How well can we measure and understand foregrounds with 21 cm experiments?

Adrian Liu[1]⋆, Max Tegmark[1]

[1]*Dept. of Physics and MIT Kavli Institute, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA*

1 December 2011

ABSTRACT

Before it becomes a sensitive probe of the Epoch of Reionization, the Dark Ages, and fundamental physics, 21 cm tomography must successfully contend with the issue of foreground contamination. Broadband foreground sources are expected to be roughly four orders of magnitude larger than any cosmological signals, so precise foreground models will be necessary. Such foreground models often contain a large number of parameters, reflecting the complicated physics that governs foreground sources. In this paper, we concentrate on spectral modeling (neglecting, for instance, bright point source removal from spatial maps) and show that 21 cm tomography experiments will likely not be able to measure these parameters without large degeneracies, simply because the foreground spectra are so featureless and generic. However, we show that this is also an advantage, because it means that the foregrounds can be characterized to a high degree of accuracy once a small number of parameters (likely three or four, depending on one's instrumental specifications) are measured. This provides a simple understanding for why 21 cm foreground subtraction schemes are able to remove most of the contaminants by suppressing just a small handful of simple spectral forms. In addition, this suggests that the foreground *modeling* process should be relatively simple and will likely not be an impediment to the foreground subtraction schemes that are necessary for a successful 21 cm tomography experiment.

Key words: Cosmology: Early Universe – Radio Lines: General – Techniques: Interferometric – Methods: Data Analysis

1 INTRODUCTION

In coming years, 21 cm tomography has the potential to vastly expand our knowledge of both astrophysics and cosmology, with its unique ability to be a direct probe of the Dark Ages and the Epoch of Reionization (Madau et al. 1997; Tozzi et al. 2000a,b; Iliev et al. 2002; Furlanetto et al. 2004; Loeb & Zaldarriaga 2004; Furlanetto et al. 2004; Barkana & Loeb 2005; Mack & Wesley 2008; Furlanetto et al. 2009), as well as its potential to provide constraints on cosmological parameters at an unprecedented level of accuracy (McQuinn et al. 2006; Santos & Cooray 2006; Bowman et al. 2007; Mao et al. 2008; Furlanetto et al. 2009; Wyithe et al. 2008; Chang et al. 2008; Morales & Wyithe 2010). First steps towards realizing this potential are being taken with the construction, testing, and initial science runs of low-frequency radio arrays such as PAPER (Parsons et al. 2010), LOFAR (Garrett 2009), MWA (Lonsdale et al. 2009), CRT (Peterson et al. 2006), GMRT (Paciga et al. 2011), and the Omniscope (Tegmark & Zaldarriaga 2010). However,

many obstacles still need to be overcome before cosmological measurements can be performed on the resulting data. For instance, foreground contamination (both from within our galaxy and from extragalactic sources) is expected to be at least four orders of magnitude brighter than the expected cosmological signal (de Oliveira-Costa et al. 2008), so any viable data analysis scheme must also contain a robust foreground subtraction algorithm.

Foreground subtraction is a problem that has been studied extensively in the literature, and many schemes have been proposed for tackling the issue. Early ideas focused on using the angular structure of the foregrounds to separate them from the cosmological signal (Di Matteo et al. 2002, 2004; Oh & Mack 2003; Santos et al. 2005; Zaldarriaga et al. 2004), while most recent proposals focus on line-of-sight (*i.e.* spectral) information (Wang et al. 2006; Gleser et al. 2008; Jelić et al. 2008; Harker et al. 2009; Bowman et al. 2009; Liu et al. 2009a,b; Harker et al. 2010; Petrovic & Oh 2011). By making use of spectral information, these proposals take advantage of the extremely high spectral resolution available in all 21 cm experiments, and indeed it was shown in Liu & Tegmark (2011) that because of the nature of the fore-

---

⋆ E-mail: acliu@mit.edu

grounds and the instrumental parameters, an optimal estimation of the power spectrum should involve a foreground subtraction scheme that operates primarily using frequency information.

Most line-of-sight proposals so far have been *blind* schemes[1] in the sense that they do not require any prior foreground modeling. All such proposals take advantage of the smooth nature of the foreground spectra, and separate out the rapidly fluctuating cosmological signal by (for instance) subtracting off a predetermined set of low-order polynomials (Bowman et al. 2009; Liu et al. 2009a,b) or by imposing a predetermined filter in Fourier space (Paciga et al. 2011; Petrovic & Oh 2011). The blind nature of these schemes may seem at first to be an advantage, because the low frequency ($\sim$ 50 to 300 MHz) regime of radio foregrounds is as yet fairly unconstrained observationally and most models are based on extrapolations and interpolations from other frequencies, where the instruments are optimized for different science goals. However, even if our foreground models are not entirely accurate initially, a non-blind scheme can always be performed iteratively until the models converge to the true measured foregrounds. Moreover, blind schemes do not improve as one makes better and better measurements of the foregrounds, whereas non-blind schemes will continually improve as measurements place increasingly strong constraints on foreground models. In this paper we therefore examine the foreground modeling process, and determine whether or not it will be possible to construct a foreground model that is good enough for foreground subtraction purposes if only a small number of independent parameters can be measured. In the spirit of "one scientist's noise is another scientist's signal", we also quantify the ability of 21 cm experiments to place constraints on phenomenological foreground parameters.

The rest of this paper is organized as follows. In Section 2 we introduce our foreground model, and we show in Section 3 that the foregrounds can be described to an extremely high precision using just three parameters. This is good news for 21 cm cosmology, for it implies that by measuring just a small handful of empirical parameters, it is possible to construct a foreground model that is sufficiently accurate for high precision foreground subtraction. However, this also implies that the large number of (physically motivated) parameters present in a typical foreground parameterization are redundant and highly degenerate with each other, so the prospects for using 21 cm experiments to gain a detailed understanding of foreground physics are bleak. We show this in Section 4, and summarize our conclusions in Section 5.

2 FOREGROUND AND NOISE MODEL

For the purposes of this paper, we limit our analyses to foregrounds in the frequency range 100 to 250 MHz, which roughly speaking covers the "sweet spots" of most 21 cm experiments that are designed to probe the Epoch of Reionization. At these frequencies, there are three dominant sources

of foreground contamination (Shaver et al. 1999; Wang et al. 2006):

(i) Extragalactic point sources.
(ii) Synchrotron emission from our Galaxy.
(iii) Free-free emission from our Galaxy.

Since foreground subtraction is best done using spectral information (Liu & Tegmark 2011), we will ignore the spatial structure of these foregrounds for this paper, as well as any polarization structure[2]. We now examine the spectra of each of these sources, and in particular compute their means and variances, which we define in the next section. With some minor modifications, our models are essentially those of Liu & Tegmark (2011).

2.1 Definition of the mean and the foreground covariance

Consider a multifrequency sky map of brightness temperature T from a typical 21 cm tomography experiment. The map $T(\hat{\mathbf{r}}, \nu)$ is a function of the direction in the sky $\hat{\mathbf{r}}$ as well as the frequency ν. We define a random function $x(\nu)$ such that

$$x(\nu) \equiv T(\hat{\mathbf{r}}, \nu) \qquad (1)$$

is the spectrum measured at a randomly chosen pixel centered at $\hat{\mathbf{r}}$. Since the cosmological signal is expected to be dwarfed by the foregrounds, this spectrum will essentially be a spectrum of the foregrounds.

To quantify the statistical properties of this random function, we can calculate its mean and covariance. In principle, doing so requires taking the ensemble averages. In practice, we instead take averages of the pixels of the sky map. The mean is thus given by

$$m(\nu) = \langle x(\nu) \rangle \equiv \frac{1}{\Omega} \int T(\hat{\mathbf{r}}, \nu) d\Omega, \qquad (2)$$

where Ω is the total area covered by the sky map[3]. The statistical covariance can be similarly defined in the usual way:

$$C(\nu, \nu') \equiv \langle x(\nu)x(\nu') \rangle - m(\nu)m(\nu'), \qquad (3)$$

where

$$\langle x(\nu)x(\nu') \rangle = \frac{1}{\Omega} \int T(\hat{\mathbf{r}}, \nu)T(\hat{\mathbf{r}}, \nu') d\Omega. \qquad (4)$$

2.2 Extragalactic point sources

Extragalactic point sources can be thought of as consisting of two populations. The first consists of bright, isolated point sources that can be resolved by one's instrument. Following previous foreground studies, we assume that these bright point sources have already been removed, prior to the main

---

[1] See Liu & Tegmark (2011) for an example that is *not* blind.

[2] See de Oliveira-Costa et al. (2008) for a discussion of the spatial structure of foregrounds, and Bernardi et al. (2010) and Geil et al. (2010) for discussions of polarized foregrounds and their removal.
[3] Note that this sky map need not cover all 4π steradians of the full sky. For instance, in an estimation of the power spectrum one may choose to use only the data from the cleanest parts of the sky.

foreground cleaning step that attempts to subtract off the rest of the foregrounds. Techniques such as forward modeling (Bernardi et al. 2011) and peeling have been explored for this purpose, and peeling simulations suggest that the bright sources can be removed down to $S_{max} \sim 10$ to $100\,\text{mJy}$ (Pindor et al. 2011). To be conservative we will use $S_{max} = 100\,\text{mJy}$ in all calculations that follow.

Below S_{max} is the second population of extragalactic point sources, consisting of a "confused" continuum of unresolved point sources. Along a given line-of-sight, we imagine the number of sources to be Poisson distributed with an average of $n\Omega_{pix}$ sources, where n is the number of sources per steradian and Ω_{pix} is the pixel size. We model the spectrum of each source as a power law with a random spectral index α drawn from a Gaussian distribution

$$p(\alpha) = \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} \exp\left[-\frac{(\alpha-\alpha_0)^2}{2\sigma_\alpha^2}\right], \qquad (5)$$

where α_0 is the mean spectral index (with its numerical value to be determined later) and $\sigma_\alpha = 0.5$ (Tegmark et al. 2000). If all of the unresolved point sources had the same flux S_* at some fiducial frequency $\nu_* \equiv 150\,\text{MHz}$, the mean intensity would be

$$m^{ps}(\nu) = \left(\frac{A_\nu}{\Omega_{pix}}\right)(n\Omega_{pix}S_*)\int\left(\frac{\nu}{\nu_*}\right)^{-\alpha}p(\alpha)d\alpha, \qquad (6)$$

where the quantity A_ν/Ω_{pix} converts our expression from having flux units to temperature units, and is given by

$$\left(\frac{A_\nu}{\Omega_{pix}}\right) = \frac{\lambda^2}{2k_B\Omega_{pix}} \approx 1.4{\times}10^{-6}\left(\frac{\nu}{\nu_*}\right)^{-2}\left(\frac{\Omega_{pix}}{1\,\text{sr}}\right)^{-1}\text{mJy}^{-1}\,\text{K}, \qquad (7)$$

where λ is the wavelength, k_B is Boltzmann's constant, and Ω_{pix} is the pixel solid angle. Of course, in reality the sources do not all have the same flux. To take this into account, we extrapolate a source count distribution from an empirical study done at higher flux levels:

$$\frac{dn}{dS_*} = B\left(\frac{S_*}{880\,\text{mJy}}\right)^{-\gamma}, \qquad (8)$$

where following Di Matteo et al. (2002) we take $B = 4.0\,\text{mJy}^{-1}\text{Sr}^{-1}$ and $\gamma = 1.75$ as our fiducial values[4]. Integrating over the distribution, Equation 6 becomes

$$m^{ps}(\nu) = (17.4x_{max}^{2-\gamma}\,\text{K})\left(\frac{B}{4.0\,\text{mJy}^{-1}\text{Sr}^{-1}}\right)\times$$
$$\left(\frac{2-\gamma}{0.25}\right)^{-1}\left(\frac{\nu}{\nu_*}\right)^{-\alpha_{ps}+\frac{\sigma_\alpha^2}{2}\ln\left(\frac{\nu}{\nu_*}\right)}, \qquad (9)$$

where $x_{max} \equiv S_{max}/880\,\text{mJy}$, and $\alpha_{ps} \equiv \alpha_0 + 2$ takes on a fiducial value of 2.5 to match measurements from the Cosmic Microwave Background (Tegmark et al. 2000). Note also that this implies $\alpha_0 \approx 0.5$, which is consistent with both Toffolatti et al. (1998) and Jackson (2005).

---

[4] Despite the fact that the source count distribution was obtained by extrapolation, we expect that whatever the true distribution is, it should be well-approximated by a power law. This is because the key quantity is the integral of the source count, which is dominated by the brightest sources of the population. In that regime, we can linearize the distribution in log-log space, giving a power law in S_*.

To compute the covariance of the distribution, we once again begin by considering a population of point sources of brightness S_* whose number density is determined by Poisson statistics (thus giving a result proportional to nS_*^2). We then integrate over the source count distribution to get

$$C^{ps}(\nu,\nu') = \frac{A_\nu A_{\nu'}}{\Omega_{pix}}\int_0^{S_{max}}\frac{dn}{dS_*}S_*^2 dS_*\left(\frac{\nu\nu'}{\nu_*^2}\right)^{-\alpha_0+\frac{\sigma_\alpha^2}{2}\ln\left(\frac{\nu\nu'}{\nu_*^2}\right)}$$
$$= (4274x_{max}^{3-\gamma}\,\text{K}^2)\left(\frac{\Omega_{pix}}{10^{-6}\,\text{Sr}}\right)^{-1}\left(\frac{B}{4.0\,\text{mJy}^{-1}\text{Sr}^{-1}}\right)$$
$$\times\left(\frac{3-\gamma}{1.25}\right)^{-1}\left(\frac{\nu\nu'}{\nu_*^2}\right)^{-\alpha_{ps}+\frac{\sigma_\alpha^2}{2}\ln\left(\frac{\nu\nu'}{\nu_*^2}\right)}. \qquad (10)$$

Note that no further subtraction of the mean term is necessary, since it was implicitly accomplished when we invoked Poisson statistics.

2.3 Galactic Synchrotron Radiation

For Galactic synchrotron radiation, we imagine the foreground spectrum in each pixel to be well fit by a power law with spectral index α, but that the value of the spectral index may vary from pixel to pixel. In a given pixel, the spectrum is thus

$$x(\nu) = A_{sync}\left(\frac{\nu}{\nu_*}\right)^{-\alpha}, \qquad (11)$$

where $A_{sync} = 335.4\,\text{K}$ (Wang et al. 2006). Similar to the point sources, we assume that the indices in different pixels to be Gaussian distributed, only this time with a mean of $\alpha_{sync} = 2.8$ and a standard deviation of $\Delta\alpha_{sync} = 0.1$ (Wang et al. 2006). Performing the same integral as for the point sources, we obtain

$$m^{sync}(\nu) = A_{sync}\left(\frac{\nu}{\nu_*}\right)^{-\alpha_{sync}+\frac{\Delta\alpha_{sync}^2}{2}\ln\left(\frac{\nu}{\nu_*}\right)}. \qquad (12)$$

Forming the foreground covariance using Equation 3, we obtain

$$C^{sync}(\nu,\nu') = A_{sync}^2\left(\frac{\nu\nu'}{\nu_*^2}\right)^{-\alpha_{sync}+\frac{\Delta\alpha_{sync}^2}{2}\ln\left(\frac{\nu\nu'}{\nu_*^2}\right)}$$
$$-m^{sync}(\nu)m^{sync}(\nu'). \qquad (13)$$

2.4 Free-free Emission

Free-free emission can be modeled in much the same way as the synchrotron radiation:

$$m^{ff}(\nu) = A_{ff}\left(\frac{\nu}{\nu_*}\right)^{-\alpha_{ff}+\frac{\Delta\alpha_{ff}^2}{2}\ln\left(\frac{\nu}{\nu_*}\right)} \qquad (14)$$

$$C^{ff}(\nu,\nu') = A_{ff}^2\left(\frac{\nu\nu'}{\nu_*^2}\right)^{-\alpha_{ff}+\frac{\Delta\alpha_{ff}^2}{2}\ln\left(\frac{\nu\nu'}{\nu_*^2}\right)}$$
$$-m^{ff}(\nu)m^{ff}(\nu'). \qquad (15)$$

but with $A_{ff} = 33.5\,\text{K}$, $\alpha_{ff} = 2.15$, and $\Delta\alpha_{ff} = 0.01$ (Wang et al. 2006).

Table 1. Free parameters in our foreground model and their fiducial values.

Parameter	Description	Fiducial Value
B	Source count normalization	$4.0\,\mathrm{mJy}^{-1}\,\mathrm{Sr}^{-1}$
γ	Source count power-law index	1.75
α_{ps}	Point source spectral index	2.5
σ_α	Point source index spread	0.5
A_{sync}	Synchrotron amplitude	335.4 K
α_{sync}	Synchrotron spectral index	2.8
$\Delta\alpha_{sync}$	Synchrotron index coherence	0.1
A_{ff}	Free-free amplitude	33.5 K
α_{ff}	Free-free spectral index	2.15
$\Delta\alpha_{ff}$	Free-free index coherence	0.01

2.5 Total Foreground Contribution

To obtain total contribution to the mean signal, we simply sum the means of the various components:

$$m(\nu) = m^{ps}(\nu) + m^{sync}(\nu) + m^{ff}(\nu). \qquad (16)$$

For the total covariance we sum the individual covariances:

$$C(\nu,\nu') = C^{ps}(\nu,\nu') + C^{sync}(\nu,\nu') + C^{ff}(\nu,\nu'). \qquad (17)$$

In total, then, our foreground model consists of 10 free parameters, which are listed in Table 1 along with their fiducial values.

2.6 Noise model

In general, the noise level in a given pixel of a sky map produced by a radio telescope/interferometer scales as

$$\sigma_{noise} \propto \frac{\lambda^2 T_{sys}}{A_e \sqrt{\Delta\nu\Delta t}}, \qquad (18)$$

where $\Delta\nu$ is the channel width of a frequency bin, Δt is the integration time, A_e is the collecting area of an antenna element, T_{sys} is the system temperature, and λ is the wavelength. To keep the discussion in this paper as general possible, we do not explicitly model the effective area and the system temperature, since both depend on one's specific instrument in complicated ways. Instead, we take a minimalistic approach and simply anchor our noise to levels that are expected for current generation 21 cm tomography experiments. From Bowman et al. (2009), we know that with 360 hours of integration at a channel width of 40 KHz, the MWA has a single pixel noise level at 158 MHz that is approximately 330 mK. Scaling this to our fiducial values, we obtain

$$\sigma_{noise} \sim (39\,\mathrm{mK}) \left(\frac{1\,\mathrm{MHz}}{\Delta\nu}\right)^{\frac{1}{2}} \left(\frac{1000\,\mathrm{hrs}}{\Delta t}\right)^{\frac{1}{2}}. \qquad (19)$$

In the numerical studies conducted in this paper, we will always be keeping Δt fixed at 1000 hrs. The frequency bin width $\Delta\nu$ will also be fixed at its fiducial value of 1 MHz, except when explicitly stated (for instance when we investigate the dependence of our results on $\Delta\nu$).

In the next section, we will be working almost exclusively in dimensionless units where the foreground power

(quantified by the diagonal $\nu = \nu'$ elements of the foreground covariance) is equal to unity across all frequencies (see Equation 22). In such units, the noise can be taken to be approximately white *i.e.* frequency independent. To see this, note that the effective area A_e of an antenna scales as λ^2, which means that the quantity λ^2/A_e has no frequency dependence, and σ_{noise} depends on frequency only because of T_{sys}. Typical 21 cm experiments are sky noise dominated (Morales & Wyithe 2010), which means that T_{sys} should be dominated by the foreground temperature, and therefore should have roughly the same frequency dependence as the foregrounds do. However, this frequency dependence was precisely the dependence that our choice of units was designed to null out. The noise thus becomes white to a good approximation, and is effectively a noise-to-signal ratio (henceforth denoted by κ). Note that while we adopt this approximation for the rest of the paper, it is by no means crucial, and in general one should always use units where the foreground covariance is white, even if in such units the noise is chromatic. For further discussion of this point, please see Section 3.2.

3 THE EASE OF CHARACTERIZING FOREGROUNDS

While the foreground parameters described in Table 1 are certainly conventional choices, they are not the most economical, in the sense that many of them are redundant. That this is the case is not surprising, since so many different foreground sources can be accurately described by spectra that deviate only slightly from power laws. In Section 3.1 we will reparametrize our foreground model using a principal component analysis. In Section 3.4 we will determine the effective number of principal components that need to be measured to accurately describe the foreground spectra, and find the number to typically be three or four.

3.1 Eigenforeground Modes

Since a measured foreground spectrum will necessarily be discrete, we begin by discretizing the mean and the covariance of our foreground model from Section 2, so that

$$\mathbf{m}_\alpha \equiv m(\nu_\alpha) \qquad (20)$$

and

$$\mathbf{C}_{\alpha\beta} \equiv C(\nu_\alpha, \nu_\beta), \qquad (21)$$

where the indices run from 1 to N_c, the total number of frequency channels in one's instrument. We define a correlation matrix

$$\mathbf{R}_{\alpha\beta} \equiv \frac{\mathbf{C}_{\alpha\beta}}{\sqrt{\mathbf{C}_{\alpha\alpha}\mathbf{C}_{\beta\beta}}}, \qquad (22)$$

and work with it instead the covariance matrix.

We now perform a principal component analysis on the foregrounds[5]. That is, we rewrite the correlation matrix in

---

[5] We perform the principal component analysis using the correlation matrix **R** instead of the covariance matrix **C** because the challenge with foreground modeling is to successfully describe the

a basis of "eigenforeground" vectors[6], where each eigenforeground vector $\mathbf{v}_n$ satisfies the eigenvalue equation

$$\mathbf{R}\mathbf{v}_n = \lambda_i \mathbf{v}_n. \tag{23}$$

Normalizing the eigenforeground vectors to unity and forming a matrix $\mathbf{V}$ where the columns of the matrix are the normalized eigenvectors, the correlation matrix can be expressed as

$$\mathbf{R} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathbf{t}} = \sum_{n=1}^{N_c} \lambda_n \mathbf{v}_n \mathbf{v}_n^t, \tag{24}$$

where $\mathbf{\Lambda} \equiv \mathrm{diag}\{\lambda_1, \lambda_2, \lambda_3, \dots\}$. Note that since $\mathbf{R}$ is real and symmetric, $\mathbf{V}^{\mathbf{t}} = \mathbf{V}^{-1}$, so $\mathbf{V}\mathbf{V}^{\mathbf{t}} = \mathbf{V}^{\mathbf{t}}\mathbf{V} = \mathbf{I}$. In the last equality of Equation 24, we see that each eigenvalue λ_n measures the contribution of its corresponding eigenforeground $\mathbf{v}_n$ to the total foreground variance.

In Section 3.4, we will seek to describe measured foreground spectra in terms of eigenforeground components. In other words, we wish to find the weight vector $\mathbf{a}$ (with component a_k for the kth eigenforeground) in the equation

$$\mathbf{y} \equiv \mathbf{x} + \mathbf{n} = \sum_{k=1}^{N_c} a_k \mathbf{v}_k + \mathbf{n} = \mathbf{V}\mathbf{a} + \mathbf{n}, \tag{25}$$

where $\mathbf{y}$, $\mathbf{x}$, and $\mathbf{n}$ are all vectors of length equal to the number of channels N_c, containing the measured spectrum of foregrounds, the true foregrounds, and the noise respectively. The vector $\mathbf{x}$, for instance, is simply a discretized and whitened version of $x(\nu)$ in Equation 1. Once an estimate $\hat{\mathbf{a}}$ of $\mathbf{a}$ has been determined, one can multiply by $\mathbf{V}$ to obtain an estimator $\hat{\mathbf{x}}$ of the foreground spectrum. We will find that by measuring a small number of parameters, we can characterize foreground spectra to a very high precision, thanks to the special properties of the eigenforegrounds, which we describe in Section 3.2.

3.2 Features of the Eigenforegrounds

Since the foreground spectra do not contain any sharp features[7], we expect their rather featureless frequency dependences to be well-described by a small number of eigenforegrounds, *i.e.* we expect the eigenvalues λ_n to be large only for the first few values of n. Performing the analysis using the foreground model of Section 2 for a noiseless experiment with 50 equally spaced frequency channels going from 100 MHz to 200 MHz, we see in Figure 1 that the eigenvalues do fall off rapidly with mode number; indeed, the fall-off is exponential (a fact that we will explain later in this section),

fine relative perturbations about the smooth predictable power law spectrum. Since the relative fluctuations are quantified by $\mathbf{R}$, not $\mathbf{C}$, we should correspondingly use $\mathbf{R}$ for our principal component analysis.

[6] Throughout this paper, we will use Greek indices to denote different frequencies and lowercase Latin indices to denote different foreground components/modes. In Section 4 we will use uppercase Latin indices to denote the different foreground parameters listed in Table 1.

[7] Following Morales et al. (2006); Gleser et al. (2008), we assume that narrowband contaminants such as terrestrial radio stations and radio recombination lines have been excised from the data prior to this point in the analysis.
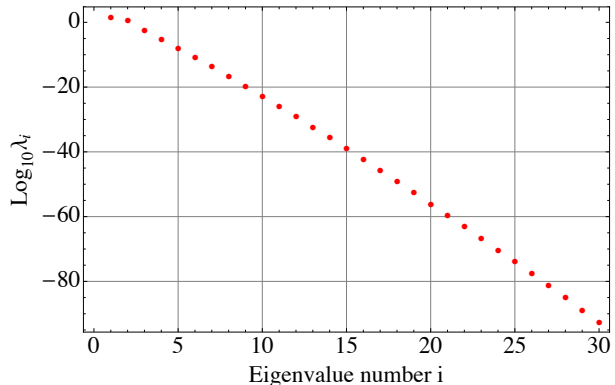

Figure 1. Eigenvalues of $\mathbf{R}$ with no noise for an experiment with 50 frequency channels, equally spaced from 100 MHz to 200 MHz.
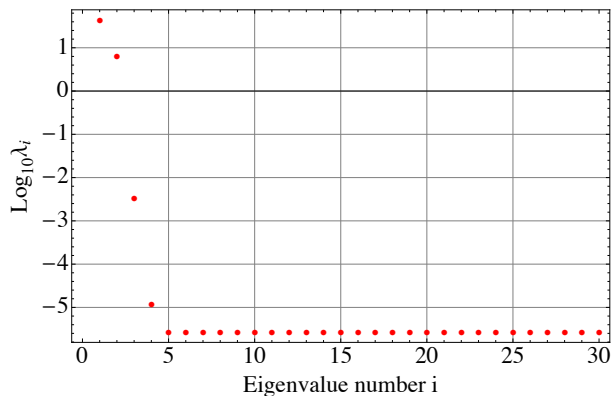

Figure 2. Eigenvalues of $\mathbf{R}$ for an experiment with 50 frequency channels, equally spaced from 100 MHz to 200 MHz, and noise levels given by the fiducial model of Section 2.6.

and in going from the first eigenvalue to the third there is a drop of more than three orders of magnitude. This means that if we select as our foreground parameters the expansion coefficients a_k of Equation 25, we can account for almost all of the foreground signal by measuring just a few numbers. In fact, in a realistic experiment it is impossible to measure more than the first few eigenvalues, as the foreground signal quickly becomes subdominant to the instrumental noise. This can be seen in Figure 2, where we once again show the eigenvalues, but this time including the noise model of Section 2.6. After the fourth eigenforeground, we see that the eigenvalues hit a noise floor because by then the eigenmodes are essentially measuring the noise. Note also that the large drop in magnitude of successive eigenvalues makes the qualitative results of this paper robust to our assumption that the noise is white in our non-dimensionalized units — because the various eigenmodes contribute to the total foreground at such different levels, a slight chromaticity in the noise will not change the mode number at which the eigenmodes hit the noise floor in Figure 2. Violating our assumption will thus have very little effect on our results.

It should also be pointed out that in an analysis where the cosmological signal is included and is larger than the instrumental noise (as would be the case for a long integration time experiment), one would expect to hit a "signal floor" rather than a noise floor. The break in the eigenvalue spec-
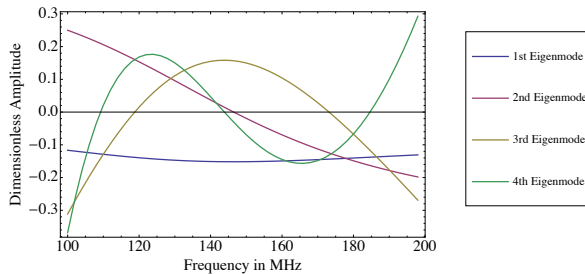
Figure 3. First few eigenvectors of $\mathbf{R}$ ("eigenforegrounds") for an experiment spanning a frequency range from $100\,\mathrm{MHz}$ to $200\,\mathrm{MHz}$.

Figure 4. First few eigenvectors of $\mathbf{R}$ ("eigenforegrounds") for an experiment spanning a frequency range from $100\,\mathrm{MHz}$ to $200\,\mathrm{MHz}$, but with the $\mathbf{C}_{\alpha\alpha}$ normalization factor restored so that the eigenforegrounds have units of temperature.

trum in going from the exponential decay of the foreground dominated region to a flatter signal region can potentially be used as a diagnostic for separating foregrounds from signal. To properly investigate the robustness of this method, however, requires detailed simulations of the signal. Since the focus of this paper is foreground modeling (and not signal extraction), we defer such an investigation to future work.

In Figure 3, we show the first few foreground eigenvectors, and in Figure 4 we restore the frequency-dependent normalization factors that were divided out in Equations 22. In both cases the eigenvectors are the ones defined by Equation 23 *i.e.* for the noiseless case, and henceforth we will only be using these noiseless eigenvectors. Even though we will include noise in our subsequent analysis, the use of the noiseless eigenvectors represents no loss of generality because the eigenvectors simply form a convenient set of basis vectors that span the space. In addition, with our assumption of white noise (in the units defined by Equation 22), the inclusion of noise alters only the eigenvalues, not the eigenvectors.

Several eigenforeground features are immediately apparent from the plots. First, the nth eigenforeground is seen to have $n-1$ nodes. This is due to the fact that correlation matrix $\mathbf{R}$ is Hermitian (since it is real and symmetric), so Equation 23 takes the mathematical form of a time-independent Schrödinger equation for a one-dimensional system, and the node theorem of quantum mechanics applies. As a consequence, each successive foreground eigenmode probes a more rapidly fluctuating component of the spectrum, which explains the rapid fall in eigenvalues seen in Figure 1 — since foregrounds are such smooth functions of frequency, the more rapidly oscillating eigenmodes are simply not required.

Another characteristic feature of Figure 3 is that the nth eigenforeground appears to look like a polynomial of order $(n-1)$, albeit with some slight deviations (the "linear" 1st eigenmode, for instance, has a small curvature to it). This approximate polynomial behavior explains the success of line-of-sight foreground subtraction schemes (Wang et al. 2006; Bowman et al. 2009; Liu et al. 2009a,b) that subtract low-order polynomials from foreground spectra — by subtracting out low-order polynomials, one is subtracting the modes with the largest eigenvalues, and since the eigenvalues fall so quickly (indeed, exponentially) with mode number, the result is that most of the foregrounds are cleaned out by the process. Together, Figures 3 and 4 also explain why it was found in Liu et al. (2009b) that polynomial fore-
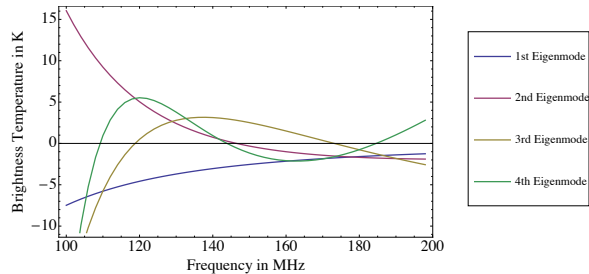
ground cleaning performs well only over narrow ($\sim 1\,\mathrm{MHz}$) frequency ranges. In Figure 4, it is seen that over large frequency ranges the eigenforegrounds do not behave like polynomials, and that the polynomial behavior is only evident after dividing out a power-law-like normalization. In other words, the eigenforegrounds in Figure 4 still have a rough power law dependence to their spectrum, and thus are not well-fit by polynomials. Only if the frequency range is narrow will the power law dependence have a negligible effect, and so only in such a scenario would polynomial subtraction do well. Fortunately, there is a simple solution to this problem — if one wishes to perform polynomial subtraction over a wide frequency range, one can still do so successfully by first dividing out a fiducial spectrum (in the same way as we did in writing down Equation 22) before performing the fits[8]. Put another way, we simply have to perform the fits in Figure 3 (where the modes *always* look like polynomials) rather than in Figure 4.

3.3 Understanding the Eigenforegrounds

Despite the fact that it can be useful (in our whitened units) to think of the eigenforegrounds as polynomials of successively higher order, we caution that this interpretation is not exact. For instance, the first eigenmode in Figure 3 is clearly not precisely constant, and contains a small quadratic correction. Indeed, we will now show that it is more useful to think of the whitened eigenmodes as functions that would be sinusoids in ν were it not for an instrument's finite frequency range and spectral resolution.

Suppose that unresolved point sources were the sole contributor to our measured foregrounds. Working in a continuous description (to avoid finite bandwidth and resolution), the analog of Equation 22 takes the form[9]

$$R(\nu, \nu') = \frac{C(\nu, \nu')}{\sigma(\nu)\sigma(\nu')}, \qquad (26)$$

---

[8] Alternatively, one can fit foregrounds over a large frequency range by using polynomials in *logarithmic* frequency space. As noted in Bowman et al. (2009), however, one may want to avoid doing this because of the interferometric nature of most 21 cm experiments — since interferometers are not sensitive to the mean emission, negative values will be measured in certain pixels, making it problematic to take the logarithm.

[9] Aside from the omission of the noise term here, what follows is precisely the toy model introduced in Liu & Tegmark (2011).

where $C(\nu, \nu')$ is given by Equation 10 and

$$\sigma(\nu) \equiv [C(\nu, \nu)]^{\frac{1}{2}} = \left(\frac{\nu}{\nu_*}\right)^{-\alpha_{ps} + \sigma_\alpha^2 \ln(\nu/\nu_*)} . \qquad (27)$$

The resulting correlation function $R(\nu, \nu')$ is given by

$$
\begin{aligned}
R(\nu, \nu') &= \exp\left[-\frac{\sigma_\alpha^2}{2}(\ln \nu - \ln \nu')^2\right] \\
&\approx \exp\left[-\frac{(\nu - \nu')^2}{2\nu_c^2}\right],
\end{aligned} \qquad (28)
$$

where we have Taylor expanded about ν_* to obtain an unnormalized Gaussian with a *coherence length* of $\nu_c \equiv \nu_*/\sigma_\alpha \ln \nu_*$. Despite the fact that this expression was derived by only considering point sources, we find that it is an excellent approximation even when the full foreground model is used, provided one uses a longer coherence length to reflect the smoother Galactic synchrotron and free-free components ($\sim 64.8\,\mathrm{MHz}$ gives a good fit to the foreground model of Section 2).

Since $R(\nu, \nu')$ now just depends on the difference $\nu - \nu'$, the continuous analog of Equation 23 takes the form of a convolution:

$$\int R(\nu - \nu') f_n(\nu') d\nu' = \lambda_n f_n(\nu) \qquad (29)$$

where f_n is the nth eigenforeground spectrum. Because convolution kernels act multiplicatively in Fourier space, the f_i's are a family of sinusoids. Indeed, plugging the ansatz $f_n(\nu) = \sin(\gamma_n \nu + \phi)$ into Equation 29 yields

$$\int_{-\infty}^{\infty} \exp\left[-\frac{(\nu - \nu')^2}{2\nu_c}\right] \sin(\gamma_n \nu' + \phi) d\nu' = \lambda_n \sin(\gamma_n \nu + \phi), \qquad (30)$$

where the eigenvalue is given by

$$\lambda_n = \sqrt{2\pi\nu_c^2} \exp\left(-2\nu_c^2 \gamma_n^2\right). \qquad (31)$$

The eigenforegrounds shown in Figure 3 should therefore be thought of as a series of sinusoid-like functions that deviate from perfect sinusoidal behavior only because of "edge effects" arising from the finite frequency range of an experiment. As expected from the exponential form of Equation 31, the modes quickly decrease in importance with increasing wavenumber γ_n, and the more coherent the foregrounds (the larger ν_c is), the fewer eigenfunctions are needed to describe the foreground spectra to high accuracy.

One discrepancy between the eigenvalue behavior in our analytic treatment and the numerical results earlier on in the section is its dependence on wavenumber γ_n. Analytically, Equation 31 predicts that γ_n should appear quadratically in the exponent, whereas numerically Figure 1 suggests something closer to a linear relationship (provided one imagines that the mode number is roughly proportional to γ_n, which should be a good approximation at high mode numbers). We find from numerical experimentation that part of this discrepancy is due to the finite frequency range edge effects, and that as one goes from a very small range to a very large range, the dependence on γ_n steepens from being linear in the exponential to being a power law in the exponential, with the index of the power law never exceeding 2. In any case, the role of Equation 31 is simply to provide an intuitive understanding of the origin of the exponential fall drop in eigenvalues, and in the frequency ranges applicable to 21 cm

tomography, we have verified numerically that a linear exponential fall-off fits the foreground model well. In particular, we can parametrize the eigenvalues with two parameters A and B, such that

$$\lambda_n \approx B \exp\left(-An\right), \qquad (32)$$

and find that $A = 6.44$ and $B = 5.98 \times 10^5$ fits the foreground model well for an instrument with a frequency range of 100 to 200 MHz and frequency bins of $\Delta\nu = 2\,\mathrm{MHz}$. We caution, however, that these parameters vary with the frequency range, frequency bin width, and foreground model, and in general one must fit for A and B for each specific set of parameters. This is what we do to generate the numerical results of the following subsection. Generically, though, A will always be somewhat larger than unity, so the foreground spectra will always be dominated by the first few eigenmodes.

3.4 Eigenforeground Measurements

So far, we have established that foreground spectra should be describable to a very high accuracy using only a small number of principal foreground components, with the importance of the kth foreground component quantified by the a_k coefficient in $\mathbf{a}$, which was defined in Equation 25. In a practical measurement, however, the presence of noise means that one does not know the true value of $\mathbf{a}$, but instead must work with an estimator $\hat{\mathbf{a}}$ that is formed from the data. There are many different ways to form this estimator, and one possibility would be to minimize the quantity

$$\chi^2 \equiv (\mathbf{y} - \mathbf{V}\hat{\mathbf{a}})^{\mathbf{t}} \mathbf{N}^{-1} (\mathbf{y} - \mathbf{V}\hat{\mathbf{a}}), \qquad (33)$$

where $\mathbf{N}$ is defined as $\langle \mathbf{n}\mathbf{n}^{\mathbf{t}} \rangle$ using the noise vector $\mathbf{n}$ of Equation 25. This least-squares minimization (which is optimal if the noise is Gaussian) yields

$$\hat{\mathbf{a}}^{LS} = [\mathbf{V}^{\mathbf{t}}\mathbf{N}^{-1}\mathbf{V}]^{-1} \mathbf{V}^{\mathbf{t}}\mathbf{N}^{-1}\mathbf{y} = \mathbf{V}^{-1}\mathbf{y} = \mathbf{V}^{\mathbf{t}}\mathbf{y}, \qquad (34)$$

since the number of eigenmodes (*i.e.* the length of $\hat{\mathbf{a}}^{LS}$) is equal to the number of frequency channels (*i.e.* the length of $\mathbf{y}$), so that $\mathbf{V}$ is a square orthogonal matrix and $\mathbf{V}^t = \mathbf{V}^{-1}$. Equation 34 states that even in the presence of noise, the least squares prescription calls for us to follow the same procedure as we would if there were no noise, namely, to take the dot product of both sides of Equation 25 with each eigenforeground vector.

Defining an error vector $\varepsilon \equiv \delta\mathbf{a} = \hat{\mathbf{a}} - \mathbf{a}$ as the difference between the true $\mathbf{a}$ and the estimator $\hat{\mathbf{a}}$, we could instead choose to minimize the diagonal quantities $\langle |\varepsilon_i|^2 \rangle$. This corresponds to so-called Wiener filtering (Tegmark 1997), and the estimator is given by

$$\hat{\mathbf{a}}^{Wiener} = \mathbf{S}\mathbf{V}^{\mathbf{t}} \left[\mathbf{V}\mathbf{S}\mathbf{V}^{\mathbf{t}} + \mathbf{N}\right]^{-1} \mathbf{y}, \qquad (35)$$

where $\mathbf{S} \equiv \langle \mathbf{a}\mathbf{a}^{\mathbf{t}} \rangle$ is the signal covariance matrix for the vector $\mathbf{a}$.

In Sections 3.4.1 and 3.4.2, we examine Wiener filtering and least-squares minimization. We will find that the estimated foreground spectrum $\hat{\mathbf{x}} = \mathbf{V}\hat{\mathbf{a}}$ from Wiener filtering is expected to be closer to the true foreground spectrum, and that the least-squares method can be adapted to give similar results, although with errors larger than for Wiener filtering.

3.4.1   Wiener Filtering

With the Wiener filtering technique, our estimate of the foreground spectrum is given by

$$\hat{\mathbf{x}}^{Wiener} = \mathbf{V}\hat{\mathbf{a}}^{Wiener} = \mathbf{V}\mathbf{S}\mathbf{V}^{\mathbf{t}}\left[\mathbf{V}\mathbf{S}\mathbf{V}^{\mathbf{t}} + \mathbf{N}\right]^{-1}\mathbf{y} \equiv \mathbf{W}\mathbf{y} \tag{36}$$

Understanding Wiener filtering thus comes down to understanding the filter $\mathbf{W}$. We first rewrite the expression slightly:

$$\mathbf{W} \equiv \mathbf{V}\mathbf{S}\mathbf{V}^{\mathbf{t}}\left[\mathbf{V}\mathbf{S}\mathbf{V}^{\mathbf{t}} + \mathbf{N}\right]^{-1} = \mathbf{V}\mathbf{S}\left[\mathbf{S} + \mathbf{V}^{\mathbf{t}}\mathbf{N}\mathbf{V}\right]^{-1}\mathbf{V}^{\mathbf{t}}, \tag{37}$$

where we have made liberal use of the fact that $\mathbf{V}^{\mathbf{t}} = \mathbf{V}^{-1}$. From Section 2.6, we know that $\mathbf{N}$ is proportional to the identity if we use whitened units. In the notation of that section, we have $\mathbf{N} = \kappa^2\mathbf{I}$, so

$$\mathbf{W} = \mathbf{V}\mathbf{S}\left[\mathbf{S} + \kappa^2\mathbf{I}\right]^{-1}\mathbf{V}^{\mathbf{t}}. \tag{38}$$

Now consider $\mathbf{S}$:

$$\begin{aligned}\mathbf{S} &= \mathbf{V}^{\mathbf{t}}\mathbf{V}\mathbf{S}\mathbf{V}^{\mathbf{t}}\mathbf{V} = \mathbf{V}^{\mathbf{t}}\langle(\mathbf{V}\mathbf{a})(\mathbf{V}\mathbf{a})^{\mathbf{t}}\rangle\mathbf{V} \\ &= \mathbf{V}^{\mathbf{t}}\langle\mathbf{x}\mathbf{x}^{\mathbf{t}}\rangle\mathbf{V} = \mathbf{V}^{\mathbf{t}}\mathbf{R}\mathbf{V} = \mathbf{\Lambda},\end{aligned} \tag{39}$$

where $\mathbf{x}$ denotes the true whitened foreground spectrum, as it did in Equation 25. In the penultimate step we used the fact that $\langle\mathbf{x}\mathbf{x}^{\mathbf{t}}\rangle$ is precisely our whitened foreground covariance, *i.e.* $\mathbf{R}$, and in the last step we used Equation 24. Inserting this result into our Wiener filter gives

$$\mathbf{W}_{ij} = \left(\mathbf{V}\mathbf{\Lambda}\left[\mathbf{\Lambda} + \kappa^2\mathbf{I}\right]^{-1}\mathbf{V}^{\mathbf{t}}\right)_{ij} = \sum_{l,m}^{N_c}\mathbf{V}_{il}\left(\frac{\lambda_l}{\lambda_l + \kappa^2}\right)\delta_{lm}\mathbf{V}^{\mathbf{t}}{}_{mj}, \tag{40}$$

where as before N_c is the number of frequency channels (which is equal to the number of eigenforegrounds). This in turn means that our estimator takes the form

$$\begin{aligned}\hat{\mathbf{x}}_i^{Wiener} &= \sum_j^{N_c}\mathbf{V}_{ij}\left(\frac{\lambda_j}{\lambda_j + \kappa^2}\right)(\mathbf{V}^{\mathbf{t}}\mathbf{y})_j \\ &= \sum_j^{N_c}\mathbf{V}_{ij}w_j\hat{\mathbf{a}}_j^{LS} = \sum_j^{N_c}w_j\hat{\mathbf{a}}_j^{LS}\mathbf{v}_j(\nu_i), \end{aligned} \tag{41}$$

where $\mathbf{v}_j$ is the jth eigenforeground as defined in Equation 23 and $w_j \equiv \lambda_j/(\lambda_j + \kappa^2)$ can be thought of as "Wiener weights" for the jth eigenforeground. Since $w_j \approx 1$ for $\lambda_j \gg \kappa^2$ and $w_j \approx 0$ for $\lambda_j \ll \kappa^2$, this weighting factor preserves eigenmodes that have high signal to noise while suppressing those that have low signal to noise. In words, the Wiener filtering procedure thus amounts to first performing a least-squares fit to obtain estimates for the eigenforeground coefficients, but then in the reconstruction of the foreground spectrum, downweighting modes that were only measured with low signal-to-noise.

In Figure 5 we show the Wiener weights for the first ten eigenmodes for an experiment with 50 frequency channels, equally spaced from $100\,\text{MHz}$ to $200\,\text{MHz}$, and noise levels given by the fiducial model of Section 2.6. The weights are seen to be small after the first few eigenmodes, suggesting that most of the information in our foreground spectrum estimate comes from a mere handful of parameters. The other parameters (*i.e.* eigenforeground coefficients) are too noisy to have much constraining power. Since the Wiener weights tend to unity in the limit of large signal-to-noise, we can
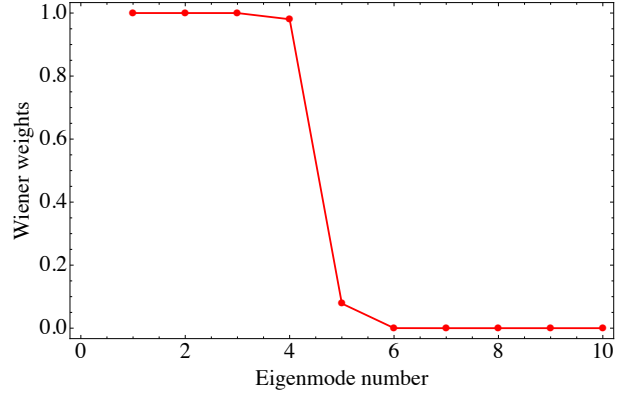

Figure 5. First few Wiener weights (defined as $w = \lambda/(\lambda + \kappa^2)$, where λ is the eigenvalue of a foreground mode and κ is the whitened noise) for an experiment with 50 frequency channels, equally spaced from $100\,\text{MHz}$ to $200\,\text{MHz}$, and noise levels given by the fiducial model of Section 2.6.

define an effective number of measurable parameters n_{eff} by summing the Wiener weights:

$$n_{\text{eff}} \equiv \sum_i^{N_c} w_i. \tag{42}$$

For the fiducial scenario in Figure 5, the numerical value of n_{eff} is 4.06.

In general, n_{eff} depends on both the nature of the foregrounds and the noise level. If the noise levels were high, the difficulty in measuring the higher (and therefore subdominant) eigenmodes would result in those measurements being noise dominated. The Wiener filter would thus suppress such modes, driving n_{eff} down. In the opposite limit where the noise levels are low, one would expect n_{eff} to be higher, but still to be relatively small. This is because we know from Section 3.2 that the foreground spectra are relatively simple functions that are dominated by the first few eigenmodes, and thus even if the noise levels were low enough for the higher modes to be measurable, it is unnecessary to give them very much weight. In general, n_{eff} varies roughly logarithmically with increasing foreground-to-noise ratio since the foreground eigenvalues drop exponentially. This nicely explains one of the results from Liu & Tegmark (2011). There it was found that in performing foreground subtraction by deweighting foreground contaminated line-of-sight Fourier modes, the number of modes that needed to be deweighted increased only logarithmically with the foreground-to-noise ratio. We now see that this is simply a consequence of Equation 32.

In Figure 6, the red/grey dotted line shows n_{eff} as a function of the total frequency range of an experiment. In changing the frequency range, we keep the integration time and the channel width $\Delta\nu$ constant so that the noise level remains unchanged. Any change in n_{eff} therefore reflects the nature of the foregrounds, and what one sees in Figure 6 is that as the total frequency range increases, more and more modes are needed when estimating the spectrum. This is because effects like the spread in the spectral indices are apparent only over large frequency ranges, so increasing the frequency range makes foregrounds more complicated to model.
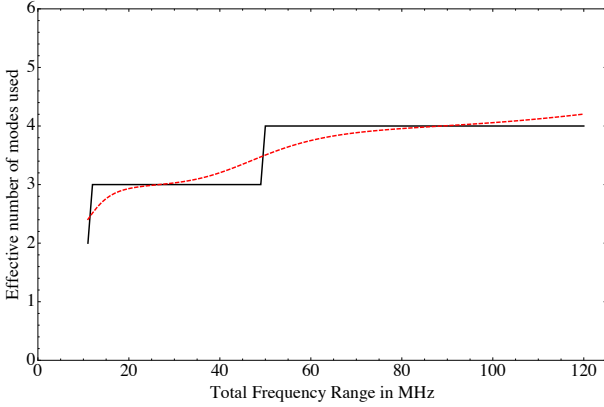
Figure 6. Shown with the red/grey dashed curve, the effective number $n_{\rm eff}$ of foreground parameters used (defined by Equation 42) in the Wiener filtering method. In solid black is the optimal m (Equation 48 adjusted for the fact that m must be an integer) for the truncated least-squares method. In both cases the behavior is shown as a function of the total frequency range of an instrument, with channel width and integration time (and thus noise level) held constant at $1\,{\rm MHz}$ and $1000\,{\rm hrs}$, respectively.

To quantify the success of a foreground model generated from Wiener filtered measurements, we consider the quantity $\delta\hat{\mathbf{x}}^w \equiv \hat{\mathbf{x}}^{\rm Wiener} - \mathbf{x}$:

$$
\begin{aligned}
\delta\hat{\mathbf{x}}_i^w &= \sum_j^{N_c} \mathbf{V}_{ij}(w_j\hat{\mathbf{a}}_j^{\mathbf{LS}} - \mathbf{a}_j) \\
&= \sum_j^{N_c} \mathbf{V}_{ij}(w_j - 1)\mathbf{a}_j - \sum_{j,k}^{N_c} \mathbf{V}_{ij}w_j\mathbf{V}_{kj}\mathbf{n}_k \\
&= \sum_j^{N_c}(w_j - 1)\mathbf{a}_j\mathbf{v}_j - \sum_j^{N_c}\mathbf{W}_{ij}\mathbf{n}_j
\end{aligned}
\tag{43}
$$

where in the penultimate step we substituted $\hat{\mathbf{a}}^{LS} = \mathbf{a} + \mathbf{V^t n}$ (which follows from Equations 25 and 34), and in the ultimate step recognized that $\sum_j \mathbf{V}_{ij}w_j\mathbf{V}_{kj}$ is the Wiener filter $\mathbf{W}$ of Equation 40. The covariance of this quantity is

$$
\begin{aligned}
\langle\delta\hat{\mathbf{x}}^w\delta\hat{\mathbf{x}}_w^t\rangle &= \sum_{i,j}^{N_c}\langle\mathbf{a}_i\mathbf{a}_j\rangle(w_i - 1)(w_j - 1)\mathbf{v}_i\mathbf{v}_j^t + \mathbf{W}\langle\mathbf{nn^t}\rangle\mathbf{W^t} \\
&= \sum_i^{N_c}\lambda_i(w_i - 1)^2\mathbf{v}_i\mathbf{v}_i^t + \mathbf{WNW^t} \\
&= \sum_i^{N_c}\lambda_i(w_i - 1)^2\mathbf{v}_i\mathbf{v}_i^t + \kappa^2\sum_i^{N_c}w_i^2\mathbf{v}_i\mathbf{v}_i^t
\end{aligned}
\tag{44}
$$

where in the penultimate step we used $\langle\mathbf{a}_i\mathbf{a}_j\rangle = \mathbf{S}_{ij} = \mathbf{\Lambda}_{ij} = \lambda_i\delta_{ij}$, and in the last step used Equation 40 as well as $\mathbf{N} = \kappa^2\mathbf{I}$. Note that there are no cross terms in going from Equation 43 to 44 because such terms would be proportional to $\langle\sum_i a_i\mathbf{v}_i\mathbf{n^t}\rangle = \langle\mathbf{xn^t}\rangle$, which is assumed to be zero because the foregrounds and the noise are uncorrelated. The diagonal terms in Equation 44 give the expected mean-square measurement error at a particular frequency. Averaging over frequency to obtain a figure-of-merit for the entire

spectrum, our mean-square measurement error becomes

$$
\begin{aligned}
\varepsilon_{\rm Wiener}^2 &\equiv \overline{\langle\delta\hat{\mathbf{x}}^w\delta\hat{\mathbf{x}}_w^t\rangle} \\
&= \frac{1}{N_c}\sum_\alpha^{N_c}\sum_i^{N_c}\lambda_i(w_i - 1)^2\mathbf{v}_i^2(\nu_\alpha) \\
&\quad + \frac{\kappa^2}{N_c}\sum_\alpha^{N_c}\sum_i^{N_c}w_i^2\mathbf{v}_i^2(\nu_\alpha) \\
&= \frac{1}{N_c}\sum_i^{N_c}\lambda_i(w_i - 1)^2 + \frac{\kappa^2}{N_c}\sum_i^{N_c}w_i^2,
\end{aligned}
\tag{45}
$$

where in the last step we permuted the sums and used the fact that the eigenforeground vectors are normalized. From this expression, we can see that there are essentially two sources of error, each represented by one of the terms in Equation 45. The first term is only significant for the higher order eigenmodes, where $w_i \approx 0$. With each element in the sum proportional to λ_i, this term quantifies the error incurred by heavily de-weighting the higher order modes (which are de-weighted so much that they are essentially omitted from the estimator). Of course, we have no choice but to omit these modes, because their measurements are so noisy that there is essentially no foreground information contained in them. The second term is significant only for the first few eigenforegrounds, which have $w_i \approx 1$. Being proportional to κ^2, this term quantifies the error induced by instrumental noise in the measurement of the lower order modes (which are included in the estimator).

In Figure 7, we show (with the red/grey dashed line) the fractional error in an estimate of a foreground spectrum as a function of the total frequency range of an instrument. The channel width $\Delta\nu$ and integration time Δt are held constant, so there is no change in the noise level. From the plot, we see that as the frequency range of the instrument increases, one is able to construct an increasingly accurate estimator of the foreground spectrum. To understand why this is so, note that as the frequency range increases, there are two effects at play — first, a larger frequency range means more complicated foregrounds, possibly increasing the errors in the fits; however, a larger frequency range with fixed channel width results in more data points to fit to, as well as a longer "lever arm" with which to probe spectral features, thus reducing the errors. What we see in Figure 7 is that the second outweighs the first. This is unsurprising, since we saw from Figure 6 that as the frequency range is increased, the number of modes required to describe the foregrounds increases rather slowly, suggesting that the foreground spectra are only getting marginally more complicated.

3.4.2 Truncated Least-Squares

In a conventional least-squares fitting of a foreground spectrum, one forms an estimator for the spectrum by taking $\hat{\mathbf{a}}^{\mathbf{LS}}$ and multiplying by $\mathbf{V}$. In other words, one fits for all the eigenforeground coefficients (*i.e.* the vector $\mathbf{a}$) and multiplies each coefficient by the spectrum of the corresponding eigenforeground. This, however, is a suboptimal procedure for estimating the true spectrum because the higher order eigenmodes have very low signal-to-noise and measurements of them tend to be noise dominated. Such modes should therefore be excluded (or heavily downweighted), which was what
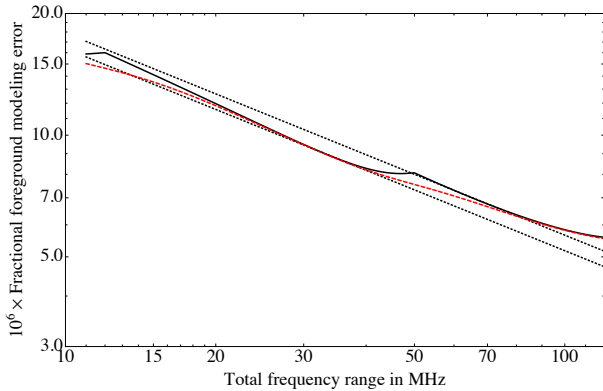
Figure 7. Expected error on measured foregrounds divided by the root-mean-square (r.m.s.) foreground intensity to give a fractional foreground modeling error. In dashed red/grey is the error for the Wiener filtering of Section 3.4.1 (the square root of Equation 45 divided by the r.m.s.). In solid black is the error for the truncated least-squares method of Section 3.4.2 (Equation 49 divided by the r.m.s. and adjusted for the fact that m must be an integer). In both cases we have plotted the errors on a log-log scale as functions of the total frequency range for an instrument with a fixed 1 MHz channel width and 1000 hrs of integration time, ensuring a constant noise level. The black dotted lines are included for reference, and are proportional to $1/\sqrt{N_c \Delta \nu}$.

the Wiener filtering of the previous section accomplished. In this section we explore a simpler method where we measure all eigenforeground coefficients, but include only the first m eigenmodes in our estimate of the spectrum, truncating the eigenmode expansion so that the noisier measurements are excluded.

In this truncated least-squares scheme, our estimate of the spectrum for the ith frequency channel takes the form

$$\hat{\mathbf{x}}_i^{LS} = \sum_j^m \hat{\mathbf{a}}_j^{LS} \mathbf{v}_j(\nu_i), \qquad (46)$$

where m is an integer to be determined later. Written in this way, we see that we can reuse all the expressions derived in the previous section as long as we set $w_i = 1$ for $i \leqslant m$ and $w_i = 0$ for $i > m$. Put another way, the truncated-least squares method approximates Wiener filtering by replacing the plot of Wiener weights shown in Figure 5 with a step function. Using Equation 45, we thus see that the mean-square error for the truncated least-squares method is

$$\begin{aligned} \varepsilon_{\mathrm{LS}}^2 &= \frac{1}{N_c} \sum_{i=1}^{N_c} \lambda_i (w_i - 1)^2 + \frac{\kappa^2}{N_c} \sum_{i=1}^{N_c} w_i^2 \\ &= \frac{1}{N_c} \sum_{i=m+1}^{N_c} \lambda_i + \frac{\kappa^2}{N_c} \sum_{i=1}^{m} 1 \\ &\approx \frac{B}{N_c} \exp\left[-A(m+1)\right] + m \frac{\kappa^2}{N_c}, \qquad (47) \end{aligned}$$

where in the last step we used the fact that the eigenvalues fall off in a steep exponential (Equation 32) that decays quickly enough that we can to a good approximation omit all but the first term in the sum.

From Equation 47, we see that the truncated least-squares method gives large errors for extreme values of m, both large and small — at large m, many modes are in-

cluded, making the exponential term in the error small, but at the cost of large noise contamination from the second term; at low m, the noise term is small, but too few eigenforegrounds are included in the estimator of the spectrum, and the first term is large. To obtain the best possible error from the truncated least-squares method, we must therefore choose an intermediate value of m that minimizes ε_{LS}^2. Differentiating with respect to m and setting the result to zero allows one to solve for the optimal m:

$$m_{opt} = \frac{1}{A} \ln\left(\frac{AB}{\kappa^2}\right) - 1, \qquad (48)$$

and inserting this into our expression for the error gives a best error of

$$\varepsilon_{LS}^{best} \sim \frac{\kappa_{fid}}{\sqrt{N_c \Delta \nu \Delta t}} \left[\frac{1}{A} \left[\ln\left(\frac{AB\sqrt{\Delta t \Delta \nu}}{\kappa_{fid}^2}\right) + 1 \right] - 1 \right]^{\frac{1}{2}}, \qquad (49)$$

where we have made the substitution $\kappa \to \kappa_{fid}/\sqrt{\Delta t \Delta \nu}$ to emphasize the scalings with integration time Δt and channel width $\Delta \nu$. Note that Equation 49 is only approximate, since m must take on an integer value.

In Figure 6, the solid black curve shows m_{opt} as a function of $N_c \Delta \nu$, fitting numerically for A and B (Equation 32) for each $N_c \Delta \nu$. The channel width $\Delta \nu$ and Δt are held at 1 MHz and 1000 hrs respectively, as described in Section 2.6. Since $\Delta \nu$ and Δt are constant, the normalized noise level κ is also constant, and the variation in m_{opt} is due purely to changes in A and B. This variation is seen to be quite weak, and we see that the optimal number of components to solve for is always a rather small number, and that this number increases only slowly with the number of channels N_c (or equivalently with the total frequency range $N_c \Delta \nu$, since the channel width $\Delta \nu$ is being held constant). This is unsurprising, given that the first term in Equation 47 decays exponentially while the second term rises only linearly, forcing the optimal m towards small m. With any current-generation experiment, we find m_{opt} to be no larger than 4. Also note that m_{opt} behaves like a discretized version of n_{eff} from the Wiener filtering, as is expected.

In Figure 7, we show how the optimal m values of Figure 6 translate into the error $\varepsilon_{total}^{best}$, again as a function of $N_c \Delta \nu$. The error is divided by the root-mean-square value of the foregrounds over the entire spectrum to give a rough percentage error, and is plotted as the solid black line, while reference lines proportional to $1/\sqrt{N_c \Delta \nu}$ are shown in dotted black. The error is seen to decrease mostly with the inverse square root of the total frequency range $N_c \Delta \nu$, as one would expect from the prefactor in Equation 49. The deviation from this behavior is due to the fact that the second part of Equation 49 depends on A and B, which in turn vary with N_c. This, however, is a rather weak effect, since the relevant parts of Equation 49 look a lot like our expression for m_{opt}, which we know from Figure 6 rises slowly. The result is that deviations from a $1/\sqrt{N_c \Delta \nu}$ scaling occur only when one is close to the transition in m_{opt} (which remember, must be an integer). Again, we see that that truncated least-squares method closely approximates the Wiener filtering method, with only slightly larger errors.

With both methods, we find that there is very little change in the error if one changes the channel width $\Delta \nu$ while correspondingly adjusting N_c so that $N_c \Delta \nu$ (*i.e.* the

total frequency range of the instrument) is kept constant. For the truncated least-squares method, such changes keep the prefactor in Equation 49 fixed, and any variations in the error are due purely to the corrections in the second part of the equation. Numerically, the lack of sharp spectral features in the foregrounds means that these corrections are found to be completely subdominant. This implies that to an excellent approximation, the errors in our measured foregrounds are dependent only on the total frequency range of our instrument, and are independent of how we bin our data — binning more coarsely results in a lower noise per frequency bin, but this is canceled out (to a very high precision) by the larger number of bins with which to perform our fits[10]. From this, we see that the decrease in errors seen in Figure 7 (where we increased the number of data points by increasing the total frequency range of the instrument) is due not to intrinsically better fits, but rather to a longer "lever arm" over which to probe foreground characteristics.

3.4.3 Which method to use?

In summary, we can see from Figure 6 that it is indeed the case that 21 cm foregrounds can be accurately characterized using just a small number ($\sim$ 3 or 4) of independent components. From Figure 7, we see that both Wiener filtering and the truncated least-squares method allow foregrounds to be estimated to an accuracy of roughly one part in 10^{-5} to 10^{-6}. This is fortunate since the cosmological signal is expected to be $\sim 10^{-4}$ smaller than the foregrounds, so a failure to reach at least that level of precision would ruin any prospects of a cosmological measurement.

With the Wiener filtering and the truncated least-squares giving such similar results, for most applications it should not matter which method is used. However, if one requires the very best foreground model achievable, then Wiener filtering should be used, since it was derived by minimizing the error. On the other hand, the least-squares method has the advantage of being simpler, and in addition is more immune to inaccuracies in foreground modeling. This is because Wiener filtering explicitly involves the foreground covariance $\mathbf{S}$, and for the method to minimize the modeling error, it is important that the foreground-to-noise ratio be accurate. On the other hand, the foregrounds only enter the least squares method via the choice of basis (*i.e.* through $\mathbf{V}$), and since this basis spans the vector space, the role of the foreground model in least-squares fitting is simply that of a prior model. Ultimately, though, the Wiener filtering method's need for an estimate of the foreground covariance is unlikely to be an obstacle in practice. As demonstrated in de Oliveira-Costa et al. (2008), it is possible to derive a principal component basis (and the corresponding eigenvalues) empirically.

4 THE DIFFICULTY IN MEASURING PHYSICAL PARAMETERS

In the previous section, we saw that the foregrounds can be accurately described using only a small number of foreground eigenmodes. Put another way, there are very few distinguishing features in radio foreground spectra, so there are really only a few independent parameters in the data. In this section, we will see that this places severe constraints on our ability to measure the physical parameters listed in Table 1.

We employ a Fisher matrix formalism to estimate the best possible error bars on measurements of these parameters. To do so, we imagine that one has already used Equation 34 to compute an estimator $\hat{\mathbf{a}}$ of the expansion coefficient vector from the data[11]. The precise value of this estimator will vary because of the random nature of instrumental noise, whose effect can be quantified by computing the error covariance matrix of $\hat{\mathbf{a}}$. Since both the Wiener filtering method and the truncated least-squares method require as a first step a full least-squares estimation of $\mathbf{a}$, we can use the standard formula for the error covariance Σ for least-squares fitting (Tegmark 1997):

$$\Sigma \equiv \langle [\hat{\mathbf{a}} - \langle \mathbf{a} \rangle][\hat{\mathbf{a}} - \langle \mathbf{a} \rangle]^{\mathbf{t}} \rangle = [\mathbf{V}^{\mathbf{t}} \mathbf{N}^{-1} \mathbf{V}]^{-1} = \kappa^2 \mathbf{I}, \quad (50)$$

where like before we used the fact that in our whitened units, $\mathbf{N} = \kappa^2 \mathbf{I}$. Once again, the angle brackets $\langle \dots \rangle$ denote an ensemble average, or (equivalently from the standpoint of practical observations) an average over many independent lines-of-sight. With this covariance, the probability distribution L of $\hat{\mathbf{a}}$ is given by

$$L(\hat{\mathbf{a}}; \boldsymbol{\Theta}) = \frac{1}{\sqrt{(2\pi)^{N_c} \det \boldsymbol{\Sigma}}} \exp \left[-\frac{1}{2} [\hat{\mathbf{a}} - \langle \mathbf{a} \rangle]^t \boldsymbol{\Sigma}^{-1} [\hat{\mathbf{a}} - \langle \mathbf{a} \rangle] \right],$$
$$(51)$$

where $\boldsymbol{\Sigma}$ is the error covariance matrix, and $\boldsymbol{\Theta} \equiv (\theta_1, \theta_2, \theta_3, \dots)$ is a vector of model parameters (such as the parameters in Table 1), which enters our expression because the true expansion coefficient vector $\mathbf{a}$ depends on these parameters. Interpreted as a function of $\boldsymbol{\Theta}$ for a given measurement $\hat{\mathbf{a}}$, L is the so-called likelihood function for the parameters that we wish to constrain. The tightness of our constraint is closely related to the Fisher information matrix, which is defined as

$$\mathbf{F}_{AB} \equiv \left\langle \frac{\partial^2 \mathcal{L}}{\partial \theta_A \partial \theta_B} \right\rangle, \quad (52)$$

where $\mathcal{L} \equiv -\ln L$. If our estimator for the parameters is unbiased, *i.e.*

$$\langle \boldsymbol{\Theta} \rangle = \boldsymbol{\Theta}_0, \quad (53)$$

where $\boldsymbol{\Theta}_0$ is the true parameter vector, then the Cramer-Rao inequality states that the error bars on θ_A (defined by

---

[10] This is provided one makes the (usually excellent) assumption that the number of frequency bins is much larger than the number of independent foreground modes, seen in this section and the previous one to be 3 or 4.

[11] Instead of considering the expansion coefficients, one could instead deal with the foreground spectrum $\mathbf{x}$ and its estimator $\hat{\mathbf{x}}$ directly. However, unlike for the expansion coefficients, the spectra themselves have the unfortunate property of having non-diagonal error covariances (*e.g.* Equation 43 for Wiener filtering), which makes the subsequent analysis and interpretation in this section more cumbersome.
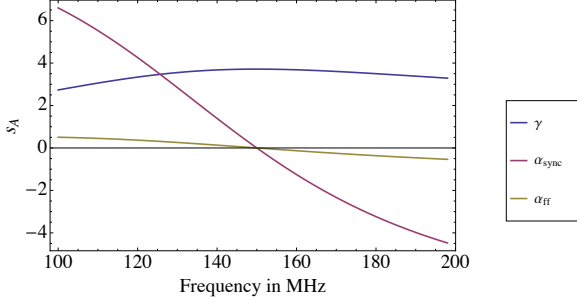
Figure 8. Parameter derivative vectors $\mathbf{s}_A$ (Equation 59) for γ, α_{sync}, and $\Delta\alpha_{sync}$.
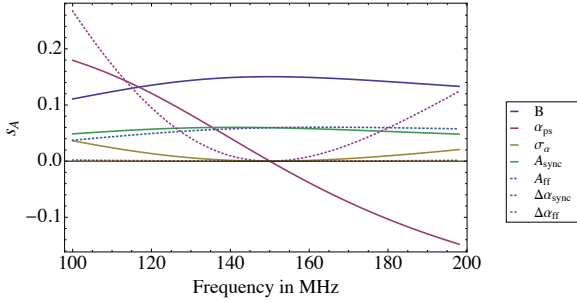

Figure 9. Parameter derivative vectors $\mathbf{s}_A$ (Equation 59) for B, α_{ps}, σ_α, A_{sync}, A_{ff}, α_{ff}, and $\Delta\alpha_{ff}$.

the standard deviation $\Delta\theta_A \equiv \sqrt{\langle\theta_A^2\rangle - \langle\theta_A\rangle^2})$ satisfy

$$\Delta\theta_A \geqslant (\mathbf{F}^{-1})_{AA}^{1/2} \qquad (54)$$

if we estimate all the parameters jointly from the data. Computing the Fisher matrix thus allows us to estimate the ability of an experiment to constrain physical parameters, with the covariance between the parameter estimates equal to $\mathbf{F}^{-1}$ if the data analysis is done in an optimal fashion.

Let us now compute the Fisher matrix for the foreground parameters listed in Table 1. Inserting Equation 51 into Equation 52 and performing some matrix algebra, one can show (Tegmark et al. 1997) that the Fisher matrix reduces to

$$\mathbf{F}_{AB} = \frac{1}{2}\mathrm{Tr}\left[\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_A}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_B}\right] + \frac{\partial\langle\mathbf{a^t}\rangle}{\partial\theta_A}\boldsymbol{\Sigma}^{-1}\frac{\partial\langle\mathbf{a}\rangle}{\partial\theta_B}. \qquad (55)$$

Referring back to our expression for $\boldsymbol{\Sigma}$, we see that the first term vanishes because $\boldsymbol{\Sigma}$ depends only on the noise level and not on the physical parameters. This gives

$$\mathbf{F}_{AB} = \frac{1}{\kappa^2}\left(\frac{\partial\langle\mathbf{a^t}\rangle}{\partial\theta_A}\right)\left(\frac{\partial\langle\mathbf{a}\rangle}{\partial\theta_B}\right). \qquad (56)$$

Note that this expression depends on the mean vector $\langle\mathbf{a}\rangle$, not the estimator $\hat{\mathbf{a}}$. This is because the Fisher matrix formalism tells us what the error bars are for the *optimal* method, whatever that method happens to be. We thus do not expect $\hat{\mathbf{a}}$ to appear in $\mathbf{F}_{AB}$, for if it did we would have the freedom to plug in a possibly sub-optimal estimator of our choosing, and by construction the Fisher matrix formalism contains information about the optimal errors.

With $\langle\mathbf{a}\rangle$ signifying the true expansion coefficient vector, we know from Equations 2 and 25 that

$$\langle\mathbf{a}\rangle = \mathbf{V^t}\langle\mathbf{x}\rangle = \mathbf{V^t}\mathbf{m}. \qquad (57)$$

Inserting this into our expression for the Fisher matrix, we obtain[12]

$$\mathbf{F}_{AB} = \frac{1}{\kappa^2}\left(\frac{\partial\mathbf{m^t}}{\partial\theta_A}\right)\mathbf{V}\mathbf{V^t}\left(\frac{\partial\mathbf{m}}{\partial\theta_B}\right) = \frac{1}{\kappa^2}\mathbf{s}_A\cdot\mathbf{s}_B, \qquad (58)$$

where we have used the fact that $\mathbf{V^t} = \mathbf{V}^{-1}$ since the eigenforegrounds are orthogonal, and have defined

$$\mathbf{s}_A \equiv \frac{\partial\mathbf{m}}{\partial\theta_A}. \qquad (59)$$

Each component of Equation 58 can be geometrically interpreted as a dot product between two $\mathbf{s}$ vectors in an N_c-dimensional space. As $\mathbf{m}$ encodes the whitened foreground spectrum (Equation 22), each $\mathbf{s}$ vector quantifies the change in the expected foreground spectrum with respect to a physical parameter. And since the final covariance matrix on the parameter errors is given by $\mathbf{F}^{-1}$, the greater the dot product between two different $\mathbf{s}$ vectors, the more difficult it is to measure the two corresponding physical parameters. The forms of different $\mathbf{s}_A$'s thus give intuition for the degeneracies in a set of parameters.

Shown in Figures 8 and 9 are plots of the $\mathbf{s}_A$ vectors for the foreground parameters shown in Table 1. For clarity, we have separated the derivatives into two plots that have different vertical scales. Many of the parameters derivatives have similar shapes, suggesting that they will have a large "dot product" in Equation 58 and therefore large degeneracies between them. Note that the overall normalization of the curves is irrelevant as far as degeneracies are concerned. This is because two parameters with identically shaped curves but different normalizations are still completely degenerate as one can perfectly compensate for changes in one parameter with smaller (or larger) changes in the other. Thus, to quantify the degree of degeneracy, we can form a matrix of normalized dot products between the $\mathbf{s}_A$ vectors, where the magnitude of each vector is individually normalized to unity. The results are shown in Table 2, where we see that the parameters form three degenerate groups — the normalization parameters (B, γ, A_{sync}, A_{ff}), forming a degenerate 4×4 block of $\approx$1's in the top left corner; the spectral index parameters (α_{ps}, α_{sync}, α_{ff}), forming a degenerate block in the middle; and the frequency coherence parameters (σ_α, $\Delta\alpha_{sync}$, $\Delta\alpha_{ff}$), forming a degenerate block in the bottom right corner. Indeed, computing the eigenvalues of the normalized Fisher matrix (*i.e.* of Table 2), one finds only three eigenvalues of order unity and higher, with the next largest eigenvalue of order 10^{-3}. This suggests that there are really only three independent foreground parameters that can be measured.

To gain an intuition for what these independent parameters quantify, consider the first few eigenvectors of the normalized Fisher matrix $\widehat{\mathbf{F}}$ shown in Table 2, that is, vectors that satisfy

$$\widehat{\mathbf{F}}\theta'_\mathbf{n} = \lambda_n\theta'_\mathbf{n}. \qquad (60)$$

These parameter eigenvectors are listed in Table 3, where

---

[12] We have implicitly assumed that the whitening procedure performed in Equations 22 was with respect to some *fiducial* foreground model, so that $\mathbf{V}$ does not depend on the foreground parameter vector $\boldsymbol{\Theta}$.

Table 2. Dimensionless dot products between $\mathbf{s}_A$ vectors (Equation 59) for the foreground parameters listed in Table 1, or equivalently, a normalized version of the Fisher matrix (given by Equation 58) so that the diagonal elements are unity. The derivatives were evaluated at the fiducial foreground parameters for the foreground model described in Section 2. The frequency range of the experiment was taken to go from 100 MHz to 200 MHz, with 50 equally spaced channels. The matrix is symmetric by construction, so the bottom left half has been omitted for clarity.

	B	γ	A_{sync}	A_{ff}	α_{ps}	α_{sync}	α_{ff}	σ_α	$\Delta\alpha_{sync}$	$\Delta\alpha_{ff}$
B	1.00	1.00	0.998	0.998	0.0603	0.111	-0.00425	0.683	0.664	0.705
γ	—	1.00	0.999	0.998	0.0603	0.111	-0.00425	0.683	0.664	0.705
A_{sync}	—	—	1.00	0.993	0.113	0.163	0.0489	0.696	0.680	0.713
A_{ff}	—	—	—	1.00	-0.00413	0.0465	-0.0684	0.661	0.639	0.688
α_{ps}	—	—	—	—	1.00	0.997	0.996	0.333	0.389	0.254
α_{sync}	—	—	—	—	—	1.00	0.987	0.399	0.454	0.322
α_{ff}	—	—	—	—	—	—	1.00	0.245	0.303	0.164
σ_α	—	—	—	—	—	—	—	1.00	0.998	0.996
$\Delta\alpha_{sync}$	—	—	—	—	—	—	—	—	1.00	0.987
$\Delta\alpha_{ff}$	—	—	—	—	—	—	—	—	—	1.00

Table 3. Eigenvectors (Equation 60) of the normalized Fisher matrix (Table 2). Each row represents an eigenvector, and going from top to bottom the eigenvectors are arranged in descending value of eigenvalue.

	B	γ	A_{sync}	A_{ff}	α_{ps}	α_{sync}	α_{ff}	σ_α	$\Delta\alpha_{sync}$	$\Delta\alpha_{ff}$
θ'_1	0.368	0.368	0.374	0.36	0.14	0.164	0.108	0.366	0.365	0.366
θ'_2	0.184	0.184	0.155	0.219	-0.532	-0.522	-0.542	-0.053	-0.088	-0.005
θ'_3	0.283	0.283	0.293	0.275	0.167	0.13	0.215	-0.442	-0.429	-0.455
θ'_4	0.006	0.006	0.002	0.01	-0.007	-0.046	-0.049	0.089	0.665	-0.739
θ'_5	-0.001	-0.001	0.001	0.	0.002	0.088	-0.087	-0.81	0.472	0.327
θ'_6	-0.001	-0.001	-0.018	0.019	0.81	-0.293	-0.506	0.004	-0.041	0.008
θ'_7	0.247	0.248	-0.125	-0.369	-0.097	0.646	-0.534	0.05	-0.081	-0.072
θ'_8	0.408	0.41	-0.236	-0.58	0.057	-0.411	0.311	-0.033	0.049	0.043
θ'_9	0.181	0.114	-0.823	0.524	0.004	0.048	0.026	0.	-0.003	0.
θ'_{10}	0.7	-0.712	0.038	-0.026	0.	-0.002	-0.001	0.	0.	0.

each row gives the weighted average of the original parameters that one must take to form the "eigenparameters". To see what foreground parameters are well characterized by 21 cm tomography, we look at the first few eigenvectors, which can be measured with the highest signal-to-noise ratio:

$$\theta'_1 \approx 0.4(B + \gamma + \sigma_\alpha + A_{sync} + A_{ff} + \Delta\alpha_{sync} + \Delta\alpha_{ff})$$
$$+0.1(\alpha_{ps} + \alpha_{ff}) + 0.2\alpha_{sync}$$
$$\theta'_2 \approx 0.2(B + \gamma + A_{ff} + A_{sync}) - 0.1(\sigma_\alpha + \Delta\alpha_{sync})$$
$$-0.5(\alpha_{ps} + \alpha_{sync} + \alpha_{ff}) - 0.02\Delta\alpha_{ff}$$
$$\theta'_3 \approx 0.3(B + \gamma + A_{sync} + A_{ff}) + 0.2(\alpha_{ps} + \alpha_{ff}) + 0.1\alpha_{sync}$$
$$-0.4(\sigma_\alpha + \Delta\alpha_{sync}) - 0.5\Delta\alpha_{ff}, \qquad (61)$$

where for clarity the coefficients have been rounded to one significant figure (see Table 3 for more precise values). The first eigenparameter weights all the parameters equally except for the spectral indices, which are somewhat downweighted. The first eigenparameter is thus roughly an "everything but spectral indices" parameter. Crudely speaking, this parameter is a generalized normalization parameter because it is mostly comprised of normalization parameters (B, γ, A_{sync}, A_{ff}) and frequency coherence parameters, whose spectral effects are not apparent until the frequency range is large. Examining the third eigenvector, we see that it is most heavily weighted towards the frequency coherence parameters, suggesting that we do have an independent parameter that acts as a "generalized frequency coherence". However, being the third eigenvector, our ability to make measure-

ments of it is lower than with the first eigenvector or the second eigenvector, which is a "generalized spectral index".

To see which degrees of freedom we cannot constrain in our foreground model, we consider the last couple of eigenparameters:

$$\theta'_9 \approx 0.18B + 0.11\gamma - 0.82A_{sync} + 0.52A_{ff}$$
$$\theta'_{10} \approx 0.7B - 0.71\gamma, \qquad (62)$$

where we have omitted all terms with coefficients less than 0.1. The last eigenvector is dominated by B and γ, which appear with almost equal and opposite magnitudes. This suggests that differences between B and γ are extremely hard to measure. In the penultimate eigenvector, the sum of coefficients for B, γ, and A_{ff} is roughly equal and opposite to the coefficient of A_{sync}. This implies that the collective difference between A_{sync} and the linear combination of B, γ, and A_{ff} is also difficult to tease out from the data, but less so than the difference between B and γ.

That we can only measure a small number of foreground parameters independently is perhaps unsurprising, given that in Figure 1 we saw that most of the foreground power comes from a small number of eigenforeground modes that lack distinctive features. Note that since κ enters our expression for the foreground parameter Fisher matrix (Equation 58) as an overall multiplicative constant, instrumental noise has no effect on the foreground parameter degeneracies we examined in this section. The parameter degeneracies are due purely to the form of the foregrounds, and have nothing to do with the noise. On the other hand, when quantifying the effective number of parameters in the Wiener

filter (Equation 42) or when solving for the optimal number of eigenforeground modes to measure in the truncated least-squares method (Equation 48), our expressions explicitly depend on the noise level. For instance, in the limit of a noiseless experiment the best characterization of foregrounds is obtained by measuring all the eigenmodes, not the three or four modes found in Sections 3.4.1 and 3.4.2. In general, however, the results of this section show that one must be able to measure at least three eigenmodes at a reasonable signal-to-noise in order to adequately model the foreground spectrum. If Equation 42 or Equation 48 suggest that fewer eigenmodes should be used in the model, it is simply because one's experiment is too noisy to allow the foregrounds to be measured accurately.

5　CONCLUSIONS

In this paper, we have shown that despite the complicated dependence that low-frequency radio foregrounds may have on physical parameters, the resulting spectra are always rather generic and featureless. The bad news from this came in Section 4, where we saw that even the most careful foreground spectrum measurements are unlikely to yield interesting constraints on foreground physics, thanks to high levels of degeneracies between different foreground parameters. This, however, is good news for those who simply consider foregrounds an impediment to 21 cm tomography. In Section 3 we saw that it was precisely because the foreground spectra were so featureless that they could be characterized to a greater accuracy than necessary for foreground subtraction using only three or four independent parameters. This bodes well for 21 cm astrophysics and cosmology, for it suggests that extremely detailed and physically motivated foreground models will not be necessary for successful foreground cleaning.

REFERENCES

Barkana R., Loeb A., 2005, Ap. J., 626, 1

Bernardi G., de Bruyn A. G., Harker G., Brentjens M. A., Ciardi B., Jelić V., Koopmans L. V. E., Labropoulos P., Offringa A., Pandey V. N., Schaye J., Thomas R. M., Yatawatta S., Zaroubi S., 2010, AA, 522, A67+

Bernardi G., Mitchell D. A., Ord S. M., Greenhill L. J., Pindor B., Wayth R. B., Wyithe J. S. B., 2011, MNRAS, pp 140–+

Bowman J. D., Morales M. F., Hewitt J. N., 2007, Ap. J., 661, 1

Bowman J. D., Morales M. F., Hewitt J. N., 2009, ApJ, 695, 183

Chang T., Pen U., Peterson J. B., McDonald P., 2008, Phys. Rev. Lett., 100, 091303

de Oliveira-Costa A., Tegmark M., Gaensler B. M., Jonas J., Landecker T. L., Reich P., 2008, MNRAS, 388, 247

Di Matteo T., Ciardi B., Miniati F., 2004, MNRAS, 355, 1053

Di Matteo T., Perna R., Abel T., Rees M. J., 2002, ApJ, 564, 576

Furlanetto S. R., Lidz A., Loeb A., McQuinn M., Pritchard J. R., Alvarez M. A., Backer D. C., Bowman J. D., Burns J. O., Carilli C. L., Cen R., Cooray A., Gnedin N., Greenhill L. J., Haiman Z., Hewitt J. N., et al. 2009, in Astro2010: The Astronomy and Astrophysics Decadal Survey Vol. 2010 of Astronomy, Astrophysics from the Highly-Redshifted 21 cm Line. pp 83–+

Furlanetto S. R., Lidz A., Loeb A., McQuinn M., Pritchard J. R., Shapiro P. R., Alvarez M. A., Backer D. C., Bowman J. D., Burns J. O., Carilli C. L., Cen R., Cooray A., Gnedin N., Greenhill L. J., Haiman Z., Hewitt J. N., et al. 2009, in Astro2010: The Astronomy and Astrophysics Decadal Survey Vol. 2010 of Astronomy, Cosmology from the Highly-Redshifted 21 cm Line. pp 82–+

Furlanetto S. R., Sokasian A., Hernquist L., 2004, MNRAS, 347, 187

Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004, Ap. J., 613, 16

Garrett M. A., 2009, ArXiv e-prints: 0909.3147

Geil P. M., Gaensler B. M., Wyithe J. S. B., 2010, ArXiv e-prints: 1011.2321

Gleser L., Nusser A., Benson A. J., 2008, MNRAS, 391, 383

Harker G., Zaroubi S., Bernardi G., Brentjens M. A., de Bruyn A. G., Ciardi B., Jelić V., Koopmans L. V. E., Labropoulos P., Mellema G., Offringa A., Pandey V. N., Pawlik A. H., Schaye J., Thomas R. M., Yatawatta S., 2010, MNRAS, 405, 2492

Harker G., Zaroubi S., Bernardi G., Brentjens M. A., de Bruyn A. G., Ciardi B., Jelić V., Koopmans L. V. E., Labropoulos P., Mellema G., Offringa A., Pandey V. N., Schaye J., Thomas R. M., Yatawatta S., 2009, MNRAS, 397, 1138

Iliev I. T., Shapiro P. R., Ferrara A., Martel H., 2002, ApJL, 572, L123

Jackson C., 2005, PASA, 22, 36

Jelić V., Zaroubi S., Labropoulos P., Thomas R. M., Bernardi G., Brentjens M. A., de Bruyn A. G., Ciardi B., Harker G., Koopmans L. V. E., Pandey V. N., Schaye J., Yatawatta S., 2008, MNRAS, 389, 1319

Liu A., Tegmark M., 2011, Phys. Rev. D, 83, 103006

Liu A., Tegmark M., Bowman J., Hewitt J., Zaldarriaga M., 2009a, MNRAS, 398, 401

Liu A., Tegmark M., Zaldarriaga M., 2009b, MNRAS, 394, 1575

Loeb A., Zaldarriaga M., 2004, Physical Review Letters, 92, 211301

Lonsdale C. J., Cappallo R. J., Morales M. F., Briggs F. H., Benkevitch L., Bowman J. D., Bunton J. D., Burns S., Corey B. E., Desouza L., Doeleman S. S., Derome M., Deshpande A., Gopala M. R., Greenhill L. J., Herne D. E., Hewitt J. N., et al. 2009, IEEE Proceedings, 97, 1497

Mack K. J., Wesley D. H., 2008, ArXiv e-prints: 0805.1531

Madau P., Meiksin A., Rees M. J., 1997, Ap. J., 475, 429

Mao Y., Tegmark M., McQuinn M., Zaldarriaga M., Zahn O., 2008, PRD, 78, 023529

McQuinn M., Zahn O., Zaldarriaga M., Hernquist L., Furlanetto S. R., 2006, Ap. J., 653, 815

Morales M. F., Bowman J. D., Hewitt J. N., 2006, Ap. J., 648, 767

Morales M. F., Wyithe J. S. B., 2010, Ann. Rev. Astron. Astrophys., 48, 127

Oh S. P., Mack K. J., 2003, MNRAS, 346, 871

Paciga G., Chang T.-C., Gupta Y., Nityanada R., Odegova J., Pen U.-L., Peterson J. B., Roy J., Sigurdson K., 2011, MNRAS, 413, 1174

Parsons A. R., Backer D. C., Foster G. S., Wright M. C. H., Bradley R. F., Gugliucci N. E., Parashare C. R., Benoit E. E., Aguirre J. E., Jacobs D. C., Carilli C. L., Herne D., Lynch M. J., Manley J. R., Werthimer D. J., 2010, The Astronomical Journal, 139, 1468

Peterson J. B., Bandura K., Pen U. L., 2006, ArXiv Astrophysics e-prints: astro-ph/0606104

Petrovic N., Oh S. P., 2011, MNRAS, pp 292–+

Pindor B., Wyithe J. S. B., Mitchell D. A., Ord S. M., Wayth R. B., Greenhill L. J., 2011, PASA, 28, 46

Santos M. G., Cooray A., 2006, PRD, 74, 083517

Santos M. G., Cooray A., Knox L., 2005, ApJ, 625, 575

Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, A & A, 345, 380

Tegmark M., 1997, Ap. J. Let., 480, L87+

Tegmark M., Eisenstein D. J., Hu W., de Oliveira-Costa A., 2000, ApJ, 530, 133

Tegmark M., Taylor A. N., Heavens A. F., 1997, Ap. J., 480, 22

Tegmark M., Zaldarriaga M., 2010, Phys. Rev. D, 82, 103501

Toffolatti L., Argueso Gomez F., de Zotti G., Mazzei P., Franceschini A., Danese L., Burigana C., 1998, MNRAS, 297, 117

Tozzi P., Madau P., Meiksin A., Rees M. J., 2000a, Ap. J., 528, 597

Tozzi P., Madau P., Meiksin A., Rees M. J., 2000b, Nuclear Physics B Proceedings Supplements, 80, C509+

Wang X., Tegmark M., Santos M. G., Knox L., 2006, ApJ, 650, 529

Wyithe J. S. B., Loeb A., Geil P. M., 2008, MNRAS, 383, 1195

Zaldarriaga M., Furlanetto S. R., Hernquist L., 2004, ApJ, 608, 622