Draft version February 15, 2016
Preprint typeset using LaTeX style emulateapj v. 12/16/11

OPTIMIZED BEAM SCULPTING WITH GENERALIZED FRINGE-RATE FILTERS

Aaron R. Parsons[1,2], Adrian Liu[1,3], Zaki S. Ali[1], Carina Cheng[1]

Draft version February 15, 2016

ABSTRACT

We generalize the technique of fringe-rate filtering, whereby visibilities measured by a radio interferometer are re-weighted according to their temporal variation. As the Earth rotates, radio sources traverse through an interferometer's fringe pattern at rates that depend on their position on the sky. Capitalizing on this geometric interpretation of fringe rates, we employ time-domain convolution kernels to enact fringe-rate filters that sculpt the effective primary beam of antennas in an interferometer. As we show, beam sculpting through fringe-rate filtering can be used to optimize measurements for a variety of applications, including mapmaking, minimizing polarization leakage, suppressing instrumental systematics, and enhancing the sensitivity of power-spectrum measurements. We show that fringe-rate filtering arises naturally in minimum variance treatments of many of these problems, enabling optimal visibility-based approaches to analyses of interferometric data that avoid systematics potentially introduced by traditional approaches such as imaging. Our techniques have recently been demonstrated in Ali et al. (2015), where new upper limits were placed on the 21 cm power spectrum from reionization, showcasing the ability of fringe-rate filtering to successfully boost sensitivity and reduce the impact of systematics in deep observations.

1. INTRODUCTION

In recent years, low-frequency radio interferometers have undergone dramatic changes in design. These transformations have been driven by new science applications such as 21 cm cosmology, where one uses the highly redshifted emission from the 21 cm hyperfine transition of neutral hydrogen to map our early Universe. Observers in 21 cm cosmology seek to measure small fluctuations (both spatially and spectrally) in a dim, diffuse background that is obscured by bright foreground emission orders of magnitude brighter in brightness temperature. This stands in contrast to many traditional observations in radio astronomy, which more usually target bright, compact objects in front of a dim background, often over a small selection of frequencies. These differences have led to the design, construction, and usage of new interferometers that only have moderate angular resolution, but are comprised of a large number of receiving elements with wide fields of view operating over a wide bandwidth. Examples of new interferometers that broadly fit some or all of this description include the Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER; Parsons et al. 2010), the Murchison Widefield Array (MWA; Tingay et al. 2013; Bowman et al. 2013), the LOw Frequency ARray (LOFAR; van Haarlem et al. 2013), the Canadian Hydrogen Intensity Mapping Experiment (CHIME; Bandura et al. 2014), the MIT Epoch of Reionization experiment (MITEoR; Zheng et al. 2014), the Large Aperture Experiment to Detect the Dark Ages (LEDA; Greenhill & Bernardi 2012), and the Hydrogen Epoch of Reionization Array (HERA; Pober et al. 2014). Further deviating from conventional array designs, the PAPER, MITEoR, CHIME, HERA projects have also maximized sensitivity by choosing to place their antenna elements in regular, redundant grids (Parsons et al. 2012a).

With new interferometer designs, it is natural to expect new approaches to data analysis. In this paper, we critically examine methods for time integration. Integrating in time is a crucial step for the high-sensitivity applications of modern low-frequency radio astronomy. Consider the measurement of the high-redshift 21 cm power spectrum as an example application. At the relevant redshifts ($z \sim 6$ to 20), theoretical models suggest that this cosmological signal will be faint — on the order of 1 mK in brightness temperature. The noise power spectrum on such measurements reaches comparable magnitudes only after long integration ($\gtrsim 1000$ hrs) on instruments optimized for such a measurement (Harker et al. 2010; Parsons et al. 2012a; Beardsley et al. 2013; Pober et al. 2014), and even then, often only for the largest spatial modes (depending on the instrument). Long time-integrations are therefore crucial not only for generating the requisite sensitivity for a detection of the cosmological signal, but also to allow faint systematics to be detected and excised from the data.

In this paper, we extend ideas introduced in Parsons & Backer (2009) (as well as similar ideas in Roshi & Perley 2003, Offringa et al. 2012 and Shaw et al. 2013a) to optimize the process of combining time-ordered data. The key realization is that fringe-rate—the Fourier dual to time—is a more natural space to enact time-averaging. Traditional time-averaging (say, a running box-car average) is equivalent to multiplying by a sinc filter in the fringe-rate domain. Generalizing this process, the convolution theorem ensures that time integration can be achieved by weighting the data in the fringe-rate domain. The fringe-rate domain provides a natural basis for time-averaging interferometric data because astronomical sources are locked to the celestial sphere, and therefore appear at predictable fringe-rates in the data. In particular, for a given interferometric baseline, there exists a maximum allowable fringe-rate, beyond which there is only instrumental noise. Fringe-rate filtering al-

[1] Astronomy Dept., U. California, Berkeley, CA
[2] Radio Astronomy Lab., U. California, Berkeley, CA
[3] Berkeley Center for Cosmological Physics, Berkeley, CA

lows the clean elimination of such noise-like modes.

We place a particular emphasis on the geometric interpretation of fringe-rate filtering, where weightings in the fringe-rate domain result in changes to an interferometer's spatial response, effectively allowing different portions of the sky to be selected by carefully chosen fringe-rate filters. These filters can be optimized for a number of different applications, including the measurement of cosmological power spectra, the reduction of polarization leakage, and the downweighting of contaminating sources far from the central regions of the sky that one is attempting to observe. Importantly, these filters can be implemented on a per-baseline basis, providing a different view of systematics in the data, which are often easier to identify when described baseline-by-baseline, instead of being mixed together in an image-domain map. However, we will also show that optimally weighted mapmaking can also be more conveniently conceptualized in a mathematical framework operating in the fringe-rate basis.

The rest of the paper is organized as follows. In Section 2 we provide a general overview of fringe-rate filtering, establishing an essential geometric intuition for the process. The specific implementation that we use for the simulations in this paper are described in Section 3. Section 4 describes how fringe-rate filtering can be optimized for various applications. We pay specific attention to the problem of mapmaking in Section 5, and show that fringe-rate filtering arises naturally in that context as well. We summarize our conclusions in Section 6.

2. OVERVIEW OF PRINCIPLE OF FRINGE-RATE FILTERING

Generally, the interferometric response V at frequency ν for two antennas in a radio interferometer is described by the visibility function[4]

$$V_{b\nu}(t) = \int d\Omega\, I_\nu(\hat{\mathbf{r}}) A_\nu(\hat{\mathbf{r}}, t) \exp\left[-i2\pi \frac{\nu}{c}\mathbf{b}(t)\cdot\hat{\mathbf{r}}\right], \quad (1)$$

where I_ν is the specific intensity of the sky in the direction $\hat{\mathbf{r}}$, A_ν is the geometric mean of the primary beam power patterns of the constituent antennas (henceforth known as "the primary beam"), $\mathbf{b}(t)$ is the baseline vector separating the two antennas in question (which is time-dependent since the baselines rotate with the Earth), and ν is the spectral frequency. We adopt the convention that our coordinate system is fixed to the celestial sphere, because it will be convenient for our algebraic manipulations later. However, it is equally valid to understand the time-variation of the visibilities as arising from the movement of astronomical sources through the primary beam and the fringes arising from a baseline, which are fixed to a topocentric coordinate system. For drift-scan telescopes like PAPER, CHIME, or HERA, this view is particularly powerful because then the primary beam and the fringe pattern are locked to one another, and may together be considered an enveloped fringe pattern that gives rise to time-variation in $V_{b\nu}(t)$ as the Earth rotates.

The rate at which angular structure on the sky moves relative to the fringe pattern—the *fringe rate*—depends on the declination and hour angle. As an example, Figure 1 illustrates the real component of the phase variation in the fringe pattern of a 30-m east-west baseline deployed at -30° latitude. Although fringes are evenly spaced in $l \equiv \sin\theta_x$, the distance a source that is locked to the celestial sphere travels through the fringe pattern depends on its position on the sphere. This is illustrated in Figure 1 by arrows that indicate the motion of sources at differing declinations over the course of two hours near transit. As the source and the fringe pattern move relative to one another, $V_\nu(t)$ oscillates with an amplitude that is determined by the strength of the source and the amplitude of the beam response, and a frequency that corresponds to the number of fringe periods traversed in a given time interval. Hence, the frequency or *fringe-rate* of oscillations in $V_\nu(t)$ ranges from a maximum at $\delta = 0^\circ$ to zero at $\delta = -90^\circ$, and can even become negative for emission from the far side of the celestial pole.

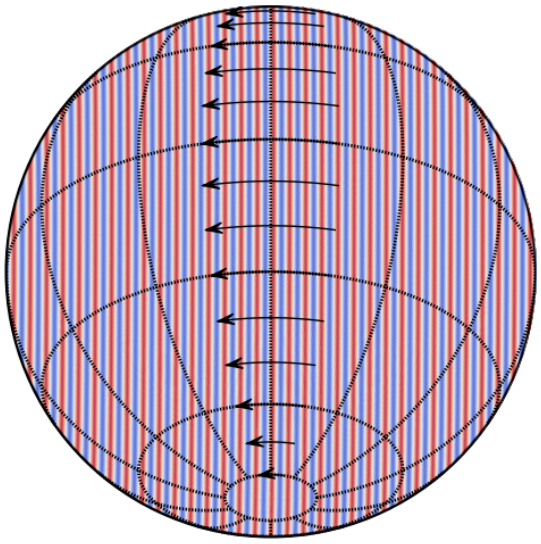

FIG. 1.— The fringe pattern at 150 MHz of a fiducial 30-m east-west baseline, overlaid with arrows indicating the distance traversed by sources at various declinations over a two-hour time span centered at transit. In a fixed time interval, sources near declination $\delta = 0^\circ$ traverse more fringe periods than sources nearer to the celestial poles. This gives rise to different fringe rates that can be used to distinguish sources in a time-series measured with a single baseline.

Let us now derive this intuition mathematically, assuming a drift-scan telescope. To sort our time-variable visibilities into different fringe-rates f, we take the Fourier transform of our visibility over a short interval of time centered at time t to get

$$\overline{V}_{b\nu}(f,t) = \int d\Omega I_\nu(\hat{\mathbf{r}}) \int dt' \gamma(t'-t) A_\nu(\hat{\mathbf{r}}, t') e^{-i2\pi\left[(t'-t)f + \frac{\nu}{c}\mathbf{b}_{t'}\cdot\hat{\mathbf{r}}\right]}, \quad (2)$$

where we have introduced the notation $\mathbf{b}_t \equiv \mathbf{b}(t)$, and γ is a tapering function for the Fourier transform in time, which we assume peaks when its argument is zero, in essence shifting the origin of our transform to time t. If the characteristic width of γ is relatively short, one is effectively examining the visibility over short timescales, during which its time-dependence will likely be dominated by features on the sky moving relative to fringes,

[4] In this section, we omit the instrumental noise contribution to the measured visibilities in order to avoid notational clutter.

and not the movement of the primary beam through the celestial sphere. We may therefore say that for short periods of time, $A_\nu(\hat{\mathbf{r}}, t') \approx A_\nu(\hat{\mathbf{r}}, t)$. Additionally, we may take the time-dependence of the baselines to leading order, with

$$\begin{aligned}\mathbf{b}_{t'} &\approx \mathbf{b}_t + \left.\frac{d\mathbf{b}}{dt}\right|_{t'=t}(t'-t) + \dots \\ &= \mathbf{b}_t - (\mathbf{b}_t \times \boldsymbol{\omega}_\oplus)(t'-t) + \dots \end{aligned} \quad (3)$$

where $\boldsymbol{\omega}_\oplus$ is the angular velocity vector of the Earth's rotation. In the last equality, we used the fact that the time-dependence of the baselines are not arbitrary, but instead are tied to the Earth's rotation, transforming the time derivative into a cross-product with $\boldsymbol{\omega}_\oplus$, as one does in the analysis of solid rotating bodies. Inserting these approximations into Equation (2) yields

$$\begin{aligned}\overline{V}_{b\nu}(f,t) &= \int d\Omega I_\nu(\hat{\mathbf{r}}) A_\nu(\hat{\mathbf{r}},t) e^{-i2\pi\frac{\nu}{c}\mathbf{b}_t\cdot\hat{\mathbf{r}}} \\ &\quad \cdot \int dt' \gamma(t'-t) e^{i2\pi[\frac{\nu}{c}(\mathbf{b}_t\times\boldsymbol{\omega}_\oplus)\cdot\hat{\mathbf{r}} - f](t'-t)} \\ &= \int d\Omega I_\nu(\hat{\mathbf{r}}) A_\nu(\hat{\mathbf{r}},t) e^{-i2\pi\frac{\nu}{c}\mathbf{b}_t\cdot\hat{\mathbf{r}}} \tilde{\gamma}\left[\frac{\nu}{c}(\mathbf{b}_t\times\boldsymbol{\omega}_\oplus)\cdot\hat{\mathbf{r}} - f\right], \end{aligned} \quad (4)$$

where $\tilde{\gamma}$ is the inverse Fourier transform of γ. To the extent that $\gamma(t)$ can be chosen to be relatively broad without violating our approximations, $\tilde{\gamma}$ will be peaked around the point where its argument is zero. Its presence in Equation (4) therefore acts approximately like a delta function, selecting portions of the sky that have $\hat{\mathbf{r}}$ satisfying the condition $f \approx \hat{\mathbf{r}}\cdot\mathbf{b}\times\boldsymbol{\omega}_\oplus\nu/c$.

In words, what the above derivation shows is that, as claimed, different fringe-rates correspond to different parts of the sky. This is illustrated in Figure 2, which shows shaded bands of constant fringe-rate for the same baseline as the one simulated in Figure 1. In general, contours of constant fringe-rate correspond to locations on the sky $\hat{\mathbf{r}}$ that have the same degenerate combination of $\hat{\mathbf{r}}\cdot\boldsymbol{\omega}_\oplus\times\mathbf{b}$. Note that this combination can also be rewritten as $\boldsymbol{\omega}_\oplus\cdot\mathbf{b}\times\hat{\mathbf{r}}$ or $\mathbf{b}\cdot\hat{\mathbf{r}}\times\boldsymbol{\omega}_\oplus$ by cyclic permutation. Thus, if any two of $\mathbf{b}$, $\omega_\oplus$, and $\hat{\mathbf{r}}$ are parallel, their cross product—and hence the fringe-rate—will be zero. For example, the fringe-rate for astronomical sources located at either celestial pole will always be zero, since $\hat{\mathbf{r}}$ would then be parallel to $\boldsymbol{\omega}_\oplus$. Similarly, a north-south only baseline located at the equator would have $\mathbf{b}$ parallel to $\boldsymbol{\omega}_\oplus$, resulting in $f = 0$ because in such a scenario the fringes would have no azimuthal dependence, and thus there would be no fringe-crossings as the Earth rotates relative to the sky.

Because different fringe-rates correspond to different parts of the sky, we may effectively select different portions of the sky by picking different linear combinations of fringe-rates. To see this, imagine decomposing our data into fringe-rates, and then applying a weighting function $w(f)$ before Fourier transforming back to the time domain. The result is

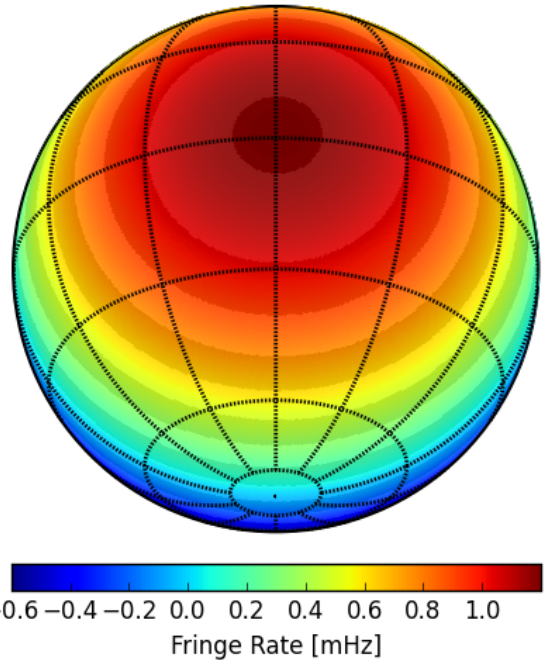

FIG. 2.— Fringe rate as a function of sky position, corresponding to the fringe pattern illustrated in Figure 1. Fringe rates peak at 1.09 mHz at $\delta = 0°$, hit zero at the south celestial pole, and become negative on the far side of the pole. Grey shading indicates the approximate angular regions that correspond to alternating fringe-rate bins, assuming a fringe-rate transform taken over a two-hour time series.

$$\begin{aligned}V^{\rm filt}_{b\nu}(t',t) &= \int df w(f) \overline{V}_{b\nu}(f,t) e^{i2\pi(t'-t)f} \\ &= \int d\Omega I_\nu(\hat{\mathbf{r}}) A_\nu(\hat{\mathbf{r}},t) e^{-i2\pi\frac{\nu}{c}\mathbf{b}_t\cdot\hat{\mathbf{r}}} \\ &\quad \times \int df e^{i2\pi f(t'-t)} w(f)\tilde{\gamma}\left[\frac{\nu}{c}(\mathbf{b}_t\times\boldsymbol{\omega}_\oplus)\cdot\hat{\mathbf{r}} - f\right]. \end{aligned} \quad (5)$$

Now, suppose we implement this filter in a sliding manner (a convolution) in time. That is, we repeat this process with the fringe-rate transform centered on each instant in time. With this, we become interested in only $t' = t$, so the final set of filtered visibilities takes the form

$$V^{\rm filt}_{b\nu}(t) = \int d\Omega\, I_\nu(\hat{\mathbf{r}}) A^{\rm eff}_\nu(\hat{\mathbf{r}},t) \exp\left[-i2\pi\frac{\nu}{c}\mathbf{b}(t)\cdot\hat{\mathbf{r}}\right], \quad (6)$$

which is precisely the same as our original measurement equation, except the primary beam has been replaced by an *effective primary beam*, defined as

$$A^{\rm eff}_\nu(\hat{\mathbf{r}},t) \equiv A_\nu(\hat{\mathbf{r}},t)(w * \tilde{\gamma})\left[\frac{\nu}{c}(\mathbf{b}_t\times\boldsymbol{\omega}_\oplus)\cdot\hat{\mathbf{r}}\right], \quad (7)$$

with * signifying a convolution. We thus see that by judiciously selecting fringe-rate weights, one can effectively reshape one's beam. In general, however, we cannot do so with perfect flexibility. This can be seen by once again examining the combination $\hat{\mathbf{r}}\cdot\mathbf{b}\times\boldsymbol{\omega}_\oplus$. For any given instant, $\mathbf{b}\times\boldsymbol{\omega}_\oplus$ picks out a particular direction on the celestial sphere. A ring of locations $\hat{\mathbf{r}}$ on the sky at a constant angle with respect to this direction will have

the same value of $\hat{\mathbf{r}} \cdot \mathbf{b} \times \boldsymbol{\omega}_\oplus$, and therefore the same fringe-rate. As a result, contours of constant fringe-rate always form rings on the sky, as illustrated in Figure 2. By weighting different fringe-rates, one can effectively "turn off" (or less harshly, to simply downweight) whole contours, but never portions of a contour.

Aside from modifying the shape of one's beam, fringe-rate filtering can also be used to integrate visibilities in time. For example, if $w(f)$ is chosen in a way that suppresses high fringe rates, the effect in the time domain will be a low-pass filter that (among other features) has the rough effect of averaging together data. Enacting the time-averaging in the fringe-rate domain is particularly helpful for differentiating between noise- and signal-like modes in the time-series data. To see this, recall that the relative compactness of the $\tilde{\gamma}$ term in Equation (4) implies that an astronomical source located at $\hat{\mathbf{r}}$ will preferentially appear at a fringe rate of $f \approx \hat{\mathbf{r}} \cdot \mathbf{b} \times \boldsymbol{\omega}_\oplus \nu / c$ in the data. Since $\hat{\mathbf{r}} \cdot \mathbf{b} \times \boldsymbol{\omega}_\oplus$ can never exceed $b\omega_\oplus$, the maximal fringe-rate that can be achieved by a source locked to the celestial sphere is $f_{\rm max} = b\omega_\oplus \nu / c$, where $\omega_\oplus \equiv |\boldsymbol{\omega}_\oplus|$ and $b \equiv |\mathbf{b}|$. Data at even higher fringe rates will likely be noise- rather than signal-dominated and may be filtered out safely with no loss of signal. This is a more tailored approach to reducing time-ordered data than simply averaging visibilities together in time. The latter can be viewed as a boxcar convolution in the time domain, which corresponds to applying a sinc filter in fringe-rate space. With wings that only decay as $1/f$, a sinc filter tends to incorporate data from the noise-dominated high fringe rate modes. A fringe-rate filter, in contrast, can be more carefully tailored to enhance modes that are sourced by actual emission from the celestial sphere.

In this section, we have provided some basic intuition for fringe-rate filtering, and have highlighted how it can be used for reshaping one's primary beam as well as to combine time-ordered data. In fact, these two applications are often intimately linked, since optimal prescriptions for combining time-ordered data ("mapmaking") involve re-weighting data by the primary beam (Tegmark 1997; Morales & Matejek 2009a; Dillon et al. 2015). We will return to this in Section 5, where we will see that the fringe-rate filtering is a natural way to approach mapmaking in interferometric observations.

3. IMPLEMENTATION

In this section, we discuss a practical implementation of the aforementioned ideas. We will use this implementation in simulations later in the paper to illustrate various applications of fringe-rate filtering. In keeping with its origins as a tool for analyzing data from the PAPER array, we simulate a model array based on PAPER, deployed at a latitude of $-30°$ and featuring the beam response pattern characteristic of PAPER dipole elements (Parsons et al. 2008; Pober et al. 2012). The PAPER beam response pattern is illustrated in the left-most panel of Figure 7. For these simulations, we also choose a specific baseline to examine: a pair of antennas separated by 30 m in the east-west direction, deployed at a latitude of $-30°$, and observing at 150 MHz. This baseline, hereafter referred to as our *fiducial baseline*, corresponds to the most repeated (and hence, most sensitive) baseline length measured by the PAPER array in the maximum-redundancy array configuration it uses for power spectral measurements (Parsons et al. 2012a, 2014; Ali et al. 2015). As such, our simulations demonstrate the performance of fringe-rate filtering in the context of the specific instrument configuration that has recently been used to place the current best upper limits on 21 cm emission from cosmic reionization (Parsons et al. 2014; Jacobs et al. 2014; Ali et al. 2015).

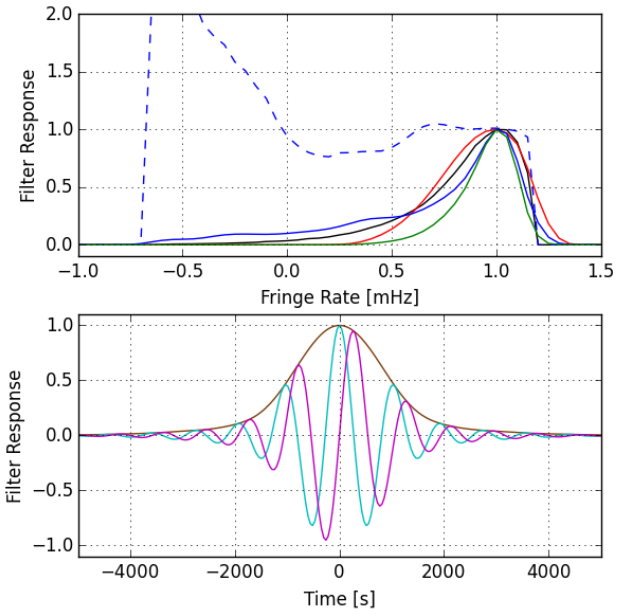

FIG. 3.— Top: the optimal power-spectrum sensitivity weighting in fringe-rate space for the XX polarization beam of our fiducial baseline (black) is overlaid with the simple parametrization for optimal weighting (red) used in Ali et al. (2015), which excises fringe rates at risk for contamination by crosstalk, and applies a Blackman-Harris window in the time domain to generate a compact time-domain convolution kernel with regions weighted below 5% of the peak response suppressed. Also illustrated are the weightings for matching the YY to the XX polarization beam (blue dashed) and for reducing polarization leakage (blue solid) described in Section 4.2, as well as a weighting that balances the needs of sensitivity, polarization leakage, crosstalk removal, and off-axis foregrounds (green). Bottom: the time-domain convolution kernel corresponding to the red curve in the top panel. Real and imaginary components are illustrated in cyan and magenta, respectively, with the absolute amplitude illustrated in brown.

In the following sections, we will derive a number of different fringe-rate weights, each optimized for a different application. Often, these optimized weights depend on the detailed properties of one's instrument, and can therefore only be computed numerically, not analytically. For example, we will find in Section 4.1 that the optimal fringe-rate weights for power spectrum estimation involve computing the root-mean-square (RMS) primary beam profile over contours of constant fringe rate on the sky (such as those shown in Figure 2). An example of the binning of the RMS beam response in fringe rate is given by the black curve in the top panel of Figure 3. A realistic primary beam will frequently require empirical modeling beyond analytic forms, making it generally difficult to derive a completely analytic expression for an optimized fringe-rate filter profile. However, in the interest of being able to rapidly generate filters as a function of varying baseline lengths and observing frequencies, we

frequently fit analytic forms (such as truncated Gaussians) to the numerical profiles, as shown by the red curve in the top panel of Figure 3. As long as the numerical profiles take the optimized forms that we will derive in Section 4, small deviations arising from an imperfect analytic fit are unlikely to significantly shift the final error properties of one's measurements. With the discussion of power spectrum measurements in Section 4.1, for example, we minimize the noise variance by varying the fringe-rate weights. Because our analytic fits to these weights start from a local minimum in noise variance, any deviations in the weighting profile will only induce small second-order increases in the final error bars.

The next step in implementing the fringe-rate filter is translating the analytic filter profile in fringe-rate space into a time-domain kernel that can be used to convolve the simulated time series of visibilities. In effect, we implement the fringe-rate filter as a finite impulse response (FIR) filter. The convolution kernel corresponding to this FIR filter is shown in the lower panel of Figure 3. Applying the fringe-rate filter as an FIR filter in the time domain, as opposed to directly multiplying the desired filter to Fourier-transformed visibilities, has the advantage that flagged or missing data can be naturally excluded from the filter by neglecting FIR taps (coefficient multiplies) that target the missing data. The summed output of the FIR filter are then renormalized to account for the missing samples. Another advantage of the FIR implementation of the fringe-rate filter is the potential for windowing the time-domain filter profile. While time-domain windowing causes further deviations from the ideal fringe-rate filter profile, it can be used to produce a more compact time-domain kernel. Reducing the number of time-domain samples used in the FIR filter improves the computational efficiency of the filter, helps limit the number of samples potentially corrupted by spurious systematics such as radio frequency interference, and reduces boundary effects. In Ali et al. (2015), a 12-hour data set was filtered in this manner, with boundary effects limited to an hour on each end. The final analysis proceeded on 8.3 hours selected from within the region not affected by boundary effects.

4. APPLICATIONS

In Section 3, we discussed how a fringe-rate filter can be implemented in practice once a particular form for the filter is selected. In this section, we optimize the selection of filters (or equivalently, of fringe-rate weights) for various applications in low-frequency radio interferometry. The key to this optimization will be the insight from Section 2, namely that the effect of fringe-rate filtering can be regarded as both a time integration and a modification of the spatial response of the primary beam on a per-baseline basis. Turning this around, one can identify the optimal primary beam needed for one's observations, and then reverse engineer the set of fringe-rates needed to achieve this beam in what is essentially a "beam sculpting" operation. For concreteness, we will focus here on 21 cm cosmology, but many of the ideas presented here are easily translatable to other applications of radio interferometry.

4.1. *Minimizing thermal noise errors in power spectrum measurements*

In Parsons et al. (2012a) and Parsons et al. (2014), it was shown that estimates of the three-dimensional power spectrum of 21 cm brightness temperature fluctuations could be obtained from a single baseline by Fourier transforming visibility data along the frequency axis (forming a "delay spectrum"), and then taking the absolute square of the results. Here, we will show how fringe-rate weights can be chosen to maximize the sensitivity of a single-baseline-derived power spectrum.

We begin by considering a generalization of the derivation in Parsons et al. (2014), where it was assumed that the primary beams of all elements in the interferometer are identical. We now consider the possibility of probing the power spectrum via a cross-correlation of two baselines with different primary beams. To be clear, our eventual discussion will be based on the analysis of fringe-rate filtered visibilities from a *single* baseline. However, from Section 2, we saw that to a good approximation, selecting different fringe-rates is equivalent to observing the sky with different effective beams. Thus, the cross-correlation of visibilities from two different fringe-rate bins is mathematically identical to cross-correlating two baselines with different beams. To begin, suppose that the ith baseline consists of antenna elements with primary beam $A_i(\hat{\mathbf{r}},\nu)$. The delay-transformed visibility takes the form

$$\widetilde{V}_i(\mathbf{u},\eta) = \frac{2k_B}{\lambda^2}\int d^2\mathbf{u}'\, d\eta'\, \widetilde{A}_i(\mathbf{u}-\mathbf{u}',\eta-\eta')\widetilde{T}(\mathbf{u}',\eta'), \quad (8)$$

where η is the Fourier dual to frequency[5], $\widetilde{A}_i$ is the Fourier transform of $A_i(\hat{\mathbf{r}},\nu)$ in both the angular and spectral directions, $\widetilde{T}$ is the brightness temperature field in Fourier space, k_B is Boltzmann's constant, λ is the central observation frequency, and $\mathbf{b} = \mathbf{u}\lambda$, with $\mathbf{u} \equiv (u,v)$. From this, we can see that two baselines with different primary beams, but located at the same location on the uv plane have a delay-spectrum cross-correlation given by

$$\begin{aligned}\langle \widetilde{V}_i(\mathbf{u},\eta)\widetilde{V}_j(\mathbf{u},\eta)^*\rangle &= \left(\frac{2k_B}{\lambda^2}\right)^2 \int d^2\mathbf{u}'\, d\eta' P(\mathbf{u}',\eta') \\ &\quad \times \widetilde{A}_i(\mathbf{u}-\mathbf{u}',\eta-\eta')\widetilde{A}_j^*(\mathbf{u}-\mathbf{u}',\eta-\eta') \\ &\approx P(\mathbf{u},\eta)\left(\frac{2k_B}{\lambda^2}\right)^2 \int d\Omega d\nu A_i(\hat{\mathbf{r}},\nu)A_j(\hat{\mathbf{r}},\nu), \quad (9)\end{aligned}$$

where angular brackets $\langle \dots \rangle$ denote an ensemble average over possible realizations of a random temperature field. In the first equality, we assumed that this field is a translation-invariant Gaussian random field specified by a power spectrum $P(\mathbf{u},\eta)$, so that

$$\langle \widetilde{T}(\mathbf{u},\eta)\widetilde{T}^*(\mathbf{u}',\eta')\rangle = \delta^D(\mathbf{u}-\mathbf{u}')\delta^D(\eta-\eta')P(\mathbf{u},\eta). \quad (10)$$

In the second equality, we made the approximation that

[5] This equation can be derived by Fourier transforming Equation (1) along the frequency axis and re-expressing the angular integral in uv coordinates assuming the flat-sky approximation. However, it also makes the crucial assumption that one can neglect the frequency-dependent nature of the mapping of baseline $\mathbf{b}$ to $\mathbf{u}$ coordinate. In practice, this is only a reasonable approximation for short baselines (Parsons et al. 2012b; Liu et al. 2014a) such as those used for power spectrum analyses in the PAPER experiment (Parsons et al. 2014; Jacobs et al. 2014; Ali et al. 2015).

for reasonably broad primary beams, $\widetilde{A}_i$ and $\widetilde{A}_j$ tend to be rather localized, which allows the comparatively broader $P(\mathbf{u},\eta)$ to be factored out of the integral.[6] Following this, we used Parseval's theorem to rewrite the integral over $(\mathbf{u},\eta)$ space as an integral over solid angle and frequency.

Rearranging Equation (9) gives an expression for the true power spectrum in terms of the cross-correlation function of two delay-space visibilities. With real data, however, one cannot perform the ensemble average on the left-hand side of Equation (9). Omitting this ensemble average, the copy of the power spectrum on the right-hand side becomes an *estimator* $\widehat{P}$ of the true power spectrum P. Introducing the definition

$$\Omega_{ij}^{pp} \equiv \frac{1}{B}\int d\Omega d\nu A_i(\hat{\mathbf{r}},\nu)A_j(\hat{\mathbf{r}},\nu), \tag{11}$$

where B is the bandwidth over which observations are made, our estimator takes the form

$$\widehat{P}(\mathbf{k}) = \left(\frac{\lambda^2}{2k_B}\right)^2 \frac{X^2Y}{\Omega_{ij}^{pp}B}\widetilde{V}_i(\mathbf{u},\eta)\widetilde{V}_j(\mathbf{u},\eta)^*, \tag{12}$$

where we have written the power spectrum in terms of cosmological Fourier coordinates $\mathbf{k}$, which are related to the interferometric Fourier coordinates by $(Xk_x, Xk_y, Yk_z) \equiv 2\pi(u,v,\eta)$, picking up an extra factor of X^2Y in the process,[7] with

$$X \equiv \frac{c}{H_0}\int_0^z \frac{dz'}{E(z')};\ E(z) \equiv \sqrt{\Omega_m(1+z)^3+\Omega_\Lambda}. \tag{13}$$

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, Ω_Λ is the normalized dark energy density, and

$$Y \equiv \frac{c(1+z)^2}{\nu_{21}H_0E(z)}, \tag{14}$$

where $\nu_{21} \approx 1420\,\mathrm{MHz}$ is the rest frequency of the 21 cm line. In the special case where there is just a single primary beam, we may set $i=j$ and drop the subscripts for brevity, and Equation (12) reduces to

$$\widehat{P}(\mathbf{k}) = \left(\frac{\lambda^2}{2k_B}\right)^2 \frac{X^2Y}{\Omega_{pp}B}|\widetilde{V}(\mathbf{u},\eta)|^2, \tag{15}$$

where

$$\Omega_{pp} \equiv \frac{1}{B}\int d\Omega d\nu |A(\hat{\mathbf{r}},\nu)|^2, \tag{16}$$

which is the relation found in Parsons et al. (2014).

Having established these results, let us re-interpret Equation (12) as an estimator for the power spectrum from the cross-multiplication of two different discretized fringe rate bins (as opposed to the cross-multiplication of baselines with different primary beams). We are free to re-interpret our estimator in this way because of the discussion in Section 2, where we showed that each visibility could be thought of as being comprised of different fringe rate contributions, each probing a different ring on the celestial sphere. Each fringe-rate therefore has its own effective primary beam, enabling our re-interpretation. That Equation (12) involves the cross-multiplication of visibilities after they have been delay-transformed over the frequency axis is not a problem, since the Fourier transforms required to enact the delay transform and the fringe-rate transform commute with one another.

Equation (12) allows a power spectrum to be estimated from the cross-multiplication of any pair of fringe-rate bins. To increase signal-to-noise on the measurement, however, one ought to form all possible cross-multiplied pairs, which can then combined into a single power spectrum estimate via a weighted average. Suppressing the arguments of $\widehat{P}$ and $\widetilde{V}$ for notational cleanliness, we can write

$$\widehat{P} = \sum_{ij} g_{ij}\widetilde{V}_i\widetilde{V}_j^*, \tag{17}$$

where g_{ij} is the weight assigned to the cross-multiplication of the ith and jth fringe-rate bins. Our goal is to select weights that minimize the error bars on the final power spectrum estimate.

For our optimization exercise, assume that errors are due to instrumental thermal noise only. If the ith fringe-rate bin has a noise contribution of n_i, the noise contribution to our estimator is

$$\widehat{P}_{\mathrm{noise}} = \sum_{ij} g_{ij}n_in_j^*. \tag{18}$$

The error bar corresponding to this noise contribution is given by the square root of its variance, which takes the form

$$\begin{aligned}\mathrm{Var}(\widehat{P}_{\mathrm{noise}}) &\equiv \langle\widehat{P}_{\mathrm{noise}}^2\rangle - \langle\widehat{P}_{\mathrm{noise}}\rangle^2 \\ &= \sum_{ijkm} g_{ij}g_{km}\left[\langle n_in_j^*n_kn_m^*\rangle - \langle n_in_j^*\rangle\langle n_kn_m^*\rangle\right] \\ &= \sum_{ij} g_{ij}g_{ji}\sigma^4,\end{aligned} \tag{19}$$

where in the last equality we assumed that the noise is Gaussian, enabling the fourth moment term to be written as a sum of second moment (variance) terms. We further assumed that the real and imaginary components of the noise are uncorrelated with each other and between different fringe-rate bins, so that if $n_i \equiv a_i + ib_i$, we have $\langle a_ia_j\rangle = \langle b_ib_j\rangle = \delta_{ij}\sigma^2/2$ and $\langle a_ib_j\rangle = 0$ for all i and j.

In minimizing the noise variance, care must be taken to ensure that there is no signal loss in the power spectrum estimation. To do so, we first note that taking the ensemble average of $\widehat{P}$ gives

$$\langle\widehat{P}\rangle = \sum_{ij} g_{ij}\langle\widetilde{V}_i\widetilde{V}_j^*\rangle = S\sum_{ij} g_{ij}\Omega_{ij}P, \tag{20}$$

where we used an ensemble-averaged version of Equation (12) to relate the true cross-correlation to the true power spectrum, and defined $S \equiv (B/X^2Y)(2k_B/\lambda^2)^2$. Ensuring that there is no signal loss is thus tantamount to requiring that $S\sum_{ij} g_{ij}\Omega_{ij} = 1$, so that $\langle\widehat{P}\rangle = P$. We may impose this constraint by introducing a Lagrange multiplier λ in our minimization of the noise variance,

[6] Although see Section IV A of Liu et al. (2014b) for some limitations of this approximation.

[7] See, e.g., Liu et al. (2014a) for a detailed derivation.

minimizing

$$\mathcal{L} = \sum_{ij} g_{ij} g_{ji} - \lambda \sum_{ij} w_{ij} \Omega_{ij}, \quad (21)$$

where both σ^4 and S have been absorbed into our definition of λ. Differentiating with respect to each element and setting the result to zero gives an optimized weight given by $g_{km} \propto \Omega_{km}$, and normalizing according to our constraint yields

$$g_{km} = \frac{\Omega_{km}}{S \sum_{ij} \Omega_{ij}^2}. \quad (22)$$

To make intuitive sense of this, let us make a few more approximations. The key quantity here is Ω_{ij}, which we can see from Equation (11) is the overlap integral between the effective primary beams of the ith and jth fringe-rate bins. In Section 2, we saw that if one takes the fringe-rate Fourier transform over a wide enough window in time, different fringe-rates map to different portions of the sky with relatively little overlap. If this is indeed the case, Ω_{ij}^{pp} vanishes unless $i = j$. Defining

$$\mu_i \equiv \frac{1}{B} \int d\Omega d\nu A_i(\hat{\mathbf{r}}, \nu)^2, \quad (23)$$

we have $\Omega_{ij}^{pp} \equiv \delta_{ij} \mu_i$, so our optimal estimator for the power spectrum (combining Equations 17, 22, and 23) reduces to

$$\widehat{P} = \frac{1}{S \sum_j \mu_j^2} \sum_i \mu_i |\widetilde{V}_i|^2. \quad (24)$$

Suppose we now define $\mu_i^{1/2} \widetilde{V}_i$ to be an optimally weighted visibility in fringe-rate space. Transforming back to the time domain using Parseval's theorem, one obtains

$$\widehat{P}(\mathbf{u}, \eta) = \left(\frac{\lambda^2}{2k_B} \right)^2 \frac{X^2 Y}{B \sum_j \mu_j^2} \sum_i |\widetilde{V}^{\rm opt}(\mathbf{u}, \eta; t_i)|^2, \quad (25)$$

where the optimally filtered visibility in the time domain $\widetilde{V}^{\rm opt}(\mathbf{u}, \eta; t_i)$ is given by

$$\begin{aligned} \widetilde{V}^{\rm opt}(\mathbf{u}, \eta; t_i) &\equiv \sum_i e^{i2\pi f t} \mu_i^{1/2} \widetilde{V}_i \\ &= \sum_i e^{i2\pi f t} \left[\frac{1}{B} \int d\Omega d\nu A_i(\hat{\mathbf{r}}, \nu)^2 \right]^{\frac{1}{2}} \widetilde{V}_i. \end{aligned} \quad (26)$$

This is a rather interesting result, in that the optimal power spectrum estimator for a single baseline interferometer consists of a squared statistic (i.e., one with no phase information) integrated in time. This may seem counterintuitive, particularly if one is accustomed to more conventional techniques where images are formed from the visibilities and averaged down before any squaring steps. There, it is crucial to average in time *before* squaring, because data from different time steps can be sourced by the same Fourier modes on the celestial sphere. Integrating before squaring allows information from these modes to be coherently averaged together (since phase information has yet to be discarded), resulting in instrumental noise that integrates down as $1/\sqrt{t}$. This then becomes a $1/t$ dependence for the error bars on the final (squared) power spectrum results, and is a much quicker reduction of instrumental noise than if the data had been squared first, which would have resulted in a $1/\sqrt{t}$ dependence on the power spectrum errors.

In our derivation, we showed that the optimal power spectrum estimator can in fact be obtained by squaring before integrating, *provided* the power spectra formed at each time instant are first fringe-rate filtered with weights $\mu_i^{1/2}$, i.e., where each fringe-rate is weighted by the RMS primary beam within the corresponding constant-fringe-rate contour on the sky. Essentially, the pre-processing step of fringe-rate filtering (with these specific weights) replaces the independent time samples with a set of correlated visibilities that have effectively already been coherently integrated in time. Note that these weights are *not* generally the same as the ones derived in Section 5.3 for optimal mapmaking, where measurements will essentially be weighted by an additional factor of the primary beam in fringe-rate space, rather than by the RMS beam weighting suggested here. Put another way, to obtain the full power spectrum sensitivity of an interferometer, it is insufficient to simply square the Fourier amplitudes outputted from a map, even if the mapmaking was optimized to minimize the error bars of the map. Forming the power spectrum in such a way would be equivalent to restricting g_{ij} to a form separable in i and j. This restriction precludes the form for w_{ij} given by Equation (22), which minimizes the error bars of the power spectrum.

Importantly, the result that we have derived here applies only when one is attempting to measure a power spectrum with a single baseline (or multiple baselines with the same baseline vector $\mathbf{b}$). This is a reasonable limit to work in for arrays such as PAPER, where a large fraction of the array's sensitivity comes from instantaneously redundant baselines (Parsons et al. 2012a). For arrays that have less instantaneous redundancy, it becomes more important to combine data from multiple baselines. If multiple baselines are involved, Equation (25) no longer reduces to a single sum over the time axis. Said differently, it is no longer true that the full sensitivity of an array can be obtained by averaging together time-slice-by-time-slice estimates of the power spectrum estimation from fringe-rate filtered data. Instead, the optimal estimator involves a double sum over time, since with multiple baselines of roughly the same length, it is possible for baselines to rotate into one another on the uv plane. That is, rotation synthesis becomes an important contribution to an interferometer's sensitivity.

We note that following fringe-rate filtering, the normalization of the power spectrum estimator must be modified accordingly. This can be seen in Equation (25), where the scalar quantities in front of the sum are different than those found in Equation (15), which is applicable to non-fringe rate filtered data. More generally, suppose we consider an auto-correlation-only estimator of the form

$$\widehat{P} = \sum_i w_i^2 |\widetilde{V}_i|^2 = \sum_i |w_i \widetilde{V}_i|^2, \quad (27)$$

where if w_i is set to $\mu_i^{1/2}$, we recover the optimized estimator derived above. Keeping w_i arbitrary here will be useful in later sections, when we examine estimators that are purposely non-optimal as far as instrumental

noise sensitivity is concerned, but may provide better final results due to a better mitigation of systematics. Like before, we may impose the condition that $\langle \widehat{P} \rangle = P$, which requires that the weights satisfy $S \sum_i w_i^2 \mu_i = 1$ (where we have once again invoked the approximation that $\Omega_{ij}^{pp} = 0$ if $i \neq j$). Equivalently, we may leave our weights initially unnormalized as one forms the combination $w_i \widetilde{V}_i$ in Equation (27), instead compensating with a normalizing denominator. Writing the estimator in this way, one obtains

$$\widehat{P} = \frac{\sum_i |w_i \widetilde{V}_i|^2}{S \sum_j w_j^2 \mu_j} = \left(\frac{\lambda^2}{2k_B}\right)^2 \frac{X^2 Y}{B} \frac{\sum_i |\widetilde{V}^{\rm filt}(t_i)|^2}{\sum_j w_j^2 \mu_j}, \quad (28)$$

where in the last equality we re-inserted the definition of S and invoked Parseval's theorem in the numerator to write the fringe-rate filtered visibilities in the time domain. Our expression now takes precisely the same form as Equation (15), except with fringe-rate filtered visibilities $\widetilde{V}^{\rm filt}(t)$ instead of the original visibilities, and $\sum_j w_j^2 \mu_j$ as a normalizing beam area instead of the integrated square beam Ω_{pp}. That the estimator can be written in such a similar form is unsurprising, since we showed in Equation (6) that fringe-rate filtered visibilities are essentially the same as the original visibilities but with modified primary beams. Indeed, the term $\sum_j w_j^2 \mu_j$ can alternatively be computed by simply evaluating the integral for Ω_{pp} but with the effective primary beam of Equation (7) instead of the original primary beam. Explicitly, the relevant integral is $\frac{1}{B} \int d\Omega d\nu A_\nu^{\rm eff}(\hat{\mathbf{r}}, \nu)^2$, which becomes

$$\frac{1}{B} \int d\Omega d\nu A_\nu(\hat{\mathbf{r}}, \nu)^2 \, w^2 \left(\frac{\nu}{c} \mathbf{b}_t \times \boldsymbol{\omega}_\oplus \cdot \hat{\mathbf{r}} \right), \quad (29)$$

where we made the approximation that $\tilde{\gamma}$ is a reasonably compact (delta function-like) function in Equation (7). Suppose we now evaluate this integral by splitting the sky into rings of constant fringe-rate, i.e., regions within which we have $\mathbf{b}_t \times \boldsymbol{\omega}_\oplus \cdot \hat{\mathbf{r}}$ equal to a constant. By construction, the function w is constant within each of these regions, and denoting its value in the jth region as w_j, the integral becomes

$$\sum_j w_j^2 \frac{1}{B} \int d\Omega d\nu A_j(\hat{\mathbf{r}}, \nu)^2 = \sum_j w_j^2 \mu_j, \quad (30)$$

completing our proof that the normalization factor in Equation (28) is simply Ω_{pp} computed with the effective primary beam.

To be conservative, however, it is important to verify our analytic results using numerical simulations, since a number of approximations were made in our derivations. We use simulated observations of individual point sources positioned at declinations in 5° increments, traversing through the primary beam and fringe pattern as a function of time. As illustrated in Figure 4, we inject point sources of unity flux density and apply the FIR implementation of fringe-rate filtering described in Section 3 to the visibility time-series. The change in amplitude of the filtered visibilities at each point along the time axis can then be identified with a specific position on the sky for that point source. We bin these responses on a HEALpix spherical grid (Górski et al. 2005) to determine the effective beam response at that location. We perform these simulations using both an isotropic primary beam (in Figure 4) and with the PAPER primary beam (in Figure 5). The effective beam responses that we recover are shown in Figure 6. The artifacts that appear every 5° in declination are a result the binned beam model imperfectly interpolating between source tracks at different declinations. Aside from these artifacts, which we emphasize are associated with the reconstruction of the beam model and not with any features exhibited along a source track, these results compare well to the results shown in Figure 7, where we show the effective beam response by looking up the filter response corresponding to the fringe rate at each sky position, i.e., by using Equation (7). As these simulations illustrate, the approximations we have made in our analytic derivation of the effect of fringe-rate filtering on the effective beam response are valid, with errors dominated by the effects of binning and interpolating between the 5° intervals in declination between point-source simulation tracks when reconstructing an effective beam.

As a final check, we perform simulations assigning flux to pixels on the sky with a Gaussian random distribution and simulating the visibilities measured as a function of time for our fiducial baseline. We compute the power spectrum amplitude for this simulated signal and a fringe-rate filtered version according to Equation (15), omitting the primary beam term Ω_{pp} in both cases. We then compute the ratio of the unnormalized $P(k)$ measurements before/after fringe-rate filtering and compare this ratio to the change in effective beam area associated with the chosen fringe-rate filter. Dividing the signal loss associated with fringe-rate filtering by the change in beam area, we obtain a ratio of 1.016 ± 0.011 (1σ), indicating that these ratios are in agreement within the sample noise of our simulation. This numerically legitimizes the approach advocated above, where we used approximate analytic arguments to argue that Equation (15) can be used for power spectrum estimation provided the effective beam squared integral is used instead of the beam squared integral for Ω_{pp}.

4.2. *Minimizing polarization leakage*

In the previous section, we maximized our sensitivity to the power spectrum under the assumption that the measurements were limited by instrumental noise. In practice, however, there may be other sources of noise or systematics that limit our constraints. One example of this is the cross-contamination between Stokes terms in interferometric polarization measurements. Minimizing such contamination is of importance for 21 cm cosmology experiments that rely on the spectral axis to probe the line-of-sight direction at cosmological distances. For these experiments, Faraday rotation combines with a spurious coupling between Stokes terms (typically Q to I) to produce polarization leakage whose spectral structure poses a worrisome foreground to the cosmological signal (Jelić et al. 2008, 2010, 2014; Bernardi et al. 2013; Moore et al. 2013, 2015). Current interferometers targeting the 21cm signal at cosmological distances (LOFAR, MWA, PAPER, HERA, CHIME, LEDA) all employ linearly polarized feeds, primarily because of their ease of construction and ability to co-locate elements sensitive to

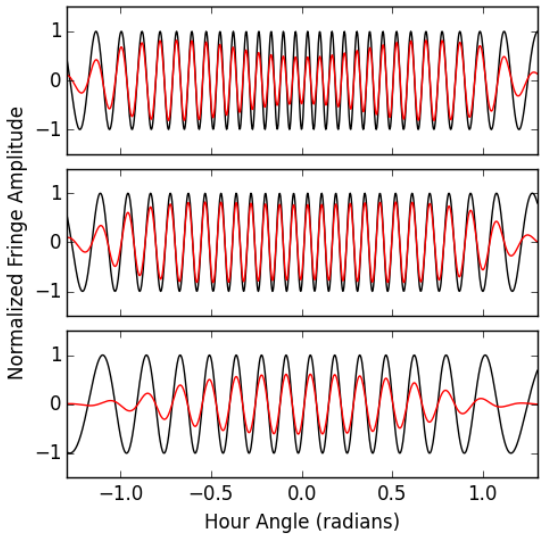

FIG. 4.— The real component of the fringe amplitude simulated for our fiducial baseline deployed at a latitude of −30° before (black) and after (red) the application of the fringe-rate filter described in Section 3. From top to bottom, the panels illustrate fringes for point sources passing through the fringe pattern at declinations of 0°, −30°, and −60°, respectively. In this simulation, antenna elements have isotropic primary beams.

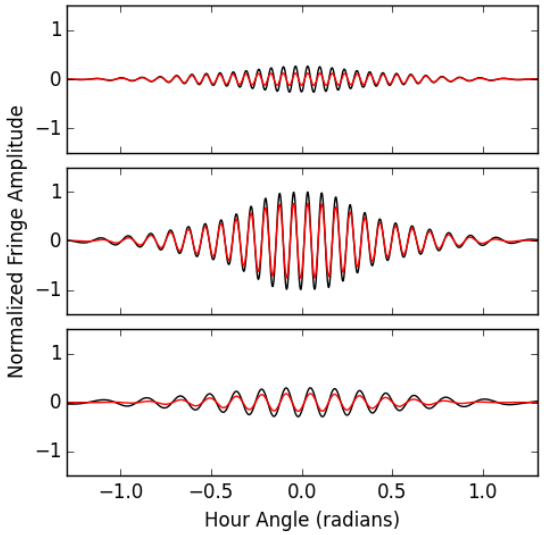

FIG. 5.— Same as Figure 4, but for antenna elements with the PAPER primary beam (Parsons et al. 2010; Pober et al. 2012).

orthogonal polarizations. However, orthogonal linearly polarized feeds in practice have primary beam responses that do not match. As described in Moore et al. (2013) and Jelić et al. (2010), if left uncorrected, the unmatched beam response between visibilities $V_{\rm XX}$ and $V_{\rm YY}$ measuring the XX and YY polarization products, respectively, is the dominant source of polarization leakage in the Stokes I measurement $V_{\rm I} \equiv (V_{\rm XX}+V_{\rm YY})/2$ for linearly polarized feeds.

With an accurate beam model, it is trivial to rescale $V_{\rm XX}$ and $V_{\rm YY}$ so that the XX and YY beam responses match in one particular direction (typically the zenith). Their sum, $V_{\rm I}$, then represents a perfect probe of the Stokes I parameter in that chosen direction, but will contain contamination from $V_{\rm Q} \equiv (V_{\rm XX} - V_{\rm YY})/2$ in directions where the XX and YY beam responses do not match. More precisely, suppose the XX polarization product has an antenna power beam of $A_{\rm XX}$, while the YY polarization product has an antenna power beam of $A_{\rm YY}$. The primary beam that we have denoted $A(\hat{\mathbf{r}})$ in previous sections is given by $(A_{\rm XX} + A_{\rm YY})/2$, which will be labeled as $A^I(\hat{\mathbf{r}})$ in this section to emphasize its meaning as the antenna response to the Stokes I sky. Defining $A^Q(\hat{\mathbf{r}}) \equiv (A_{\rm XX} - A_{\rm YY})/2$, the Stokes I visibility takes the form (see, e.g., Moore et al. 2015)

$$V_{b\nu}(t) = \int d\Omega\, I_\nu(\hat{\mathbf{r}}) A^I_\nu(\hat{\mathbf{r}},t) \exp\left[-i2\pi\frac{\nu}{c}\mathbf{b}(t)\cdot\hat{\mathbf{r}}\right] + \int d\Omega\, Q_\nu(\hat{\mathbf{r}}) A^Q_\nu(\hat{\mathbf{r}},t) \exp\left[-i2\pi\frac{\nu}{c}\mathbf{b}(t)\cdot\hat{\mathbf{r}}\right], \quad (31)$$

where Q_ν denotes the Stokes Q sky at frequency ν. The second term represents the Q to I leakage in the visibility measurement, which does not vanish if $A_{\rm XX} \neq A_{\rm YY}$, i.e., if there are any asymmetries in the beams. For a single baseline at a given instant in time, this leakage term cannot be eliminated. The heart of the problem is the impossibility of creating a match between a pair of two-dimensional functions (the XX and YY beam responses) with a single degree of freedom (the amplitude of $V_{\rm XX}$ relative to $V_{\rm YY}$). In order to improve the match between polarization beams in interferometric measurements, many interferometric measurements from distinct points in the uv plane will have to be combined with appropriate weights to effect a re-weighting of the sky along two dimensions.

A commonly used technique for correcting the mismatch between the XX and YY polarization beams is to separately image these polarization products, correct each pixel in each image using modeled beam responses, and then to sum the corrected images together to form a Stokes I map (e.g., Sullivan et al. 2012; Bernardi et al. 2013; Asad et al. 2015). Mathematically, this technique is identical to convolving the sampled uv plane by the Fourier transform of the directionally-dependent correction applied in the image domain. For an ideal array that samples the uv plane at scales significantly finer than the aperture of a single element, this technique can in principle perfectly correct mismatches between the XX and YY polarization beams. In practice, the success that can be achieved with this technique depends strongly on an array's uv sampling pattern.

Take, for example, the case of a sparsely sampled uv plane where the spacing between uv samples is much greater than the aperture scale of a single element. In this case, the beam correction described above convolves each uv sample with a kernel whose size scales roughly as the size of the aperture of a single element in wavelengths. Since this kernel is much smaller than the spacing between uv samples, each point in the convolved uv plane is dominated by the product of a kernel weight and a single visibility measurement. As such, for a chosen uv coordinate, the level of leakage in the Stokes I uv plane effectively reduces to what can be achieved using a single number to rescale $V_{\rm XX}$ and $V_{\rm YY}$ before summing.

In image domain, the baseline fringe pattern has been

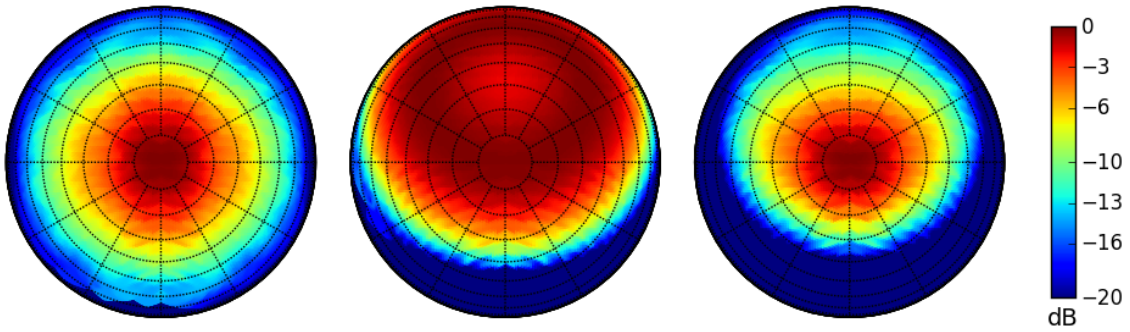

FIG. 6.— The effective primary beam response of a baseline, as reconstructed from point-source simulations described at the end of Section 4.1 and illustrated in Figures 4 and 5. The left panel shows that PAPER's model beam response is recovered from unfiltered visibilities; the center panel illustrates the beam weighting that results from applying a fringe-rate filter tuned to optimize sensitivity for power spectrum measurements, assuming an isotropic primary beam; the right panel shows the effective response after applying this fringe-rate filter to data including PAPER's model beam response. The periodic structure along the vertical axis is an artifact of reconstructing beams from point sources spaced every 5° in declination; we emphasize that they are not associated with any fundamental structure associated with fringe-rate filtering.

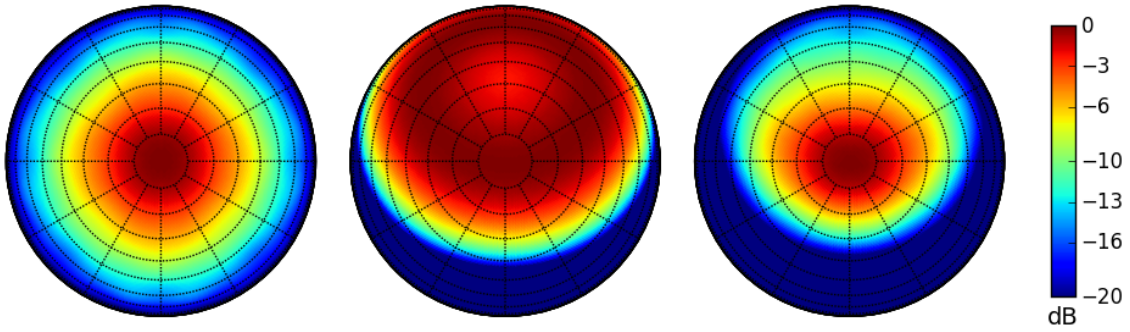

FIG. 7.— The model primary beam response of a baseline, simulated analytically using the interpretation of fringe-rate filtering as a spatial filter acting along fringe-rate contours. Panels follow the same order as in Figure 6.

integrated across the sky weighted by two different (XX, YY) beam responses to produce a visibility. This means that an interferometric baseline for these two polarizations samples the uv plane convolved by two different kernels — the Fourier transforms of the XX, YY beams, respectively. Because the resulting $V_{\rm XX}$ and $V_{\rm YY}$ visibilities no longer retain any direction-dependent information (they were integrated over angle) $V_{\rm XX}$ and $V_{\rm YY}$ cannot be re-weighted in a way that produces pure Stokes I response in all directions, except in the trivial case where the YY beam differs from the XX beam by a direction-independent multiplicative factor. Said another way, given a single degree of freedom in the relative weighting of $V_{\rm XX}$ and $V_{\rm YY}$, it is not possible to guarantee a pure Stokes I response in more than one direction.

As mentioned above, it is common practice in synthesis imaging to grid visibilities in the uv plane with convolving kernels that contain direction-dependent information. Examples include W-projection kernels (Cornwell et al. 2003), the Fourier transform of the primary beam used in optimal map making (Morales & Matejek 2009b; Bhatnagar et al. 2008), and (relevant here) the Fourier transform of the inverse of direction-dependent Mueller matrices (Tasse et al. 2013). Indeed, in the case of a completely sampled uv plane, the inverse Mueller beam kernel would perfectly match the XX and YY beams. However, as the sampling of the uv plane becomes sparser, the removal of samples removes degrees of freedom in weighting as a function of direction. In the limiting case of a very sparse array where the convolving kernels of uv samples do not overlap, we revert to the single direction-independent weighting factor above.

Another way of thinking about this is to realize that, by gridding visibilities in the uv plane with a convolving kernel, we are attempting to reassign sky intensity from the visibility, which was integrated over in equation 1, back to its original location so that it may be re-weighted to match the XX and YY polarization responses. This inversion is only well posed in the case of a completely sampled uv plane. As we lose samples, we lose the ability to reconstruct the original sky intensity, and hence, to perfectly match the XX and YY polarization responses as a function of direction. In the case of a sparsely sampled uv plane, it is only possible to construct a pure Stokes I response as a function of direction if one has sufficient additional information about the sky (for example, if it sparsely populated by point sources) that one can reconstruct the missing modes and hence, fully invert the uv plane.

For cases where uv sampling falls somewhere between the sparse and the oversampled cases described above, the level of primary beam correction that can be realized

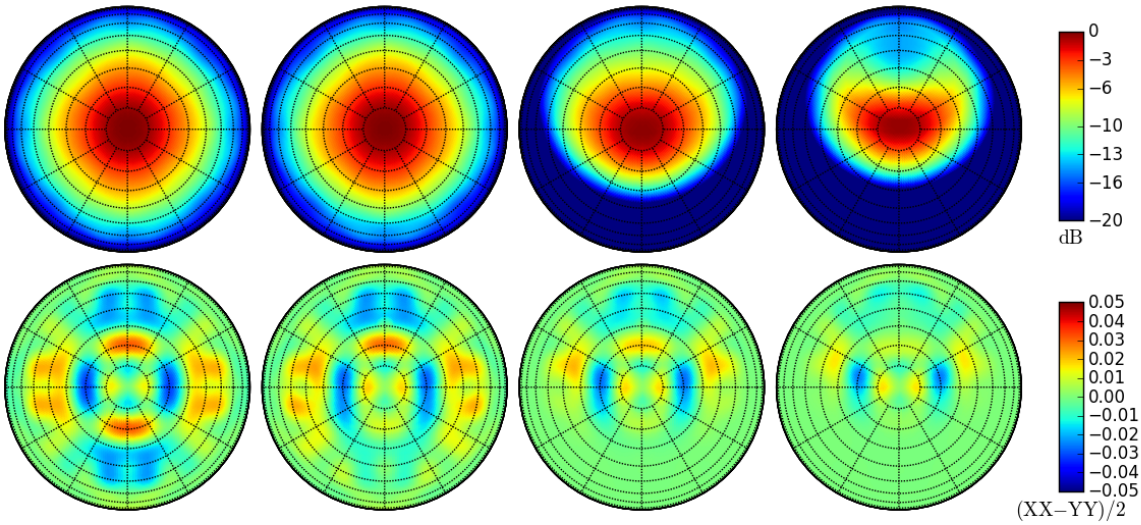

FIG. 8.— Beam response patterns illustrating the application of fringe-rate filtering to minimizing polarization leakage. Panels from left to right illustrate the response before matching XX and YY polarization beams, the response after such matching, the subsequent application of an optimal power-spectrum sensitivity filter, and the application of a filter optimizing both sensitivity and polarization match. The top row depicts the Stokes I beam response in logarithmic units; the bottom row shows the polarization match $(\mathrm{XX}-\mathrm{YY})/2$ in linear units. The fringe-rate filters optimizing sensitivity and polarization match correspond to the red and green curves in Figure 3, respectively. The third column most closely corresponds to the fringe-rate filter applied in Ali et al. (2015).

is more complicated. Ultimately, the Fourier relationship between the uv plane and the image dictates that samples that are nearby to one another in the uv plane enable primary beam corrections on the largest angular scales, while samples that are farther apart contribute to corrections on finer angular scales, with the orientation of the samples relative to one another dictating the axis along which such corrections take effect in image domain. Typically, earth-rotation synthesis is required to sample the uv plane densely enough to allow for effective beam correction for diffuse structure[8] configurations are not dense enough to fully correct the beam even then. One particularly relevant case that falls in this last category are many of the maximum redundancy configurations currently favored by several 21 cm cosmology experiments for their sensitivity benefits (Parsons et al. 2012a, 2014).

However, even in the single-baseline case, earth-rotation synthesis provides dense uv sampling along one direction: the direction the baseline traverses in the uv plane. The appropriate convolution kernel can combine samples along this track so as to correct the primary beam mismatch along one axis. Of course, what we have just described—a convolution kernel acting along a time series of samples from a single baseline—is precisely fringe-rate filtering. Put another way, since fringe-rate filtering has the effect of modifying one's primary beams, it is possible to tailor fringe-rates to improve the match between the XX and YY polarization beams. In the case of sparse array sampling, fringe-rate filtering reproduces what can be achieved by independently imaging the polarization products. While this is not as effective at mitigating polarization leakage as can be achieved through imaging in the dense sampling case, we will now show that it nonetheless represents a substantial improvement over the naive summing of XX and YY visibility measurements.

Suppose for every frequency channel in our observations, we compute the RMS $A_{\mathrm{XX}}(\hat{\mathbf{r}})$ along the spatial ring corresponding to each fringe-rate bin, using the relation $f \approx \hat{\mathbf{r}}\cdot\mathbf{b}\times\boldsymbol{\omega}_{\oplus}\nu/c$ derived in Section 2. The result is a one-dimensional beam profile in fringe-rate space. Repeating the same exercise for A_{YY}, one can then form the ratio between the two profiles, quantifying the mismatch between the XX and YY beams in a one-dimensional projection. This ratio is plotted for PAPER as the blue dashed curve in the top panel of Figure 3. If one then uses this curve as a set of fringe-rate weights for the V_{YY}, the resulting effective beams will be more closely matched to one another, which we will demonstrate later in this section when we quantitatively estimate polarization leakage for various fringe-rate filtering schemes. Note that we use the RMS rather than the beam itself because ultimately we seek a measurement of the power spectrum, which is a squared quantity.

In the second column of Figure 8, we show the effective beam (top row) and the beam mismatch (i.e., $A^{Q}(\hat{\mathbf{r}})$; bottom row) after performing our beam matching procedure. One sees that the mismatch is slightly mitigated compared to the original unweighted case (leftmost column). However, substantial power remains in the mismatched beam $A^{Q}(\hat{\mathbf{r}})$ because fringe-rate filtering can only alter beam shapes in a one-dimensional family of fringe-rate rings (as we discussed in Section 2). A perfect matching would require a set of two-dimensional weights.

Interestingly, the beam matching can be improved

[8] As mentioned earlier in this section, it is possible to correct for direction-dependent polarization beams toward sparse point sources using more widely separated visibility samples. It is diffuse structure (which is more localized in the uv plane) that poses the greatest challenge.

with one further weighting that is not specifically targeted at polarization matching, namely the sensitivity-optimized weighting proposed in Section 4.1. These were the weights applied in Ali et al. (2015), and the results of applying these weights to combined V_I visibilities following the beam-matching procedure are shown in the third column of Figure 8. While this extra weighting step was not designed to deal with polarization in mind, it is still effective at mitigating leakage because it down-weights portions of the sky that have some of the worst beam mismatches.

Finally, one may use knowledge of the mismatched beam $A^Q(\hat{\mathbf{r}})$ to further inform one's weighting. In particular, one may take the scheme that we have outlined so far (applying fringe-rate filters to first match XX and YY beams, then to maximize sensitivity), and add one final filter that downweights portions of the sky where the mismatched beam is known to be large. In the fourth column of Figure 8, we additionally weight each fringe-rate by the ratio of the RMS $A^I(\hat{\mathbf{r}})$ fringe-rate beam profile to the RMS $A^Q(\hat{\mathbf{r}})$ fringe-rate beam profile. The effective fringe-rate weights that are applied to the Stokes I visibilities are given by the green curve of Figure 3. From Figure 8, we see that this hybrid sensitivity- and leakage-optimized weighting significantly reduces the beam mismatch.

In minimizing polarization leakage, one must be careful not to enact an overly aggressive scheme that narrows the effective beam to such an extent that there is a substantial loss of power spectrum sensitivity. To quantify this trade-off, consider a generalization of Equation (9) to include polarization leakage. Performing a derivation similar to the one in Section 4.1 yields

$$\langle \widetilde{V}_i(\mathbf{u},\eta)\widetilde{V}_j(\mathbf{u},\eta)^*\rangle \approx S\left[P_I(\mathbf{k})\Omega_{ij}^{II} + P_Q(\mathbf{k})\Omega_{ij}^{QQ}\right], \quad (32)$$

where $S \equiv (B/X^2Y)(2k_B/\lambda^2)^2$ and $(Xk_x, Xk_y, Yk_z) \equiv 2\pi(u,v,\eta)$, as before. The indices i and j again index fringe-rate bins. Beam integrals Ω^{QQ} and Ω^{II} are defined analogously to Equation (11), but with superscripts indicating the Stokes components. The Stokes I power spectrum P_I is the same as the power spectrum P defined in Section 4.1, with the new notation serving only to distinguish it from the Stokes Q power P_Q. In deriving Equation (32), we made two key assumptions. The first is that the I and Q contributions to the sky are on average uncorrelated. Empirically, this appears to be the case (Wieringa et al. 1993; Gaensler et al. 2001; Bernardi et al. 2003), and physically, one expects this to be so since foreground interstellar medium clouds affect polarized and unpolarized emission differently. The second assumption is that the polarized sky is describable by a power spectrum. At some level, one expects this assumption to fail, as there exist bright polarized sources that are not accounted for in a random realization of some power spectrum. However, since our ultimate goal is to measure power spectra, it is reasonable to define an effective Stokes Q power spectrum.

Like before, rather than considering individual correlations between different fringe-rate bins, one may insert fringe-rate filtered delay-space visibilities $\widetilde{V}^{\rm filt}$ on the left-hand side of Equation (32). The same equation then holds with the fringe-rate bin indices omitted (since the fringe-rates have already been combined in a weighted combination) and Ω^{QQ} and Ω^{II} replaced by the square integral of the *effective* beams, which we denote $\Omega^{QQ}_{\rm eff}$ and $\Omega^{II}_{\rm eff}$, respectively. One sees then that a suitable estimator of the Stokes I power spectrum is

$$\widehat{P}_I(\mathbf{k}) = \frac{|\widetilde{V}^{\rm filt}(\mathbf{u},\eta)|^2}{S\Omega^{II}_{\rm eff}}. \quad (33)$$

Ensemble-averaging both sides and inserting the fringe-rate filtered version of Equation (32) yields

$$\langle \widehat{P}_I(\mathbf{k})\rangle = P_I(\mathbf{k}) + \frac{\Omega^{QQ}_{\rm eff}}{\Omega^{II}_{\rm eff}} P_Q(\mathbf{k}). \quad (34)$$

The second term is the polarization leakage in our final power spectrum estimate. The key quantity is $\Omega^{QQ}_{\rm eff}/\Omega^{II}_{\rm eff}$, which quantifies the effectiveness of one's polarization leakage suppression scheme. On the other hand, if the fringe-rate filtered visibilities have an instrumental noise variance of $\sigma^2_{\rm filt}$ (as we assumed in the previous section), the instrumental noise errors $\Delta\widehat{P}_I$ in our power spectrum estimator are

$$\Delta\widehat{P}_I(\mathbf{k}) \equiv \left(\mathrm{Var}\left[\widehat{P}_I(\mathbf{k})\right]\right)^{\frac{1}{2}} = \frac{\sigma^2_{\rm filt}}{S\Omega^{II}_{\rm eff}}. \quad (35)$$

For instrumental noise sensitivity, then, the crucial quantity is $\sigma^2_{\rm filt}/\Omega^{II}_{\rm eff}$. In Table 1 we list the value of this metric (rightmost column) for the various weighting schemes considered in this section, normalized to the value obtained with no fringe-rate filtering. We also show the integrated power beam $\int A^{\rm eff}(\hat{\mathbf{r}})d\Omega$, the square-integrated power beam $\Omega^{II}_{\rm eff} \equiv \int A^{\rm eff}(\hat{\mathbf{r}})^2 d\Omega$, the effective (noise equivalent) integration time implied by each fringe-rate filter, and the fractional power spectrum leakage $\Omega^{QQ}_{\rm eff}/\Omega^{II}_{\rm eff}$.

At this point, we should emphasize that because fringe-rate filtering performs a weighted combination of samples over a wide time interval, the instrumental noise errors $\Delta\widehat{P}_I$ are substantially correlated between nearby times samples. The number of independent modes preserved for power spectrum analysis corresponds to the number and weighting of the filtered fringe-rate modes. Even without filtering, quantifying the number of sampled modes is non-trivial, as the time dependence of a baselines projection (and hence, uv mode sampling) toward a patch of sky varies substantially over the wide fields of view of low-frequency interferometers. As discussed in Parsons et al. (2014), to accurately account for the statistical interdependence of these modes in either the fringe-rate filtering or square time-domain integration case, it is necessary to use bootstrapping to determine errorbars.

With only a polarization-matching filter applied weighting the YY polarization beam to best match the XX polarization beam (second row of Table 1; second column of Figure 8), we see that the fractional polarization leakage is reduced from 1.64×10^{-3} to 1.26×10^{-3}. Applying an additional sensitivity weighting step from Section 4.1 (third row of Table 1; third column of Figure 8) further reduces the fractional polarization leakage to 0.88×10^{-3}, and though there is a 25% reduction in effec-

TABLE 1
Sensitivity and polarization leakage metrics for different fringe-rate filters.

	Integrated power beam [sr]	Square-integrated power beam $\Omega^{II}_{\rm eff}$ [sr]	Effective integration time [s]	Polarization leakage ($\Omega^{QQ}_{\rm eff}/\Omega^{II}_{\rm eff}$)	Normalized power spectrum sensitivity
No fringe-rate filter	0.74	0.32	420	1.64×10^{-3}	1.
Polarization-matched fringe-rate filter	0.74	0.32	420	1.26×10^{-3}	1.
Sensitivity-optimized fringe-rate filter	0.51	0.24	2930	0.88×10^{-3}	1.9
Polarization and sensitivity-optimized	0.36	0.15	5570	0.68×10^{-3}	1.7

tive beam area, the power spectrum sensitivity is boosted by almost a factor of two. This improvement takes into account the improvement in sensitivity of each independent power-spectrum sample, but also the corresponding decrease in the number of independent time samples available. (Recall that fringe-rate filtering effectively averages together nearby time samples, and thus returns a time series where instrumental noise is no longer independent from time slice to time slice). Finally, applying another filter to minimize leakage (fourth row of Table 1; fourth column of Figure 8; green curve of Figure 3), results in another $\sim 25\%$ decrease in polarization leakage with only a $\sim$10% degradation in sensitivity. This degradation is associated with the longer integration times required to produce the narrow fringe-rate weighting profile needed for polarization matching. This results in sensitive regions of the beam being underweighted because of their leakage.

Our final, best-performing weighting scheme for minimizing polarization was in some sense a rather arbitrary weighting. This is unavoidable, since the power spectrum of low-frequency polarized emission is unknown at fine angular scales. We now show how our final weighting step may be optimized if the polarized power spectrum is known. Suppose we estimate the power spectrum by forming weighted averages of the visibility cross-correlations between all pairs of fringe-rate bins. The general form for such an estimator is given by Equation (17). However, as we found in Section 4.1, one may approximate the different fringe-rate bins as being uncorrelated. The same approximation can be made here, allowing us to define μ^I and μ^Q by letting $\Omega^{II}_{ij} \equiv \delta_{ij}\mu^I_i$ and $\Omega^{QQ}_{ij} \equiv \delta_{ij}\mu^Q_i$. With no correlations between fringe-rate bins, it becomes sufficient to once again use Equation (27). Taking the ensemble average of Equation (27) but this time including polarization leakage terms, i.e., inserting Equation (32) rather than Equation (9) gives

$$\langle \widehat{P} \rangle = \left(S \sum_i w_i^2 \mu_i^I \right) P_I + \left(S \sum_i w_i^2 \mu_i^Q \right) P_Q. \quad (36)$$

If, as before, we require our weights to be normalized so that $S\sum_i w_i^2 \mu_i^I = 1$, the first term gives an unbiased estimator of the Stokes I power spectrum. The second term is the bias in our final power spectrum estimate due to leakage from Q to I. Ideally, the weights are chosen to mitigate this contribution. However, in attempting to minimize this systematic, one must be careful not to pick weights that dramatically amplify the instrumental noise contribution. We therefore choose to minimize an overall "variance" that is the noise variance of Equation (19) plus the square of the Stokes Q power spectrum bias (essentially adding the error bars from instrumental noise and polarization leakage in quadrature). In other words, we minimize

$$\mathcal{L} = \sigma^4 \mathbf{m}\cdot\mathbf{m} + (SP_Q)^2 \left(\mathbf{m}\cdot\boldsymbol{\mu}^Q\right)^2 - \lambda \mathbf{m}\cdot\boldsymbol{\mu}^I, \quad (37)$$

where we have introduced a Lagrange multiplier λ to enforce our normalization constraint, and have grouped the different fringe-rate weights into a vector $\mathbf{m} \equiv (w_1^2, w_2^2, \dots)$, with the beam overlap integrals similarly grouped into vectors $\boldsymbol{\mu}^I$ and $\boldsymbol{\mu}^Q$. Now, notice that if we define

$$\mathbf{H} \equiv \sigma^4 \mathbf{I} + (SP_Q)^2 \boldsymbol{\mu}^Q (\boldsymbol{\mu}^Q)^t, \quad (38)$$

our expression can be written as

$$\mathcal{L} = \mathbf{m}^t \mathbf{H} \mathbf{m} - \lambda \mathbf{m}^t \boldsymbol{\mu}^I. \quad (39)$$

Differentiating this, setting the result to zero, and solving for the normalized weights gives

$$\mathbf{m} = \frac{\mathbf{H}^{-1}\boldsymbol{\mu}^I}{S(\boldsymbol{\mu}^I)^t \mathbf{H}^{-1} \boldsymbol{\mu}^I}. \quad (40)$$

This expression involves $\mathbf{H}^{-1}$, and can be evaluated explicitly in our case using the Woodbury formula, giving

$$\mathbf{H}^{-1} = \sigma^{-4} \left[\mathbf{I} - \frac{P_Q^2 \boldsymbol{\mu}^Q (\boldsymbol{\mu}^Q)^t}{P_N^2 + P_Q^2 \boldsymbol{\mu}^Q \cdot \boldsymbol{\mu}^Q} \right], \quad (41)$$

where P_N is the noise power spectrum, which is equal to the RMS visibility noise divided by by S. If the sky were completely unpolarized ($P_Q = 0$) or there were no polarization leakage due to mismatched beams ($\boldsymbol{\mu}^Q = 0$), then $\mathbf{H}^{-1} \propto \mathbf{I}$, and the optimal weighting would take the form $\mathbf{m} \propto \boldsymbol{\mu}^I$. Recalling that each component of a $\boldsymbol{\mu}$ vector corresponds to a different fringe rate, we see that $\mathbf{m} \propto \boldsymbol{\mu}^I$ is precisely the result derived in Section 4.1, where each fringe-rate bin was weighted by the integrated square of the beam within the bin.

With polarization leakage, Equation (40) shows that as a first step, one should still weight by the square-integrated beam within each fringe-rate bin. However, one should then further weight by the more complicated form for $\mathbf{H}^{-1}$ given by Equation (41). To gain some intuition for this operation, consider the limit of zero instrumental noise. One then obtains

$$\mathbf{H}^{-1}\Big|_{P_N=0} \propto \mathbf{I} - \boldsymbol{\mu}^Q (\boldsymbol{\mu}^Q \cdot \boldsymbol{\mu}^Q)^{-1} (\boldsymbol{\mu}^Q)^t, \quad (42)$$

which we immediately recognize as a projection matrix that projects out the vector $\boldsymbol{\mu}^Q$. Since the components of $\boldsymbol{\mu}^Q$ encode the polarization leakage response in various fringe-rate bins, $\mathbf{H}^{-1}$ projects out linear combinations of fringe rates that are indicative of Q to I leakage due to

beam asymmetries. At the other extreme, if the polarization power spectrum is weak compared to the noise power spectrum, we have

$$\mathbf{H}^{-1}\Big|_{P_N \gg P_Q} \propto \mathbf{I} - \left(\frac{P_Q}{P_N}\right)^2 \boldsymbol{\mu}^Q (\boldsymbol{\mu}^Q)^t, \qquad (43)$$

which is similar to a projection operator, but with the unwanted or projected piece tempered by the square of the small parameter P_Q/P_N. Intuitively, even modes that are contaminated by polarization leakage contain cosmological information, and it is advantageous to avoid too drastic a projection if possible. An overly aggressive subtraction of leakage modes results in a loss of cosmological signal, which when corrected for by the altered normalization of the power spectrum, results in a magnification of the error bars. For the general case of intermediate noise levels, the $\mathbf{H}^{-1}$ matrix smoothly interpolates between the two extremes.

While mathematically optimal, the fringe-rate weighting proposed here is currently difficult to put into practice. This is because the Q power spectrum P_Q has yet to be positively measured at low frequencies. For example, recent measurements in Moore et al. (2015) provide only upper limits on P_Q, at least at the k-scales that are the most promising for a first detection of the 21cm signal.

4.3. *Minimizing Instrumental Systematics and Off-Axis Foregrounds*

A final application of beam sculpting with fringe-rate filters targets the suppression of systematics in data. We will consider two systematics: additive phase terms associated with instrumental crosstalk, and sidelobes associated with celestial emission outside of the primary field of interest. Both of these applications are closely aligned with the original application of fringe-rate filters described in Parsons & Backer (2009). Reduction of sidelobe contamination (as well as radio frequency interference mitigation) using a similar technique was also discussed in Offringa et al. (2012).

For the purposes of this discussion, we consider instrumental crosstalk to be a spurious correlation introduced between otherwise uncorrelated signals as a result of electromagnetic coupling in the instrument (typically between adjacent, unshielded signal lines) or because a non-celestial source has injected a correlated signal (e.g. switching noise on power supplies). Although crosstalk can be suppressed using phase switching (Ryle 1952), it is always present at some level in interferometric observations. If it is temporally stable, however, it is possible to significantly suppress crosstalk in data by averaging visibilities over a long period (so that the fringing celestial signal washes out) and then subtracting the average complex additive offset from the data. This technique has long been applied to, e.g., PAPER observations (Parsons et al. 2010; Pober et al. 2013; Jacobs et al. 2014; Parsons et al. 2014; Ali et al. 2015).

As a time-domain filter, this crosstalk removal technique can also naturally be understood as a notch filter for removing signals with zero fringe-rate. Because crosstalk removal uses a finite time interval for computing the average, applying this notch fringe-rate filter has the effect of removing emission from the region of sky corresponding to the zero fringe-rate bin. As illustrated in Figure 2, for our fiducial baseline, this corresponds to the unshaded region intersecting the south celestial pole. For PAPER, this region is sufficiently low in the beam that its removal has little impact, but in general, subsequent analysis of crosstalk-removed data may require accounting for the beam-sculpting effects of the crosstalk removal filter.

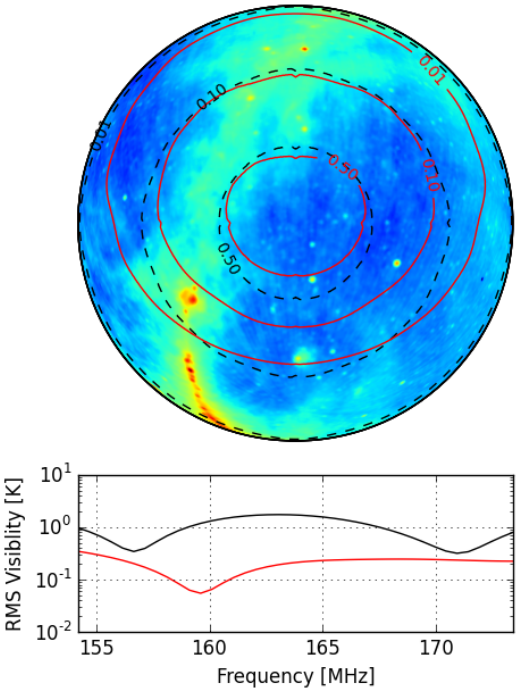

FIG. 9.— Top: A projection of the global sky model of de Oliveira-Costa et al. (2008) as viewed from latitude −30° at a local sidereal time of 05:44, overlaid with the PAPER beam response contours before (black) and after (red) the application of a fringe-rate filter optimized for sensitivity. Bottom: a simulation of the corresponding RMS visibilities for a 30-m east-west baseline before (black) and after (red) the same fringe-rate filtering. The reduced presence of foregrounds in the simulated visibilities after fringe-rate filtering can be understood as coming from the spatial filter enacted by fringe-rate filtering.

Thus, when considering instrumental systematics, there may be additional criteria that influence one's choice of fringe-rate filter besides optimizing signal-to-noise; one may choose to excise the zero fringe-rate bin to improve data quality at a very modest cost to sensitivity. Similarly, it is common to encounter situations where celestial emission that is low in the primary beam is bright enough to introduce undesirable sidelobe structure or other systematics in observations targeting an area nearer to beam center. In this case, one may again find it desirable to depart from optimal weighting derived in Section 4.1 by further down-weighting regions of low sensitivity in order to gain improvements in foreground systematics. This application of fringe-rate filtering is particularly relevant for 21cm cosmology experiments where approximately Gaussian signals are overlaid with highly non-Gaussian foregrounds. Fringe-rate filters that are informed by the angular structure in foreground models can substantially suppress foreground systematics while having little impact on a statistically isotropic

Gaussian signal.

In Figure 9 we illustrate this technique using simulations of a single baseline. The sky model that we employ is that of de Oliveira-Costa et al. (2008), and in the top panel of the figure we show the field of view of our simulated interferometer at a particular instant in time. Primary beam contours for the PAPER instrument are shown in dashed black, and one sees that there is substantial power from bright Galactic plane emission. In solid red are the contours for the effective primary beam following the application of a power-spectrum optimized fringe-rate filter (see Section 4.1). The reduced effective beam size means that less of the Galactic plane contributes to the measured power. This can be seen in the bottom panel of Figure 9, where we show the corresponding RMS visibilities. The red curve is clearly reduced compared to the black curve, indicating reduced foreground contamination. If desired, the contamination can be further mitigated by picking a different set of fringe-rate weights that downweight the edges of the primary beam even more. However, this extra foreground suppression must be carefully balanced with the increase in instrumental noise error bars that inevitably results from the narrowing of the effective beam beyond the size implied by the optimized procedure of Section 4.1.

5. FRINGE-RATE FILTERING AS MAPMAKING FROM TIME-ORDERED DATA

In the previous sections, we have focused on applications of fringe-rate filtering that operate on a single baseline basis. These applications are particularly powerful for maximally redundant arrays such at PAPER, which have most of their sensitivity concentrated in multiple identical copies of a small handful of baseline types. By design, maximally redundant arrays are not optimized for imaging, which instead require arrays that sample a large number of unique baselines. In this section, we turn our attention to such arrays, tackling the imaging (i.e., mapmaking) problem for multi-baseline arrays. We will find once again that the fringe-rate space is particularly well-suited for implementing time integration for interferometric data.

Suppose our time-ordered visibilities are grouped into a measurement vector $\mathbf{v}$ of length $N_b N_t$, where N_b is the number of baselines, and N_t is the number of snapshots taken in time. If we represent the true sky as a vector $\mathbf{x}$ of length $N_{\rm pix}$, and our instrument's response as a matrix $\mathbf{A}$ of size $N_b N_t \times N_{\rm pix}$, the measurement equation is given by

$$\mathbf{v} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{44}$$

where $\mathbf{n}$ is a noise vector. Note that in this general form, Equation (44) is not basis-specific. For example, while it is often useful to think of $\mathbf{x}$ as a vector containing a list of temperatures in a set of pixels on the sky (hence the variable name $N_{\rm pix}$), it is equally valid to employ another basis, such as spherical harmonics. Similarly, while we call $\mathbf{v}$ the time-ordered data, it need not be a time series, and in fact, we will see that a description in fringe-rate space is in fact quite natural.

Given our measurement $\mathbf{v}$, the optimal estimator $\hat{\mathbf{x}}$ of the true sky $\mathbf{x}$ is given by (Tegmark 1997; Morales & Matejek 2009a; Dillon et al. 2015)

$$\hat{\mathbf{x}} = \mathbf{M}\mathbf{A}^\dagger\mathbf{N}^{-1}\mathbf{v}, \tag{45}$$

where $\mathbf{M}$ is some invertible matrix chosen by the data analyst, the dagger signifies an adjoint, and $\mathbf{N}$ is the noise covariance matrix, defined as $\langle \mathbf{n}\mathbf{n}^\dagger \rangle$, with angled brackets denoting an ensemble average. Again, our vector/matrix expressions are basis-independent, so even though the formation of $\hat{\mathbf{x}}$ is often described as "mapmaking", it need not correspond to spatial imaging in the traditional sense of the word. A similar approach can also be found in "A-projection" algorithms (Bhatnagar et al. 2008, 2013; Tasse et al. 2013).

The error bars on the estimator $\hat{\mathbf{x}}$ are obtained by computing the square root of the diagonal elements of the covariance $\mathbf{\Sigma}$, which is given by

$$\mathbf{\Sigma} \equiv \langle (\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^\dagger \rangle = \mathbf{M}\mathbf{A}^\dagger\mathbf{N}^{-1}\mathbf{A}\mathbf{M}^\dagger. \tag{46}$$

With a suitable choice of $\mathbf{M}$, the estimator given by Equation (45) minimizes the variance. Regardless of one's choice, however, Equation (45) can be shown to be lossless (Tegmark 1997), in the sense that any quantities (such as power spectra) formed further downstream in one's analysis will have identically small error bars whether one forms these data products from $\hat{\mathbf{x}}$ or chooses to work with the larger and more cumbersome set of original data $\mathbf{v}$.

In principle, Equation (45) is all that is needed to optimally estimate the true sky. One simply forms the relevant matrices and performs the requisite matrix inversions and multiplications. However, this is computationally infeasible in practice, given that modern-day interferometers are comprised of a large number of baselines operating over long integration times, resulting in rather large matrices. This is what motivated the authors of Shaw et al. (2013b) to propose their m-mode formalism, essentially rendering many of the relevant matrices sparse, making them computationally easy to manipulate. While the m-mode formalism is a general framework that can be used to solve a variety of problems (such as mitigating foreground contamination), our goal here is to develop similarly convenient techniques for the mapmaking problem (i.e., the formation of $\hat{\mathbf{x}}$), with much detail devoted to the intuition behind how our optimal estimator operates for an interferometer.

5.1. *The general sub-optimality of time integration*

We begin by showing that it is suboptimal to make maps by integrating visibilities in time. Writing out Equation (1) for the visibility $V_{b\nu}(t)$ with an explicit coordinate system, we have

$$V_{b\nu}(t) = \int A_\nu(\hat{\mathbf{r}}, t) I_\nu(\hat{\mathbf{r}}) \exp\left[-i2\pi\left(\frac{b_y}{\lambda}\cos\eta\sin\delta\right)\right] \times \exp\left[-i2\pi\left(\frac{b_0}{\lambda}\cos\delta\sin(\alpha - \omega_\oplus t)\right)\right] d\Omega + n(t), \tag{47}$$

where $n(t)$ is the instrumental noise, α and δ are the right ascension and declination, respectively, η is the geographic latitude of the array, and $b_0 \equiv \sqrt{b_x^2 + b_y^2 \sin^2\eta}$, where b_x and b_y are the east-west and north-south baseline lengths, respectively. We have chosen our definition

of $t = 0$ to conveniently absorb an arbitrary constant phase. Like before, we are assuming that the primary beam is fixed with respect to local coordinates and translates azimuthally on the celestial sphere. We additionally assume that the baseline is phased to zenith. In other words, Equation (47) describes an interferometer observing in a drift-scan mode.

To see how integrating in time may be suboptimal, consider a simplified, purely pedagogical thought experiment where our interferometer consists of a single east-west baseline ($b_y = 0$) situated at the equator ($\eta = 0$). For the primary beam, suppose we have a beam that is extremely narrow in the polar direction, so that $A_\nu(\hat{\mathbf{r}}, t) \equiv \delta^D(\delta) A^\alpha_\nu(\alpha - \omega_\oplus t)$, where δ^D signifies a Dirac delta function. Plugging these into restrictions into our equation, we obtain

$$V_{b\nu}(t) = \int A^\alpha_\nu(\alpha - \omega_\oplus t) I_\nu\left(\delta = 0, \alpha\right) \times \exp\left[-i2\pi \frac{b_x}{\lambda} \sin(\alpha - \omega_\oplus t)\right] d\alpha + n(t). \quad (48)$$

For a single baseline, the function $V_{b\nu}(t)$ is precisely the continuous version of the discrete data vector $\mathbf{v}$. To obtain $\mathbf{v}$, then, one would simply sample $V_{b\nu}(t)$ discretely in time. For a multi-baseline array, forming $\mathbf{v}$ involves following the above procedure for each baseline, and then concatenating the resulting vectors to form a single long $\mathbf{v}$ vector. To make our analytic manipulations more convenient, however, we will keep t a continuous variable, so that $\mathbf{v}$ is a hybrid quantity, discrete in baseline but continuous in time. Acting on $\mathbf{v}$ by a matrix then involves summing over baselines and integrating over time.

Identifying $n(t)$ and $I_\nu(\theta = \pi/2, \varphi)$ as the continuous versions of $\mathbf{n}$ and $\mathbf{x}$ respectively, the rest of Equation (48)'s integrand can be interpreted as the continuous version of $\mathbf{A}$. We can model the noise covariance between baselines b and b', at times t and t' as

$$N_{bb'}(t, t') = \sigma^2 \delta_{bb'} \delta^D(t - t'), \quad (49)$$

where σ is an RMS noise level assumed to be uncorrelated in time and uncorrelated between baselines.

To see how the optimal prescription of Equation (45) combines information from different times, we need only evaluate $\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v}$, for the $\mathbf{M}$ matrix has no time index, so its application has no impact on how time-ordered data is combined. In our toy model, we have

$$\left(\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v}\right)_\alpha = \sum_b \int \frac{dt}{\sigma^2} A^\alpha_\nu(\alpha - \omega_\oplus t)\, e^{i2\pi \frac{b_x}{\lambda} \sin(\alpha - \omega_\oplus t)} V_{b\nu}(t), \quad (50)$$

where the α variable serves as the continuous version of a discrete vector index. This expression shows that the optimal, minimum variance prescription does not call for the integration of visibilities in time. Instead, our expression calls for the *convolution* of the visibility data with a kernel that is specified by the primary beam shape and the baseline.

Now, recall from our application of the convolution theorem in previous sections that for interferometric data, convolution in time is equivalent to multiplication in fringe-rate space. Equation (50) therefore suggests that the optimal way to combine different time samples is to express visibilities in fringe-rate space, and then to weight different fringe-rates appropriately before summing. We will develop this method for mapmaking more generally in Section 5.3.

5.2. *The special case where integrating in time is optimal*

Before proceeding, it is instructive to establish the special case where time integration is the optimal technique for an initial data reduction step in mapmaking, since it is frequently employed in the literature. An inspection of Equation (50) shows that were it not for the time-dependence in the primary beam and the time-dependence of the sky moving through a baseline's fringes, the optimal recipe would indeed reduce to an integration of visibilities in time. Finding the limit where time integration is optimal is then equivalent to finding a special case where the aforementioned time-dependences vanish.

Recall that in our previous example, the primary beam had a time-dependence only because our thought-experiment consisted of a drift-scan telescope, whose measurement equation was written in coordinates fixed to the celestial sphere. Instead of this, suppose one had a narrow primary beam that tracked a small patch of the sky. The primary beam would then have a fixed shape in celestial coordinates, and $A_\nu(\hat{\mathbf{r}}, t)$ would simply become $A_\nu(\hat{\mathbf{r}})$ in Equation (47). To attempt to nullify the time-dependence of fringes sweeping across the celestial sphere, one may phase the visibilities in a time-dependent way, essentially tracking the center of the patch as it moves across the sky. Putting this all together and assuming that the primary beam is narrow enough to justify a flat-sky approximation, the measurement equation becomes

$$V_{b\nu}(t) = \int A_\nu(\hat{\mathbf{r}}) I_\nu(\hat{\mathbf{r}}) \exp\left[-i2\pi \left(\frac{b_0}{\lambda} \sin(\alpha - \omega_\oplus t)\right)\right] \times \exp\left[-i2\pi \left(\frac{b_y}{\lambda} \cos\eta \sin\delta\right) + i\psi(t)\right] d\Omega + n(t), \quad (51)$$

where we have assumed for simplicity that the center of our small field is directly above the equator, and that a time-dependent phase $\psi(t)$ has been applied. With this, the optimal combination of time-ordered data becomes

$$\left(\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v}\right)^{\text{flat}}_{(\delta,\alpha)} = \frac{A_\nu(\delta, \alpha)}{\sigma^2} e^{-i2\pi \frac{b_y}{\lambda} \cos\eta \sin\delta} \times \sum_b \int dt\; e^{-i2\pi \frac{b_x}{\lambda} \sin(\alpha - \omega_\oplus t) + i\psi(t)} V_{b\nu}(t). \quad (52)$$

This is still not quite a simple average in time because there is no choice of $\psi(t)$ that can cancel out the time-dependence of $\sin(\alpha - \omega_\oplus t)$ for all φ and all t. Another way to phrase the problem is to note that even in the flat-sky approximation, one cannot expand Taylor expand $\sin(\alpha - \omega_\oplus t)$ over long observation times. With short observations, however, an expansion is justified. Performing this expansion, invoking the narrow field-of-view approximation in the α direction, and picking

$\psi(t) = -2\pi\frac{b_x}{\lambda}\omega_\oplus t$ gives

$$\left(\mathbf{A}^\dagger\mathbf{N}^{-1}\mathbf{v}\right)^{\rm flat,short}_{(\delta,\alpha)} = \frac{A_\nu(\delta,\alpha)}{\sigma^2}e^{-i2\pi\left(\frac{b_y}{\lambda}\cos\eta\cos\delta+\frac{b_x}{\lambda}\sin\alpha\right)} \times\sum_b\int dt\, V_{b\nu}(t), \quad (53)$$

which is a simple averaging in time. In short, then, integrating in time is an optimal way to combine time-ordered data if a number of criteria are met: the flat-sky approximation must hold (which can be achieved, for example, by having a small primary beam), the primary beam must track the field, the visibilities must be phased to track the center of the field, and the observations must be short.

5.3. *Optimized fringe-rate filtering for imaging*

We now proceed to derive the optimal prescription for combining time-ordered data, which will naturally give rise to fringe-rate filtering. Because we will *not* be invoking the same approximations as we did in previous sections, we will begin with Equation (47). From our toy example (Equation 50), we know that fringe-rate space (the Fourier dual of time) is a promising space in which to combine time-ordered data. Here, we make the assumption that we are dealing with a full sidereal day's worth of data, so that the visibility in fringe-rate space is given by

$$\overline{V}_{b\nu}(f) \equiv \frac{1}{T_\oplus}\int_{-T_\oplus/2}^{T_\oplus/2} dt\exp\left(-2\pi i f t\right)V_{b\nu}(t), \quad (54)$$

where f is the fringe-rate, and $T_\oplus = 2\pi/\omega_\oplus$ is the Earth's rotation period. It is natural to work in fringe-rate bins such that the nth bin is given by $f_n \equiv n/T_\oplus$, where n is an integer. The measurement in the nth bin is then given by

$$\overline{V}_{b\nu}(f_n) = \int d\Omega\, T(\hat{\mathbf{r}})e^{-i2\pi\frac{b_y}{\lambda}\cos\eta\sin\delta} \times\int_{-\frac{T_\oplus}{2}}^{\frac{T_\oplus}{2}}\frac{dt}{T_\oplus}\,B(\hat{\mathbf{r}},t)e^{-i\frac{2\pi nt}{T_\oplus}+i\frac{2\pi b_0}{\lambda}\cos\delta\sin(\omega_\oplus t-\alpha)}, \quad (55)$$

where we have temporarily omitted the additive noise term to avoid mathematical clutter. To proceed, we make some simplifying assumptions (although only some of which are absolutely required). First, assume that we are once again considering a drift-scan instrument. If the primary beam shape is approximately separable, we can then say

$$B(\hat{\mathbf{r}},t) \equiv B_\delta(\delta)B_\alpha(\alpha-\omega_\oplus t), \quad (56)$$

where B_α is a function with period 2π. Taking advantage of this periodicity, we can write the beam as

$$B(\hat{\mathbf{r}},t) = B_\delta(\delta)\sum_q\overline{B}_q e^{-iq\alpha}e^{iq\omega_\oplus t}, \quad (57)$$

where $\overline{B}_q \equiv \int\frac{d\alpha}{2\pi}B_\alpha(\alpha)e^{iq\alpha}$ is the q^{th} Fourier coefficient. Plugging this into Equation (55) and making the substitution $\psi \equiv \omega_\oplus t - \varphi$, one obtains

$$\overline{V}_{b\nu}\,(f_n) = \int d\Omega\, T(\hat{\mathbf{r}})B_\delta(\delta)e^{-i2\pi\frac{b_y}{\lambda}\cos\eta\sin\delta} \times\sum_q\frac{\overline{B}_q e^{-in\varphi}}{2\pi}\int_{-\pi-\varphi}^{\pi+\varphi}d\psi\, e^{i(q-n)\psi+i\frac{2\pi b_0}{\lambda}\cos\delta\sin\psi}. \quad (58)$$

Now, the integral over ψ is of a periodic function over one period. We may therefore freely shift the limits of the integral by a constant amount without affecting the result. In particular, we may remove the $+\varphi$ terms in the limits (the only restriction being that having performed a φ-dependent shift, it is no longer legal to permute the various integrals), and the result is a standard integral form for a Bessel function J of the first kind:

$$\overline{V}_{b\nu}(f_n) = \int\frac{d\Omega}{2\pi}\,T(\hat{\mathbf{r}})B_\delta(\delta)e^{-i2\pi\frac{b_y}{\lambda}\cos\eta\sin\delta}e^{-in\alpha} \times\sum_q\overline{B}_q J_{n-q}\left(\frac{2\pi b_0}{\lambda}\cos\delta\right). \quad (59)$$

Several features are of note here. For wide primary beams, $\overline{B}_q$ is sharply peaked around $q = 0$, so the terms following the sum over q essentially amount to $J_n(2\pi b_0\cos\delta/\lambda)$. Now, notice that the argument of the Bessel function is bounded, always lying between $\pm 2\pi b_0/\lambda$. For large n (high fringe-rate bins), then, one can use the small argument asymptotic form for J_n,

$$J_n\left(\frac{2\pi b_0\cos\delta}{\lambda}\right) \approx \frac{1}{n!}\left(\frac{\pi b_0\cos\delta}{\lambda}\right)^n, \quad (60)$$

which is a sharply decreasing function of n for large n. This means that there must be very little sky signal at high fringe-rate bins, which is yet another reflection of the salient feature that we have emphasized throughout this paper: high fringe-rate bins constitute noise-dominated modes, since the Earth's rotation rate works in conjunction with the baseline length to impose a maximum fringe-rate for sources locked to the celestial sphere.

Putting together the optimal prescription as we did above, we have

$$\left(\mathbf{A}^\dagger\mathbf{N}^{-1}\mathbf{v}\right)_{\delta,\alpha} = \frac{B^*_\delta(\delta)\cos\delta}{2\pi\sigma^2}\sum_{b,n}e^{in\alpha}e^{i2\pi\frac{b_y}{\lambda}\cos\eta\sin\delta} \times\sum_q\widetilde{B}^*_q J_{n-q}\left(\frac{2\pi b_0}{\lambda}\cos\delta\right)\overline{V}_{b\nu}(f_n). \quad (61)$$

In words, this recipe instructs us to move into fringe-rate space (where the sky emission is already concentrated in f_n) and to further downweight by $\sum_q\widetilde{B}^*_q J_{n-q}\left(\frac{2\pi b_0}{\lambda}\cos\delta\right)$, which, as we have argued above, is small for high fringe rates. Thus, filtering away the high fringe-rates is the optimal way to combine time-ordered data from an interferometer.

In summary, we see that downweighting high fringe-rates once again appears as an integral part of an optimal recipe, this time for mapmaking. While we made some assumptions (such as a separable beam) for analytic convenience, our qualitative conclusions are general. To generalize our treatment, one can simply return to Equation (55), numerically evaluating the integral over time. One

can then read off an expression for **A** from the remaining integral, and proceed as before.

6. CONCLUSION

In this paper, we have revisited the concept of filtering the visibility time-series measured by an interferometric baseline that was presented in Parsons & Backer (2009). Using a mapping between the timescale of variation in visibility data and position on the sky for a chosen baseline, we have seen that the rectangular time windows typically used when integrating visibilities are almost always sub-optimal (particularly for wide-field, drift-scan instruments), and motivate filtering on the basis of fringe rate as a step for optimally combining time-ordered visibility data. In Section 5, we found that fringe-rate filtering also naturally arises as part of optimized mapmaking prescriptions such as those described in Tegmark (1997), Morales & Matejek (2009b), or Dillon et al. (2015).

We also showed that fringe-rate filtering can alternately be interpreted as a per-baseline operation for sculpting the primary beam along one dimension. Using analytic derivations and simulations, we highlight several important applications of such beam sculpting. One key application for 21cm cosmological experiments starved for sensitivity is the ability to re-weight visibility data according to the signal-to-noise ratio in each fringe-rate bin. As shown in Section 4.1, pre-processing the visibilities in such a way allows an optimal (minimum variance) estimate of the power spectrum to be formed by squaring visibilities time-slice-by-time-slice before averaging the results together. This somewhat surprising result allows a data analyst to deal directly with visibilities until the final squaring step, allowing for cleaner diagnoses of instrumental systematics while avoiding potential analysis-induced systematics associated with gridding data on the uv plane. Other important applications include improving the match between polarization beam to reduce polarization leakage, and down-weighting areas low in the primary beam to reduce systematics from off-axis foregrounds.

In Ali et al. (2015), the fringe-rate filtering techniques presented here are applied to observations from the PAPER array as part of their power-spectrum analysis pipeline. The results highlight the power of fringe-rate filtering in 21cm cosmology applications. Given its efficiency, flexibility, and close alignment with the natural observing basis of radio interferometers, we anticipate that fringe-rate filtering may be an important analysis tool for current 21cm experiments, as well as future instruments such as the Hydrogen Epoch of Reionization Array (HERA; Pober et al. 2014) and the Square Kilometre Array (SKA; Carilli 2014).

7. ACKNOWLEDGMENTS

It gives us great pleasure to thank James Aguirre, Josh Dillon, Gianni Bernardi, Daniel Jacobs, Saul Kohn, David Moore, Miguel Morales, and Jonathan Pober for helpful discussions. This research was supported in part by the NSF CAREER award No. 1352519 and the NSF AST grant No. 1129258.

REFERENCES

Ali, Z. S., et al. 2015, ArXiv e-prints
Asad, K. M. B., et al. 2015, ArXiv e-prints
Bandura, K., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9145, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 22
Beardsley, A. P., et al. 2013, MNRAS, 429, L5
Bernardi, G., Carretti, E., Cortiglioni, S., Sault, R. J., Kesteven, M. J., & Poppi, S. 2003, ApJ, 594, L5
Bernardi, G., et al. 2013, ApJ, 771, 105
Bhatnagar, S., Cornwell, T. J., Golap, K., & Uson, J. M. 2008, A&A, 487, 419
Bhatnagar, S., Rau, U., & Golap, K. 2013, ApJ, 770, 91
Bowman, J. D., et al. 2013, PASA, 30, 31
Carilli, C. L. 2014, ArXiv e-prints
Cornwell, T. J., Golap, K., & Bhatnagar, S. 2003, W-Projection: A New Algorithm for Non-Coplanar Baselines, EVLA Memo 67
de Oliveira-Costa, A., Tegmark, M., Gaensler, B. M., Jonas, J., Landecker, T. L., & Reich, P. 2008, MNRAS, 388, 247
Dillon, J. S., et al. 2015, Phys. Rev. D, 91, 023002
Gaensler, B. M., Dickey, J. M., McClure-Griffiths, N. M., Green, A. J., Wieringa, M. H., & Haynes, R. F. 2001, ApJ, 549, 959
Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., & Bartelmann, M. 2005, ApJ, 622, 759
Greenhill, L. J., & Bernardi, G. 2012, ArXiv e-prints
Harker, G., et al. 2010, MNRAS, 405, 2492
Jacobs, D. C., et al. 2014, ArXiv e-prints
Jelić, V., Zaroubi, S., Labropoulos, P., Bernardi, G., de Bruyn, A. G., & Koopmans, L. V. E. 2010, MNRAS, 409, 1647
Jelić, V., et al. 2008, MNRAS, 389, 1319
—. 2014, A&A, 568, A101
Liu, A., Parsons, A. R., & Trott, C. M. 2014a, Phys. Rev. D, 90, 023018
—. 2014b, Phys. Rev. D, 90, 023019
Moore, D., et al. 2015, ArXiv e-prints
Moore, D. F., Aguirre, J. E., Parsons, A. R., Jacobs, D. C., & Pober, J. C. 2013, ApJ, 769, 154
Morales, M. F., & Matejek, M. 2009a, MNRAS, 400, 1814
—. 2009b, MNRAS, 400, 1814
Offringa, A. R., de Bruyn, A. G., & Zaroubi, S. 2012, MNRAS, 422, 563
Parsons, A., Pober, J., McQuinn, M., Jacobs, D., & Aguirre, J. 2012a, ApJ, 753, 81
Parsons, A., et al. 2008, PASP, 120, 1207
Parsons, A. R., & Backer, D. C. 2009, AJ, 138, 219
Parsons, A. R., Pober, J. C., Aguirre, J. E., Carilli, C. L., Jacobs, D. C., & Moore, D. F. 2012b, ApJ, 756, 165
Parsons, A. R., et al. 2010, AJ, 139, 1468
—. 2014, ApJ, 788, 106
Pober, J. C., et al. 2012, AJ, 143, 53
—. 2013, ApJ, 768, L36
—. 2014, ApJ, 782, 66
Roshi, D. A., & Perley, R. A. 2003, ArXiv Astrophysics e-prints
Ryle, M. 1952, Royal Society of London Proceedings Series A, 211, 351
Shaw, J. R., Sigurdson, K., Pen, U.-L., Stebbins, A., & Sitwell, M. 2013a, ArXiv e-prints
—. 2013b, ArXiv e-prints
Sullivan, I. S., et al. 2012, ApJ, 759, 17
Tasse, C., van der Tol, S., van Zwieten, J., van Diepen, G., & Bhatnagar, S. 2013, A&A, 553, A105
Tegmark, M. 1997, ApJ, 480, L87
Tingay, S. J., et al. 2013, PASA, 30, 7
van Haarlem, M. P., et al. 2013, A&A, 556, A2
Wieringa, M. H., de Bruyn, A. G., Jansen, D., Brouw, W. N., & Katgert, P. 1993, A&A, 268, 215
Zheng, H., et al. 2014, MNRAS, 445, 1084