Epoch of reionization window. I. Mathematical formalism

Adrian Liu[*]

Department of Astronomy, UC Berkeley, Berkeley, California 94720, USA
and Berkeley Center for Cosmological Physics, UC Berkeley, Berkeley, California 94720, USA

Aaron R. Parsons

Department of Astronomy, UC Berkeley, Berkeley, California 94720, USA
and Radio Astronomy Laboratory, UC Berkeley, Berkeley, California 94720, USA

Cathryn M. Trott

International Centre for Radio Astronomy Research, Curtin University, Bentley, Western Australia, Australia
and ARC Centre of Excellence for All-Sky Astrophysics (CAASTRO), Curtin University,
Bentley, Western Australia 6009, Australia

The 21 cm line provides a powerful probe of astrophysics and cosmology at high redshifts, but unlocking the potential of this probe requires the robust mitigation of foreground contaminants that are typically several orders of magnitude brighter than the cosmological signal. Recent simulations and observations have shown that the smooth spectral structure of foregrounds combines with instrument chromaticity to contaminate a "wedge"-shaped region in cylindrical Fourier space. While previous efforts have explored the suppression of foregrounds within this wedge, as well as the avoidance of this highly contaminated region, all such efforts have neglected a rigorous examination of the error statistics associated with the wedge. Using a quadratic estimator formalism applied to the interferometric measurement equation, we provide a framework for such a rigorous analysis (incorporating a fully covariant treatment of errors). Additionally, we find that there are strong error correlations at high spatial wave numbers that have so far been neglected in sensitivity derivations. These error correlations substantially degrade the sensitivity of arrays relying on contributions from long baselines, compared to what one would estimate assuming uncorrelated errors.

I. INTRODUCTION

Modern cosmological observations have produced exquisite constraints on both the initial and final conditions of structure formation in our Universe. Initial conditions have now been probed to high significance with a large number of cosmic microwave background experiments [1,2], while at low redshifts, a combination of galaxy surveys and traditional astronomical measurements provide the final conditions [3]. Still missing from these direct observations, however, are the intermediate epochs that bridge the gap between early and late times. For example, despite tremendous recent progress in high-redshift galaxy observations, details regarding the formation of the first luminous objects and their effects on the intergalactic medium (IGM) during the epoch of reionization (EoR) remain uncertain.

In the next few years, direct observations of the EoR will be made possible by measurements of the redshifted 21 cm hyperfine transition of neutral hydrogen (see, e.g.,

Refs. [4–7] for reviews). At the relevant redshifts, the intensity field of the 21cm brightness temperature depends on a rich variety of different astrophysical effects, such as fluctuations in the ionization and spin states of the IGM, as well as cosmological quantities such as the underlying dark matter density field and peculiar velocity gradients. A map of the 21 cm intensity field at redshifts $z \sim 6$ and above would therefore be a rich probe of EoR physics, including the nature of the first luminous sources (such as their typical mass and luminosity scales), their ionizing and heating efficiency, and feedback processes on the IGM, among other effects. Such a mapping can be accomplished in three dimensions, since the spectral nature of a 21 cm measurement provides redshift (and therefore line-of-sight distance) information, while the angular directions are mapped using traditional imaging. Because of this, the 21 cm line allows access to a large fraction of our Universe's comoving volume, potentially allowing futuristic measurements to move beyond astrophysics and into the measurement of fundamental cosmological parameters [8–10].

There are currently a number of experiments aimed at mapping the fluctuations of the cosmological 21 cm

[*]acliu@berkeley.edu

signal, including the Giant Metrewave Radio Telescope Epoch of Reionization experiment (GMRT-EoR [11]), the Low Frequency Array (LOFAR [12]), the Murchison Widefield Array (MWA [13]), and the Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER [14]). These interferometer arrays have yet to make a positive detection of the cosmological signal, with the primary challenges being foreground contamination and the high sensitivity requirements. To increase sensitivity, these experiments are primarily targeting binned, statistical measures of the brightness temperature field such as the power spectrum. Recent progress has resulted in a number of increasingly stringent upper limits [11,15,16], and proposed next-generation instruments such as the Hydrogen Epoch of Reionization Array [17] and the Square Kilometer Array [18] promise to yield extremely high significance measurements.

In addition to achieving the required sensitivity, observations targeting the redshifted 21 cm line must also contend with foreground contaminants. In the relevant frequency ranges (roughly $\sim$100 to 200 MHz, corresponding to $z \sim 13$ to 6), there exist a large number of non-cosmological sources of radio emission that contaminate measurements. These include sources such as the diffuse synchrotron radiation from our own Galaxy, as well as extragalactic point sources, whether they are bright and resolved or part of a dim and unresolved background. The brightness temperature of foregrounds is expected to be 10^5 times greater than theoretical expectations for the amplitude of the cosmological signal. A detection of the reionization power spectrum will therefore be challenging without a robust foreground mitigation strategy.

Historically, cosmic microwave background (CMB) experiments have had to deal with similar problems of foreground contamination. However, strategies for foreground cleaning that have been developed for the CMB cannot be directly applied to 21 cm cosmology for two reasons. First, CMB experiments typically operate at higher frequencies, where foregrounds are not as bright. In fact, microwave-frequency foregrounds are subdominant to the CMB away from the Galactic plane. In addition, CMB experiments measure anisotropies over a two-dimensional surface, with different observation frequencies providing consistency checks and a set of redundant measurements that can be used for foreground isolation. The three-dimensional mapping of the 21 cm line, on the other hand, contains unique cosmological information at every frequency, which makes it more difficult to remove foregrounds in a way that does not result in the loss of cosmological signal [19].

With CMB techniques unlikely to succeed without modification, a number of alternate foreground mitigation strategies have been suggested for 21 cm cosmology. These include spectral polynomial fitting [20–23], Wiener filtering [24], principal component analyses [11,25–27],

nonparametric subtractions [28–30], Fourier-mode or delay filtering [31,32], frequency stacking [33], Karhunen-Loéve eigenmode projection [34,35], and inverse covariance weighting [15,36,37]. The vast majority of these approaches rely on the fact that foreground sources are expected to be spectrally smooth, while the cosmological EoR signal is expected to fluctuate rapidly with frequency [38]. The cosmological signal can therefore be extracted by isolating spectrally smooth components from the data.

Recently, however, a complication to this simple picture was realized, in what has been colloquially termed the "foreground wedge." Consider a cylindrically binned power spectrum measurement, i.e., one where Fourier amplitudes are squared and binned in annuli specified by wave numbers perpendicular to the line of sight $k_\perp$ and wave numbers parallel to the line of sight $k_\parallel$. Because the line-of-sight direction is equivalent to the spectral axis of an interferometer, one might have naively expected smooth spectrum foregrounds to be sequestered to only the lowest $k_\parallel$. However, this neglects the fact that interferometers are inherently chromatic instruments, with a given baseline probing finer spatial scales (higher $k_\perp$) at higher frequencies. This coupling of spectral and spatial information is sometimes coined mode mixing and results in the leakage of information from low to high $k_\parallel$. This effect is particularly pronounced at high $k_\perp$, where the modes are typically probed by longer baselines, which are more chromatic. Putting everything together, theoretical studies and simulations [32,39–44] have shown that foregrounds are expected to leak out of the lowest $k_\parallel$ into a characteristic wedge that is schematically shown in Fig. 1. Observations with PAPER and MWA have confirmed this basic picture [45], including its evolution with frequency [15].

The foreground wedge is both a blessing and a curse. At the sensitivity levels that have been achieved by current experiments, observations have seen a sharp dropoff in foregrounds beyond the wedge [45]. Theoretical calculations and simulations have shown that such a dropoff is the natural consequence of geometric limitations [32], provided the foregrounds are spectrally smooth. If further integration reveals low-level foregrounds that are spectrally unsmooth, their influence will leak beyond what is typically labeled as the edge of the wedge. However, if foregrounds continue to be reasonably smooth, the fact that physical considerations limit the extent of wedge implies that there must exist an "EoR window": a region in Fourier space that is *a priori* expected to be foreground-free. The existence of the EoR window thus enables a relatively robust foreground avoidance strategy, where a detection of the power spectrum can be made simply by avoiding measurements within the wedge. On the other hand, such a conservative approach forces one to work at higher k than if the chromatic effects had not caused the wedge in the first place. This is unfortunate because the ratio of the cosmological signal to instrumental noise
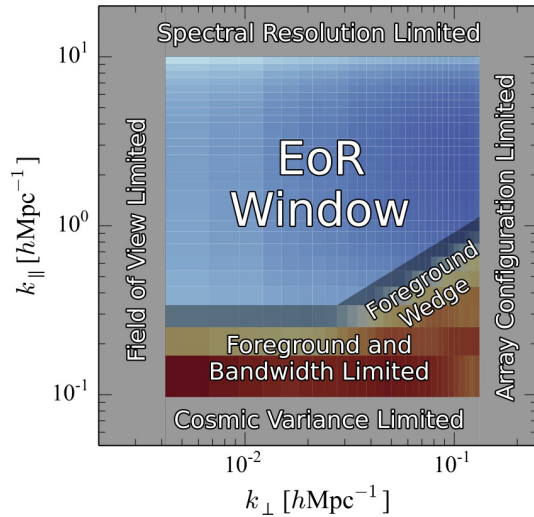
FIG. 1 (color online). A schematic of the EoR window in the cylindrical $k_\perp k_\parallel$ Fourier plane. At the lowest $k_\perp$, errors increase because of limits on an instrument's field of view. High $k_\perp$ modes are probed by the longest baselines of an interferometer array, and the sensitivity drops to zero beyond $k_\perp$ scales corresponding to these baselines. Spectral resolution limits the sensitivity at large $k_\parallel$. The lowest $k_\parallel$ are in principle limited by cosmic variance, but in practice the larger concern is limited bandwidth and the foreground contamination, which intrinsically resides at low $k_\parallel$. As one moves towards higher $k_\perp$, however, the foregrounds leak out to higher $k_\parallel$ in a characteristic shape known as the "foreground wedge." The remaining parts of the Fourier plane are thermal-noise dominated, allowing (with a large collecting area or a long integration time) a clean measurement of the power spectrum in this "EoR window."

typically peaks at low k, which means that if it were possible to work within the wedge (or to at least push back its boundaries a little), one would be able to make higher significance detections of the power spectrum. Indeed, in Ref. [17] it was suggested that working within the wedge can increase the detection significance anywhere from a factor of 2 to 6 (depending on the interferometer's configuration), with corresponding decreases in the error bars on astrophysical parameters.

Given the potentially high payoff associated with pushing back the influence of the wedge (or equivalently, enlarging the EoR window), it is important to have a statistically rigorous framework for describing the wedge. In this paper, we provide just such a mathematical framework. Since there already exists an extensive literature on the foreground wedge and the EoR window, it is worth summarizing the ways in which this paper builds upon and extends previous results. Works such as Refs. [39–41] describe instrumental simulations that take one particular realization of foregrounds and propagate them through a power spectrum estimation pipeline. They therefore only probe the *mean* power spectrum, and not the scatter (i.e., the errors) about this mean. In Ref. [43], a statistical treatment of point source populations was

considered. While error bars were computed, off-diagonal error correlations (i.e., covariances) in the final measurements were neglected. Potential error correlations are important particularly because Ref. [43] considered the application of various tapering functions to their Fourier transforms, and certain choices can result in significant correlations between Fourier bins. We consider full foreground covariances, a full treatment of instrumental effects (such as having a non-top-hat beam), a full treatment of data analysis choices such as tapering functions, and a fully covariant propagation of errors. We build upon Ref. [42], which made use of Monte Carlo methods to propagate errors. Our treatment is more analytic, allowing us to capture the large dynamic range needed to accurately compute the error statistics in a measurement where the foregrounds are many orders of magnitude brighter than the signal. This is made computationally feasible by our use of the delay spectrum approach introduced in Ref. [32], where input frequency spectra are sorted into a set of time-delay τ modes via a per-baseline Fourier transform. However, unlike in some works where the delays are used as an *approximation* for line-of-sight Fourier modes (an assumption that is only valid for short baselines, as we will discuss in Sec. II), we use delays strictly as a convenient choice of *basis*. This basis makes it computationally possible for us to deal directly with visibilities in our formalism (bypassing any mapmaking steps), which avoids gridding artifacts in our numerical results. We also take into full account the correlations between partially overlapping baselines and therefore rigorously treat the possible complications that were highlighted in Ref. [44]. A related treatment pertaining to lower-redshift 21 cm intensity mapping experiments (though focusing less on the details of the wedge) can be found in Ref. [35]. While our fiducial calculations are centered around instruments targeting the EoR, the techniques developed in this paper are equally applicable to cosmological 21 cm at lower (or higher) redshifts.

We accomplish our goals by making use of the quadratic estimator formalism, which was adapted for 21 cm power spectrum measurements in Refs. [36,37] and applied to real data in Ref. [15]. However, appropriate "wedge effects" were not incorporated into the formalism, an omission that we rectify in this paper. Placing everything in the quadratic estimator formalism enables a systematic computation of the aforementioned error statistics, as well as a systematic study of the optimality (or lack thereof) of various power spectrum estimators. In the following paper (Ref. [46], henceforth "paper II"), we will take advantage of this to examine the extent to which statistical methods can enlarge the EoR window.

With our fully covariant treatment, we find that the wedge is not simply a region of large foreground errors and biases, but also as a marker for error correlations: at $k_\perp$ values where the wedge is a dominant effect, the errors tend

TABLE I.   Dictionary of scalars and functions. The "context" column gives equation references, typically either their defining equation or their first appearance in the text.

Quantity	Meaning/definition	Context
Basic quantities		
$\mathbf{u}$	Fourier dual to angular direction $\boldsymbol{\theta}$	Eq. (4)
η	Fourier dual to ν (i.e., spectral wave number)	Eq. (4)
τ	Delay, i.e., Fourier dual to ν (or spectral wave number) for a single baseline	Eq. (3)
$V(\mathbf{b}, \nu)$	Visibility measured by baseline $\mathbf{b}$ at frequency ν	Eq. (2)
$\tilde{V}(\mathbf{b}, \tau)$	Delay-space visibility by baseline $\mathbf{b}$ at delay τ	Eq. (3)
Instrumental parameters		
$\mathbf{b}$	Baseline vector	Eq. (1)
$\boldsymbol{\theta}_0$	Characteristic width of primary beam	Eq. (1)
A	Primary beam function	Eq. (1)
$\tilde{A}$	Spatial Fourier transform of primary beam function	Eq. (5b)
$\tilde{A}_{b\parallel}$	Profile of $\tilde{A}$ parallel to baseline vector direction	Eq. (7)
$\tilde{A}_{b\perp}$	Profile of $\tilde{A}$ perpendicular to baseline vector direction	Eq. (7)
B_{chan}	Frequency channel width	Eq. (2)
γ	Frequency channel profile	Eq. (2)
$\tilde{\gamma}$	Fourier transform of frequency channel profile γ	Eq. (8)
B_{band}	Bandwidth corresponding to depth of cosmological volume	Eq. (3)
Ω_{pp}	Integrated beam squared area	Eq. (50)
T_{sys}	System temperature	Eq. (49)
t	Total integration time	Eq. (49)
$n(b)$	Number baselines	Eq. (49)
Sky		
$I(\boldsymbol{\theta}, \nu)$	Sky brightness temperature at angle $\boldsymbol{\theta}$ and frequency ν	Eq. (1)
$\tilde{I}(\mathbf{u}, \eta)$	Fourier transform of the sky temperature at angular wave number $\mathbf{u}$ and line-of-sight wave number η	Eq. (4)
$P(u, \eta)$	Cylindrically binned power spectrum at angular wave number u and line-of-sight wave number η	Eq. (A3)
C_ℓ^X	Angular power spectrum of foregrounds,	Eq. (46)
ν_c^X	Frequency coherence length of foregrounds (for both C_ℓ^X and ν_c^X, $X = $ diff for diffuse Galactic emission and $X = $ ps for point sources)	Eq. (46)
Data analysis		
ϕ	Bandpass or tapering function	Eq. (3)
$\tilde{\phi}$	Fourier transform of bandpass or tapering function ϕ	Eq. (9)
$h(\mathbf{u}, \eta; \mathbf{b}, \tau)$	Visibility response at delay τ of baseline $\mathbf{b}$ to sky mode on spatial scale $\mathbf{u}$ and spectral scale η	Eq. (17)
$g(u, \eta; b, \tau)$	Same as $h(\mathbf{u}, \eta; \mathbf{b}, \tau)$ but integrated over direction on uv plane perpendicular to baseline vector	Eq. (30)

to be strongly correlated. With strongly correlated errors, the number of independently measurable Fourier modes is reduced, suggesting that previous sensitivity estimates (such as those in Refs. [17,47,48]) may be overly optimistic, particularly for arrays that make use of long baselines (such as LOFAR or GMRT). In fact, the rewards for working within the wedge may be overrated as a result of this, but of course this cannot be quantified without a rigorous way to compute the error statistics of the wedge—hence the present paper.

The rest of this work is organized as follows. In Sec. II we examine the measurement equation of an interferometer in detail, paying special attention to chromatic effects. This provides a first noncovariant preview of the foreground wedge, which we generalize to an approximate, but fully covariant description in Sec. IV, following a review of the quadratic estimator formalism in Sec. III. In Sec. V we

discard the approximations made in Sec. IV in a full numerical implementation of our formalism. We summarize our conclusions in Sec. VII. Because a large number of mathematical quantities are defined in this paper, we provide dictionaries in Tables I and II for the reader's convenience.

II. A NONCOVARIANT PREVIEW OF THE FOREGROUND WEDGE

In this section, we will derive the Fourier-space foreground wedge from first principles. We begin with the visibility V measured by a baseline $\mathbf{b}$ at frequency ν:

$$V(\mathbf{b}, \nu) = \int I(\boldsymbol{\theta}, \nu) A\left(\frac{\boldsymbol{\theta}}{\theta_0}\right) \exp\left(-i2\pi \frac{\nu}{c} \mathbf{b} \cdot \boldsymbol{\theta}\right) d^2\boldsymbol{\theta}, \quad (1)$$

TABLE II.   Dictionary of vectors and matrices. The quantities shown here are grouped into three categories: those that exist in the vector space of the visibility measurements (indexed, for example, by baseline and delay), those that exist in the vector space of band powers (indexed by Fourier wave numbers), and those that bridge the two spaces. In the column giving the length and dimensions, N_{b1} denotes the number of baselines, N_ν the number of frequency channels, and N_{bands} the number of bins in Fourier space (i.e., the number of band powers).

Quantity	Components	Meaning/definition	Length/dimensions	Context
Quantities in measurement space				
x	$\mathbf{x}_i$	Serialized data vector of visibilities	$N_{b1}N_\nu$	Eq. (14)
C	$\mathbf{C}_{ij}$	Total covariance matrix $\mathbf{C} \equiv \langle \mathbf{x}\mathbf{x}^\dagger \rangle$	$N_{b1}N_\nu \times N_{b1}N_\nu$	Eq. (15)
N	$\mathbf{N}_{ij}$	Noise covariance matrix	$N_{b1}N_\nu \times N_{b1}N_\nu$	Eq. (15)
S	$\mathbf{S}_{ij}$	Signal covariance matrix	$N_{b1}N_\nu \times N_{b1}N_\nu$	Eq. (16)
Quantities in power spectrum space				
p	p_α	Serialized vector of true band powers (i.e., power spectrum at various grid points in Fourier space)	N_{bands}	Eq. (19)
p̂	$\hat{p}_\alpha$	Estimator for band powers derived from measurements	N_{bands}	Eq. (22)
Σ	$\Sigma_{\alpha\beta}$	Error covariance of estimated band powers, i.e., $\Sigma_{\alpha\beta} \equiv \langle \hat{p}_\alpha \hat{p}_\beta \rangle - \langle \hat{p}_\alpha \rangle \langle \hat{p}_\beta \rangle$	$N_{bands} \times N_{bands}$	Eq. (23)
Σ̄	$\bar{\Sigma}_{\alpha\beta}$	Error correlation of estimated band powers	$N_{bands} \times N_{bands}$	Eq. (52)
W	$\mathbf{W}_{\alpha\beta}$	Window function matrix	$N_{bands} \times N_{bands}$	Eq. (24)
b	$\mathbf{b}_\alpha$	Power spectrum estimator bias	N_{bands}	Eq. (26)
M	M_α	Power spectrum estimator normalization	N_{bands}	Eq. (28)
Hybrid quantities				
$\mathbf{E}_\alpha$	$(\mathbf{E}_\alpha)_{ij}$	Estimator matrix for quadratic estimator of of band power p_α, i.e., $\hat{p}_\alpha = \mathbf{x}\mathbf{E}_\alpha\mathbf{x}$	$N_{b1}N_\nu \times N_{b1}N_\nu$ for each $\alpha = 1$ to N_{bands}	Eq. (22)
$\mathbf{C}_{,\alpha}$	$(\mathbf{C}_{,\alpha})_{ij}$	Response of total covariance to the αth band power, i.e., $\mathbf{C}_{,\alpha} \equiv \partial\mathbf{C}/\partial p_\alpha$	$N_{bl}N_\nu \times N_{bl}N_\nu$ for each $\alpha = 1$ to N_{bands}	Eq. (20)

where $\boldsymbol{\theta}$ is the angular sky position, $I(\boldsymbol{\theta}, \nu)$ is the sky temperature and $A(\boldsymbol{\theta}/\theta_0)$ is the primary beam,[1] with θ_0 denoting its characteristic width. For notational simplicity in this section, we will omit the instrumental noise contribution to the visibility, but of course it is always implicitly present. Our (arbitrary) convention for A is that it is dimensionless and is normalized so that $A(0) = 1$. In what follows, we will see that the foreground wedge arises from the fact that the product of ν and $\boldsymbol{\theta}$ appears in the complex exponential. Fourier transforms in $\boldsymbol{\theta}$ are therefore coupled to ν and vice versa, leading to the "mode-mixing" phenomena coined in Ref. [41] and ultimately the wedge.

---

[1]In general, the primary beam will depend on frequency, although for some instruments (such as PAPER) the antennas are intentionally designed to minimize the frequency dependence of the beam [14]. In this paper, we will neglect the frequency dependence, because our goal is not to provide results pertaining to a particular instrument, but instead to provide a rigorous understanding of how foregrounds enter an interferometric power spectrum measurement. Including a frequency-dependent beam makes many of our analytic manipulations more difficult, which obscures the key physical effects that give rise to the foreground wedge. We note, however, that our general strategy of incorporating an interferometer's measurement equation into the quadratic estimator formalism is one that is capable of including frequency-dependent beams, albeit at a slightly greater computational cost.

To mimic the discreteness of frequency channels in a real instrument, we introduce a function γ that describes the response of a single frequency channel. Our measurement equation therefore becomes

$$V(\mathbf{b}, \nu) = \int \frac{d\nu'}{B_{chan}} d^2\boldsymbol{\theta} I(\boldsymbol{\theta}, \nu') A\left(\frac{\boldsymbol{\theta}}{\theta_0}\right)$$
$$\times \gamma\left(\frac{\nu - \nu'}{B_{chan}}\right) e^{-i2\pi\frac{\nu'}{c}\mathbf{b}\cdot\boldsymbol{\theta}}, \qquad (2)$$

with B_{chan} denoting the width of a frequency channel, and γ normalized such that $\int_{-\infty}^{\infty} \gamma(x) dx = 1$.

So far, we have no choice in the matter, in that the results of Eq. (2) are handed to us by the instrument. Moving onto data analysis, however, there is considerable freedom as to how one proceeds. For example, we may choose to move into delay space, which is accomplished by taking the Fourier transform in frequency of the spectrum measured by a single baseline (a "delay transform"):

$$\tilde{V}(\mathbf{b}, \tau) = \int V(\mathbf{b}, \nu) \phi\left(\frac{\nu - \nu_0}{B_{band}}\right) e^{-i2\pi\nu\tau} d\nu, \qquad (3)$$

where τ is the delay (with units of time), B_{band} is the bandwidth over which we wish to compute a power spectrum, and ν_0 is the central frequency of our band.
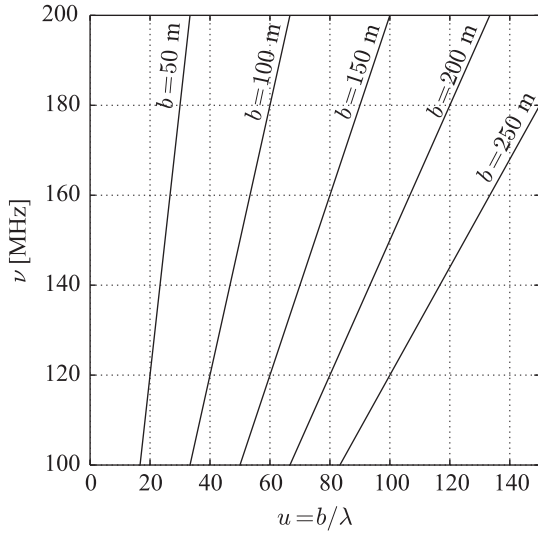
FIG. 2. Angular Fourier coordinate u probed by a variety of baseline lengths, plotted as a function of observing frequency. Taking a Fourier transform of the frequency spectrum of data from a single baseline essentially amounts to taking a Fourier transform along a solid line in this figure. This is a good approximation to taking a Fourier transform along the "true" frequency axis for short baselines.

The function ϕ is normalized so that $\phi(0) = 1$ and captures both the bandpass of our instrument and any tapering that one may wish to impose near the band edges. (In what follows, we will therefore use the terms "tapering function" and "bandpass" interchangeably to describe ϕ.) Precisely what form the edge tapering takes is a data analysis choice, and as shown in Ref. [43], choosing a good tapering function minimizes leakage of foregrounds in the Fourier plane. Later on, we will extend the work of Ref. [43] to self-consistently incorporate the effect that a tapering function has on error covariances. Aside from the peculiarities of certain tapering functions, working in delay space is simply a change of basis and, as emphasized in Ref. [32], represents no loss of generality. In later sections, we will find that delay space is a particular efficient basis to work in, one that makes many of the large matrices required for power spectrum estimation sparse (see Appendix C for details).

At this point our data are in a basis specified by baseline vector $\mathbf{b}$ and delay mode τ. This basis closely approximates the Fourier basis that the power spectrum inhabits, but the correspondence is not exact and in general must be accounted for. Concretely, suppose we let $\mathbf{u}$ be the Fourier dual of $\boldsymbol{\theta}$ and η be the Fourier dual of ν. Since the transverse comoving distance can be used to convert angles on the sky to transverse comoving distance, we have $\mathbf{u} \propto \mathbf{k}_{\perp}$ (the exact conversion is given in Appendix A, where we define our Fourier conventions). Similarly, the observed frequency of a spectral line can be mapped to a comoving line-of-sight distance, so $\eta \propto k_{\parallel}$. At a particular

frequency, a single baseline $\mathbf{b}$ of an interferometer roughly measures the spatial Fourier mode specified by wave number $\mathbf{u} = \mathbf{b}/\lambda$ [this can be seen by applying Parseval's theorem to Eq. (1) and assuming that the primary beam is wide]. Across different frequencies, however, we see that a single baseline probes different spatial Fourier modes. A Fourier transform of a single baseline's spectrum [Eq. (3)] is thus a Fourier transform along one of the solid lines in Fig. 2, where we show the chromaticity of various baselines. As pointed out in Ref. [32], with short baselines it is an excellent approximation to say that $\eta \sim \tau$. In this paper, we do not make this "delay approximation," as we consider baselines of all lengths.[2]

To properly account for the mapping between the "true" Fourier coordinates $(\mathbf{u}, \eta)$ to our visibilities expressed in the $(\mathbf{b}, \tau)$ basis, we can express the sky temperature $I(\boldsymbol{\theta}, \nu)$ in terms of its true Fourier transform, $\tilde{I}(\mathbf{u}, \eta)$:

$$I(\boldsymbol{\theta}, \nu) = \int \tilde{I}(\mathbf{u}, \eta) e^{i2\pi(\mathbf{u}\cdot\boldsymbol{\theta}+\eta\nu)} d\eta d^2u. \tag{4}$$

Inserting Eqs. (2) and (4) into Eq. (3) gives

$$\tilde{V}(\mathbf{b}, \tau) = \int d^2u \, d\eta \tilde{I}(\mathbf{u}, \eta) \int d^2\theta A\left(\frac{\boldsymbol{\theta}}{\theta_0}\right)$$
$$\times \int d\nu \frac{d\nu'}{B_{\text{chan}}} e^{-i2\pi\frac{\nu'}{c}\mathbf{b}\cdot\boldsymbol{\theta}} \gamma\left(\frac{\nu - \nu'}{B_{\text{chan}}}\right)$$
$$\times \phi\left(\frac{\nu - \nu_0}{B_{\text{band}}}\right) e^{-i2\pi\nu\tau} e^{i2\pi(\mathbf{u}\cdot\boldsymbol{\theta}+\eta\nu')} \tag{5a}$$

$$= \int d^2u \, d\eta \tilde{I}(\mathbf{u}, \eta)\theta_0^2 \int d\nu \frac{d\nu'}{B_{\text{chan}}}$$
$$\times \tilde{A}\left[\theta_0\left(\mathbf{u} - \frac{\nu'}{c}\mathbf{b}\right)\right] \gamma\left(\frac{\nu - \nu'}{B_{\text{chan}}}\right)$$
$$\times \phi\left(\frac{\nu - \nu_0}{B_{\text{band}}}\right) e^{i2\pi(\nu'\eta - \nu\tau)}. \tag{5b}$$

The integral over ν' can be simplified because $\tilde{A}$ is a much broader function (with a characteristic width of $c\mathbf{b}/\theta_0$) than γ (which has a characteristic width of B_{chan}). For a typical EoR experiment, one might have $\nu_0 = 150$ MHz and $B_{\text{chan}} = 50$ kHz. Even with a rather conservative $\theta_0 = 1$ rad, $\tilde{A}$ is wider than γ for any baseline shorter than $\sim 3000\lambda$. With compact arrays giving the highest sensitivity [17,47], very few interferometers that are optimized for an EoR measurement have any relevant sensitivity on baselines this long. As a result, $\tilde{A}$ can be factored out of the integral and evaluated at $\nu' = \nu$, leaving a Fourier transform

---

[2]Note that the distinction between τ modes and η modes implies that our tapering functions differ slightly from those examined in Ref. [43]. In this paper, the tapering functions are applied to the per-baseline Fourier transform, rather than the Fourier transform along the true frequency axis of Fig. 2.

over the ν' variable. Defining $\tilde{\gamma}$ to be the Fourier transform of γ, this leaves

$$
\begin{aligned}
\tilde{V}(\mathbf{b}, \tau) \approx \int d^2\mathbf{u}\, d\eta\, \tilde{I}(\mathbf{u}, \eta)\, \theta_0^2 \tilde{\gamma}(B_{\mathrm{chan}}\eta) \\
\times \int d\nu \tilde{A}\left[\theta_0\left(\mathbf{u} - \frac{\nu}{c}\mathbf{b}\right)\right] \\
\times \phi\left(\frac{\nu - \nu_0}{B_{\mathrm{band}}}\right) e^{i2\pi\nu(\eta-\tau)}.
\end{aligned} \tag{6}
$$

With our approximations, then, the effect of having frequency channels with nonzero width is to envelope our response in η: limited spectral resolution makes the array less sensitive to high η modes (i.e., rapidly fluctuating spectral modes).

To further simplify our expression, it is useful to orient our $\mathbf{u} \equiv (u, v)$ axes so that the u axis is in the same direction as the baseline vector $\mathbf{b}$. If we further assume that the antenna's footprint on the uv plane is separable, i.e.,

$$
\tilde{A}\left[\theta_0\left(\mathbf{u} - \frac{\nu}{c}\mathbf{b}\right)\right] = \tilde{A}_{b\|}\left[\theta_0\left(u - \frac{\nu}{c}b\right)\right]\tilde{A}_{b\perp}(\theta_0 v), \tag{7}
$$

then our expression becomes

$$
\begin{aligned}
\tilde{V}(\mathbf{b}, \tau) \approx \int d^2\mathbf{u}\, d\eta\, \tilde{I}(\mathbf{u}, \eta)\, \theta_0^2 \tilde{A}_{b\perp}(\theta_0 v)\tilde{\gamma}(B_{\mathrm{chan}}\eta) \\
\times \int d\nu \tilde{A}_{b\|}\left[\theta_0\left(u - \frac{\nu}{c}b\right)\right] \\
\times \phi\left(\frac{\nu - \nu_0}{B_{\mathrm{band}}}\right) e^{i2\pi\nu(\eta-\tau)}.
\end{aligned} \tag{8}
$$

To proceed beyond this point requires specific forms for $\tilde{A}_{b\|}$ and γ. However, it is instructive to examine various limits. For an instrument with short baselines and/or narrow fields of view satisfying $b\theta_0 \ll c/B_{\mathrm{band}}$, $\tilde{A}_{b\|}$ is a slowly varying function that can be factored out of the integral, yielding

$$
\begin{aligned}
\tilde{V}(\mathbf{b}, \tau)\Big|_{b\theta_0 \ll c/B_{\mathrm{band}}} \approx \int d^2\mathbf{u}\, d\eta\, \tilde{I}(\mathbf{u}, \eta)\, \theta_0^2 B_{\mathrm{band}} \\
\times \tilde{A}\left[\theta_0\left(\mathbf{u} - \frac{\nu_0}{c}\mathbf{b}\right)\right]\tilde{\gamma}(B_{\mathrm{chan}}\eta) \\
\times \tilde{\phi}[B_{\mathrm{band}}(\eta - \tau)]e^{i2\pi\nu_0(\eta-\tau)},
\end{aligned} \tag{9}
$$

where we used Eq. (7) to recombine $\tilde{A}_{b\|}$ and $\tilde{A}_{b\perp}$. This gives the "usual" description of an interferometric measurement: each baseline samples a portion of the sky in $uv\eta$ space, defined by the antenna's footprint on the uv plane and the Fourier transform of the bandpass shape in η, enveloped by the Fourier transform of a frequency channel's profile. On the other hand, for an instrument with long baselines and/or

wide fields of view satisfying $b\theta_0 \gg c/B_{\mathrm{band}}$, it is ϕ that is broad compared to $\tilde{A}_{b\|}$, and Eq. (8) is well approximated by[3]

$$
\begin{aligned}
\tilde{V}(\mathbf{b}, \tau)\Big|_{b\theta_0 \gg c/B_{\mathrm{band}}} \approx \int d^2\mathbf{u}\, d\eta\, \tilde{I}(\mathbf{u}, \eta)\, \frac{c\theta_0}{b}\tilde{A}_{b\perp}(\theta_0 v) \\
\times \tilde{\gamma}(B_{\mathrm{chan}}\eta)\phi\left(\frac{uc/b - \nu_0}{B_{\mathrm{band}}}\right) \\
\times A_{b\|}\left[\frac{c}{b\theta_0}(\eta - \tau)\right]e^{i2\pi(\eta-\tau)\frac{uc}{b}}.
\end{aligned} \tag{10}
$$

We see that in this limit, the bandpass shape ϕ and the primary beam $A_{b\|}$ swap roles: the bandpass acts as the convolution kernel on the uv plane, while the primary beam acts as the convolution kernel in the η direction. When $b\theta \sim c/B_{\mathrm{band}}$, the full expression given by Eq. (8) interpolates between the two extremes given by Eqs. (9) and (10); in general, the bandpass shape and the primary beam both play a part in determining the uv plane and η-direction convolution kernels. This point was emphasized in Ref. [32].

Returning to the long-baseline limit of Eq. (10), we can see our first glimpse of the foreground wedge. Suppose one were dealing with flat-spectrum foregrounds, where $\tilde{I}(u, \eta) = \tilde{I}_0(u)\delta(\eta)$, so that there is no signal beyond $\eta = 0$. With such a sky, the measurement becomes proportional to $A_{b\|}(c\tau/b\theta_0)$ and $|\tilde{V}(\mathbf{b}, \tau)|^2 \propto A_{b\|}^2(c\tau/b\theta_0)$. Suppose we now make use of the delay approximation [32], where the quantity $|\tilde{V}(\mathbf{b}, \tau)|^2$ can be treated as an estimate of the power spectrum $P(u, \eta)$ at $u \approx \nu_0 b/c$ and $\eta \approx \tau$. The result is

$$
P(u, \eta) \propto A_{b\|}^2(\nu_0\eta/u\theta_0). \tag{11}
$$

Thus, even flat spectrum sources (which would naively only have power at $\eta = 0$) give a nonzero measured power spectrum at higher η. If the primary beam is zero beyond some argument value (which is always true in some sense, since A is identically zero below the horizon), then the power extends only to a finite region on the $u\eta$ plane. For example, if we define θ_0 to be the angle at which A drops to zero (we are free to do so, since θ_0 is simply a characteristic scale that we have so far kept general as simply "some characteristic scale"), Eq. (11) predicts that the line

$$
\eta = \frac{\theta_0}{\nu_0}u \tag{12}
$$

---

[3]As one factors ϕ out of the integral, the function can no longer enforce the limits on the ν integral that reflect the finite bandwidth of the instrument. However, the $\tilde{A}$ function that remains inside the integral is by construction narrow. It is therefore perfectly acceptable to keep $-\infty$ and $+\infty$ as the limits of integration.

should be a sharp boundary between zero and nonzero power. Switching from angular or spectral Fourier coordinates u and η to comoving spatial wave numbers $k_\perp$ and $k_\parallel$ using the relations in Appendix A gives

$$k_\parallel = \frac{H_0 D_c E(z)\theta_0}{c(1+z)} k_\perp, \qquad (13)$$

where $E(z) \equiv \sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}$, D_c is comoving line-of-sight distance, H_0 is the Hubble parameter, Ω_m is the normalized matter density, and Ω_Λ is the normalized dark energy density. This is precisely the usual formula given for the edge of the foreground wedge (e.g., in Refs. [15,32,41,42]). As a function of η (or $k_\parallel$), the wedge has a profile given by the square of the primary beam profile, scaled by u in such a way that power is found at higher η for higher u.

In presenting a preview of the wedge, we have made a number of assumptions that will be relaxed in the following sections. For example, we will no longer approximate τ modes as η modes. Nor will we assume that each baseline cleanly samples just a single value of u. In fact, our use of the delay approximation here was somewhat inappropriate—while one may always take a delay *transform*, we have seen (e.g., from Fig. 2) that the delay *approximation* (i.e., assuming $\tau \sim \eta$) ought to work well only if baselines are short. By using Eq. (10) instead of Eq. (9), however, we are expressly working in the long-baseline limit. It is thus important to emphasize that the preview shown here is included only to build intuition, and a much more rigorous treatment will be presented in later sections. There, we will also generalize to a sky with an arbitrary power spectrum. While there will be minor alterations to details of the foreground wedge, the basic picture will remain intact.

In the strictest sense, the derivation that we have just presented is nothing new. We have simply rederived a number of results (e.g., the existence of the wedge) that are already known in the literature. However, our rederivations have been part of an analytic formalism, confirming a number of numerical results while bringing together many known (but previously separate) features in a unified framework. In the next section, we will extend this framework to include a fully covariant description of the wedge, including correlated errors in such a way that extends the quadratic power spectrum estimation techniques of Ref. [36] to include wedge physics. Setting up such a framework allows one to systematically examine the statistical properties of various power spectrum estimators in light of the wedge, a study that we perform in paper II.

III. QUADRATIC ESTIMATOR FORMALISM

Until now, our focus has been on the measurement of a visibility, which is *linear* in temperature. The power spectrum, however, is a quantity that depends quadratically on temperature. In this section we very briefly review the mathematical machinery that makes possible the fully covariant description of the EoR window we will present in Secs. IV and V. The basis of our discussion will be the quadratic estimator formalism, which has a long history in the CMB and galaxy survey literature (e.g., Refs. [49–51]) and was explicitly adapted for 21 cm cosmology in Refs. [36,37].

The central quantity that encodes our instrument (and therefore the statistical properties of our estimators) is the data covariance matrix $\mathbf{C}$. To form the covariance matrix, imagine that our input data (organized by baseline and delay mode) are serialized into a data vector $\mathbf{x}$, i.e.,

$$\mathbf{x} = \begin{pmatrix} \tilde{V}(\mathbf{b}_1, \tau_1) \\ \tilde{V}(\mathbf{b}_1, \tau_2) \\ \vdots \\ \tilde{V}(\mathbf{b}_2, \tau_1) \\ \tilde{V}(\mathbf{b}_2, \tau_2) \\ \vdots \end{pmatrix}. \qquad (14)$$

The covariance matrix is then given by $\mathbf{C} \equiv \langle \mathbf{xx}^\dagger \rangle$. Importantly, we will keep track of all off-diagonal elements in the matrix, so that all correlations between different baselines and different delay (or spectral) modes are taken into account. Knowledge of these correlations will allow us to formulate both a covariant description of the foreground wedge and the tools to fight its contaminating influence.

Although we omitted noise contributions in the previous section for notational cleanliness, they of course contribute to the variance captured by $\mathbf{C}$. Assuming that instrumental noise is uncorrelated with sky signals, the noise appears as an additive term to the sky covariance, so that we can write $\mathbf{C}$ as

$$\mathbf{C} \equiv \mathbf{N} + \mathbf{S}, \qquad (15)$$

where $\mathbf{S}$ is the sky signal portion of the covariance. Inserting Eq. (6) into our general definition of the data covariance, we obtain

$$\mathbf{S}_{ij} = \int d^2\mathbf{u}\,d\eta\, P(\mathbf{u},\eta) h(\mathbf{u},\eta;\mathbf{b}_i,\tau_i) h^*(\mathbf{u},\eta;\mathbf{b}_j,\tau_j), \qquad (16)$$

where we have defined

$$h(\mathbf{u},\eta;\mathbf{b},\tau) \equiv \theta_0^2 \tilde{\gamma}(B_{\text{chan}}\eta) \int d\nu \tilde{A} \left[\theta_0 \left(\mathbf{u} - \frac{\nu}{c}\mathbf{b} \right) \right]$$
$$\times \phi\left(\frac{\nu - \nu_0}{B_{\text{band}}} \right) e^{i2\pi\nu(\eta-\tau)}, \qquad (17)$$

and have made use of the definition of the power spectrum $P(\mathbf{u},\eta)$:

$$\langle \tilde{I}(\mathbf{u},\eta)\tilde{I}^*(\mathbf{u}',\eta')\rangle \equiv P(\mathbf{u},\eta)\delta(\mathbf{u}-\mathbf{u}')\delta(\eta-\eta'). \quad (18)$$

As we have written it, our power spectrum is expressed in terms of $\mathbf{u}$ and η, instead of the more common combination of $\mathbf{k}_\perp$ and $k_\parallel$. Doing so minimizes the number of cosmological quantities in our expressions, since $\mathbf{u}$ and η are the more "natural" quantities from the perspective of the instrument. This represents no loss of generality, and indeed, in Sec. V we will express our results in terms of $\mathbf{k}_\perp$ and $k_\parallel$.

To form a practical estimator for the power spectrum, it is necessary to discretize. Assuming that the power spectrum possesses cylindrical symmetry[4] (so that the power depends only on the magnitudes of $\mathbf{u}$ and η), one can imagine binning the $uv\eta$ space into a series of annuli, each specified by a radius $|u| = \sqrt{u^2+v^2}$ and "vertical distance" $|\eta|$ away from the $\eta = 0$ plane. Each annulus can then be represented as a small cell on a $|u|\eta$ plane (which we will henceforth call the $u\eta$ plane to conform to convention).[5] As long as these cells are made sufficiently small, the power spectrum can be approximated by a constant *band power* p_α in each cell, with α indexing a serialized list of locations on the $u\eta$ plane. With this approximation, our covariance can be written compactly as

$$\mathbf{C} = \mathbf{N} + \sum_\alpha p_\alpha \mathbf{C}_{,\alpha}, \quad (19)$$

where $\mathbf{C}_{,\alpha} \equiv \partial\mathbf{C}/\partial p_\alpha$ is the response of the covariance to αth band power and is given by

$$(\mathbf{C}_{,\alpha})_{ij} = \int_{V_\alpha} d^2\mathbf{u}\,d\eta\,h(\mathbf{u},\eta;\mathbf{b}_i,\tau_i)h^*(\mathbf{u},\eta;\mathbf{b}_j,\tau_j), \quad (20)$$

with V_α denoting the annular volume that is binned into the αth $u\eta$ cell. If desired, the sky contribution of the

covariance can be further divided into separate contributions from foregrounds and the cosmological signal:

$$\mathbf{C} = \mathbf{N} + \mathbf{C}_{\mathrm{fg}} + \sum_\alpha p_\alpha^{\mathrm{sig}}\mathbf{C}_{,\alpha}, \quad (21)$$

where $\mathbf{C}_{\mathrm{fg}}$ is the foreground covariance and p_α^{sig} is the αth band power of the cosmological signal only. Equation (21) is more general than Eq. (19), because the latter implicitly assumes that $\mathbf{C}_{\mathrm{fg}}$ is given by $\sum_\alpha p_\alpha^{\mathrm{fg}}\mathbf{C}_{,\alpha}$, where p_α^{fg} is a set of foreground band powers. This assumption holds true only when the foregrounds are describable as a power spectrum.

In the quadratic estimator formalism, the band powers are extracted by forming weighted pairwise combinations of the data vector $\mathbf{x}$. In particular, one can form a quadratic estimator $\hat{p}_\alpha$ of the true band power p_α by computing[6]

$$\hat{p}_\alpha = \mathbf{x}^\dagger \mathbf{E}^\alpha \mathbf{x}, \quad (22)$$

where $\mathbf{E}^\alpha$ is an estimator matrix of weights to be used for weighting pairwise products of the data when estimating the αth band power. To see how this works, consider (as a toy example) a noiseless cosmological survey with uncorrelated real-space measurements in a three-dimensional volume. Further suppose that one's goal is to measure the unbinned, three-dimensional power spectrum $P(\mathbf{k})$, i.e., $\hat{p}_\alpha = \hat{P}(\mathbf{k}_\alpha)$. If $\mathbf{x}$ is expressed in a real-space basis (so that it is simply a serialized list of real-space voxel intensities), a sensible choice for the estimator matrix would be $\mathbf{E}_{ij}^\alpha \propto e^{-i\mathbf{k}_\alpha\cdot(\mathbf{r}_i-\mathbf{r}_j)}$, where $\mathbf{r}_i$ and $\mathbf{r}_j$ are the position vectors of the ith and jth voxels, respectively. We therefore see that in this example, the role of $\mathbf{E}^\alpha$ is to take a Fourier transform of the data. If one desires estimates of a *binned* power spectrum (for example, one where statistical isotropy allows the binning of power over shells of constant $|\mathbf{k}|$), the relevant $\mathbf{E}^\alpha$ for different $\mathbf{k}_\alpha$ are simply averaged together in each bin.

Now suppose instead that our data are expressed in a Fourier basis, so that each element of the data vector $\mathbf{x}$ represents the Fourier amplitude at some location in $\mathbf{k}$ space. The estimator matrix is then even simpler, and in fact becomes diagonal, with $\mathbf{E}_{ij}^\alpha = \delta_{\alpha i}\delta_{ij}$. In our current application, the data are organized by baseline $\mathbf{b}$ and delay τ. As discussed above, $\mathbf{b}$ and τ closely approximate the Fourier wave numbers $\mathbf{u}$ and η in some regimes, but the correspondence is not perfect. For our application we would therefore expect $\mathbf{E}^\alpha$ to be diagonal dominant, but not be perfectly diagonal. (For an explicit form, see Secs. IV and V.)

In general, correlated errors and other instrumental effects (such as the ones that we seek to model in this paper) make the estimator matrices more complicated than

---

[4]The true power spectrum of the cosmological signal is of course spherically symmetric (thanks to statistical isotropy) and thus can be further binned. However, systematics such as foregrounds tend to be cylindrically symmetric, since the instrument probes the line-of-sight direction differently than it does the angular directions [52]. Another effect to consider (though it is beyond the scope of this paper) is that of redshift-space distortions, which will break spherical symmetry. The cylindrical power spectrum is therefore a useful diagnostic quantity to compute prior to the formation of a spherical power spectrum. Note that in general, we need not make any assumptions about symmetry in our estimation formalism. If desired, our formalism can be used to estimate $P(\mathbf{u},\eta)$ without even the cylindrical binning step. The presence of symmetry, while useful for increasing signal to noise, is much less important than other assumptions such as the smoothness of foregrounds. (Although see Sec. VI for a discussion of strategies for dealing with unsmooth foregrounds.)

[5]We note an unfortunate bit of notation: u is used both to denote the first coordinate on the uv plane as well as the magnitude of the $\mathbf{u}$ vector. Unfortunately, both usages are standard.

[6]Throughout this paper, we use hats to denote estimators of quantities.

they were in our pedagogical examples. They will typically involve the $\mathbf{C}_{,\alpha}$ matrices, since those provide the link between the "input" space that the data covariance inhabits and the "output" space of band powers. The detailed form of the family of $\mathbf{E}_\alpha$ matrices is a choice made by the data analyst, and different choices yield estimators with different statistical properties. One such property is the error covariance $\Sigma_{\alpha\beta} \equiv \langle \hat{p}_\alpha \hat{p}_\beta \rangle - \langle \hat{p}_\alpha \rangle \langle \hat{p}_\beta \rangle$ of our estimated band powers, which is given by[7]

$$\Sigma_{\alpha\beta} = 2\mathrm{tr}[\mathbf{C}\mathbf{E}^\alpha\mathbf{C}\mathbf{E}^\beta], \qquad (23)$$

a result that can be derived by direct substitution of Eq. (22) into the definition of the error covariance. The error bars on our band power estimates are given by $\Delta p_\alpha \equiv (\Sigma_{\alpha\alpha})^{\frac{1}{2}}$, but it is important to note that the error covariance contains much more information than just the error bars: off-diagonal elements of the covariance encode correlations between different $u\eta$ cells of the cylindrical power spectra. With a fully covariant formulation of the errors, it is possible to overresolve in $u\eta$ space, evading the commonly made assumption that $u\eta$ cells are independent (i.e., have a diagonal covariance matrix) so long as they are more than $1/B_{band}$ apart in the η direction and separated by more than the width of $\tilde{A}$ in the u direction. While this assumption is prevalent in the 21 cm cosmology literature for reasons of computational simplicity, we will see that it is one that should be avoided. More precisely, we will see in Sec. V that the errors on the $u\eta$ plane are not independent at high $k_\perp$ (including the foreground wedge region), necessitating a full accounting of the entire error covariance matrix and not just its diagonal elements.

In addition to the error covariance, the quadratic estimator formalism also allows the computation of biases and window functions. Taking the expectation value of Eq. (22), recalling that $\mathbf{C} \equiv \langle \mathbf{x}\mathbf{x}^\dagger \rangle$, and inserting Eq. (21) gives

$$\hat{p}_\alpha = \sum_\beta \mathrm{tr}[\mathbf{E}^\alpha\mathbf{C}_{,\beta}]p_\beta^{\mathrm{sig}} + \mathrm{tr}[\mathbf{E}^\alpha(\mathbf{N} + \mathbf{C}_{\mathrm{fg}})]$$

$$\equiv \sum_\beta \mathbf{W}_{\alpha\beta}p_\beta^{\mathrm{sig}} + b_\alpha, \qquad (24)$$

where we have defined the window function matrix

$$\mathbf{W}_{\alpha\beta} \equiv \mathrm{tr}[\mathbf{E}^\alpha\mathbf{C}_{,\beta}] \qquad (25)$$

and the contamination bias

$$b_\alpha = \mathrm{tr}[\mathbf{E}^\alpha(\mathbf{N} + \mathbf{C}_{\mathrm{fg}})]. \qquad (26)$$

From Eq. (24), we see that each row of the window function matrix gives a *window function* that specifies the linear combination of the true band powers that each estimate of a band power represents. Typically, the $\mathbf{E}^\alpha$ matrix is normalized such that each row of $\mathbf{W}$ sums to unity, allowing the linear combinations to be interpreted as weighted averages.[8]

The contamination bias represents an additive bias to the estimated power spectrum that arises due to residual noise and foregrounds in the data. In practice, cross-correlation techniques—such as forming cross-power spectra between odd and even time samples of data, as was done in Ref. [15], or between different subsets of redundant baselines in an array, as was done in Ref. [16]—allow the noise bias to be eliminated without any explicit bias subtraction. The bias that one needs contend with is therefore solely comprised of the foreground bias:

$$b_\alpha = b_\alpha^{\mathrm{fg}} = \mathrm{tr}[\mathbf{E}^\alpha\mathbf{C}_{\mathrm{fg}}]. \qquad (27)$$

If a perfect foreground model is available, this expected level of this bias can be computed and subtracted from the power spectrum estimate. However, because a detailed knowledge of the low-frequency sky is as yet unavailable, this subtraction step is often omitted to avoid oversubtractions that destroy cosmological information. Instead, one simply hopes that the bias is small in regions of the $u\eta$ plane where one wishes to make a power spectrum measurement.

In the following sections, the error covariance Σ, the window function matrix $\mathbf{W}$, and the bias b_α are the quantities that will provide us with a detailed, covariant picture of the EoR window and the foreground wedge. The bias essentially captures the power spectrum contribution from noise and foreground contaminants and corresponds to the foreground wedge signatures seen in various simulations in the literature. The window functions provide an alternate view of the wedge: the wedge can be thought of as a leakage of power from low to high η (or equivalently, $k_\parallel$) modes that becomes increasingly pronounced as u (or $k_\perp$) increases. For a wedge to exist, window functions for band

---

[7]While we write all vector and matrix quantities in boldface, it is important to note that there are two different vector spaces at work here. The error covariance Σ and the window function matrix $\mathbf{W}$ defined later inhabit the output vector space indexed by locations on the $u\eta$ plane, unlike matrices such as $\mathbf{N}$ and $\mathbf{C}$, which inhabit the input vector space indexed by baseline and delay. Hybrid quantities include $\mathbf{C}_{,\alpha}$ and $\mathbf{E}^\alpha$, which can either be thought of as a family of matrices in the input vector space or as rank-3 tensors in a combined space.

[8]In the signal processing literature, the tapering function ϕ that we introduced in the previous section is often called a "window function." This is a conceptually separate use of the term, and we avoid the signal processing nomenclature in order to be consistent with the cosmology literature. For us, a window function will always refer to the linear combination of the true power spectrum that forms a particular band power estimate, and never the tapering function.

powers centered at high η and high u must have tails that extend to low η, where foregrounds live. Finally, the error covariance provides an estimate of the error bars throughout the Fourier plane (including the wedge region), as well as a quantification of how the chromatic response of an interferometer can cause error correlations between otherwise uncorrelated $u\eta$ cells.

IV. A COVARIANT DESCRIPTION OF THE FOREGROUND WEDGE FOR A BASIC ESTIMATOR

In this section, we use the quadratic estimator formalism to derive the foreground wedge and the EoR window for a "basic" estimator of the power spectrum. We will make a number of approximations for the sake of analytical tractability, leaving an exact numerical treatment to Sec. V. The goal here is to formalize the discussion from Sec. II to obtain a fully covariant, analytic description of power spectrum properties at high u and low η, where the foreground wedge resides. This will provide a basic picture of the foreground challenges that we face, setting the stage for paper II, where we look at how these challenges can be mitigated with better estimators.

The relatively simple estimator that we will examine in this paper is specified by the relation

$$\mathbf{E}^{\alpha} = M_{\alpha}\mathbf{N}^{-1}\mathbf{C}_{,\alpha}\mathbf{N}^{-1}, \quad (28)$$

where $\mathbf{N}$ is the instrumental noise covariance and M_{α} is a normalizing scalar[9] for each band power α. This choice gives an estimator that is quite similar to the crude $|\tilde{V}(\mathbf{b},\tau)|^2$ estimator discussed informally in Sec. II. The principal difference between what we will consider here and our previous estimator is the presence of $\mathbf{C}_{,\alpha}$. The role of $\mathbf{C}_{,\alpha}$ is twofold. Its first purpose is to complete a signal-to-noise weighting of our data: the copies of $\mathbf{N}^{-1}$ downweight noisy data, while $\mathbf{C}_{,\alpha}$ (by virtue of its being the derivative of $\mathbf{C}$) upweights the high signal portions. The second purpose is to map the data from the input baseline $\mathbf{b}$ and delay τ space to the output $u\eta$ space. Recall from Sec. II that while $\mathbf{b}$ and τ approximate $\mathbf{u}$ and η, respectively, the correspondence is not perfect. Applying $\mathbf{C}_{,\alpha}$ completes the transition to Fourier space, a fact that will become more apparent when we write down an explicit form for the matrix.

Studying the basic estimator given by Eq. (28) is worthwhile because it is approximately equivalent to the methods used in a number of state-of-the-art 21 cm power spectrum pipelines for analyzing observations and

simulations [43,44,53]. These pipelines typically use an optimal mapmaking approach [54,55] to first go from visibilities to a gridded $uv\eta$ data cube of Fourier amplitudes. The complex magnitudes of these amplitudes are then squared and binned to estimate power spectrum band powers. (Note that while the mapmaking may be optimal in this case, the subsequent power spectrum estimation is not.) In Appendix B, we will prove that in the limit of infinitely fine bins in Fourier space, such pipelines are equivalent to estimating power spectra directly from the visibilities using Eq. (28) and the quadratic estimator formalism. Our numerical results will therefore be roughly representative of those seen in the aforementioned pipelines, but with fewer gridding artifacts because we go straight from visibilities to power spectra.

For analytical tractability here and numerical tractability in later sections, we will use an approximate form for the covariance matrix:

$$\mathbf{C}_{ij} = \int du d\eta P(u,\eta) g(u,\eta;b_i,\tau_i) g^*(u,\eta;b_j,\tau_j), \quad (29)$$

where

$$g(u,\eta;b_i,\tau_i) \equiv k\tilde{\gamma}(B_{\mathrm{chan}}\eta)$$
$$\times \int d\nu \tilde{A}_{b\parallel}\left[\theta_0\left(u - \frac{\nu}{c}b_i\right)\right]$$
$$\times \phi\left(\frac{\nu - \nu_0}{B_{\mathrm{band}}}\right)e^{i2\pi\nu(\eta-\tau_i)} \quad (30)$$

with

$$k \equiv \theta_0\left(\int dq \tilde{A}_{b\perp}^2(\theta_0 q)\right)^{\frac{1}{2}}, \quad (31)$$

and we have once again omitted the instrumental noise contribution to the covariance for simplicity. Superficially, this looks quite similar to Eqs. (16) and (17). However, in this case we have gone beyond simply forming $\mathbf{C} \equiv \langle \mathbf{x}\mathbf{x}^{\dagger}\rangle$ from Eq. (8), in that we have performed the integral over $\tilde{A}_{b\perp}$. This is not always permissible and represents a subtle additional approximation: we have assumed that baselines that are similar in length but very different in orientation have a negligible correlation with each other, and that those with similar orientations are correlated as though they were identical in orientation. In other words, we assume that although two baselines can be completely uncorrelated, partially redundant, or perfectly redundant in the direction of the baseline vector,[10] overlaps between baselines in the

---

[9]In this paper, we do not consider the more general possibility of a matrix-based normalization, where instead of a simple multiplicative normalization, one multiplies the unnormalized band power estimates $\hat{p}'_{\beta}$ by a matrix $\mathbf{M}$ to form the normalized band powers $\hat{p}_{\alpha}$, i.e., $\hat{p}_{\alpha} = \sum_{\beta}\mathbf{M}_{\alpha\beta}\hat{p}'_{\beta}$. For details, please see Ref. [15] or paper II.

[10]Note that our ability to accommodate the full continuum of redundancy from zero to perfect redundancy along the direction of the baseline vector allows us to capture possible subtleties related to partial redundancy effects like those highlighted in Ref. [44].

transverse direction are treated in a binary fashion, so that the overlap is either zero or perfect. This assumption was inherited from the derivation of Eq. (8), which required a reorientation of the axes of the uv plane so that the u axis would lie along the direction of the baseline. While this can always be done for a single baseline, the covariance matrix **C** encodes correlations between different baselines, which may be oriented differently. It is thus strictly speaking incorrect to form a covariance matrix from Eq. (8), and in principle one should use Eq. (6) instead. For the purposes of intuition, however, we may continue with our approximate expression as long as we remember that distant baselines have negligible correlation.

As we previously mentioned, $\mathbf{C}_{,\alpha}$ provides the crucial link between the input data and the output Fourier space. It therefore forms a crucial component of any error statistic. Working in the limit of a continuous (rather than discrete) set of band powers, we may differentiate **C** with respect to $P(u_\alpha, \eta_\alpha)$ to obtain

$$(\mathbf{C}_{,\alpha})_{ij} \equiv \frac{\partial \mathbf{C}_{ij}}{\partial P(u_\alpha, \eta_\alpha)} = g(u_\alpha, \eta_\alpha; b_i, \tau_i) g^*(u_\alpha, \eta_\alpha; b_j, \tau_j),$$
(32)

where we used the fact that $P(u, \eta)$ can be written as $\int du_\alpha d\eta_\alpha P(u_\alpha, \eta_\alpha) \delta(u - u_\alpha) \delta(\eta - \eta_\alpha)$. Inserting this into Eq. (28) provides a concrete example of the general proof of equivalence given in Appendix B. One sees that each copy of g acts on a noise-weighted copy of the data vector **x**. Examining Eq. (30) reveals that the action of g is to Fourier transform the delay spectrum back into the frequency domain, apply another copy of the tapering functions, grid the result at the appropriate location on the uv plane, and then Fourier transform in frequency again, before adjusting for frequency channel discretization by an additional weighting in η. This is precisely the procedure that one would follow with a mapmaking algorithm in $uv\eta$ space [55]. The result is then squared to form a power spectrum. While this particular example may at first sight seem to render the delay basis obsolete (since the first action of g is to transform back to a frequency spectrum), it is important to remember that in a more realistic case, one may be unable approximate the band powers as being continuous. For example, at small u and η, bin sizes may be comparable to the values of u and η themselves. Many of the algebraic simplifications used in this section then become inapplicable, necessitating full numerical manipulations of the relevant matrices, which are typically more computationally efficient in the delay basis (as we discuss in Appendix C).

Continuing with our approximation scheme for this section, however, Eq. (32) is particularly convenient for computing our suite of error statistics because it is separable. Taking advantage of this, the window functions for our basic estimator reduce to

$$\mathbf{W}_{\alpha\beta} = \text{tr}[\mathbf{E}^\alpha \mathbf{C}_{,\beta}] \propto \text{tr}[\mathbf{C}_{,\alpha} \mathbf{C}_{,\beta}] = |\mathbf{g}^{\alpha\dagger} \cdot \mathbf{g}^\beta|^2$$

$$\approx |\sum_b \int d\tau g^*(u_\alpha, \eta_\alpha; b, \tau) g(u_\beta, \eta_\beta; b, \tau)|^2, \quad (33)$$

where we defined the shorthand $\mathbf{g}^\alpha \equiv g(u_\alpha, \eta_\alpha; b, \tau)$, and in the last step assumed that our delay bins were fine enough to be approximated as being continuous. This form for the window function matrix has a straightforward geometric interpretation: when estimating the αth band power, one probes a mixture of the true band powers; the amount of the βth band that is included in the estimate of the αth band is given by the overlap of our interferometer's response to the αth and βth bands. We stress, however, that this simple form does not hold when one considers more complicated estimators such as the ones that we will consider in paper II.

Let us now compute some example window functions. Just as we did in Sec. II, we can gain some analytic intuition by working in the short- and long-baseline limits. For short baselines satisfying $b\theta_0 \ll c/B_{\text{band}}$ (or equivalently, if the time delay τ of a signal between the antennas of the baseline satisfies $\tau \ll 1/\theta_0 B_{\text{band}}$), we may invoke the same approximations that led to Eq. (9) and say that

$$g^{\text{short b1}}(u, \eta; b, \tau) \propto \tilde{\gamma}(B_{\text{chan}}\eta)\tilde{A}_{b\|}\left[\theta_0\left(u - \frac{\nu_0}{c}b\right)\right]$$

$$\times \tilde{\phi}[B_{\text{band}}(\eta - \tau)]e^{i2\pi\nu_0(\eta - \tau)}. \quad (34)$$

Inserting Eq. (34) into Eq. (33) and evaluating the integral over τ yields

$$W_{u_\alpha, \eta_\alpha}^{\text{short b1}}(u_\beta, \eta_\beta) \propto \tilde{\gamma}^2(B_{\text{chan}}\eta_\alpha)\tilde{\gamma}^2(B_{\text{chan}}\eta_\beta)$$

$$\times \left(\sum_b \tilde{A}_{b\|}\left[\theta_0\left(u_\alpha - \frac{\nu_0}{c}b\right)\right]\right.$$

$$\times \tilde{A}_{b\|}\left[\theta_0\left(u_\beta - \frac{\nu_0}{c}b\right)\right]\Bigg)^2$$

$$\times (\widetilde{\phi^2}[B_{\text{band}}(\eta_\alpha - \eta_\beta)])^2, \quad (35)$$

where $\widetilde{\phi^2}$ signifies the Fourier transform of ϕ^2, not the square of $\tilde{\phi}$. This expression is in line with what one might intuitively expect from interferometry: the spatial u dependence of the window function is controlled by the primary beam (or more precisely, its Fourier transform), while the spectral η dependence is controlled by the bandpass.

On the other hand, with long baselines satisfying $b\theta_0 \gg c/B_{\text{band}}$ we can use the approximations that led to Eq. (10), obtaining

$$g^{\text{long b1}}(u, \eta; b, \tau) \propto \tilde{\gamma}(B_{\text{chan}}\eta)\phi\left(\frac{uc/b - \nu_0}{B_{\text{band}}}\right) \int d\nu \tilde{A}_{b\parallel}\left[\theta_0\left(u - \frac{\nu}{c}b\right)\right]e^{i2\pi\nu(\eta-\tau)}. \tag{36}$$

Once again, we may insert this into Eq. (33) to get

$$W^{\text{long b1}}_{u_\alpha, \eta_\alpha}(u_\beta, \eta_\beta) \propto \tilde{\gamma}^2(B_{\text{chan}}\eta_\alpha)\tilde{\gamma}^2(B_{\text{chan}}\eta_\beta)\left|\sum_b \phi\left(\frac{u_\alpha c/b - \nu_0}{B_{\text{band}}}\right)\phi\left(\frac{u_\beta c/b - \nu_0}{B_{\text{band}}}\right)\right.$$
$$\left.\times \int d\nu \tilde{A}^*_{b\parallel}\left[\theta_0\left(u_\alpha - \frac{\nu}{c}b\right)\right]\tilde{A}_{b\parallel}\left[\theta_0\left(u_\beta - \frac{\nu}{c}b\right)\right]e^{-i2\pi\nu(\eta_\alpha-\eta_\beta)}\right|^2. \tag{37}$$

Now, consider first the η dependence (rather than the u dependence) of the window functions, since our principal worry is that smooth, low-η foregrounds might scatter to higher η because of the instrument's chromaticity. If we suppress the u dependence by setting $u = u_\alpha = u_\beta$, we obtain

$$W^{\text{long b1}}_{\eta_\alpha}(\eta_\beta) \propto \tilde{\gamma}^2(B_{\text{chan}}\eta_\alpha)\tilde{\gamma}^2(B_{\text{chan}}\eta_\beta)\left|\sum_b \frac{c}{b} e^{i2\pi\frac{uc}{b}(\eta_\beta-\eta_\alpha)}\phi^2\left(\frac{\frac{uc}{b} - \nu_0}{B_{\text{band}}}\right)(A_{b\parallel} * A_{b\parallel})\left[\frac{c}{b\theta_0}(\eta_\beta - \eta_\alpha)\right]\right|^2. \tag{38}$$

Equation (38) predicts that with our basic estimator, foregrounds should appear in the now-familiar wedge in $u\eta$ space. To see this, consider the additive foreground bias for a hypothetical single-baseline interferometer. Since we are concerned with foregrounds, the most relevant regions of the $u\eta$ plane will be the low η regions. We may therefore safely ignore the $\tilde{\gamma}^2$ terms, since at low η they will be approximately unity anyway. If we imagine that such foregrounds are described by a power spectrum p^{fg}_α, the foreground covariance is given by $\sum_\alpha p^{\text{fg}}_\alpha \mathbf{C}_{,\alpha}$, and the foreground contribution b^{fg}_α to the bias in Eq. (26) is given by

$$b^{\text{fg}}_\alpha = \text{tr}[\mathbf{E}^\alpha \mathbf{C}_{\text{fg}}] = \sum_\beta \text{tr}[\mathbf{E}^\alpha \mathbf{C}_{,\beta}]p^{\text{fg}}_\beta = \sum_\beta \mathbf{W}_{\alpha\beta}p^{\text{fg}}_\beta \tag{39a}$$

$$\propto \int \left((A_{b\parallel} * A_{b\parallel})\left[\frac{c}{b\theta_0}(\eta_\alpha - \eta_\beta)\right]\right)^2 P^{\text{fg}}(\eta_\beta)d\eta_\beta, \tag{39b}$$

where we have suppressed the u dependence of the foreground power spectrum for notational cleanliness. Now, if we approximate the foregrounds as being completely comprised of flat spectrum sources, then $P^{\text{fg}}(\eta) \propto \delta(\eta)$, and our foreground bias becomes

$$b^{\text{fg}}(\eta) \propto \left((A_{b\parallel} * A_{b\parallel})\left[\frac{c}{b\theta_0}\eta\right]\right)^2. \tag{40}$$

This equation provides a general mathematical form for the profile of the foreground wedge, as a function of η, and reduces to previously derived special cases for the wedge profile in the limit of top-hat primary beam shapes and bandpasses [42]. Had we retained the ϕ^2 terms, they would have enforced the condition that $u \approx b\nu_0/c$. We therefore see that foreground contamination ought to leak to higher η for higher values of u, since those are probed only by larger b. Importantly, we emphasize that this foreground wedge feature is basis independent, in that the final result makes no mention of delays. Our choice in this paper to express visibilities and covariances using a delay basis (rather than, say, a frequency basis, i.e., spectra) is a choice that is computationally convenient (see Appendix C for details) but is fundamentally an arbitrary one. The same is true regarding our indexing of measurements by baseline. [The baseline length b that appears in Eq. (40) is an expression of array configuration rather than basis choice; indeed, Eq. (33) shows that all baseline indices are summed over.] *The foreground wedge is purely a function of our instrument's design and the form of our power spectrum estimator.*

With multiple baselines, Eq. (38) contains cross terms between different baselines. Working once again at low η and high u (long baselines), one has

$$W_{\eta_\alpha}(\eta_\beta) \propto \sum_{b_i} \frac{1}{b_i^4} \phi^4 \left(\frac{\frac{uc}{b_i} - \nu_0}{B_{\text{band}}} \right) \left((A_{b\parallel} * A_{b\parallel}) \left[\frac{c}{b_i \theta_0} (\eta_\alpha - \eta_\beta) \right] \right)^2 + \sum_{b_i} \sum_{b_j > b_i} \frac{2}{b_i^2 b_j^2} \phi^2 \left(\frac{\frac{uc}{b_i} - \nu_0}{B_{\text{band}}} \right) \phi^2 \left(\frac{\frac{uc}{b_j} - \nu_0}{B_{\text{band}}} \right)$$

$$\times \cos \left[2\pi \left(\frac{1}{b_i} - \frac{1}{b_j} \right) (\eta_\alpha - \eta_\beta) u \right] (A_{b\parallel} * A_{b\parallel}) \left[\frac{c}{b_i \theta_0} (\eta_\alpha - \eta_\beta) \right] (A_{b\parallel} * A_{b\parallel}) \left[\frac{c}{b_j \theta_0} (\eta_\alpha - \eta_\beta) \right]. \quad (41)$$

While this expression is certainly more complicated than the one we had before, the same basic picture holds: the window functions can be quite broad in the η direction, thus allowing foregrounds to be scattered from low to medium values of η. But in general, the resulting contamination is still enveloped by terms like $(A_{b\parallel} * A_{b\parallel}) [\frac{c}{b\theta_0} \eta]$, which limits the possible contamination at high η.

Let us now turn briefly to the behavior of the window functions as a function of u. With the short-baseline limit already provided by Eq. (35), we once again focus on the long-baseline limit. Setting $\eta = \eta_\alpha = \eta_\beta$ to isolate the u dependence, Eq. (37) becomes

$$W_{u_\alpha}^{\text{long b1}}(u_\beta) \propto \tilde{\gamma}^4 (B_{\text{chan}} \eta) \left| \widetilde{A_{b\parallel}^2} [\theta_0 (u_\beta - u_\alpha)] \sum_b \frac{c}{b\theta_0} \right.$$

$$\left. \times \phi \left(\frac{u_\alpha c/b - \nu_0}{B_{\text{band}}} \right) \phi \left(\frac{u_\beta c/b - \nu_0}{B_{\text{band}}} \right) \right|^2, \quad (42)$$

where $\widetilde{A_{b\parallel}^2}$ denotes the Fourier transform of the square of $A_{b\parallel}$, not the square of the Fourier transform. As expected, a baseline b roughly probes a u scale equal to $b\nu_0/c$. Additionally, the window functions peak at $u_\alpha = u_\beta$, a fact that is enforced by the appearance of the two copies of ϕ as a product, as well as by the presence of $\widetilde{A_{b\parallel}^2}[\theta_0 (u_\beta - u_\alpha)]$. The width of the window function in the u direction therefore depends on both the primary beam and the bandpass function. The $\widetilde{A_{b\parallel}^2}$ term has a characteristic width of θ_0^{-1}, while the ϕ functions have a characteristic width of $B_{\text{band}} b/c$. Now, the window function involves the product of these functions, which means that the width of its central portion will be determined mostly by the *narrower* of the two contributions. In the long-baseline limit that we are working in, $\theta_0^{-1} \ll B_{\text{band}} b/c$ by definition, so the $\widetilde{A_{b\parallel}^2}$ part is the narrower contribution and determines the width of central portion of the window function. We thus predict that the central width is just θ_0^{-1} and does not depend on the u (or equivalently, $k_\perp$) value on which the window function is centered.

However, simply characterizing the width of the central peak is insufficient for our purposes. As emphasized throughout this paper, the large dynamic range that exists between the bright foreground emission and the dim cosmological signal means that it is important to accurately capture the weak, low-level wings of the window functions,

away from the central peak. These wings will be controlled by the broader contribution in Eq. (42), namely, the product of the bandpasses. As stated above, the bandpasses have a characteristic width $B_{\text{band}} b/c$, and since baselines of length b probe spatial scales given by $u \sim \nu_0 b/c$, our window function wings will have a width Δu of

$$\Delta u = \frac{B_{\text{band}}}{\nu_0} u. \quad (43)$$

The widths of the window function wings are therefore proportional to u and grow as u increases. Equivalently, since $\Delta u \propto u$, the fractional wing width is constant, and the wings will appear to have the same width on a logarithmic u (or $k_\perp$) scale. This is intuitively unsurprising, as longer baselines probe a greater spread of spatial scale due to their greater chromaticity, and the width of this spread is proportional to the baseline length (see Fig. 2). Since a given u mode is mostly accessed by baselines of length $b \sim u\lambda$, one is then driven to the conclusion that $\Delta u \propto u$.

V. A NUMERICAL MODEL OF A BASIC ESTIMATOR

Having made various approximations in the previous section to enable an analytic treatment of the foreground wedge, we will now discard most of these approximations in lieu of an exact numerical treatment of our basic estimator. We will find that the basic picture that we presented above remains unchanged.

A. Instrument and foreground model

The model instrument that we consider in this paper is intended to reflect a typical design for an interferometer optimized for 21 cm power spectrum (rather than one that is intended to function as a general-purpose low-frequency radio observatory). Maximizing power spectrum sensitivity requires antennas to be placed in a way that yields a large number of short, identical baselines [17,47]. With this in mind, we perform our computations for a square, 20 by 20 array of antennas, with a 14 m spacing between adjacent antennas. Each antenna is assumed to have a Gaussian primary beam, with a full width at half maximum (FWHM) of 40.5° that is approximated as frequency independent in the $B_{\text{band}} = 8$ MHz band that we consider. The frequency width of each individual spectral channel is set at $B_{\text{chan}} = 50$ kHz. With no loss of qualitative generality,

we consider only observations centered around $\nu_0 = 150$ MHz. The formalism presented in this paper applies to all redshifts accessible to a 21 cm interferometer, and none of the "lessons learned" in our analysis are substantially changed by examining a different redshift.

For computational simplicity, we take the tapering function $\phi[(\nu - \nu_0)/B_{\text{band}}]$ of our delay transform [Eq. (3)] to be Gaussian, even though previous studies in the literature have argued for more desirable choices such as Blackman-Harris function or a Blackman-Nuttal function [16,43]. Using a Gaussian allows us to compute analytically compute the ν integral in our measurement equation [Eq. (8)], giving

$$\tilde{V}(\mathbf{b}, \tau) = 2\pi\theta_0^2 B_{\text{band}} \int d^2\mathbf{u} d\eta \tilde{I}(\mathbf{u}, \eta) \exp\left[-\frac{2\pi^2\theta_0^2(u - \nu_0 b/c)^2}{1 + (2\pi\theta_0 b B_{\text{band}}/c)^2}\right] \exp\left[-\frac{2\pi^2 B_{\text{band}}^2(\eta - \tau)^2}{1 + (2\pi\theta_0 b B_{\text{band}}/c)^2}\right]$$

$$\times \exp\left[i2\pi B_{\text{band}}(\eta - \tau)\frac{4\pi^2\theta_0^2(u - \nu_0 b/c)(bB_{\text{band}}/c)}{1 + (2\pi\theta_0 b B_{\text{band}}/c)^2}\right]\frac{\tilde{A}_{b\perp}(\theta_0 v)\tilde{\gamma}(B_{\text{chan}}\eta)}{\sqrt{1 + (2\pi\theta_0 b B_{\text{band}}/c)^2}}e^{i2\pi\nu_0(\eta - \tau)}, \quad (44)$$

where the characteristic scale of the beam θ_0 is given by the standard deviation of our Gaussian beam, which in our case is $\theta_0 = \text{FWHM}/\sqrt{8\ln 2} = 17.2°$. Intuitively, one sees that each delay mode of each baseline probes a reasonably localized region in $uv\eta$ space. One also sees that there exists a complex exponential term that mixes spatial and spectral information, which is to be expected given the chromatic nature of an interferometer's synthesized beam.

Following this, we form the covariance matrix $\mathbf{C}_{ij} \equiv \langle \tilde{V}(\mathbf{b}_i, \tau_i)\tilde{V}^*(\mathbf{b}_j, \tau_j)\rangle$ under the same coherency approximation as the one we employed in the previous section: two baselines may have any amount of overlap on the uv plane in the direction parallel to their baseline vector, but are either completely nonoverlapping or perfectly overlapping in the direction perpendicular to the baseline vector. This allows the integral over v (defined to be the direction on the uv plane perpendicular to a pair of correlated baselines) to be evaluated analytically. The sky signal portion of our covariance matrix is then

$$\mathbf{S}_{ij} = 4\pi^{\frac{5}{4}} B_{\text{band}}^2 \theta_0^4 e^{i2\pi\nu_0(\tau_j - \tau_i)} \int du d\eta \frac{P(u, \eta)}{\sqrt{1 + \alpha_i^2}\sqrt{1 + \alpha_j^2}}\tilde{\gamma}^2\left(B_{\text{chan}}\eta\right)\exp\left[-2\pi^2\theta_0^2\left(\frac{(u - \nu_0 b_i/c)^2}{1 + \alpha_i^2} + \frac{(u - \nu_0 b_j/c)^2}{1 + \alpha_j^2}\right)\right]$$

$$\times \exp\left[-2\pi^2 B_{\text{band}}^2\left(\frac{(\eta - \tau_i)^2}{1 + \alpha_i^2} + \frac{(\eta - \tau_j)^2}{1 + \alpha_j^2}\right)\right]\exp\left[i4\pi^2\left(\alpha_i\frac{(\eta - \tau_i)(u - \nu_0 b_i/c)}{1 + \alpha_i^2} - \alpha_j\frac{(\eta - \tau_j)(u - \nu_0 b_j/c)}{1 + \alpha_j^2}\right)\right], \quad (45)$$

where $\alpha_i \equiv 2\pi\theta_0 B_{\text{band}} b_i/c$ and similarly for α_j. We take the frequency channel response γ to be a Gaussian, which makes its Fourier transform $\tilde{\gamma}$ also a Gaussian. Computing $\mathbf{C}_{,\alpha}$ is very similar to computing $\mathbf{S}$. Since instrumental noise is random and does not depend on the power spectrum, we have $\mathbf{C}_{,\alpha} = \mathbf{S}_{,\alpha}$. To find $\mathbf{C}_{,\alpha}$, then, we simply need to evaluate the integrals in Eq. (45), but with u and η integration limits chosen to match to the band in question, rather than being $-\infty$ and $+\infty$.

Having described our instrument and how it manifests itself in the covariance of our measurements, the final ingredient that we require for our numerical calculations is a model for the total power spectrum $P(u, \eta)$. We model the total power spectrum as the sum of the cosmological power spectrum and a foreground power spectrum. For the cosmological power spectrum, we use the spherically symmetric power spectrum provided in Ref. [56] and assume statistical isotropy to compute the cylindrical power spectrum needed for our covariance. As for the foregrounds, we consider a relatively simple two-component power spectrum model P_{fg}:

$$P_{\text{fg}}(u, \eta) = A(C_{\ell = 2\pi u}^{\text{diff}} e^{-\nu_c^{\text{diff}}|\eta|} + C_{\ell = 2\pi u}^{\text{ps}} e^{-\nu_c^{\text{ps}}|\eta|}), \quad (46)$$

where A is an overall normalization, C_ℓ^{diff} is the angular power spectrum of the diffuse Galactic emission, and ν_c^{diff} is its frequency coherence length. The corresponding quantities for point sources are given by $C_{\ell = 2\pi u}^{\text{ps}}$ and ν_c^{ps}. The η dependence of this parametric form is motivated by the mathematical results of Ref. [26]. In that work, empirically motivated models of foreground spectra were put through an eigenmode analysis, and it was found that the resulting set of eigenmodes were essentially Fourier modes in frequency (i.e., η modes). The eigenvalue spectra were well fit by a linear exponential in η, with a coherence frequency of 64.8 MHz. For simplicity, we adopt this value for both ν_c^{ps} and ν_c^{diff}.

To model the angular structure of the diffuse Galactic emission, we use the GLOBAL SKY MODEL (GSM) software [57] to generate a model of the sky at 150 MHz. We then compute the angular power spectrum of this model, which we find to be well fit by

$$C_\ell^{\text{diff}} \propto \begin{cases} \exp(a_1\ell + a_2\ell^2) & \text{for } \ell \le 8 \\ b_1\ell^{b_2} & \text{for } \ell > 8, \end{cases} \qquad (47)$$

with $a_1 = -1.450$, $a_2 = 0.1003$, $b_1 = 0.7666$, and $b_2 = -2.365$. For notational simplicity, we have omitted the overall normalization of C_ℓ^{diff}, as it can be absorbed into A. We note that while only the high ℓ (power-law) portion of the angular power spectrum is typically modeled in foreground studies, it is crucial to include the low ℓ behavior as well. To see this, note that Eq. (45) predicts a substantial overlap between the response of the shortest baselines and the $u = 0$ mode of the power spectrum. This is simply reflecting the fact that the sky is not infinite in extent. Thus, even though the autocorrelation and zero-spacing baseline products are typically discarded from interferometric data, the instrument may still be sensitive to the zero mode of the sky.

For the point source contribution to foregrounds, we neglect clustering for simplicity, and therefore take $C_{\ell=2\pi u}^{\text{ps}}$ to be a constant. We fix this constant by assuming that at $\ell = 1000$, the amplitude of the point source angular power spectrum is roughly a factor of 10 smaller than that of the diffuse emission [58]. We therefore set $C_{\ell=2\pi u}^{\text{ps}} = 0.1\, C_{\ell=1000}^{\text{diff}}$. Using a simple, ℓ-independent angular power spectrum for point sources is not a required approximation for our formalism, and this assumption can be easily relaxed. In principle, including clustering would boost the point source power at low ℓ modes [59]. In practice, however, the low ℓ regime is dominated by the diffuse Galactic emission anyway, and we do not expect that an inclusion of clustering would qualitatively impact our numerical results.

We fix the overall normalization A of our foregrounds by considering the zero mode of the power spectrum. We require that

$$P(u = 0, \eta = 0) = B_{\text{band}}\theta_0^2\bar{I}_{\text{GSM}}^2, \qquad (48)$$

where $\bar{I}_{\text{GSM}} = 433$ K is the mean temperature of our GSM foreground template.

Finally, we must add the noise contribution to our covariance. The computation of a noise covariance matrix $\mathbf{N}$ is rather subtle, given the assumptions that we have made above regarding baselines that overlap in directions perpendicular to their baseline vectors. We first sort the baselines of our array by baseline length into 54 equally spaced bins. If only one baseline fell into each bin, the noise variance assigned to each bin would be [47]

$$\mathbf{N}_{ii}\bigg|_{\text{single b1}} = \frac{\Omega_{\text{pp}}}{2t}B_{\text{band}}T_{\text{sys}}^2, \qquad (49)$$

where t is the integration time (taken to be 520 h), T_{sys} is the system temperature, and

$$\Omega_{\text{pp}} \equiv \int A^2(\boldsymbol{\theta}/\theta_0)d^2\theta. \qquad (50)$$

Note that our expression for the noise variance differs from equations that are commonly seen in the literature in two ways. First, our variance is proportional to B_{band}. This is simply due to the fact that we are working in a delay basis rather than a frequency basis. In addition, the beam area used here is the integrated *square* of the beam profile (rather than just the integral of the beam profile itself), which was shown in Ref. [16] to be the correct beam area to use for this calculation. With our Gaussian primary beam, $\Omega_{\text{pp}} = \pi\theta_0^2$. We assume a sky-noise dominated instrument and set $T_{\text{sys}} = \bar{I}_{\text{GSM}}$.

We now adjust for the fact that each baseline length bin contains more than just a single baseline. If there are $n(b_i)$ baselines within a particular bin, we simply divide the noise variance by $n(b_i)$, assuming that a combination of instantaneous redundancy and rotation synthesis allows a large fraction of the baselines to be combined coherently, prior to forming a power spectrum. (In practice, this is a somewhat optimistic assumption, given that baselines that are dissimilar in orientation can only be combined statistically, and not coherently.) Further assuming that the noise covariance matrix is diagonal, our final form for $\mathbf{N}$ is

$$\mathbf{N}_{ij} = \frac{1}{n(b_i)}\frac{\Omega_{\text{pp}}}{2t}B_{\text{band}}T_{\text{sys}}^2\delta_{ij}, \qquad (51)$$

where the Kronecker delta function ensures that instrumental noise is uncorrelated both between baselines and between delay bins.

Admittedly, the instrument and noise models presented in this section are quite crude and make use of a large number of simplifying assumptions. The assumptions regarding instrumental noise and rotation synthesis, in particular, are quite optimistic. However, the numerical computations that we perform in this paper are not designed to be definitive sensitivity calculations. Rather, the goal is to use our rough model—which captures the essential features of large-amplitude smooth-spectrum foregrounds and lower-amplitude broadband instrumental noise—to gain some statistically rigorous intuition for the EoR window. For further technical details regarding our exact implementation (e.g., for information about bin sizes), we refer the reader to Appendix C.

B. Window functions and foreground bias

We now examine the statistical properties of our basic power spectrum estimator. We will find that with our basic estimator, the foregrounds appear in a wedge in $k_\perp k_\parallel$ space, consistent with findings in the previous literature.

In Fig. 3, we show the expected foreground bias from Eq. (27) in terms of the cylindrical Fourier coordinates[11] $k_\perp$ and $k_\parallel$. This shows the expected level of additive foreground bias in an estimate of the power spectrum. This foreground bias is the quantity that is most directly comparable to previous studies where a single set of simulated foregrounds are propagated through a power spectrum estimation pipeline. We see that our results are consistent with such studies as well as the analytic arguments of the previous section: the foregrounds are mostly sequestered to low $k_\parallel$ modes at low values of $k_\perp$, rising to higher $k_\parallel$ modes at high $k_\perp$ in a wedgelike pattern. As predicted by Eq. (13), the edge of the wedge is defined by a line with unit logarithmic slope. Beyond this edge, the contamination drops off sharply to give a clean EoR window.

Importantly, we emphasize that the characteristic shapes seen here are not tied to the specifics of our foreground model, beyond the fact that foregrounds are spectrally smooth. Instead, the foreground wedge is tied to the chromatic nature of our instrument and the properties of our basic estimator. While this may be difficult to establish definitively with a simulation, the covariant formalism that we make use of in this paper allows the foreground models to be easily disentangled from the way they interact with the instrument and the power spectrum estimation pipeline. For example, in Fig. 4, we show some examples of window functions on the $k_\perp k_\parallel$ plane. From Eq. (25), we know that these functions depend only on our choice of estimator (through $\mathbf{E}_\alpha$) and the instrument's response (through $\mathbf{C}_{,\beta}$). Thus, any signatures of the wedge that are independent of the foreground model should be apparent in the window functions.

Consider first the leftmost plot from Fig. 4, which shows three window functions that are all centered on the $k_\parallel$ value of 0.25 $h\mathrm{Mpc}^{-1}$, but different $k_\perp$ values. As one moves to higher $k_\perp$, the window functions become increasingly elongated[12] in the $k_\parallel$ direction. With long tails to low $k_\parallel$ (where foreground emission naturally resides), this implies
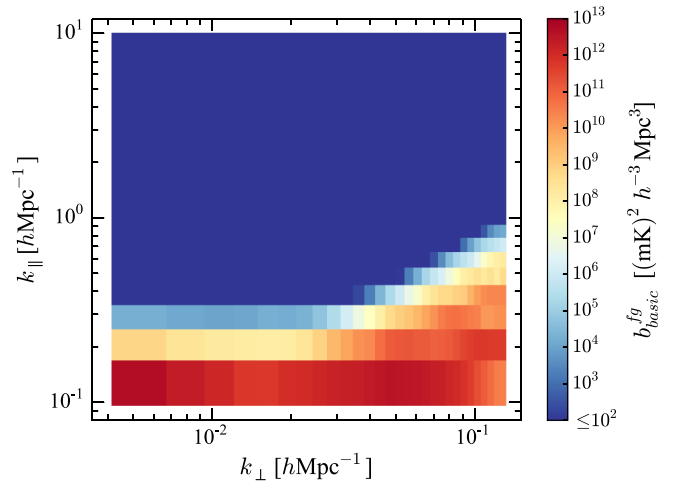

FIG. 3 (color online). Expected foreground bias [Eq. (27)] for the basic estimator defined in Sec. IV, where $\mathbf{E}_\alpha \propto \mathbf{N}^{-1}\mathbf{C}_{,\alpha}\mathbf{N}^{-1}$. For this basic estimator, the chromatic nature of the instrument results in foreground contamination in the form of a characteristic wedge at high $k_\perp$.

a leakage of smooth foregrounds from low to high $k_\parallel$. Since this effect is most pronounced at high $k_\perp$, the result is precisely a wedgelike structure. To guide the eye, the black lines on each plot show the edge of the wedge as predicted[13] by Eq. (13) with $\theta_0 = \pi/2$.

Window functions centered at higher $k_\parallel$ values (central and rightmost plots in Fig. 4) also develop elongations as one moves from low to high $k_\perp$, although the effect is visually subtle due to our logarithmic plotting. These elongations are slightly less important for smooth foregrounds, as even the elongated tails are not quite long enough to reach the lowest $k_\parallel$ for window functions that are centered at high $k_\parallel$. However, such effects may be important if foregrounds turn out to contain unsmooth (high $k_\parallel$) components (see Sec. VI for a brief discussion of this).

Except for at the lowest $k_\perp$ values, the width of the window functions in the $k_\perp$ direction also appears to increase with increasing $k_\perp$. Plotted on the logarithmic axes of Fig. 4, it is visually clear that at intermediate to high $k_\perp$ the window functions have a roughly constant *logarithmic* width, confirming the proportional increase in width with $k_\perp$ that we predicted in Eq. (43).

Our analytic, covariant treatment of power spectrum statistics allows us to compute window functions to the high dynamic range shown in Fig. 4. This is crucial given that foregrounds are expected to be ~10^{10} times brighter in

---

[11]In the mathematical formalism that we have developed so far, we have preferred to use u and η as our Fourier coordinates perpendicular and parallel to the line of sight, respectively. However, to better relate our results to theoretical predictions, we convert to using the more conventional Fourier coordinates $k_\perp$ and $k_\parallel$ when displaying our results. For details about this conversion and our Fourier conventions, please see Appendix A.

[12]This elongation is real and not just an artifact of our logarithmic $k_\perp k_\parallel$ axes. We know this because the set of three windows shown in each plot of Fig. 4 are chosen to be centered on the same $k_\parallel$.

---

[13]Our use of $\theta_0 = \pi/2$ to define the edge of the wedge is somewhat arbitrary, given that there is nothing special about $\theta_0 = \pi/2$ in our flat-sky approximation. However, as emphasized in Ref. [32], in a proper curved-sky treatment it is the natural scale to consider, since the primary beam of an instrument must vanish at the horizon.
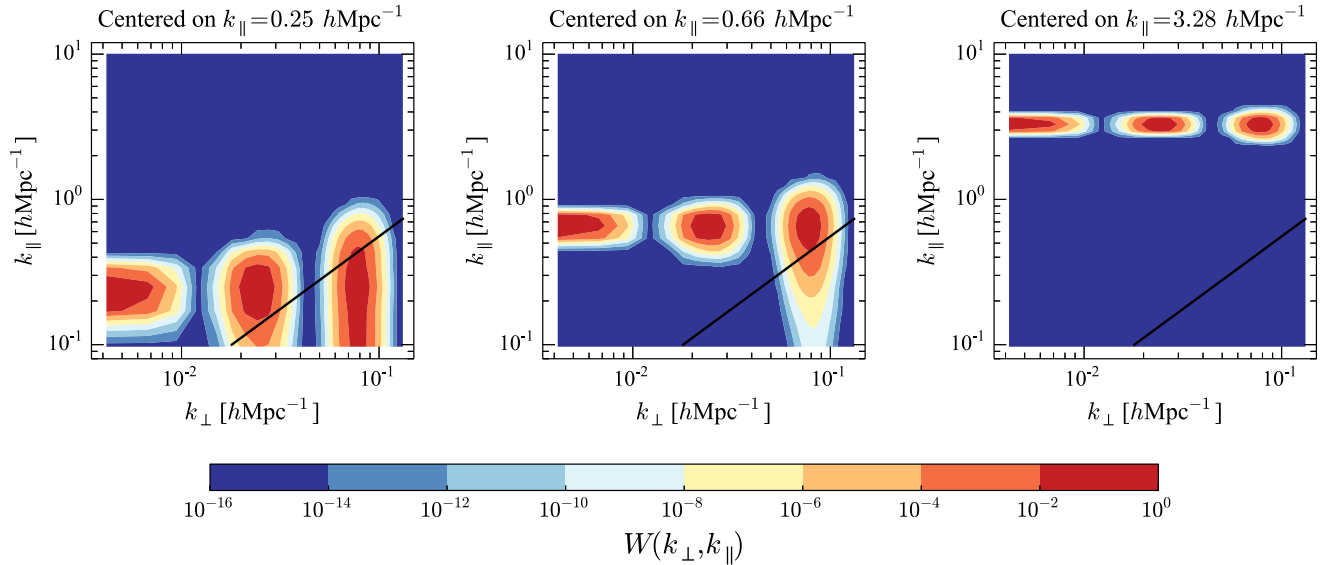
FIG. 4 (color online).   Sample window functions on the $k_\perp k_\parallel$ plane for the basic estimator defined by $\mathbf{E}_\alpha \propto \mathbf{N}^{-1}\mathbf{C}_{,\alpha}\mathbf{N}^{-1}$. The leftmost plot shows a set of window functions centered at $k_\parallel = 0.25\ h\mathrm{Mpc}^{-1}$; the middle plot shows a set centered at $k_\parallel = 0.66\ h\mathrm{Mpc}^{-1}$; the rightmost plot at $k_\parallel = 3.28\ h\mathrm{Mpc}^{-1}$. Within each plot, the window functions are centered, from left to right, at $k_\perp = 0.0042\ h\mathrm{Mpc}^{-1}$, $k_\perp = 0.024\ h\mathrm{Mpc}^{-1}$, and $k_\perp = 0.078\ h\mathrm{Mpc}^{-1}$. The black solid line in each plot indicates the rough extent of the foreground wedge [Eq. (13) with $\theta_0 = \pi/2$]. For window functions centered at high $k_\perp$, there is substantial elongation in the $k_\parallel$ direction, causing higher $k_\parallel$ modes to pick up foreground power, resulting in a foreground-contaminated wedge.

power (i.e., in temperature-squared units) compared to the cosmological signal. It is therefore essential to capture the low-level tails of window functions. Conveniently, once the windows have been computed, the foreground bias for a different foreground power spectrum can be easily determined using Eq. (39a).

C. Error bars and error covariance

In addition to the foreground bias, we may quantify the error covariance in estimates of the power spectrum. In Fig. 5, we show the square root of the diagonal elements of the error covariance matrix [Eq. (23)], i.e., the power spectrum error bars. These error bars capture more than just thermal noise errors and include contributions from the foreground covariance. Indeed, one sees that the foreground wedge appears not just in the bias that we discussed previously, but also in the form of increased error bars. Outside the wedge, the error bars are dominated by thermal noise and are quite low, in what constitutes the EoR window. Note that the thermal noise contribution did not appear in the bias, since we have assumed that cross correlations have eliminated the noise bias. Towards the smallest $k_\perp$, the errors increase by a small amount due to cosmic variance. This effect is typically negligible unless $k_\parallel$ is also small but does play a small role since our model has a rather low noise level. (Again, we emphasize that our goal is not to perform a definitive sensitivity calculation.) At the highest $k_\parallel$, the errors also rise slightly. This is due to the finite spectral resolution of our instrument, which is

self-consistently included in our error bars via the spectral channel profile γ in our formalism.

Beyond just the error bars, our formalism also delivers the off-diagonal elements of the error covariance Σ, which encode the error correlations between different $k_\perp k_\parallel$ cells. As emphasized in Ref. [15], these correlations need to be quantified if one wishes to accurately propagate errors from the cylindrical power spectrum $P(k_\perp, k_\parallel)$ to the spherical
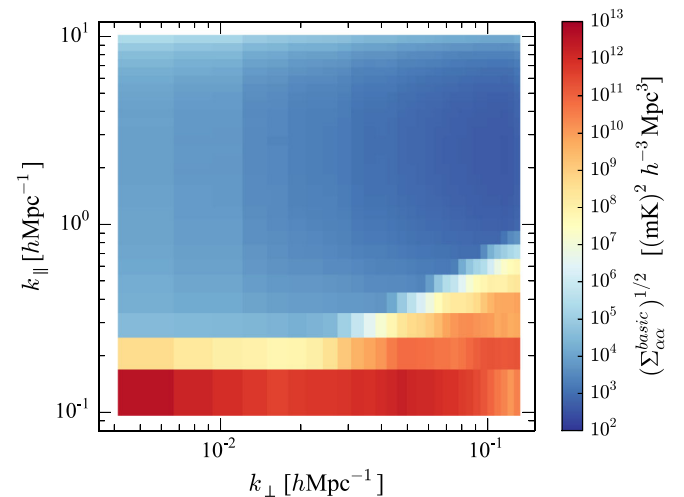

FIG. 5 (color online).   Expected foreground error bars [given by $(\Sigma_{\alpha\alpha})^{1/2}$ for the αth band power or $k_\perp k_\parallel$ cell] for the basic estimator defined by $\mathbf{E}_\alpha \propto \mathbf{N}^{-1}\mathbf{C}_{,\alpha}\mathbf{N}^{-1}$. The foreground wedge also shows up in the error bars of a power spectrum measurement.
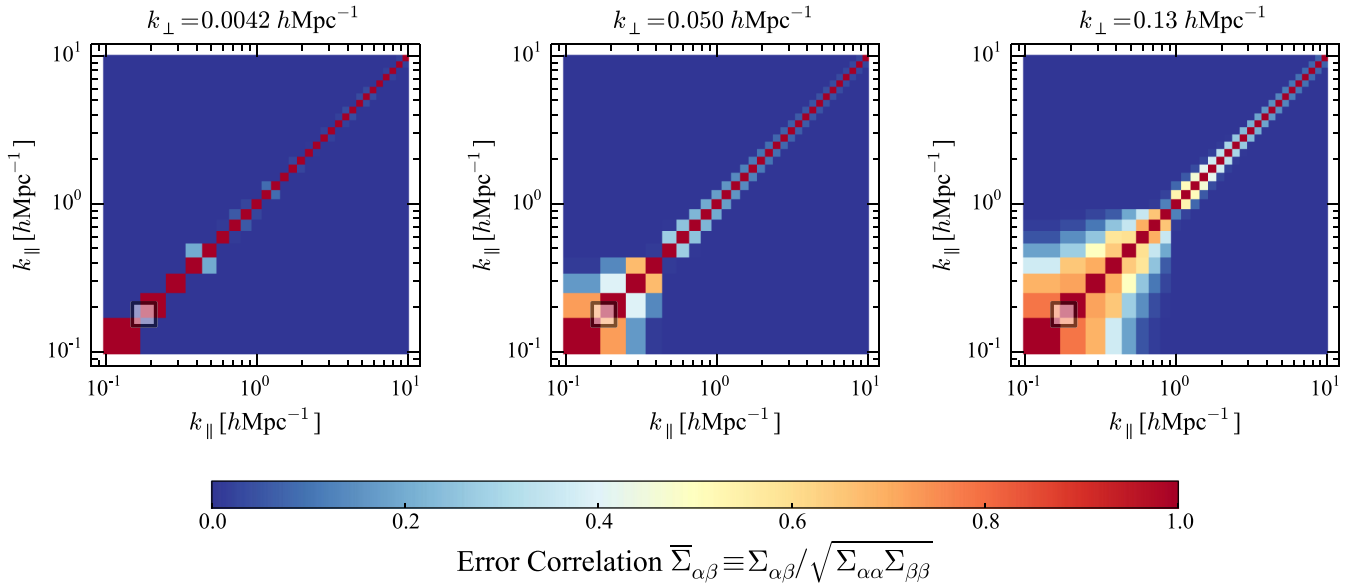
FIG. 6 (color online). Sections of the measurement error correlation matrix [Eq. (52)] for our basic $\mathbf{E}_\alpha \propto \mathbf{N}^{-1}\mathbf{C}_{,\alpha}\mathbf{N}^{-1}$ estimator. Each plot shows the measurement error correlation between all $k_\parallel$ bins for fixed $k_\perp$ (from left to right, $k_\perp = 0.0042\ h\text{Mpc}^{-1}$, $0.050\ h\text{Mpc}^{-1}$, and $0.13\ h\text{Mpc}^{-1}$). It is often assumed in the literature that errors in two $k_\parallel$ cells are uncorrelated if the cells are more than $\Delta k_\parallel \sim 2\pi \frac{H_0 E(z)}{c(1+z)^2} \frac{1}{B_{\text{band}}}$ apart. Overlaid near the bottom left corner of each plot is a semitransparent square of size $\Delta k_\parallel \times \Delta k_\parallel$. Comparing the sizes of these squares to the width of the off-diagonal correlations, one sees that such an assumption holds only at low $k_\perp$. At high $k_\perp$ the errors are highly correlated, reducing the number of independently measurable modes.

power spectrum $P(k)$. To capture this information, we consider the correlation matrix, defined as

$$\bar{\Sigma}_{\alpha\beta} \equiv \frac{\Sigma_{\alpha\beta}}{\sqrt{\Sigma_{\alpha\alpha}\Sigma_{\beta\beta}}}, \tag{52}$$

which is essentially a whitened version of the error covariance. Examining $\bar{\Sigma}$ instead of Σ allows the correlations rather than the larger errors within the wedge to be the dominant feature. To further aid visualization, we focus on just small portions of the correlation matrix. Whereas the full correlation matrix would relate all $k_\perp k_\parallel$ coordinates to all other such coordinates, in Fig. 6 we fix $k_\perp$ at three separate values and consider the correlations between different $k_\parallel$ coordinates.

Immediately obvious is the fact that the error correlations form qualitatively different structures at different values of $k_\perp$. To calibrate our expectations, we include in each plot a semitransparent square of size $\Delta k_\parallel \times \Delta k_\parallel$, where

$$\Delta k_\parallel \sim 2\pi \frac{H_0 E(z)}{c(1+z)^2} \frac{1}{B_{\text{band}}}. \tag{53}$$

This is the scale over which errors are expected to be correlated in the $k_\parallel$ direction. It is derived by making the assumption that for a survey with bandwidth B_{band}, the error correlation scale in η should be roughly $1/B_{\text{band}}$, and expressing this scale in cosmological Fourier coordinates.

In the leftmost plot of Fig. 6, where $k_\perp$ is fixed at the low value of $0.0042\ h\text{Mpc}^{-1}$, we see that our rough expectations are correct. Having chosen our Fourier cell sizes with the survey volume in mind (see Appendix C for details), the cells are seen to be essentially uncorrelated.

This simple picture breaks down, however, as we move towards higher $k_\perp$, where the effects of the foreground wedge are more pronounced. Figure 6 shows that increasing $k_\perp$ (which essentially means moving deeper and deeper into the wedge) causes different $k_\parallel$ cells to become increasingly correlated. To understand why error correlations are to be expected, consider what it would take for the errors to be uncorrelated in $k_\parallel$. Suppose one had an achromatic instrument with noise properties that were uncorrelated and uniform (i.e., white) between all frequency channels. Moving into $k_\parallel$ space by way of a Fourier transform does not induce any error correlations, because uncorrelated white noise has the same statistical properties in all bases. On the other hand, for our measurement we have an inherently chromatic instrument, which makes our noise chromatic. Fourier transforming nonwhite noise will result in noise correlations, even if the noise was uncorrelated to begin with. Additionally, interferometers are chromatic in a very specific way, with longer baselines more chromatic (as illustrated in Fig. 2). Since higher $k_\perp$ are probed by longer baselines, error correlations should increase with $k_\perp$, as seen in Fig. 6.

Viewed together, Figs. 5 and 6 suggest that the characteristic $k_\parallel$ correlation scale coincides roughly with the $k_\parallel$ extent of the wedge. This suggests that the instrumental effects that caused the wedge have also decreased the number of independent measurements, thus decreasing overall signal to noise. (Note, however, that the correlated errors persist even at high $k_\parallel$. Though somewhat challenging to see with our logarithmic axes and hybrid binning, one sees that even there the off-diagonal $k_\parallel$ correlations are greater for large $k_\perp$ modes.) That there is a rough matching of scales between the wedge and our error correlation scale is not entirely surprising (though not given in the *a priori* sense[14]), since that window functions and covariances are closely related to one another. Recall that our window functions exhibited long elongations in $k_\parallel$ as we moved towards higher $k_\perp$. With extremely broad window functions, our instrument essentially smoothed over a large number of modes, and it is unsurprising that the errors in nearby bins ended up being positively correlated.

We can quantify the error correlations in more detail by defining an effective number of independent cells $N_{\rm eff}$, where

$$N_{\rm eff} \equiv \frac{N_c^2}{\sum_{\alpha,\beta}^{N_c} \bar{\Sigma}_{\alpha\beta}}, \qquad (54)$$

with N_c being the number of Fourier cells that enter into a simple unweighted averaging of Fourier modes. By construction, if all the modes that one averages over are independent, $N_{\rm eff}$ equals N_c, whereas $N_{\rm eff} = 1$ if the modes are perfectly correlated. In Fig. 7, we consider the effect of averaging Fourier modes along the $k_\parallel$ axis and show $N_{\rm eff}$ as a function of N_c, with N_c increasing from unity (when only the lowest $k_\parallel$ mode is included in the average) to $N_c = 30$ (the total number of $k_\parallel$ cells in our computation). We do this for two different constant $k_\perp$ slices on the Fourier plane, $k_\perp = 0.0042 \ h{\rm Mpc}^{-1}$ and $0.13 \ h{\rm Mpc}^{-1}$ (corresponding to the leftmost and rightmost plots of Fig. 6, respectively). For low $k_\perp$, one sees the linear relationship $N_{\rm eff} \approx N_c$ regardless of how many $k_\parallel$ cells are included in the averaging. For high $k_\perp$, however, $N_{\rm eff}$ increases only very slowly at first, as one averages together the highly correlated modes within the wedge. The increase in $N_{\rm eff}$ is linear only at higher $k_\parallel$, where our hybrid binning becomes logarithmic, and each cell encompasses a greater extent in $k_\parallel$, eventually exceeding the error correlation length. In our

FIG. 7 (color online). Effective number of independent cells $N_{\rm eff}$ as a function of N_c, the number of Fourier cells included in an averaging of $k_\parallel$ modes (going from low to high $k_\parallel$). The approximately linear relationship seen for the low $k_\perp$ curve is indicative of uncorrelated modes. On the other hand, the high $k_\perp$ curve approaches linearity only when high $k_\parallel$ modes dominate the average. Its initially slow increase at low $k_\parallel$ shows that the modes are highly correlated. Thus, for power spectrum sensitivity calculations to be reliable, one must take error correlations into account or risk overestimating an instrument's sensitivity.

formalism, the correlated errors within each of these larger cells have already been self-consistently averaged over, giving independent cells that contribute linearly to $N_{\rm eff}$. Regardless of the specific of one's power spectrum estimation formalism, it is crucial to take into account error correlations if averaging together bins that are narrower than the error correlation length, or if the bins are wider than this, to ensure that the implicit averaging performed within the bin is done correctly [rather than relying on possibly incorrect *a priori* assumptions such as that implied by Eq. (53)].

For the particular computational setup in this paper, one sees about a factor of 2 reduction in $N_{\rm eff}$ for the highest $k_\perp$. Going to even higher $k_\perp$ causes even greater reductions in sensitivity (compared to simple expectations). The correlations discussed here are particularly important for experiments proposing to make measurements deep within the wedge. As discussed above, the extent of the wedge in $k_\parallel$ provides the characteristic error correlation length between different $k_\parallel$ cells. It therefore follows that errors are highly correlated whenever one chooses to work within the wedge. Though such measurements may be well motivated by the fact that the cosmological-signal-to-thermal-noise ratio is largest at low k, previous studies that established this have typically assumed the correlation length given by Eq. (53).

---

[14]While window functions and error covariances are related to one another, they are conceptually separate entities, and window function widths do not in general coincide perfectly with error correlation lengths. In paper II, for example, we will see an example of an estimator with completely uncorrelated errors, but whose window functions continue to have a nonzero width. It is also possible to write down estimators that artificially force the window functions to be delta functions, with the corresponding errors becoming *anticorrelated* [15,35].
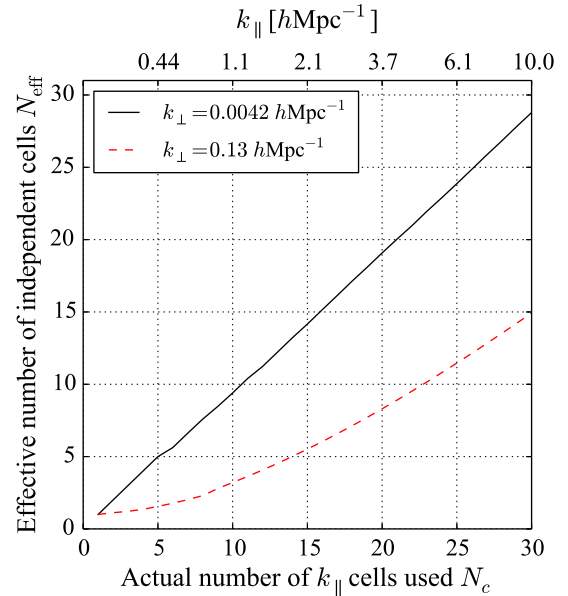
The resulting signal-to-noise ratios may therefore have been overestimated.

We emphasize that the error correlations seen in this section exist whether or not one's power spectrum estimation pipeline includes a direct subtraction of modeled foregrounds from the input data. To see this, suppose a direct foreground subtraction scheme reduces the foreground covariance by some constant multiplicative constant, so that $\mathbf{C}_{\mathrm{fg}} \rightarrow \varepsilon \mathbf{C}_{\mathrm{fg}}$, where $0 < \varepsilon \leq 1$. If one assumes that thermal noise is negligible compared to foreground residuals at low k after a long time integration, the result of our reduced $\mathbf{C}_{\mathrm{fg}}$ will be a corresponding decrease in the amplitude of the final bias and the error bars. However, the plots of error correlation in Fig. 6 will remain unchanged, since the correlation is insensitive to an overall scaling. The number of independent modes will therefore still decrease in the manner discussed above, reducing sensitivity. Since this loss of sensitivity is the most pronounced at high $k_\perp$, it is particularly important to take into account for arrays that make use of long baselines, such as LOFAR or GMRT.

VI. UNSMOOTH FOREGROUNDS?

In many prior works on 21 cm cosmology, the assumption of spectrally smooth foregrounds is considered crucial to one's ability to perform foreground subtraction. At various points in this paper, we too have assumed that foregrounds are smooth, and incorporating this assumption into our general framework gave rise to various predictions, such as the existence of the EoR window. However, since the power spectrum estimation framework itself does not require smooth foregrounds, a number of our key results would survive a (hypothetical) discovery of unsmooth foreground sources. We now briefly discuss how the problem of foreground mitigation would change if such an unfortunate discovery were to be made.

Because the window functions encode only the mapping of the true power spectrum to the estimated power spectrum and do not depend on the actual power spectrum, they do not rely on the assumption of smoothness. Irrespective of whether the foregrounds are smooth or not, the window functions accurately describe how foreground power is smeared out on the $k_\perp k_\parallel$ plane by our instrument. If the foregrounds are smooth, their influence is limited to the wedge. This makes foreground mitigation easy, as their avoidance requires no more than a simple cut on the Fourier plane. If the foregrounds are not as smooth as expected, the EoR window will be smaller, but its exact size can still be predicted by convolving the (now unsmooth) model of our foregrounds with the same window functions as before. Forthcoming data from various experiments at higher sensitivity will allow further foreground modeling and—with the help of the window functions—an accurate determination of the extent of the foreground wedge. Encouragingly, recent theoretical calculations have shown that in physically motivated models of synchrotron emission, foreground spectra tend to be smooth even under the most pessimistic of assumptions [60].

For the sake of argument, however, let us consider a worst-case scenario where foregrounds are discovered to be sufficiently unsmooth for the EoR window to be drastically reduced in size. In such a scenario, a number of strategies can be employed for foreground subtraction. First, foregrounds can be modeled and subtracted to the best of one's ability in the visibility data. Following that, a more sophisticated estimator (one that downweights the data not by $\mathbf{N}$ but by the total covariance $\mathbf{C}$ to account for uncertainty in the foregrounds) can be used. Finally, the foreground bias can be subtracted from the power spectrum using Eq. (39a), and window function decorrelation techniques can also be used in an attempt to increase the size of the EoR window. We explore a number of these techniques in paper II [46].

VII. CONCLUSIONS

In any measurement of the redshifted 21 cm power spectrum, foreground contamination is a serious concern. Fortunately, observations and various theoretical studies have shown that despite complications arising from the inherently chromatic nature of an interferometric measurement, smooth spectrum foregrounds occupy a characteristic wedge region in cylindrical $k_\perp k_\parallel$ Fourier space. The complement of this region is expected to be relatively foreground-free, forming an EoR window where measurements might be made.

While there exists an extensive literature on the topic, previous studies have typically focused on how the foreground wedge manifests itself in the mean power spectrum signal. However, the same physical effects that cause the wedge in the power spectrum also affect the associated error statistics, such as the error covariance and the window functions. An examination of some of these statistics was performed in Ref. [42] using Monte Carlo methods. In this paper, we have provided a complementary treatment by deriving a rigorous, fully covariant mathematical description of the foreground wedge and the EoR window. While our methods require the numerical *evaluation* of some matrix expressions, they differ from previous work in that they do not require numerical *simulations* of interferometric measurements, since the underlying framework is largely analytic. This makes it possible to compute error statistics with very high dynamic range, which is crucial since the foregrounds are expected to dwarf both the instrumental noise and cosmological signal.

Our formalism takes advantage of the delay spectrum techniques introduced in Ref. [32] to achieve computational savings, and in fact it is the use of the delay basis that makes our covariant, high dynamic range calculations numerically feasible. However, we reemphasize that this is merely a choice of basis, and that our results are independent of this choice. This was shown explicitly in

Sec. IV, when we developed a description of the foreground wedge in terms of window functions. Our description decouples the causes of the wedge—which depend only on the chromatic nature of the instrument and the specific form of our power spectrum estimator—from the detailed nature of the foreground emission. Independent of foreground properties, window functions that are centered at high $k_\perp$ will typically develop long tails towards low $k_\parallel$. The wedge then results from the additional assumption that foregrounds are spectrally smooth, so that strong signals from low $k_\parallel$ are transferred to higher $k_\parallel$ by the long tails. Once the window functions have been computed, however, our formalism allows such assumptions to be relaxed.

With a fully covariant framework, we are able to track all error correlations in our numerical computations. We find that measurements made at high $k_\perp$ have highly correlated errors, effectively reducing the number of independent measurements that can be made in that part of Fourier space. This is particularly important for sensitivity forecasts that rely heavily on measurements made within the wedge, since the wedge's extent in Fourier space is roughly on the same scale as that of the error correlations. Previous studies have typically neglected error correlations, assuming that errors are independent as long as the spatial Fourier cells are of the same size as an antenna's uv footprint, and the spectral Fourier cells are on the order of $1/B_{\text{band}}$. Our work suggests that this is likely to be too optimistic an assumption. At the highest $k_\perp$ considered in our numerical computations $(k_\perp = 0.13\ h\text{Mpc}^{-1})$, for example, error correlations reduce the number of independent modes by approximately a factor of 2. This effect will be even more pronounced at even higher $k_\perp$, which are probed by experiments with extremely long baselines. Since the chromatic effects that caused the wedge are closely related to those that cause error correlations, it will be crucial in future research to address the question of exactly how far the wedge can be pushed back (or equivalently, how much one can expand the EoR window). In paper II, we use the formalism of this paper to explore statistical methods for enlarging the EoR window [46].

In this paper, our goal was to provide a rigorous treatment of the wedge. Previous treatments have typically made different simplifying assumptions. These include neglecting partially redundant baselines, approximating delay modes as η modes, making assumptions about baseline length, assuming top-hat primary beams, neglecting binning artifacts, or assuming that errors are uncorrelated on the Fourier plane. Our framework discards all of these approximations simultaneously, and it is gratifying to see that the basic picture of the EoR window as a naturally foreground-free region of Fourier space remains unchanged. This bodes well for foreground avoidance efforts that aim to detect the EoR by working outside the wedge, making it possible for 21 cm cosmology to open a new window into the high redshift universe using only

existing data analysis techniques, with even more transformative results possible with further advances that expand the EoR window.

APPENDIX A: FOURIER CONVENTIONS

In this Appendix, we define our Fourier conventions. In sections where we establish formalism, we typically use a Fourier convention with factors of 2π in the exponent, so that the sky $I(\boldsymbol{\theta}, \nu)$ and its Fourier transform $\tilde{I}(\mathbf{u}, \eta)$ are related by

$$\tilde{I}(\mathbf{u}, \eta) = \int_{-\infty}^{\infty} I(\boldsymbol{\theta}, \nu) e^{-i2\pi(\mathbf{u}\cdot\boldsymbol{\theta}+\eta\nu)} d^2\boldsymbol{\theta} d\nu \qquad (A1)$$

and

$$I(\boldsymbol{\theta}, \nu) = \int_{-\infty}^{\infty} \tilde{I}(\mathbf{u}, \eta) e^{i2\pi(\mathbf{u}\cdot\boldsymbol{\theta}+\eta\nu)} d^2\mathbf{u} d\eta, \qquad (A2)$$

where $\mathbf{u}$ is the Fourier dual to $\boldsymbol{\theta}$ and η is the Fourier dual to ν. Correspondingly, the power spectrum $P(\mathbf{u}, \eta)$ is defined as

$$\langle \tilde{I}(\mathbf{u}, \eta)\tilde{I}^*(\mathbf{u}', \eta')\rangle \equiv P(\mathbf{u}, \eta)\delta(\mathbf{u} - \mathbf{u}')\delta(\eta - \eta'). \qquad (A3)$$

Adopting this Fourier convention is convenient for developing a mathematical description of the foreground wedge because the Fourier transform closely mimics the definition of a visibility [Eq. (1)].

Theoretical studies, however, typically use different coordinates and a different Fourier convention. In cosmological coordinates, the Fourier transform is given by

$$\tilde{T}(\mathbf{k}_\perp, k_\parallel) = \int_{-\infty}^{\infty} T(\mathbf{r}_\perp, r_\parallel) e^{-i(\mathbf{k}_\perp \cdot \mathbf{r}_\perp + k_\parallel r_\parallel)} d^2\mathbf{r}_\perp dr_\parallel, \quad (A4)$$

and the inverse transform is given by

$$T(\mathbf{r}_\perp, r_\parallel) = \int_{-\infty}^{\infty} \tilde{T}(\mathbf{k}_\perp, k_\parallel) e^{i(\mathbf{k}_\perp \cdot \mathbf{r}_\perp + k_\parallel r_\parallel)} \frac{d^2\mathbf{k}_\perp dk_\parallel}{(2\pi)^3}, \quad (A5)$$

where $T(\mathbf{r}_\perp, r_\parallel)$ is the sky temperature at comoving position $\mathbf{r}_\perp$ perpendicular to the line of sight and $r_\parallel$ parallel to the line of sight. The power spectrum $\bar{P}$ is defined as

$$\langle \tilde{T}(\mathbf{k}_\perp, k_\parallel) \tilde{T}^*(\mathbf{k}'_\perp, k'_\parallel) \rangle = (2\pi)^3 \bar{P}(\mathbf{k}_\perp, k_\parallel) \delta(\mathbf{k} - \mathbf{k}'), \quad (A6)$$

where $\mathbf{k} \equiv (\mathbf{k}_\perp, k_\parallel)$.

Because the angular positions and frequencies can be mapped to transverse and line-of-sight comoving distances, respectively, the two Fourier conventions can be related to one another. Defining $\mathbf{r}_\perp$ as the comoving distance perpendicular to the line of sight, we have

$$\mathbf{r}_\perp = D_c \boldsymbol{\theta}, \quad (A7)$$

with

$$D_c \equiv \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')}; \qquad E(z) \equiv \sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}, \quad (A8)$$

where c is the speed of light, z is the redshift of observation, H_0 is the Hubble parameter, Ω_m is the normalized matter density, and Ω_Λ is the normalized dark energy density. The line-of-sight direction is more subtle. When forming a power spectrum, one typically includes only data from a relatively narrow range in redshift. Otherwise, cosmological evolution invalidates the assumption of a translation-invariant temperature field, which is needed in the definition of a power spectrum. What matters, then, is not the mapping between frequency and the total comoving line-of-sight distance, but rather, the *local* relation between differences in frequency $\Delta\nu$ and differences in distance $\Delta r_\parallel$ at the redshift of observation:

$$\Delta r_\parallel = \frac{c}{H_0\nu_{21}} \frac{(1+z)^2}{E(z)} \Delta\nu, \quad (A9)$$

where $\nu_{21} \equiv 1420$ MHz is the rest frequency of the 21 cm line. Having established this mapping we may (in what is perhaps an abuse of notation) recenter our coordinates so that $\Delta\nu \to \nu$ and $\Delta r_\parallel \to r_\parallel$. Such a recentering introduces a constant phase shift in our Fourier transforms, which has no bearing on quadratic statistics such as the power spectrum.

Making the identification between $T(\mathbf{r}_\perp, r_\parallel)$ and $I(\boldsymbol{\theta}, \nu)$ using our mappings, we may compare Eqs. (A1) and (A4) to conclude that

$$\mathbf{k}_\perp = \frac{2\pi\mathbf{u}}{D_c}; \qquad k_\parallel = \frac{2\pi\nu_{21}H_0E(z)}{c(1+z)^2}\eta. \quad (A10)$$

This allows the power spectra defined under the different Fourier conventions to be related to one another:

$$\bar{P}(\mathbf{k}_\perp, k_\parallel) = \frac{c(1+z)^2 D_c^2}{\nu_{21}H_0E(z)} P(\mathbf{u}, \eta). \quad (A11)$$

In this paper, we will plot all numerical results in terms of cosmological Fourier coordinates $\mathbf{k}_\perp$ and $k_\parallel$, even though we perform all our computations in terms of $\mathbf{u}$ and η. We use WMAP9 cosmological parameters: $\Omega_m = 0.28$, $\Omega_\Lambda = 0.72$, and $H_0 = 69.7 \frac{\text{km/s}}{\text{Mpc}}$ [1].

APPENDIX B: EQUIVALENCE OF GRIDDED AND VISIBILITY-BASED APPROACHES TO BASIC POWER SPECTRUM ESTIMATION

In this Appendix, we prove that estimating the power spectrum directly from visibilities using the basic estimator of Sec. IV is equivalent to first gridding the visibilities in $uv\eta$ space and estimating the power spectrum by squaring and binning the results.

In Sec. IV we considered a quadratic power spectrum estimator of the form

$$\hat{p}_\alpha \propto \mathbf{x}^\dagger \mathbf{N}^{-1} \mathbf{C}_{,\alpha} \mathbf{N}^{-1} \mathbf{x}, \quad (B1)$$

where $\mathbf{N}$ is the instrumental noise covariance matrix, $\mathbf{x}$ is the data vector (containing all spectral information from all baselines), and $\mathbf{C}_{,\alpha} \equiv \partial\mathbf{C}/\partial p_\alpha$ is the response of the covariance matrix to the αth band power. Working in the limit of continuous band powers and differentiating Eq. (16) with respect to $P(u_\alpha, \eta_\alpha)$ gives

$$(\mathbf{C}_{,\alpha})_{ij} \equiv \frac{\partial C_{ij}}{\partial P(u_\alpha, \eta_\alpha)} = h(u_\alpha, \eta_\alpha; b_i, \tau_i) h^*(u_\alpha, \eta_\alpha; b_j, \tau_j). \quad (B2)$$

This allows us to simplify our expression for the power spectrum:

$$\hat{p}_\alpha \propto |\mathbf{h}^{\alpha\dagger} \mathbf{N}^{-1} \mathbf{x}|^2, \quad (B3)$$

where $\mathbf{h}_i^\alpha \equiv h(u_\alpha, \eta_\alpha; b_i, \tau_i)$. If we imagine writing each $\mathbf{h}^\alpha$ vector as a column of a larger $\mathbf{H}$ matrix, the result can be compactly rewritten as

$$\hat{\mathbf{p}}_\alpha \propto |(\mathbf{H}^\dagger \mathbf{N}^{-1} \mathbf{x})_\alpha|^2, \qquad (B4)$$

where we have similarly grouped the band powers into a vector $\hat{\mathbf{p}}$.

We now show that essentially the same estimator results if one first uses the visibilities to form a (Fourier space) map of the sky, which is then squared to form the power spectrum. Comparing Eqs. (6) and (17), we see that our measurement equation can be written as

$$\mathbf{x} = \mathbf{Hs} + \mathbf{n}, \qquad (B5)$$

where $\mathbf{x}$ is our data vector like before, $\mathbf{n}$ is the instrumental noise contribution to the data (with covariance $\langle \mathbf{nn}^\dagger \rangle$ given by the instrumental noise covariance $\mathbf{N}$ discussed above), and $\mathbf{s}$ is a discretized "map" of the Fourier sky, with elements given by $\tilde{I}(\mathbf{u}, \eta)$ evaluated at predefined grid points.

In the generalized mapmaking problem, one seeks to use the data $\mathbf{x}$ to form an estimator $\hat{\mathbf{s}}$ of the true sky $\mathbf{s}$. This is accomplished in a lossless manner [54] by the estimator

$$\hat{\mathbf{s}} = \mathbf{R}\mathbf{H}^\dagger \mathbf{N}^{-1} \mathbf{x}, \qquad (B6)$$

where $\mathbf{R}$ is an invertible matrix. As discussed in Ref. [55] the combination $\mathbf{H}^\dagger \mathbf{N}^{-1} \mathbf{x}$ constitutes a dirty map, and the role of $\mathbf{R}$ is to normalize and/or deconvolve this map. If $\mathbf{R}$ is taken to be diagonal, then its role is merely one of normalization. More complicated forms for $\mathbf{R}$ mix different pixels of the dirty map, in principle allowing the dirty map to be deconvolved. For example, the choice $\mathbf{R} \equiv [\mathbf{H}^\dagger \mathbf{N}^{-1} \mathbf{H}]^{-1}$ deconvolves the instrumental beam perfectly, giving an estimator with the property $\langle \hat{\mathbf{s}} \rangle = \mathbf{s}$ (so that each pixel in the estimated map is on average probing only the corresponding pixel in the true map, rather than a linear combination of pixels in the pattern of a remaining point-spread function).

Suppose one forgoes deconvolution by picking $\mathbf{R}_{ij} \equiv r_i \delta_{ij}$ and then proceeds to form an estimate of the power spectrum by squaring the complex magnitude of the resulting map estimator $\hat{\mathbf{s}}$. The result is

$$|\hat{\mathbf{s}}_i|^2 = r_i^2 |(\mathbf{H}^\dagger \mathbf{N}^{-1} \mathbf{x})_i|^2 \propto \hat{\mathbf{p}}_i, \qquad (B7)$$

thereby proving that the basic power spectrum estimator that we examined in Sec. IV is equivalent to one where one forms a $uv\eta$ space dirty map from visibilities and then squares the result.

We stress that the proof that we have just presented is basis independent, in the sense that our data vector need not be indexed by baseline and delay. For example, one may choose to deal with frequency spectra rather than delay

spectra, in which case the data vector would be indexed by baseline and frequency channel. The resulting $\mathbf{h}^\alpha$ vectors would no longer be given by Eq. (17), but the proof shown here would be unchanged.

Crucially, the proof shown here assumed an infinitely finely discretized Fourier space. In practice, this will only be a good approximation on small scales (large Fourier wave numbers), where the difference in wave number Δk between neighboring discretized bins is small compared to the magnitudes of the wave numbers k themselves. In paper II we will formalize this assumption by importing the Feldman-Kaiser-Peacock approximation that is commonly used in galaxy surveys.

Importantly, we emphasize that while squaring a normalized $uv\eta$ dirty map is a perfectly reasonable way to estimate a power spectrum, it is by no means optimal. Indeed, Eqs. (B4) and (B7) are provably nonoptimal, and better estimators are explored in paper II. Instead, the estimators considered here (and therefore in Sec. IV) are intended to be representative of simple, "first pass" methods [43,44,53], and their statistical properties provide basic pictures of the challenges that one faces.

APPENDIX C: TECHNICAL COMPUTATIONAL DETAILS

In this Appendix, we provide further details pertaining to the numerical computations described in Sec. V.

1. Binning

In the quadratic estimator formalism used in this paper, there are two sets of discretizations: a discretization of the input data and a discretization of the output power spectra.

For all the computations performed here, we have chosen to express the input data in a basis parameterized by baselines and delay. In other words, each component of the input data vector $\mathbf{x}$ corresponds to a different baseline and delay pair. For computational tractability, we bin baselines together into 50 linear bins of width 5 m, with the first bin centered about 10 m and the last bin centered about 255 m. (There are in fact some slightly longer baselines in our array. However, we discard them in our analysis for reasons of numerical stability.) The delay bins are again linear and range from -200 to 198.75 μs in 320 bins. This gives delay increments of 0.125 μs, equal to the natural bin size of $1/B_{\text{band}}$. Note that this equality is by no means a requirement. If computational resources are not a concern, it may be preferable to slightly overresolve. This is perfectly legitimate despite the fact that the bins will no longer be independent, since the inclusion of the channel profile γ and our tracking of full covariance information allows the nonindependence to be self-consistently captured.

Given that a baseline of length b roughly probes modes with $u \sim b\nu_0/c$, and that a delay bin τ roughly probes $\eta \sim \tau$, a logical way to discretize the output $u\eta$ space would be to

use linear bins that matched the input bins (but with u bins scaled by an appropriate factor of ν_0/c). Such a scheme would be the most appropriate for matching the specifications of the instrument. However, the cosmological power spectrum is expected to evolve on logarithmic k scales. Thus, a linear binning is computationally wasteful at high k, where a large number of bins are used to resolve a power spectrum that does not evolve very much. On the other hand, a logarithmic binning scheme that is appropriate at high k will tend to be computationally wasteful at low k, where one would be overresolving the instrumental response. As a compromise, we use a hybrid binning scheme that is roughly linear at low k and roughly logarithmic at high k. In this scheme, the $(n+1)$th boundary of the u bins u_{n+1} is given by

$$u_{n+1} = 1.036u_n + 2.5. \qquad (C1)$$

Similarly, the $(n+1)$th boundary of the η bins η_{n+1} is given by

$$\eta_{n+1} = 1.095\eta_n + 0.125 \ \mu\text{s}. \qquad (C2)$$

At low u and low η the additive terms dominate, yielding bin boundaries that are spaced in an approximately linear fashion well suited to the instrumental specifications. At high u and η the multiplicative terms dominate, giving logarithmic bins that are a good fit for theoretical expectations. For both u and η, we use 30 bins, giving a total of 900 $u\eta$ band powers. The bottom edge of the lowest u bin is at $u = 3$, while the bottom edge of the lowest η bin is at $\eta = 0.12 \ \mu\text{s}$.

2. Sparseness and computational shortcuts

The methods and computations presented in this paper are basis independent. By this, we mean that while our final goal is to estimate a power spectrum (and its associated error statistics) on the $k_\perp k_\parallel$ plane, our input data may be expressed in any basis that we find convenient. We now elaborate on our reasons for working in a baseline or delay basis.

As an example, consider the evaluation of **C**. With the Gaussian beams and tapering functions used in Sec. V, the covariance **C** is given by Eq. (45). For parts of the matrix corresponding to short baselines, one can see by inspection that the matrix will be diagonal dominant, with a large number of off-diagonal elements that are close to zero. The $\mathbf{C}_{,\alpha}$ matrices are even more sparse, since many of the diagonal elements (those that do not satisfy $u \approx \nu_0 b_i/c$ or $\eta \approx \tau_i$) will also be zero. In our computations, we skip the evaluation of matrix elements that are expected to be small. This represents significant savings in computation time, given that with our binning scheme each matrix measures 16000×16000, and each element requires numerically integrating a two-dimensional integral given by Eq. (45). Moreover, with 900 band powers, this process must be repeated 900 times for each of the $\mathbf{C}_{,\alpha}$ matrices.

In our implementation, we set off-diagonal matrix elements of **C** to zero if the integrand is suppressed by at least 10^{-12} relative to the relevant diagonal elements everywhere over the integration volume. For the $\mathbf{C}_{,\alpha}$ matrices, we apply the additional constraint that a diagonal element is to be skipped if the integrand is attenuated by 10^{-12} or more compared to a different diagonal element that satisfies $u_\alpha \approx \nu_0 b/c$ and $\eta_\alpha \approx \tau$, where u_α and η_α are the $u\eta$ values corresponding to the αth band.

The sparseness that we have described here is a direct result of our using a baseline or delay basis. In contrast, parameterizing the spectral information in a frequency basis results in substantially denser matrices, since the data are highly correlated between frequencies [42]. The delay transform roughly isolates spectral information by η mode. This isolation is imperfect in the long-baseline limit, as we saw in Eq. (10). The matrices are therefore still dense for elements corresponding to long baselines, but the sparseness that is available with short baselines provides enough savings to enable the full propagation of covariant information without resorting to Monte Carlo methods, which can sometimes be slow to converge to the dynamic range displayed in Sec. V. Finally, we note that in an application of our methods to real measurements, the delay transform operates individually on each baseline [32] and therefore can be applied with negligible computational cost to the input data.

[1] G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R. Nolta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright, Astrophys. J. Suppl. Ser. **208**, 19 (2013).

[2] P. A. R. Ade *et al.* (Planck Collaboration), arXiv:1303.5062 [Astron. Astrophys. (to be published)].

[3] C. P. Ahn *et al.*, Astrophys. J. Suppl. Ser. **211**, 17 (2014).

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

[5] M. F. Morales and J. S. B. Wyithe, Annu. Rev. Astron. Astrophys. **48**, 127 (2010).

[6] J. R. Pritchard and A. Loeb, Rep. Prog. Phys. **75**, 086901 (2012).

[7] A. Loeb and S. R. Furlanetto, *The First Galaxies in the Universe* (Princeton University, Princeton, NJ, 2013).

[8] M. McQuinn, O. Zahn, M. Zaldarriaga, L. Hernquist, and S. R. Furlanetto, Astrophys. J. **653**, 815 (2006).

[9] J. D. Bowman, M. F. Morales, and J. N. Hewitt, Astrophys. J. **661**, 1 (2007).

[10] Y. Mao, M. Tegmark, M. McQuinn, M. Zaldarriaga, and O. Zahn, Phys. Rev. D **78**, 023529 (2008).

[11] G. Paciga, J. G. Albert, K. Bandura, T.-C. Chang, Y. Gupta, C. Hirata, J. Odegova, U.-L. Pen, J. B. Peterson, J. Roy, J. R. Shaw, K. Sigurdson, and T. Voytek, Mon. Not. R. Astron. Soc. **433**, 639 (2013).

[12] S. Yatawatta *et al.*, Astron. Astrophys. **550**, A136 (2013).

[13] S. J. Tingay *et al.*, Proceedings of the Astronomical Society of Australia **30**, e007 (2013).

[14] 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).

[15] J. S. Dillon, A. Liu, C. L. Williams, J. N. Hewitt, M. Tegmark, E. H. Morgan, A. M. Levine, M. F. Morales, S. J. Tingay, G. Bernardi, J. D. Bowman, F. H. Briggs, R. C. Cappallo, D. Emrich, D. A. Mitchell, D. Oberoi, T. Prabu, R. Wayth, and R. L. Webster, Phys. Rev. D **89**, 023002 (2014).

[16] A. R. Parsons, A. Liu, J. E. Aguirre, Z. S. Ali, R. F. Bradley, C. L. Carilli, D. R. DeBoer, M. R. Dexter, N. E. Gugliucci, D. C. Jacobs, P. Klima, D. H. E. MacMahon, J. R. Manley, D. F. Moore, J. C. Pober, I. I. Stefan, and W. P. Walbrugh, Astrophys. J. **788**, 106 (2014).

[17] 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).

[18] G. Mellema *et al.*, Exp. Astron. **36**, 235 (2013).

[19] A. Liu, J. R. Pritchard, M. Tegmark, and A. Loeb, Phys. Rev. D **87**, 043002 (2013).

[20] X. Wang, M. Tegmark, M. G. Santos, and L. Knox, Astrophys. J. **650**, 529 (2006).

[21] A. Liu, M. Tegmark, and M. Zaldarriaga, Mon. Not. R. Astron. Soc. **394**, 1575 (2009).

[22] J. D. Bowman, M. F. Morales, and J. N. Hewitt, Astrophys. J. **695**, 183 (2009).

[23] A. Liu, M. Tegmark, J. Bowman, J. Hewitt, and M. Zaldarriaga, Mon. Not. R. Astron. Soc. **398**, 401 (2009).

[24] L. Gleser, A. Nusser, and A. J. Benson, Mon. Not. R. Astron. Soc. **391**, 383 (2008).

[25] G. Paciga, T.-C. Chang, Y. Gupta, R. Nityanada, J. Odegova, U.-L. Pen, J. B. Peterson, J. Roy, and K. Sigurdson, Mon. Not. R. Astron. Soc. **413**, 1174 (2011).

[26] A. Liu and M. Tegmark, Mon. Not. R. Astron. Soc. **419**, 3491 (2012).

[27] 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. Lett. **763**, L20 (2013).

[28] G. Harker, S. Zaroubi, G. Bernardi, M. A. Brentjens, A. G. de Bruyn, B. Ciardi, V. Jelić, L. V. E. Koopmans, P. Labropoulos, G. Mellema, A. Offringa, V. N. Pandey, J. Schaye, R. M. Thomas, and S. Yatawatta, Mon. Not. R. Astron. Soc. **397**, 1138 (2009).

[29] E. Chapman, F. B. Abdalla, G. Harker, V. Jelić, P. Labropoulos, S. Zaroubi, M. A. Brentjens, A. G. de Bruyn, and L. V. E. Koopmans, Mon. Not. R. Astron. Soc. **423**, 2518 (2012).

[30] E. Chapman, F. B. Abdalla, J. Bobin, J.-L. Starck, G. Harker, V. Jelić, P. Labropoulos, S. Zaroubi, M. A. Brentjens, A. G. de Bruyn, and L. V. E. Koopmans, Mon. Not. R. Astron. Soc. **429**, 165 (2013).

[31] N. Petrovic and S. P. Oh, Mon. Not. R. Astron. Soc. **413**, 2103 (2011).

[32] A. R. Parsons, J. C. Pober, J. E. Aguirre, C. L. Carilli, D. C. Jacobs, and D. F. Moore, Astrophys. J. **756**, 165 (2012).

[33] J. Cho, A. Lazarian, and P. T. Timbie, Astrophys. J. **749**, 164 (2012).

[34] J. R. Shaw, K. Sigurdson, U.-L. Pen, A. Stebbins, and M. Sitwell, Astrophys. J. **781**, 57 (2014).

[35] J. R. Shaw, K. Sigurdson, M. Sitwell, A. Stebbins, and U.-L. Pen, arXiv:1401.2095.

[36] A. Liu and M. Tegmark, Phys. Rev. D **83**, 103006 (2011).

[37] J. S. Dillon, A. Liu, and M. Tegmark, Phys. Rev. D **87**, 043005 (2013).

[38] S. P. Oh and K. J. Mack, Mon. Not. R. Astron. Soc. **346**, 871 (2003).

[39] A. Datta, J. D. Bowman, and C. L. Carilli, Astrophys. J. **724**, 526 (2010).

[40] H. Vedantham, N. Udaya Shankar, and R. Subrahmanyan, Astrophys. J. **745**, 176 (2012).

[41] M. F. Morales, B. Hazelton, I. Sullivan, and A. Beardsley, Astrophys. J. **752**, 137 (2012).

[42] C. M. Trott, R. B. Wayth, and S. J. Tingay, Astrophys. J. **757**, 101 (2012).

[43] N. Thyagarajan *et al.*, Astrophys. J. **776**, 6 (2013).

[44] B. J. Hazelton, M. F. Morales, and I. S. Sullivan, Astrophys. J. **770**, 156 (2013).

[45] J. C. Pober, A. R. Parsons, J. E. Aguirre, Z. Ali, R. F. Bradley, C. L. Carilli, D. DeBoer, M. Dexter, N. E. Gugliucci, D. C. Jacobs, P. J. Klima, D. MacMahon, J. Manley, D. F. Moore, I. I. Stefan, and W. P. Walbrugh, Astrophys. J. Lett. **768**, L36 (2013).

[46] A. Liu, A. R. Parsons, and C. M. Trott, following paper, Phys. Rev. D **90**, 023019 (2014).

[47] A. Parsons, J. Pober, M. McQuinn, D. Jacobs, and J. Aguirre, Astrophys. J. **753**, 81 (2012).

[48] A. P. Beardsley *et al.*, Mon. Not. R. Astron. Soc. **429**, L5 (2013).

[49] M. Tegmark, Phys. Rev. D **55**, 5895 (1997).

[50] J. R. Bond, A. H. Jaffe, and L. Knox, Phys. Rev. D **57**, 2117 (1998).

[51] M. Tegmark, A. J. S. Hamilton, M. A. Strauss, M. S. Vogeley, and A. S. Szalay, Astrophys. J. **499**, 555 (1998).

[52] M. F. Morales and J. Hewitt, Astrophys. J. **615**, 7 (2004).

[53] G. Bernardi *et al.*, Astrophys. J. **771**, 105 (2013).

[54] M. Tegmark, Astrophys. J. Lett. **480**, L87 (1997).

[55] M. F. Morales and M. Matejek, Mon. Not. R. Astron. Soc. 400, 1814 (2009).

[56] R. Barkana, Mon. Not. R. Astron. Soc. 397, 1454 (2009).

[57] A. de Oliveira-Costa, M. Tegmark, B. M. Gaensler, J. Jonas, T. L. Landecker, and P. Reich, Mon. Not. R. Astron. Soc. 388, 247 (2008).

[58] M. G. Santos, A. Cooray, and L. Knox, Astrophys. J. 625, 575 (2005).

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

[60] G. Bernardi, M. McQuinn, and L. J. Greenhill, arXiv:1404.0887.