free-free emission, and extragalactic free-free emission. We have done so based off of the models described in [12] and with the help of former UW postdoctoral researcher Le Zhang [15]. The power spectrum model used to describe the four main astrophysical foreground sources excluding strong point sources is given by

$$C_f(\ell,\nu,\nu') = A\left(\frac{1000}{\ell}\right)^{\beta}\left(\frac{\nu_f^2}{\nu\nu'}\right)^{2\alpha}\exp\Big(-\frac{\ln^2(\nu/\nu')}{2\xi^2}\Big)\,, \tag{1}$$

where ν_f is 130 MHz. The four different astrophysical foreground types and their respective parameters are shown in the table in Fig. 4. However, [12] argues that poisson fluctuations begin to dominate the point source foregrounds for point sources above 0.1 mJy, and these are not included in the model in [15]. Therefore, we use a model for extragalactic point sources in [12] to describe clustered point sources below 0.1 mJy and then populate the map with unclustered strong point sources from 0.1 mJy to 10 Jy using the method described below.

	$A(\mathrm{mK}^2)$	β	α	ξ	$\Delta\alpha$
extragalactic point sources	57.0	1.1	2.07	1.0	0.2
extragalactic free-free	0.014	1.0	2.10	35	0.03
Galactic synchrotron	700	2.4	2.80	4.0	0.15
Galactic free-free	0.088	3.0	2.15	35	0.03

Figure 4: The four different astrophysical foreground types included in our simulation and their respective parameters from the models described in [12] . Each astrophysical foreground type is uncorrelated with the others and with the 21-cm signal. These astrophysical foregrounds are assumed to be spectrally smooth given by the power law description of their power spectra in Eq. 1. Image courtesy of [12].

For our strong point source population, we turn to [8]. We poisson sample point sources over the flux range for S[mJy] $\in$ [0.1, 10,000] from the following distribution

$$\log\Big[\frac{(\mathrm{dN/dS})}{\mathrm{S}^{-2.5}}\Big] = \sum_{\mathrm{i}=0}^{6}\mathrm{a_i}\left[\log\Big(\frac{\mathrm{S}}{\mathrm{mJy}}\Big)\right]^{\mathrm{i}}, \tag{2}$$

where $a_0 = 0.841$, $a_1 = 0.540$, $a_2 = 0.364$, $a_3 = -0.063$, $a_4 = -0.107$, $a_5 = 0.052$, and $a_6 = -0.007$.

We then randomly assign a spectral index to each strong point source from a Gaussian distribution with an average index given by

$$\overline{\gamma} = -0.83 - 0.13\log_{10}(\mathrm{S[mJy]}/1000)\,, \tag{3}$$

and a standard deviation of

$$\delta\gamma = 0.2\,, \tag{4}$$

where these statistical parameters come from those motivated in [12]. The results of this additional modeling of strong extragalactic point sources into the simulation are shown in Fig. 5(a).

4. Simulating GBT Instrumental Systematics and Noise

Second, an instrumental systematic refers to any systematic error introduced in an observer's data as produced by the instrument's response to it's detection environment. For our primary focus being in application to the GBT radio

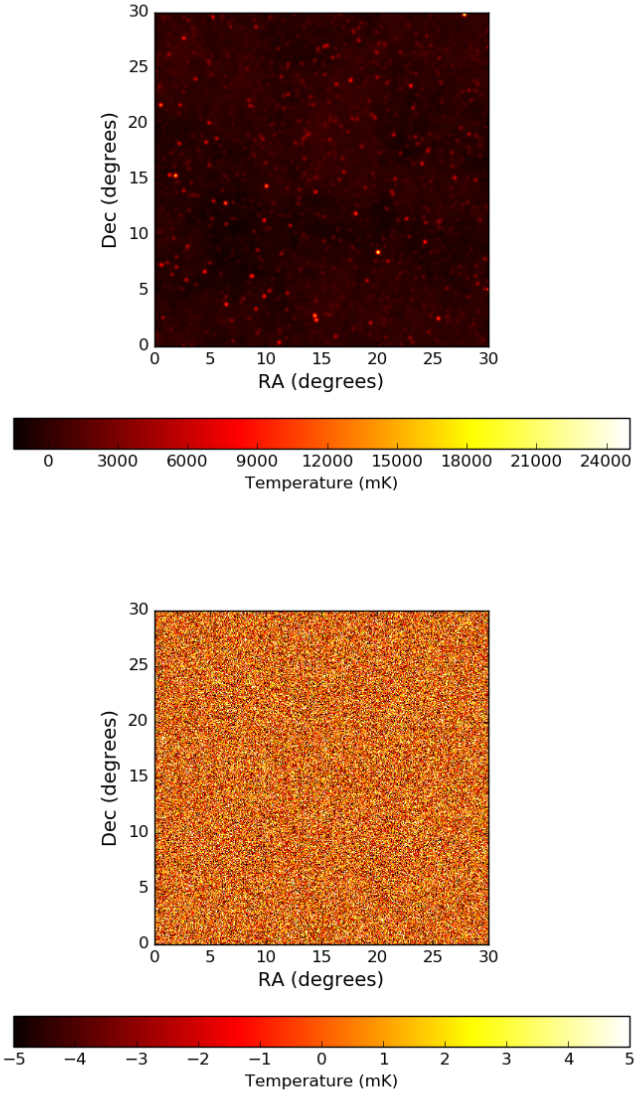

Figure 5: (a) Frequency slice at frequency channel 100 of 256 of a simulated gaussian convolved map with all four main foregrounds and gaussian random noise. This map now includes simulated strong point sources. These strong point sources are shown by the bright red and yellow dots in the map. Note that the temperature scaling is at kelvin level in parallel with the magnitude of foregrounds seen in real GBT beam data (Figure 3(a)). (b) Frequency slice at frequency channel 100 of 256 of the simulated map in (a) after we have applied our SVD foreground cleaning procedure by removing the 10 largest SVD modes. This foreground cleaning procedure is described in more detail in Section 5. We see that this technique allows us to subtract off these foregrounds to get down to the $\sim$5 mK level of the 21 cm signal in this simulation.

telescope, one example of this effect is beam convolution causing what is referred to as "mode-mixing". This instrumental systematic causes known smooth foregrounds, such as galactic synchrotron emission, to become un-smooth, and, hence, making it harder to distinguish this very bright un-smooth foreground from the random, un-smooth 21 cm signal. Our second main area of progress has been in incorporating three main instrumental systematics into our simulations. These three main instrumental systematics are beam convolution effects, bandpass calibration effects, and system noise.

4.1. Airy Beam Pattern Convolution Currently, successful 21 cm intensity mapping observational campaigns involving members of our collaboration make use of single-dish radio telescopes such as the GBT [10], [13] or the Parkes Radio Telescope in Australia [1]. These observational campaigns have incorporated into their pipelines the procedure of convolving their radio beam maps to a common beam resolution in order to minimize mode mixing. To do this, they assume the radio beam corresponds to a gaussian shape fitted by looking at bright point sources. This is simply because the measurements of the beam pattern only have high signal to noise in the main lobe of the beam. However, in reality, the real radio beam will have side-lobes that are frequency dependent. The current pipeline tries to convolve the maps to a common resolution because the size of the fitted gaussian beam is smaller at higher frequency. This procedure assumes that the beam is a gaussian profile because that is easier computationally and also because the empirical beam pattern is not measured well enough to have a more precise empirically motivated model. As such, we have explored the possibility of convolving our simulated GBT radio beam map to a fitted airy disk beam pattern, which does take into account the more realistic GBT beam having side-lobes.

The formula for intensity of the beam as a function of angle for an airy disk beam pattern is given by

$$I(\theta) = I_0 \left(\frac{2J_1(k\ R\sin(\theta)}{k\ R\sin(\theta)} \right)^2, \tag{5}$$

where I_0 is the maximum intensity of the pattern at the airy disc center, $k = 2\pi/\lambda$, and R is the radius of the aperture (in this case, $2R = 100$ meters for the GBT). J_1 is a spherical Bessel function. Our simulated data cube covers the frequency range of 700 to 900 MHz in 256 channels for the range in λ. Our airy beam function is plotted in Fig. 6 for three common frequencies used in the GBT 21 cm intensity mapping beam data, namely, 700, 800, and 900 MHz.

4.2. Gaussian Random Noise and Bandpass Fluctuations After we have convolved to a frequency dependent Airy disk beam, a more realistic approximation than our current GBT beam model, we then add two separate realizations of uncorrelated gaussian random noise to the maps. For this, we assume a 25 kelvin system temperature, as is the case for the real GBT receiver, and an observation time per pixel equivalent to that in the 2013 GBT auto-power results [13]. Following this, we then correlate maps with two different noise realizations. Once we have simulated the gaussian random noise into our maps, we then model in fluctuations in the bandpass seen at GBT. We do this by approximating fluctuations in the bandpass using gaussian statistics with a normalized mean fluctuation and a standard deviation of these fluctuations of 2%.

5. Discussion and Conclusion

Given our core goal of accurately simulating the effects of astrophysical foregrounds and instrumental systematics on 21 cm intensity mapping efforts using the GBT, we discuss the above simulation results in light of the current procedure used by the GBT intensity mapping collaboration to remove astrophysical foregrounds and account for such instrumental systematics. This foreground removal technique is referred to as the Singular Value Decomposition (SVD) Principal Component Analysis (PCA) method [4], [10], [13]. In brief, one first obtains maps of the sky that include astrophysical foregrounds, instrumental systematics, detector noise, and the 21 cm signal. Then, one creates a covariance matrix with this map data and performs an eigenvalue decomposition of this matrix. An assessment of the frequency structure of the eigenvectors then allows one to distinguish, and, further, remove the astrophysical foreground or instrumental systematic from the bulk radio signal. Following this, the ultimate goal is hopefully reached as one may now use this remaining signal to determine the HI power spectrum.

A pictorial representation of the SVD process described above as applied to real GBT data is shown in Fig. 7. As one

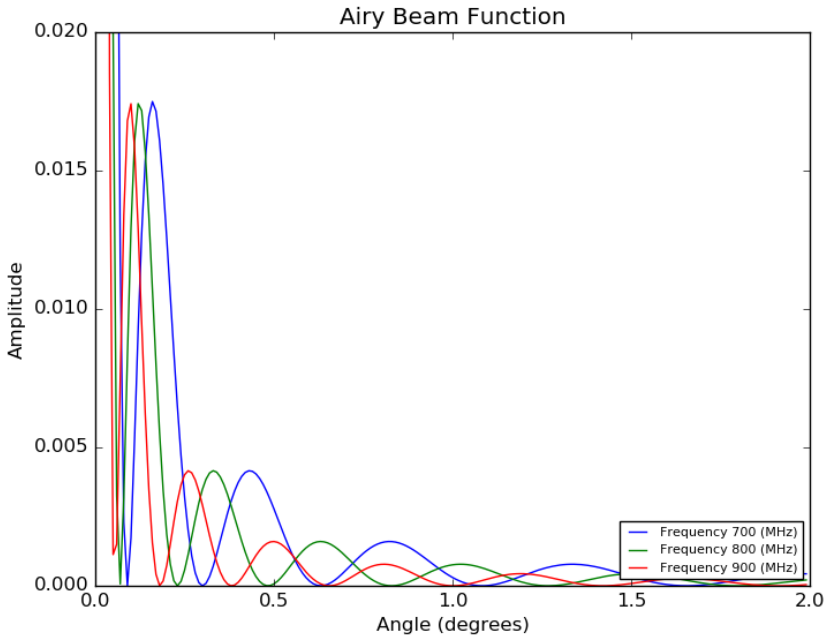

Figure 6: Airy disk beam pattern amplitude shown for three common frequencies used in the GBT 21 cm intensity mapping beam data, namely, 700 (blue), 800 (green), and 900 (red) MHz plotted versus angle (in degrees). We have used this frequency dependent airy beam function as a means to model the side-lobes of the real GBT radio beam.

can see from Fig. 7, the simulations that we have created thus far with the additional strong point sources and added airy beam convolution are on the right track towards replicating the real astrophysical foregrounds and instrumental systematics at the GBT. In our case, it appears that we have not added in enough instrumental systematics to mix the SVD modes of our simulated foregrounds. This lacking is manifest in the fact that the simulation SVD eigenvalue spectrum in Fig. 7(a) is able to get down to the level of the 21 cm signal ($\sim 10^{-4}$ in map-space amplitude) using less SVD modes (10 modes) than the real GBT SVD eigenvalue spectrum shown in Fig. 7(b).

The ideal goal of the simulation is an exact replication of the SVD eigenspectrum shown in Fig. 7(b) with the simulation SVD eigenspectrum shown in Fig. 7(a). This would correspond to an exact understanding of all of the relevant astrophysical foregrounds and instrumental systematics currently seen at the GBT in 21 cm intensity mapping. One would then use the information from such a simulation to account for the effects of the now identified astrophysical foregrounds and instrumental systematics.

As for the goal of testing the GBT collaboration's current foreground removal technique, our main results are that the current foreground removal technique is robust against the astrophysical foregrounds and instrumental systematics included in our simulations; although, we do lose a bit of 21 cm signal in performing this subtraction. These results can be directly visualized in Fig. 5. After applying our SVD foreground cleaning technique (removing the 10 largest SVD modes) to our best simulated map including the 4 main astrophysical foregrounds and 21 cm signal, along with instrumental systematics in Fig. 5(a), we are left with the map in Fig. 5(b). The resulting cleaned map is at the temperature level ($\sim$5mK) necessary to probe the underlying 21 cm signal in these simulations.

For future work, a comparison with current GBT results suggests that we must add more complex astrophysical foregrounds and more numerous instrumental systematics into to the simulation. For example, we seek to add in polarized foregrounds which have a frequency dependent polarization angle due to Faraday rotation. Furthermore, when these polarized foregrounds interact with the GBT, polarization leakage into intensity occurs, and, this instrumental effect couples this polarized foreground signal to an un-smooth spectrum, again, making it difficult to remove them from the bulk radio signal.

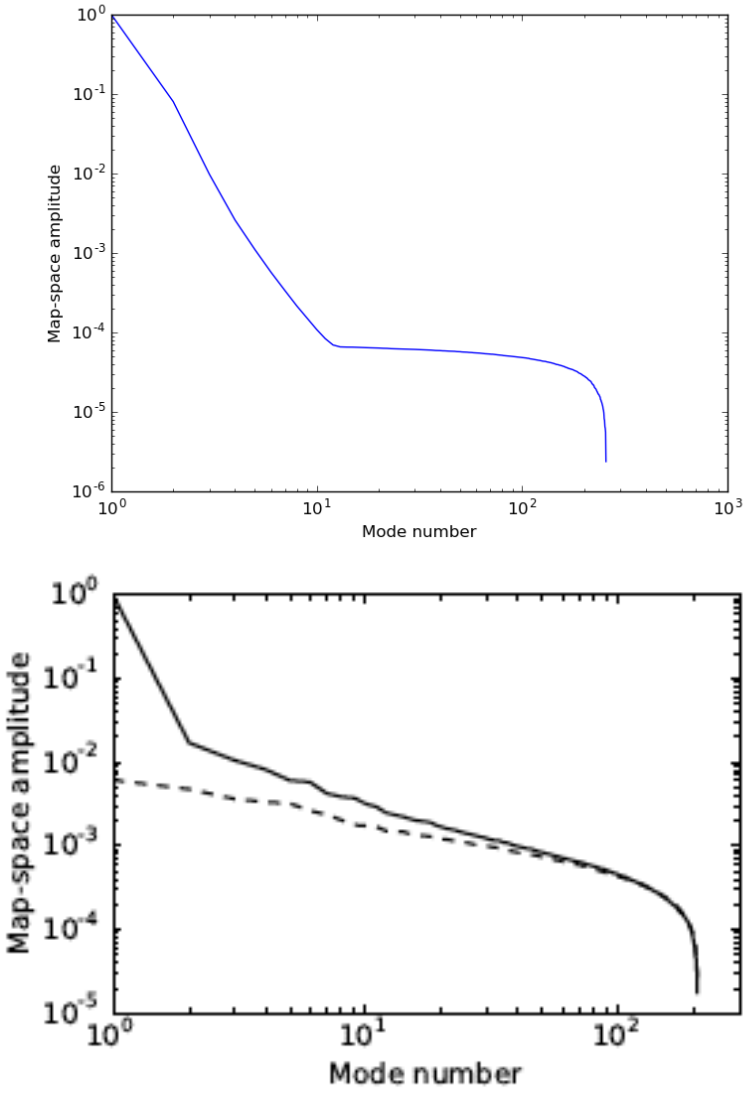

Figure 7: (a) Square root of the covariance eigenvalue spectrum corresponding to our simulations of the foregrounds and instrumental response seen using the GBT. The eigenvalue spectrum is normalized in accordance with the largest foreground mode. In this case, the largest foreground mode corresponds to the smooth frequency structure associated with the mean galactic synchrotron emission across the map. (b) Square root of the covariance eigenvalue spectrum corresponding to real GBT measurements [14], normalized in accordance with the largest foreground mode. Image courtesy of [14].

Acknowledgments

B.L.H would like to acknowledge the generous support of the Wisconsin Space Grant Consortium through an undergraduate Research Fellowship. B.L.H would also like to thank his advisors on this research project, Professor Peter Timbie and graduate student Christopher Anderson, for their insightful comments and supervision during the completion of this research project.

References

[1] Anderson, C. J., Luciw, N. J., & Li, Y.-C., et al. 2017, MNRAS submitted

[2] Bowman, J. D., Cairns, I., & Kaplan, D. L., et al. 2013, PASA, 30, 28

[3] Burns, J. O., Lazio, J., & Bale, S., et al. 2011, AdSpR, 49, 433B

[4] Chang, T.-C., Pen, U.-L., & Bandura, K., et al. 2010, Nature, 466, 463

[5] Chen, X. 2012, International Journal of Modern Physics Conference Series, 12, 256

[6] Cianciara, A. J., Anderson, C. J., & Chen, X., et al. 2017, under revision (JAI)

[7] DeBoer, D. R., Parsons, A. R., & Aguiree, J. E., et al. 2016, arXiv:1606.07473

[8] Huynh, M. T., Jackson, C. A., & Norris, R. P., et al. 2005, AJ, 130, 1373

[9] Kouveliotou, C., Agol, E., & Batalha, N., et al. 2014, arXiv:1401.3741

[10] Masui, K. W., Switzer, E. R., & Banavar, N., et al. 2013, ApJ, 763, 1

[11] Newburgh, L. B., Bandura, K., & Bucher, M. A., et al. 2016, in Proc. SPIE, 9906, 99065X

[12] Santos, M. G., Cooray, A., & Knox, L., 2005, ApJ, 625, 575

[13] Switzer, E. R., Masui, K. W., & Bandura, K., et al. 2013, MNRAS, 434, L46

[14] Switzer, E. R., Chang, T.-C., & Masui, K. W., et al. 2015, ApJ, 815, 1

[15] Zhang, L., Bunn, E. F., & Karakci, A., et al. 2016, ApJS, 222, 3