FRINGE-RATE FILTERING

Aaron R. Parsons[1,2], Adrian Liu[1], Zaki S. Ali[1],
Draft version January 8, 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 Reionization[3]. At the relevant redshifts ($z \sim 6$ to 20), theoretical models suggest that this cosmological signal will be faint, on the order of $1\,\text{mK}$ in brightness temperature, while a typical instrument's system temperature is typically $\sim 100\,\text{K}$. 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 interferometric 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 mapmaking. Delay/DDR paper. Also look at Andre's stuff.

- Outline the paper.

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

$$\tilde{V}(\tau) = \int W(\nu) \cdot S(\nu) \cdot V(\nu) e^{-2\pi i \nu \tau} \, d\nu, \qquad (1)$$

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

$$V_\nu(t) = \int I_\nu(\hat{s}, t) \cdot A_\nu(\hat{s}) \cdot e^{-2\pi i \vec{b} \cdot \hat{s}/c}, \qquad (2)$$

where I is the specific intensity of the sky in the direction of the topocentric unit vector $\hat{s}$, A is the geometric mean of the primary beam power patterns of the constituent antennas, $\vec{b}$ is the baseline vector separating the two antennas in question, and ν is the spectral frequency. Interferometric responses (visibilities) for ground-based instruments are generally a function of time, t, even for a static sky because the Earth's rotation makes the expression of I in topocentric coordinates a time-dependent quantity.

Assuming that antenna positions are fixed to the Earth, the fringe pattern of an interferometer — the variation in phase response across the sky — is fixed in topocentric coordinates. Hence, time variation in $V_\nu(t)$ can be regarded as resulting from the motion of spatial structures in I through the fringe pattern and primary beam response pattern. Moreover, for interferometers such as PAPER, CHIME, and HERA that do not point, but instead drift-scan the sky as it rotates by, the primary beam response pattern and the phase response pattern are locked to one another, and may together be considered the fringe pattern that gives rise to time variation in $V_\nu(t)$ as the Earth rotates.

The rate at which angular structure in I moves through a fringe pattern depends both on 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 $\delta = 0°$ to zero at $\delta = -90°$, and can even become negative for emission from the far side of the celestial pole.

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 (Bhatnagar et al. 2008; Morales & Matejek 2009b) and the w component of the baseline (Cornwell et al. 2003). As noted in Sullivan et al. (2012) and Morales & Matejek (2009b), 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
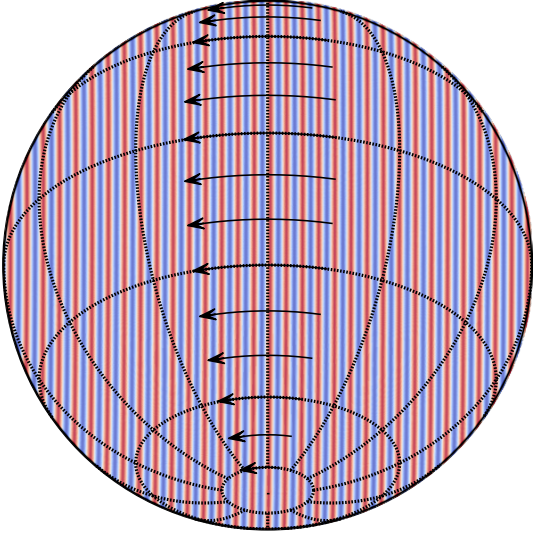
Fig. 1.— The fringe pattern at 150 MHz of a 30-m east-west baseline, overlaid with arrows indicating the distance traversed by sources at various declinations over a two-hour span centered at transit. In a fixed time interval, sources near $\delta = 0°$ traverse more fringe periods than sources nearer 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.
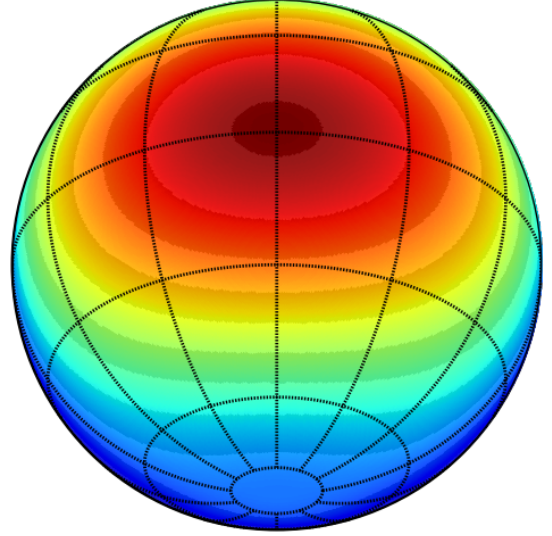
Fig. 2.— Fringe rate as a function of sky position, corresponding to the fringe pattern illustrated in Figure 1. Fringe rates peak at XXX mHz at $\delta = 0°$, hit zero at the south celestial pole, and become negative on the far side of the pole. Grey shading indicates the approximate angular regions that correspond to alternating fringe-rate bins, assuming a fringe-rate transform taken over a two-hour time series.

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 Hazelton et al. (2013) by applying a carefully tailored fringe-rate filter to each time series of visibility spectra.

(Parsons & Backer 2009), (Shaw et al. 2013b),

3. COMBINING TIME-ORDERED DATA IN AN OPTIMAL WAY FOR MAPMAKING

[XXX: Say somewhere that unless otherwise stated, considering a drift-scan telescope.] [XXX: Don't forget to define hats as estimators, once and for all.] In this section, our goal is to examine how time-ordered visibilities from an interferometer should be best combined into information about the sky (such as an image-domain map). Contrary to expectations, we will find that it is suboptimal to pursue the traditional, straightforward approach of averaging data in time, in the sense that such a procedure gives rise to larger-than-necessary error bars. Instead, employing an unbiased, minimum-variance prescription naturally yields the technique of *fringe-rate filtering*, the subject of this paper.

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

$$\mathbf{v} = \mathbf{A}\mathbf{x} + \mathbf{n}, \qquad (3)$$

where $\mathbf{n}$ is a noise vector. Note that in this general form, Equation (3) 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 (Tegmark 1997; Morales & Matejek 2009a)

$$\hat{\mathbf{x}} = \left[\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}\right]^{-1} \mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v}, \qquad (4)$$

where $\mathbf{N}$ 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. Regardless of the basis we work in, this estimator is unbiased, i.e.

$$\langle \hat{\mathbf{x}} \rangle = \mathbf{x}, \qquad (5)$$

and has minimum variance, which is given by

$$\boldsymbol{\Sigma} \equiv \langle (\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^\dagger \rangle = \left[\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}\right]^{-1}. \qquad (6)$$

The estimator given by Equation (4) can also be proved 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 $\mathbf{v}$.

[XXX: add that the unbiased condition holds only when the matrix is invertible.]

In principle, Equation (4) 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. (2013a) 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 $T(\hat{\mathbf{r}})$:

$$V_b(t) = \int B(\hat{\mathbf{r}}, t) T(\hat{\mathbf{r}}) \exp\left[-i2\pi\left(\frac{b_y}{\lambda}\cos\eta\cos\theta\right)\right]$$
$$\times \exp\left[-i2\pi\left(\frac{b_0}{\lambda}\sin\theta\sin(\varphi-\omega_\oplus t)\right)\right] d\Omega + n(t), \quad (7)$$

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, $\omega_\oplus$ 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 other words, Equation (7) 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 $B(\hat{\mathbf{r}}, t) \equiv \delta(\theta - \pi/2) B_\varphi(\varphi - \omega_\oplus t)$. Plugging these into restrictions into our equation, we obtain

$$V_b(t) = \int B_\varphi(\varphi - \omega_\oplus t) T\left(\theta = \frac{\pi}{2}, \varphi\right)$$
$$\times \exp\left[-i2\pi\frac{b_x}{\lambda}\sin(\varphi - \omega_\oplus t)\right] d\varphi + n(t). \quad (8)$$

Now suppose we have a large number of different baselines, so that we have multiple copies of this equation, representing a time series for each baseline. Grouping these time series into a vector gives the continuous version of $\mathbf{v}$. We can similarly identify $n(t)$ and $T(\theta = \pi/2, \varphi)$ as the continuous versions of $\mathbf{n}$ and $\mathbf{x}$ respectively, with the rest of the Equation (8)'s integrand

as the continuous version of $\mathbf{A}$. We can model the noise covariance between baselines b and b', at times t and t' as

$$N_{bb'}(t, t') = \sigma^2 \delta_{bb'} \delta(t - t'), \quad (9)$$

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 (4) combines information from different times, we need only evaluate $\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v}$, for the subsequent application of $\left[\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}\right]^{-1}$ only serves to undo the effects of an instrument's synthesized beam once the time series have already been combined. In our toy model, we have

$$\left(\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v}\right)_\varphi = \sum_b \int \frac{dt}{\sigma^2} B_\varphi(\varphi - \omega_\oplus t) e^{i2\pi\frac{b_x}{\lambda}\sin(\varphi - \omega_\oplus t)} V_b(t),$$
$$(10)$$

where the φ 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 (10) 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 (10) 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 (7). 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

$$V_b(t) = \int B(\hat{\mathbf{r}})T(\hat{\mathbf{r}}) \exp\left[-i2\pi\left(\frac{b_0}{\lambda}\sin(\varphi - \omega_\oplus t)\right)\right]$$
$$\times \exp\left[-i2\pi\left(\frac{b_y}{\lambda}\cos\eta\sin\theta\right) + i\psi(t)\right] d\Omega + n(t), \quad (11)$$

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

$$\left(\mathbf{A}^\dagger\mathbf{N}^{-1}\mathbf{v}\right)^{\text{flat}}_{(\theta,\varphi)} = \frac{B(\theta,\varphi)}{\sigma^2}e^{i2\pi\frac{b_y}{\lambda}\cos\eta\sin\theta}$$
$$\times \sum_b \int dt\, e^{i2\pi\frac{b_x}{\lambda}\sin(\varphi - \omega_\oplus t) - i\psi(t)} V_b(t). \quad (12)$$

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 $\sin(\varphi - \omega_\oplus t)$ 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 $\sin(\varphi - \omega_\oplus t)$ 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

$$\left(\mathbf{A}^\dagger\mathbf{N}^{-1}\mathbf{v}\right)^{\text{flat,short}}_{(\theta,\varphi)} = \frac{B(\theta,\varphi)}{\sigma^2}e^{i2\pi\left(\frac{b_y}{\lambda}\cos\eta\sin\theta + \frac{b_x}{\lambda}\sin\varphi\right)}$$
$$\times \sum_b \int dt\, V_b(t), \quad (13)$$

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 (7). From our toy example [Equation (10)], 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

$$\widetilde{V}_b(f) \equiv \frac{1}{T_\oplus} \int_{-T_\oplus/2}^{T_\oplus/2} dt \exp\left(-2\pi i f t\right) V_b(t), \quad (14)$$

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

$$\widetilde{V}_b(f_n) = \int d\Omega\, T(\hat{\mathbf{r}})e^{-i2\pi\frac{b_y}{\lambda}\cos\eta\cos\theta}$$
$$\times \int_{-\frac{T_\oplus}{2}}^{\frac{T_\oplus}{2}} \frac{dt}{T_\oplus} B(\hat{\mathbf{r}},t)e^{-i\frac{2\pi nt}{T_\oplus} + i\frac{2\pi b_0}{\lambda}\sin\theta\sin(\omega_\oplus t - \varphi)}, \quad (15)$$

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

$$B(\hat{\mathbf{r}},t) \equiv B_\theta(\theta)B_\varphi(\varphi - \omega_\oplus t), \quad (16)$$

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

$$B(\hat{\mathbf{r}},t) = B_\theta(\theta)\sum_q \widetilde{B}_q e^{-iq\varphi}e^{iq\omega_\oplus t}, \quad (17)$$

where $\widetilde{B}_q \equiv \int \frac{d\varphi}{2\pi}B_\varphi(\varphi)e^{iq\varphi}$ is the qth Fourier coefficient. Plugging this into Equation (15) and making the substitution $\psi \equiv \omega_\oplus t - \varphi$, one obtains

$$\widetilde{V}_b(f_n) = \int d\Omega\, T(\hat{\mathbf{r}})B_\theta(\theta)e^{-i2\pi\frac{b_y}{\lambda}\cos\eta\cos\theta}$$
$$\times \sum_q \frac{\widetilde{B}_q e^{-in\varphi}}{2\pi}\int_{-\pi-\varphi}^{\pi+\varphi} d\psi\, e^{i(q-n)\psi + i\frac{2\pi b_0}{\lambda}\sin\theta\sin\psi}. \quad (18)$$

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:

$$\widetilde{V}_b(f_n) = \int \frac{d\Omega}{2\pi}\, T(\hat{\mathbf{r}})B_\theta(\theta)e^{-i2\pi\frac{b_y}{\lambda}\cos\eta\cos\theta}e^{-in\varphi}$$
$$\times \sum_q \widetilde{B}_q J_{n-q}\left(\frac{2\pi b_0}{\lambda}\sin\theta\right). \quad (19)$$

Several features are of note here. For wide primary beams, $\widetilde{B}_q$ is sharply peaked around $q = 0$, so the terms following the sum over q essentially amount to $J_n(2\pi b_0\sin\theta/\lambda)$. 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,

$$J_n\left(\frac{2\pi b_0\sin\theta}{\lambda}\right) \approx \frac{1}{n!}\left(\frac{\pi b_0\sin\theta}{\lambda}\right)^n, \quad (20)$$

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

$$\left(\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v}\right)_{\theta,\varphi} = \frac{B_\theta^*(\theta)\sin\theta}{2\pi\sigma^2} \sum_{b,n} e^{in\varphi} e^{i2\pi\frac{b_y}{\lambda}\cos\eta\cos\theta}$$
$$\times \sum_q \widetilde{B}_q^* J_{n-q}\left(\frac{2\pi b_0}{\lambda}\sin\theta\right) \widetilde{V}_b(f_n). (21)$$

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 $\sum_q \widetilde{B}_q^* J_{n-q}\left(\frac{2\pi b_0}{\lambda}\sin\theta\right)$, 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.

[XXX: Talk about how we're only going to use these to guide intuition, and more numerical stuff later in the paper.]

3.4. *The possibility of deconvolution*

If one has a sufficient number of baselines (roughly speaking, a "fully-filled *uv*-plane"), then the application of $\left[\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}\right]^{-1}$ should perfectly deconvolve instrumental beams in our estimator of the sky [Equation (4)] and leave a delta-function effective beam. Indeed, this is precisely what Equation (5) states. However, in practice the matrix $\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}$ may not be invertible, and the best that one can do is to produce an estimate that is convolved (or only partially deconvolved) with an effective beam.

The question of whether one has a sufficient density of baselines to deconvolve out the beam is most easily answered by writing $\mathbf{A}$ and $\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}$ in convenient bases. We first rewrite Equation (19) so that it expresses a given baseline's fringe-rate measurements in terms of azimuthal Fourier modes $a_m(\theta)$ of the sky $T(\theta,\varphi) = \sum_m a_m(\theta) e^{im\varphi}$. The measurement equation becomes

$$\widetilde{V}_b(f_n) = \sum_m \int d\theta\, B_\theta(\theta)\sin\theta\, e^{-i2\pi\frac{b_y}{\lambda}\cos\eta\cos\theta}$$
$$\times \sum_q \widetilde{B}_q J_{n-q}\left(\frac{2\pi b_0}{\lambda}\sin\theta\right)\delta_{mn} a_m(\theta). (22)$$

The Kronecker delta function appears because for a drift-scan telescope, the time coordinate uniformly maps to the azimuthal angle φ (or equivalently, the Fourier index m uniformly maps to fringe-rate bin), as emphasized in Shaw et al. (2013a). From the expression we can immediately read off the form of $\mathbf{A}$ in this basis, and inserting it into $\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}$ results in

$$\left(\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}\right)_{m,\theta;m'\theta'} = \frac{\delta_{mm'}}{\sigma^2}\sum_b \lambda_b^m h_b(\theta,m) h_b^*(\theta',m'), (23)$$

where we have once again taken $\mathbf{N}$ to be diagonal, with

entries σ^2, and have made the definitions

$$g_b(\theta,m) \equiv B_\theta^*(\theta)\sin\theta \exp\left(i2\pi\frac{b_y}{\lambda}\cos\eta\cos\theta\right)$$
$$\times \sum_q \widetilde{B}_q^* J_{m-q}\left(\frac{2\pi b_0}{\lambda}\sin\theta\right) (24)$$

$$h_b(\theta,m) \equiv \frac{g_b(\theta,m)}{\sqrt{\lambda_b^m}} (25)$$

$$\lambda_b^m \equiv \int_0^\pi |g_b(\theta,m)|^2 d\theta. (26)$$

Whether or not $\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}$ is invertible for a given baseline distribution is manifestly clear from the form of Equation (23). Imagining $\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}$ as a discrete matrix with each side cycling through polar angle θ more rapidly than azimuthal Fourier mode number m, we see that our matrix is block diagonal. Invertibility then hinges on satisfying two requirements. First, none of the blocks can be identically zero. While in principle this can never happen, in practice this will be the case for sufficiently high m, since m can be thought of as a fringe-rate bin index, and very high fringe-rates are physically unattainable unless one has very long baselines. Put another way, the arguments in the previous section for a high fringe-rate cutoff ensure that λ_b^m gets very small for large m. For invertibility, then, one needs to either have a cutoff in m (which then defines a resolution scale in the φ direction), or sufficiently long baselines.

The second requirement for invertibility comes from the fact that a block diagonal matrix's inverse is again block diagonal, with each block replaced by its inverse. We therefore require the invertibility of each block, which is obtained by setting $m = m'$ in Equation (23), and is indexed by (θ,θ'). Now, notice that were we to discretize our expression, it would have precisely the form of an eigenvalue decomposition $\sum_i \lambda_i \mathbf{w}_i \mathbf{w}_i^\dagger$ of an arbitrary matrix, where λ_i is the matrix's i^{th} eigenvalue and $\mathbf{w}_i$ is the ith eigenvector (satisfying the normalization condition $\mathbf{w}_i^\dagger \cdot \mathbf{w}_i = 1$). To ensure invertibility, the number of eigenvectors must equal the side length of the matrix. For us, this requirement corresponds to having the number of unique baselines be equal to or greater than the number of angular resolution elements in the θ direction. Intuitively, the same declinations map to different fringe-rates for different baselines, and thus with enough baselines, one is able to take the appropriate linear combination of different baselines to select a specific declination and nothing else. Importantly, this is conceptually separate from having a variety of baseline lengths in the north-south direction and performing synthesis imaging in the declination direction. An examination of the relevant expressions shows that our argument holds even if all baselines are of a purely east-west orientation; the ability to localize in θ can be acquired by combining fringe rates appropriately.

Having derived an expression for $\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}$, we can also compute the error covariance $\mathbf{\Sigma}$ of our maps. If we have enough baselines for invertibility, Equation (6) says that

$$\mathbf{\Sigma}_{m,\theta;m'\theta'} = \left[\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}\right]^{-1}_{m,\theta;m'\theta'}$$
$$= \sigma^2 \delta_{mm'} \sum_b \frac{1}{\lambda_b^m} h_b(\theta,m) h_b^*(\theta',m'). (27)$$

As expected, azimuthal modes corresponding to low fringe-rate bins can be measured with high signal-to-noise.

4. COMBINING TIME-ORDERED DATA WITHOUT MAPMAKING

Until now, our discussion of fringe-rate filtering has centered around the mapmaking problem. However, our intuitive understanding of fringe-rate filtering—that it differentiates instrumental noise from celestial signals by identifying high fringe-rates as being noise-dominated—holds much more generally. In this section, we show how fringe-rate filtering can be used a de-noising tool for visibilities, bypassing the image domain entirely. This allows one to take advantage an important benefit of mapmaking, namely, the reduction of instrumental noise through the coherent combination of time-ordered data, while avoiding some of its drawbacks such as the introduction of curved-sky or gridding artifacts. Moreover, for many applications (such as power spectrum estimation), it is more natural to work with interferometric visibilities directly, provided they can be de-noised.

To de-noise our visibilities, suppose (for a brief moment) that one did in fact produce an estimator $\hat{\mathbf{x}}$ of the sky in the image domain. One could then transform this estimator back into time-ordered visibilities $\mathbf{v}_{\text{filt}}$ by simply computing $\mathbf{v}_{\text{filt}} \equiv \mathbf{A}\hat{\mathbf{x}}$. Such a set of visibilities would represent de-noised visibilities since Equation (4) is essentially a best-fit solution based on noisy data, which means that $\mathbf{A}\hat{\mathbf{x}}$ is simply the predicted measurements from the best-fit model. Now, if these predicted measurements were actually generated by first forming a map, one would still incur all the difficulties in imaging that one had hoped to avoid. To deal with this issue, we note that the relationship between the original visibilities and the de-noised ones are given by

$$\mathbf{v}_{\text{filt}} = \mathbf{A} \left[\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A} \right]^{-1} \mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v} \equiv \mathbf{W}\mathbf{v}. \qquad (28)$$

The matrix $\mathbf{W}$ can be thought of as a de-noising filter, and if it can be computed analytically we will have bypassed imaging entirely.

Fortunately, an analytic treatment is possible if one makes the observation that the matrices "internal" to $\mathbf{W}$ can be expressed in any basis without affecting the final result. Expressing the $\mathbf{A}$ matrix in a fringe-rate/azimuthal Fourier mode/polar angle/baseline basis, i.e. the basis implied by Equation (22), a little algebra reveals that $\mathbf{W}$ takes the form

$$\mathbf{W}_{n,b;n',b'} = \delta_{nn'} \sum_{b''} \frac{c_{bb'}(n)c^*_{b'b''}(n)}{(\lambda^n_{b''})^2}, \qquad (29)$$

where we have again assumed white noise, and have defined

$$c_{bb'}(n) \equiv \int_0^\pi g^*_b(\theta, n) g_{b'}(\theta, n) d\theta. \qquad (30)$$

[XXX: may need to be clearer about what basis the final result is in.] The various matrix elements of $\mathbf{W}$ provide the linear combinations of data from various baselines and fringe-rate bins that should be taken to de-noise the data. Different fringe-rate bins do not mix, whereas the data from different baselines are combined to de-noise a given fringe-rate bin. Intuitively, a given fringe-rate

is measured by different baselines with different levels of sensitivity (for instance, longer baselines more "naturally" sample higher fringe-rates), and $\mathbf{W}$ takes this into account to combine data in an optimal way. Note that with our analytical prescription, all relevant quantities can be precomputed given a beam model, and no imaging is required to combine information from different baselines.

4.1. De-noising visibilities from a single baseline

While Equation (29) is useful for combining information from different baselines, it is unable to de-noise data from a single baseline. Indeed, one can see that $\mathbf{W}$ reduces to the identity matrix in the single baseline case. To understand this, recall that for a drift-scan interferometer there is a one-to-one, monotonic correspondence fringe-rate and azimuthal Fourier number. Thus, while high fringe-rate bins are quite *likely* to be noise-dominated, if the sky contained a strong signal in a high azimuthal Fourier mode, it would in principle be *possible* for the bin to be signal-dominated. Put another way, high fringe-rate bins are not necessarily noise-dominated and discardable, since they each access unique information about the sky.

In order to de-noise visibilities from a single baseline, it is necessary to place a prior on the nature of the sky signal. For instance, suppose one had reliable prior knowledge of the sky in the form of a signal covariance matrix $\mathbf{S} \equiv \langle \mathbf{x}\mathbf{x} \rangle$. One could then form a Wiener-filtered estimate of $\mathbf{x}$:

$$\hat{\mathbf{x}}_{\text{Wiener}} = \left[\mathbf{S}^{-1} + \mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A} \right]^{-1} \mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v}, \qquad (31)$$

which by construction minimizes $\langle |\hat{\mathbf{x}}_{\text{Wiener}} - \mathbf{x}|^2 \rangle$, the expected error (Tegmark 1997). Like before, we can then proceed to multiply by $\mathbf{A}$ to arrive at a set of filtered visibilities:

$$\mathbf{v}_{\text{Wiener}} \equiv \mathbf{A}\hat{\mathbf{x}}_{\text{Wiener}}. \qquad (32)$$

As an example, suppose the sky consisted of a continuum of unresolved and unclustered point sources. The signal covariance matrix would then be of the form $\mathbf{S} = \sigma_s^2 \mathbf{I}$, where σ_s is some root-mean-square signal magnitude. In the low signal-to-noise limit[4], where the second term in Equation 31's inverse term can be neglected, the de-noising filter $\mathbf{W}^{\text{Wiener}}$ becomes

$$\mathbf{W}^{\text{Wiener}}_{nn'} = \delta_{nn'} \frac{\sigma_s^2}{\sigma_n^2} \int_0^\pi |g_b(\theta, n)|^2 d\theta. \qquad (33)$$

Each element of this diagonal matrix specifies a weighting for a fringe-rate bin. The integral quantifies the response of the baseline to the particular fringe-rate bin, and when multiplied by σ_s^2, gives the signal in the bin. (Note that σ_s and σ_n have different units, so that $\mathbf{W}^{\text{Wiener}}$ is dimensionless). With σ_n^2 in the denominator, the Wiener filter enacts a signal-to-noise weighting in fringe-rate space. Given that $|g_b(\theta, n)|^2$ decays rapidly at high n (as discussed in previous sections), the Wiener filter does precisely what we intuitively expect it to do: it down-weights high fringe-rates that are likely noise-dominated. Our placement of a flat signal prior essentially precluded

---

[4] What turns out to be a signal-to-noise weighting in our special case simply becomes a signal-over-signal-plus-noise weighting in the general case.

the possibility that a strong signal in high azimuthal Fourier modes could result in a signal-dominated measurement at high fringe-rate bins, allowing us to safely downweight those bins without sacrificing signal.

Care must be taken, however, when using a set of visibilities that have been Wiener-filtered. Because fringe-rate filtering is effectively a way to enact time-integration, the resulting time-series of filtered visibilities will possess correlated errors and will no longer be independent. This is expected and desirable in the sense that it is what allows the Wiener filter to minimize errors. However, this comes at a cost. Recall that fringe-rates map directly to azimuthal Fourier mode numbers. Thus, if one were to use the Wiener-filtered data to measure (say) a power spectrum, the result would be biased low, as Equation (33) gives fringe-rate weights that are all less than one. Instead, we envision the Wiener-filtered visibilities as one where the lower noise levels allow systematics to be more clearly identified.[5]

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 $\hat{p}_\alpha$ for the αth band of the power spectrum:

$$\hat{p}_\alpha = N_\alpha \hat{\mathbf{x}}^\dagger \mathbf{C}^{-1} \mathbf{Q}_\alpha \mathbf{C}^{-1} \hat{\mathbf{x}}, \qquad (34)$$

where $\mathbf{C} \equiv \langle \hat{\mathbf{x}} \hat{\mathbf{x}}^\dagger \rangle = \mathbf{S} + \mathbf{\Sigma}$ is the *total* (signal plus noise) covariance matrix of the sky map, and N_α is a normalization factor, given by[6]

$$N_\alpha = \text{tr} \left[\mathbf{C}^{-1} \mathbf{Q}_\alpha \mathbf{C}^{-1} \mathbf{Q}_\alpha \right]. \qquad (35)$$

---

[5] It is admittedly possible to construct an unbiased power spectrum estimator from Wiener-filtered visibilities: one simply renormalizes any fringe-rate bins that were biased low by the Wiener-filter. The only subtlety is that this renormalization cannot be performed individually on different fringe-rate bins, since the diagonal nature of $\mathbf{W}^{\text{Wiener}}$ would result in the identity matrix. Instead, we must make the observation that a typical power spectrum estimator takes advantage of statistical isotropy to sum together the power estimates of individual azimuthal modes (or equivalently, fringe rates) to form power estimates on a particular angular scale. As long as the final estimate on each angular scale is properly normalized, the result will be unbiased. However, it will not be *optimal* (unlike the estimator that we will derive in Section 5) in the sense that the error bars will not be as small as can be possibly achieved by the data set.

[6] 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.

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

$$\mathbf{Q}_\alpha \equiv \frac{\partial \mathbf{C}}{\partial p_\alpha}. \qquad (36)$$

The inverse-covariance weighting enacted by $\mathbf{C}^{-1}$ makes Equation (34) 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 (4), which states that $\hat{\mathbf{x}} = \left[\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A} \right]^{-1} \mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v}$. 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 (34), we obtain

$$\hat{p}_\alpha = N_\alpha \hat{\mathbf{v}}^\dagger \mathbf{N}^{-1} \mathbf{A} \mathbf{Q}_\alpha \mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v} \qquad (37)$$

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

$$\mathbf{C}_{\hat{\mathbf{r}}, \hat{\mathbf{r}}'} = [\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}]^{-1} + \sum_{\ell m} Y_{\ell m}(\hat{\mathbf{r}}) Y_{\ell m}^*(\hat{\mathbf{r}}') C_\ell, \qquad (38)$$

and thus

$$\mathbf{Q}_\ell \equiv \frac{\partial \mathbf{C}}{\partial C_\ell} = \sum_m Y_{\ell m}(\hat{\mathbf{r}}) Y_{\ell m}^*(\hat{\mathbf{r}}'), \qquad (39)$$

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 $\mathbf{Q}_\ell$, even though this was unnecessary for $\mathbf{C}^{-1}$.

Substituting Equations (21) and (39) into Equation (37), we obtain

$$\hat{C}_\ell \propto \sum_{m=-\infty}^{\infty} \left| \sum_b w_{\ell m}^b \widetilde{V}_b(f_m) \right|^2, \qquad (40)$$

where

$$w_{\ell m}^b \equiv \begin{cases} 2\pi \int d(\cos\theta) N_\ell^m P_\ell^m(\cos\theta) g_b(\theta, m), & \text{for } -\ell \le m \le \ell \\ 0, & \text{otherwise,} \end{cases} \qquad (41)$$

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

$$N_\ell^m = \sqrt{\frac{2\ell + 1}{4\pi} \frac{(\ell - m)!}{(\ell + m)!}} \qquad (42)$$

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:

$$\widehat{C}_\ell \propto \int_{-T_\oplus/2}^{T_\oplus/2} \frac{dt}{T_\oplus} \left| \sum_b V_{\ell b}^{\mathrm{filt}}(t) \right|^2, \qquad (43)$$

where

$$V_{\ell b}^{\mathrm{filt}}(t) \equiv \sum_m \exp\left[i\frac{2\pi mt}{T_\oplus} \right] w_{\ell m}^b \widetilde{V}_b(f_m). \qquad (44)$$

Together, Equations 43 and 44 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 (40), 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 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 –¿ m-mode correspondence more throughout.]
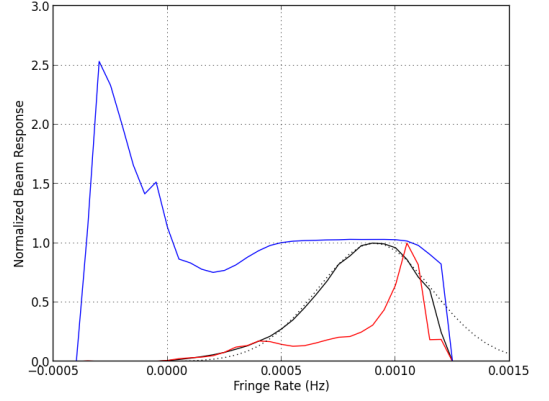
6. APPLICATIONS

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 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 response[7]. 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 illustrate the applications of fringe-rate filters in this section, we make use of simulated observations at 150 MHz with the fiducial instrument described in §**??**.

FIR filter. Analytic form for smoothness vs. frequency.

---

[7] It is worth noting that, as described in Parsons & Backer (2009), 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 limit its application in this context for 21cm cosmological applications
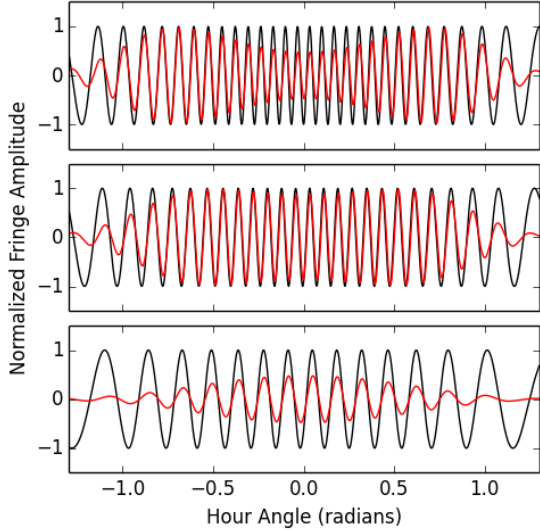
FIG. 4.— Fringe amplitude before (black) and after (red) the application of the fringe-rate filter described in §??. From top to bottom, the panels illustrate the real component of the visibility measured by a 30-m east-west baseline deployed at a latitude of $-30°$ for point sources passing through the fringe pattern at declinations of $0°$, $-30°$, and $-60°$, respectively. In this simulation, antenna elements have isotropic primary beams.

Desireability of a compact footprint. Data gaps and edge issues.

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.

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 $\delta = 0°$, decreasing to 0 at $\delta = -90°$, and for an array such as PAPER deployed near -30° S latitude, reaching a minimum of $\approx -f_{\max}/2$ at $\delta = -60°$ on the far side 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 Morales & Matejek (2009b), and have many features in common with m-mode synthesis described in Shaw et al. (2013b).

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

$$f_{\max} = \frac{|\mathbf{b}_{\mathrm{eq}}|}{c} \omega_\oplus \nu, \tag{45}$$

where $f_{\max}$ is the maximum fringe rate, $\mathbf{b}_{\mathrm{eq}}$ is the baseline vector projected parallel to the equatorial plane, c is the speed of light, $\omega_\oplus$ 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.

$$\widehat{P}(\mathbf{k}_{t\tau}) = \left(\frac{\lambda^2}{2k_{\mathrm{B}}}\right)^2 \frac{X^2 Y}{\Omega B} \left\langle \tilde{V}_i(\tau,t)\tilde{V}_j^*(\tau,t) \right\rangle_{i<j}, \tag{46}$$

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 $\frac{h^{-3}\ \mathrm{Mpc}^3}{\mathrm{sr \cdot Hz}}$, Ω is the angular area[8], B is the bandwidth, $\langle \ldots \rangle_{i<j}$ indicates the ensemble average over instantaneously redundant baseline measurements indexed by i, j, and $\tilde{V}(\tau,t)$

---

[8] As described in detail in Appendix 9, the angular area used to normalize high-redshift 21cm power spectrum measurements (e.g., Ω in Equation 46) is proportional to the integral of the squared beam power over angular area (Ω_{PP}; equation 55). This contrasts the standard beam area (Ω_{P}; equation 56) that is used to relate flux density to a brightness temperature. Since Equation 46 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.
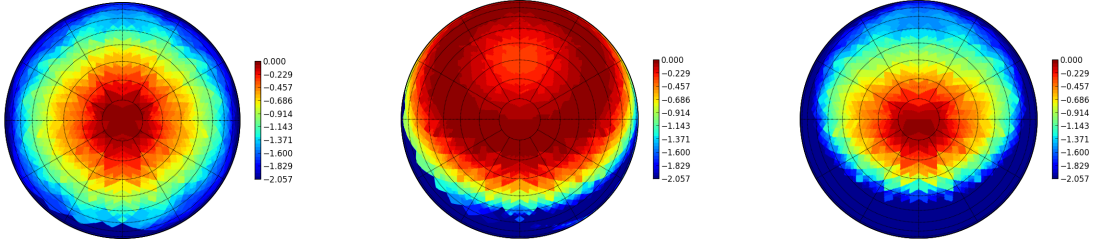
FIG. 5.— The effective primary beam response of a baseline, as determined from the simulations in §6.2. Panels indicate reconstructions of PAPER's model beam response used in the simulation (left), the beam weighting that results from the application of a fringe-rate filter weighted to optimize SNR for 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.

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.

$$\widehat{\Delta}_{21}^2(k) = \frac{k^3}{2\pi^2} \left\langle \widehat{P}(\mathbf{k}_{t\tau}) \right\rangle_{|\mathbf{k}_{t\tau}|=k}, \qquad (47)$$

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 mK^2 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*

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 measure-

ment $V_{\mathrm{I}} \equiv (V_{\mathrm{XX}} + V_{\mathrm{YY}})/2$ 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 $V_{\mathrm{Q}} \equiv (V_{\mathrm{XX}} - V_{\mathrm{YY}})/2$ 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 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 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
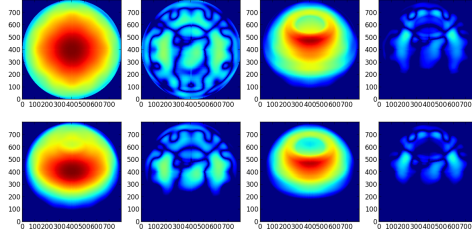
Fig. 6.—

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 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; **?**; **?**; **?**).

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 desireable 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 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 **?** 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

Similar geometric limits apply to the variation of visibilities in the time dimension. As described in Equation 7 of Parsons & Backer (2009), the rate of change of the geometric delay versus time — that is, delay-rate — is given by

$$\frac{d\tau_g}{dt} = -\omega_\oplus \left(\frac{b_x}{c} \sin H + \frac{b_y}{c} \cos H \right) \cos \delta, \qquad (48)$$

where $\mathbf{b} = (b_x, b_y, b_z)$ is the baseline vector expressed in equatorial coordinates, $\omega_\oplus$ 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

$$-\omega_\oplus \frac{|b_y|}{c} \leq \frac{d\tau_g}{dt} \leq \omega_\oplus \frac{|b_y|}{c}. \qquad (49)$$

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 Parsons & Backer 2009) for a model visibility consisting of a single point source located at the point of maximum delay-rate, which gives us

$$\tilde{V}(\nu, f) \approx \tilde{A}(\nu, f) * \tilde{S}(\nu) * \int e^{2\pi i \omega_\oplus \frac{b_y \nu}{c} t} e^{-2\pi i f t} dt$$

$$\approx \tilde{A}(\nu, f) * \tilde{S}(\nu) * \delta_D(\frac{b_y}{c} \omega_\oplus \nu - f), \qquad (50)$$

where f is the fringe-rate of the source[9], $\tilde{A}(\nu, f)$ 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 45.

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 band (i.e. evaluating Equation 45 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.

9. ON CALCULATING BEAM AREAS FOR NORMALIZING POWER-SPECTRUM MEASUREMENTS

We begin by examining the integrated volume, $\mathbb{V}$, used to normalize the 3D Fourier transform in Equation 3 of

---

[9] Delay rate is equivalent to the frequency-integrated fringe rate.

P12a. We express this volume in observing coordinates as

$$\mathbb{V} = \Omega B \cdot X^2 Y, \tag{51}$$

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

$$\tilde{V}^2(u,v,\eta) = \left(\frac{2k_B}{\lambda^2}\right)^2 \left[\int dl\ dm\ d\nu A(l,m,\nu)T(l,m,\nu)e^{-2\pi i(ul+vm+\eta\nu)}\right] \times$$

$$\left[\int dl'\ dm'\ d\nu' A^*(l',m',\nu')T^*(l',m',\nu')e^{2\pi i(ul'+vm'+\eta\nu')}\right] \tag{52}$$

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

$$\tilde{V}_{21}^2(u,v,\eta) \approx \left(\frac{2k_B}{\lambda^2}\right)^2 \frac{B}{X^2 Y} \widehat{P}(\mathbf{k}) \int dl\ dm\ |A(l,m)|^2. \tag{53}$$

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):

$$\tilde{V}_{21}^2(u,v,\eta) \approx \left(\frac{2k_B}{\lambda^2}\right)^2 \frac{\Omega B}{X^2 Y} \widehat{P}_{21}(\mathbf{k}). \tag{54}$$

As this shows, the relevant beam area in Equation 54 is

the power-square beam, Ω_{PP}, given by

$$\Omega_{PP} \equiv \int dl\ dm\ |A(l,m)|^2. \tag{55}$$

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

$$\Omega_P \equiv \int dl\ dm A(l,m), \tag{56}$$

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)

$$\Omega' \equiv \Omega_P^2 / \Omega_{PP} \tag{57}$$

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 a 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 46) to $\widehat{P}(\mathbf{k})$, 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.

REFERENCES

Bhatnagar, S., Cornwell, T. J., Golap, K., & Uson, J. M. 2008, A&A, 487, 419
Cornwell, T. J., Golap, K., & Bhatnagar, S. 2003, W-Projection: A New Algorithm for Non-Coplanar Baselines, EVLA Memo 67
Hazelton, B. J., Morales, M. F., & Sullivan, I. S. 2013, ArXiv e-prints
Morales, M. F., & Matejek, M. 2009a, MNRAS, 400, 1814
—. 2009b, MNRAS, 400, 1814
Parsons, A. R., & Backer, D. C. 2009, AJ, 138, 219

Parsons, A. R., et al. 2010, AJ, 139, 1468
Shaw, J. R., Sigurdson, K., Pen, U.-L., Stebbins, A., & Sitwell, M. 2013a, ArXiv e-prints
—. 2013b, ArXiv e-prints
Sullivan, I. S., et al. 2012, ApJ, 759, 17
Tegmark, M. 1997, ApJ, 480, L87