Coaxing cosmic 21 cm fluctuations from the polarized sky using *m*-mode analysis

J. Richard Shaw,[1,*] Kris Sigurdson,[2] Michael Sitwell,[2] Albert Stebbins,[3] and Ue-Li Pen[1]

[1]*Canadian Institute for Theoretical Astrophysics, 60 Saint George Street, Toronto, Ontario M5S 3H8, Canada*
[2]*Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada*
[3]*Theoretical Astrophysics Group, Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA*

In this paper we continue to develop the *m*-mode formalism, a technique for efficient and optimal analysis of wide-field transit radio telescopes, targeted at 21 cm cosmology. We extend this formalism to give an accurate treatment of the polarized sky, fully accounting for the effects of polarization leakage and cross polarization. We use the geometry of the measured set of visibilities to project down to pure temperature modes on the sky, serving as a significant compression, and an effective first filter of polarized contaminants. As in our previous work, we use the *m*-mode formalism with the Karhunen-Loève transform to give a highly efficient method for foreground cleaning, and demonstrate its success in cleaning realistic polarized skies observed with an instrument suffering from substantial off axis polarization leakage. We develop an optimal quadratic estimator in the *m*-mode formalism which can be efficiently calculated using a Monte Carlo technique. This is used to assess the implications of foreground removal for power spectrum constraints where we find that our method can clean foregrounds well below the foreground wedge, rendering only scales $k_\parallel < 0.02 \, h\,\mathrm{Mpc}^{-1}$ inaccessible. As this approach assumes perfect knowledge of the telescope, we perform a conservative test of how essential this is by simulating and analyzing data sets with deviations about our assumed telescope. Assuming no other techniques to mitigate bias are applied, we find we recover unbiased power spectra when the per-feed beamwidth to be measured to 0.1%, and amplifier gains to be known to 1% within each minute. Finally, as an example application, we extend our forecasts to a wideband 400–800 MHz cosmological observation and consider the implications for probing dark energy, finding a pathfinder-scale medium-sized cylinder telescope improves the Dark Energy Task Force figure of merit by around 70% over Planck and Stage II experiments alone.

I. INTRODUCTION

Recent years have seen a surge in excitement at the promise of radio cosmology. By using low frequency observations of the 21 cm line we can survey the distribution of neutral hydrogen throughout large volumes of the Universe. Radio interferometers provide an efficient and cost effective method for doing this.

This transformation of radio interferometers into survey instruments has been driven by recent technological advances, particularly in the cheap low-noise amplifiers required for mobile phones, and the constant progress of Moores law making large, high bandwidth correlators economical. By correlating a large number of low cost feeds in a compact area we can produce a telescope ideally suited for wide-field surveys.

There are three main epochs we can observe: low redshift ($z \lesssim 4$), where we observe the large scale emission from unresolved galaxies, a technique termed *intensity mapping* [1,2]; the *Epoch of Reionization* ($z \sim 6$–10) where the neutral intergalactic medium is eaten away by the first ionizing sources [3]; and perhaps even the primordial structure in the *dark ages* ($z \gtrsim 30$), though observations at these very low frequencies ($\nu < 50$ MHz) will be extremely challenging [4]. These eras are of huge cosmological importance, a fact reflected in the large number of current and planned experiments targeting 21 cm observations, with GMRT [5], HERA [6], LOFAR [7], MWA [8], MITEoR [9] and PAPER [10] targeting the Epoch of Reionization and BAOBAB [11], BAORadio [12], BINGO [13], CHIME [14], EMBRACE/EMMA [15], Ooty [16], Parkes [17] and Tianlai [18] aiming at the low redshift intensity mapping era.

In this paper we will focus on the low redshift intensity mapping epoch, though most of the results and techniques we describe apply equally well at higher redshifts. Observations at these low redshifts probe the expansion history of the Universe through measurements of the baryon acoustic oscillation (BAO) [19–21]. Our radio observations are complimentary to BAO measurements from Galaxy surveys [22,23], and can probe a larger volume at high redshift with a completely different set of systematics.

---
[*]jrs65@cita.utoronto.ca

To make effective use of this new generation of radio interferometers, we must develop new methods of interpreting and analyzing their data. Progress has accelerated in recent years with many developments [24–31].

In a previous paper [32] we developed a new technique for the analysis of data from these experiments called the *m-mode formalism*. This method departs from the usual interferometric analysis—making no flat-sky or small field approximations—at the expense of being limited to transit telescopes for which it is an exact treatment. It also brings computational advantages by allowing us to break the data into uncorrelated *m* modes, making it feasible to treat the full statistics of the data. This opens up the possibility of performing optimal mapmaking, foreground subtraction and power spectrum estimation, which would be extremely difficult otherwise.

Perhaps the foremost challenge facing 21 cm cosmology is the presence of bright astrophysical radio sources at frequencies below 1.4 GHz which are around 6 orders of magnitude brighter than the 21 cm signal. This emission comes mainly from synchrotron radiation, which is spectrally smooth, and in principle this allows it to be separated from the 21 cm as it is described by a small number of modes [33] and these can simply be removed. The remaining modes, which have significant spectral structure, are assumed to be free of contamination. Unfortunately, this picture is complicated by the realities of radio observation:

a. Frequency dependent beams lead to mixing of angular structure into spectral structure which contaminates the foreground clean modes [34]. This problem, known as *mode mixing*, means that looking at *only* the frequency direction of our data is insufficient to separate these two signals.

b. Synchrotron emission from our Galaxy is highly polarized, and though the total intensity is spectrally smooth, Faraday rotation by the magnetic interstellar medium means that the polarized emission is not. Unfortunately, the complicated polarization response of real telescopes irreversibly mixes some fraction of the polarized sky, introducing significant frequency fluctuations [35]. As the emission comes from a range of Faraday depths, we cannot simply derotate the emission.

Fundamentally there are still the same number of large foreground modes, mode mixing only makes them harder to identify. In [32] we developed a foreground removal technique based on the Karhunen-Loève (KL) transform. This uses the full covariance statistics of the contaminating foregrounds to find an optimal separation from the 21 cm signal, fully accounting for this mode mixing effect. However, the technique presented there was limited in two important ways: no attempt was made to address the problem of polarized foregrounds, and it assumed that we have full knowledge of the properties of our instrument, including the full polarized response of the primary beam, and any per-feed amplitude gains and phase shifts introduced in the receiver system. In this paper we continue to develop both the *m*-mode formalism and the KL transform for foreground cleaning, with particular emphasis on investigating these two limitations.

We start by extending the *m*-mode formalism to give a full treatment of polarization (Sec. II), and discuss how the unpolarized approach of [32] is a limiting case (Sec. III). The example telescope we use throughout is described in Sec. IV, and its harmonic space sensitivity is examined in Sec. V. Next we take a careful look at the geometry of the measured *m* modes (Sec. VI), leading us to a technique which both efficiently compresses the data and effectively removes polarized contamination. We give an overview of the Karhunen-Loève scheme for foreground removal in Sec. VII, and demonstrate its effectiveness on simulated polarized skies. In Sec. VIII we construct an optimal power spectrum estimator in the *m*-mode formalism, which we use to study the performance of our foreground filter (Sec. IX). In Sec. X we use this estimator to show how instrumental uncertainties give rise to power spectrum biases. Finally we forecast the performance of our example telescope at measuring the expansion history of the Universe and constraining the nature of dark energy (Sec. XI).

II. POLARIZED TRANSIT TELESCOPE ANALYSIS

In this section we develop a fully polarized version of the *m*-mode formalism, a new method for analyzing transit interferometers that was first introduced in a previous paper [32]. That treatment encapsulates all the essential ideas but avoids the added complexity of tracking the polarization, and is a useful introduction to the full treatment given here. Polarized descriptions of full-sky interferometry have been given elsewhere (notably [36,37]), but here we develop the transit telescope limit.

Any transit telescope can be viewed as a collection of feeds, fixed relative to the ground frame. Each feed, F_i, measures a combination of the electric field $E_a(\hat{\boldsymbol{n}})$ coming from various directions in the sky. In order to accurately treat the polarization when the response varies over the sky, we need to be able to keep track of the contribution from each direction to the electric field at a point. In order to do this we define $\boldsymbol{\varepsilon}$ as the electric field density in a frequency interval $d\nu$ and solid angle $d^2\hat{n}$ by

$$d\boldsymbol{E} = (\mu_0 c)^{1/2} \boldsymbol{\varepsilon}(\hat{\boldsymbol{n}}, \nu) d^2\hat{n} d\nu. \tag{1}$$

With this definition the Poynting flux is conveniently written as

$$\boldsymbol{S}_p = \frac{1}{\mu_0 c} \boldsymbol{E} \times \boldsymbol{H}$$

$$= \int d^2\hat{n} \, d^2\hat{n}' \, d\nu \, d\nu' \, \hat{\boldsymbol{n}} \langle \boldsymbol{\varepsilon}(\hat{\boldsymbol{n}}) \cdot \boldsymbol{\varepsilon}(\hat{\boldsymbol{n}}') \rangle. \tag{2}$$

Radio emission from the sky is generally incoherent and so we can write the correlations of $\boldsymbol{\varepsilon}$ explicitly in terms of the Stokes parameters

$$
\begin{aligned}
&\langle \varepsilon_a(\hat{\boldsymbol{n}},\nu)\varepsilon_b^*(\hat{\boldsymbol{n}}',\nu)\rangle \\
&= \frac{2k_B}{\lambda^2}\delta(\hat{\boldsymbol{n}}-\hat{\boldsymbol{n}}')\delta(\nu-\nu') \\
&\times [\mathcal{P}_{ab}^T T(\hat{\boldsymbol{n}}) + \mathcal{P}_{ab}^Q Q(\hat{\boldsymbol{n}}) + \mathcal{P}_{ab}^U U(\hat{\boldsymbol{n}}) + \mathcal{P}_{ab}^V V(\hat{\boldsymbol{n}})],
\end{aligned} \quad (3)
$$

where the indices are over basis vectors transverse to the line of sight. As in the unpolarized case we are more interested in the brightness temperature on the sky than the intensity, and so we have written Eq. (3) to make that explicit (thus Q, U and V are polarization brightness temperatures). The polarization tensors $\mathcal{P}_{ab}^X$ are related to the Pauli matrices (on an orthonormal basis), specifically

$$
\mathcal{P}_{ab}^T = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathcal{P}_{ab}^Q = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},
$$

$$
\mathcal{P}_{ab}^U = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \mathcal{P}_{ab}^V = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \quad (4)
$$

The standard basis vectors to use in spherical geometry are the polar and azimuthal directions, $\hat{\theta}$ and $\hat{\phi}$, as these allow spin spherical harmonics to be used straightforwardly to decompose the polarization field.

Any feed on the telescope measures a weighted combination of the electric field coming from each direction on the sky. In particular we need to keep track of the antennas sensitivity to the orientation of the incoming electric field. We will write the measured signal at the ith feed as F_i, which is given by

$$
F_i(\phi) = \int d^2\hat{n} A_i^a(\hat{\boldsymbol{n}};\phi)\varepsilon_a(\hat{\boldsymbol{n}})e^{2\pi i\hat{\boldsymbol{n}}\cdot\boldsymbol{u}_i(\phi)}, \quad (5)
$$

and is directly proportional to the voltage induced in the circuit. Here, and onwards, we will implicitly sum over the polarization index a. The antenna reception pattern A_i^a is a vector quantity describing the electric field response in a given direction. The response $A \propto \boldsymbol{l}_{\text{eff}}$, the effective antenna length (choosing them as equal would make F_i be the antenna voltage). We normalize A such that the normalized antenna power pattern $P_n(\hat{\boldsymbol{n}}) = |A(\hat{\boldsymbol{n}})|^2$, ensuring that the solid angle of the beam is

$$
\Omega_i = \int d^2\hat{n}|A(\hat{\boldsymbol{n}})|^2. \quad (6)
$$

In Eq. (5) we have also included an exponential factor which gives the phase relative to an arbitrary reference point, in this $\boldsymbol{u}_i$, as the distance from the feed to the reference point in wavelengths. As both this and the antenna orientation change with the Earths rotation relative to the sky, we write them as functions of ϕ, the rotation angle.

The fundamental quantity in radio interferometry is the cross-correlation between two feeds, the *visibility* $V_{ij} = \langle F_i F_j^*\rangle$. Using Eqs. (3) and (5) we can write down exactly what a visibility measures, explicitly keeping track of the different sky polarizations to give

$$
\begin{aligned}
V_{ij}(\phi) = \int &[B_{ij}^T(\hat{\boldsymbol{n}};\phi)T(\hat{\boldsymbol{n}}) + B_{ij}^Q(\hat{\boldsymbol{n}};\phi)Q(\hat{\boldsymbol{n}}) \\
&+ B_{ij}^U(\hat{\boldsymbol{n}};\phi)U(\hat{\boldsymbol{n}}) + B_{ij}^V(\hat{\boldsymbol{n}};\phi)V(\hat{\boldsymbol{n}})]d^2\hat{n} + n_{ij}(\phi).
\end{aligned} \quad (7)
$$

Note that from here onwards the symbol V denotes two different quantities, the visibility $V_{ij}(\phi)$ and the Stokes V sky field $V(\hat{\boldsymbol{n}})$. The distinction will be clear from the context. In the above equation the beam transfer functions B_{ij}^X encode all the information about the optics and geometry of the instrument. They are given by

$$
B_{ij}^X(\hat{\boldsymbol{n}};\phi) = \frac{2}{\Omega_{ij}}A_i^a(\hat{\boldsymbol{n}};\phi)A_j^{b*}(\hat{\boldsymbol{n}};\phi)\mathcal{P}_{ab}^X e^{2\pi i\hat{\boldsymbol{n}}\cdot\boldsymbol{u}_{ij}(\phi)}, \quad (8)
$$

where $\Omega_{ij} = \sqrt{\Omega_i\Omega_j}$. The measured visibilities contain noise, which we include as an additional term n_{ij}. In this work we will assume that the noise from different antennas and frequencies is uncorrelated (we discuss the statistics in more detail in Appendix A).

We normalize our visibilities so that they are the correlated antenna temperature in the noiseless limit (in particular the autocorrelation *is* the antenna temperature). The factor of 2 in the definition of the transfer function ensures that for an unpolarized sky with uniform brightness T_b, the measured autocorrelation $V_{ii} = T_b$.

The above equations (7) and (8) are completely exact. The general approach to interferometric analysis is to approximate the above to a 2D Fourier transform, which is valid for small fields of view. For wide-field observations we can attempt to relax this with techniques such as mosaicing [38] and w projection [39] though this quickly becomes complicated. In our case we are interested in a specific class of *transit* interferometers intended for surveys. However, as these instruments are extremely wide field, this approach is limiting. Instead we will try a different route, restricting our domain to transit telescopes, but otherwise attempting to keep the analysis exact.

To continue, we decompose into spherical harmonics, as they are a natural way of representing fluctuations on the sky. As polarization is not a scalar field we must expand Q and U in spin-2 harmonics $Y_{lm}^{(\pm 2)}(\hat{\boldsymbol{n}})$ (the Stokes V field transforms as a scalar). This yields

$$T(\hat{\boldsymbol{n}}) = \sum_{lm} a^T_{lm} Y_{lm}(\hat{\boldsymbol{n}}), \qquad (9)$$

$$Q(\hat{\boldsymbol{n}}) + iU(\hat{\boldsymbol{n}}) = \sum_{lm} a^{(+2)}_{lm} Y^{(+2)}_{lm}(\hat{\boldsymbol{n}}), \qquad (10)$$

$$Q(\hat{\boldsymbol{n}}) - iU(\hat{\boldsymbol{n}}) = \sum_{lm} a^{(-2)}_{lm} Y^{(-2)}_{lm}(\hat{\boldsymbol{n}}), \qquad (11)$$

$$V(\hat{\boldsymbol{n}}) = \sum_{lm} a^V_{lm} Y_{lm}(\hat{\boldsymbol{n}}). \qquad (12)$$

The polarized beam transfer matrices also transform as spin fields, and so we decompose them in the same way, with

$$B^T_{ij}(\hat{\boldsymbol{n}}; \phi) = \sum_{lm} B^T_{ij,lm}(\phi) Y^*_{lm}(\hat{\boldsymbol{n}}), \qquad (13)$$

$$B^Q_{ij}(\hat{\boldsymbol{n}}; \phi) - iB^U_{ij}(\hat{\boldsymbol{n}}; \phi) = \sum_{lm} B^{(+2)}_{ij;lm}(\phi) Y^{(+2)*}_{lm}(\hat{\boldsymbol{n}}), \qquad (14)$$

$$B^Q_{ij}(\hat{\boldsymbol{n}}; \phi) + iB^U_{ij}(\hat{\boldsymbol{n}}; \phi) = \sum_{lm} B^{(-2)}_{ij;lm}(\phi) Y^{(-2)*}_{lm}(\hat{\boldsymbol{n}}), \qquad (15)$$

$$B^V_{ij}(\hat{\boldsymbol{n}}; \phi) = \sum_{lm} B^V_{ij,lm}(\phi) Y^*_{lm}(\hat{\boldsymbol{n}}). \qquad (16)$$

Note that we have decomposed with the complex conjugates of the spin harmonics. This allows us to use the orthogonality of the (spin) spherical harmonics to rewrite the visibility equation (7) as

$$V_{ij}(\phi) = \sum_{lm} \left[B^T_{ij;lm}(\phi) a^T_{lm} + \frac{1}{2} B^{(+2)}_{ij;lm}(\phi) a^{(+2)}_{lm} \right.$$
$$\left. + \frac{1}{2} B^{(-2)}_{ij;lm}(\phi) a^{(-2)}_{lm} + B^V_{ij;lm}(\phi) a^V_{lm} \right] + n_{ij}(\phi). \qquad (17)$$

Though this has completely transformed the problem into harmonic space, it will be more convenient if we change into the conventional E and B mode decomposition as they are real scalar fields [40]. This can be done by making the standard substitutions

$$a^{(+2)}_{lm} = -(a^E_{lm} + ia^B_{lm}), \qquad (18)$$

$$a^{(-2)}_{lm} = -(a^E_{lm} - ia^B_{lm}) \qquad (19)$$

as well as the corresponding changes for the beam matrices

$$B^{(+2)}_{ij;lm} = -(B^E_{ij;lm} - iB^B_{ij;lm}), \qquad (20)$$

$$B^{(-2)}_{ij;lm} = -(B^E_{ij;lm} + iB^B_{ij;lm}), \qquad (21)$$

leaving the visibility as

$$V_{ij}(\phi) = \sum_{lm} \left[B^T_{ij;lm}(\phi) a^T_{lm} + B^E_{ij;lm}(\phi) a^E_{lm} \right.$$
$$\left. + B^B_{ij;lm}(\phi) a^B_{lm} + B^V_{ij;lm}(\phi) a^V_{lm} \right] + n_{ij}(\phi). \qquad (22)$$

In the above the harmonic coefficients are now all the transforms of real scalar fields (the $B^X_{ij;lm}$ are the complex conjugates of the spherical harmonic coefficients).

As we are considering only transit telescopes, the visibilities are periodic in sidereal time, ϕ, and so Fourier transforming Eq. (22) is an obvious next step:

$$V_{ij;m} = \int \frac{d\phi}{2\pi} V_{ij}(\phi) e^{-im\phi} \qquad (23)$$

yields Fourier coefficients, $V_{ij;m}$, which we refer to as m *modes*. They will become the key quantity in our analysis. As the visibility is a *complex* time stream, the positive and negative ms are independent measurements.

As the ϕ dependence simply rotates the functions about the polar axis the transfer function is trivially $B^X_{ij;lm}(\phi) = B^X_{ij;lm}(\phi = 0)e^{im\phi}$. The integral over the exponential factors generates the Kroenecker delta $\delta_{mm'}$ and removes the summation over m entirely, and we can write the m modes as

$$V_{ij;m} = \sum_{l} [B^T_{ij;lm} a^T_{lm} + B^E_{ij;lm} a^E_{lm}$$
$$+ B^B_{ij;lm} a^B_{lm} + B^V_{ij;lm} a^V_{lm}] + n_{ij;m}. \qquad (24)$$

Though slightly hidden, this a property of the *convolution theorem*. For a transit telescope the visibility time stream is an azimuthal convolution of the beam and sky. This means its Fourier conjugate, the m modes, are products of the individual Fourier modes (with a remaining summation over the l index). This equation fully describes how the measured visibilities are related to the polarized sky that we are observing in Fourier space.

One way to gain more insight into the m modes is to realize that the visibility we see is a complex time series, which when we take its Fourier transform splits the time series into right and left moving waves (positive and negative m respectively). In our case the time variable is ϕ, the Earths rotation. As the beam on the sky is a Fourier mode, a correlated beam pointing south of the Celestial North Pole only produces modes moving in one direction as the Earth rotates. However pointing the same beam beyond the north pole (that is north of it as defined in the ground frame), produces the modes moving in the other direction as the Fourier mode on the sky now moves in the opposite direction with respect to the Earths rotation. One important consequence of this is that if we use the freedom

to choose the order of the feed pairs such that the baseline vectors point towards the east, positive m modes are produced from the part of the beam below the pole, and negative m modes come from the part above. If the primary beam does not extend over the pole only positive ms are produced, though a small amount of negative ms are seen because of the effect of the primary beam.

In fact, while the positive and negative m modes may be independent measurements they are still observations of the same sky—for a real field $a_{lm} = a^*_{l,-m}$—and thus both V^{ij}_m and V^{ij*}_{-m} measure the same harmonics on the sky. It will be useful to change our notation to make this fact transparent.

Let us separate out the positive and negative m parts by defining

$$B^{X,+}_{ij;lm} = B^X_{ij;lm} \qquad n^+_{ij;m} = n_{ij;m} \qquad (25)$$

$$B^{X,-}_{ij;lm} = (-1)^m B^{X*}_{ij;l,-m} \qquad n^-_{ij;m} = n^*_{ij;-m} \qquad (26)$$

which is valid for $m \geq 0$. Additionally to prevent double counting the $m = 0$ measurement we need to set $B^{X,-}_{ij;l0} = n^-_{ij;0} = 0$. For brevity of notation, we will introduce a label α which indexes both the positive and negative m parts of all included feed pairs ij, such that any particular α specifies exactly the values of $ij, \pm$ (exactly how α is packed is unimportant). This gives the final form of the m-mode visibility equation that we use as the basis of this work,

$$V_{\alpha;m} = \sum_l [B^T_{\alpha;lm} a^T_{lm} + B^E_{\alpha;lm} a^E_{lm}$$
$$+ B^B_{\alpha;lm} a^B_{lm} + B^V_{\alpha;lm} a^V_{lm}] + n_{\alpha;m}. \qquad (27)$$

As in [32] we can write this equation in an explicit matrix form which will allow us to simplify the notation. The beam transfer matrices above can be written in an explicit matrix notation

$$(\mathbf{B}^X_m)_{(\alpha\nu)(l\nu')} = B^{X,\nu}_{\alpha;m} \delta_{\nu\nu'} \qquad (28)$$

where the row index labels all baseline (α) and frequency combinations (ν), whereas the column index is over all multipole (l) and frequencies (ν'). Similarly we can define vectors for the visibilities and harmonic coefficients

$$(\mathbf{v}_m)_{(\alpha\nu)} = V^\nu_{\alpha;m} \qquad (\mathbf{a}^X_m)_{(l\nu)} = a^{X\nu}_{lm}. \qquad (29)$$

To keep track of the different polarization states we define the block matrix and vector

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_T & | & \mathbf{B}_E & | & \mathbf{B}_B & | & \mathbf{B}_V \end{pmatrix}, \qquad \mathbf{a} = \begin{pmatrix} \mathbf{a}_T \\ \mathbf{a}_E \\ \mathbf{a}_B \\ \mathbf{a}_V \end{pmatrix} \qquad (30)$$

such that

$$v = \mathbf{B}a + n. \qquad (31)$$

This is the essence of the m-mode formalism: a simple, linear matrix relation that exactly describes the whole measurement process for a transit interferometer. As we will discuss in Sec. V both the number of m modes and the dimensionality of the $\mathbf{B}$ matrices is bounded by the physical size of the instrument. This means that we can easily apply all the standard tools of statistical signal processing without even remembering that we are dealing with an interferometer. In the following sections we do this with gusto.

Despite this being an interferometry paper, the uv plane has not been mentioned at all so far. Though it is prevalent in many interferometric applications, as both an extremely useful aid for physical understanding and for computational efficiency (by virtue of the FFT), the m-mode formalism does not make use of it. Eschewing the uv plane is part of its power, helping it to work trivially for wide-field analysis, and focusing us on only the measured degrees of freedom. However, it comes at the cost of making it difficult to have concrete physical interpretations of the process.

III. UNPOLARIZED LIMIT

In [32] we developed an unpolarized formalism because it gives a simpler problem to analyze, both conceptually and computationally. However, under certain assumptions it is directly equivalent to the full polarized case.

For a telescope with dual polarized antennas, let us suppose that we can engineer our telescope optics such that the field patterns of the two feeds (labeled X and Y) obey two constraints. First, that their normalized power patterns are equal everywhere:

$$|\mathbf{A}_X|^2 = |\mathbf{A}_Y|^2 = A^2, \qquad (32)$$

and second that their polarization orientations are orthogonal all over the sky:

$$\mathbf{A}_X \cdot \mathbf{A}_Y = 0. \qquad (33)$$

Under these constraints there is only one relevant linear combination of the four XX, YY, XY and YX visibilities that is sensitive to the total intensity, the average of the XX and YY visibilities,

$$V_u = \frac{1}{2}(V_{XX} + V_{YY})$$
$$= \frac{1}{\Omega} \int d^2\hat{n} A^2(\hat{n}) e^{2\pi i \hat{n} \cdot u} T(\hat{n}) + \frac{1}{2}(n_{XX} + n_{YY}). \qquad (34)$$

Because of the properties of the polarization matrices, this is not sensitive to the different polarization modes Q, U and

V, while all the orthogonal combinations are insensitive to the total intensity T.

In this limit, the combination V_u is equivalent to the unpolarized formalism given in [32], if we relabel the noise terms such that $n = (n_{XX} + n_{YY})/2$. Provided the noise terms are uncorrelated this reduces the power spectrum down by a factor of 2—that is the unpolarized system temperature is $T_{\text{sys},u} = T_{\text{sys},p}/\sqrt{2}$.

IV. CYLINDER TELESCOPES

Cylinder telescopes are interferometric arrays consisting of one or more parabolic cylindrical reflectors. They have a long history in radio astronomy, with well-known facilities like the Molongo Synthesis Telescope [41], and the Ooty Radio Telescope [42]. Though advances in amplifier technology meant they steadily lost favor to dish-based interferometers, interest in them has recently been revived. Reasons are twofold: the development of cheap, room temperature, low-noise amplifiers has dramatically improved sensitivity, and 21 cm intensity mapping has provided an application for which they are ideally suited.

Intensity mapping requires a large collecting area in a compact region to achieve high brightness sensitivity, which cylinders can provide cheaply. Additionally cylinder telescopes are a cost effective way of surveying large amounts of sky at high speed [43]. And while arrays of dipoles provided a bigger instantaneous field of view, the large number of elements required at a fixed angular resolution makes the receiver and correlation hardware increasingly expensive.
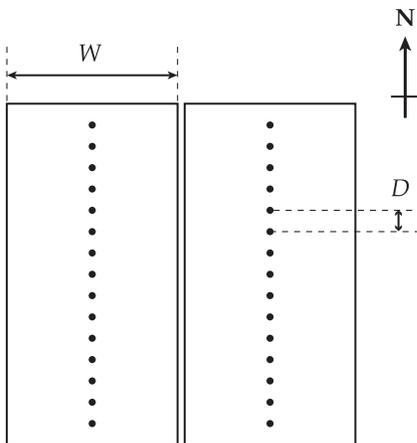
FIG. 1. A schematic of a cylinder telescope, consisting of two cylinders aligned north south on the ground. Each cylinder is of width W, and has N_{feeds} regularly spaced a distance D apart. In this paper we will only consider cylinders which are touching, making the total width of the array $2W$. The cylinders are assumed to be long enough that there are no optical differences between feeds at the edge and in the center of the array.

Each cylinder has a parabolic cross section such that they focus only in one direction. In the layout we assume (see Fig. 1), this gives a long and thin beam on the sky, extending nearly from horizon to horizon in the north-south direction but which is only around 2 degrees wide east west. Feeds are spaced along the axis of each cylinder—when correlated these provide resolution in the N-S direction. As the telescope operates as a transit telescope this means that the entire visible sky is observed once per sidereal day.

In this paper we illustrate the m-mode formalism using a medium-sized cylinder telescope, similar to the CHIME Pathfinder. Table I lists the parameters of this example instrument.

A. Beam model

In the m-mode formalism knowledge of the primary beams of our instrument is crucial. In our model we assume an arrangement such that at each location there are two perpendicular dipoles: the X feed where the dipole is aligned across the cylinder (pointing east), and Y feed where the dipole is aligned along the axis (pointing north). In both cases the feeds hang below a conducting ground plane which stops the beam spilling above the cylinder (which is assumed to have a focal ratio [f-ratio] of 1/4).

Solving for the beam on the sky for a feed placed in a parabolic cylinder is a complex problem (for one approach see [44,45]). Crudely the cylinder acts in two ways: in the parabolic direction it focuses the antenna beam to a diffraction limited beam on the sky; in the orthogonal direction it acts like a mirror, inverting the antenna beam. Rather than trying to accurately solve for the beam, we try to capture these two effects. We will break the model down into the product of two 1D functions: a function for the E-W direction, calculated by illuminating the cylinder with the dipole beam, and solving for the diffraction in the Fraunhofer limit, and a N-S function which is just the reflected feed amplitude in the N-S direction. We will also model the polarization direction as being the same as that of an unfocused dipole (in spherical coordinates, for a dipole along the polar axis, the polarization direction is $\hat{\boldsymbol{\theta}}$).

TABLE I. Parameters of the example cylinder telescope.

Parameters	Value
Number of cylinders	2
Cylinder width [m]	20
Feeds per cylinder	64 (dual-pol)
Feed spacing [m]	0.3
T_{sys} [K]	50
Bandwidth [MHz]	400–800
Channel width [MHz]	2.5
Number of Channels	160 (in groups of 40)
Telescope Latitude	45°

First we model the beam amplitude for the unfocused dipole in the E-plane and H-plane as taking the form

$$A_D(\theta; \theta_W) = \exp\left(-\frac{\ln 2}{2} \frac{\tan^2\theta}{\tan^2\theta_W}\right), \qquad (35)$$

where θ_W is the full width at *half power* of the beam. For a horizontal dipole mounted a distance $\lambda/4$ over a conducting ground plane (see [46] Sec. 4.7), we can exactly calculate the widths in the H-plane ($\theta_H = 2\pi/3$) and E-plane ($\theta_E \approx 0.675\theta_H$). We use these values for our fiducial beam model, though we will vary them later in this paper.

In the E-W direction we are solving the Frauhofer diffraction problem of a cylinder feed illuminating an aperture of a finite width. This has the solution

$$A_F(\theta; \theta_W, W) \propto \int_{-\frac{W}{2}}^{\frac{W}{2}} A_D\left(2\tan^{-1}\left(\frac{2x}{W}\right); \theta_W\right) e^{-ikx\sin\theta} dx$$

$$\propto \int_{-1}^{1} e^{-\frac{\ln 2}{\tan^2\theta_W}\frac{u^2}{1-u^2} - i\frac{\pi W}{\lambda}u\sin\theta} du \qquad (36)$$

where we have used the fact that for a cylinder with a f-ratio of $1/4$ a ray striking a distance x from the cylinder center reflects by an angle $\theta = 2\tan^{-1}(2x/W) = 2\tan^{-1}u$ where W is the cylinder width.

Putting these components together, our overall beam model can be written as the product of three functions. For the X feed

$$A_a^X(\hat{\boldsymbol{n}}) = A_F(\sin^{-1}(\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{x}}); \theta_E, W)$$
$$\times A_D(\sin^{-1}(\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{y}}); \theta_H) p_a(\hat{\boldsymbol{n}}; \hat{\boldsymbol{x}}) \qquad (37)$$

and for the Y feed

$$A_a^Y(\hat{\boldsymbol{n}}) = A_F(\sin^{-1}(\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{x}}); \theta_H, W)$$
$$\times A_D(\sin^{-1}(\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{y}}); \theta_E) p_a(\hat{\boldsymbol{n}}; \hat{\boldsymbol{y}}) \qquad (38)$$

where the vectors $\hat{\boldsymbol{x}}$ is a unit vector transverse to the cylinder, pointing east, and $\hat{\boldsymbol{y}}$ is along the cylinder, pointing north. The function p_a gives the unit vector polarization direction on the sky for a dipole in direction $\hat{\boldsymbol{d}}$

$$\hat{p}_a(\hat{\boldsymbol{n}}; \hat{\boldsymbol{d}}) = \frac{1}{(1 - (\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{d}})^2)^{1/2}} [\hat{\boldsymbol{d}} - (\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{d}})\hat{\boldsymbol{n}}]_a. \qquad (39)$$

In Fig. 2 we illustrate the on-sky beam for our example telescope. We plot the response of an "instrumental Stokes I," constructed from the combination of $XX + YY$ polarization, to Stokes I and polarizated emission on the sky. The response to Stokes I on the sky is given by

$$R_{I \to I} = (A_X^a A_X^b + A_Y^a A_Y^b)\mathcal{P}_{ab}^I. \qquad (40)$$
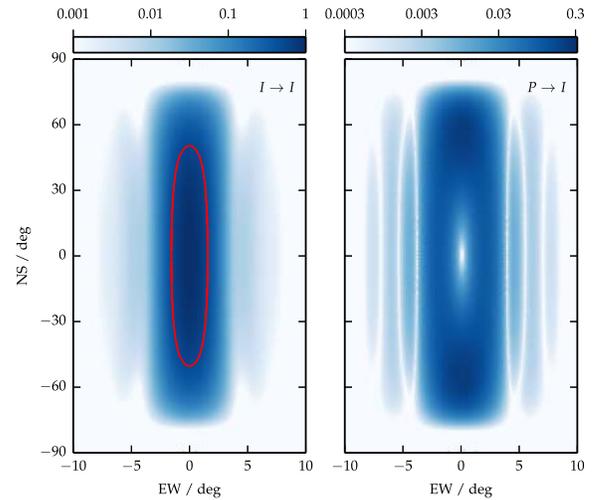
FIG. 2 (color online). The primary beam of the cylinder telescope forms a long strip on the sky from north to south. This figure illustrates the transfer into an instrumental Stokes I ($XX + YY$ polarizations), from the total intensity on the sky (left panel), and from the polarized sky only (right panel). The red contour in the top panel marks the half power point of the beam.

As a measure of the response to polarized radiation we use

$$R_{P \to I}^2 = \sum_{P \in \{Q,U,V\}} [(A_X^a A_X^b + A_Y^a A_Y^b)\mathcal{P}_{ab}^P]^2. \qquad (41)$$

For a beam with no polarization leakage, this response is zero. Though our example has no leakage on axis, Fig. 2 clearly shows that there is significant pickup of polarization away from the beam center.

V. SENSITIVITY IN HARMONIC SPACE

The geometry of an interferometer on the ground determines its angular sensitivity on the sky, with the total size of the optical system determining the smallest scales that can be measured. This limits the number of harmonic modes on the sky that we are able to measure, reducing Eq. (27) to finite sums.

The set of spherical harmonics that a given baseline is sensitive to can be found by expanding a plane wave on the sky:

$$e^{2\pi i\hat{\boldsymbol{n}} \cdot \boldsymbol{u}} = \sum_{lm} [4\pi i^l j_l(2\pi|\boldsymbol{u}|) Y_{lm}^*(\hat{\boldsymbol{u}})] Y_{lm}(\hat{\boldsymbol{n}}) \qquad (42)$$

where the part in square brackets is the coefficient in a spherical harmonic expansion of the plane wave. The amplitude of a spherical harmonic function can be conveniently written in terms of integrals of Bessel functions [[47], Sec. 5.4]. For large $l \gg 1$ we find

$$|Y_{lm}(\theta,\phi)|^2 = \frac{2l+1}{4\pi} \int_0^\infty \left[J_m\left(\frac{t \sin\theta}{2} \right) \right]^2 J_{2l+1}(t) dt$$

$$\approx \frac{l}{2\pi} J_m(l \sin\theta)^2 \tag{43}$$

where we have used the approximation that $\lim_{n\to\infty} J_n(x) = \delta(x-n)$. Combining this with Eq. (42) shows that the magnitude of the spherical harmonic coefficients of a plane wave are

$$|a_{lm}|^2 = 8\pi l j_l(2\pi|u|)^2 J_m(l \sin\theta_u)^2. \tag{44}$$

In particular this shows that the coefficients are effectively bounded in a triangle by $l < 2\pi|u|$ and $|m| < l \sin\theta_u$ because of the exponential decay of the Bessel functions for large order.

The highest frequency Fourier mode measured on a sky by an individual baseline comes from the maximum distance between illuminated areas on the correlated antennas (this is the largest distance from the origin in the uv plane). Following through from Eq. (44), we expect the range of measurable modes to be $l < 2\pi d_{\max}/\lambda$ and $m < 2\pi d_{\text{E-W}}/\lambda$, where $d_{\max}$ is the largest distance associated with the baselines and $d_{\text{E-W}}$ the largest in the E-W direction.

Let us consider our cylinder (see Fig. 1). A feed on the cylinder effectively illuminates the whole width of the cylinder, but a very short distance along its axis. This makes the largest E-W distance of all feed pairs $N_{\text{cyl}}W$, and the largest N-S distance $N_{\text{feeds}}D$. In terms of spherical harmonic coefficients on the sky, we are limited to

$$l < \frac{2\pi}{\lambda} \sqrt{(N_{\text{cyl}}W)^2 + (N_{\text{feeds}}D)^2}, \tag{45}$$

$$m < \frac{2\pi}{\lambda} N_{\text{cyl}}W. \tag{46}$$

Though this result is correct for a cylinder telescope, for an interferometer with a compact field of view, pointing away from the celestial equator, it needs modifying. As before the resolution in the E-W direction is determined by the maximum distance $d_{\text{E-W}}$, however, if the primary beam does not cross the equator this resolution corresponds to a larger fraction of the circle of constant declination at that point. As the m mode corresponds to the Fourier mode in the azimuthal direction, this means that the limit on m is in fact $m < 2\pi \cos\delta\, d_{\text{E-W}}/\lambda$, where δ is the declination of the point in the primary beam closest to the celestial equator.

To look at the sensitivity of the telescope in more detail we can calculate the Fisher matrix of the a_{lm} coefficients (we discuss the interpretation of Fisher matrices in detail in Sec. VIII). For Gaussian noise the likelihood function for the a_{lm}s is

$$\mathcal{L}(\boldsymbol{a};\boldsymbol{v}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{v} - \mathbf{B}\boldsymbol{a})^\dagger \mathbf{N}^{-1}(\boldsymbol{v} - \mathbf{B}\boldsymbol{a}) \right). \tag{47}$$

From this we can calculate the Fisher matrix for a particular m:

$$\mathcal{F}_{ll'} = -\left\langle \frac{\partial^2}{\partial a_l^T \partial a_{l'}^T} \ln\mathcal{L} \right\rangle$$

$$= [\mathbf{B}_T^\dagger \mathbf{N}^{-1} \mathbf{B}_T]_{ll'}. \tag{48}$$

We expect that in general this matrix will be singular and hence we cannot find the covariance matrix of the a_{lm} coefficients by finding $\mathcal{F}^{-1}$. One obvious source of this is that the interferometer does not see the whole sky—any declination less than $\delta = -45°$ is below the horizon—and this manifests itself as correlated combinations of a_{lm}s that we cannot separate. Additionally the angular resolution falls off towards the horizon meaning that we do not have uniform sensitivity across the sky.

The consequence of this is obvious from simply counting the degrees of freedom involved. For the example telescope, there are 762 unique baselines each of which gives a noisy complex measurement of the sky. However, we are sensitive up to $l_{\max} \sim 400$ for each polarization, giving $4(l_{\max} - m)$ complex degrees of freedom on the sky, so there must be some combinations about which we have no information.

Sensitivity in (l,m) space — $\log_{10}\left[\mathcal{F}^{-1/2} / \mu\text{K} \right]$
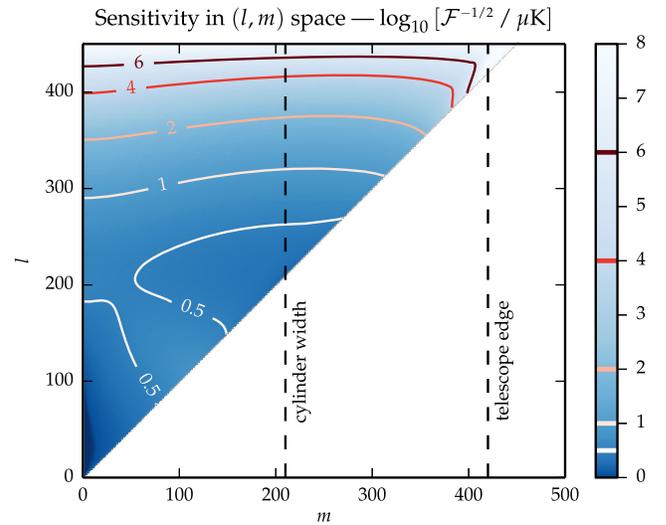
FIG. 3 (color online).   Sensitivity of the array to temperature, derived from the square-root-inverse function of the diagonal elements of the Fisher matrix $(\mathcal{F}_{(lm)(lm)})^{-1/2}$. The plot above shows the $\log_{10}$ of the sensitivity in units of μK. The sensitivity to the three remaining Stokes parameters is largely identical. The dashed black lines mark the m corresponding to the separation between the cylinders, and the total width of the cylinders. As we would expect the sensitivity peaks in m at the zero separation, and the single cylinder separation. It then falls off rapidly at the edge of the telescope.

As in general we cannot determine the covariance matrix of the a_{lm}, we will use the Fisher matrix itself to describe the sensitivity. In Fig. 3 we show the square-root-inverse function of the diagonal elements of the Fisher matrix at each m for the example telescope, this gives an illustration of the amount of information we have about any spherical harmonic mode.

VI. SVD PROJECTION

For 21 cm cosmology we are only interested in deriving real properties of the *unpolarized* sky. As we shall see this is usually of much lower dimension than the space of measurements made by an interferometer, leaving a large number of redundant degrees of freedom which are just filled by the instrumental noise. Eliminating these would allow us to significantly compress the data space, without losing useful information. In Fig. 4 we illustrate the geometry of the measured visibilities. The matrix **B** wholly describes the mapping between the sky and the measured visibilities, and understanding its structure is the key to isolating the important degrees of freedom.

To start with let us concentrate on how to reduce to only the degrees of freedom on the sky (ignoring their polarizations for now). The matrix **B** tells us how a subspace of the spherical harmonics a map into a subspace in visibility space v. This visibility subspace (shown by the plane in the Fig. 4), is termed the *image* of **B**. The subspace of
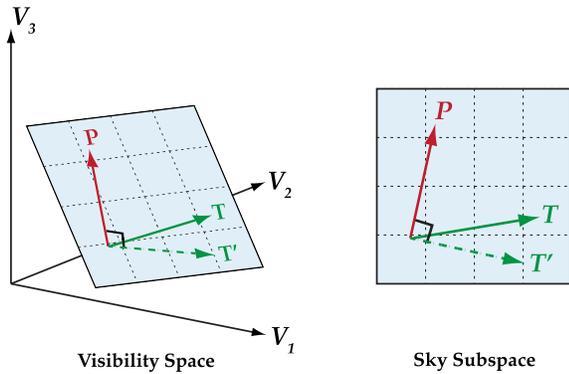
FIG. 4 (color online). The information about the sky does not spread throughout the space of visibilities but is contained in a subspace, a linear combination of the measured signals which does not span the whole visibility space. Directions orthogonal to this subspace are excited only by the instrumental noise, and contain no information about the sky. The left panel illustrates the geometry of the full visibility space, showing the sky subspace as a plane. In the right panel we show only the sky plane. Within this sky subspace, there are yet lower dimensional subspaces that the total intensity (labeled T) and polarized (P) signals get mapped to. However, they need not be orthogonal, an effect we must take into account. One way of treating this is to project onto the space orthogonal to polarization (labeled T'), this eliminates polarized contamination at the expense of some sensitivity to total intensity. This is discussed in detail in the text.

visibilities orthogonal to the image, is called the *cokernel*. The cokernel has no mapping to the sky, and so measuring this subspace yields no useful information. By projecting our data onto the image, we remove the cokernel and compress our data by retaining only the relevant degrees of freedom. In Fig. 4 this corresponds to projecting onto the plane, eliminating the perpendicular dimensions.

The number of retained degrees of freedom is given by the dimensionality of the image—that is, the *rank* of **B**—and cannot exceed the number of measured modes on the sky. For a single frequency and m, the rank is guaranteed to be less than the total number of spherical harmonics required to describe the polarized sky, that is $4(l_{\max} - m)$. However, in the case of incomplete sky coverage, we cannot measure all spherical harmonic modes independently, and this coupling means that the rank is likely to be reduced to around $4f_{\mathrm{sky}}(l_{\max} - m)$, where f_{sky} is the fraction of sky observed.

These bounds on the rank of **B** depend only on the physical size of the telescope, and not details of the feed distribution. For compact interferometers with little redundancy, the number of feed pairs rapidly exceeds the rank of the matrix. As they cannot be making independent measurements of the sky, there are many linear combinations which contain only noise. Projecting onto the image removes these and gives a large computational saving by dramatically decreasing the degrees of freedom we must consider.

To find the image of **B** we can use the singular value decomposition (SVD). However, first we will prewhiten the visibilities with respect to the instrumental noise. This transforms it to be uncorrelated with unit variance and can be done by multiplying them with $\mathbf{N}^{-\frac{1}{2}}$, a matrix such that $\mathbf{N}^{-\frac{1}{2}}(\mathbf{N}^{-\frac{1}{2}})^{\dagger} = \mathbf{N}^{-1}$. As **N** is positive definite this factorization always exists and can be found by Cholesky factorization or eigendecomposition. This leaves Eq. (31) as

$$\mathbf{N}^{-\frac{1}{2}}v = \mathbf{N}^{-\frac{1}{2}}\mathbf{B}a + \mathbf{N}^{-\frac{1}{2}}n. \qquad (49)$$

We then take the SVD of the whitened beam transfer matrix

$$\mathbf{N}^{-\frac{1}{2}}\mathbf{B} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\dagger}. \qquad (50)$$

The matrix **U** defines the image and cokernel, given by columns of **U** corresponding to nonzero and zero singular values respectively. In practice many singular values are numerically small but not precisely zero, giving modes which are either nonzero because of numerical precision or simply carry a very small but nonzero amount of information about the sky. In this case we separate the image and cokernel using a numerical threshold. We partition the columns of the matrix **U** into two matrices $\mathbf{U}_I$ and $\mathbf{U}_N$ which give the image and cokernel respectively. To compress our data we simply filter with the matrix $\mathbf{U}_I$ to give $v' = \mathbf{U}_I^{\dagger}\mathbf{N}^{-\frac{1}{2}}v$.

While this filtering can yield a large compression, we should note that it preserves all the information about the sky. However, the cosmological signal we are interested in is purely unpolarized and requires only $\sim f_{\rm sky}(l_{\rm max} - m)$ modes per frequency and m to describe it. This suggests that we should be able to improve our compression by around another factor of 4.

As a first attempt we might consider projecting onto the image of $\mathbf{B}_T$, the total intensity transfer matrix, rather than the full $\mathbf{B}$. In Fig. 4 this would correspond to projecting straight onto the T vector, rather than just the plane.

Unfortunately as illustrated in Fig. 4, the image of the total intensity need not be orthogonal to the subspace containing the polarized image. This is a manifestation of polarization leakage. In this case by doing this we lose the ability to differentiate between polarized and unpolarized signals from the sky, resulting in catastrophic leakage of polarized foregrounds.

A resolution to this problem is to project not onto the image of $\mathbf{B}_T$ but to perform another projection, this time onto the polarization cokernel. In Fig. 4 this is equivalent to projecting onto the vector T'. By doing this we ensure that there is no leakage of the polarized sky into our compressed data, at the expense of throwing away information about the total intensity signal that lies in the overlap between the two spaces.

To project out the polarized signal, we first construct the polarization transfer matrix

$$\mathbf{B}_{\rm pol} = \left(\; \mathbf{B}_E \; \middle| \; \mathbf{B}_B \; \middle| \; \mathbf{B}_V \; \right), \tag{51}$$

then we use this to isolate the polarization cokernel in the sky compressed basis by performing another SVD

$$\mathbf{U}_I^{\dagger} \mathbf{N}^{-\frac{1}{2}} \mathbf{B}_{\rm pol} = \mathbf{U}_{\rm pol} \Sigma_{\rm pol} \mathbf{V}_{\rm pol}^{\dagger}. \tag{52}$$

As before we separate into the image and cokernel of this matrix, by dividing up $\mathbf{U}_{\rm pol}$ into $\mathbf{U}_{{\rm pol},I}$ and $\mathbf{U}_{{\rm pol},N}$ respectively. As before the separation onto the two spaces is not exact, but done through a numerical threshold. By projecting our data space onto the cokernel we achieve this final compression.

Overall we have applied three transformations to our data:

a. Whiten the instrumental noise by applying $\mathbf{N}^{-\frac{1}{2}}$.
b. Project onto the sky subspace by using $\mathbf{U}_I^{\dagger}$.
c. Project out the polarized sky using $\mathbf{U}_{{\rm pol},N}^{\dagger}$.

Combined these define a new basis in which to consider our data. One which strives to preserve as much of the relevant information as possible, while vastly reducing the number of degrees of freedom we must consider. We define our filtered visibility data as

$$\bar{v} = \mathbf{U}_{{\rm pol},N}^{\dagger} \mathbf{U}_I^{\dagger} \mathbf{N}^{-\frac{1}{2}} v. \tag{53}$$

We can write a modified version of the measurement equation (31) which relates this to the sky signal

$$\bar{v} = \bar{\mathbf{B}} a + \bar{n} \tag{54}$$

where we have defined

$$\bar{\mathbf{B}} = \mathbf{U}_{{\rm pol},N}^{\dagger} \mathbf{U}_I^{\dagger} \mathbf{N}^{-\frac{1}{2}} \mathbf{B}, \tag{55}$$

$$\bar{n} = \mathbf{U}_{{\rm pol},N}^{\dagger} \mathbf{U}_I^{\dagger} \mathbf{N}^{-\frac{1}{2}} n. \tag{56}$$

As the columns of $\mathbf{U}_I$ and $\mathbf{U}_{{\rm pol},N}$ are orthonormal, the instrumental noise still has the identity covariance $\langle \bar{n} \bar{n}^{\dagger} \rangle = \bar{\mathbf{N}} = \mathbf{I}$. In Fig. 5 we show the singular values of the new mapping matrix $\bar{\mathbf{B}}$, clearly illustrating that we are only sensitive to a small number of modes on the sky.

In order to visualize our data we will want to make maps from our filtered data set. For Gaussian distributed instrumental noise it is straightforward to make maximum-likelihood maps of the sky as discussed in [32]. As we have whitened the instrumental noise, our data has a likelihood function

$$\mathcal{L}(a; \tilde{v}) \propto \exp\left(-\frac{1}{2} |\bar{v} - \bar{\mathbf{B}} a|^2 \right) \tag{57}$$

and thus we can solve for the maximum-likelihood solution using the Moore-Penrose pseudoinverse, giving our best estimate of the spherical harmonics simply as

SVD spectrum ($\log_{10} \Sigma/\Sigma_{\rm max}$)
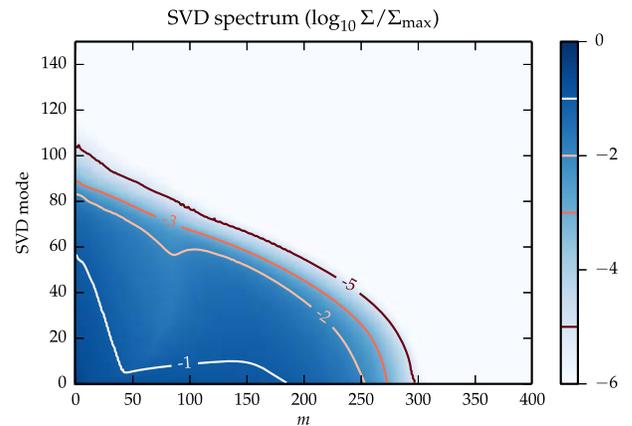
FIG. 5 (color online). The singular values of $\bar{\mathbf{B}}$ for the 400 MHz channel after removal of the polarized modes. Large singular values represent modes on the sky that are well measured. We see that at each m there are less than 100 measured degrees of freedom from the sky, with the spectrum dropping off very steeply beyond this. This is a significant saving, before compression there are twice as many modes as there are unique baselines, including positive and negative ms. In our example there are 762 unique baselines (without autocorrelations), so there would be $\sim$1500 modes.

$$\hat{\boldsymbol{a}} = \bar{\mathbf{B}}^+ \bar{\boldsymbol{v}}. \tag{58}$$

As in [32], to make a full map of the sky, we simply use this estimator on a per-m and per frequency basis and collate the estimates. We can then perform an inverse spherical harmonic transform to produce sky maps at each frequency. As we have projected onto the polarization cokernel, the data do not contain any information about the polarized sky. Combined with the minimum power property of the Moore-Penrose pseudoinverse this means that the polarized spherical harmonics will be zero.

VII. FOREGROUND REMOVAL WITH THE KARHUNEN-LOÈVE TRANSFORM

The foremost challenge for any 21 cm intensity mapping experiment is separating the cosmological signal from astrophysical contaminants which are around 10^4–10^6 times larger. The primary sources are the diffuse synchrotron emission from our own Galaxy and emission from extragalactic point sources [48]. All significant foregrounds are expected to be spectrally smooth [33], however, the 21 cm signal decorrelates quickly as each frequency corresponds to a different spatial slice. This gives an opportunity to separate the two.

Conceptually foreground removal is simple—we just remove the smooth frequency component from our observations. Unfortunately the reality is far from straightforward. The large dynamic range between the amplitude of the foregrounds and the 21 cm signal makes several effects extremely problematic.

(i) *Mode mixing*

In a real experiment the shape of the beam on the sky will vary with the observed frequency, driven by the optical effects of using a fixed physical aperture or feed spacing. Even if the angular fluctuations on the sky were frequency independent as we scan through in frequency the beam structure changes, and this introduces variations of our measurements with frequency.

(ii) *Model uncertainties*

Astrophysical foregrounds are poorly constrained at the small angular and frequency scales that will be probed by upcoming 21 cm intensity mapping experiments. While there exist theoretical and phenomenological models of this regime, a successful foreground removal method should be robust to uncertainties in the foreground statistics. Though most effort has focused on the uncertainties in the two-point correlations, we must also ensure that higher order moments do not impair our analysis.

Given these complications, we would prefer a foreground removal method to be conservative, throwing away potentially useful information in order to be robust to them. It is better to be cautiously correct than precisely wrong.

Accepting that we may prefer to lose information about the 21 cm signal in order to be unbiased by residual foregrounds, we would still like to perform the best job we can, requiring that we are statistically optimal.

(iii) *Statistically optimal*

Whatever space the foregrounds are removed in we must be able to keep track of the statistics of both the instrumental noise, and the foreground residuals in order to be able to optimally perform subsequent stages, notably power spectrum estimation.

This latter point is especially pertinent for any technique that operates directly in map space. It is not only difficult to express the pixel-pixel correlations caused by the measurement process (especially with noise), but similarly difficult to project these back after any foreground cleaning has been performed.

In a previous paper [32] we developed a foreground removal technique that addresses these three issues. It does this by explicitly taking into account the statistics of both the signal and foregrounds in the basis that they are measured. In this section we give an overview of this method in the context of the polarized analysis presented here.

A. Stationary statistics

Understanding the statistics of our measured data is an essential ingredient in all but the most basic analysis if we make best use of the data. For intensity mapping experiments, our data have three components: the 21 cm signal which we are trying to extract, the foregrounds and instrumental noise. The statistics of instrumental noise live in the visibility space, the basis of our measurements. However the other components are naturally represented on the sky, and must be projected into this space using Eq. (58).

In this work we treat the sky as a statistically isotropic field with a two-point function

$$\langle a_{lm\nu'} a^*_{l'm'\nu'} \rangle = C_l(\nu, \nu') \delta_{ll'} \delta_{mm'}, \tag{59}$$

which we write in matrix form as $\mathbf{C}_{\text{sky}}$ defined as

$$[\mathbf{C}_{\text{sky}}]_{(l\nu)(l'\nu')} = C_l(\nu, \nu') \delta_{ll'}. \tag{60}$$

This quantity can be projected into the SVD basis for a given m using the transfer matrix $\bar{\mathbf{B}}$, which means the final two-point function can be written as

$$\bar{\mathbf{C}} = \bar{\mathbf{B}} \mathbf{C}_{\text{sky}} \bar{\mathbf{B}}^\dagger + \bar{\mathbf{N}}. \tag{61}$$

As the measurement process itself does not mix m modes, provided the statistics of the sky do not couple them (which is the case for a statistically isotropic sky), then the covariance of the data is block diagonal in m. This brings huge computational savings, and makes a full

analysis tractable [32]. Clearly the observed sky is not statistically isotropic, with our own Galaxy varying wildly across the sky. However, as discussed in [32], this does not seem to diminish the effectiveness of the analysis.

These savings come because we can then operate on each block independently. For instance to diagonalize a covariance [an $O(N^3)$ operation] we can save around a factor of m_{max}^2 in computation by diagonalizing each block separately, and as we only need store the diagonal blocks storage is reduced by a factor of m_{max}.

B. Foreground removal

Any foreground removal method aims to find a subset of the data within which there is significantly more 21 cm signal than astrophysical foregrounds. Most techniques are linear, and they can be thought of as constructing a new linear basis for the data which localizes the two components into distinct regions. Unfortunately, in the presence of mode mixing, it is not obvious how to select a basis which separates the two components—what we introduce here is a method which can automatically generate it.

The signal covariances of the signal and foregrounds describe how their respective power is distributed and correlated within the measured data. It is these correlations that make the foreground fluctuations superficially seem much larger than those of the signal. In fact we expect them to be driven by a very small number of very highly correlated modes, and we would like to change to a basis where this is apparent. This can be achieved by use of the Karhunen-Loève transform (often called the signal-noise eigendecomposition), which has a long history in cosmology [49–51]. This transform simultaneously diagonalizes both the signal and foreground covariance matrices, generating an uncorrelated set of modes. This makes comparing the amount of signal and foreground power in each mode trivial.

Performing this transform requires covariance matrices for the signal and foregrounds. The signal matrix, **S**, contains only the 21 cm signal that we want to extract:

$$\bar{\mathbf{S}} = \bar{\mathbf{B}}\mathbf{C}_{21}\bar{\mathbf{B}}^\dagger \tag{62}$$

whereas the noise covariance contains the astrophysical foregrounds:

$$\bar{\mathbf{F}} = \bar{\mathbf{B}}\mathbf{C}_f\bar{\mathbf{B}}^\dagger. \tag{63}$$

This requires models for the statistics of both the signal and the foregrounds. The signal is modeled as a simple Gaussian random field for the 21 cm emission from unresolved galaxies, whereas the foreground model includes both the synchrotron emission from our Galaxy

and the contribution from a background of extragalactic point sources. The details of both are discussed in Appendix C.

Using these two matrices we can construct the Karhunen-Loève eigenbasis (see Appendix B for details on the process). This gives us a set of statistically uncorrelated eigenmodes, and corresponding eigenvalues. Writing the eigenvectors in a matrix row-wise gives the transformation matrix to diagonalize the covariances. By convention the signal covariance transforms to

$$\bar{\mathbf{S}} \to \bar{\mathbf{S}}' = \mathbf{P}\bar{\mathbf{S}}\mathbf{P}^\dagger = \Lambda, \tag{64}$$

where Λ is the diagonal matrix of the eigenvalues, and the foreground covariance becomes

$$\bar{\mathbf{F}} \to \bar{\mathbf{F}}' = \mathbf{P}\bar{\mathbf{F}}\mathbf{P}^\dagger = \mathbf{I}. \tag{65}$$

Hence, in the new basis the eigenvalues λ give the ratio of signal to foreground power. In practice the S/F spectrum is steep, with a quick transition from foreground dominated to signal dominated modes [32].

Transforming a visibility vector into the new basis is done by simply applying

$$\bar{v}' = \mathbf{P}\bar{v}. \tag{66}$$

To isolate the 21 cm signal we want to select modes which contain little foreground contamination, which can be done by picking modes with eigenvalue (S/F power) greater than some threshold. This forms a reduced basis within which the remaining modes have negligible contamination by foregrounds. To project into this basis we define the matrix $\mathbf{P}_s$ which contains only the rows from **P** corresponding to eigenvalues greater than the threshold *s*. In Fig. 6 we illustrate how the signal and foreground modes appear when projected back onto the sky.

For the purpose of power spectrum estimation (see next section) we will only require forward estimators (where we project quantities into the KL basis) and knowing $\mathbf{P}_s$ will suffice. However, for visualizing our results, we want to be able to transform back to the sky (by way of the measured visibilities). This requires us to use an inverse to map from the truncated KL basis back to the visibilities. Unfortunately because the KL modes are nonorthogonal it is ambiguous how to project back into the higher dimensional space. One obvious choice would be to make further use of the Moore-Penrose pseudoinverse. This returns a vector in the visibility space which is a linear combination of the retained signal modes while preserving their projected amplitudes. However, because the full set of modes is not orthogonal the resulting vector has a nonzero foreground amplitude (see Fig. 7 for a visual illustration).

A far better choice is to generate the full inverse $\mathbf{P}^{-1}$ and remove columns corresponding to the rejected modes (we denote this matrix $\mathbf{P}_{-s}$). This is equivalent to projecting into

Foreground Mode   Signal Mode
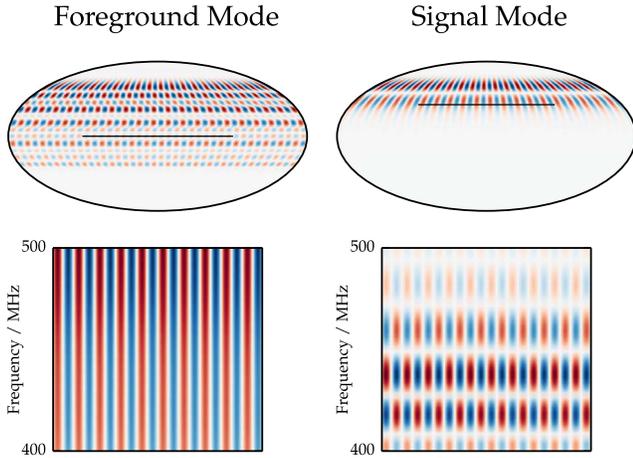
FIG. 6 (color online). Here we plot two KL modes (with $m = 20$) as they would look on the sky: one of the most foregroundlike modes ($S/F = 4 \times 10^{-13}$), and one of the most signal-like ($S/F = 170$). The lower, square panels show the frequency direction of a cut through the sky indicated on the upper panel with a black line. Though they are derived in visibility space, when projected back to the sky, they appear as we would expect with the foreground mode having a smooth frequency spectrum, and the signal mode oscillating. Modes at either end of the spectrum, like the ones plotted, are easy to interpret, this is not generally true of the intermediate modes.

the full KL basis, zeroing the foreground contaminated modes, and the using the full inverse to return the visibility space. The distinction with the pseudoinverse is shown in Fig. 7.

To demonstrate the foreground removal process we project separate realizations of the total intensity foregrounds, polarized foregrounds (showing Stokes Q only), and the 21 cm signal, through the filtering process (see Fig. 8). We show the original simulations, and the total
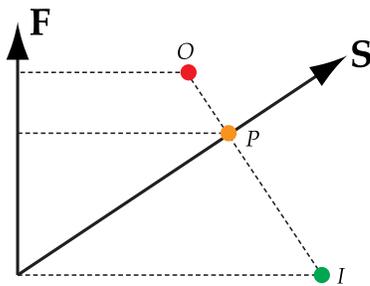
FIG. 7 (color online). To remove foregrounds from our data (point O), we separate our space into two subspaces of foreground contaminated modes, and signal modes (denoted by **F** and **S**). These spaces are not guaranteed to be orthogonal. Inverting with the pseudoinverse gives the linear combination of signal vectors with the same amplitude, however, the resulting vector P is clearly contaminated by foregrounds (as the projection onto **F** is nonzero). The full inverse gives point I, which has the same projection onto **S**, but contains no foregrounds, however, it is necessarily a combination of both **S** and **F**.

intensity maps made from the foreground filtered visibilities of each component. This illustrates how the foreground amplitude is dramatically reduced by the process, while the signal retains its overall character.

C. Double-KL transform

So far we have neglected the effects of instrumental noise. To add the instrumental noise back in we simply transform all noise contributions into the new basis. Writing the total noise contribution as $\mathbf{N}_{\text{all}} = \bar{\mathbf{F}} + \bar{\mathbf{N}}$, the matrix in the truncated basis is

$$\mathbf{N}^{\text{all}} \rightarrow \mathbf{N}_s^{\text{all}} = \mathbf{P}_s(\bar{\mathbf{F}} + \bar{\mathbf{N}})\mathbf{P}_s^{\dagger} \tag{67}$$

$$= \mathbf{I} + \mathbf{P}_s\mathbf{N}\mathbf{P}_s^{\dagger}. \tag{68}$$

Though this transform ensures that our foreground contamination remains minimal, as the transformed instrumental noise matrix will not remain diagonal this gives a correlated component between all our modes. However, for further analysis it will be particularly useful if the set of modes we use in our calculation is uncorrelated. By making a further KL transformation on the foreground removed signal $\mathbf{S}_s = \Lambda_s$, and total noise $\mathbf{N}_t^{\text{all}}$ covariance matrices, we find a new transformation matrix $\mathbf{Q}$ which maps into a basis where this is true. We will apply a further cutoff to this, including only modes with a signal to total noise ratio greater than s to give a transform $\mathbf{Q}_t$.

For notational convenience we will write the total transformation in terms of a single matrix $\mathbf{R} = \mathbf{Q}_t\mathbf{P}_s$, having chosen suitable values for the two cutoffs s and t. Quantities in this final basis we denote with tildes, for example a visibility mapped into this basis is $\tilde{\nu} = \mathbf{R}\bar{\nu}$, and a covariance is $\tilde{\mathbf{C}} = \mathbf{R}\bar{\mathbf{C}}\mathbf{R}^{\dagger}$. We will denote the signal covariance $\tilde{\mathbf{S}} = \tilde{\Lambda}$, and the total noise covariance (including foregrounds) as $\tilde{\mathbf{N}} = \mathbf{I}$.

VIII. POWER SPECTRUM ESTIMATION

In cosmology we are primarily interested not in the individual structures we see, but in their global properties. It is these statistical observations which tell us about the fundamental nature of the Universe. The quantity we are most interested in is the power spectrum which encodes most of the cosmological information in its shape and evolution. In particular for 21 cm intensity mapping it allows us to measure the position of the BAOs, which in turn can shed light on the time evolution of dark energy [19].

In order to determine the power spectrum shape we first need to parametrize it. We choose to model the two-dimensional, real-space comoving power spectrum, describing it as a linear summation of different basis functions
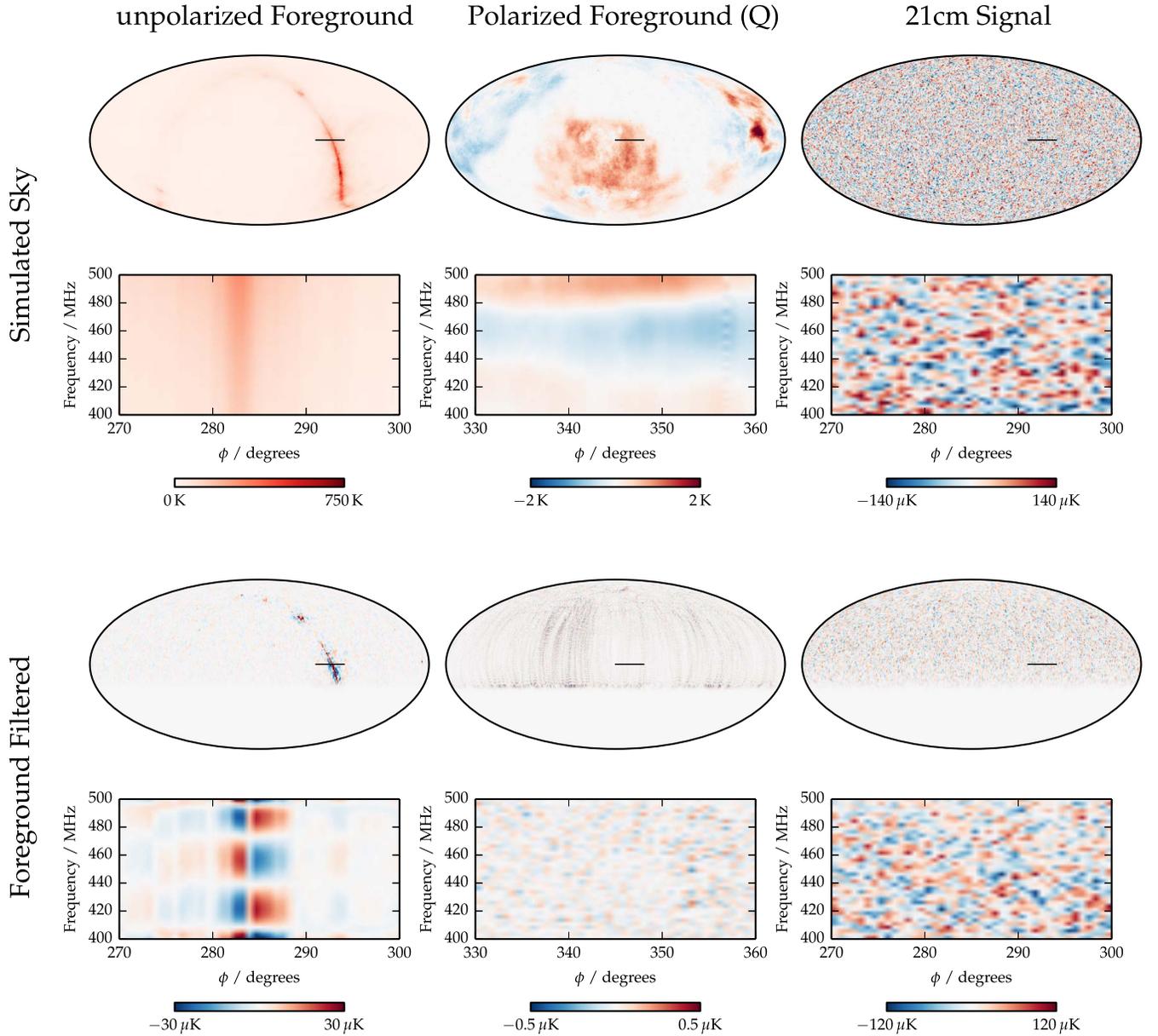
FIG. 8 (color online).   This figure illustrates the process of foreground removal on simulations of the radio sky. The top row of panels shows sky maps of the individual components: unpolarized foregrounds, polarized foregrounds (showing Stokes Q only) and the 21 cm signal. On the bottom row we show the total intensity maps we would make after foreground cleaning visibilities from our example telescope. Particularly, the lower middle panel is showing the leakage of polarized foregrounds into the total intensity. Again, the square panels show the frequency direction for the slice of the sky indicated with a black line. Both the polarized and unpolarized foregrounds become substantially suppressed, whereas the 21 cm signal is largely unaffected. In this example we have discarded modes with $S/F < 10$. This leaves a clear correspondence between the original signal simulation and the foreground subtracted signal, while leaving the foreground residuals over ten times smaller in amplitude.

$$P(\boldsymbol{k}) = \sum_a p_a P_a(\boldsymbol{k}). \tag{69}$$

In this paper we decompose k space into bands in $k_{\parallel}$ and $k_{\perp}$, such that each band represents a ring around the line of sight axis in the full three-dimensional k space.

We can calculate the accuracy we could achieve measuring the power spectrum using the Fisher information matrix, which provides a method for predicting our ability to constrain arbitrary sets of parameters, and has become the essential tool in cosmology for forecasting. The Fisher matrix is defined as

$$F_{ab} = -\left\langle \frac{\partial^2}{\partial p_a \partial p_b} \log \mathcal{L}(\boldsymbol{p}; \tilde{\boldsymbol{\nu}}) \right\rangle_{\tilde{\boldsymbol{\nu}}} \tag{70}$$

where $\mathcal{L}$ is the likelihood function and the p_a are the parameters we are trying to forecast. In the limit that we are measuring the Fisher information for the true parameters $\boldsymbol{p} = \boldsymbol{p}_0$ that generate the data $\tilde{\boldsymbol{v}}$, and the priors are uniform in this region, the inverse $\mathbf{F}^{-1}$ gives a lower bound on the errors of any unbiased estimator (the Cramér-Rao bound), and can be viewed as a forecast for the covariance of the p_a.

Let us specialize this to the case of estimating the power spectrum. After projection into the foreground cleaned basis we assume that the remaining modes follow a complex Gaussian distribution with zero mean. This assumption should be reasonable provided we have successfully removed the modes containing any significant foreground contribution—it is these modes which contain the most non-Gaussian contributions. In this case the Fisher information matrix of a single m mode for a set of parameters p_a is

$$F_{ab}^{(m)} = \mathrm{Tr}[\tilde{\mathbf{C}}_a \tilde{\mathbf{C}}^{-1} \tilde{\mathbf{C}}_b \tilde{\mathbf{C}}^{-1}], \tag{71}$$

where $\tilde{\mathbf{C}}_a$ is the linear response of the data covariance to a change in p_a, that is

$$\tilde{\mathbf{C}}_a = \frac{\partial \tilde{\mathbf{C}}}{\partial p_a} = \frac{\partial}{\partial p_a} \langle \tilde{\boldsymbol{v}} \tilde{\boldsymbol{v}}^\dagger \rangle. \tag{72}$$

For power spectrum forecasting, the p_a are the amplitudes of our power spectrum bands [see Eq. (69)]. To calculate the response $\tilde{\mathbf{C}}_a$ we need to project the band functions $P_a(\boldsymbol{k})$ into the cleaned basis. First, the spatial representation $P_a(\boldsymbol{k})$ must be turned into a multifrequency angular power spectrum $C_l(\nu, \nu') = \langle a_{lm}(\nu) a_{lm}^*(\nu') \rangle$. We do this by using a simple linear flat-sky prescription which includes the effects of both redshift distortion and structure growth (see Appendix C). We denote the matrix representation of the angular power spectrum basis function as $\mathbf{C}_a$. This must be projected into the KL basis

$$\tilde{\mathbf{C}}_a = \mathbf{R} \mathbf{B} \mathbf{C}_a \mathbf{B}^\dagger \mathbf{R}^\dagger. \tag{73}$$

In practice explicitly calculating the $\tilde{\mathbf{C}}_a$ this way is computationally very expensive. We will discuss a fast Monte Carlo alternative for calculating the Fisher matrix later in this section. In the constructed eigenbasis $\tilde{\mathbf{C}} = \tilde{\Lambda} + \mathbf{I}$ is exactly diagonal, however, $\tilde{\mathbf{C}}_a$ can have off-diagonal elements.

As there is no coupling between them, the total Fisher information for the whole data set is simply the sum over the individual m modes

$$F_{ab} = \sum_m F_{ab}^{(m)}. \tag{74}$$

The Fisher matrix gives us the ability to forecast how well we can possibly measure the power spectrum, but it does not tell us how to go about estimating that power spectrum. We will use the quadratic power spectrum estimator of [52,53]. This is an *optimal* estimator in that it achieves the Cramér-Rao bound giving it the lowest possible variance. We will give an overview of this estimator below, though we encourage the reader to look at the original papers for more detail.

As our data set is made up of a large number of independent m modes, for simplicity we will start with the power spectrum estimator for the whole data set, and then break it down into individual modes which can be calculated simply.

For notational simplicity, it is most convenient to start with a related estimator

$$\hat{q}_a = \tilde{\boldsymbol{v}}^\dagger \mathbf{E}_a \tilde{\boldsymbol{v}}, \tag{75}$$

where the quantities on the right-hand side include all ms. This forms a weighted combination of all the quadratic pairs $\tilde{\boldsymbol{v}} \tilde{\boldsymbol{v}}^\dagger$. Our actual power spectrum estimator is built out of linear combinations of the q estimator:

$$\hat{p}_a = \sum_b M_{ab}(\hat{q}_b - b_b). \tag{76}$$

In this b_b subtracts the additive bias from the instrumental and foreground noise, and the mixing matrix M_{ab} takes linear combinations such that $\hat{p}_a$ is related to the actual power spectrum. Our estimator will have minimum variance with the choice

$$\mathbf{E}_a = \tilde{\mathbf{C}}^{-1} \tilde{\mathbf{C}}_a \tilde{\mathbf{C}}^{-1}, \tag{77}$$

where again these matrices include all ms. For a detailed derivation of this weighting, see [52]. Requiring $\hat{p}_a$ to be an unbiased estimator of the power spectrum we can determine the noise bias term

$$b_a = \mathrm{Tr} \mathbf{E}_a \tilde{\mathbf{N}}. \tag{78}$$

Our remaining choice is that of the mixing matrix M_{ab} which gives the exact link between our estimator and the "true" power spectrum. In particular, we care about the expectation of the estimator

$$\langle \hat{p}_a \rangle = \sum_b W_{ab} p_b, \tag{79}$$

which we have written in terms of a window function W_{ab} which mixes the power spectrum bands. Using Eqs. (75) and (77) we find that

$$\langle \hat{q}_a - b_a \rangle = \mathrm{Tr}[\mathbf{E}_a(\tilde{\mathbf{C}} - \tilde{\mathbf{N}})] \tag{80}$$

$$= \sum_b F_{ab} p_b \tag{81}$$

and combining this with Eq. (79) gives the window function as

$$W_{ab} = \sum_c M_{ac} F_{cb}. \tag{82}$$

We fix the normalization by requiring that $\sum_b W_{ab} = 1$. Our choice of the mixing matrix M_{ab} also affects the covariance of the estimator, giving

$$\text{Cov}(\hat{p}_a, \hat{p}_b) = \sum_{cd} M_{ac} M_{bd} F_{ab}, \tag{83}$$

where we have used the fact that $\text{Cov}(\hat{q}_a, \hat{q}_b) = F_{ab}$.

There are three common choices for the mixing matrix M_{ab} [54]:

a. *Unwindowed*: Choosing the window function to be the identity means that $\langle \hat{p}_a \rangle = p_a$. This corresponds to $M_{ab} = F_{ab}^{-1}$. This is the most natural choice, however it gives highly correlated errors bars.

b. *Uncorrelated*: To decorrelate the estimator covariance, we choose $M_{ab} = [\sum_b F_{ab}^{1/2}]^{-1} F_{ab}^{-1/2}$. This leads to uncorrelated estimates, but leads to mildly spread window functions [31].

c. *Minimum variance*: The minimum variance estimator requires that the mixing matrix is diagonal $M_{ab} = [\sum_c F_{ac}]^{-1}$, and gives window functions with moderate spread.

We are generally interested in the Unwindowed estimator and we will use this in our forecasts, however, for convergence reasons that we discuss later, we will also use the minimum variance estimator when estimating power spectrum biases.

Naive calculation of this estimator is problematic because of the large dimensionality of the data. However, we can trivially exploit the independence of the individual m modes to simplify this calculation. Noting that the covariance matrices in Eq. (77) are block diagonal in m because they are statistically independent, the weight matrix $\mathbf{E}_a$ is also block diagonal. This means we can rewrite the q estimator as a sum of separate estimators for each m:

$$\hat{q}_a = \sum_m \hat{q}_a^{(m)}, \tag{84}$$

with

$$\hat{q}_a^{(m)} = \tilde{\mathbf{v}}_m^\dagger \mathbf{E}_a^{(m)} \tilde{\mathbf{v}}_m, \tag{85}$$

where the $\mathbf{E}_a^{(m)}$ are the diagonal blocks of $\mathbf{E}_a$, and $\tilde{\mathbf{v}}_m$ is the data for each m mode. Similarly we will also break up the bias terms into contributions from each m. The total bias

$$b_a = \sum_a b_a^{(m)} \tag{86}$$

where the individual

$$b_a^{(m)} = \text{Tr} \mathbf{E}_a^{(m)} \tilde{\mathbf{N}}^{(m)}. \tag{87}$$

Unfortunately exact calculations of the Fisher matrix $F_{ab}^{(m)}$ and the bias $b_a^{(m)}$ are computationally difficult. While many aspects of the calculation can be simplified by the fact that $\tilde{\mathbf{C}}$ is diagonal, the need to explicitly construct covariances of $\tilde{\mathbf{C}}_a$ and $\tilde{\mathbf{N}}$ in the KL basis is still prohibitive. To avoid this, we follow [31,54] and construct a Monte Carlo scheme to evaluate the Fisher matrix.

The key to this scheme is that evaluating the q estimator is quick as we do not need to explicitly construct any large matrices. This is achieved by constructing the intermediate vector

$$\mathbf{w} = \bar{\mathbf{B}}^\dagger \mathbf{R}^\dagger \tilde{\mathbf{C}}^{-1} \tilde{\mathbf{v}}, \tag{88}$$

which can be efficiently evaluated from right to left. The q estimator is then

$$\hat{q}_a^{(m)} = \mathbf{w}^\dagger \mathbf{C}_a \mathbf{w}. \tag{89}$$

As $\mathbf{C}_a$ is block diagonal in l this can be quickly evaluated. To estimate the Fisher matrix we draw many random realizations of our data set to which we apply the q estimator. Then, noting that the covariance of $\hat{q}_a$ is

$$\text{Cov}(\hat{q}_a^{(m)}, \hat{q}_b^{(m)}) = F_{ab}^{(m)} \tag{90}$$

we can evaluate the sample covariance of our q samples to form an estimate of the Fisher matrix. To estimate the bias term we use the fact that

$$b_a^{(m)} = \langle \tilde{\mathbf{n}}^\dagger \mathbf{E}_a \tilde{\mathbf{n}} \rangle = \langle q_a^{(m)} \rangle_{\tilde{\mathbf{N}}}, \tag{91}$$

and take the average of the q estimator under random realizations of the noise.

This Monte Carlo scheme converges rapidly enough that it is effective for forecasting. Unavoidably there will be small of-diagonal terms in the estimated Fisher matrix which do not converge exactly, and these errors can become amplified when taking powers to construct the mixing matrix M_{ab}. These errors remain small enough that they are not apparent when performing power spectrum estimation on data close to the fiducial model, and in most cases this Monte Carlo technique is still sufficient. However, for data significantly biased from the fiducial model this could be significant, particularly when using the unwindowed estimator which requires the inverse of the Fisher matrix.

Specifically for the unwindowed estimator we can see that the problem is that the effective window function of the power spectrum, $W_{ab}^{(\text{eff})}$, for a particular Monte Carlo estimate of the Fisher matrix $F_{ab}^{(\text{mc})}$ is

$$W_{ab}^{(\text{eff})} = \sum_c (F^{(\text{mc})})_{ac}^{-1} F_{cb} \neq \delta_{ab}. \qquad (92)$$

The size of the off-diagonal elements in this effective window function determines how much the power is spuriously spread around. Though we do not have access to the true Fisher matrix in order to calculate $W^{(\text{eff})}$, we can estimate the size of the off-diagonal terms by using multiple Monte Carlo estimate, in particular using one estimated with a much larger number of samples in place of the true Fisher matrix F_{ab}. For our default choice of 2000 samples, this leads to off-diagonal terms of order 10^{-2} to 10^{-3}, suggesting that biases of 10^2 to $10^3 \sigma$ will produce significant spurious power.

However, using the minimum variance estimator, which does not require us to calculate any powers of the Fisher matrix, alleviates this problem as it only requires the diagonals of the Fisher matrix. This is the route we take when dealing with the biased data we will find in Sec. X.

IX. DISCUSSION

A. Polarized foreground removal

Foreground cleaning inevitably throws away measured information about the sky, and is guaranteed to reduce our sensitivity to the 21 cm signal we are seeking. As our primary interest is to measure the 21 cm power spectrum it is vital that we understand how foreground cleaning methods affect our power spectrum errors. Over the previous sections we have developed the tools to tackles this: in Sec. II we saw how the *m*-mode formalism gives us a simple and efficient description of the measurement process; Sec. VII developed an effective foreground cleaning method based on the KL transform that allows us to easily track the statistics of our data through the cleaning; and in the previous section (Sec. VIII) we constructed an optimal estimator for the power spectrum, and forecast its errors using the Fisher matrix. Here, we combine these to forecast the performance of our example telescope in the presence of foregrounds.

In Fig. 9 we show the power spectrum errors for observations of the 400–500 MHz band with our example telescope. We forecast three distinct sets of foregrounds: *no* foregrounds; foregrounds with *no polarized* emission; and *partially polarized* foregrounds. We use values of the foreground amplitudes and spectral correlation that are representative of those in our Galaxy, these models are described in detail in Appendix C. In particular the latter includes the effects of Faraday rotation, especially emission from a range of Faraday depths within our Galaxy, that produce significant spectral structure in the polarized emission.

Clearly the dominant effect of foreground removal in both cases is that we become insensitive to power at low $k_\parallel$, with a slight increase in the errors across k space. This is in
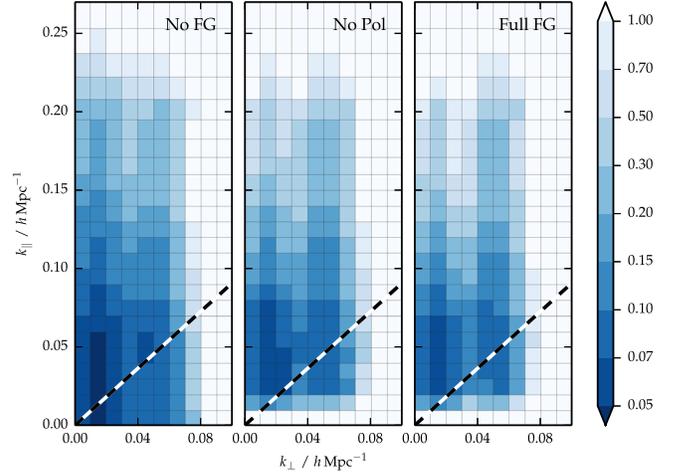
FIG. 9 (color online). Forecast errors on the power spectrum as a fraction of its fiducial value for the 400–500 MHz band. The three panels show the predicted errors without foregrounds (left), with unpolarized foregrounds (center) and with polarized foregrounds (right). The dashed line indicated the predicted bound of the "foreground wedge," showing that with perfect knowledge of our instrument foregrounds can be successfully cleaned well into this region.

line with our expectations that the foregrounds contaminate the large scale frequency modes corresponding to small $k_\parallel$, though we discuss how this relates to the foreground wedge of [29,55] later.

Polarized foregrounds are removed primarily by the action of the SVD filter described in Sec. VI. This leads to only a slight worsening of the errors compared to the case of unpolarized foregrounds only. One concern could be that the SVD filter does not discriminate between polarized modes on the basis of the magnitude of their contamination (as would be done by a KL-based filter), it removes them all. This approach is not perfectly optimal, and could be improved by allowing all polarization modes to propagate through and let the KL filter determine which to remove. In tests on smaller examples, this approach yields no significant improvement, but due to computational limitations cannot be demonstrated in the example in this work.

In all the cases illustrated in Fig. 9 there are clear peaks in the sensitivity in the $k_\perp$ direction that correspond to those seen in Fig. 3, and a rapid dropoff as we approach the resolution limit of the telescope. Additionally at low $k_\perp$ we can see there is a reduction in sensitivity caused by the sample variance of the small number of large scale angular modes.

B. Foreground wedge

Previous studies of the performance of 21 cm experiments in the face of large astrophysical foregrounds have found the bulk of the contamination to lie in a wedge shaped region of $k_\parallel < \beta k_\perp$ (for an experiment dependent

constant β), termed the foreground wedge [29,55,56]. In these studies, the complement of this region remains largely free of contamination, and is thought to provide the best chance for observing cosmological 21 cm radiation (in the context of Epoch of Reionization observation this region is called the *EoR window*).

Important progress has been made in recent years understanding the source of this contamination [29,55]: spectrally smooth radio emission is observed at a delay which depends on the baseline length, and distance of the emission from the phase center, the phase rotation with frequency from this delay appears like fluctuations along the line of sight. This argument leads us to predict that spectrally smooth sources contribute power within a region

$$k_\parallel < \Delta\theta\left(\chi(z)\frac{H(z)}{c(1+z)}\right)k_\perp, \tag{93}$$

where $\chi(z)$ is the comoving distance to redshift z, and $\Delta\theta$ is the maximum observable distance from the beam center.

In Fig. 9 we mark the boundary of the foreground wedge for our example telescope. While foreground removal makes us insensitive to small $k_\parallel$ there is no discernible variation of this with $k_\perp$, and we can observe well into the foreground wedge. Clearly there is no fundamental loss of information about the entire wedge. Though the distinction between the information lost here, and the whole wedge is small for our example, for a larger telescope with higher angular resolution the difference will be significant.

As pointed out in [33] we expect the foregrounds along each line of sight to be described by only a small number of eigenmodes (with those beyond five contributing less than 10^{-10} in power). Though mode mixing may make these modes appear to contribute power throughout the foreground wedge, fundamentally there are only a small number of them. The KL transform projects these eigenmodes forward into the data basis while keeping track of how their angular structure correlates different baselines. If our knowledge of the telescope is perfect, we can use these modes to exactly project out the large foreground contributions to the data.

If our knowledge of the telescope is not perfect as in our forecasts (e.g. Fig. 9), we cannot perfectly remove the foregrounds. We investigate this in the following section.

X. AN UNCERTAIN WORLD

So far we have demonstrated that the 21 cm signal can be separated from the astrophysical foregrounds in a way which does not distort our measurement of the underlying power spectrum. This assumed an ideal instrument about which our knowledge was perfect in every sense, conditions that a real telescope will not meet. There are many sources of nonideality—primary beam response, amplifier gains, cable delays and noise temperatures are just a few—each of which could distort our measurements. We can divide these nonidealities into two classes:

a. *Known deviations* from the design can be incorporated into our analysis to keep it unbiased and optimal, though our ultimate sensitivity may change relative to the design.

b. *Unknown deviations* from our best model of the instrument cannot be corrected and will lead to bias from both foreground leakage and using a biased power spectrum estimator.

The second class of deviations is the most serious, and so for these effects we would like to know how large our uncertainty can be before it matters, or more precisely before it is significant compared to the statistical errors.

As our ability to separate signal and foregrounds requires detailed knowledge of our instrument, we can form a naive expectation of the allowed uncertainty from the dynamic range between signal and foregrounds. In the smooth frequency modes where foregrounds dominate, they are around 10^5 times brighter than the 21 cm (10 K versus 0.1 mK), and so we expect that knowing our instrumental gains and beam shapes to 10^{-5} accuracy should be sufficient.

In this section we aim to test two particular forms of uncertainty that we can parametrize simply in our model telescope to see if the requirements are as stringent as 10^{-5}. Our approach is to assume that our example telescope represents our best knowledge about the state of the system, which we use to generate our foreground cleaning filter and our power spectrum estimator. We then generate a *corrupted time stream* corresponding to the observations the true telescope would make. By analyzing this time stream with the filters generated for the example telescope we can see at what point imperfect knowledge leads to significant power spectrum biasing.

A. Gain fluctuations

A receiver system turns the input antenna voltage into a signal which can be measured and correlated. In the process of doing this a complex gain may be applied, and while this can be corrected for, this generally leaves unknown residuals in the data. This gain residual is unique to each feed and may be time and frequency dependent.

We model gain fluctuations on a feed by feed basis, as a complex perturbation around a nominal gain of unity. The perturbed feed input is

$$F_i{}' = (1 + \Delta g_i)F_i \tag{94}$$

where the perturbation Δg is a complex Gaussian random variable with variance $\langle \Delta g_i \Delta g_i^* \rangle = \sigma_g^2$. These combine to give corrupted visibilities

$$V_{ij}' = (1 + \Delta g_i)(1 + \Delta g_j^*) V_{ij}. \qquad (95)$$

In our model we do not allow the gain to fluctuate in frequency, enforcing each antennas gain to be frequency independent. However, we do allow the gains to fluctuate in time, assuming that each 60 s sample has a separate uncorrelated gain residual. Over the two years of integration, the errors on each coadded sample are reduced by a factor of $\sqrt{N_{\mathrm{days}}}$.

We start with the base time stream to which we have added random gain fluctuations with $\sigma_g = 10\%$, 1% and 0.1% in each 60 s period. These time streams are then analyzed with the fiducial analysis products that assume no gain fluctuations. In Fig. 10 we show the power spectrum biases corresponding to each level of gain fluctuation. We have used the minimum variance estimator discussed in Sec. VIII, the results appear similar if we use the unwindowed estimator (albeit noisier). The bias, which is caused by foreground leakage from the imperfect calibration, is mostly located within the foreground wedge. This is in line with our expectation from [29,55] which indicate that leakage from imperfect foreground cleaning will be concentrated in this region. However, there are significant discrepancies from this picture that seem to be related to the array geometry (such as the line $k_\perp = 0.03\ h\ \mathrm{Mpc}^{-1}$) that may require more detailed study to understand intuitively [57].

We can see that the bias becomes negligible for error residuals of around 1%. Over the course of the two years observation this corresponds to a tolerance on gain fluctuations of $\sim 2 \times 10^{-4}$ for each synthetic beam ($\sim 1°$).
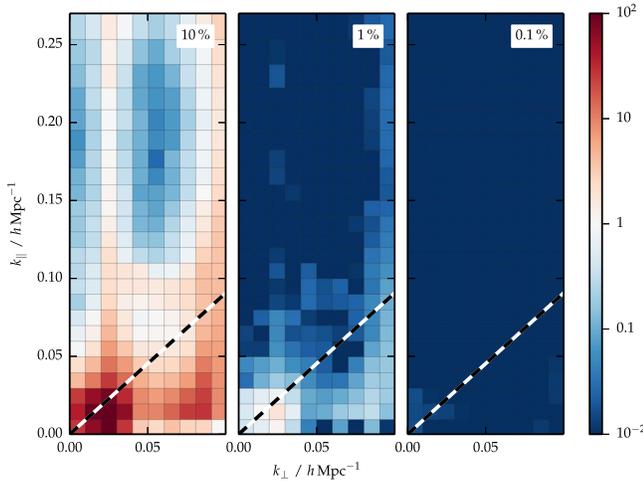
FIG. 10 (color online). Biasing of the power spectrum from complex gain perturbations with amplitude $\sigma_g = 10\%$, 1% and 0.1%, again for observations of 400–500 MHz. The bias is given as a fraction of the statistical error. Regions where this ratio is less than one (shown in blue) indicate where the systematic errors are subdominant compared to the statistical errors. Again we indicate the foreground wedge with a dashed line, however in this case we note that most of the bias lies within this region.

This required tolerance is significantly less than the 10^{-5} naively expected. This difference is due to the fact that we repeatedly measure the same sky because our array is highly redundant (with typical redundancies of ~ 30) allowing us to average down the effect of gain fluctuations, reducing the precision required on an individual baseline.

This level of precision should be achievable with techniques such as redundant baseline calibration [58]. Our analysis assumes that the residuals are Gaussian and independent in time, such that they quickly average down with repeated measurements. In practice there may be a component of the residuals from $1/f$ noise with large correlation times which make this assessment more difficult. We leave investigation of such effects for future studies.

B. Unknown primary beam

One of the key inputs to our analysis is an accurate model of each feeds primary beam. In particular we need the electric field response at each position on the sky, given by the quantity $A_a(\hat{n})$. Generally this quantity can only be determined by calibrating from observations of the sky (for instance by holography). As this process is challenging and time consuming, we would like to know how precise the calibration must be.

Here, we use the parametrization of the primary beam given in Sec. IV A. We use the fiducial model of the dipoles beam, $\theta_H = 2\pi/3$, $\theta_E = 0.7\theta_H$ (this is the same as the example used throughout). However, we will perturb the E-plane widths of each antenna around the fiducial model by an amount $\Delta\theta_E^i$. Increasing θ_E has the effect of making the primary beam of the X feed slightly narrower, and the Y feed longer (decreasing it does the opposite). It also reduces the difference in response between the X and Y feeds, reducing the expected amount of polarization leakage. This is demonstrated in Fig. 11 where we show the effect on the Stokes I and polarized response to changes in θ_E to the X and Y feeds. In particular we show the derivatives of $R_{I\to I}$ [Eq. (40)] and $R_{P\to I}$ [Eq. (41)] with respect to θ_E^X and θ_E^Y.

To calculate the changes to the data we need to propagate these primary beam changes through to the beam transfer matrices. At linear order in the $\Delta\theta_E^i$ the perturbed beam transfer functions are

$$B_{ij}^X = B_{ij}^X + \frac{dB_{ij}^X}{d\theta_E^i}\Delta\theta_E^i + \frac{dB_{ij}^X}{d\theta_E^j}\Delta\theta_E^j \qquad (96)$$

where the derivatives are related to the primary beam derivatives by

$$\frac{dB_{ij}^X}{d\theta_E^k} = -\frac{d\ln\Omega_{ij}}{d\theta_E^k}B_{ij}^X$$
$$+ \frac{2}{\Omega_{ij}}\left[\frac{dA_i^a}{d\theta_E^k}A_j^{b*} + A_i^a\frac{dA_j^{b*}}{d\theta_E^k}\right]\mathcal{P}_{ab}^X\, e^{2\pi i\hat{n}\cdot u_{ij}}. \qquad (97)$$
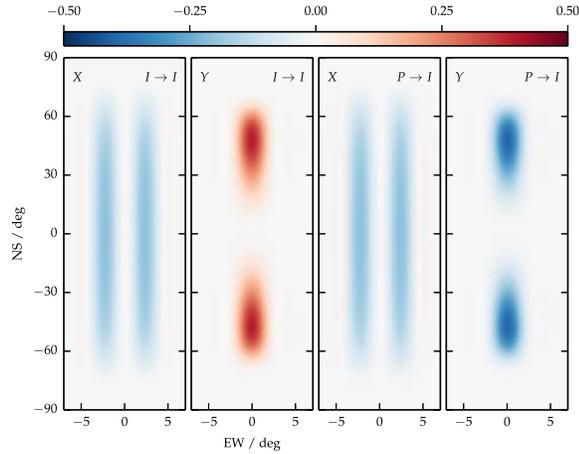
FIG. 11 (color online).   The response of the primary beam to fractional changes in the X and Y dipole E-plane widths. Similar to Fig. 2 we illustrate the transfer from the total intensity and polarized sky, into an instrumental Stokes I combination, however, here we show the derivative with respect to changes in the E-plane width of the X and Y feeds. The first two plots show the change of the total intensity response with changes in the E-plane of the X and Y dipoles; the second two plots show the changes in the polarization response, again corresponding to changes in the X and Y feeds. For instance a 1% change in each dipoles width changes each response by 1% of the corresponding plot (to first order).

The derivative of the composite beam solid angle is

$$\frac{d \ln \Omega_{ij}}{d\theta_E^i} = \frac{d \ln \Omega_{ji}^*}{d\theta_E^i} = \frac{1}{2\Omega_i} \int d^2\hat{n} \, \frac{dA_i^a}{d\theta_E^k} A_j^{b*} \mathcal{P}_{ab}^I. \quad (98)$$

By treating the primary beam derivatives $dA_i^a/d\theta_E^k$ as a modified beam, we can use Eq. (97) to calculate time streams for the beam perturbed beam transfers. We then use Eq. (96) to apply the effects of arbitrary combinations of perturbations to θ_E for each antenna.

We draw a set of Gaussian distributed values for the width of each feed, θ_E^i. We vary standard deviation (10%, 1% and 0.1% of the fiducial model) and use it to generate synthetic data with perturbed beam widths. We propagate the analysis of these corrupted time streams all the way through to the power spectrum, assuming the fiducial configuration. In Fig. 12 we show the results for the minimum variance estimator. Again we see that the bias is mostly concentrated in the foreground wedge region. The bias can be significant (compared to statistical errors) if our beam knowledge is imperfect, though it has mostly disappeared in the case where we know the beamwidth to 0.1%.

This analysis suggests that if the beamwidth were the only varying parameter, in the absence of other bias mitigating techniques, we would need to measure it to $\sim 10^{-3}$ accuracy. However, as the beam derivative is
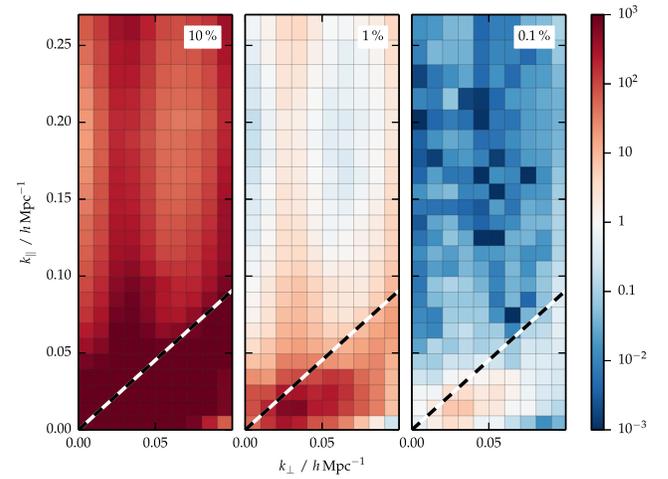
FIG. 12 (color online).   Power spectrum biasing for 10%, 1% and 0.1% shifts from the fiducial E-plane width. These biases are given in units of σ for each band, values greater than one indicate where this systematic error dominates the statistical error. These are the biases of the minimum variance estimator (so as to avoid issues with the power spectrum deconvolution). Here we can see that unknown fluctuations in the beamwidth of more than 0.1% give rise to significant power spectrum biases.

typically of order 0.1 (see Fig. 11), this can be seen as a precision of around 10^{-4} on the beam itself, similar to the gain fluctuations, and still a lower precision than our expectation of around 10^{-5}. We can attribute this to the fact that our power spectrum estimation is dependent on a complicated combination of all the primary beams, and this averages down the fluctuations in the same manner as we expect for the gain fluctuations.

Clearly a realistic description of the beam must contain much more than a simple beamwidth, but this indicates the accuracy to which we must strive to map the primary beam of each feed. This level of precision will be challenging, though not unprecedented, with similar accuracies achieved by holographic means [59].

XI. FULL BANDWIDTH FORECASTS

Experiments such as CHIME are targeted at measuring the evolution of dark energy over a large range of redshift. As an example application of this method we show in this section forecasts for the example cylinder telescope (similar in size to the CHIME Pathfinder but smaller than the full CHIME) across a full octave in bandwidth of 400–800 MHz, corresponding to a redshift range of $z \approx 0.8$–2.6. This is broken up into four 100 MHz subbands to illustrate the changes with frequency.

In Fig. 13 we show the power spectrum forecasts for each of the four 100 MHz subbands. This clearly illustrates the increase in sensitivity as we move to higher frequency, particularly at large $k_\perp$ where the increased angular resolution combines with the decreased observation distance to
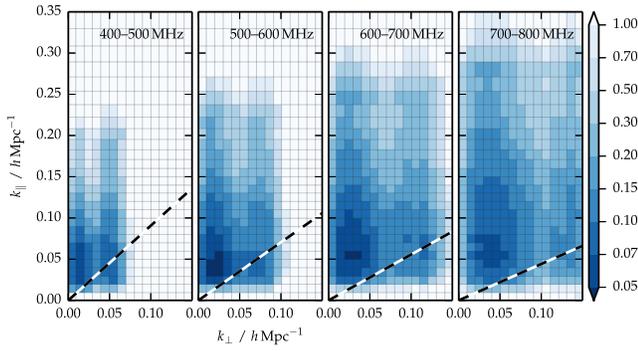
FIG. 13 (color online).   This figure shows the fractional error on each power spectrum bin, for different frequency bands between 400 and 800 MHz. This clearly illustrates the increasing angular resolution as we move to higher frequencies. Again, the red dashed lines indicate the location of the foreground wedge.

FIG. 14 (color online).   Constraints on the expansion history as a function of redshift, shown relative to a fiducial ΛCDM cosmology. The red line shows the mean expansion history predicted from the Planck constraints on w_0, w_a [60] (combined with Union 2 supernovae data [61]). The grey lines show a selection of expansion histories *randomly* drawn from the Planck + Union2 posterior distribution to illustrating the trajectories which are consistent with those data sets. The black errorbars show what could be achieved with a medium sized cylinder experiment which allows us to place constraints directly on $D_M(z)$ and $H_0(z)$. As we can see, the best discrimination comes at low redshift from the 600–800 MHz bands (assuming a w_0, w_a cosmology).

dramatically increase the spatial resolution. There is an additional boost at large $k_\parallel$ where the constant frequency corresponds to a decreasing line of sight distance. We can also see how the double peaked structure in sensitivity (discussed in Sec. IX) changes with frequency, with the peaks moving outwards and broadening as expected from the increasing resolution. However, the dropoff at small $k_\perp$ barely increases in size as it comes from the contribution of sample variance which does not change with the increased angular resolution (it does shift slightly because a fixed angular scale maps a smaller spatial scale at higher frequency).

The effect of foreground cleaning is similar across all bands, with it removing sensitivity for $k_\parallel < 0.02 \; h\,\mathrm{Mpc}^{-1}$. We do not expect the number of modes used to describe the foregrounds along a particular line of sight to vary significantly with the small shifts in frequencies between the bands, and this should translate into a similar loss of power spectrum sensitivity for each band.

To constrain the dark energy equation of state, we will use the measured power spectrum in each band to determine the apparent scale of the baryon acoustic oscillation as a function of redshift. The angular and line of sight scales respectively constrain the transverse comoving distance $D_M(z)$ and the Hubble parameter $H(z)$. These give two distinct probes of the expansion history as a function of redshift. In Fig. 14 we illustrate how measurements from our example telescope could be used to improve current constraints from Planck.

In Fig. 15 we show the predicted constraints on the dark energy equation of state in the w_0-w_a parametrization. We describe how these are derived from the power spectrum forecasts in Appendix E. This gives a figure of merit (FOM) [62] of 7 for the telescope and Planck, and 88 if we add in Stage II experiments. This is an improvement by around 70% from Planck and Stage II only (FOM of 53). If there were no loss in sensitivity due to foreground cleaning, the FOM increases to 21 and 135 respectively.

FIG. 15 (color online).   Constraints on the dark energy equation of state. We show the constraints for the example cylinder with Planck only (large, red), and with Planck and Stage II experiments (smaller, blue). The lighter and darker contours for each illustrate the 2σ and 1σ bounds respectively.

XII. CONCLUSION

In this paper we have improved and extended the *m*-mode formalism for analyzing observations from transit radio interferometers. In particular, we have extended the formalism to include a complete description of polarization (see Sec. II). This allows us to characterize observations of

the real polarized sky including the effects of instrumental polarization. Including these effects is crucial when making wide-field multifrequency observations with a polarization-dependent sky response. Furthermore, by considering the geometry of the measured data in the vector space of the observations, we have developed a simple SVD projection that not only yields a significant data compression, but also acts as an effective filter to suppress polarized foreground contamination (Sec. VI).

In the limit of statistically isotropic foregrounds, each m mode is independent of the others with no statistical coupling between them. Thus the m-mode formalism, because it allows each mode to be treated independently, allows for a compact and computationally efficient representation for statistics of our data. We have exploited this to develop the KL transform as a technique for the removal of astrophysical foregrounds, which otherwise appears to be extremely challenging using other methods (Sec. VII). We believe this is the first technique shown to be effective at the removal of polarized foregrounds to below the signal level while using a telescope model with realistic amounts of polarization leakage (see Fig. 8).

Within the m-mode formalism we have constructed an optimal quadratic estimator for the 21 cm power spectrum that is computationally efficient and takes into account the full statistics of the data, including the effects of the foreground cleaning (Sec. VIII). This has allowed us to forecast the performance of a medium sized cylinder transit telescope (Sec. IX)—similar in size to the CHIME Pathfinder telescope currently under construction. We show that the KL transform is able to clean foregrounds well into the foreground wedge, demonstrating that there is no fundamental information loss within the region, with foreground cleaning limiting our measurements only in a smaller band $k_\parallel \lesssim 0.02 \; h \, \mathrm{Mpc}^{-1}$. In fact, we find that even the removal of the polarized fore-grounds gives a minimal reduction in the expected ability to constrain the power spectrum.

While our results are encouraging, the m-mode formalism does make simplifying assumptions and the impact of these assumptions needs to be tested when analyzing real experiments. For instance, for the analysis to be tractable, we assume the statistics of the data are stationary under rotation of the Earth. This is expected of the 21 cm signal itself, but is not expected to be true of both for the foregrounds where the Galaxy is heavily anisotropic (though in our simulations this does not seem prevent us from suppressing foregrounds consistent with the actual structure of the Galaxy), and for the instrumental effects where $1/f$ noise, radio frequency interference, and thermal fluctuations make the behavior of the instrument time dependent. The m-mode formalism also assumes perfect knowledge of the telescope, including amplifier gains, and fully characterized beams. In Sec. X we have investigated how these uncertainties, if ignored, would lead to significant biases in the measured power spectrum, and placed limits on how well we must know these to faithfully recover the power spectrum. We find that random complex gain variations can have an amplitude of up to 1% (on one minute time scales), before they cause any significant power spectrum shifts. Similarly, using the beamwidth as a simple parametrization of our uncertainty we find that we must know the width of the primary beam of each feed to around 0.1% to avoid bias. These precisions are less stringent than naive expectations from the dynamic range between the signal and foregrounds (around 10^{-5}). Though challenging, requiring effort and innovation, they should be achievable.

One avenue to further loosen these calibration requirements is to follow the same philosophy we take with foreground removal and conservatively identify, and remove, the modes which are particularly susceptible to this miscalibration. Even in the case where we perturb the nominal beamwidth by an unknown number of order 10% there is a significant fraction of the KL modes that do not get biased appreciably. It is conceivable that through Monte Carlo modeling of beam uncertainties the highly corruptible KL modes could be found and excised prior to estimating the power spectrum (at the cost of increased error bars). Alternatively we could pursue a more targeted approach by incorporating these instrumental uncertainties into the noise model, and using the KL filter to remove them. We leave investigations of these and other bias mitigating techniques for future work.

When our feed spacing is larger than the Nyquist criterion at a particular wavelength (for a beam stretching to the horizon this is $> \lambda/2$), we cannot uniquely localize a source on the sky. This aliasing effect causes us to form multiple images when mapmaking and, while not leading to biases, gives a degradation in power spectrum errors. For the example cylinder telescope used here, this occurs at $\nu > 500$ MHz. While an investigation of this effect is beyond the scope of this paper we do not expect it is a fundamental limitation and believe that this degradation may be alleviated by moving away from a fully uniform feed spacing.

The pipeline we have developed for performing the m-mode analysis described in this paper is publicly available [63]. The tools created for modeling and simulating the radio sky are available at the same location.

_____

[1]See http://healpix.sourceforge.net/.

package [64]. Computations were performed on the GPC supercomputer at the SciNet HPC Consortium. SciNet is funded by the Canada Foundation for Innovation under the auspices of Compute Canada; the Government of Ontario; the Ontario Research Fund—Research Excellence; and the University of Toronto.

APPENDIX A: NOISE POWER SPECTRUM

The sensitivity of a radio receiver is a well studied problem [46,65,66]. For a single feed the power received in a frequency interval $\Delta\nu$ is simply related to the antenna temperature $P = g^2 k_B T_a \Delta\nu$ (in the absence of noise). In our notation the antenna temperature for a single feed is simply equal to its autocorrelation $T_a = V_{ii}$. However, we need to extend this to the case of the correlation of two separate antennas. Provided that the power $P \propto \langle F_i F_j^* \rangle$ for both the autocorrelation $i = j$ and cross-correlation $i \neq j$ cases, the signal observed is

$$P = g_i g_j^* k_B V_{ij} \Delta\nu, \qquad (A1)$$

where the real and imaginary parts of P contain the cosine and sinelike correlations. The same conclusion can be reached by following through the correlation of the induced voltage from each antenna using the effective length. With our normalization $l_{\text{eff}}^i = l_{\text{max}}^i A_i$, with l_{max} the maximum length anywhere on the sky.

Beyond the astrophysical signal there are other contributions to the observed power. This noise may come from many sources such as the ground or the atmosphere, or the receiver system itself. For the autocorrelation of a single feed the instantaneous noise power defines the system temperature

$$P = g^2 k_B T_{\text{sys}} \Delta\nu. \qquad (A2)$$

When consider the cross-correlation between different feeds, provided the noise at both is uncorrelated, there is no additional power observed in the mean of the signal. However, the noise does contribute to the fluctuations about the mean. If we average a frequency channel of width $\Delta\nu$ over a rectangular window of time length τ, we find the mean power observed is

$$\bar{P} = g_i g_j^* k_B \Delta\nu (V_{ij} + \delta_{ij} T_{\text{sys},i}). \qquad (A3)$$

The fluctuations in the amplitude have standard deviation

$$\sigma_P = g_i g_j^* k_B \Delta\nu \sqrt{\frac{T_{\text{sys},i} T_{\text{sys},j}}{\tau \Delta\nu}}. \qquad (A4)$$

See [65,66] for a detailed calculation. The fluctuations in the real and imaginary part have an equal amplitude of $\sigma_P / \sqrt{2}$. We have assumed we are in the limit where the system temperature dominates the antenna temperature, $T_{\text{sys}} \gg T_a$.

If the noise at each feed is independent, that means that the noise between different baseline pairs is uncorrelated. The variance that we would ascribe to the measurement of a particular visibility ij at a particular time (after the averaging) is

$$\sigma_{ij}^2 = \frac{T_{\text{sys},i}(\nu) T_{\text{sys},j}(\nu)}{\tau \Delta\nu}. \qquad (A5)$$

To calculate the m-mode power spectrum of fluctuations N_m we first calculate the noise correlation function. Assuming that it is white noise, and again using a rectangular window function the correlation function is

$$\zeta_{ij}(t) = \langle n_{ij}(t') n_{ij}^*(t' - t) \rangle = \sigma_{ij}^2 \text{tri}(t/\tau), \qquad (A6)$$

where the triangle function $\text{tri}(x) = 1 - |x|$ for $|x| < 1$. To calculate the noise power spectrum we simply Fourier transform this quantity. As we need to consider the problem in terms of the Earths rotation, we identify distinct sidereal days as independent measurements of the sky and treat the averaged noise as periodic. Similarly we can identify redundant baselines as independent measurements of the same quantity. Only the diagonal elements of the noise matrix corresponding to the same frequency and baseline are nonzero. The discrete power spectrum of the noise, defined by $\langle n_{ij}^m n_{ij}^{m'*} \rangle = N_{ij}^m \delta_{mm'}$, is

$$N_{ij}^m = \frac{T_{\text{sys},i}(\nu) T_{\text{sys},j}(\nu)}{N_{\text{day}} N_{\text{red}} t_{\text{sid}} \Delta\nu} \text{sinc}^2\left(\pi \frac{m\tau}{t_{\text{sid}}}\right), \qquad (A7)$$

where N_{day} is the number of sidereal days that have been observed. Usually we would want the integration length to be smaller than any angular scale we are interested in, in this limit $m\tau \ll t_{\text{sid}}$, and the sinc factor is ~ 1.

APPENDIX B: KARHUNEN-LOÈVE TRANSFORM

Let us write our measurement as a vector $\boldsymbol{x}$, where each dimension corresponds to a measured degree of freedom. We can write $\boldsymbol{x}$ as

$$\boldsymbol{x} = \boldsymbol{s} + \boldsymbol{n} \qquad (B1)$$

where $\boldsymbol{s}$ and $\boldsymbol{n}$ are respectively the signal we are interested in and some generalized form of noise (in the case of 21 cm this may include the foregrounds). These components have covariance matrices

$$\langle \boldsymbol{s}\boldsymbol{s}^\dagger \rangle = \mathbf{S}, \qquad \langle \boldsymbol{n}\boldsymbol{n}^\dagger \rangle = \mathbf{N}. \qquad (B2)$$

We are free to transform the measurement vector as we wish, $\boldsymbol{x}' = \mathbf{R}\boldsymbol{x}$, provided we are careful to update all the statistics we make use of. In our case we are interested

in the two-point statistics and so it is sufficient to transform the covariance matrix $\mathbf{X}' = \langle (\mathbf{R}x)(\mathbf{R}x)^\dagger \rangle = \mathbf{R}\mathbf{X}\mathbf{R}^\dagger$. The KL transform takes advantage of this to produce simultaneous eigenmodes of the signal and noise covariances.

We start by making the eigendecomposition of the noise matrix

$$\mathbf{N} = \mathbf{R}_1^\dagger \mathbf{N}' \mathbf{R}_1 \tag{B3}$$

where $\mathbf{R}_1$ is the unitary matrix of eigenvectors (stacked row by row), and $\mathbf{N}'$ is the diagonal matrix of eigenvalues. Using this we can transform the data vector $x' = \mathbf{R}_1 x$, which produces a new signal covariance

$$\mathbf{S}' = \langle s's'^\dagger \rangle = \langle (\mathbf{R}_1 s)(\mathbf{R}_1 s)^\dagger \rangle = \mathbf{R}_1 \mathbf{S}\mathbf{R}_1^\dagger \tag{B4}$$

and reduces the noise matrix to $\mathbf{N}'$. As the new noise matrix consists solely of positive diagonal elements $(\mathbf{N}')_{ii} = \lambda_i^N$, a further transformation $x'' = \mathbf{R}_2 x'$, where $\mathbf{R}_2 = \mathbf{N}'^{-\frac{1}{2}}$, reduces the noise matrix to the identity $\mathbf{N}'' = \mathbf{I}$. The signal matrix is transformed to

$$\mathbf{S}'' = \mathbf{R}_2 \mathbf{R}_1 \mathbf{S}\mathbf{R}_1^\dagger \mathbf{R}_2^\dagger. \tag{B5}$$

Applying any unitary transformation to the data will leave the noise covariance as the identity. We use this freedom to diagonalize the signal covariance by eigendecomposition $\mathbf{S}'' = \mathbf{R}_3^\dagger \Lambda \mathbf{R}_3$, leaving the total transformation on the data as

$$x \to \tilde{x} = \mathbf{R}_3 \mathbf{R}_2 \mathbf{R}_1 x. \tag{B6}$$

Overall this has changed the covariance matrices to

$$\mathbf{S} \to \Lambda, \tag{B7}$$

$$\mathbf{N} \to \mathbf{I}. \tag{B8}$$

By making this transformation we have simultaneously diagonalized the correlations of both the signal and the noise, mapping the latter to the identity matrix. In particular, the elements of Λ give the signal to noise ratio of each mode. With no hidden correlations this basis allows us to cleanly filter data by simply throwing away modes with signal to noise ratio below some threshold. This is equivalent to zeroing the corresponding elements of $\tilde{x}$.

Rather than explicitly constructing the three transformations, it is mathematically equivalent to finding the solutions to the generalized eigenvalue problem

$$\mathbf{S}x = \lambda \mathbf{N}x, \tag{B9}$$

with the eigenvectors forming the transformation matrix, and the eigenvalues giving the elements of the signal covariance Λ. This approach is simpler and computationally more efficient.

APPENDIX C: STATISTICAL MODELS

As discussed in Sec. VII to use the Karhunen-Loève transform to perform foreground cleaning we require models of the two-point statistics of both the 21 cm signal and the foreground contaminants. For computational efficiency these models must be isotropic and so we only need to specify the angular power spectrum

$$C_l^{XY}(\nu, \nu') = \langle a_{lm}^X(\nu) a_{lm}^{Y*}(\nu') \rangle, \tag{C1}$$

for all the pairs of the four polarization components $X, Y \in \{T, E, B, V\}$.

1. Astrophysical foregrounds

Our foreground models are based on [48]. However we only include the dominant two components, the Galactic synchrotron emission and extragalactic point sources. In both cases the angular power spectrum is of the form

$$C_l(\nu, \nu') = A \left(\frac{l}{100} \right)^{-\alpha} \left(\frac{\nu\nu'}{\nu_0^2} \right)^{-\beta} e^{-\frac{1}{2\xi_l^2} \ln^2(\nu/\nu')}. \tag{C2}$$

The original models were calibrated for forecasting observations of the Epoch of Reionization. In [32] we recalibrated them for the high frequency, all of the sky observations we are concerned with in this paper. However for this work we also need to specify the correlations of the polarized parts of the foregrounds. We assume that the dominant source of polarized emission is our own Galaxy (ignoring the polarization of point sources) and model the polarized emission as being a statistical fraction f_{pol} of the unpolarized emission

$$C_l^{EE}(\nu, \nu') = C_l^{BB}(\nu, \nu') = f_{\mathrm{pol}}^2 C_l^{TT}(\nu, \nu'). \tag{C3}$$

In addition we assume that the polarized emission is uncorrelated such that $C_l^{TE} = C_l^{TB} = C_l^{EB} = 0$, and that there is no circular polarization from the Galaxy $C_l^{VV} = 0$. Our fiducial polarization fraction is $f_{\mathrm{pol}} = 0.5$. We list the parameters for these models in Table II.

2. 21 cm signal

On large scales the 21 cm brightness temperature is a biased tracer of the matter density field [67] with a power spectrum P_{T_b} given by

$$P_{T_b}(\boldsymbol{k}; z, z') = \bar{T}_b(z)\bar{T}_b(z')(b + f\mu^2)^2 P_m(k; z, z') \tag{C4}$$

where b is the bias and $P_m(k; z, z') = P(k)D_+(z)D_+(z')$ is the real-space matter power spectrum. The evolution of the perturbations is given by the growth factor $D_+(z)$

TABLE II. Parameters for our foreground power spectrum model given in Eq. (C2). These are based on the models of [48], adapted to the intensity mapping regime in [32].

Component	Polarization	A (K^2)	α	β	ζ
Galaxy	TT	6.6×10^{-3}	2.80	2.8	4.0
	EE, BB	1.65×10^{-3}	2.80	2.8	4.0
Point sources	TT	3.55×10^{-4}	2.10	1.1	1.0

normalized such that $D_+(0) = 1$, with the growth rate $f = d \ln D_+ / d \ln a$ (that is the logarithmic derivative of the growth factor D_+). The mean brightness temperature is assumed to take the form

$$\bar{T}_b(z) = 0.1 \left(\frac{\Omega_{\mathrm{HI}}}{0.33 \times 10^{-4}} \right) \left(\frac{\Omega_m + (1+z)^{-3}\Omega_\Lambda}{0.29} \right)^{-1/2}$$
$$\times \left(\frac{1+z}{2.5} \right)^{1/2} \mathrm{mK} \qquad \text{(C5)}$$

given in [1]. In [68] they determine the degenerate product $\Omega_{\mathrm{HI}}b = 0.62 \times 10^{-3}$, which we use in this work. As the redshift distortions break the $\Omega_{\mathrm{HI}}b$ degeneracy we fix $b = 1$.

For use in our foreground filter, we require the angular power spectrum of the 21 cm brightness temperature [69,70]. This can be calculated from the real-space power spectrum equation (C4), but is computationally difficult, generally requiring double integration over highly oscillatory functions for each ν, ν' pair. To speed this up we use the flat-sky approximation from [70]:

$$C_l(z, z') = \frac{1}{\pi \chi \chi'} \int_0^\infty dk_\parallel \cos (k_\parallel \Delta \chi) P_{T_b}(\boldsymbol{k}; z, z') \qquad \text{(C6)}$$

where χ and χ' are the comoving distances to redshift z and z' and their difference is denoted by $\Delta \chi = \chi - \chi'$. The wave vector $\boldsymbol{k}$ has components $k_\parallel$ and $l/\bar{\chi}$ in the directions parallel and perpendicular to the line of sight ($\bar{\chi}$ is the mean of χ and χ'). This approximation is accurate to the 1% level for $l > 10$ [70].

We use this method not only for calculating the signal covariance function, but also the band functions required for the power spectrum. To determine each $\mathbf{C}_a$ we simply apply Eqs. (C4) and (C6), with $P_m(\boldsymbol{k}) = P_a(\boldsymbol{k})$.

APPENDIX D: SIMULATING ALL-SKY RADIO EMISSION

Testing of the m-mode formalism and the foreground cleaning with the Karhunen-Loève transform requires the use of synthetic sky maps. For it to be realistic these simulated maps must capture the essential properties of the 21 cm signal and foreground components. In this section we briefly describe how these simulations are generated.

1. 21 cm signal

Assuming the cosmological 21 cm emission is Gaussian on the scales of interest, the angular power spectrum given in the previous section [Eq. (C6)] completely specifies its fluctuations. Maps of the sky can be generated by drawing Gaussian realizations of the power spectrum, using Cholesky decomposition to produce the correct frequency correlation structure, and then adding in the mean temperature given by Eq. (C5).

2. Extragalactic point sources

We construct our point source simulations from three components: a population of real bright point sources ($S > 10$ Jy at 151 MHz); a synthetic population of dimmer sources down to 0.1 Jy at 151 MHz; and an unresolved background of dimmer sources ($S < 0.1$ Jy) modeled as a Gaussian random field. This last component dramatically reduces the number of sources we must directly generate.

The unresolved background is generated by drawing a Gaussian realization from the point source model detailed in Table II. The random source catalog is constructed by drawing from the point source distribution of [71] and scattering the sources randomly over the sky. The intrinsic polarization of each point source is determined by

$$Q(\nu) + iU(\nu) = pI(\nu) \qquad \text{(D1)}$$

where the polarization fraction p is a complex Gaussian random variable with standard deviation σ_p. This standard deviation is equal to the average polarization fraction of sources in the catalog, we set $\sigma_p = 5\%$.

The population of real bright point sources is generated by matching the Very Large Array Low-frequency Sky Survey (VLSS) at 74 MHz [72] against the National Radio Astronomy Observatory VLA Sky Survey (NVSS) at 1.4 GHz [73]. We only include sources interpolated to be brighter than 10 Jy at 151 MHz. Each source is assigned the polarization as measured by NVSS, and is extrapolated to other frequencies. In this work we have also assumed that the six sources above 100 Jy (at 600 MHz) have been removed from the time stream to high accuracy.

The polarization of an extragalactic source is Faraday rotated as it passes through the magnetized interstellar medium in our Galaxy, generating oscillatory frequency structure in the polarization. To apply this, we use the Faraday depth map of [74] to rotate the polarization angle of our background sources.

3. Galactic synchrotron intensity

In this work we continue to use the prescription developed in a previous paper [32] to generate constrained simulations of the total intensity of synchrotron emission from our Galaxy. These maps are formed from two distinct components:

a. A large scale base map produced by extrapolating the Haslam map[2] with a spectral index map from [75].

b. A randomly generated map that adds in fluctuations in frequency and on small angular scales. This is constrained to be zero on the scales constrained by the Haslam map, and is designed to smoothly extrapolate the angular fluctuations of the Haslam map to smaller scales, and reproduce the anisotropic fluctuations on small scale power across the sky [76].

The procedure for generating these two components is described in detail in [32], with the only change being the spectral index map used.

4. Galactic synchrotron polarization

To test our foreground removal and analysis we need to be able to create simulated multifrequency maps of our Galaxy, and in particular its polarization structure. As the observed radiation has been omitted across a range of Faraday depths, unlike extragalactic sources, this is challenging. One approach is to make use of the increasingly sophisticated models of the Galactic magnetic field structure [77], and electron distribution [78], to create realistic large scale simulations of the polarization structure [79]. However, we instead appeal to the ideas of Faraday rotation measure synthesis [80] to rapidly create simulations that capture the important effects.

Rotation measure synthesis attempts to link the wavelength dependent polarization rotation to the structure along the line of sight. Polarized radiation emitted at a distance r from us is Faraday rotated by an amount $\phi\lambda^2$ before it reaches us, with the Faraday depth

$$\phi(\mathbf{r}) = \int_0^r n_e(\mathbf{r}')\mathbf{B}(\mathbf{r}') \cdot d\mathbf{r}'. \tag{D2}$$

The key idea in Faraday rotation measure synthesis is to not directly probe the physical structure of emission, but to probe the structure as a function of Faraday depth. In this case we can just think of the observed polarized emission in a given direction $P(\hat{\mathbf{n}}, \lambda^2)$ as being the summation of the emission at all Faraday depths $F(\hat{\mathbf{n}}, \phi, \lambda^2)$, rotated by the correct wavelength dependent amount

$$P(\hat{\mathbf{n}}, \lambda^2) = \int F(\hat{\mathbf{n}}, \phi, \lambda^2) e^{2i\phi\lambda^2} d\phi. \tag{D3}$$

In [80] the idea was to use multiwavelength observations to invert this Fourier relation, and constrain the structure of $F(\hat{\mathbf{n}}, \phi, \lambda^2)$. However, we will attempt to use well motivated assumptions about the emission in Faraday space to construct simulations of polarized skies. As in [80] we

presume that the Faraday space emission $F(\hat{\mathbf{n}}, \phi, \lambda^2)$ is separable in its spectral dependence, such that $F(\hat{\mathbf{n}}, \phi, \lambda^2) = f(\hat{\mathbf{n}}, \phi)s(\hat{\mathbf{n}}, \lambda^2)$. This flattens the spectrum of f so that we can still use the Faraday synthesis formalism for it. We take the spectral function $s(\hat{\mathbf{n}}, \lambda^2)$ from the unpolarized emission.

We start with a simple model of the emission from the Galaxy, assuming that along any line of sight the emission comes from many independent synchrotron regions, each of fixed brightness ΔT. In a direction with total brightness temperature T, there are $N = T/\Delta T$ such regions. We assume that the emitting regions are scattered across a range in Faraday depth. With no reason to favor positive or negative Faraday depths, we assume this distribution is zero mean. From observations of extragalactic point sources we know the Faraday depth to the edge of our Galaxy [74] and this gives us a measure of the range of Faraday depths within the Galaxy. Combining these properties the distribution is modeled as a zero-mean Gaussian with a width $\sigma_\phi(\hat{\mathbf{n}})$ which is determined from the Faraday rotation data. We determine $\sigma_\phi(\hat{\mathbf{n}})$ by taking the Faraday depth map of [74], taking its absolute value, and smoothing with a FWHM of 10°. Each of these regions has a small width in ϕ over which its polarization is coherent. We call this coherence length ξ_ϕ, and note that it determines the size of structures in Faraday space.

To determine the polarization structure we start by calculating the number of emitting regions within a range ϕ to $\phi + \Delta\phi$. This is given by

$$\Delta N = \frac{N}{(2\pi\sigma_\phi^2)^{1/2}} e^{-\frac{1}{2}(\frac{\phi}{\sigma_\phi})^2} \Delta\phi. \tag{D4}$$

As each region is independent we assume they all have a randomly distributed complex polarization, drawn from a Gaussian distribution with variance $(\alpha_p\Delta T)^2$, where α_p is the polarization fraction. Within this range in Faraday depth the polarizations add up like a random walk, giving the expected total root-mean-square polarization as

$$\alpha_p\Delta T\Delta N^{1/2} = (8\pi)^{1/4}\alpha_p\Delta_T \left[\frac{1}{(4\pi\sigma_\phi^2)^{1/2}} e^{-\frac{1}{4}(\frac{\phi}{\sigma_\phi})^2}\right] \left(\frac{N\sigma_\phi}{\Delta\phi}\right)^{\frac{1}{2}} \Delta\phi. \tag{D5}$$

This gives the expected magnitude of the emission at each position in Faraday space, showing that even in Faraday space we see depolarization because of the incoherent combination of multiple Faraday sources at a single depth.

This suggests we model the emission as two factors:

$$f(\hat{\mathbf{n}}, \phi) = w(\hat{\mathbf{n}}, \phi)c(\hat{\mathbf{n}}, \phi). \tag{D6}$$

The first $w(\hat{\mathbf{n}}, \phi)$ is a positive envelope function which defines the region, and amplitude of emission in Faraday depth.

$$w(\hat{n}, \phi) \propto \frac{A}{\sqrt{4\pi\sigma_\phi^2}} e^{-\frac{1}{4}\left(\frac{\phi}{\sigma_\phi}\right)^2} (T\sigma_\phi)^{1/2} \qquad \text{(D7)}$$

The second $c(\hat{n}, \phi)$ is a random field that gives fluctuations in the complex polarization as a function of Faraday depth, this should be highly correlated on scales $\Delta\phi \ll \xi_\phi$, and uncorrelated on scales $\Delta\phi \gg \xi_\phi$. We model this as a Gaussian random field drawn with an angular power spectrum

$$C_l(\phi, \phi') \propto \left(\frac{l}{100}\right)^{-\alpha} \exp\left(-\frac{(\phi - \phi')^2}{2\zeta^2}\right). \qquad \text{(D8)}$$

The angular dependence is chose to match that of the total intensity model equation (C2).

The normalization of these functions is degenerate with the value of ΔT. We fix the combination by considering what happens at high frequency observations where Faraday rotation is much less important. In this limit the polarization fraction is determined by the incoherent addition of the polarization of the emitting regions and is $\sim\alpha_p(\Delta T/T)^{1/2}$. We choose $\alpha_p = 2/3$ which is the intrinsic polarization of synchrotron with a spectral index of the electron energy distribution $\gamma = 5/3$. Polarized maps from the WMAP satellite at 23 GHz [81] indicate that the Galaxy is 20% polarized at high latitudes, we use this fact to determine the overall normalization.

The only remaining degree of freedom is the correlation length of the emitting regions in Faraday space, ξ_ϕ. The size of an emitting region in Faraday space will grow towards the Galactic center because of the increased magnetic field strengths. We construct a crude model

$$\xi_\phi = \min(\sigma_\phi/20, 3 \, \text{rad m}^{-2}), \qquad \text{(D9)}$$

chosen to visually reproduce the amount of depolarization seen in 1.4 GHz polarization maps [82–84].

This gives all the necessary ingredients to draw a realization of $f(\hat{n}, \phi)$ which we can Fourier transform and scale by the spectral function $s(\hat{n}, \lambda^2)$ to produce the polarized emission using Eq. (D3). All these operations are performed on a regular grid in λ^2, which is extended beyond the desired frequency range to negate edge effects. The resulting series of maps are then interpolated onto the required frequency slices.

In Fig. 16 we show the polarization fraction and the frequency correlation length derived from a simulation between 400–600 MHz. Though the model we have constructed here is crude, and based on unrealistic assumptions about the Galactic emission it exhibits the properties we would expect from the real Galactic emission:

a. Emission is from a range of Faraday depths, rather than a single screen.

Polarization Fraction

Correlation length
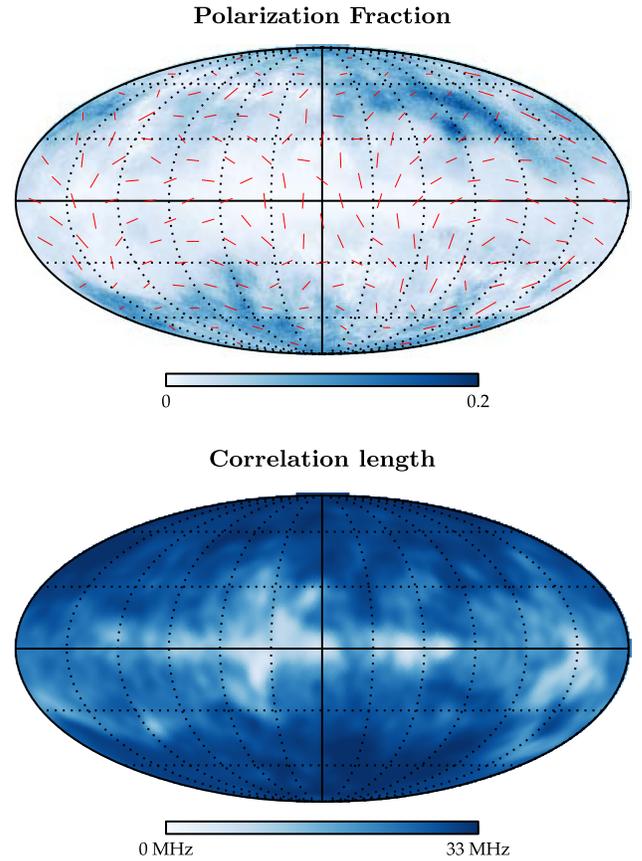
FIG. 16 (color online). The top panel shows the polarization direction and fraction of the 600 MHz slice of a polarized simulation of the Galaxy. This clearly demonstrates the effect of Faraday depolarization towards the Galactic center. The lower panel shows the effective correlation length as measured across the sky, smoothed on 10° scales.

b. In the Galactic plane there is substantial depolarization, but at high latitudes the polarization fraction is around that of 23 GHz.

c. Frequency decorrelation on lengths that we would predict from the Faraday rotation over the Galaxy, going to near zero in the Galactic center where the emission goes up to large Faraday depths.

APPENDIX E: DISTANCE MEASUREMENTS

By extracting the BAO signal from the 21 cm power spectrum, measurements of the Hubble rate $H(z)$ and transverse comoving distance $D_M(z)$ can be made. The BAO manifests itself as a preferred separation in the two-point correlation function at distances $s_\perp$ perpendicular to the line of sight and $s_\parallel$ parallel to the line of sight. The fractional errors on $s_\perp$ and $s_\parallel$ are equivalent to the fractional errors on the combinations s/D_M and sH, respectively, where s is the comoving sound horizon at the drag epoch. Thus, if s is well known (for example from observations of the cosmic microwave background) then

measurements of $s_\perp$ and $s_\parallel$ put observational constraints on D_M and H.

To project uncertainties in the power spectrum onto D_M and H, we first transform the Fisher matrix for the power spectrum $\mathbf{F}$ into the Fisher matrix $\mathbf{F}_s$ for the parameters $\theta_s = (\ln s_\perp^{-1}, \ln s_\parallel)$ via the Jacobian $(J_s)_{ij} = \partial P(\mathbf{k}_i)/\partial(\theta_s)_j$, where the Fisher matrices are related by $\mathbf{F}_s = \mathbf{J}_s^T \mathbf{F} \mathbf{J}_s$. $\mathbf{J}_s$ is evaluated using a fiducial cosmological model.

To calculate $\mathbf{J}_s$, we follow [85], where the effect of the baryons on the power spectrum is modeled by an additive term P_b to the otherwise "wiggles free" power spectrum that is approximated as

$$P_b(\mathbf{k}) = \sqrt{8\pi^2} A_0 P_{0.2} \text{sinc}(x)$$
$$\times \exp\left[-(k/k_s)^{1.4} - (k\Sigma_{nl})^2/2\right] \quad (E1)$$

where $x = \sqrt{(k_\perp s_\perp)^2 + (k_\parallel s_\parallel)^2}$, k_s is the Silk scale and $P_{0.2}$ is the linear power spectrum evaluated at $k = 0.2\,h\,\text{Mpc}^{-1}$. In this expression, A_0 is a normalization constant, taken to be $A_0 = 0.5817$. Σ_{nl} is the nonlinear dampening scale given by

$$\Sigma_{nl}^2 = (1 - \mu^2)\Sigma_\perp^2 + \mu^2 \Sigma_\parallel^2 \quad (E2)$$

where $\Sigma_\parallel = \Sigma_\perp(1 + f)$, $\Sigma_\perp = \Sigma_0(G(z)/G(0))$, G is the growth function and f is the growth rate. We follow [86] and assume that we may partially reconstruct parts of the

BAO signal degraded by nonlinear effects for modes with high signal to noise and set the effective nonlinear dampening scale to be $\Sigma_0 = 4.70(\sigma_8/0.9)\,h^{-1}\,\text{Mpc}$. By differentiating Eq. (E1) with respect to the variables θ_s, one can form $\mathbf{J}_s$ and subsequently evaluate $\mathbf{F}_s$. Note that $\mathbf{F}_s$ is equivalent to the Fisher matrix for the variables $\theta_d = (\ln(D_M(z)/s), \ln(sH(z)))$.

The Fisher matrix $\mathbf{F}_s$ can be transformed again into the Fisher matrix $\mathbf{F}_{DE}$ for the cosmological parameters $\theta_{DE} = (w_0, w_a, \Omega_\Lambda, \Omega_k, \omega_m, \omega_b)$ by use of the Jacobian $(\mathbf{J}_{DE})_{ij} = \partial(\theta_d)_i/\partial(\theta_{DE})_j$, which as before is evaluated using a fiducial cosmological model. In θ_{DE}, the equation of state w has been parametrized as

$$w(z) = w_0 + w_a \frac{z}{1+z}. \quad (E3)$$

Note that the ω_b dependence in $\mathbf{F}_{DE}$ comes from the comoving sound horizon s, present in both terms of θ_d, which is dependent on the baryon to photon ratio $R_b = 3\rho_b/4\rho_\gamma$. The Fisher matrix for the dark energy parameters is then formed as $\mathbf{F}_{DE} = \mathbf{J}_{DE}^T \mathbf{F}_s \mathbf{J}_{DE}$. Constraint contours in the w_0-w_a plane can be found by marginalizing over the other variables in θ_{DE}, which in this case amounts to inverting $\mathbf{F}_{DE}$ to get the covariance matrix, removing the rows and columns corresponding to the marginalized variables, and inverting once more to recover the marginalized Fisher matrix for (w_0, w_a).

[1] T.-C. Chang, U.-L. Pen, J. B. Peterson, and P. McDonald, Phys. Rev. Lett. **100**, 091303 (2008).

[2] A. Loeb and J. S. B. Wyithe, Phys. Rev. Lett. **100**, 161301 (2008).

[3] S. R. Furlanetto, S. P. Oh, and F. H. Briggs, Phys. Rep. **433**, 181 (2006).

[4] A. Loeb and M. Zaldarriaga, Phys. Rev. Lett. **92**, 211301 (2004).

[5] G. Swarup, S. Ananthakrishnan, V. K. Kapahi, A. P. Rao, C. R. Subrahmanya, and V. K. Kulkarni, Curr. Sci. **60**, 95 (1991).

[6] J. C. Pober, A. Liu, J. S. Dillon, J. E. Aguirre, J. D. Bowman, R. F. Bradley, C. L. Carilli, D. R. DeBoer, J. N. Hewitt, D. C. Jacobs, M. McQuinn, M. F. Morales, A. R. Parsons, M. Tegmark, and D. J. Werthimer, Astrophys J **782**, 66 (2014).

[7] M. P. van Haarlem, M. W. Wise, A. W. Gunst, G. Heald, J. P. McKean, J. W. T. Hessels, A. G. de Bruyn, R. Nijboer, J. Swinbank, R. Fallows *et al.*, Astron. Astrophys. **556**, A2 (2013).

[8] C. J. Lonsdale *et al.*, Proc. IEEE **97**, 1497 (2009).

[9] H. Zheng *et al.*, arXiv:1309.2639.

[10] A. R. Parsons, D. C. Backer, G. S. Foster, M. C. H. Wright, R. F. Bradley, N. E. Gugliucci, C. R. Parashare, E. E. Benoit, J. E. Aguirre, D. C. Jacobs, C. L. Carilli, D. Herne, M. J. Lynch, J. R. Manley, and D. J. Werthimer, Astron. J. **139**, 1468 (2010).

[11] J. C. Pober, A. R. Parsons, D. R. DeBoer, P. McDonald, M. McQuinn, J. E. Aguirre, Z. Ali, R. F. Bradley, T.-C. Chang, and M. F. Morales, Astron. J. **145**, 65 (2013).

[12] R. Ansari, J.-E. Campagne, P. Colom, C. Magneville, J.-M. Martin, M. Moniez, J. Rich, and C. Yèche, C. R. Phys. **13**, 46 (2012).

[13] R. A. Battye, I. W. A. Browne, C. Dickinson, G. Heron, B. Maffei, and A. Pourtsidou, Mon. Not. R. Astron. Soc. **434**, 1239 (2013).

[14] Canadian Hydrogen Intensity Mapping Experiment (CHIME), 2013.

[15] G. W. Kant, P. D. Patel, S. J. Wijnholds, M. Ruiter, and E. van der Wal, IEEE Trans. Antennas Propag. **59**, 1990 (2011).

[16] S. Saiyad Ali and S. Bharadwaj, J. Astrophys. Astron. **35**, 157 (2014).

[17] Parkes, CSIRO Parkes Observatory, http://www.parkes.atnf.csiro.au/.

[18] X. Chen, Int. J. Mod. Phys. Conf. Ser. **12**, 256 (2012).
[19] C. Blake and K. Glazebrook, Astrophys. J. **594**, 665 (2003).
[20] W. Hu and Z. Haiman, Phys. Rev. D **68**, 063004 (2003).
[21] H.-J. Seo and D. J. Eisenstein, Astrophys. J. **598**, 720 (2003).
[22] C. Blake et al., Mon. Not. R. Astron. Soc. **418**, 1707 (2011).
[23] L. Anderson et al., Mon. Not. R. Astron. Soc. **427**, 3435 (2012).
[24] S. T. Myers, C. R. Contaldi, J. R. Bond, U.-L. Pen, D. Pogosyan, S. Prunet, J. L. Sievers, B. S. Mason, T. J. Pearson, A. C. S. Readhead, and M. C. Shepherd, Astrophys. J. **591**, 575 (2003).
[25] M. Tegmark and M. Zaldarriaga, Phys. Rev. D **79**, 083530 (2009).
[26] A. R. Parsons and D. C. Backer, Astron. J. **138**, 219 (2009).
[27] A. Liu, M. Tegmark, S. Morrison, A. Lutomirski, and M. Zaldarriaga, Mon. Not. R. Astron. Soc. **408**, 1029 (2010).
[28] A. Liu and M. Tegmark, Phys. Rev. D **83**, 103006 (2011).
[29] A. R. Parsons, J. C. Pober, J. E. Aguirre, C. L. Carilli, D. C. Jacobs, and D. F. Moore, Astrophys. J. **756**, 165 (2012).
[30] R. Ansari, J. E. Campagne, P. Colom, J. M. Le Goff, C. Magneville, J. M. Martin, M. Moniez, J. Rich, and C. Yèche, Astron. Astrophys. **540**, A129 (2012).
[31] J. S. Dillon, A. Liu, and M. Tegmark, Phys. Rev. D **87**, 043005 (2013).
[32] J. R. Shaw, K. Sigurdson, U.-L. Pen, A. Stebbins, and M. Sitwell, Astrophys. J. **781**, 57 (2014).
[33] A. Liu and M. Tegmark, Mon. Not. R. Astron. Soc. **419**, 3491 (2012).
[34] A. Liu, M. Tegmark, and M. Zaldarriaga, Mon. Not. R. Astron. Soc. **394**, 1575 (2009).
[35] D. F. Moore, J. E. Aguirre, A. R. Parsons, D. C. Jacobs, and J. C. Pober, Astrophys. J. **769**, 154 (2013).
[36] J. Kim, Mon. Not. R. Astron. Soc. **375**, 625 (2007).
[37] J. D. McEwen and A. M. M. Scaife, Mon. Not. R. Astron. Soc. **389**, 1163 (2008).
[38] M. A. Holdaway, in Synthesis Imaging in Radio Astronomy II, Astronomical Society of the Pacific Conference Series Vol. 180, edited by G. B. Taylor, C. L. Carilli, and R. A. Perley (Astronomical Society of the Pacific, San Francisco, 1999), p. 401.
[39] S. Bhatnagar, T. J. Cornwell, K. Golap, and J. M. Uson, Astron. Astrophys. **487**, 419 (2008).
[40] M. Zaldarriaga and U. Seljak, Phys. Rev. D **55**, 1830 (1997).
[41] B. Y. Mills, Proceedings of the Astronomical Society of Australia **4**, 156 (1981).
[42] G. Swarup, N. V. G. Sarma, M. N. Joshi, V. K. Kapahi, D. S. Bagri, S. H. Damle, S. Ananthakrishnan, V. Balasubramanian, S. S. Bhave, and R. P. Sinha, Nature (London) **230**, 185 (1971).
[43] J. B. Peterson, K. Bandura, and U. L. Pen, arXiv:astro-ph/0606104.
[44] C. Craeye, in Proceedings of the IEEE Antennas and Propagation Society International Symposium, Washington, DC, 2005, Vol. 2B (IEEE, New York, 2005), p. 449.
[45] C. Craeye, in Proceedings of the IEEE Antennas and Propagation Society International Symposium, San Diego, 2008 (IEEE, New York, 2008), p. 1.
[46] C. A. Balanis, Antenna Theory: Analysis and Design, 3rd ed. (Wiley, Hoboken, NJ, 2005).
[47] D. Varshalovich, A. Moskalev, and V. Khersonskiĭ, Quantum Theory of Angular Momentum: Irreducible Tensors, Spherical Harmonics, Vector Coupling Coefficients, 3nj Symbols (World Scientific, Singapore, 1988).
[48] M. G. Santos, A. Cooray, and L. Knox, Astrophys. J. **625**, 575 (2005).
[49] J. R. Bond, Phys. Rev. Lett. **74**, 4369 (1995).
[50] M. Tegmark, A. N. Taylor, and A. F. Heavens, Astrophys. J. **480**, 22 (1997).
[51] E. F. Bunn, M. Zaldarriaga, M. Tegmark, and A. de Oliveira-Costa, Phys. Rev. D **67**, 023501 (2003).
[52] M. Tegmark, Phys. Rev. D **55**, 5895 (1997).
[53] J. R. Bond, A. H. Jaffe, and L. Knox, Phys. Rev. D **57**, 2117 (1998).
[54] N. Padmanabhan, U. Seljak, and U. L. Pen, New Astron. **8**, 581 (2003).
[55] M. F. Morales, B. Hazelton, I. Sullivan, and A. Beardsley, Astrophys. J. **752**, 137 (2012).
[56] A. Datta, J. D. Bowman, and C. L. Carilli, Astrophys. J. **724**, 526 (2010).
[57] B. J. Hazelton, M. F. Morales, and I. S. Sullivan, Astrophys. J. **770**, 156 (2013).
[58] A. Liu, M. Tegmark, S. Morrison, A. Lutomirski, and M. Zaldarriaga, Mon. Not. R. Astron. Soc. **408**, 1029 (2010).
[59] G. R. Harp et al., IEEE Trans. Antennas Propag. **59**, 2004 (2011).
[60] P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday et al. (Planck Collaboration), Astron. Astrophys. **571**, A16 (2014).
[61] N. Suzuki et al., Astrophys. J. **746**, 85 (2012).
[62] A. Albrecht, G. Bernstein, R. Cahn, W. L. Freedman, J. Hewitt, W. Hu, J. Huth, M. Kamionkowski, E. W. Kolb, L. Knox, J. C. Mather, S. Staggs, and N. B. Suntzeff, arXiv:astro-ph/0609591.
[63] See http://github.com/radiocosmology.
[64] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, Astrophys. J. **622**, 759 (2005).
[65] P. C. Crane and P. J. Napier, in Synthesis Imaging in Radio Astronomy, Astronomical Society of the Pacific Conference Series Vol. 6, edited by R. A. Perley, F. R. Schwab, and A. H. Bridle (Astronomical Society of the Pacific, San Francisco, 1989), p. 139.
[66] J. M. Wrobel and R. C. Walker, in Synthesis Imaging in Radio Astronomy II, Astronomical Society of the Pacific Conference Series Vol. 180, edited by G. B. Taylor, C. L. Carilli, and R. A. Perley (Astronomical Society of the Pacific, San Francisco, 1999), p. 171.
[67] K. W. Masui, E. R. Switzer, N. Banavar, K. Bandura, C. Blake, L.-M. Calin, T.-C. Chang, X. Chen, Y.-C. Li, Y.-W. Liao, A. Natarajan, U.-L. Pen, J. B. Peterson, J. R. Shaw, and T. C. Voytek, Astrophys. J. **763**, L20 (2013).
[68] E. R. Switzer, K. W. Masui, K. Bandura, L.-M. Calin, T.-C. Chang, X.-L. Chen, Y.-C. Li, Y.-W. Liao, A. Natarajan, U.-L. Pen, J. B. Peterson, J. R. Shaw, and T. C. Voytek, Mon. Not. R. Astron. Soc. **434**, L46 (2013).

[69] A. Lewis and A. Challinor, Phys. Rev. D **76**, 083005 (2007).

[70] K. K. Datta, T. R. Choudhury, and S. Bharadwaj, Mon. Not. R. Astron. Soc. **378**, 119 (2007).

[71] T. Di Matteo, R. Perna, T. Abel, and M. J. Rees, Astrophys. J. **564**, 576 (2002).

[72] A. S. Cohen, W. M. Lane, W. D. Cotton, N. E. Kassim, T. J. W. Lazio, R. A. Perley, J. J. Condon, and W. C. Erickson, Astron. J. **134**, 1245 (2007).

[73] J. J. Condon, W. D. Cotton, E. W. Greisen, Q. F. Yin, R. A. Perley, G. B. Taylor, and J. J. Broderick, Astron. J. **115**, 1693 (1998).

[74] N. Oppermann *et al.*, Astron. Astrophys. **542**, A93 (2012).

[75] M.-A. Miville-Deschênes, N. Ysard, A. Lavabre, N. Ponthieu, J. F. Macías-Pérez, J. Aumont, and J. P. Bernard, Astron. Astrophys. **490**, 1093 (2008).

[76] L. La Porta, C. Burigana, W. Reich, and P. Reich, Astron. Astrophys. **479**, 641 (2008).

[77] R. Jansson and G. R. Farrar, Astrophys. J. Lett. **761**, L11 (2012).

[78] J. M. Cordes and T. J. W. Lazio, arXiv:astro-ph/0207156.

[79] A. Waelkens, T. Jaffe, M. Reinecke, F. S. Kitaura, and T. A. Enßlin, Astron. Astrophys. **495**, 697 (2009).

[80] M. A. Brentjens and A. G. de Bruyn, Astron. Astrophys. **441**, 1217 (2005).

[81] A. Kogut, J. Dunkley, C. L. Bennett, O. Doré, B. Gold, M. Halpern, G. Hinshaw, N. Jarosik, E. Komatsu, M. R. Nolta, N. Odegard, L. Page, D. N. Spergel, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright, Astrophys. J. **665**, 355 (2007).

[82] M. Wolleben, T. L. Landecker, W. Reich, and R. Wielebinski, Astron. Astrophys. **448**, 411 (2006).

[83] J. C. Testori, P. Reich, and W. Reich, Astron. Astrophys. **484**, 733 (2008).

[84] W. Reich and P. Reich, in *Proceedings of the International Astronomical Union Symposium, Puerto Santiago, Tenerife, Spain, 2008*, Vol. 259, edited by K. G. Strassmeier, A. G. Kosovichev, and J. E. Beckman (International Astronomical Union, Paris, 2009), p. 603.

[85] H.-J. Seo and D. J. Eisenstein, Astrophys. J. **665**, 14 (2007).

[86] H.-J. Seo, S. Dodelson, J. Marriner, D. Mcginnis, A. Stebbins, C. Stoughton, and A. Vallinotto, Astrophys. J. **721**, 164 (2010).