Draft version June 27, 2013
Preprint typeset using LaTeX style emulateapj v. 03/07/07

FRINGE-RATE FILTERING

Aaron R. Parsons[1,2], Adrian Liu[1],
Draft version June 27, 2013

ABSTRACT

Subject headings:

1. INTRODUCTION

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. THE MAPMAKING PROBLEM: COMBINING TIME-ORDERED DATA IN AN OPTIMAL WAY

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

$$\mathbf{v} = \mathbf{A}\mathbf{x} + \mathbf{n}, \qquad (2)$$

where $\mathbf{n}$ is a noise vector. Note that in this general form, Equation (2) is not basis-specific. For example, while it is often useful to think of $\mathbf{x}$ as a vector containing a list of temperatures in a set of pixels on the sky (hence the variable name $N_{\rm pix}$), it is equally valid to employ another basis, such as spherical harmonics. Similarly, while we call $\mathbf{v}$ the time-ordered data, it need not be a time series, and in fact, 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)

$$\hat{\mathbf{x}} = \left[\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}\right]^{-1} \mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v}, \qquad (3)$$

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.

$$\langle \hat{\mathbf{x}} \rangle = \mathbf{x}, \qquad (4)$$

and has minimum variance, which is given by

$$\mathbf{\Sigma} \equiv \langle (\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^\dagger \rangle = \left[\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}\right]^{-1}. \qquad (5)$$

The estimator given by Equation (3) 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 (3) 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.

2.1. *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 $T(\hat{\mathbf{r}})$:

$$V_b(t) = \int d\Omega \, B(\hat{\mathbf{r}}, t) T(\hat{\mathbf{r}}) \exp\left[-i2\pi \left(\frac{b_0}{\lambda} \sin\theta \sin(\varphi - \omega_\oplus t) + \frac{b_y}{\lambda}\right.\right. \qquad (6)$$

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

[1] Astronomy Dept., U. California, Berkeley, CA
[2] Radio Astronomy Lab., U. California, Berkeley, CA
[3] Dept. of Physics and Astronomy, U. Pennsylvania, Philadelphia, PA
[8] School of Earth and Space Exploration, Arizona State U., Tempe, AZ

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 (6) 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

$$V_b(t) = \int d\varphi \, B_\varphi(\varphi - \omega_\oplus t) \exp\left[-i2\pi \frac{b_x}{\lambda} \sin(\varphi - \omega_\oplus t)\right] T\left(\theta = \frac{\pi}{2}, \varphi\right) n(t). \tag{7}$$

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 (7)'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

$$N_{bb'}(t, t') = \sigma^2 \delta_{bb'} \delta(t - t'), \tag{8}$$

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 (3) 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

$$\left(\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v}\right)_\varphi = \frac{1}{\sigma^2} \sum_b \int dt \, B_\varphi(\varphi - \omega_\oplus t) \, e^{i2\pi \frac{b_x}{\lambda} \sin(\varphi - \omega_\oplus t)} V_b(t), \tag{9}$$

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 (9) 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 2.3, demonstrating that its optimality is not a peculiarity of our current pedagogical example, but is instead applicable in all situations.

2.2. *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 (9) 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 (6). 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

$$V_b(t) = \int d\Omega \, B(\hat{\mathbf{r}}) T(\hat{\mathbf{r}}) \exp\left[-i2\pi \left(\frac{b_0}{\lambda} \sin(\varphi - \omega_\oplus t) + \frac{b_y}{\lambda} \cos\eta \sin\right.\right. \tag{10}$$

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

$$\left(\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v}\right)_{(\theta,\varphi)} \Bigg|_{\text{flat sky}} = \frac{B(\theta, \varphi)}{\sigma^2} e^{i2\pi \frac{b_y}{\lambda} \cos\eta \sin\theta} \sum_b \int dt \; e^{i2\pi \frac{b_x}{\lambda} \mathrm{s}} \tag{11}$$

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

$$\left(\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{v}\right)_{(\theta,\varphi)} \Bigg|_{\text{flat sky, short obs.}} = \frac{B(\theta, \varphi)}{\sigma^2} e^{i2\pi \left(\frac{b_y}{\lambda} \cos\eta \sin\theta + \frac{b_x}{\lambda} \sin\right.} \tag{12}$$

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.]

2.3. *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 (6). From our toy example [Equation (9)], 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

$$\widetilde{V}_b(f) \equiv \frac{1}{T_\oplus}\int_{-T_\oplus/2}^{T_\oplus/2} dt \exp\left(-2\pi i f t\right) V_b(t), \tag{13}$$

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

$$\widetilde{V}_b(f_n) = \int d\Omega\, T(\hat{\mathbf{r}}) e^{-i2\pi\frac{b_y}{\lambda}\cos\eta\cos\theta}\int_{-\frac{T_\oplus}{2}}^{\frac{T_\oplus}{2}}\frac{dt}{T_\oplus}\, B(\hat{\mathbf{r}},t)\exp\left[-i\frac{2\pi n t}{T_\oplus} + i\frac{2\pi b_0}{\lambda}\sin\theta\sin\left(\omega_\oplus t - \varphi\right)\right], \tag{14}$$

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

$$B(\hat{\mathbf{r}},t) \equiv B_\theta(\theta) B_\varphi(\varphi - \omega_\oplus t), \tag{15}$$

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

$$B(\hat{\mathbf{r}},t) = B_\theta(\theta)\sum_q \widetilde{B}_q e^{-iq\varphi} e^{iq\omega_\oplus t}, \tag{16}$$

where $\widetilde{B}_q \equiv \int \frac{d\varphi}{2\pi} B_\varphi(\varphi) e^{iq\varphi}$ is the q^{th} Fourier coefficient. Plugging this into Equation (14) and making the substitution $\psi \equiv \omega_\oplus t - \varphi$, one obtains

$$\widetilde{V}_b(f_n) = \int d\Omega\, T(\hat{\mathbf{r}}) B_\theta(\theta) e^{-i2\pi\frac{b_y}{\lambda}\cos\eta\cos\theta}\sum_q \frac{\widetilde{B}_q e^{-in\varphi}}{2\pi}\int_{-\pi-\varphi}^{\pi+\varphi} d\psi\, e^{i(n-q)\psi} e^{i\frac{2\pi b_0}{\lambda}\sin\theta\sin\psi}. \tag{17}$$

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 as 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:

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

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,

$$J_n\left(\frac{2\pi b_0\sin\theta}{\lambda}\right) \approx \frac{1}{n!}\left(\frac{\pi b_0 \sin\theta}{\lambda}\right)^n, \tag{19}$$

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

$$\left(\mathbf{A}^\dagger\mathbf{N}^{-1}\mathbf{v}\right)_{\theta,\varphi} = \frac{B_\theta^*(\theta)\sin\theta}{2\pi\sigma^2}\sum_{b,n} e^{in\varphi} e^{i2\pi\frac{b_y}{\lambda}\cos\eta\cos\theta}\sum_q \widetilde{B}_q^* J_{n-q}\left(\frac{2\pi b_0}{\lambda}\sin\theta\right) \tag{20}$$

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.]

2.4. *Effective beam patterns*

Unlike the narrow-beam toy example of Section 2.1, Equation (18) reveals that in general, the fringe-rate response of a celestial source depends on the declination of the source. Sources near the celestial pole tend to have low fringe rates, whereas those near the celestial equator tend to have higher fringe rates, as one can see in Figure [XXX: insert figure refrence]. Applying a filter in fringe-rate space therefore geometrically selects certain parts of the sky, effectively narrowing the primary beam of the observation. To see this, consider an interferometer baseline of length 16λ, located at the equator. Neglecting the effects of the instrument's primary beam, the contribution of sources to $\widetilde{V}_b(f_n)$ at various θ is given by $J_n(2\pi b_0 \sin\theta/\lambda)\sin\theta$. This function has a broad response over a large range of θ, as one can see in the left column of Figure 1. However, this response is highly oscillatory away from zero declination ($\theta = \pi/2$) and therefore cancels out near celestial poles, particularly for low fringe-rate bins. To quantify this, we plot in the right panels of Figure 1 the cumulative effect of sources at various θ, i.e.

$$I_n(\theta) \equiv \int_0^\theta J_n\left(\frac{2\pi b_0\sin\gamma}{\lambda}\right)\sin\gamma\, d\gamma, \tag{21}$$

and ask which range of θ values contribute most to the final result $\widetilde{V}_b(f_n) \propto I_n(\pi/2)$. One sees that for the low ($n = 5$) fringe-rate bin, most of the contribution to $\widetilde{V}_b(f_n)$ comes from a narrow region around $\theta \approx \pi/2$, whereas for the higher fringe-rate, a broader range of declinations contribute to the final result. A fringe-rate filter that downweights high fringe-rates therefore narrows the effective beam.

We now compute more rigorously quantify the notion of an effective beam. Note that 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 (3)] and leave a delta-function effective beam. Indeed, this is precisely what Equation (4) 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 (18) 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

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

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

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

where we have once again taken $\mathbf{N}$ to be diagonal, with entries σ^2, and have made the definitions

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

$$h_b(\theta,m) \equiv \frac{g_b(\theta,m)}{\int_0^\pi g_b(\gamma,m)d\gamma} \tag{25}$$

$$\lambda_b^m \equiv \left|\int_0^\pi g_b(\gamma,m)d\gamma\right|^2. \tag{26}$$

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 i^{th} 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 element 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 (5) says that

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

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

In the event of non-invertibility, on the other hand, one may elect to forgo the deconvolution step, and to simply normalize the estimator of the sky appropriately. Put another way, as an alternative to Equation (3), we can instead use

$$\hat{\mathbf{x}} = \mathbf{M}\mathbf{A}^t\mathbf{N}^{-1}\mathbf{v} \tag{28}$$

as our estimator, where $\mathbf{M}$ is a diagonal normalization matrix. Inserting Equation (2) into this and taking the ensemble average gives

$$\langle\hat{\mathbf{x}}\rangle = \mathbf{M}\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}\mathbf{x} \equiv \mathbf{B}^{\rm eff}\mathbf{x}, \tag{29}$$

where we have assumed that the instrumental noise is unbiased, so that $\langle\mathbf{n}\rangle = 0$, and have defined an effective beam matrix $\mathbf{B}^{\rm eff}$ that is proportional to $\mathbf{A}^\dagger\mathbf{N}^{-1}\mathbf{A}$, which we have already computed in an (m,θ) basis. If we were to re-express our result in a (φ,θ) (i.e., image) basis, each row of $\mathbf{B}^{\rm eff}$ would give the weighted average of the true sky that forms the estimator of the sky temperature in some direction—precisely what one would define as an effective beam. To ensure that these beams and our estimator are properly normalized, we require that each row sums to unity, which implicitly defines $\mathbf{M}$. Importantly, our definition of the effective beam includes both the effects of the primary beam and the synthesized beam.

For intuition and simplicity, however, suppose we stay in the (m,θ) basis and keep m fixed while we examine

the behavior of the effective beam as a function of θ for a pure east-west baseline. Ignoring irrelevant normalization constants, the result from Equation (23) is

$$B^{\rm eff}(\theta) \propto B_\theta(\theta)\sin\theta\sum_q \widetilde{B}_q J_{m-q}\left(\frac{2\pi b_0}{\lambda}\sin\theta\right). \quad (30)$$

The Bessel function provides an additional tapering of the primary beam, just as we heuristically predicted at the beginning of this section. This is not the result of deconvolution, since we assumed no deconvolution in this treatment. The narrowing of the beam is due purely to an optimal combination of time-ordered data.

Finally, with a limited number of baselines, one may also perform a partial deconvolution. For example, if one replaces the uncomputable quantity $\left[\mathbf{A}^\dagger\mathbf{N}^{-1}\mathbf{A}\right]^{-1}$ by a pseudoinverse, the eigenmodes that are present in the data will be deconvolved, while those that are missing will be set to zero (Shaw et al. 2013a). The result is a partially narrowed beam that converges to a delta function in both sky directions as baselines are added until full invertibility is attained.

2.5. *Denoising data sets*

Talk about how we can think of the denoising as a pre-processing step.

[XXX: Maybe discuss computational cost for narrow field telescopes vs. widefield?]

[XXX: Since it originated from Equation (3), our proposal of fringe-rate filtering is guaranteed to be lossless.]

[XXX; Mention somewhere that this is very closely related to what Richard has.]

3. FRINGE-RATE FILTERING

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).

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 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. 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 ($f_{\rm max}$) 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 $\approx -f_{\rm max}/2$ 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

$$f_{\rm max} = \frac{|\mathbf{b}_{\rm eq}|}{c}\omega_\oplus\nu, \quad (31)$$

where $f_{\rm max}$ is the maximum fringe rate, $\mathbf{b}_{\rm eq}$ 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 §**??**), $f_{\rm max} = 1.3$ 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,

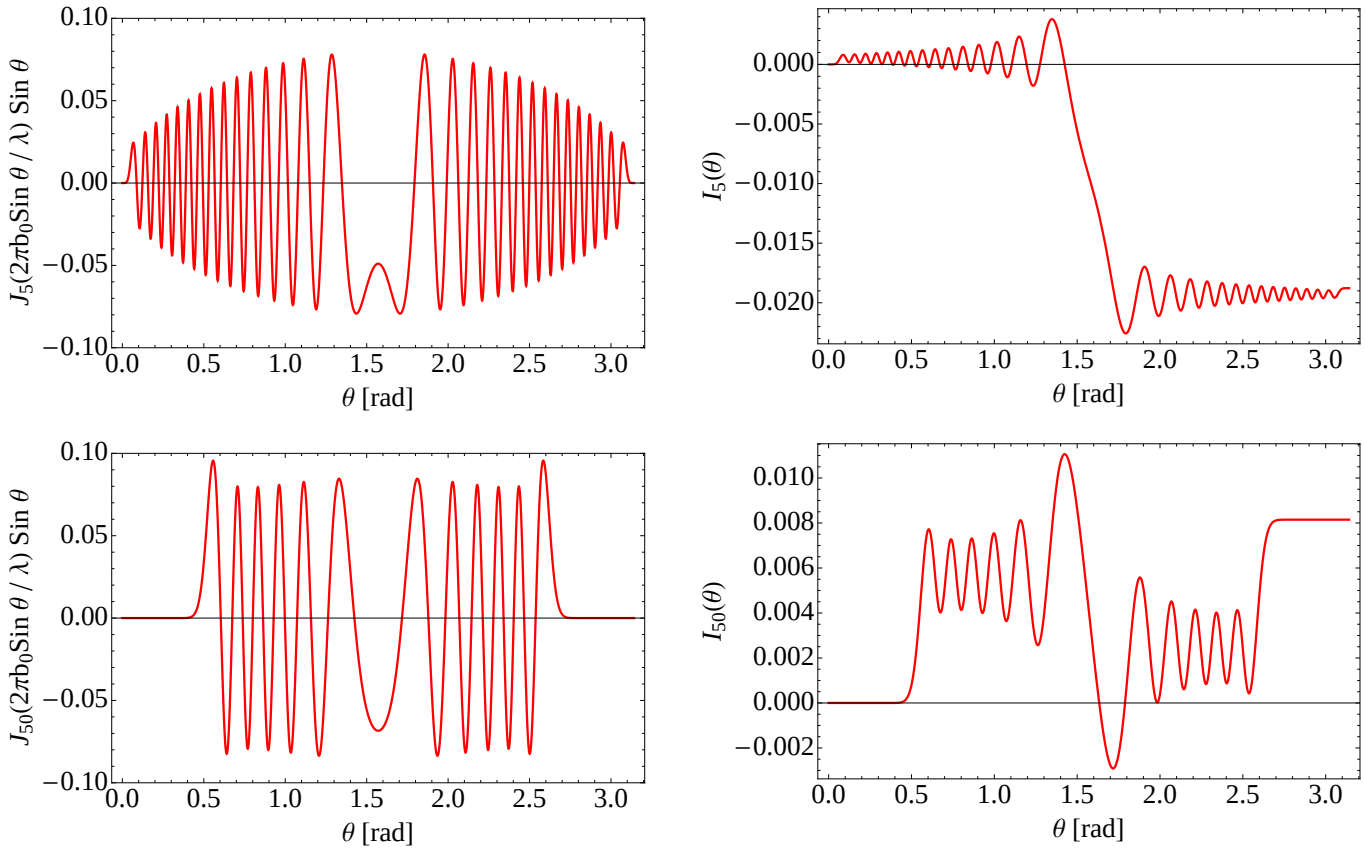

FIG. 1.— Left column: responses of selected fringe rate bins to different θ (and therefore different declinations). Right column: cumulative response integrated from 0 to θ, i.e. $I_n(\theta)$ as defined by Equation (21). The top row shows the $n = 5$ fringe-rate bin, while the bottom row shows the $n = 50$ fringe-rate bin, with the n^{th} bin defined to have a fringe rate of $f_n \equiv n/T_\oplus$. For these illustrative plots, the primary beam was ignored, and the baseline under consideration was a single 16λ baseline with an east-west orientation at the equator. From the right column, one sees that lower fringe rates derive most of their celestial signal from a narrow region in the sky, so low-pass fringe-rate filtering suggested in Section 2.3 causes a narrowing of the effective primary beam.

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.

$$\widehat{P}(\mathbf{k}_{t\tau}) = \left(\frac{\lambda^2}{2k_\mathrm{B}}\right)^2 \frac{X^2Y}{\Omega B} \left\langle \tilde{V}_i(\tau,t)\tilde{V}_j^*(\tau,t)\right\rangle_{i<j}, \quad (32)$$

which follows from equation 12 of P12a, with λ being the observing wavelength, k_B is Boltzmann's constant, X^2Y is a cosmological scalar with units of $\frac{h^{-3}\ \mathrm{Mpc}^3}{\mathrm{sr\cdot Hz}}$, Ω is the angular area[5], B is the bandwidth, $\langle\ldots\rangle_{i<j}$ indicates the ensemble average over instantaneously redundant baseline measurements indexed by i, j, and $\tilde{V}(\tau,t)$ 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.

[5] As described in detail in Appendix 10, the angular area used to normalize high-redshift 21cm power spectrum measurements (e.g., Ω in Equation 32) is proportional to the integral of the squared beam power over angular area (Ω_{PP}; equation 41). This contrasts the standard beam area (Ω_P; equation 42) that is used to relate flux density to a brightness temperature. Since Equation 32 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.

$$\widehat{\Delta}^2_{21}(k) = \frac{k^3}{2\pi^2}\left\langle \widehat{P}(\mathbf{k}_{t\tau})\right\rangle_{|\mathbf{k}_{t\tau}|=k}, \quad (33)$$

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 §3, 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.

3.1. *Crosstalk*

Crosstalk removal proceeds by subtracting the 1-hr time-average of the visibilities for each baseline from each integration. This process, described in Parsons et al. (2010), distinguishes oscillating fringes associated with celestial emission from the static phase bias associated

with crosstalk. This crosstalk-removal filter essentially constitutes a high-pass fringe-rate filter, as described in §3. The width of the stop-band of the crosstalk filter is much narrower than low-pass fringe-rate filters described in that section. Nighttime data are averaged in LST over the 55-day observation using 43-second time bins matching the integration interval of the data after the compression step described in §**??**.

4. SIMULATIONS

4.1. *Point Source Simulations for Mapping Beam Response*

4.2. *Noise Simulations for Estimating Signal Loss*

Describe the simulations here.

5. BEAM SCULPTING

6. MATCHING POLARIZATION BEAMS

7. RESULTS AND DISCUSSION

8. CONCLUSION

9. 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

$$\frac{d\tau_g}{dt} = -\omega_\oplus \left(\frac{b_x}{c} \sin H + \frac{b_y}{c} \cos H \right) \cos\delta, \tag{34}$$

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

$$-\omega_\oplus \frac{|b_y|}{c} \leq \frac{d\tau_g}{dt} \leq \omega_\oplus \frac{|b_y|}{c}. \tag{35}$$

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

$$\begin{aligned} \tilde{V}(\nu, f) &\approx \tilde{A}(\nu, f) * \tilde{S}(\nu) * \int e^{2\pi i \omega_\oplus \frac{b_y \nu}{c} t} e^{-2\pi i f t} dt \\ &\approx \tilde{A}(\nu, f) * \tilde{S}(\nu) * \delta_{\mathrm{D}}(\frac{b_y}{c}\omega_\oplus \nu - f), \end{aligned} \tag{36}$$

where f is the fringe-rate of the source[6], $\tilde{A}(\nu, f)$ 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 31.

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 31 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 §3, at PAPER's latitude, delay-rates range from -$f_{\mathrm{max}}/2$ 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.

10. 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 P12a. We express this volume in observing coordinates as

$$\mathbb{V} = \Omega B \cdot X^2 Y, \tag{37}$$

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

$$\begin{aligned} \tilde{V}^2(u, v, \eta) = \left(\frac{2k_{\mathrm{B}}}{\lambda^2}\right)^2 \Bigg[\int dl\ dm\ d\nu A(l, m, \nu) T(l, m, \nu) e^{-2\pi i (ul + v \\ \Bigg[\int dl'\ dm'\ d\nu' A^*(l', m', \nu') T^*(l', m', \nu') e^{2\pi i (ul' + vm \end{aligned} \tag{38}$$

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

$$\tilde{V}_{21}^2(u, v, \eta) \approx \left(\frac{2k_{\mathrm{B}}}{\lambda^2}\right)^2 \frac{B}{X^2 Y} \widehat{P}(\mathbf{k}) \int dl\ dm\, |A(l, m)|^2. \tag{39}$$

We compare this result with the relation between the delay-transformed visibility, $\tilde{V}$, to the three-dimensional

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

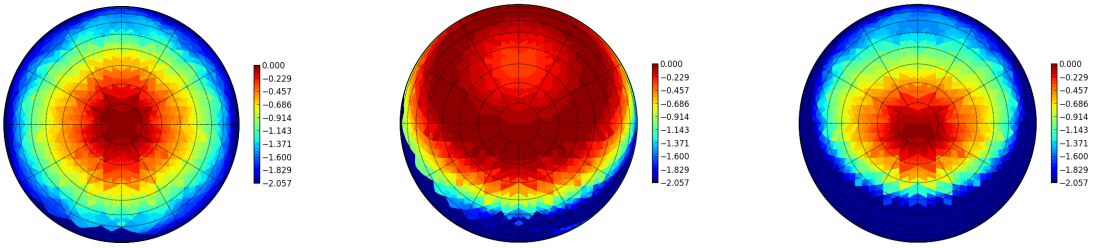

FIG. 2.— The effective primary beam response of a baseline, as determined from the simulations in §4.1. Panels indicate reconstructions of PAPER's model beam response used in the simulation (left), the beam weighting that results from the application of a fringe-rate filter weighted to optimize SNR for 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.

power spectrum of reionization, $P_{21}(\mathbf{k})$ (P12a):

$$\tilde{V}_{21}^2(u,v,\eta) \approx \left(\frac{2k_{\rm B}}{\lambda^2}\right)^2 \frac{\Omega\, B}{X^2 Y} \widehat{P}_{21}(\mathbf{k}). \quad (40)$$

As this shows, the relevant beam area in Equation 40 is the power-square beam, $\Omega_{\rm PP}$, given by

$$\Omega_{\rm PP} \equiv \int dl\ dm\, |A(l,m)|^2. \quad (41)$$

This contrasts with the standard metric for beam area — the integrated power beam — which we will call $\Omega_{\rm P}$, and is given by

$$\Omega_{\rm P} \equiv \int dl\ dm A(l,m), \quad (42)$$

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)

$$\Omega' \equiv \Omega_{\rm P}^2/\Omega_{\rm PP} \quad (43)$$

should be used in lieu of Ω, as these equations pick up two factors of $\Omega_{\rm P}$ in the conversion from Jy^2 to mK^2, along with the a factor of $\Omega_{\rm PP}$ in the denominator relating to the integrated volume. For equations that relate a measured visibility (in units of brightness temperature, e.g. Equation 32) to $\widehat{P}(\mathbf{k})$, the factor of $\Omega_{\rm P}^2$ is already applied in the conversion from units of Jy to mK, and Ω corresponds to the remaining factor of $\Omega_{\rm PP}$.

For PAPER, $\Omega_{\rm P} \approx 0.72$ sr, while $\Omega_{\rm PP}$ 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 $\Omega_{\rm P}$.

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