UNLOCKING SENSITIVITY FOR VISIBILITY-BASED ESTIMATORS OF THE 21 CM REIONIZATION POWER SPECTRUM

Yunfan Gerry Zhang[1], Adrian Liu[1,2,3], Aaron R. Parsons[1,2]

Draft version October 13, 2017

ABSTRACT

Radio interferometers designed to measure the cosmological 21 cm power spectrum require high sensitivity. Several modern low-frequency interferometers feature drift-scan antennas placed on a regular grid to maximize the number of instantaneously coherent (redundant) measurements. However, even for such maximum-redundancy arrays, significant sensitivity comes through partial coherence between baselines. Current visibility-based power spectrum pipelines, though shown to ease control of systematics, lack the ability to make use of this partial redundancy. We introduce a method to leverage partial redundancy in such power spectrum pipelines for drift-scan arrays. Our method cross-multiplies baseline pairs at a time lag and quantifies the sensitivity contributions of each pair of baselines. Using the configurations and beams of the 128-element Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER-128) and staged deployments of the Hydrogen Epoch of Reionization Array (HERA), we illustrate how our method applies to different arrays and predict the sensitivity improvements associated with pairing partially coherent baselines. As the number of antennas increases, we find partial redundancy to be of increasing importance in unlocking the full sensitivity of upcoming arrays.

1. INTRODUCTION

The Epoch of Reionization (EoR) represents the last key stage of our Universe's early evolution. The study of this event stands at the intersection of cosmology and astrophysics. Understanding this event not only serves as a scientific goal of its own, but also as a gateway to information regarding fundamental physics of inflation, neutrino mass, and the phenomenology of the first stars and galaxies (e.g., **??????**).

Observational studies of reionization, including Gunn-Peterson measurements of quasi-stellar objects (**?**) and Cosmic Microwave Background temperature and anisotropy measurements (CMB; **?**), the kinetic Sunyaev-Zel'dovich effect (**???**) and Lyman alpha galaxy observations (**???**) have given us indications of the rough time frame of reionization, but only limited constraints on the finer spatial and temporal structures. A surge of recent radio-astronomical experiments of reionization focus on measuring the "spin-flip" transition of neutral hydrogen with a characteristic wavelength of 21 cm (**??**). The 21 cm brightness temperature is a direct tracer of neutral hydrogen through the epoch of reionization. Therefore, by measuring the three dimensional distribution of the 21 cm signal we measure the full temporal and spatial variations of this event. However, before realizing full-scale 21 cm tomography, many current radio interferometric efforts aim to measure the spatial power spectrum of 21 cm brightness temperature fluctuations. Current-generation instruments include the Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER; **??**), the Murchison Widefield Array (MWA; **??**), and the Low Frequency Array (LOFAR; **?**). Next-generation instruments include the Hydrogen Epoch of Reionization Array (HERA; e.g. **????**), which is under construction, and the Square Kilometer Array Low (SKA-low; e.g. **?** and references therein), which is currently in planning stages.

The highly redshifted 21 cm signal is faint and diffuse, in contrast to the localized bright sources targeted by many traditional radio telescopes. Current 21 cm experiments are sensitivity-starved, with the experimental challenge further increased when considering foreground contamination five orders of magnitude brighter than the cosmological signal of interest. Low-frequency radio interferometers aiming to measure the 21 cm signal are thus designed differently from traditional instruments. To satisfy the sensitivity needs, modern arrays are large (ranging upwards from 100 elements) and are typically of the drift-scan (static-pointing) type to limit cost. To further improve sensitivity, experiments such as PAPER and HERA feature multiple copies of the same baselines to repeatedly measure the same Fourier signal (**?**).

Analysis pipelines for the 21 cm power spectrum typically fall into two categories. In the first, images are formed through rotation synthesis, and after a foreground mitigation step, are Fourier transformed to construct a power spectrum (**???**). An alternative technique works directly with visibilities from baselines, delay-transforming and cross-multiplying them to form the power spectrum (**?**). This technique avoids many systematics associated with combining data from different baselines and tracks the native sampling of the interferometer. An example of the visibility-based pipeline is presented in **?**, which provides power spectrum measurements with the 64-element version of PAPER (henceforth as PAPER-64). However, one disadvantage of existing visibility-based pipelines is their lack of treatment of partial redundancy. Baselines that are slightly different in length and orientation "rotate into" each other at a time offset. Baselines of different lengths and orientations therefore contain partially coherent information.

yunfanz@berkeley.edu
[1] Astronomy Dept., U. California, Berkeley, CA
[2] Radio Astronomy Lab., U. California, Berkeley, CA
[3] Hubble Fellow

This is the basis of earth-rotation synthesis. While imaging based power-spectrum pipelines naturally include all redundancy information, visibility-based pipelines so far only cross multiply fully redundant baselines, i.e., baselines of the same length and orientation.

Visibility-based pipelines are not fundamentally limited to using only fully redundant baselines. In fact, most sensitivity forecasts to date do include partial redundancy (**???**). Recently, **?** proposed a visibility-based approach to produce power spectrum measurements using partially redundant baselines in sky-tracking measurements. Arrays capable of such tracking include MWA and LOFAR. Our work introduces a two-baseline power spectrum estimator that can be applied to both tracking-capable arrays and drift-scan telescopes such as PAPER, HERA, and potentially SKA-low.

The basic idea of our proposed visibility-based, two-baseline power-spectrum estimator is as follows. The Earth's rotation causes the baselines in a drift-scan array to pick up different modes of the sky with time. Rotation synthesis makes use of the rotation-induced uv coverage map to form an image. With visibility-based pipelines, the same information can be extracted. To do so, we cross-multiply time-shifted visibilities, with the proper weighting, to form power spectra. Due to the large number of elements of modern arrays, the task of cross-multiplying every baseline against every other, scaling as number of array elements to the fourth power, can be computationally formidable, and many pairs of baselines provide only negligible redundancy information. Our contribution is thus twofold. First, we introduce a formalism to estimate the power spectrum from pairs of partially-redundant baselines in a visibility pipeline. Secondly we show how to use the two-baseline estimator to automatically pre-select baseline pairs and time offsets, making the problem computationally efficient. More precisely, our formalism allows one to simultaneously identify the baselines that have strong redundancy, find the time offset that corresponds to maximal redundancy for a given pair of baselines, and quantify the sensitivity associated with cross multiplying such a pair of partially redundant baselines, which in turn is used as weight to combine measurements in a power spectrum pipeline.

The rest of this paper is organized as follows. In Section 2 we introduce some terminology and notation used in the rest of the paper. In Section 3 we introduce the formalism for weighting partially redundant baselines and derive the power spectrum estimator. In Section 4 we present numerical tests of this technique as well as the expected sensitivity improvement this method provides for HERA and PAPER-128 pipelines. With Section 5 we conclude.

2. NOTATION AND TERMINOLOGY

In order to avoid confusion and ambiguity for the rest of this paper, we introduce some terminology that may differ from what is commonly found in the literature.

We make the distinction between a *baseline*, which corresponds to two specific antennas, and a *class of baselines*, which refers to all baselines of the same length and orientation in a given array. Baselines of the same class are traditionally called "redundant baselines", because they measure the same Fourier mode in the sky. We shall call baselines in the same class *equivalent baselines*, and

reserve the *redundancy* of two baselines to mean a variable function of the relative time offsets of their visibility time series. With this terminology, equivalent baselines are fully redundant with each other simultaneously at all times. Non-equivalent baselines also have partial redundancy, and the redundancy can be maximized by shifting their time series by a relative time offset.
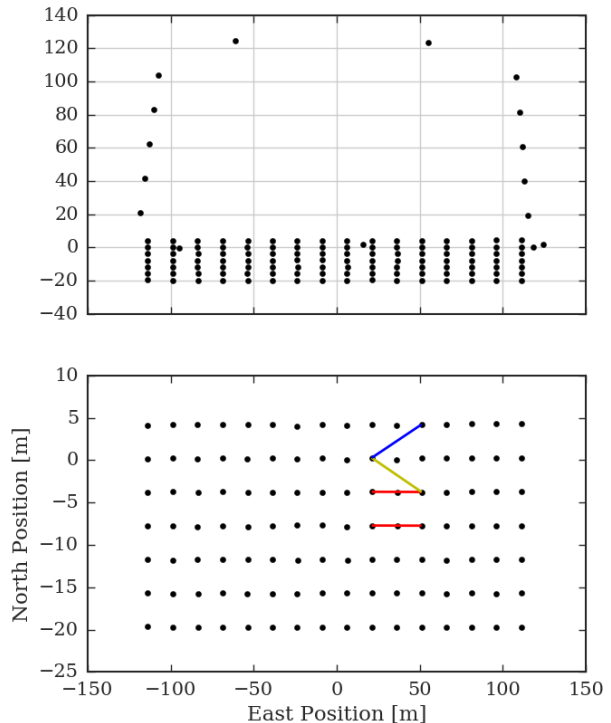

Fig. 1.— The PAPER-128 layout. Each dot corresponds to the location of an antenna. The top panel shows the antenna positions drawn to scale; the bottom panel show the antenna labels and distances of the 112-element grid, excluding the outrigger antennas. In the bottom panel, the two baselines marked with red segments are example of an equivalent pair, each with antennas separated by 2 units east and 0 unit north. Similarly, the baselines marked in blue and yellow are examples of classes that, due to the small North-South separation within the grid, are expected to be nearly equivalent to the red baselines.

We shall use the 128-element PAPER array (henceforth referred to as PAPER-128) to motivate our formalism and demonstrate our method, and extend our results to several HERA configurations in Section 4.4. The PAPER array is located in the Karoo desert in South Africa (30:43:17.5 S, 21:25:41.8 E). The layout pattern is shown in Fig. 1. The array consists of a 112-element core in a rectangular grid, and 16 "outriggers" used primarily to aid calibration. In the bottom panel, the two baselines marked in red are an example of an equivalent pair. We denote a equivalency class of baselines in the PAPER grid by their separations, in this case {2,0}, for the antennas are separated by 2 units east and 0 units north. Similarly, the baselines marked in blue and yellow are respectively examples of {2,1} and {2,−1}. Note that {2,0} and {−2,0} are the same class of baselines and should not be included twice in power-spectrum mea-

surements. Antennas in purely north-south baselines are close (4m), and hence these baselines are not suitable for sensitive measurements due to cross-coupling. On the other hand, the small North-South separation means that classes such as {2,0} and {2,1} are expected to be near-equivalent. The PAPER-64 analysis of **?** uses three classes of baselines, the PAPER-128 equivalent of which are {2, 0}, {2, 1} and {2, −1}. There, baselines are cross-multiplied within each class. This paper provides the method for inter-class multiplications. We will use the short hand notation $\{m, n : m', n'\}$ to denote a pair of baseline classes to be cross-multiplied.

3. METHOD

In this section we introduce our method to cross-multiply non-equivalent baseline classes.

3.1. *uvw tracks*

Radio interferometric observations are often described in the coordinates uvw, defined as:

$$(u, v, w) = \frac{\nu}{c}\boldsymbol{b}, \qquad (1)$$

where $\boldsymbol{b}$ is the baseline vector in Cartesian coordinates, with first and second coordinates pointing East and North, respectively, and ν is the frequency of observation, and c is the speed of light. Relative to a phase center on the sky, each baseline maps to a point in uvw space. As the Earth rotates, the points trace out tracks in the uvw space. We show in Fig. 2 uvw tracks of the three PAPER-128 baselines colored in Fig. 1, projected onto the uv and vw planes. The tracks are traced over 12 sidereal hours, at 0.15 GHz, relative to a phase center that drifts from zenith.

Equivalent baselines follow identical uvw tracks. Traditionally, we can identify redundancy of nearly equivalent baselines as crossings of the uv tracks, a two-dimensional projection of uvw, as shown in the top panel of Fig. 2. However, there are several reasons that uv track-crossing does not imply perfect redundancy. The most obvious reason is that the three-dimensional uvw tracks do not actually cross, as is evident from the vw projection in the bottom panel of Fig. 2. In general, track-crossings are not accurate enough to determine the optimal time offset for the cross multiplication of visibilities, nor can it provide an estimate of the degree of redundancy. Furthermore, for drift-scan arrays, even a hypothetical crossing in uvw space would not imply perfect redundancy. This will become evident in the next section, after we develop a more general formalism that accounts for the point spread function of primary. The relation between track-crossing and redundancy for both drift-scan and tracking measurements is further explored in Appendix A.

3.2. *Formalism*

In this section we formulate a power spectrum estimator from the product of delay-transformed visibilities from two arbitrary baselines. We shall work in the case of drift-scan telescopes and extend the results to tracking measurements.

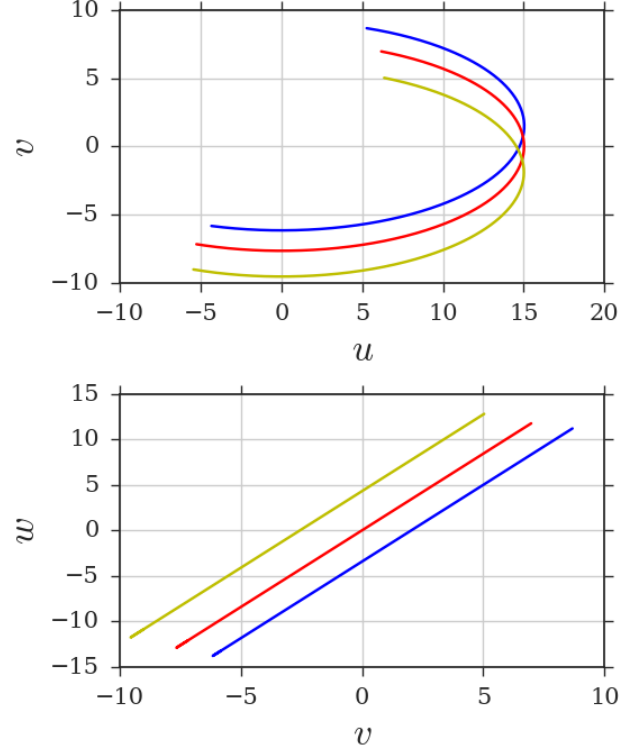We begin with the visibility as commonly defined in

FIG. 2.— Tracks of the three PAPER-128 baseline classes shown in Fig. 1, here with the same respective colors. Tracks are traced out over half a sidereal day for frequency $\nu = 0.15$ GHz and a phase center that passes through zenith. The top panel shows uv tracks with crossings among each pair of baseline classes. Projection to the vw plane in the bottom panel shows that although tracks appear to overlap in the uv plane, tracks in uvw space do not cross.

the literature (e.g. **??**):

$$
\begin{aligned}
V_\nu(\boldsymbol{b}) &= \int d\Omega A(\hat{\boldsymbol{s}}, \nu)\phi(\nu)I_\nu(\hat{\boldsymbol{s}}) \exp\left(-2\pi i \frac{\nu}{c}\boldsymbol{b}\cdot\hat{\boldsymbol{s}}\right) \\
&\approx \frac{2k_B}{\lambda^2} \int d\Omega A(\hat{\boldsymbol{s}}, \nu)\phi(\nu)T(\hat{\boldsymbol{s}}) \exp\left(-2\pi i \frac{\nu}{c}\boldsymbol{b}\cdot\hat{\boldsymbol{s}}\right),
\end{aligned}
$$
$$(2)$$

where k_B is Boltzmann's constant, λ is the mean wavelength, and $\hat{\boldsymbol{s}}$ and $d\Omega$ denote a direction in the sky and its corresponding solid angle. Inside the integral we have $\phi(\nu)$ as the frequency bandpass profile, $A(\hat{\boldsymbol{s}}, \nu)$ as the (frequency-dependent) primary beam, and I as the specific intensity, which has been related to T, the brightness temperature in the Rayleigh-Jeans limit. The beam power pattern A is dimensionless, normalized to 1 at its peak (zenith), and we assume it to be the same for all baselines. Power-spectrum measurements are typically taken from ~ 0.01 GHz centered around the corresponding redshift of interest (e.g. 0.15 GHz for redshift $z \approx 9.5$), thus imposing a sharp drop-off on $\phi(\nu)$.

Ultimately we would like to relate the observed visibility to the power spectrum, P, defined such that

$$\langle \widetilde{T}^*(\boldsymbol{k})\widetilde{T}(\boldsymbol{k}') \rangle = (2\pi)^3 \delta_D^3(\boldsymbol{k} - \boldsymbol{k}')P(k), \qquad (3)$$

where $\boldsymbol{k}$ is the cosmological wavenumber, δ_D^3 is the three-dimensional Dirac delta function, and $\widetilde{T}(\boldsymbol{k})$ is the three-dimensional Fourier transform of the brightness temper-

ature field $T(\mathbf{r})$, with $\mathbf{r}$ being the cosmological position coordinate. We begin by relating the observational coordinates[4] $\hat{s}$ and ν, to $\mathbf{r}$ and $\mathbf{k}$:

$$
\begin{aligned}
r &= \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')} \\
&\approx \frac{c}{H_0} \int_0^{z_0} \frac{dz'}{E(z')} - \frac{c(1+z_0)^2}{\nu_{21}H_0 E(z_0)} (\nu - \nu_0) \\
&\equiv X - Y\Delta\nu,
\end{aligned} \tag{4}
$$

where r is the magnitude of the vector $\mathbf{r}$, with z its corresponding redshift. $\nu_{21} = 1.420\,\mathrm{GHz}$ is the 21 cm transition rest frequency, ν_0 is a reference central frequency with corresponding redshift z_0, H_0 is the current-day Hubble constant, and

$$
E(z) = \sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}, \tag{5}
$$

with Ω_m and Ω_Λ being respectively the normalized matter and dark energy density. Inverting Eq. (4) for ν gives

$$
\nu = \frac{X - r}{Y} + \nu_0. \tag{6}
$$

This provides us with a mapping between the observational frequency and the cosmological distance in the radial direction. For the angular directions, we have in the thin-shell limit

$$
d^2 r = X^2 d\Omega. \tag{7}
$$

Note that Eq. (7) does not require the flat-sky approximation; the angular integral is still performed over the curved sky.

Eqs. (6) and (7) show that observational angular and spectral dimensions together correspond to the three dimensional volume of cosmological interest. Moreover, recall that the power spectrum in Eq. (3) is proportional to the product of Fourier transforms of the temperature field. Since Eq. (2) resembles a Fourier transform along the angular directions, it is natural to perform a further Fourier transform along the radial, or equivalently, frequency axis. The transform, called the delay transform (?), has gained popularity recently since it has been shown that foregrounds are isolated in delay space.[5] We define the delay-transformed visibility as:

$$
\begin{aligned}
V(\mathbf{b}, \tau) &\equiv \int d\nu V_\nu(\mathbf{b})\phi(\nu)\exp(2\pi i\nu\tau) \\
&= \int d\Omega d\nu B(\hat{s}, \nu)T(\hat{s}, \nu)\exp\left[-2\pi i\nu\left(\frac{\mathbf{b}\cdot\hat{s}}{c} - \tau\right)\right].
\end{aligned} \tag{8}
$$

Here the delay τ is the Fourier dual of ν. Eq. (8) expresses the delay-transformed visibility as an integral over observational coordinates $\hat{s}$ and ν. For notational simplicity we have defined the quantity

$$
B(\hat{s}, \nu) \equiv \frac{2k_B}{\lambda^2}\phi(\nu)A(\hat{s}, \nu). \tag{9}
$$

With Eqs. (6) and (7), we can rewrite the delayed-transformed visibility in cosmological coordinates as

$$
V(\mathbf{b}, \tau) = \int \frac{d^3 r}{X^2 Y} B(\mathbf{r})T(\mathbf{r})\exp\left[-2\pi i\left(\frac{\mathbf{b}}{c}\cdot\hat{\mathbf{r}} - \tau\right)\nu_r\right], \tag{10}
$$

where $d\nu = -dr/Y$ and $d^3 r = -X^2 Y d\Omega d\nu$. We have written ν_r as a reminder that ν and r are related by Eq. (6).

Existing visibility-based power-spectrum pipelines for redundant drift-scan arrays relate the power-spectrum to the conjugate square of the visibilities (???). We would like to generalize such relations by relating the power-spectrum to the product of two visibilities from two arbitrary baselines and time offsets. To do this, we must account for the fact that the beam pattern of a baseline shifts relative to the sky as the Earth rotates. Here we choose to fix the sky, and denote the rotated coordinates in the topocentric frame with the three-dimensional rotation operator $\boldsymbol{\Gamma}$:

$$
V_\psi(\mathbf{b}', \tau) = \int \frac{d^3 r}{X^2 Y} B(\boldsymbol{\Gamma}\mathbf{r})T(\mathbf{r})e^{-2\pi i\left(\frac{\mathbf{b}}{c}\cdot\boldsymbol{\Gamma}\hat{\mathbf{r}} - \tau\right)\nu_r - i\psi_\nu}, \tag{11}
$$

where we introduced a frequency-dependent phase ψ_ν, which to linear-order in ν one would typically pick to correspond to the re-phasing of the two visibilities to the same phase center.[6]

With implicit bounds of integrals from $-\infty$ to ∞, we have

$$
\begin{aligned}
&\langle V^*(\mathbf{b}, \tau)V_\psi(\mathbf{b}', \tau)\rangle \\
&= \int \frac{d^3 r d^3 r'}{(X^2 Y)^2}\langle T^*(\mathbf{r})T(\mathbf{r}')\rangle B^*(\mathbf{r})B(\boldsymbol{\Gamma}\mathbf{r}')\Phi(\mathbf{r}, \boldsymbol{\Gamma}\mathbf{r}') \\
&= \int \frac{d^3 r d^3 r'}{(X^2 Y)^2}\left(\int \frac{d^3\kappa}{(2\pi)^3}\frac{d^3\kappa'}{(2\pi)^3}\langle T^*(\boldsymbol{\kappa})T(\boldsymbol{\kappa}')\rangle e^{-i(\boldsymbol{\kappa}\cdot\mathbf{r} - \boldsymbol{\kappa}'\cdot\mathbf{r}')}\right) \\
&\quad \times B^*(\mathbf{r})B(\boldsymbol{\Gamma}\mathbf{r}')\Phi(\mathbf{r}, \boldsymbol{\Gamma}\mathbf{r}') \\
&= \int \frac{d^3 r d^3 r'}{(X^2 Y)^2}\left(\int \frac{d^3\kappa}{(2\pi)^3}P(\kappa)e^{-i\boldsymbol{\kappa}\cdot(\mathbf{r} - \mathbf{r}')}\right) \\
&\quad \times B^*(\mathbf{r})B(\boldsymbol{\Gamma}\mathbf{r}')\Phi(\mathbf{r}, \boldsymbol{\Gamma}\mathbf{r}'), \tag{12}
\end{aligned}
$$

where

$$
\begin{aligned}
\Phi(\mathbf{r}, \boldsymbol{\Gamma}\mathbf{r}') \equiv &\exp\left[i\frac{2\pi}{c}\left(\mathbf{b}\cdot\nu_r\hat{\mathbf{r}} - \mathbf{b}'\cdot\nu_{r'}\boldsymbol{\Gamma}\hat{\mathbf{r}}'\right)\right] \\
&\times \exp\left[-i2\pi\tau\left(\nu_r - \nu_{r'}\right) - i\psi_\nu\right]. \tag{13}
\end{aligned}
$$

The third equality of Eq.(15) follows from assuming the translational invariance of the statistics of the 21 cm field. Continuing, we may make the assumption that the 3D power spectrum varies negligibly over the k-space of interest, thus allowing us to pull out of the integral the power spectrum centered at

$$
k_{\mathbf{b}, \tau} \equiv 2\pi\sqrt{\left(\frac{\tau}{Y}\right)^2 + \left(\frac{\bar{b}}{\lambda X}\right)^2}, \tag{14}
$$

where $\bar{b}$ is approximately the mean of b and b'. This

---

[4] The observational direction $\hat{s}$ is equal to the cosmological direction $\hat{\mathbf{r}}$. We shall nevertheless keep the notations separate to highlight the conceptual differences and switch between the symbols depending on the context.

[5] We discuss foreground isolation in detail in Section 4.2.

[6] For many cases, the linear-order interpretation is sufficient. We keep the general term to reserve the option to determine its exact form numerically as this information comes at no additional computational cost.

gives

$$\langle V^*(\boldsymbol{b},\tau)V_\psi(\boldsymbol{b}',\tau)\rangle$$

$$\approx P(k_{\bar{b},\tau})\int \frac{d^3r d^3r'}{(X^2Y)^2}\delta_D^{(3)}(\boldsymbol{r}-\boldsymbol{r}')B^*(\boldsymbol{r})B(\boldsymbol{\Gamma}\boldsymbol{r}')\Phi(\boldsymbol{r},\boldsymbol{\Gamma}\boldsymbol{r}')$$

$$= P(k_{\bar{b},\tau})\int \frac{d^3r}{(X^2Y)^2}B^*(\boldsymbol{r})B(\boldsymbol{\Gamma}\boldsymbol{r})e^{-i2\pi\frac{\nu_r}{c}(\hat{\boldsymbol{r}}\cdot\boldsymbol{b}-\boldsymbol{\Gamma}\hat{\boldsymbol{r}}\cdot\boldsymbol{b}')-i\psi_\nu}, \quad (15)$$

where δ_D is the Dirac delta-function, and notice that the phase factor $\exp\left[-i2\pi\tau(\nu-\nu')\right]$ drops out in the end. In factoring out the power spectrum from the integral, one is essentially making the approximation that the estimator probes only a single Fourier mode. In Section 4.2 we will relax this assumption and examine the exact form of the Fourier-space footprint that is being probed, as well as its effect on foreground isolation.

Since the beam pattern and bandpass are given in $\hat{\boldsymbol{s}}$ and ν, we convert the last line of Eq. (15) back to these coordinates to get the general relation between the delay-transformed visibilities and the power spectrum:

$$\langle V^*(\boldsymbol{b},\tau)V_\psi(\boldsymbol{b}',\tau)\rangle$$

$$\approx P(k_{\bar{b},\tau})\int \frac{d\Omega d\nu}{X^2Y}B^*(\hat{\boldsymbol{s}},\nu)B(\boldsymbol{\Gamma}\hat{\boldsymbol{s}},\nu)e^{i2\pi\frac{\nu}{c}(\hat{\boldsymbol{s}}\cdot\boldsymbol{b}-\boldsymbol{\Gamma}\hat{\boldsymbol{s}}\cdot\boldsymbol{b}')-i\psi_\nu}.$$

$$(16)$$

We can therefore form the power spectrum estimator for the baseline pair $\{\boldsymbol{b}:\boldsymbol{b}'\}$:

$$\hat{P}(k_{\bar{b},\tau}) \equiv \frac{V^*(\boldsymbol{b},\tau)V_\psi(\boldsymbol{b}',\tau)}{\Theta}, \quad (17)$$

where the weight is defined as

$$\Theta \equiv \int d\nu\,\Theta_\nu, \quad (18)$$

with

$$\Theta_\nu \equiv e^{-i\psi_\nu}\int \frac{d\Omega}{X^2Y}B^*(\hat{\boldsymbol{s}},\nu)B(\boldsymbol{\Gamma}\hat{\boldsymbol{s}},\nu)e^{i2\pi\frac{\nu}{c}(\hat{\boldsymbol{s}}\cdot\boldsymbol{b}-\boldsymbol{\Gamma}\hat{\boldsymbol{s}}\cdot\boldsymbol{b}')}.$$

$$(19)$$

Notice that Θ has no dependence on τ. We point out that although all our derivations focused on drift-scan telescopes, we can get the analogous result for tracking measurements simply by noticing that for a tracking primary beam with radial symmetry we have:

$$\Theta_\nu \equiv e^{-i\psi_\nu}\int \frac{d\Omega}{X^2Y}B^*(\hat{\boldsymbol{s}},\nu)B(\hat{\boldsymbol{s}},\nu)e^{i2\pi\frac{\nu}{c}(\hat{\boldsymbol{s}}\cdot\boldsymbol{b}-\boldsymbol{\Gamma}\hat{\boldsymbol{s}}\cdot\boldsymbol{b}')}.$$

$$(20)$$

Roughly speaking, Eqs. (17) through (19) tell us that the product of visibilities at a time offset is proportional to the power spectrum times the Fourier transform of the cross multiplied beam patterns. As a check, when applied to equivalent baselines, $\boldsymbol{b}=\boldsymbol{b}'$, the peak correlation occurs at $\Delta t=0$ and $\hat{\boldsymbol{s}}=\boldsymbol{\Gamma}\hat{\boldsymbol{s}}$, in which case Eq. (16) reduces to Eq. (B9) of **?**. With Eq. (17) and Eq. (18) we can, for any given pair of baseline classes and time offsets, estimate the degree of redundancy, here represented by Θ. This allows us to achieve our goals stated in the introduction: to identify candidate baseline pairs with significant redundancy, to find the time offset that maximizes redundancy, and to quantify the degree of such redundancy. We can do all the above simply by computing the weight Θ from Eq. (18) for various time offsets.
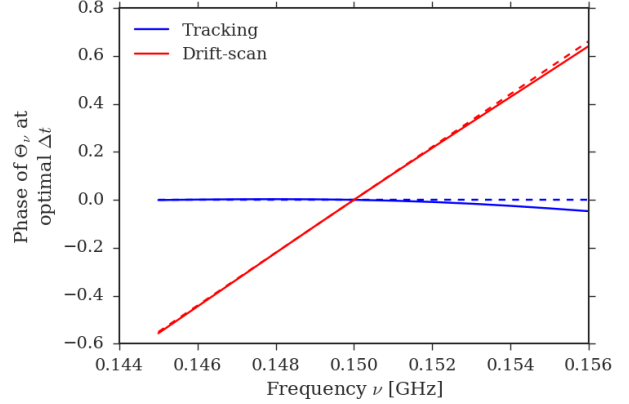

FIG. 3.— Frequency dependent peak phases for PAPER-128 baseline pair $\{2,0:2,1\}$. Both the drift-scan case and a hypothetical tracking baseline are shown in solid lines. The first-order, linear effects are shown in dashed lines. We have also fixed the global phase in both cases to 0 at 0.15 GHz. The drift-scan case exhibits linear behavior and conforms well to the first order effect of a shift in delay space due to the movement of zenith.

3.3. Rephasing

If ψ_ν were set to zero in Eq. (19), Θ_ν at the peak of correlation (as a function of time) is generally complex, and often far from real. Furthermore, this phase of peak correlation is inevitably frequency dependent. This frequency dependence would lead to destructive interference when we integrate over frequency, unless we correct Θ_ν by a phase ψ_ν.

Selecting ψ_ν to be a linear function of ν allows one to cancel the decoherence due to the two visibilities having different phase centers. By default the correlators of a drift-scan array phase the two visibilities both to zenith at the same time. When they are cross-multiplied with a time lag, the visibilities must be rephased before the delay transform to account for the movement of the phase center. The effect of the phase thus roughly corresponds in a shift of the delay mode measured, and we should expect ψ_ν to be approximately a linear function of ν:

$$\psi_\nu \sim 2\pi\Delta\tau\nu, \quad (21)$$

where $\Delta\tau$ is a shift in delay corresponding to the phase center movement, i.e.

$$\Delta\tau = \frac{\boldsymbol{b}'}{c}\left(\boldsymbol{\Gamma}\hat{\boldsymbol{s}}-\hat{\boldsymbol{s}}\right). \quad (22)$$

In Fig. 3 we show the phases of Θ_ν at the optimal time offset, i.e. when $|\Theta_\nu|$ is maximized. The phases are shown for the baseline pair $\{2,0:2,1\}$ of PAPER-128. The drift-scan phase dependence is compared with that of the same baseline with hypothetical tracking-elements (Eq. (19) and Eq. (20)). In the drift-scan case, we indeed see a linear relation corresponding to a delay of $\Delta\tau \approx 15.7$ ns. In the tracking case, since the phase center is fixed to the sky, only second order effects are observed. The origin of the second order effects can be seen as due to the w-term, or more precisely the fact that uvw tracks do not cross when their two-dimensional projections do. We refer the reader to Appendix A for further explanation. In both cases, the full effects are encapsulated in the phase of Θ_ν and can be thus determined empirically without any
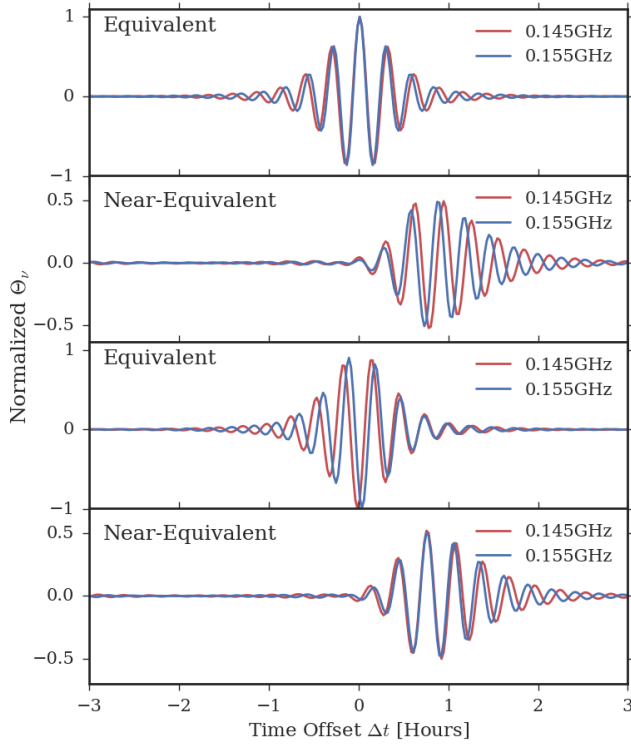
FIG. 4.— Comparisons of the peak phases of two different frequencies. Specifically shown are the real parts of Θ_ν. First and third panels shows equivalent baselines, second and fourth show a pair of nearly equivalents. The top two panels have zero phase shift ($\psi_\nu = 0$), and the bottom two are rephased to a time offset of 0.78 hours. The first and last panels thus show a coherently rephased series that would add constructively at the respective peaks.

additional computation.

To illustrate the effect of rephasing, we compare in Fig. 4 real parts of Θ_ν of both equivalent and nearly equivalent baseline pairs for two channels: 0.145 GHz and 0.155 GHz. The top two panels have zero rephasing, and the bottom two are linearly rephased to a differential delay of $\Delta\tau \approx 15.7$ ns, as we determined above. The first and third panels show the equivalent baseline pairs {2,0:2,0}, and second and fourth panels show {2,0:2,1}. Summing over frequency without proper rephasing leads to destructive interference and sensitivity loss. The wider the frequency profile, the more destructive the interference would be. Only after rephasing to the correct time offset for each baseline pairs can we constructively combine the frequency channels.

We see from Fig. 4 that although the amplitude of correlations match up for all time offsets, the phase would only locally match. The need for rephasing can thus be understood as a symptom of the underlying spatial decoherence; coherence at the phase center does not extend to the entire beam pattern (see Section 4.1). While integrating over frequencies and sky-direction in Eq. (18), there are necessarily sky-directions that do not add coherently. This decoherence across the beam pattern cannot be removed and plays a fundamental role in determining how much sensitivity one can recover from nearly equivalent baselines.

4. ANALYSIS

In this section we delve into visualizations and analyses of our method, and explore sensitivity contributions of various baseline classes for PAPER and HERA array configurations.

4.1. *Visualization*

So far for clarity and generality we have avoided the traditional formulation of radio astronomy in terms of the uv-plane. To gain some visual intuition of the formalism in Section 3, we show explicit beam fringe patterns of the two baselines and their interactions. For this section, we let $(l, m, \sqrt{1 - l^2 - m^2}) \equiv \hat{s}$ and examine a single frequency of 0.15 GHz.

At the time offset of maximum redundancy, we expect two baselines to have to same fringe pattern in both frequency and phase. Due to the time offset, however, the beam centers would be slightly shifted with respect to each other. In the top panels of Fig. 5, we show the real parts of peak-normalized beam-fringe patterns. The top left and middle panels show the beam-fringe patterns for the baseline classes {2,0} and {2,1}, offset by 0.78 hours. More specifically, displayed are the real parts of $g_b \equiv B(\hat{s}, \nu) \exp\left[-i2\pi\frac{\nu}{c}\hat{s} \cdot \boldsymbol{b}\right]$ and $g_{b'} \equiv B(\boldsymbol{\Gamma}\hat{s}, \nu) \exp\left[-i2\pi\frac{\nu}{c}\boldsymbol{\Gamma}\hat{s} \cdot \boldsymbol{b'}\right]$, respectively. The top right panel show their conjugate product $g_b^* g_{b'} e^{-i\psi_\nu} = B^*(\hat{s}, \nu) B(\boldsymbol{\Gamma}\hat{s}, \nu) \exp\left[i2\pi\frac{\nu}{c}(\hat{s} \cdot \boldsymbol{b} - \boldsymbol{\Gamma}\hat{s} \cdot \boldsymbol{b'}) - i\psi_\nu\right]$, where we have chosen global phase ψ_ν such that the peak of the conjugate product is real. The product of the beam fringe patterns in the top right shows that the fringes cancel out as we expect. Comparing with Eq. (19), we see that $\Theta_\nu = \int d\Omega g_b^* g_{b'} e^{-i\psi_\nu}$ is the sum over the image shown in the top right panel. At the optimal time offset, the fringe patterns of the two baselines cancel out, leading to a substantial weight Θ_ν. The frequency-summed weight Θ is constructed with rephasing as described in Section 3.3 so that the conjugate product (top-right panel) for different frequencies align in phase at their peaks. Note it is not possible to align the phase of $g_b^* g_{b'} e^{-i\psi_\nu}$ for all (l, m, ν) due to spatial decoherence across the beam as discussed at the end of Section 3.3.

Similarly, we show the correlation of the two baselines in terms of the overlap of their point-spread functions in the uv-plane. The bottom-left and middle panels are the peak-normalized instantaneous uv coverage of two baselines ($\tilde{g}_b$ and $\tilde{g}_{b'}$) while the bottom right panel is their product. We see from the bottom right panel that the cross-multiplied baselines' recovered power is concentrated at $u \approx 15$. The product of the uv point-spread functions shown in the bottom-right panel should not be confused with the Fourier transform of the beam-fringe conjugate product shown in the top-right panel.

4.2. *Chromaticity*

A crucial concern when measuring the 21 cm power spectrum is the possibility of astrophysical foreground contamination. Because foregrounds are orders of magnitude larger than the cosmological signal, even small residuals from an imperfect isolation of these foregrounds can masquerade as a false detection. Recent literature treatments of this problem have shown that if one combines
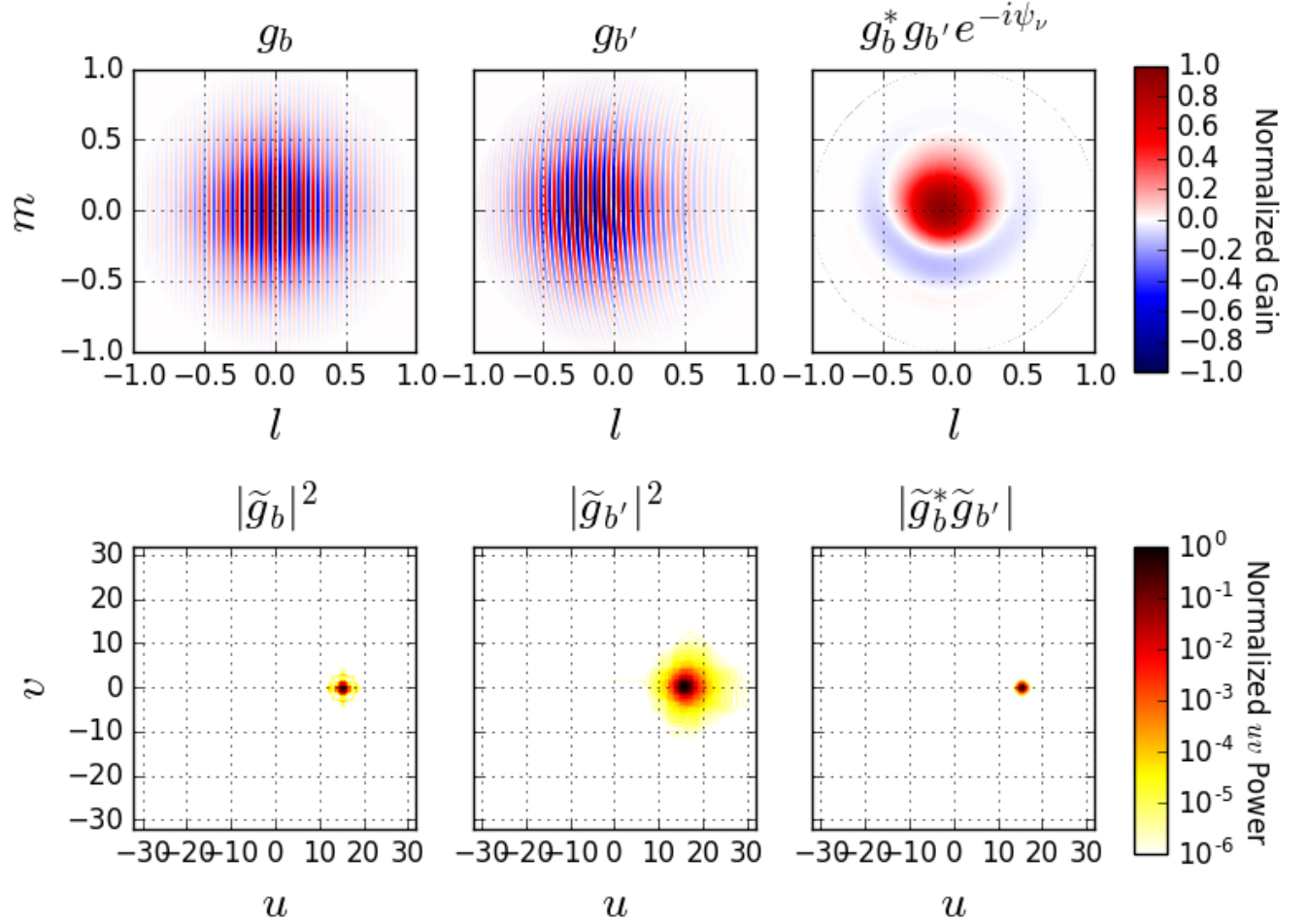
FIG. 5.— PAPER-128 beam-fringe patterns and their point-spread functions on the uv-plane for a frequency of 0.15 GHz. Selected baselines have 30-meter East-West separations. The first baseline, shown in the first column, is purely East-West, while the second baseline, shown in the middle column, has 4-meter North-South separation and has been rotated with a time offset of 0.78 hours. Panels in the top row show the real parts of the beam-fringe patterns of the baselines (left and middle), and their conjugate product (right). The beam-fringe values are normalized such that the peak of the original beam is unity. The bottom-left and middle panels show the peak-normalized uv point spread of the two beam-fringe patterns. The product of the beam spreads displayed in the bottom-right panel shows that power concentrated at $u \approx 15$ is recovered.

the mathematical properties of interferometric measurement with the empirical fact that foregrounds are spectrally smooth, a clean separation of foregrounds and cosmological signal can be achieved in harmonic space. These studies suggest that foregrounds preferentially appear in harmonic modes of the sky corresponding to fluctuations that are angularly fine but spectrally smooth, a region in harmonic space colloquially known as "the wedge" (**??????????????????**). A foreground mitigation strategy can then consist mainly of avoiding those modes by working in the complementary region known as "the EoR window". However, for this strategy to be viable, it is necessary to show that one's data analysis pipeline does not corrupt the separation between the wedge and the EoR window. We will now do so for our proposed two-baseline power spectrum estimator.

Traditionally, spectral and angular dimensions are cylindrically binned with the coordinates $k_\perp$ and $k_\parallel$, where the angular direction $k_\perp$ is probed by the beam and line-of-sight direction $k_\parallel$ is probed by the frequency spectrum. Despite their usefulness, however, $(k_\perp, k_\parallel)$ are local coordinates that are well defined only in the narrow field-of-view limit. As discussed in previous sections, the combination of information from nearly equivalent baselines for wide-field telescopes is inherently a three-dimensional problem. Although the estimator of Section 3 does not have such limitations, to discuss foreground signatures with possible curved sky effects, we must employ a set of basis functions where angular fluctuations are encoded by spherical harmonics, and radial fluctuations by spherical Bessel functions.

? discusses in detail power-spectrum analyses using spherical harmonics. In such a basis, modes in the sky are indexed by the magnitude k of the wavevector $\mathbf{k}$ and spherical harmonic indices (ℓ, m). In other words, one defines

$$\overline{T}_{\ell m}(k) \equiv \sqrt{\frac{2}{\pi}} \int d\Omega dr r^2 j_\ell(kr) Y_{\ell m}^*(\hat{\mathbf{r}}) T(\hat{\mathbf{r}}), \qquad (23)$$

where $Y_{\ell m}$ is a spherical harmonic function and j_ℓ is the ℓth-order spherical Bessel function of the first kind. The power spectrum is then related to these spherical harmonic Bessel modes via the relation

$$\langle \overline{T}_{\ell m}(k) \overline{T}_{\ell' m'}^*(k') \rangle = \frac{\delta^D(k-k')}{k^2} \delta_{\ell\ell'} \delta_{mm'} P(k). \qquad (24)$$

A delay-transformed visibility $V(\mathbf{b}, \tau)$ is related to the spherical Fourier-Bessel modes by

$$V(\mathbf{b}, \tau) = \sqrt{\frac{2}{\pi}} \sum_{\ell m} \int dk\, k^2 g_{\ell m}(k; \mathbf{b}, \tau) \overline{T}_{\ell m}(k), \qquad (25)$$

where

$$g_{\ell m}(k; \mathbf{b}, \tau) = \int d\Omega d\nu B(\hat{\mathbf{r}}, \nu) Y_{\ell m}(\hat{\mathbf{r}}) j_\ell(kr) e^{i2\pi\nu(\tau - \mathbf{b}\cdot\hat{\mathbf{r}}/c)} \tag{26}$$

and the analogous quantity for V_ψ is given by

$$g_{\ell m}^{\psi}(k; \mathbf{b}, \tau) = \int d\Omega d\nu B(\mathbf{\Gamma}\hat{\mathbf{r}}, \nu) Y_{\ell m}(\hat{\mathbf{r}}) j_\ell(kr)$$
$$\times e^{i2\pi\nu(\tau - \mathbf{b}\cdot\mathbf{\Gamma}\hat{\mathbf{r}}/c) + i\psi_\nu}. \qquad (27)$$

Note that unlike the primary beam and the fringe pattern, the spherical harmonic function does *not* rotate, since it originated from a spherical harmonic expansion of the sky temperature, which is fixed. With these expressions and a little algebra combining Eqs. (24) through (27), we arrive at a form of the two-baseline estimator (compare with Eq. (16)):

$$\hat{P}(k) = \frac{\langle V^*(\mathbf{b}, \tau) V_\psi(\mathbf{b}', \tau) \rangle}{\Theta}$$
$$= \sum_\ell \int dk\, W_\ell(k; \mathbf{b}, \mathbf{b}', \tau) P(k), \qquad (28)$$

where we have defined the *window function*

$$W_\ell(k; \mathbf{b}, \mathbf{b}', \tau) \equiv \frac{2k^2}{\pi\Theta} \sum_m g_{\ell m}(k; \mathbf{b}, \tau) g_{\ell m}^{\psi*}(k; \mathbf{b}', \tau). \qquad (29)$$

By construction, the window function sums to unity when integrated over all k and summed over all ℓ. We may therefore interpret our estimator $\hat{P}$ of the power spectrum to be a weighted average over all possible ℓ and k modes, with the window function providing the weights of this average. This means that window functions can serve as indicators of foreground leakage into the EoR window. One writes down the estimator for a hypothetical measurement of a power spectrum mode within the EoR window, and additionally computes the window function for that estimator. Ideally, the window function for $\hat{P}(k)$ will be sharply peaked around k. In general, however, window functions will have wings that encroach upon other modes. If the window function is substantially non-zero at low-k modes—which is where foreground emission typically resides, given its spectral smoothness—then the EoR window will be contaminated by foregrounds.

Figure 6 shows four example window functions. The two window functions in the top plot are for equivalent baseline power-spectrum estimators, i.e., where one squares the delay spectrum from a single baseline and normalizes the result. The window function localized at low ℓ is for the baseline {2,0} of around 30 m in length, while the window function localized at high ℓ is for {12,0}, roughly 180 m in length. The window functions for near-equivalent baseline estimates are shown in the bottom plot, for {2,0:2,1} and {12,0:12,1} at low and high $\bar{b}$, respectively. In all cases, the window functions that we show are for $\tau = 703$ ns, and are generated using primary beam models for PAPER. For numerical convenience we assumed that the bandpass ϕ_ν (see Eq. (9)) takes the form of a half-period sine curve that peaks at 0.15 GHz and goes to zero 5 MHz on either side of the peak.

One sees that all four window functions are fairly localized around specific values of k, which to a good approximation are described by Eq. (14). This suggests that both equivalent- and near-equivalent-baseline delay-spectrum estimators are good estimators of the power spectrum. At higher ℓ, the window functions for both estimators become elongated in k, with stronger tails towards the low k modes where foregrounds reside. This is the phenomenology of the wedge, where at fine angular scales the foreground leakage to higher k modes are more
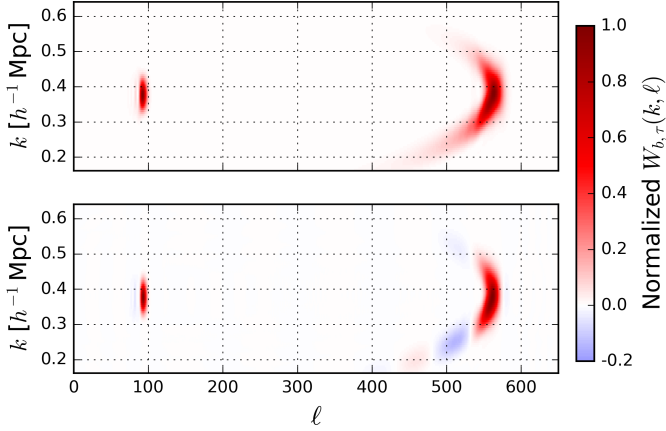
FIG. 6.— Sample window functions for delay-spectrum estimates of the power spectrum explored in this paper. The top plot shows window functions for a traditional equivalent-baseline delay-spectrum estimate, where a single baseline's delay spectrum is squared. The bottom plot shows the window functions resulting from nearly equivalent baselines. In each plot, the window function on the right corresponds to the window function for baselines of length $b \sim 180$ m in the East West direction, while the window function on the left is for $b \sim 30$ m. The general extent of the window functions for the two-baseline estimator are seen to be roughly the same as for the single-baseline estimator, suggesting that the former is just as good as the latter in keeping the EoR window free of foregrounds.

pronounced. Importantly, however, we note that the window functions for the near-equivalent-baseline estimator are no more elongated than for the equivalent-baseline estimator. One may thus conclude that the our proposed near-equivalent-baseline estimator does not result in extra foreground contamination of the EoR window.

4.3. Sensitivity

In this section we discuss the sensitivity contributions of a variety of baseline pairs from PAPER-128. Recall the sensitivity contribution of a particular baseline pair is given by Θ (Eq. (18)). From now on we use $\Theta_{bb'}$ and $\Delta t_{bb'}$ to denote the peak value and location of Θ, i.e. the maximum of the redundancy weight $|\Theta(\Delta t)|$ and the corresponding offset for a given baseline pair $\{b : b'\}$.

In the middle panel of Fig. 7 we show $\Theta_{bb'}$ and $\Delta t_{bb'}$ for a variety of baseline combinations. Baseline pairs that are mirror images of each other give approximately the same amount of redundancy ($\Theta_{bb'}$), with the opposite time offset, as expected from symmetry. For example, $\{1,0:1,1\}$ is mirror image of $\{1,0:1,-1\}$ and these two pairs of baselines give the same sensibility contribution. Thus we only show a subset of representative baseline pairs to illustrate the contributions. For a more complete result, see Fig. 9.

As is immediately clear from the figure, baseline pairs that have smaller optimal time offsets tend to have higher correlations. In other words, correlation peaks with closer to zero time offset are higher. This is expected for two reasons; one is that the longer the time offset of maximum redundancy, the more the sky has moved with respect to the primary beams and hence the smaller the overlapping patch of sky surveyed. The other reason is that smaller optimal time offset corresponds to smaller w-term-induced decoherence and hence better re-

dundancy.

To determine the actual relative contribution to sensitivity of these baseline pairs, we also have to account for the multiplicities of these baseline classes, i.e., the number of antenna pairs with the same length and orientation. We would like to estimate an effective weight $\widetilde{\Theta}_{bb'}$ that accounts for both the peak height $\Theta_{bb'}$ and the multiplicities of the baseline classes. From Fig. 7 we see for example that $\{1,0\}$ has higher multiplicity than $\{2,0\}$, or $\{1,1\}$. Assuming that each equivalent baseline delivers the same Θ_{bb}, the relative contribution to sensitivity can be estimated as follows.

First we can average the visibilities of the equivalent baselines. Since the core of PAPER-128 is a 16 by 7 antenna configuration, there are $M \equiv (16 - |m|) \times (7 - |n|)$ copies of the baseline class $\{m, n\}$. This means that if we add visibility measurements of all the equivalent baselines, we get a factor of $\sqrt{M}$ reduction in the noise level σ_N of the visibility. The sensitivity contribution of $\{m, n\}$, cross multiplied with $\{m', n'\}$ thus roughly scales as $\sqrt{(16 - |m|)(7 - |n|)(16 - |m'|)(7 - |n'|)} = \sqrt{MM'}$. For cross-multiplications of nearly equivalent baselines of types $\{m, n\}$ and $\{m', n'\}$, we get an effective weight:

$$\widetilde{\Theta}_{bb'} \propto \Theta_{bb'} \sqrt{MM'}. \qquad (30)$$

Shown in the bottom panel of Fig. 7 are the peak heights weighted by the multiplicity factor. Points that have zero optimal time offset are the equivalent baseline pairs and their weighted correlation values simply reflect the multiplicity factor. For clarity of presentation we have "folded over" the negative time offsets and combined baseline pairs that are mirror images.

With $\widetilde{\Theta}_{bb'}$, we can estimate the power spectrum by inverse covariance weighting. We again assume that the power spectrum P varies negligibly over the k-space of interest, so that measurements at different $k_{\bar{b},\tau}$ can be combined to obtain an estimator at some k_τ:

$$\hat{P}(k_\tau) = \frac{\sum_{bb'} \hat{P}(k_{\bar{b},\tau})/\sigma_P^2(bb')}{\sum_{bb'} 1/\sigma_P^2(bb')}$$
$$= \frac{\sum_{bb'} \hat{P}(k_{\bar{b},\tau})\widetilde{\Theta}_{bb'}^2}{\sum_{bb'} \widetilde{\Theta}_{bb'}^2}, \qquad (31)$$

where the sum is over classes of baseline pairs, and the power spectrum noise covariance, σ_P^2, is proportional to the effective weight, $\widetilde{\Theta}$.

We define the estimator sensitivity to be the square root of the inverse of the total power spectrum noise variance Σ_P^2:

$$\rho \propto 1/\Sigma_P \propto \rho_0 \sqrt{\sum_{bb'}^N \widetilde{\Theta}_{bb'}^2}, \qquad (32)$$

where, if σ_S^2 and σ_N^2 are the characteristic signal and noise levels of a single-baseline visibility, $\rho_0 \equiv \sigma_S^2/\sigma_N^2$ is the signal to noise ratio.

The scaling in Eq. (30) was an approximate one for simplicity of motivation. As we derive in Appendix B, this weight should be corrected by a factor proportional to ρ_0:
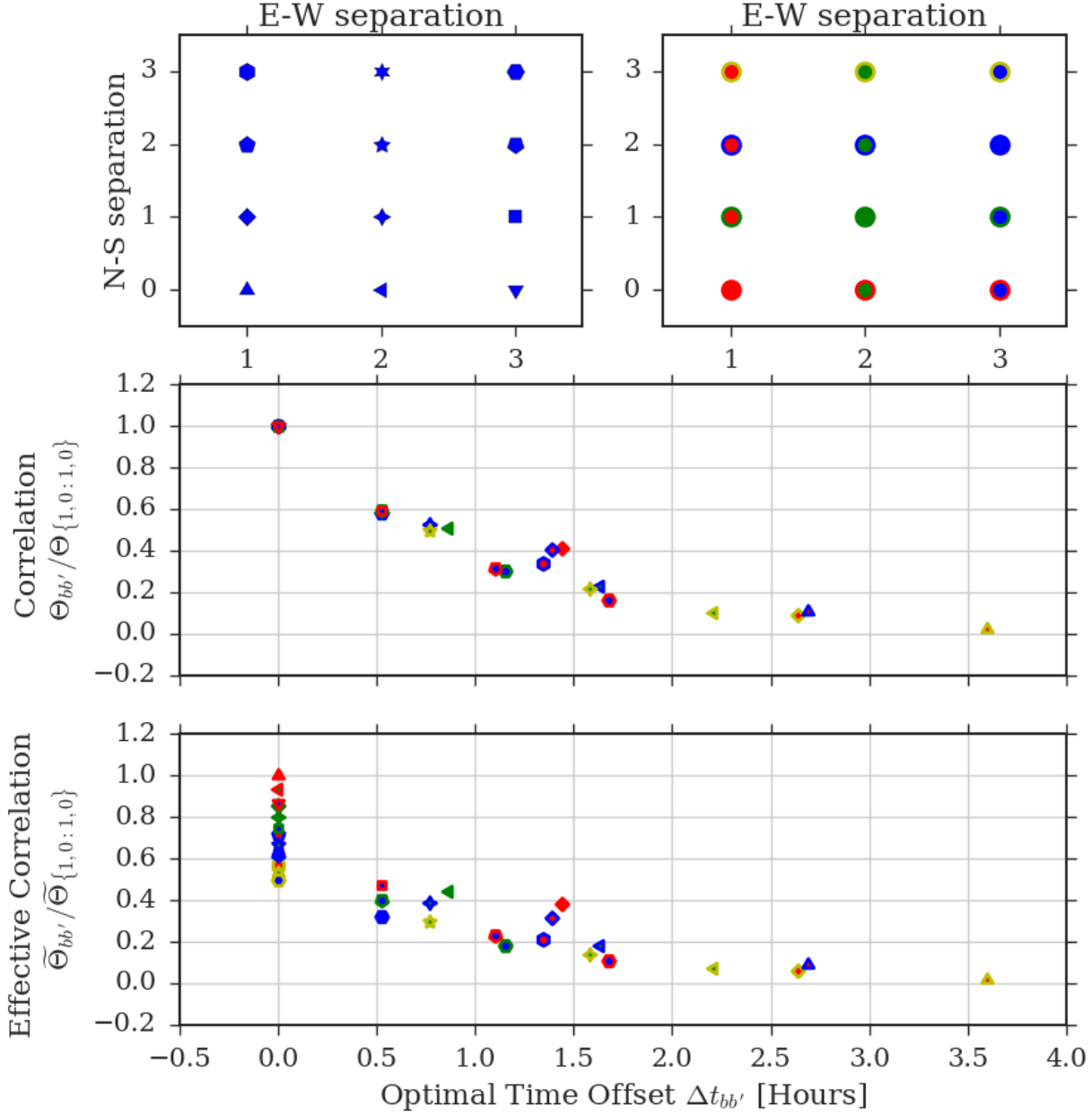
FIG. 7.— Relative sensitivity contributions of selected baseline combinations in PAPER-128. Each labeled point in the middle and bottom panels correspond to a cross-correlated baseline pair {m,n:p,q}. The shape of the symbol encodes the first baseline {m,n}, as displayed in the top left legend panel. The edge and face colors encode the second baseline {p,q}, as displayed in the top right legend panel. The middle panel shows the peak height (Θ) of each baseline combination, while the bottom panel multiplies the heights by the corresponding multiplicities as in Eq. (30). In both the middle and bottom panels, we have chosen to fix the value of {1,0:1,0} to unity.
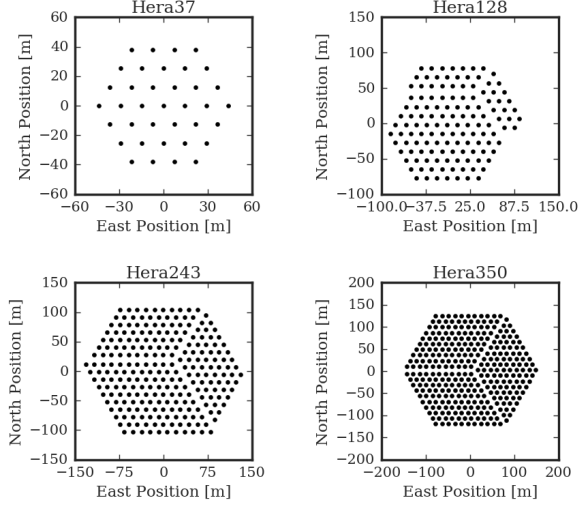
FIG. 8.— Planned Hydrogen Epoch of Reionization Array antenna configurations. HERA-37 is expected to complete and start collecting data in August of 2017, while the other three are planned configurations in the upcoming phases. For HERA-350, only the 320 elements in the core are shown.

$$\widetilde{\Theta}_{bb'} = \frac{\Theta_{bb'}\sqrt{MM'}}{\sqrt{1 + \rho_0\,(M + M')}}. \qquad (33)$$

For a given ρ_0, Eq. (33) quantifies the relative sensitivity contribution of a baseline pair $\{b : b'\}$. Assuming a reionization signal of $\sim 30\,\mathrm{mK}$, observation centered at $0.15\,\mathrm{GHz}$ ($z = 8.5$), and 120 days of integration with PAPER antennas, we have roughly (see Eq.(20) in ?)

$$\rho_0 \sim 0.001 \left[\frac{\bar{b}}{40\,\mathrm{m}}\right] \left[\frac{0.1h\,\mathrm{Mpc}^{-1}}{k}\right]^3, \qquad (34)$$

where $\bar{b}$ is the average baseline length between the pair.

4.4. *Array Configuration Comparisons*

We run our algorithm over all possible baseline-pairs of PAPER-128, HERA-37, HERA-128, HERA-243 and HERA-350. The HERA antenna configurations are shown in Fig. 4.2. The hexagonal design is the densest pattern of antenna-packing. The larger arrays are fractured with a "gap" dividing the antennas into three different groups. The gaps are designed so as to improve uv coverage and ease calibration without compromising sensitivity, but also produces many more nearly equivalent baselines than a pure hexagonal layout. The motivations behind the designs are explained in ?. Compared to PAPER-128, the hexagonal pattern of HERA lack short baselines oriented close to each other, and thus we expect to see only longer nearly equivalent baselines with high correlations. The lower multiplicities per class of baselines is compensated by the larger number of classes of baseline-pairs, especially given the gap in the larger versions.
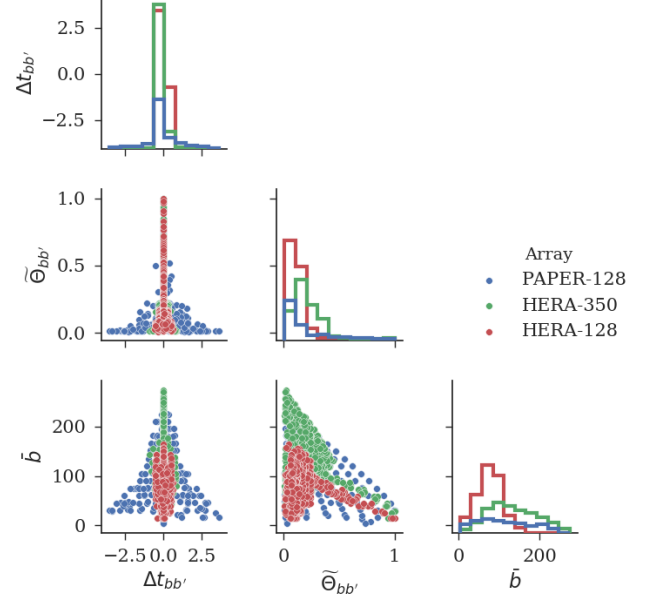

FIG. 9.— Pairplots of the top contributing baseline pairs in three arrays. Plotted properties are optimal time offset $\Delta t_{bb'}$ (in hours), effective weight $\widetilde{\Theta}_{bb'}$, and baseline length $\bar{b}$ (in meters). Only those points with $\widetilde{\Theta}_{bb'} > 0.01$ are shown (the weight for the top class of equivalent pair is normalized to 1). Scatter points of the three arrays overlap, in the order indicated by the legend.

In Fig. 9 we present pair-distributions of three different properties for baselines that contribute well to the sensitivity ($\widetilde{\Theta}_{bb'} > 0.01$, where again the weight for the top class of equivalent pair is normalized to 1). The properties shown are effective weight $\widetilde{\Theta}_{bb'}$, optimal time offset $\Delta t_{bb'}$, and average baseline length $\bar{b}$. We only show three of the mentioned arrays for visual clarity. The other two results are qualitatively similar. We point out some key features:

- $\Delta t_{bb'}$ vs. $\widetilde{\Theta}_{bb'}$: This relation is familiar from Fig. 7. The points at $\Delta t = 0$ are the equivalent baselines. HERA arrays have fewer data points with high $\Delta t_{bb'}$, due to their comparatively narrow beam. All near-equivalent pairs for HERA arrays appear at $\widetilde{\Theta}_{bb'} \lesssim 0.4$.

- $\Delta t_{bb'}$ vs. $\bar{b}$: PAPER-128 shows the trend that longer baselines correspond to lower $\Delta t_{bb'}$. HERA arrays do not exhibit this trend here because of a selection effect. Shorter HERA baselines require much longer Δt to overlap, partly because of the hexagonal structure requires $60°$ of rotation to overlap baselines, partly because of the smaller primary beam. Thus most short HERA baseline pairs are not shown in this figure, as they do not meet the selection criterion $\widetilde{\Theta}_{bb'} > 0.01$.

- $\widetilde{\Theta}_{bb'}$ vs. $\bar{b}$: In this plot the HERA arrays each show two superimposed structures. The equivalent baselines, all having the same individual weight $\Theta_{bb'}$, appear as elongated structures towards high $\widetilde{\Theta}_{bb'}$. The narrowness of these structures is an indication

of the approximately linear and one-to-one correspondence between baseline length and multiplicity. The wide structures spanning the whole range of $\bar{b}$ are the near-equivalent baseline pairs. We see that for HERA arrays, all near-equivalent pairs appear at $\widetilde{\Theta}_{bb'} \lesssim 0.4$. For PAPER-128, the existence of short near-equivalent pairs with high multiplicity partially fills the high $\widetilde{\Theta}_{bb'}$ region. Note the general trend that longer baselines tend to have lower $\widetilde{\Theta}_{bb'}$. This is due to the lower multiplicity of longer baselines. The trend is particularly obvious in the linear structure for the HERA arrays, which are the equivalent baselines. All having unity $\Theta_{bb'}$ by definition, the linearly decreasing trend of $\widetilde{\Theta}_{bb'}$ is a direct measure of the baseline multiplicity structures of the HERA array configurations.

- The top sensitivity-contributing near-equivalent pair in each array is normalized to have the same $\Theta_{bb'}$ (not shown), but those of HERA arrays have much lower $\widetilde{\Theta}_{bb'}$ than the top pair of PAPER-128. This is because they are longer baselines with lower multiplicity. In the end, these baseline classes still lead to high contributions to total sensitivity (Fig. 10) because there are many more such baseline pairs for HERA.

As Fig. 9 suggests, not all baseline pairs contribute significantly to sensitivity. Due to the large number of baseline pairs, it may not be computationally feasible to include all pairs. Having quantified the sensitivity from a given pair of baselines, we study the cumulative sensitivity of the array depending on which baseline pairs we include. By construction, we prefer the pairs with larger $\widetilde{\Theta}_{bb'}$. In Fig. 10, we plot fractional sensitivity ρ as a function of $\widetilde{\Theta}_{min}$, the minimum cutoff for $\widetilde{\Theta}_{bb'}$. In other words $\rho(\widetilde{\Theta}_{min})$ is the sensitivity when all baseline pairs with $\widetilde{\Theta}_{bb'} > \widetilde{\Theta}_{min}$ are included.

Note that here we normalized sensitivity for each array as the fraction of the total sensitivity of an array, i.e. when all baseline pairs are used. The plot therefore does not compare absolute sensitivity across different arrays. We see as expected that in all cases, using the nearly equivalent baselines leads to increasingly significant improvements as $\widetilde{\Theta}_{min}$ is lowered, or in other words, when more baseline pairs are used. We point out that PAPER results until now (??) only used the top contributing baseline pairs, essentially only exploiting the sensitivity of the $\rho(\widetilde{\Theta}_{min} \sim 1)$ scenario.

The dashed lines represent the values when only the equivalent baseline-pairs are used. The small and unfractured HERA-37, with no gap (like in HERA-350) or short nearly equivalent baselines (like in PAPER 128), will not benefit much from the nearly equivalent baselines. Near-equivalent baselines make an increasingly significant contribution to the total sensitivity as the number of array elements are increased. PAPER-128 is designed with highly redundant nearly equivalent baselines, and thus these baselines start contributing at higher $\widetilde{\Theta}_{min}$. The fractured HERA configurations will benefit even more from nearly equivalent baselines at low $\widetilde{\Theta}_{min}$ due to the presence of more classes of such pairs. The stepwise pattern seen in almost all cases are characteristic of a regular grid; as we step to lower $\widetilde{\Theta}_{bb'}$ large groups of baseline pair classes get included in "batch". Thus the "unfractured" HERA-37, with more regularity in its antenna configurations than the larger counterparts, exhibit more step patterns in fractional sensitivity.
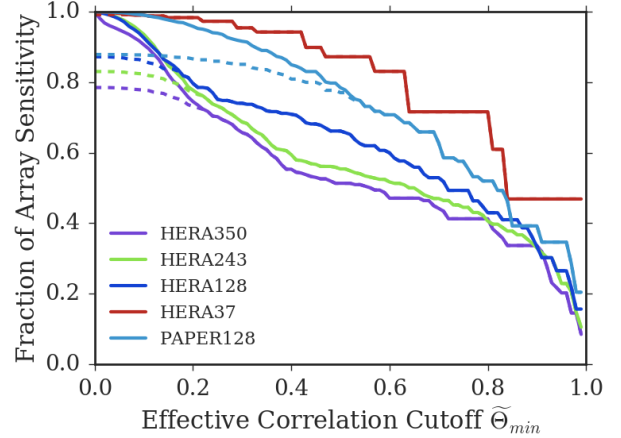

FIG. 10.— Fractional sensitivity of redundant arrays as a function of the minimum effective weight. Dashed lines represent when only equivalent baseline pairs are used, while solid lines indicate use of both the equivalent and nearly equivalent baselines are used. The vertical axis is normalized independently for each array to the total sensitivity when all baseline pairs, equivalent and near-equivalent, are used.

5. CONCLUSION

Upcoming arrays aiming at measuring the 21 cm power spectrum, including HERA and SKA-low, feature increasing number of antennas on regular grids to maximize coherent information. Having efficient pipelines to extract full sensitivity of these arrays is thus a pressing necessity.

Current visibility-based power spectrum pipelines have the advantage of easy-to-manage systematics, but have yet to unlock the full sensitivity of the arrays as they do not incorporate partially redundant baselines. We present a visibility and delay-transform based method to overcome this limitation by extracting the sensitivity contained in such partial redundancy. Applicable to both tracking and drift scan measurements, our method relies on cross-multiplications of visibilities that are offset in time. Given an antenna array configuration, our method identifies the best baseline pairs to cross-multiply and predicts the optimal time offset $\Delta t_{bb'}$, weight $\Theta_{bb'}$, and the necessary rephasing factor ψ_ν. With the predicted results one can incorporate partial redundancy into existing delay-transform based power-spectrum pipelines.

One benefit of delay-transform based power-spectrum estimation is the ability to isolate foregrounds in Fourier space. We show that our proposed estimator does not lead to extra foreground leakage into the EoR window. We then apply our estimator to a variety array configurations. We show that for arrays with increasing number of antennas, partially redundant baselines make up a more significant portion of the array's total sensitivity. The proposed power-spectrum estimator enables one to unlock the full sensitivity of a redundant radio inter-

APPENDIX

A.
TRACK-CROSSING AND W-TERM

In this section we link our results from Section 3 to the traditional views of rotation synthesis and in particular uv tracks. In describing the relative motion of the baseline relative to the sky, we have in Section 3 rotated the direction coordinate $\hat{s}$ with $\mathbf{\Gamma}$. The typical convention of rotation synthesis has the rotation operator acting on $\boldsymbol{b}'$ instead of $\hat{s}$, in which case Eq. (18) takes the equivalent form

$$\Theta \equiv \int \frac{d\Omega d\nu}{X^2 Y} B^*(\hat{s}, \nu) B(\mathbf{\Gamma}\hat{s}, \nu) e^{-i2\pi \frac{\nu}{c} \hat{s} \cdot \left(\boldsymbol{b} - \mathbf{\Gamma}^{-1} \boldsymbol{b}'\right)}, \tag{A1}$$

where $\mathbf{\Gamma}^{-1}$ is the inverse of $\mathbf{\Gamma}$. And for tracking elements we have analogously:

$$\Theta \equiv \int \frac{d\Omega d\nu}{X^2 Y} B^*(\hat{s}, \nu) B(\hat{s}, \nu) e^{-i2\pi \frac{\nu}{c} \hat{s} \cdot \left(\boldsymbol{b} - \mathbf{\Gamma}^{-1} \boldsymbol{b}'\right)}, \tag{A2}$$

Track-crossing corresponds to vanishing of the exponent

$$\boldsymbol{b} - \mathbf{\Gamma}^{-1} \boldsymbol{b}' = 0. \tag{A3}$$

If exponent vanishes, $\Theta_\nu(\Delta t)$ is real, and one would not observe any non-zero phase at its peak value. However, with the inclusion of the w-term, we see that track crossings in uv-plane do not actually imply crossing in the uvw space, and therefore the exponent or, equivalently, Eq. (A3) does not vanish and we observe non-zero, frequency dependent peak phases even in the case of tracking arrays, as shown in Fig. 3.

Furthermore, in the case of drift-scan arrays, even a crossing in uvw space does not necessarily correspond to the maximum of correlation. From Eq. (A1), we see the track-crossing condition Eq. (A3) maximizes Θ if and only if no other term in the integral depends on $\mathbf{\Gamma}$. We see that this is true for the tracking case (Eq. (A2)), but not the drift-scan case (Eq. (A1)). Only in the special case of a tracking measurement, where the w-term happens to be negligible, do uv-track crossings correspond exactly to maxima of correlations. In this case, at the crossing points the exponent in Eq. (A3) vanishes and ψ_ν is zero for all ν. In general, second-order effects due to the presence of w are observed for the tracking measurements as in Fig. 3.

B.
DERIVATION OF NOISE COVARIANCE

In this appendix we provide a derivation of the effective weight $\widetilde{\Theta}$ quoted in Section 4.3. In doing so, we assume that the power spectrum measurements from distinct baseline classes are combined by inverse variance weighting. To begin, we separate the visibility and power spectrum into signal and noise contributions:

$$\begin{aligned} V &= V_S + V_N, \\ P &= P_S + P_N. \end{aligned} \tag{B1}$$

We write the noise variance of power spectrum and visibility as:

$$\begin{aligned} \sigma_V^2 &= \langle |V_N|^2 \rangle, \\ \sigma_P^2 &= \langle P_N^2 \rangle. \end{aligned} \tag{B2}$$

One may notice that we use a single covariance for the complex visibility. It is straightforward to show that the same result holds if a separate real and imaginary components are used, as long as they are independent of each other. In fact, for simplicity and without loss of generality we shall treat the visibility as a real quantity in the rest of this derivation. Note that though we can assume $\langle V_N^m \rangle = 0$, where m is odd. The same is not true for P_N.

The variance of P constructed with visibilities V_1 and V_2 from two baseline classes can be estimated:

$$\sigma_P^2 = \langle P^2 \rangle - \langle P \rangle^2,$$
$$\propto \langle \frac{(V_{1S} + V_{1N})^2 (V_{2S} + V_{2N})^2}{\Theta^2} \rangle - \langle \frac{(V_{1S} + V_{1N})(V_{2S} + V_{2N})}{\Theta} \rangle^2$$
$$= \frac{1}{\Theta_{12}^2} \left(V_{1S}^2 \sigma_{V2}^2 + V_{2S}^2 \sigma_{V1}^2 + \langle V_{1N}^2 V_{2N}^2 \rangle \right)$$
$$= \frac{1}{\Theta_{12}^2} \left[V_S^2 (\sigma_{V2}^2 + \sigma_{V1}^2) + \sigma_{V1}^2 \sigma_{V2}^2 \right],$$
$$\text{(B3)}$$

where in the second last line we have substituted in the visibility noise variance, and have assumed independent noise between baselines. In the final line we used Wick's theorem and the fact that the signal from the two visibilities are equal.

Recall from the discussion on multiplicities that we can write

$$\sigma_V^2 = \frac{\sigma_0^2}{M}, \tag{B4}$$

where σ_0 is the single-baseline noise level. Letting $\rho_0 = V_S^2/\sigma_0^2$ be the signal to noise ratio for a single baseline, we can write

$$\sigma_P^2 \propto \frac{\sigma_0^4}{\Theta_{12}^2} \left[\rho_0 \left(\frac{1}{M_1} + \frac{1}{M_2} \right) + \frac{1}{M_1 M_2} \right]$$
$$\propto \frac{1}{\widetilde{\Theta}_{12}^2}. \tag{B5}$$

Thus we have defined a slightly modified version of the effective weight in Eq. (33):

$$\widetilde{\Theta}_{12} = \frac{\Theta_{12} \sqrt{M_1 M_2}}{\sqrt{1 + \rho_0 (M_1 + M_2)}}. \tag{B6}$$