

FRINGE-RATE FILTERING

AARON R. PARSONS^{1,2}, ADRIAN LIU¹, ZAKI S. ALI¹,
Draft version February 9, 2015

ABSTRACT

Subject headings:

1. INTRODUCTION

[XXX: Rework intro to make connection to 21cm more explicit.]

In any observation of the sky, integrating in time results in increased sensitivity. Such increased sensitivity is particularly important in applications where instrumental noise levels are expected to be high compared to the faint signals that one seeks to measure. As an example of this, in recent years multiple instruments have been built in an attempt to detect the redshifted 21 cm line from the Epoch of Reionization³. 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, while a typical instrument’s system temperature is typically ~ 100 K. Long time-integrations are therefore crucial, and as a practical matter, this is often accomplished by accumulating time-samples, either in the image domain or in Fourier space. In both cases, one is essentially making maps, which can be shown to be in principle a lossless method of data compression, and therefore an attractive method for realizing the increased sensitivity that time-integration provides.

The quest for increased sensitivity, however, has often led to instrumental designs that have made mapmaking algorithms rather unwieldy. Consider first the possibility of accumulating time-samples in the image domain, again for the special case of a 21 cm interferometer array. To increase sensitivity, such arrays typically have wide fields-of-view and are configured to yield a large number of redundant baselines. Over long time-integrations, imaging wide fields-of-view requires careful attention to curved-sky artifacts, which can become computationally expensive to control, particularly if the pixelization is taken to be extremely fine to avoid artifacts from gridding. While a fine pixelization is necessary for mapmaking to be lossless (XXX: cite), this can become wasteful for interferometers with a large number of redundant baselines, in the sense that a large number of pixels are needed to store information about a small number of Fourier modes. [XXX: Make the connection in this last sentence a little clearer.]

Mapmaking in Fourier space (or more precisely, on the uv -plane) alleviates some of these problems. Fourier space is a natural space for the accumulation of interfer-

ometric data, for there each baseline probes a relatively localized region. It is therefore a particularly economical basis for accumulating interferometric data, especially if the final goal is to produce a spatial power spectrum, or other quantities that are “native” to Fourier space. However, gridding artifacts remain, and are even more troublesome when power spectrum estimation is considered in the context of rotation synthesis. As the Earth rotates, baselines sample a series of points along tracks on the uv -plane. Nearby points are almost perfectly correlated, and ought to be integrated coherently prior to the squaring step of any power spectrum measurement, while faraway points can only be combined statistically after squaring. Unless one is willing to keep track of the level of correlation of every uv -point with every other uv -point, an arbitrary choice must be made as to how far away from each other two points can lie before they are considered incoherent. Often, this choice is encapsulated by the uv -pixel size (particularly in calculations of power spectrum sensitivity XXX: cite), with samples falling in the same cells labeled as coherent, and those in different cells labeled as incoherent. However, this can potentially lead to a loss of sensitivity in combining samples that straddle pixel boundaries. While these obstacles can in principle be overcome, it is clear that extreme care must be taken in the mapmaking process to ensure that the full sensitivity of an interferometer array is realized.

In this paper, we introduce a new method for combining time-ordered data—fringe-rate filtering—that avoids the aforementioned pitfalls while affording one the advantages traditionally reserved only when mapmaking. The essential idea is that for an interferometer operating in a drift-scan observing mode, celestial sources of emission have predictable fringe-rates

- For any instrument, important to integrate in time. This is how you get sensitivity.
- Historically (and what is conceptually the easiest) is to make a map and to bin there.
- However, mapmaking is i) difficult, especially with the widefield instruments, ii) subject to artifacts and systematics such as gridding artifacts, iii) “wasteful” for an interferometer array with high redundancy, since the image basis is not efficient if you’re only measuring a small number of modes, iv) unnecessary if you’re using an interferometer to measure a power spectrum.
- For interferometers trying to measure a power spectrum, it’d be nice to be able to stay in the natural “Fourier” space defined by an interferometer.

¹ Astronomy Dept., U. California, Berkeley, CA

² Radio Astronomy Lab., U. California, Berkeley, CA

³ In this paper, we will frequently use 21 cm interferometric arrays as example instruments with which to focus our discussion, and indeed, Sections XXX pertain only to such arrays. However, we emphasize that the central idea of this paper—that fringe-rate filtering can be used to combine time-ordered data in a way that maximizes sensitivity—is one that should be widely applicable to any interferometer.

- On the other hand, there are advantages to map-making that we want: i) coherent integrations in time, ii) the mitigation of systematics (such as polarization leakage), iii) the ability to down-weight data towards the edge of the primary beam (the instrument has already done that once, but an optimal inverse-variance weighted estimator needs to do it once again). Not clear how to do that if one doesn't go to the image-domain.
- In this paper, we introduce a method—fringe-rate filtering—that allows one to stay in visibility space but still get all the advantages of mapmaking. The paper is geared towards drift-scan telescopes, although some of the lessons are also applicable to tracking telescopes. The central idea is to Fourier-transform the time-series in order to sort the observations by fringe-rate, and then to enact a low-pass filter. Since different parts of the sky have different interferometer fringe-rates, a careful tailoring of the filter can downweight different parts of the sky, allowing one to perform the downweightings towards the edge of the primary beam.
- Fringe-rate space is particularly good for identifying high versus low signal-to-noise modes, because certain fringe-modes are physically impossible for true celestial emission (mention the super high azimuthal mode caveat here). We can eliminate those modes. This also makes it clear that time-averaging visibilities is not good. Because it's a sinc in fringe-rate space, which has some high fringe-rate components.
- Fringe-rate filters can be optimized to maximize sensitivity, assuming a noise power spectrum. But they can also be used to help fight systematics. As an example, 21cm arrays suffer from polarization leakage. This can be particularly hard to fight because of the redundancy in the arrays. Beam-sculpting, though, helps reduce this. Specific to 21cm arrays, also show that we don't mess up delay-spectrum.
- Other similar things in literature. Richard's m -mode formalism made more explicit. Optimal map-making. Delay/DDR paper. Also look at Andre's stuff.
- Outline the paper.

Further details are supplied in Appendices A and B of ?.

2. OVERVIEW OF PRINCIPLE OF FRINGE-RATE FILTERING

Generally, the interferometric response, V , for two antennas in a radio interferometer is described by the measurement equation

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

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. 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 spatial structures through the primary beam and the fringes arising from a baseline that is 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 $V_\nu(t)$ 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, as illustrated in Figure 2, 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.

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 to get

$$\tilde{V}_\nu^b(f) = \int d\Omega I_\nu(\hat{\mathbf{r}}) \int dt \gamma(t) A_\nu(\hat{\mathbf{r}}, t) e^{i2\pi \frac{\nu}{c} \mathbf{b}(t) \cdot \hat{\mathbf{r}}} e^{-i2\pi f t}, \quad (2)$$

where $\gamma(t)$ is a tapering function for the Fourier transform in time, which we can take to be centered around $t = 0$ without loss of generality. 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, $A_\nu(\hat{\mathbf{r}}, t) \approx A_\nu(\hat{\mathbf{r}}, 0) \equiv A_{\nu,0}(\hat{\mathbf{r}})$. Additionally, we may take the time-dependence of the baselines to leading order, with

$$\mathbf{b}(t) \approx \mathbf{b}(0) + \left. \frac{d\mathbf{b}}{dt} \right|_{t=0} t + \dots = \mathbf{b}_0 + (\boldsymbol{\omega}_\oplus \times \mathbf{b}_0)t + \dots \quad (3)$$

where $\boldsymbol{\omega}_\oplus$ is the angular velocity vector of the Earth's rotation, and $\mathbf{b}_0 \equiv \mathbf{b}(0)$. 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 ω_{\oplus} . Inserting these approximations into Equation (2) yields

$$\begin{aligned}\tilde{V}_{\nu}^b(f) &= \int d\Omega I_{\nu}(\hat{\mathbf{r}}) A_{\nu,0}(\hat{\mathbf{r}}) e^{i2\pi \frac{\nu}{c} \mathbf{b}_0 \cdot \hat{\mathbf{r}}} \\ &\quad \times \int dt \gamma(t) e^{i2\pi [\frac{\nu}{c} (\omega_{\oplus} \times \mathbf{b}_0) \cdot \hat{\mathbf{r}} - f] t} \\ &= \int d\Omega I_{\nu}(\hat{\mathbf{r}}) A_{\nu,0}(\hat{\mathbf{r}}) e^{i2\pi \frac{\nu}{c} \mathbf{b}_0 \cdot \hat{\mathbf{r}}} \\ &\quad \times \tilde{\gamma} \left[\frac{\nu}{c} (\omega_{\oplus} \times \mathbf{b}_0) \cdot \hat{\mathbf{r}} - f \right], \quad (4)\end{aligned}$$

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 $f \approx \hat{\mathbf{r}} \cdot \omega_{\oplus} \times \mathbf{b}_0 \nu / c$.

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 an east-west baseline located at XXX. Contours of constant fringe-rate correspond to locations on the sky $\hat{\mathbf{r}}$ that have the same degenerate combination of $\hat{\mathbf{r}} \cdot \omega_{\oplus} \times \mathbf{b}_0$. Note that this combination can also be rewritten as $\omega_{\oplus} \cdot \mathbf{b}_0 \times \hat{\mathbf{r}}$ or $\mathbf{b}_0 \cdot \hat{\mathbf{r}} \times \omega_{\oplus}$ by cyclic permutation. Thus, if any two of \mathbf{b}_0 , ω_{\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 ω_{\oplus} . Similarly, a north-south only baseline located at the equator would have \mathbf{b}_0 parallel to ω_{\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

$$\begin{aligned}V_{\text{filt},\nu}^b(t) &= \int df w(f) \tilde{V}_{\nu}^b(f) \\ &= \int d\Omega I_{\nu}(\hat{\mathbf{r}}) A_{\nu,0}(\hat{\mathbf{r}}) e^{i2\pi \frac{\nu}{c} \mathbf{b}_0 \cdot \hat{\mathbf{r}}} \\ &\quad \times \int df w(f) \tilde{\gamma} \left[\frac{\nu}{c} (\omega_{\oplus} \times \mathbf{b}_0) \cdot \hat{\mathbf{r}} - f \right], \\ &\equiv \int d\Omega I_{\nu}(\hat{\mathbf{r}}) A_{\nu,0}^{\text{eff}}(\hat{\mathbf{r}}) e^{i2\pi \frac{\nu}{c} \mathbf{b}_0 \cdot \hat{\mathbf{r}}}, \quad (5)\end{aligned}$$

where we have defined an *effective primary beam*

$$A_{\nu,0}^{\text{eff}}(\hat{\mathbf{r}}) \equiv A_{\nu,0}(\hat{\mathbf{r}}) (w * \tilde{\gamma}) \left[\frac{\nu}{c} (\omega_{\oplus} \times \mathbf{b}_0) \cdot \hat{\mathbf{r}} \right], \quad (6)$$

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 $\hat{\mathbf{r}} \cdot \omega_{\oplus} \times \mathbf{b}_0$. For any given instant, $\omega_{\oplus} \times \mathbf{b}_0$ 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 $\hat{\mathbf{r}} \cdot \omega_{\oplus} \times \mathbf{b}_0$, 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 $f \approx \hat{\mathbf{r}} \cdot \omega_{\oplus} \times \mathbf{b}_0 \nu / c$ in the data. Since $\hat{\mathbf{r}} \cdot \omega_{\oplus} \times \mathbf{b}_0$ can never exceed $\omega_{\oplus} b_0$, the maximal fringe-rate that can be achieved by a source locked to the celestial sphere is $f_{\text{max}} = \omega_{\oplus} b_0 \nu / c$.

note that because the $\tilde{\gamma}$ term in Equation (4) is relatively compact,

Weighting different parts of the sky Integrating together stuff Actually two sides of the same thing, because we're optimal mapmaking requires that.

Fringe-rate filtering amounts to decomposing into the fringe-rates, weighting fringe rates by some function $w(f)$, and then transforming back to time. In other words,

$$V_{\text{filt}}(t) = \int \frac{df}{1/T} w(f) \tilde{V}_b(f) \quad (7)$$

$$\begin{aligned}V_b(t) &= \int B(\hat{\mathbf{r}}, t) T(\hat{\mathbf{r}}) \exp \left[-i2\pi \left(\frac{b_y}{\lambda} \cos \eta \cos \theta \right) \right] \\ &\quad \times \exp \left[-i2\pi \left(\frac{b_0}{\lambda} \sin \theta \sin(\varphi - \omega_{\oplus} t) \right) \right] d\Omega + n(t), \quad (8)\end{aligned}$$

where $n(t)$ is the instrumental noise, θ and φ are polar and azimuthal angles fixed to the celestial sphere, respectively, $B(\hat{\mathbf{r}}, t)$ is the primary beam, λ is the wavelength, ω_{\oplus} is the angular frequency of the Earth's rotation, 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. With this measurement equation, 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

One way of handling this additional integration is via gridding in the uv -plane. Each measured visibility in a wide-field interferometer represents the integral over a kernel in the uv -plane that reflects the primary beam of the elements (??) and the w component of the baseline (?). As noted in ? and ?, in order to optimally account for the mode-mixing introduced by these kernels, gridding kernels must be used that correctly distribute each measurement among the sampled uv -modes, such that, in the ensemble average over many measurements

by many baselines, each uv -mode becomes an optimally weighted estimator of the actual value given the set of measurements.

However, this approach has a major shortcoming when applied to maximum-redundancy array configurations. In order to maximize sensitivity, such configurations are set up to deliberately sample identical visibilities that reflect the same combinations of modes in the uv plane, with few nearby measurements (P12a). As a result, such array configurations tend to lack enough measurements of different combinations of uv modes to permit the ensemble average to converge on the true value. Said differently, maximum-redundancy array configurations tend to produce measurement sets that, when expressed as linear combinations of uv -modes of interest, are singular.

Our alternative approach avoids this and many of the difficulties outlined in ? by applying a carefully tailored fringe-rate filter to each time series of visibility spectra.

Similar geometric limits apply to the variation of visibilities in the time dimension. As described in Equation 7 of ?, the rate of change of the geometric delay versus time — that is, delay-rate — is given by

$$\frac{d\tau_g}{dt} = -\omega_{\oplus} \left(\frac{b_x}{c} \sin H + \frac{b_y}{c} \cos H \right) \cos \delta, \quad (9)$$

where $\mathbf{b} = (b_x, b_y, b_z)$ is the baseline vector expressed in equatorial coordinates, ω_{\oplus} is the angular frequency of the earth's rotation, and H, δ are the hour-angle and declination of a point on the celestial sphere, respectively. As a result, there exists a maximum rate of change based on the length of a baseline projected to the $z = 0$ equatorial plane. For arrays not deployed near the poles, $|b_y| \gg |b_x|$ (i.e., they are oriented more along the east-west direction than radially from the polar axis), and the maximum rate of change corresponds to $H = 0$ and $\delta = 0$, where we have

$$-\omega_{\oplus} \frac{|b_y|}{c} \leq \frac{d\tau_g}{dt} \leq \omega_{\oplus} \frac{|b_y|}{c}. \quad (10)$$

For a maximum east-west baseline length in the PAPER array of 300m, $\omega_{\oplus}|b_y|/c$ is approximately 0.07 ns/s. To better elucidate the meaning of this bound, we take the Fourier transform along the time axis (see Equation 8 in ?) for a model visibility consisting of a single point source located at the point of maximum delay-rate, which gives us

$$\begin{aligned} \tilde{V}(\nu, f) &\approx \tilde{A}(\nu, f) * \tilde{S}(\nu) * \int e^{2\pi i \omega_{\oplus} \frac{b_y \nu}{c} t} e^{-2\pi i f t} dt \\ &\approx \tilde{A}(\nu, f) * \tilde{S}(\nu) * \delta_D\left(\frac{b_y}{c} \omega_{\oplus} \nu - f\right), \end{aligned} \quad (11)$$

where f is the fringe-rate of the source⁴, $\tilde{A}(\nu, f)$ indicates the Fourier transform of the antenna response along the time direction, and approximation is indicated because we assume $|b_y| \gg |b_x|$, and because the Fourier transform must involve a discrete length of time, during which our assumption that $\cos H \approx 1$ breaks down at second order. The delta function above gives rise to the expression for the maximum fringe rate in Equation 56.

This example of a source with a maximal fringe-rate serves to show that a filter may be applied in delay-rate domain, using the fact that the maximum delay-rate is bounded by the maximum fringe rate within the

⁴ Delay rate is equivalent to the frequency-integrated fringe rate.

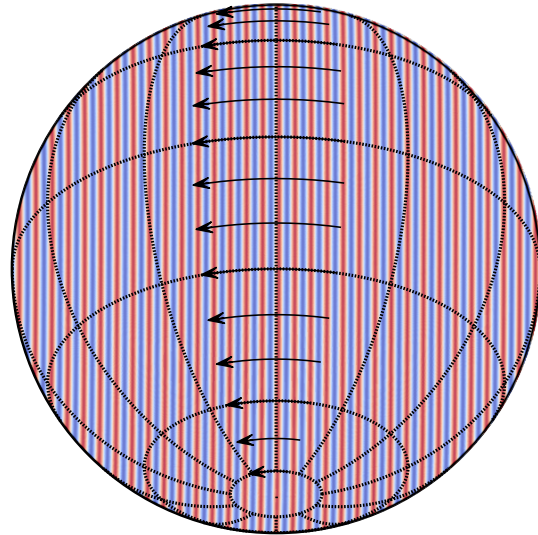


FIG. 1.— The fringe pattern at 150 MHz of a 30-m east-west baseline, overlaid with arrows indicating the distance traversed by sources at various declinations over a two-hour span centered at transit. In a fixed time interval, sources near $\delta = 0^\circ$ traverse more fringe periods than sources 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.

band (i.e. evaluating Equation 56 at the maximum ν involved in the delay transform), to remove emission that exceeds the variation dictated by array geometry for sources locked to the celestial sphere. As in the delay filtering case, assuming the geometric bounds on delay rate implicitly assumes that \tilde{A} and \tilde{S} are compact in f , which is to say that instrumental responses and celestial emission must be smooth in time; variable emission from, e.g., fast-transients will be heavily suppressed by such delay-rate filters. For PAPER, with a maximum baseline length of 300m and a maximum observing frequency of 200 MHz, the maximum delay-rate has a period of 68.5s. As described in §??, at PAPER's latitude, delay-rates range from $-f_{\max}/2$ to f_{\max} . Filtering delay-rates to this range corresponds in sensitivity to an effective integration time of 45.2 s. An example of the bounds of a delay filter in DDR space is given by the shaded magenta region in Figure ???. As in the delay filtering case, filtering along the delay-rate axis permits substantial down-sampling of the signal, which is the basis for the reduction in data volume along the time axis. We note that for the analysis in §??, we choose to use a slightly wider delay-rate filter to be conservative in our first application of this technique, corresponding to an integration time of 43 seconds.

(?), (?),

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

Having provided some basic intuition for fringe-rate filtering, we now provide a more formal treatment of the problem. In what follows, we will consider the general problem of combining time-ordered visibilities into a map of the sky. We will see that the optimal recipe for doing this is *not* to time average the visibilities, but instead, to apply a fringe-rate filter.

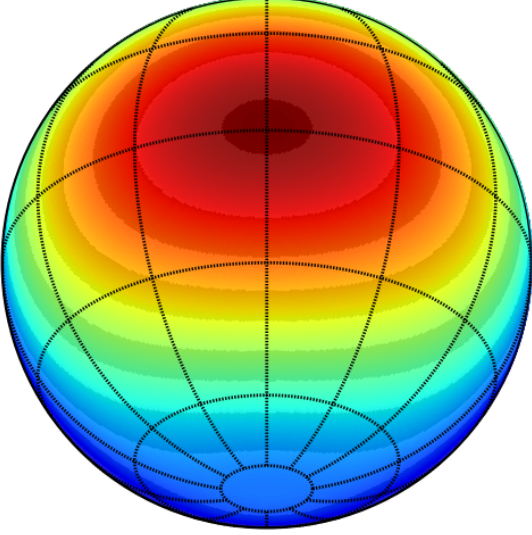


FIG. 2.— Fringe rate as a function of sky position, corresponding to the fringe pattern illustrated in Figure 1. Fringe rates peak at XXX mHz at $\delta = 0^\circ$, hit zero at the south celestial pole, and become negative on the far side of the pole. Grey shading indicates the approximate angular regions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

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 $N_b N_t \times N_{\text{pix}}$, the measurement equation is given by

$$\mathbf{v} = \mathbf{A}\mathbf{x} + \mathbf{n}, \quad (12)$$

where \mathbf{n} is a noise vector. Note that in this general form, Equation (12) 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, a central message of this paper is that an optimal data analysis prescription is more naturally phrased in terms of the Fourier dual to time, i.e. fringe-rate.

Given our measurement \mathbf{v} , the optimal estimator $\hat{\mathbf{x}}$ of the true sky \mathbf{x} is given by (??) [XXX: add a citation to Josh as well]

$$\hat{\mathbf{x}} = \mathbf{M}\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v}, \quad (13)$$

where \mathbf{M} is some invertible matrix chosen by the data analyst, and \mathbf{N} is the noise covariance matrix, defined as $\langle \mathbf{nn}^\dagger \rangle$, 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 Σ , which is given by

$$\Sigma \equiv \langle (\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^\dagger \rangle = \mathbf{M}\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A} \mathbf{M}^\dagger. \quad (14)$$

With a suitable choice of \mathbf{M} , the estimator given by Equation (13) minimizes the variance. Regardless of one’s choice, however, Equation (13) can be shown to be lossless (?), 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 (13) 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 ? 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.

3.1. The general sub-optimality of time integration

We begin by showing that it is suboptimal to make maps by integrating in time. Consider the visibility response $V_b(t)$ of an interferometer baseline b at time t to the sky $T(\hat{\mathbf{r}})$:

$$V_b(t) = \int B(\hat{\mathbf{r}}, t) T(\hat{\mathbf{r}}) \exp \left[-i2\pi \left(\frac{b_y}{\lambda} \cos \eta \sin \delta \right) \right] \times \exp \left[-i2\pi \left(\frac{b_0}{\lambda} \cos \delta \sin(\alpha - \omega_\oplus t) \right) \right] d\Omega + n(t), \quad (15)$$

where $n(t)$ is the instrumental noise, α and δ are the right ascension and declination, respectively, $B(\hat{\mathbf{r}}, t)$ is the primary beam, λ is the wavelength, ω_\oplus is the angular frequency of the Earth’s rotation, 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. With this measurement equation, 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 (15) 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 $B(\hat{\mathbf{r}}, t) \equiv \delta^{(D)}(\delta) B_\alpha(\alpha - \omega_\oplus t)$, where $\delta^{(D)}$ signifies a Dirac delta function. Plugging these into restrictions

into our equation, we obtain

$$V_b(t) = \int B_\alpha(\alpha - \omega_\oplus t) T(\delta = 0, \alpha) \times \exp \left[-i2\pi \frac{b_x}{\lambda} \sin(\alpha - \omega_\oplus t) \right] d\alpha + n(t). \quad (16)$$

For a single baseline, the function $V_b(t)$ is precisely the continuous version of the discrete data vector \mathbf{v} . To obtain \mathbf{v} , then, one would simply sample $V_b(t)$ 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 $T(\theta = \pi/2, \varphi)$ as the continuous versions of \mathbf{n} and \mathbf{x} respectively, the rest of Equation (16)'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

$$N_{bb'}(t, t') = \sigma^2 \delta_{bb'} \delta(t - t'), \quad (17)$$

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 (13) 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

$$(\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v})_\alpha = \sum_b \int \frac{dt}{\sigma^2} B_\alpha(\alpha - \omega_\oplus t) e^{i2\pi \frac{b_x}{\lambda} \sin(\alpha - \omega_\oplus t)} V_b(t), \quad (18)$$

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 (18) 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 type of fringe-rate filtering in full generality in Section 3.3, demonstrating that its optimality is not a peculiarity of our current pedagogical example, but is instead applicable in all situations.

3.2. 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 used so ubiquitously in the literature. An inspection of Equation (18) 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 $B(\hat{\mathbf{r}}, t)$ would simply become $B(\hat{\mathbf{r}})$ in Equation (15). 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

$$V_b(t) = \int B(\hat{\mathbf{r}}) T(\hat{\mathbf{r}}) \exp \left[-i2\pi \left(\frac{b_0}{\lambda} \sin(\alpha - \omega_\oplus t) \right) \right] \times \exp \left[-i2\pi \left(\frac{b_y}{\lambda} \cos \eta \cos \delta \right) + i\psi(t) \right] d\Omega + n(t), \quad (19)$$

[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

$$(\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v})_{(\delta, \alpha)}^{\text{flat}} = \frac{B(\delta, \alpha)}{\sigma^2} e^{i2\pi \frac{b_y}{\lambda} \cos \eta \cos \delta} \times \sum_b \int dt e^{i2\pi \frac{b_x}{\lambda} \sin(\alpha - \omega_\oplus t) - i\psi(t)} V_b(t). \quad (20)$$

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

$$(\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v})_{(\theta, \varphi)}^{\text{flat, short}} = \frac{B(\delta, \alpha)}{\sigma^2} e^{i2\pi \left(\frac{b_y}{\lambda} \cos \eta \cos \delta + \frac{b_x}{\lambda} \sin \alpha \right)} \times \sum_b \int dt V_b(t), \quad (21)$$

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.

[XXX: talk about how the same issues apply for a uv -plane description.]

[XXX: Maybe mention instantaneous snapshots and how it relates?]

[XXX: Fringe stopping at higher frequencies.]

3.3. 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 (15). From our toy example (Equation 18), 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

$$\tilde{V}_b(f) \equiv \frac{1}{T_{\oplus}} \int_{-T_{\oplus}/2}^{T_{\oplus}/2} dt \exp(-2\pi i f t) V_b(t), \quad (22)$$

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 n^{th} bin is given by $f_n \equiv n/T_{\oplus}$, where n is an integer. The measurement in the n^{th} bin is then given by

$$\tilde{V}_b(f_n) = \int d\Omega T(\hat{\mathbf{r}}) e^{-i2\pi \frac{b_y}{\lambda} \cos \eta \sin \delta} \times \int_{-\frac{T_{\oplus}}{2}}^{\frac{T_{\oplus}}{2}} \frac{dt}{T_{\oplus}} B(\hat{\mathbf{r}}, t) e^{-i\frac{2\pi n t}{T_{\oplus}} + i\frac{2\pi b_0}{\lambda} \cos \delta \sin(\omega_{\oplus} t - \alpha)}, \quad (23)$$

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

$$B(\hat{\mathbf{r}}, t) \equiv B_{\delta}(\delta) B_{\alpha}(\alpha - \omega_{\oplus} t), \quad (24)$$

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

$$B(\hat{\mathbf{r}}, t) = B_{\delta}(\delta) \sum_q \tilde{B}_q e^{-iq\alpha} e^{iq\omega_{\oplus} t}, \quad (25)$$

where $\tilde{B}_q \equiv \int \frac{d\alpha}{2\pi} B_{\alpha}(\alpha) e^{iq\alpha}$ is the q^{th} Fourier coefficient. Plugging this into Equation (23) and making the substitution $\psi \equiv \omega_{\oplus} t - \varphi$, one obtains

$$\tilde{V}_b(f_n) = \int d\Omega T(\hat{\mathbf{r}}) B_{\delta}(\delta) e^{-i2\pi \frac{b_y}{\lambda} \cos \eta \sin \delta} \times \sum_q \frac{\tilde{B}_q e^{-in\varphi}}{2\pi} \int_{-\pi-\varphi}^{\pi+\varphi} d\psi e^{i(q-n)\psi + i\frac{2\pi b_0}{\lambda} \cos \delta \sin \psi}. \quad (26)$$

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:

$$\tilde{V}_b(f_n) = \int \frac{d\Omega}{2\pi} T(\hat{\mathbf{r}}) B_{\delta}(\delta) e^{-i2\pi \frac{b_y}{\lambda} \cos \eta \sin \delta} e^{-in\alpha} \times \sum_q \tilde{B}_q J_{n-q} \left(\frac{2\pi b_0}{\lambda} \cos \delta \right). \quad (27)$$

Several features are of note here. For wide primary beams, \tilde{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 ,

$$J_n \left(\frac{2\pi b_0 \cos \delta}{\lambda} \right) \approx \frac{1}{n!} \left(\frac{\pi b_0 \cos \delta}{\lambda} \right)^n, \quad (28)$$

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

$$(\mathbf{A}^{\dagger} \mathbf{N}^{-1} \mathbf{v})_{\delta, \alpha} = \frac{B_{\delta}^*(\delta) \cos \delta}{2\pi \sigma^2} \sum_{b, n} e^{in\alpha} e^{i2\pi \frac{b_y}{\lambda} \cos \eta \sin \delta} \times \sum_q \tilde{B}_q^* J_{n-q} \left(\frac{2\pi b_0}{\lambda} \cos \delta \right) \tilde{V}_b(f_n). \quad (29)$$

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 $\sum_q \tilde{B}_q^* J_{n-q} \left(\frac{2\pi b_0}{\lambda} \cos \delta \right)$, 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 (24). Returning to Equations (22) and (23), 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

$$\tilde{V}_b(f_n) = \int d\Omega T(\hat{\mathbf{r}}) g_{bn}(\hat{\mathbf{r}}) \approx \frac{4\pi}{N} \sum_i T(\hat{\mathbf{r}}_i) g_{bn}(\hat{\mathbf{r}}_i), \quad (30)$$

where the key quantity is

$$g_{bn}(\hat{\mathbf{r}}) \equiv e^{-i2\pi \frac{b_y}{\lambda} \cos \eta \sin \delta} \times \int_{-\frac{T_{\oplus}}{2}}^{\frac{T_{\oplus}}{2}} \frac{dt}{T_{\oplus}} B(\hat{\mathbf{r}}, t) e^{-i\frac{2\pi n t}{T_{\oplus}} + i\frac{2\pi b_0}{\lambda} \cos \delta \sin(\omega_{\oplus} t - \alpha)}. \quad (31)$$

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

$$(\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v})_{\delta, \alpha} \propto \sum_{n, b} g_{bn}^*(\hat{\mathbf{r}}) \tilde{V}_b(f_n). \quad (32)$$

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.

4. MAPPING FRINGE-RATES TO SPATIAL LOCATIONS OVER SHORT INTEGRATION TIMES

In the previous section, we considered a general way to combine time-ordered interferometric data, and found that fringe-rate filtering provided a natural way to do so. In this section, we provide an approximate treatment of the more realistic case where one cannot integrate over an entire day. The result will provide a rigorous link to the geometric intuition discussed in Section 2.

$$V_b(t) = \int d\Omega B(\hat{\mathbf{r}}, t) T(\hat{\mathbf{r}}) \exp\left(i \frac{2\pi}{\lambda} \mathbf{b}(t) \cdot \hat{\mathbf{r}}\right) \quad (33)$$

$$\tilde{V}_b(f) = \int d\Omega T(\hat{\mathbf{r}}) \int_{-\frac{T}{2}}^{\frac{T}{2}} \frac{dt}{T} B(\hat{\mathbf{r}}, t) e^{i \frac{2\pi}{\lambda} \mathbf{b}(t) \cdot \hat{\mathbf{r}}} e^{-i2\pi f t} \quad (34)$$

For short periods of time, $B(\hat{\mathbf{r}}, t) \approx B(\hat{\mathbf{r}}, t=0) \equiv B_0(\hat{\mathbf{r}})$. Also,

$$\mathbf{b}(t) = \mathbf{b}(t=0) + \left. \frac{d\mathbf{b}}{dt} \right|_{t=0} t + \dots = \mathbf{b}_0 + (\boldsymbol{\omega}_\oplus \times \mathbf{b}_0)t + \dots \quad (35)$$

where $\boldsymbol{\omega}_\oplus$ is the angular velocity vector of the Earth's rotation, and $\mathbf{b}_0 \equiv \mathbf{b}(t)$. 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 $\boldsymbol{\omega}_\oplus$. With this,

$$\begin{aligned} \tilde{V}_b(f) &= \int d\Omega T(\hat{\mathbf{r}}) B_0(\hat{\mathbf{r}}) e^{i \frac{2\pi}{\lambda} \mathbf{b}_0 \cdot \hat{\mathbf{r}}} \\ &\times \int_{-\frac{T}{2}}^{\frac{T}{2}} \frac{dt}{T} \exp\left[i \frac{2\pi}{\lambda} \hat{\mathbf{r}} \cdot (\boldsymbol{\omega}_\oplus \times \mathbf{b}_0)t - i2\pi f t\right] \\ &= \int d\Omega T(\hat{\mathbf{r}}) B_0(\hat{\mathbf{r}}) \text{sinc}\left[\pi T \left(\hat{\mathbf{r}} \cdot \boldsymbol{\omega}_\oplus \times \frac{\mathbf{b}_0}{\lambda} - f\right)\right] e^{i \frac{2\pi}{\lambda} \mathbf{b}_0 \cdot \hat{\mathbf{r}}} \end{aligned}$$

The sinc function is peaked near zero argument, and thus its presence in the integral enforces that $f \approx \hat{\mathbf{r}} \cdot \boldsymbol{\omega}_\oplus \times \mathbf{b}_0/\lambda$, picking out only specific portions of the celestial sphere. As advertised, different fringe-rates correspond to different portions of the sky.

Fringe-rate filtering amounts to decomposing into the fringe-rates, weighting fringe rates by some function $w(f)$, and then transforming back to time. In other words,

$$V_{\text{filt}}(t) = \int \frac{df}{1/T} w(f) \tilde{V}_b(f) \quad (37)$$

5. POWER SPECTRUM ESTIMATION WITH FRINGE-RATE FILTERING

We now discuss the problem of power spectrum estimation given time-ordered data. The primary challenge, again, is to ensure that the benefits of a long time-integration (i.e. increased sensitivity) can be realized without explicitly making maps. For example, one cannot simply average over power spectrum estimates formed from every time instant of some time-ordered visibilities. Doing so would reduce the sensitivity of the final power spectrum, as coherent information between time-samples would be lost. However, it would be desirable to be able to optimally include coherent information while remaining in a time-ordered basis where systematic effects are transparent. In this section, we will show that the techniques of fringe-rate filtering make it possible to construct a power spectrum estimator that satisfies these requirements.

We start with the estimator \hat{p}_α for the α th band of the power spectrum:

$$\hat{p}_\alpha = N_\alpha \hat{\mathbf{x}}^\dagger \mathbf{C}^{-1} \mathbf{Q}_\alpha \mathbf{C}^{-1} \hat{\mathbf{x}}, \quad (38)$$

where $\mathbf{C} \equiv \langle \hat{\mathbf{x}} \hat{\mathbf{x}}^\dagger \rangle = \mathbf{S} + \boldsymbol{\Sigma}$ is the *total* (signal plus noise) covariance matrix of the sky map, and N_α is a normalization factor, given by⁵

$$N_\alpha = \text{tr}[\mathbf{C}^{-1} \mathbf{Q}_\alpha \mathbf{C}^{-1} \mathbf{Q}_\alpha]. \quad (39)$$

[XXX: grr...the Fisher matrix isn't diagonal here, so it's a little more annoying when it comes to normalization. Need to decide which norm convention to use]. Written in this general form, our estimator can be used to estimate any power spectrum, be it an angular power spectrum C_ℓ (in which case $\hat{p}_\alpha \equiv \hat{C}_\ell$ and α would index the ℓ value), or a three-dimensional power spectrum $P(\mathbf{k})$ (in which case $\hat{p}_\alpha \equiv \hat{P}(\mathbf{k}_\alpha)$). The type of power spectrum under consideration is encoded by the form of

⁵ More sophisticated normalization schemes involving matrix normalization constants are possible [XXX: cite Josh MWA paper], but for simplicity (and no loss of generality as far as our present discussion is concerned) we do not consider them here.

the \mathbf{Q}_α matrices. They are defined as the linear response of the covariance matrix \mathbf{C} to the power spectrum:

$$\mathbf{Q}_\alpha \equiv \frac{\partial \mathbf{C}}{\partial p_\alpha}. \quad (40)$$

The inverse-covariance weighting enacted by \mathbf{C}^{-1} makes Equation (38) a minimum-variance estimator (?). For notational cleanliness we have omitted any corrections to *additive* biases in the estimator, implicitly assuming that such biases have already been removed.

As it stands, our estimator is written explicitly in terms of the estimated sky maps $\hat{\mathbf{x}}$. These are related to the visibilities via Equation (13), which states that $\hat{\mathbf{x}} = [\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}]^{-1} \mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v}$. If we assume that our measurements are noise-dominated, then $\mathbf{C} \approx \mathbf{\Sigma} = [\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}]^{-1}$, so $\mathbf{C}^{-1} \approx [\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}]$. Inserting these into Equation (38), we obtain

$$\hat{p}_\alpha = N_\alpha \hat{\mathbf{v}}^\dagger \mathbf{N}^{-1} \mathbf{A} \mathbf{Q}_\alpha \mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v} \quad (41)$$

as our visibility-based estimator.

To see how fringe-rate filters relate to this expression, consider the specific example of measuring the angular power spectrum C_ℓ . In this case, we parameterize the signal covariance in terms of spherical harmonics, so that

$$\mathbf{C}_{\hat{\mathbf{r}}, \hat{\mathbf{r}}'} = [\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}]^{-1} + \sum_{\ell m} Y_{\ell m}(\hat{\mathbf{r}}) Y_{\ell m}^*(\hat{\mathbf{r}}') C_\ell, \quad (42)$$

and thus

$$\mathbf{Q}_\ell \equiv \frac{\partial \mathbf{C}}{\partial C_\ell} = \sum_m Y_{\ell m}(\hat{\mathbf{r}}) Y_{\ell m}^*(\hat{\mathbf{r}}'), \quad (43)$$

where we have renamed the index α to ℓ as a reminder that we are considering the angular power spectrum. Note that since the noise covariance does not depend on the power spectrum C_ℓ , it is necessary to include the signal covariance when deriving \mathbf{Q}_ℓ , even though this was unnecessary for \mathbf{C}^{-1} .

Substituting Equations (29) and (43) into Equation (41), we obtain

$$\hat{C}_\ell \propto \sum_{m=-\infty}^{\infty} \left| \sum_b w_{\ell m}^b \tilde{V}_b(f_m) \right|^2, \quad (44)$$

where

$$w_{\ell m}^b \equiv \begin{cases} 2\pi \int d(\cos \theta) N_\ell^m P_\ell^m(\cos \theta) g_b(\theta, m), & \text{for } -\ell \leq m \leq \ell \\ 0, & \text{otherwise,} \end{cases} \quad (45)$$

with P_ℓ^m being the associated Legendre polynomial of order [XXX: insert proper name here], and

$$N_\ell^m = \sqrt{\frac{2\ell+1}{4\pi} \frac{(\ell-m)!}{(\ell+m)!}} \quad (46)$$

is the normalization constant for spherical harmonics. Since we can equivalently interpret m not as the azimuthal Fourier mode number but as the fringe-rate bin index, it is the Fourier dual to time, our estimator can be rewritten using Parseval's theorem:

$$\hat{C}_\ell \propto \int_{-T_\oplus/2}^{T_\oplus/2} \frac{dt}{T_\oplus} \left| \sum_b V_{\ell b}^{\text{filt}}(t) \right|^2, \quad (47)$$

where

$$V_{\ell b}^{\text{filt}}(t) \equiv \sum_m \exp \left[i \frac{2\pi m t}{T_\oplus} \right] w_{\ell m}^b \tilde{V}_b(f_m). \quad (48)$$

Together, Equations 47 and 48 suggest a method for estimating power spectra. First, time-ordered visibilities are Fourier-transformed into a fringe-rate basis. Different fringe-rate bins are then weighted by the $w_{\ell m}^b$ weights, which essentially encode the “dot product” between a baseline’s response in the fringe-rate bin and the angular mode in question. The result is then transformed back into a set of fringe-rate filtered visibility time series. A power spectrum is computed for every time instant before averaging together different times. Since this procedure was derived from a minimum-variance estimator of the power spectrum, it is guaranteed to produce the smallest possible error bars.

We have thus accomplished our goal of writing down an optimal estimator for the power spectrum that works directly with visibilities. Conveniently, the integration of coherent information takes place via a fringe-rate filter, which operates on independently for different baselines. Moreover, since it returns data in its original format—a time-series—systematics localized to certain baselines and certain times can be easily isolated. Note that while the filtering is performed per-baseline, information from multiple baselines regarding the same ℓ mode is coherently included before squaring to form the power spectrum. This is most clearly seen in Equation (44), where visibilities from different baselines are simply added together, except that they are weighted by different amounts depending on how much overlap they have with the mode in question. Traditionally, taking advantage of coherent information between non-identical baselines requires imaging (whether in the image basis or as a Fourier-transformed image in the form of, say, uv -plane data). With our approach, imaging (and its attendant artifacts) are completely avoided without sacrificing sensitivity. [XXX: JUST REALIZED THIS IS FAR MORE ALGORITHMICALLY EFFICIENT THAN MAPMAKING!!!]

[XXX: Maybe also consider $l \sim 2\pi b/\lambda$ approx.] [XXX: Perhaps make connection to thingy more clear.] [XXX: Consider also playing up the weights as the “dot product” of the thing you’re interested in and the fringe-rate-space visibilities to emphasize how different baselines just need to be combined based on their overlap.] [XXX: Need to find a place to talk about effective beams and how it affects power spectrum sensitivity. Talk about how we will henceforth use simulated effective beams so that we don’t have to make all the assumptions that went into the analytic work. Should also credit Richard for the fringe-rate \leftrightarrow m-mode correspondence more throughout.]

6. APPLICATIONS

In this section, we discuss the use of fringe-rate filtering in several applications in low-frequency radio interferometry. So far, we have motivated fringe-rate filtering as a method for optimally weighting noisy data baseline by baseline. As we examined in §XXX, fringe rates map to declinations, giving fringe-rate filters an alternate interpretation as a spatial filter that modifies beam response on a per-baseline basis. The choice for this spatial filter that we have examined so far optimizes the SNR by

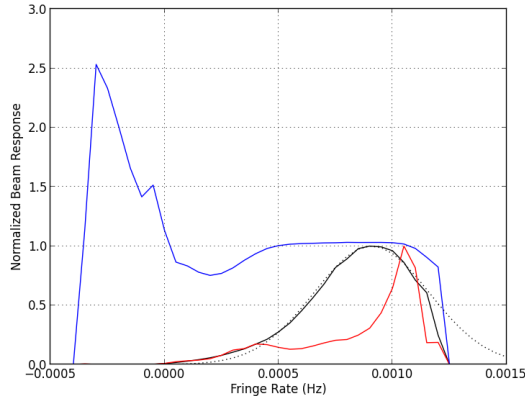


FIG. 3.—

weighting according to the beam response within fringe rate (declination) bins. However, this is not the only choice available. More generally, fringe-rate filters can be chosen to optimize any metric that varies spatially by re-weighting the interferometric response as a function of declination. In this sense, fringe-rate filtering can be regarded as a “beam sculpting” operation that can be used to tailor an interferometer’s spatial response.

As we will discuss below, fringe-rate filtering as a beam sculpting operation is a powerful tool. However, it is fundamentally limited in that it can only perform a one-dimensional re-weighting (in declination) of the two-dimensional sky response⁶. Moreover, the spatial resolution of the weighting function is limited by the temporal resolution of the fringe-rate filter as it interacts with the extent of the sky in fringe-rate space. As the presence of b_y in the expression for the spatial response of a fringe-rate bin in Equation ?? indicates, baselines with longer east-west extents discriminate between declinations with finer resolution than do shorter baselines.

Recognizing these limitations, we now explore the application of fringe-rate filters for sculpting beam response. In addition to the signal-to-noise metric addressed previously, we will explore two other metrics that fringe-rate filters can be chosen to optimize: polarization response and instrumental systematics.

6.1. Implementation

In order to demonstrate fringe-rate filtering in the applications discussed below, we will make repeated use of simulated data. In keeping with its origins targeting observations from the PAPER array, we choose for our simulations a model array based on PAPER, deployed at a latitude -30° and featuring the beam response pattern characteristic of PAPER dipole elements (??). 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 ar-

ray in the maximum-redundancy array configuration it uses for power spectral measurements (??). As such, this simulation serves to demonstrate the performance of fringe-rate filtering in the context of the specific instrument configuration that has been used to place the current best upper limits on 21cm emission from cosmic reionization.

To implement the fringe-rate filters described below, we begin with the ideal filter shape in fringe-rate space. In the case of the beam-weighted fringe-rate filter described in §6.3, this ideal filter shape takes a truncated Gaussian form where the peak and width have been fit to the projection of the beam model in fringe-rate space, as illustrated in Figure ?. The beam model projection is obtained by binning the beam power along fringe-rate contours, as illustrated for coarse fringe-rate bins in Figure ?. The analytic Gaussian form is then truncated at the maximum fringe rate to obtain a fringe-rate filter profile that matches the beam-weighted profile to within a power-averaged RMS of XXX%. While it would be possible to use the beam-weighted profile directly as the ideal fringe-rate filter profile, having an analytic form allows filters to be rapidly generated as a function of baseline length and observing frequency. Since optimal SNR weighting is relatively insensitive to small weighting errors (XXX back this up), this level of match between the computed and analytic forms of the desired fringe-rate filter is acceptable.

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 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 and helps limit the number of samples potentially corrupted by spurious systematics such as RFI.

To summarize, the simulations are based on a 30-m PAPER baseline observing at 150 MHz. While the shape of the fringe-rate filter is application-specific, we generally implement them as FIR filters, with time-domain kernels using coefficients that are spaced to match the 43-second integrations recorded in our simulated observations. Finally, we apply a window function to truncate the wings of this time-domain kernel to increase its compactness in time. As demonstrated in Figure ??, this windowing increases the power-averaged RMS deviation from the ideal filter to XXX%, but the improvement in computational cost, the reduction in impact for any spurious systematics, and the relative insensitivity of the applications described below on the specific filter shape, make this an

⁶ It is worth noting that, as described in ?, signal delay can be used for re-weighting beam response along a second axis. However, delay filtering can only be used as a spatial filter for smooth-spectrum emission, fundamentally limiting its application in this context for 21cm cosmological applications

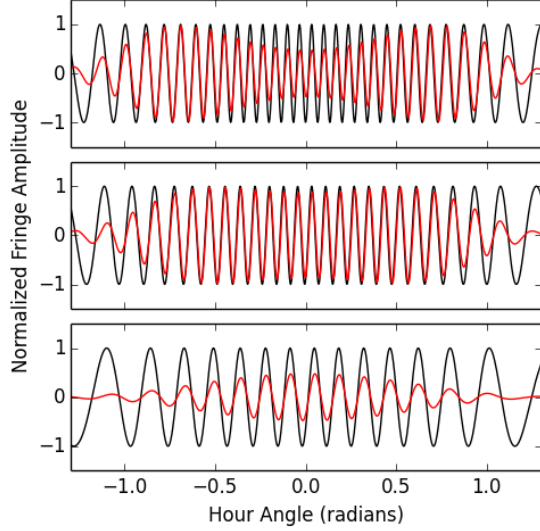


FIG. 4.— Fringe amplitude before (black) and after (red) the application of the fringe-rate filter described in §???. From top to bottom, the panels illustrate the real component of the visibility measured by a 30-m east-west baseline deployed at a latitude of -30° for point sources passing through the fringe pattern at declinations of 0° , -30° , and -60° , respectively. In this simulation, antenna elements have isotropic primary beams.

advantageous trade-off.

6.2. Point Source Simulations for Mapping Beam Response

6.3. Optimizing the Signal-to-Noise Ratio in Power Spectral Measurements

Specific application for statistically isotropic Gaussian random signals. Noise Simulations for Determining Effective Volume.

Discuss beam area as normalization that keeps signal level constant. Refer to ? Appendix A.

We begin by examining the integrated volume, \mathbb{V} , used to normalize the 3D Fourier transform in Equation 3 of P12a. We express this volume in observing coordinates as

$$\mathbb{V} = \Omega B \cdot X^2 Y, \quad (49)$$

where B is the bandwidth, Ω is the angular area, and X, Y are redshift-dependent scalars relating angle and frequency to spatial scales, respectively. Ω arises from the bounds set by $A(l, m, \nu)$, the antenna power response, on the angular extent in the integral

$$\tilde{V}^2(u, v, \eta) = \left(\frac{2k_B}{\lambda^2} \right)^2 \left[\int dl dm d\nu A(l, m, \nu) T(l, m, \nu) e^{-2\pi i(u l + v m + \eta \nu)} \right] \left[\int dl' dm' d\nu' A^*(l', m', \nu') T^*(l', m', \nu') e^{2\pi i(u l' + v m' + \eta \nu')} \right] \quad (50)$$

which is a slightly modified version of Equation 6 of P12a relating the delay-transformed visibility \tilde{V} , sampled at wavemodes u, v (the Fourier complements of angular coordinates l, m) and η (the Fourier complement of spectral frequency ν), to a temperature field T . As shown in

XXX, this reduces to

$$\tilde{V}_{21}^2(u, v, \eta) \approx \left(\frac{2k_B}{\lambda^2} \right)^2 \frac{B}{X^2 Y} \hat{P}(\mathbf{k}) \int dl dm |A(l, m)|^2. \quad (51)$$

We compare this result with the relation between the delay-transformed visibility, \tilde{V} , to the three-dimensional power spectrum of reionization, $P_{21}(\mathbf{k})$ (P12a):

$$\tilde{V}_{21}^2(u, v, \eta) \approx \left(\frac{2k_B}{\lambda^2} \right)^2 \frac{\Omega B}{X^2 Y} \hat{P}_{21}(\mathbf{k}). \quad (52)$$

As this shows, the relevant beam area in Equation 52 is the power-square beam, Ω_{PP} , given by

$$\Omega_{PP} \equiv \int dl dm |A(l, m)|^2. \quad (53)$$

This contrasts with the standard metric for beam area — the integrated power beam — which we will call Ω_P , and is given by

$$\Omega_P \equiv \int dl dm A(l, m), \quad (54)$$

This beam area metric is used to convert visibility measurements from Jy units to mK, but is incorrect for normalizing power spectra that relate to the two-point correlation function of a temperature field.

For equations that relate power-spectrum sensitivity to a system temperature (e.g. Equations 15 and 16 in P12a)

$$\Omega' \equiv \Omega_P^2 / \Omega_{PP} \quad (55)$$

should be used in lieu of Ω , as these equations pick up two factors of Ω_P in the conversion from Jy^2 to mK^2 , along with the factor of Ω_{PP} in the denominator relating to the integrated volume. For equations that relate a measured visibility (in units of brightness temperature, e.g. Equation 57) to $\hat{P}(\mathbf{k})$, the factor of Ω_P^2 is already applied in the conversion from units of Jy to mK, and Ω corresponds to the remaining factor of Ω_{PP} .

For PAPER, $\Omega_P \approx 0.72$ sr, while Ω_{PP} is 0.31 sr. Following the definition above, $\Omega' \approx 1.69$. These beam areas are calculated numerically from a beam model, but typically, Ω' is about a factor of two larger than Ω_P .

Describe simulation, refer to figures, compare results to analytic prediction.

In the last step prior to forming power spectra, we apply a fringe-rate filter to effect time-domain integration, using the effective time interval that a baseline measures a single k -mode to integrate coherently (with noise decreasing as \sqrt{t} , in units of mK), before measurements at different times represent independent modes that must be squared before further integration (with noise now decreasing as \sqrt{t} , in units of mK^2).

As outlined in Appendix ?? in the context of data compression, we take the Fourier transform of the time series in each channel and apply a low-pass filter that preserves fringe-rates that geometrically correspond to sources rotating on the celestial sphere. For a planar array with transit observations, fringe-rates vary according to declination, with fringe rates reaching a maximum (f_{max}) at $\delta = 0^\circ$, decreasing to 0 at $\delta = -90^\circ$, and for an array such as PAPER deployed near -30° S latitude, reaching a minimum of $\approx -f_{\text{max}}/2$ at $\delta = -60^\circ$ on the far side

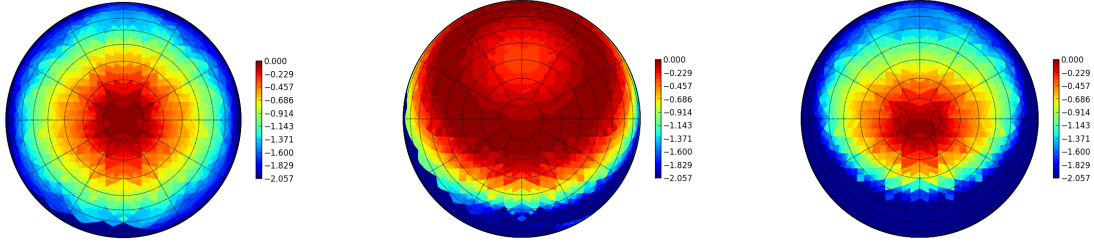


FIG. 5.— The effective primary beam response of a baseline, as determined from the simulations in §6.2. Panels indicate reconstructions of PAPER’s model beam response used in the simulation (left), the beam weighting that results from the application of a fringe-rate filter weighted to optimize SNR for a 30-m baseline with PAPER’s beam response (center), and the effective primary beam response of the baseline after the application of the fringe-rate filter.

of the south celestial pole. In order to avoid introducing undesirable frequency structure, we apply the same filter, tuned to the width set by the highest frequency of the sub-band used in the power spectrum analysis described in §??, to each channel, even though maximum fringe-rates are generally frequency-dependent. In a future paper, we will explore the idea of employing fringe-rate filters that purposely down-weight fringe-rate modes on the sky according to the expected signal-to-noise ratio in each mode. Such filters would essentially correspond to a one-dimensional implementation of the inverse primary beam uv -gridding discussed in ?, and have many features in common with m -mode synthesis described in ?.

Since thermal noise scatters equally into all fringe rate bins, applying a filter that passes only fringe rates corresponding the celestial emission has the effect of denoising the data. We apply such a filter to the data, choosing the bounds of the filter to match the geometric bounds set by a 30-m east-west baseline, according to the equation

$$f_{\max} = \frac{|\mathbf{b}_{\text{eq}}|}{c} \omega_{\oplus} \nu, \quad (56)$$

where f_{\max} is the maximum fringe rate, \mathbf{b}_{eq} is the baseline vector projected parallel to the equatorial plane, c is the speed of light, ω_{\oplus} is the angular frequency of the Earth’s rotation, and ν is the spectral frequency. At 174 MHz (the highest frequency in a 20-MHz window centered on 164 MHz that is used in §??), $f_{\max} = 1.3$ mHz, corresponding to a fringe period of 780 s. Hence, the fringe-rate filter that is applied passes fringe-rates in the range $-0.7 < f < 1.3$ mHz. The width of this filter corresponds in sensitivity to an effective integration time of 525 s. We note that this filtering could have been applied during the data compression described in §??, but was implemented separately to enable the compression to work uniformly over all baselines in the array without additional information about antenna location.

After applying this filter, we transform the data back to time domain in preparation for forming power spectra via the delay transform. It should be noted that, in time domain, the data are now heavily over-sampled; adjacent samples are no longer statistically independent. Hence, when averaging power-spectra versus time, noise will not beat down according to the strict number in samples, but rather, according to the actual number of statistically independent samples underlying the time series.

$$\hat{P}(\mathbf{k}_{t\tau}) = \left(\frac{\lambda^2}{2k_B} \right)^2 \frac{X^2 Y}{\Omega B} \left\langle \tilde{V}_i(\tau, t) \tilde{V}_j^*(\tau, t) \right\rangle_{i < j}, \quad (57)$$

which follows from equation 12 of P12a, with λ being the observing wavelength, k_B is Boltzmann’s constant, $X^2 Y$ is a cosmological scalar with units of $\frac{h^{-3} \text{ Mpc}^3}{\text{sr} \cdot \text{Hz}}$, Ω is the angular area⁷, B is the bandwidth, $\langle \dots \rangle_{i < j}$ indicates the ensemble average over instantaneously redundant baseline measurements indexed by i, j , and $\tilde{V}(\tau, t)$ is the delay-transformed visibility, expressed in terms of delay τ and time t . We use t as a subscript on \mathbf{k} to denote the different modes sampled by a baseline as the sky rotates, and τ to indicate the dependence of \mathbf{k} on the delay mode in question.

$$\hat{\Delta}_{21}^2(k) = \frac{k^3}{2\pi^2} \left\langle \hat{P}(\mathbf{k}_{t\tau}) \right\rangle_{|\mathbf{k}_{t\tau}|=k}, \quad (58)$$

where the three-dimensional symmetry of the power spectrum is invoked to average over all independent measurements of modes in a shell of $|\mathbf{k}| = k$, with independent measurements indexed here by t . As described in §??, the number of independent modes that are averaged (with noise decreasing with number of modes, M , as \sqrt{M} in mK² units; see P12a) is determined by overall observing window and the number of fringe-rate bins that are preserved in the fringe-rate filtering process. Since we have not decimated the number of integrations to the critical sampling rate corresponding to the width of the applied fringe-rate filter, M is *not* the number of integrations. However, we are free to average the power spectrum estimates for each integration, even though nearby samples do not have statistically independent noise, understanding that noise will decrease according to the number of underlying independent samples.

6.4. Polarization Response

⁷ As described in detail in Appendix ??, the angular area used to normalize high-redshift 21cm power spectrum measurements (e.g., Ω in Equation 57) is proportional to the integral of the squared beam power over angular area (Ω_{PP} ; equation 53). This contrasts the standard beam area (Ω_P ; equation 54) that is used to relate flux density to a brightness temperature. Since Equation 57 relates a measured visibility in units of brightness temperature to $P(\mathbf{k})$, a factor of Ω_P^2 has already been applied to convert Jy to mK. In this case, Ω indicates the remaining factor of Ω_{PP} , which for PAPER is 0.31 sr.

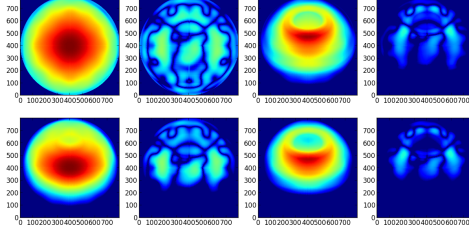


FIG. 6.—

Minimizing cross-contamination between Stokes terms in interferometric polarization measurements 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 (????). 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 ?, 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 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 polarization beams. However, the success that can be achieved with this technique depends strongly on an array's UV sam-

pling 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 (??).

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 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. It should not surprise the reader that what we have just described — a convolution kernel acting along a time series of samples from a single baseline — is an alternate description of fringe-rate filtering. Through the correct choice of fringe-rate filter, it is possible to improve the match between the XX and YY polarization beams, and 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 show in §?? that it nonetheless represents a substantial improvement over the naive summing of XX and YY visibility measurements.

6.5. 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 ?.

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 (????).

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 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 SNR weighting in a fringe-rate filter 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.

7. CONCLUSION

In this paper, we have revisited the concept of filtering the visibility time-series measured by an interferometric baseline that was presented in ?. 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 map-making, 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 ? and ?, 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 ?, 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; ?).

8. ACKNOWLEDGMENT

It gives us great pleasure to thank James Aguirre, David Moore, Danny Jacobs, Miguel Morales, and Jonathan Pober for helpful discussions. This research was supported by the National Science Foundation, award #1129258.