FRINGE-RATE FILTERING

Aaron R. Parsons[1,2], Adrian Liu[1], Zaki S. Ali[1],
Draft version February 24, 2015

ABSTRACT

1. INTRODUCTION

In recent years, low-frequency radio interferometers have gone through dramatic changes in design. These transformations have been driven by new science applications such as 21 cm cosmology, where one uses the highly redshifted 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) on top of a dim, diffuse background that itself is obscured by bright foreground sources that are 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. Additionally, these elements are often placed in regular redundant grids to maximize sensitivity. Examples of this include the Donald C. Backer Precision Array for Probing the Epoch of Reionization, XXX: decide exactly which experiments to include.

With new telescope 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 ~ 1000 hrs of integration on instruments optimized for such a measurement (Parsons et al. 2012a), and even then, only for the largest scale modes. 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) to optimize the process of combining time-ordered data. The key realization is that fringe-rates—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 low-pass filtering the data (i.e., weighting the data) in the fringe-rate domain. The fringe-rate domain provides a natural basis for time-averaging interferometeric 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 allows 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 primary beam, 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 cleaner views 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 the 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 some 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[3]

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

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

---

[1] Astronomy Dept., U. California, Berkeley, CA
[2] Radio Astronomy Lab., U. California, Berkeley, CA

[3] In this section, we omit the instrumental noise contribution to the measured visibilities in order to avoid notational clutter.
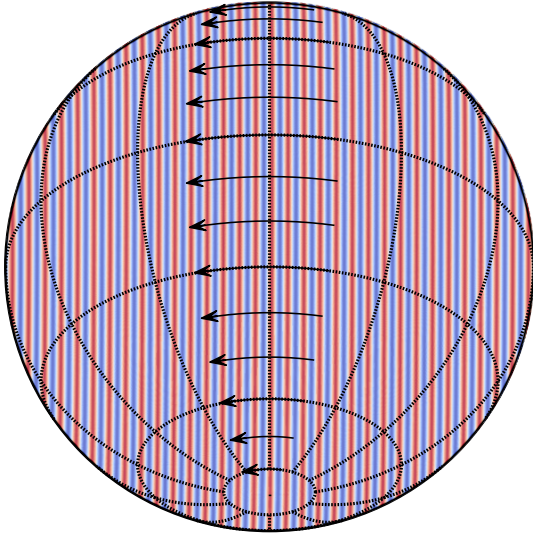
FIG. 1.— The fringe pattern at 150 MHz of a 30-m east-west baseline, overlaid with arrows indicating the distance traversed by sources at various declinations over a two-hour span centered at transit. In a fixed time interval, sources near $\delta = 0°$ traverse more fringe periods than sources nearer to the celestial poles, giving rise to different fringe rates that can be used to distinguish these sources in a timeseries measured with a single baseline.

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. Here, we have adopted 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. Though 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. The movement of a source through the fringe pattern causes $V_\nu(t)$ to oscillate 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 $\delta = 0°$

to zero at $\delta = -90°$, and can even become negative for emission from the far side of the celestial pole.

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

$$\overline{V}_{b\nu}(f,t) = \int d\Omega I_\nu(\hat{\mathbf{r}}) \int dt' \gamma(t'-t) A_\nu(\hat{\mathbf{r}},t') e^{i2\pi\left[\frac{\nu}{c}\mathbf{b}_{t'}\cdot\hat{\mathbf{r}} - f(t'-t)\right]},$$
(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, thereby centering our transform at time $t' = t$. If the characteristic width of γ is relatively short, the time-dependence of the visibility will likely be dominated by features on the sky moving relative to fringes, 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

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

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

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

where $\tilde{\gamma}$ is the 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 $f \approx \hat{\mathbf{r}}\cdot\boldsymbol{\omega}_\oplus \times \mathbf{b}\nu/c$.

In words, what the above derivation shows is that as advertised, different fringe-rates correspond to different parts of the sky. This is illustrated in Figure 2, which shows contours of constant fringe-rate for the same baseline as the one simulated in Figure 1. 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 $\mathbf{b}$, $\boldsymbol{\omega}_\oplus$, 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 **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

$$V_{b\nu}^{\text{filt}}(t', t) = \int df w(f) \overline{V}_{b\nu}(f, t) e^{i2\pi f(t'-t)}$$

$$= \int d\Omega I_\nu(\hat{\mathbf{r}}) A_\nu(\hat{\mathbf{r}}, t) e^{i2\pi \frac{\nu}{c} \mathbf{b}_t \cdot \hat{\mathbf{r}}}$$

$$\times \int df e^{i2\pi f(t'-t)} w(f) \tilde{\gamma} \left[\frac{\nu}{c} (\boldsymbol{\omega}_\oplus \times \mathbf{b}_t) \cdot \hat{\mathbf{r}} - f \right]. \quad (5)$$

Now, suppose we implement this filter in a sliding manner 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

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

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

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

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 $\hat{\mathbf{r}} \cdot \boldsymbol{\omega}_\oplus \times \mathbf{b}$. For any given instant, $\boldsymbol{\omega}_\oplus \times \mathbf{b}$ 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 $\hat{\mathbf{r}} \cdot \boldsymbol{\omega}_\oplus \times \mathbf{b}$, 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 essentially averages 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 $f \approx \hat{\mathbf{r}} \cdot \boldsymbol{\omega}_\oplus \times \mathbf{b}\nu/c$ in the data. Since $\hat{\mathbf{r}} \cdot \boldsymbol{\omega}_\oplus \times \mathbf{b}$ can never exceed $\omega_\oplus b$, the maximal fringe-rate that can be achieved by a source locked to the celestial sphere is $f_{\max} = \omega_\oplus b\nu/c$, 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
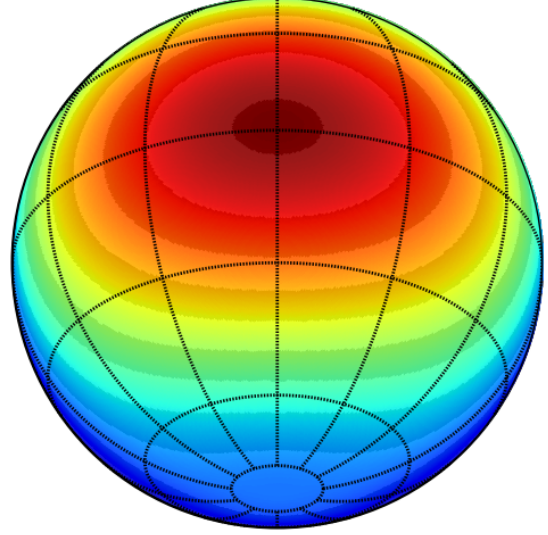

FIG. 2.— Fringe rate as a function of sky position, corresponding to the fringe pattern illustrated in Figure 1. Fringe rates peak at 1.09 mHz at $\delta = 0°$, hit zero at the south celestial pole, and become negative on the far side of the pole. Grey shading indicates the approximate angular regions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

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 in 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 particular natural way to approach mapmaking in interferometric observations.

3. IMPLEMENTATION

In this section, we discuss a practical implementation of fringe-rate filtering, which will be used in simulations to illustrate various applications of fringe-rate filtering later in the paper. In keeping with its origins targeting observations from the PAPER array, we simulate a model array based on PAPER, deployed at a latitude $-30°$ and featuring the beam response pattern characteristic of PAPER dipole elements (Parsons et al. 2008; Pober et al. 2012). For these simulations, we also choose a specific baseline to examine: a pair of antennas separated by 30 m in the east-west direction, observing at 150 MHz. This 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).

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 primary beam profile over contours of constant fringe-rate on the sky (such as those shown in Figure 2). A realistic primary beam will frequently require empirical modeling beyond analytic forms. It is therefore in general difficult to derive a completely analytic expression for an optimized fringe-rate filter profile. However, in the interests 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 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. Applying the fringe-rate filter as an FIR filter in the time domain, as opposed 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 and helps limit the number of samples potentially corrupted by spurious systematics such as radio frequency interference.

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 selec-

tion 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 $A_i(\hat{\mathbf{r}})$. 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[4], $\widetilde{A}_i$ is the Fourier transform of $A_i(\hat{\mathbf{r}})$ 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 it is implicitly assumed that the baseline length is given by $\mathbf{b} = \mathbf{u}\lambda$, with $\mathbf{u} \equiv (u, v)$. From this, we can see that two baselines with different primary beams, but located at the same location

---

[4] 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 a good approximation only 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).

on the uv plane have a delay-spectrum cross-correlation given by

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

where angular brackets $\langle \ldots \rangle$ 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 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.[5] Following this, we used Parseval's theorem to rewrite the integral over $(\mathbf{u}, \eta)$ space as an integral over angles 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

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

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

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

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 $X^2 Y$ in the process,[6] with

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

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

$$Y \equiv \frac{c(1+z)^2}{\nu_{21} H_0 E(z)}, \quad (14)$$

where $\nu_{21} \equiv 1420 \, \text{MHz}$ 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

[5] Although see Liu et al. (2014b) for some limitations of this approximation.
[6] See, e.g., (Liu et al. 2014a) for a detailed derivation.

for brevity, and Equation (12) reduces to

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

where

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

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

$$\widehat{P} = \sum_{ij} w_{ij} \widetilde{V}_i \widetilde{V}_j^*, \quad (17)$$

where w_{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

$$\widehat{P}_{\text{noise}} = \sum_{ij} w_{ij} n_i n_j^*. \quad (18)$$

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

$$\text{Var}(\widehat{P}_{\text{noise}}) \equiv \langle \widehat{P}_{\text{noise}}^2 \rangle - \langle \widehat{P}_{\text{noise}} \rangle^2$$
$$= \sum_{ijkm} w_{ij} w_{km} \left[\langle n_i n_j^* n_k n_m^* \rangle - \langle n_i n_j^* \rangle \langle n_k n_m^* \rangle \right]$$
$$= \sum_{ij} w_{ij} w_{ji} \sigma^4, \quad (19)$$

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 $\langle a_i a_j \rangle = \langle b_i b_j \rangle = \delta_{ij}\sigma^2/2$ and $\langle a_i b_j \rangle = 0$ 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

$$\langle \widehat{P} \rangle = \sum_{ij} w_{ij}\langle \widetilde{V}_i \widetilde{V}_j^* \rangle = S \sum_{ij} w_{ij}\Omega_{ij}P, \qquad (20)$$

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 $S\sum_{ij}w_{ij}\Omega_{ij} = 1$, 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, minimizing

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

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 $w_{km} \propto \Omega_{km}$, and normalizing according to our constraint yields

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

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} vanishes unless $i = j$. Defining

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

we have $\Omega_{ij} \equiv \delta_{ij}\gamma_i$, so our optimal estimator for the power spectrum reduces to

$$\widehat{P} = \frac{1}{S \sum_j \gamma_j^2} \sum_i \gamma_i |\widetilde{V}_i|^2. \qquad (24)$$

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

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

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

$$\widetilde{V}^{\text{opt}}(\mathbf{u}, \eta; t_i) \equiv \sum_i e^{i2\pi ft}\gamma_i^{1/2}\widetilde{V}_i$$

$$= \sum_i e^{i2\pi ft}\left[\frac{1}{B}\int d\Omega d\nu A_i(\hat{\mathbf{r}}, \nu)^2\right]^{\frac{1}{2}} \widetilde{V}_i. \quad (26)$$

This is a rather remarkable 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 $\gamma_i^{1/2}$, i.e., where each fringe-rate is weighted by the root-mean-square 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* in general the same as the ones derived in Section 5.3 for optimal mapmaking, where measurements were essentially weighted by an additional factor of the primary beam in fringe-rate space, rather than by the root-mean-square 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 w_{ij} to a form separable in i and j, which is a restriction that 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. That is, rotation synthesis becomes an important contribution to an interferometer's sensitivity.

In closing, 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. To ensure that one's power spectrum estimates are correctly normalized, one can simply follow the prescription in Equation (25). However, the derivation of Equation (25) involved a number of approximations. Additionally, as we discussed in Section 3, practical implementations of fringe-rate filtering may deviate from the optimal prescription outlined here. It may therefore be safer to instead compute the normalization of the power spectrum numerically. Once the fringe-rate weights have been specified, Equation (7) gives the effective beam. Since we know from Equation (6) that the measurement equation for fringe-rate filtered visibilities is the same as that for the original visibilities up to the revised beam, one can simply use Equation (15) as long as the beam area is accordingly revised.

In Figure XXX, we show... [XXX uncomment out stuff below to talk about the simulations]

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 other sources of noise or systematics that may limit our constraints. One example of this is the cross-contamination between Stokes terms in interferometric polarization measurements. Minimizing such contamination is of paramount importance for 21cm 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; 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 orthogonal polarizations. However, orthogonal linearly polarized feeds in practice have primary beam responses that do not match. As described in Moore et al. (2013), if left uncorrected, the unmatched beam response between visibilities V_{XX} and V_{YY} measuring the XX and YY polarization products, respectively, is the dominant source of polarization leakage in the Stokes I measurement $V_I \equiv (V_{XX} + V_{YY})/2$ for linearly polarized feeds.

With an accurate beam model, it is trivial to rescale V_{XX} and V_{YY} so that the XX and YY beam responses match in a chosen (typically, zenith) direction. Their sum, V_I, then represents a perfect probe of the Stoke I parameter in that chosen direction, but will contain contamination from $V_Q \equiv (V_{XX} - V_{YY})/2$ in directions where the XX and YY beam responses do not match. 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_{XX} relative to V_{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 reweighting of the sky along two dimensions.

The typical 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. **???**). 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, and 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 polariation beams. However, 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 can be no better than what can be achieved by using a single number to rescale V_{XX} and V_{YY} before summing.

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 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, although some array 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 21cm 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
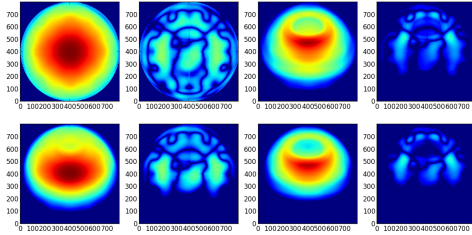
Fig. 3.—

plane as its project toward the phase center changes. 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, the result will be identical to the best that 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.

[XXX discuss results here]

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).

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 Walsh switching (XXX cite), 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; **?**; **?**, 2014).

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 a 30-m baseline observing at 150 MHz, 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.

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 undesireable 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.

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 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_{pix}, and our instrument's response as a matrix $\mathbf{A}$ of size $N_b N_t \times N_{\mathrm{pix}}$, the measurement equation is given

by

$$\mathbf{v} = \mathbf{A}\mathbf{x} + \mathbf{n}, \qquad (27)$$

where $\mathbf{n}$ is a noise vector. Note that in this general form, Equation (27) 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_{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 the most 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)

$$\hat{\mathbf{x}} = \mathbf{M}\mathbf{A}^{\dagger}\mathbf{N}^{-1}\mathbf{v}, \qquad (28)$$

where $\mathbf{M}$ is some invertible matrix chosen by the data analyst, 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.

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

$$\mathbf{\Sigma} \equiv \langle (\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^{\dagger} \rangle = \mathbf{M}\mathbf{A}^{\dagger}\mathbf{N}^{-1}\mathbf{A}\mathbf{M}^{\dagger}. \qquad (29)$$

With a suitable choice of $\mathbf{M}$, the estimator given by Equation (28) minimizes the variance. Regardless of one's choice, however, Equation (28) 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 (28) 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. (2013) 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 co-

ordinate system, we have

$$V_{b\nu}(t) = \int A_{\nu}(\hat{\mathbf{r}}, t)I_{\nu}(\hat{\mathbf{r}}) \exp\left[-i2\pi\left(\frac{b_y}{\lambda}\cos\eta\sin\delta\right)\right]$$
$$\times \exp\left[-i2\pi\left(\frac{b_0}{\lambda}\cos\delta\sin(\alpha - \omega_{\oplus}t)\right)\right]d\Omega + n(t), \quad (30)$$

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 (30) 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 $A_{\nu}(\hat{\mathbf{r}}, t) \equiv \delta^D(\delta)A_{\nu}^{\alpha}(\alpha - \omega_{\oplus}t)$, where δ^D signifies a Dirac delta function. Plugging these into restrictions into our equation, we obtain

$$V_{b\nu}(t) = \int A_{\nu}^{\alpha}(\alpha - \omega_{\oplus}t)I_{\nu}\left(\delta = 0, \alpha\right)$$
$$\times \exp\left[-i2\pi\frac{b_x}{\lambda}\sin(\alpha - \omega_{\oplus}t)\right]d\alpha + n(t). \quad (31)$$

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 (31)'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

$$N_{bb'}(t, t') = \sigma^2 \delta_{bb'}\delta(t - t'), \qquad (32)$$

where σ is an root-mean-square noise level assumed to be uncorrelated in time and uncorrelated between baselines.

To see how the optimal prescription of Equation (28) combines information from different times, we need only evaluate $\mathbf{A}^{\dagger}\mathbf{N}^{-1}\mathbf{v}$, for the $\mathbf{M}$ has no time index, so its application has no impact on how time-ordered data is combined. In our toy model, we have

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

$$(33)$$

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 the convolution theorem that convolution in time is equivalent to multiplication in the Fourier dual space of time. For an interferometer baseline, this Fourier dual space is fringe-rate. Equation (33) 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 application of fringe-rate filtering in greater generality 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, since it is frequently employed in the literature. An inspection of Equation (33) 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 (30). 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

$$V_{b\nu}(t) = \int A_\nu(\hat{\mathbf{r}}) I_\nu(\hat{\mathbf{r}}) \exp\left[-i2\pi \left(\frac{b_0}{\lambda} \sin(\alpha - \omega_\oplus t) \right) \right]$$
$$\times \exp\left[-i2\pi \left(\frac{b_y}{\lambda} \cos\eta \cos\delta \right) + i\psi(t) \right] d\Omega + n(t), \quad (34)$$

[XXX: check this equation] 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

$$\left(\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v}\right)^{\text{flat}}_{(\delta,\alpha)} = \frac{A_\nu(\delta, \alpha)}{\sigma^2} e^{i2\pi \frac{b_y}{\lambda} \cos\eta \cos\delta}$$
$$\times \sum_b \int dt\, e^{i2\pi \frac{b_x}{\lambda} \sin(\alpha - \omega_\oplus t) - i\psi(t)} V_{b\nu}(t). \quad (35)$$

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

$$\left(\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v}\right)^{\text{flat,short}}_{(\delta,\alpha)} = \frac{A_\nu(\delta, \alpha)}{\sigma^2} e^{i2\pi \left(\frac{b_y}{\lambda} \cos\eta \cos\delta + \frac{b_x}{\lambda} \sin\alpha \right)}$$
$$\times \sum_b \int dt\, V_{b\nu}(t), \quad (36)$$

which is a simple averaging in time. In short, then, integrating in time is an optimal way to combine time-ordered data only if a number of criteria are met: the flat-sky approximation must hold, 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. *Fringe-rate filtering*

We now proceed to derive the optimal prescription for combining time-ordered data, which will lead us to the technique of fringe-rate filtering. Because our derivation will *not* require any of the approximations that we have invoked so far for pedagogical reasons, we will begin with our general expression for time-ordered visibilities, Equation (30). From our toy example (Equation 33), we know that fringe-rate space (the Fourier dual of time) is a promising space in which to combine time-ordered data. Formally, measurements in this space are given by

$$\overline{V}_b(f) \equiv \frac{1}{T_\oplus} \int_{-T_\oplus/2}^{T_\oplus/2} dt \exp\left(-2\pi i f t\right) V_b(t), \quad (37)$$

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

$$\overline{V}_b(f_n) = \int d\Omega\, T(\hat{\mathbf{r}}) e^{-i2\pi \frac{b_y}{\lambda} \cos\eta \sin\delta}$$
$$\times \int_{-\frac{T_\oplus}{2}}^{\frac{T_\oplus}{2}} \frac{dt}{T_\oplus} B(\hat{\mathbf{r}}, t) e^{-i \frac{2\pi n t}{T_\oplus} + i \frac{2\pi b_0}{\lambda} \cos\delta \sin(\omega_\oplus t - \alpha)}, \quad (38)$$

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

$$B(\hat{\mathbf{r}}, t) \equiv B_\delta(\delta) B_\alpha(\alpha - \omega_\oplus t), \quad (39)$$

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

$$B(\hat{\mathbf{r}}, t) = B_\delta(\delta) \sum_q \widetilde{B}_q e^{-iq\alpha} e^{iq\omega_\oplus t}, \quad (40)$$

where $\widetilde{B}_q \equiv \int \frac{d\alpha}{2\pi} B_\alpha(\alpha) e^{iq\alpha}$ is the qth Fourier coefficient. Plugging this into Equation (38) and making the substi-

tution $\psi \equiv \omega_{\oplus} t - \varphi$, one obtains

$$\overline{V}_b(f_n) = \int d\Omega \, T(\hat{\mathbf{r}}) B_\delta(\delta) e^{-i2\pi \frac{b_y}{\lambda} \cos \eta \sin \delta}$$

$$\times \sum_q \frac{\widetilde{B}_q e^{-in\varphi}}{2\pi} \int_{-\pi-\varphi}^{\pi+\varphi} d\psi \, e^{i(q-n)\psi + i\frac{2\pi b_0}{\lambda} \cos \delta \sin \psi}. \quad (41)$$

Now, note that 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:

$$\overline{V}_b(f_n) = \int \frac{d\Omega}{2\pi} T(\hat{\mathbf{r}}) B_\delta(\delta) e^{-i2\pi \frac{b_y}{\lambda} \cos \eta \sin \delta} e^{-in\alpha}$$

$$\times \sum_q \overline{B}_q J_{n-q} \left(\frac{2\pi b_0}{\lambda} \cos \delta \right). \quad (42)$$

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,

$$J_n \left(\frac{2\pi b_0 \cos \delta}{\lambda} \right) \approx \frac{1}{n!} \left(\frac{\pi b_0 \cos \delta}{\lambda} \right)^n, \quad (43)$$

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. Intuitively, this is because sources on the celestial sphere have their fringe rates limited by the Earth's rotation period and projected baseline length b_0, making high fringe rates physically unattainable by true celestial emission. Any signals seen in high fringe-rate bins will therefore be primarily due to noise.

With celestial emission appearing only in low fringe-rate bins, it is reasonable to expect that one can reduce noise (in other words, achieving the goals of time integration) by Fourier transforming the data into fringe-rate space and downweighting (or discarding) high fringe-rate modes. This is confirmed by constructing the optimal prescription as we did above, which yields

$$\left(\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v} \right)_{\delta,\alpha} = \frac{B_\delta^*(\delta) \cos \delta}{2\pi \sigma^2} \sum_{b,n} e^{in\alpha} e^{i2\pi \frac{b_y}{\lambda} \cos \eta \sin \delta}$$

$$\times \sum_q \widetilde{B}_q^* J_{n-q} \left(\frac{2\pi b_0}{\lambda} \cos \delta \right) \overline{V}_b(f_n) (44)$$

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, low-pass fringe-rate filtering is the optimal way to combine time-ordered data from an interferometer.

Before proceeding, let us summarize the essential features of fringe-rate filtering, away from the approximation of a separable beam in Equation (39). Returning to Equations (37) and (38), we see that because the sky $T(\hat{\mathbf{r}})$ is not a function of time, the Fourier transform into fringe-rate space acts only on the beam and the fringe pattern. We can therefore write

$$\overline{V}_b(f_n) = \int d\Omega \, T(\hat{\mathbf{r}}) g_{bn}(\hat{\mathbf{r}}), \quad (45)$$

where the key quantity is

$$g_{bn}(\hat{\mathbf{r}}) \equiv e^{-i2\pi \frac{b_y}{\lambda} \cos \eta \sin \delta}$$

$$\times \int_{-\frac{T_\oplus}{2}}^{\frac{T_\oplus}{2}} \frac{dt}{T_\oplus} B(\hat{\mathbf{r}}, t) e^{-i\frac{2\pi nt}{T_\oplus} + i\frac{2\pi b_0}{\lambda} \cos \delta \sin(\omega_\oplus t - \alpha)}. \quad (46)$$

The term in front of the integral corresponds to the fringes in the declination direction. These are static as the Earth rotates, and therefore are unaffected by the Fourier transform. The second half of the exponent within the integral corresponds to fringes in the azimuthal direction, with the $\cos \delta$ term arising because of the spherical geometry shown in Figure 1, where sources near the equator traverse fringes more quickly than those near the pole. Examining the integral, we see that there are two contributions to a non-zero fringe-rate. First, the primary beam sweeps across the celestial sphere, crossing fringes. In addition, the changing orientation of baselines relative to the celestial sphere induces a fringe-rate. Over a short time-integration (which is likely a more realistic scenario than integrating over $T_\oplus$ as we have assumed so far) we may approximate this second contribution by Taylor expanding the $\sin(\omega_\oplus t - \alpha)$ term. This yields a time-dependence that may be absorbed into the $\exp(-i2\pi nt/T_\oplus)$ term as an additive term to the fringe-rate bin n. What remains is then simply a phase-shifted Fourier transform of the primary beam. In other words, $g_{bn}(\hat{\mathbf{r}})$ selects portions of the primary beam where sources on the sky have a fringe-rate of $n/T_\oplus$. Now, since $\overline{V}_b(f_n) = \int d\Omega \, T(\hat{\mathbf{r}}) g_{bn}(\hat{\mathbf{r}})$, each row of the matrix $\mathbf{A}$ corresponds to $g_{bn}(\hat{\mathbf{r}})$ for a different baseline b and/or fringe rate bin n, and our optimal prescription for combining time-ordered data becomes

$$\left(\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v} \right)_{\delta,\alpha} \propto \sum_{n,b} g_{bn}^*(\hat{\mathbf{r}}) \overline{V}_b(f_n). \quad (47)$$

One sees that the procedure essentially calls for a weighting that emphasizes the fringe rate bins that fall on bright portions of the primary beam. As seen in our toy example, this preferentially favors lower fringe-rates over higher ones.

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 show that the rectangular time windows typically used when integrating visibilities are almost always sub-optimal, and motivate filtering on the basis of fringe

rate as step for optimally combining time-ordered visibility data. In §**??**, we show that fringe-rate filtering indeed can represent a computationally efficient first step for optimal mapmaking, particularly for telescopes with sparsely sampled apertures, or for interferometers with wide fields of view where gridding in the uv plane incurs a significant computational cost.

We also show that fringe-rate filtering can alternately be interpreted as a per-baseline operation for sculpting the primary beam along the declination direction. 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 SNR in each fringe-rate bin. This operation, which is effectively a one-dimensional case of the optimal beam weighting described in Morales & Matejek (2009b) and Bhatnagar et al. (2008), can improve the sensitivity achieved in a per-baseline power spectral analysis by a factor of several while avoiding many of the systematics associated with gridding data in the uv plane. Other important application 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 is likely to be an important analysis tool for current 21cm experiments, as well as future instruments such as the Hydrogen Epoch of Reionization Array (HERA; **?**) and the Square Kilometre Array (SKA; **?**).

REFERENCES

Ali, Z. S., et al. 2015, ArXiv e-prints
Bhatnagar, S., Cornwell, T. J., Golap, K., & Uson, J. M. 2008, A&A, 487, 419
Dillon, J. S., et al. 2015, Phys. Rev. D, 91, 023002
Jacobs, D. C., et al. 2014, ArXiv e-prints
Jelić, V., et al. 2008, MNRAS, 389, 1319
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
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
Shaw, J. R., Sigurdson, K., Pen, U.-L., Stebbins, A., & Sitwell, M. 2013, ArXiv e-prints
Tegmark, M. 1997, ApJ, 480, L87