CHARACTERIZING SIGNAL LOSS IN THE 21 CM REIONIZATION POWER SPECTRUM:
A REVISED STUDY OF PAPER-64

Carina Cheng[1,◇], Aaron R. Parsons[1,2], Matthew Kolopanis [3], Daniel C. Jacobs[3], Adrian Liu[1,4,†], Saul A. Kohn[5], James E. Aguirre[5], Jonathan C. Pober[6], Zaki S. Ali[1], Gianni Bernardi[7,8,9], Richard F. Bradley[10,11,12], Chris L. Carilli[13,14], David R. DeBoer[2], Matthew R. Dexter[2], Joshua S. Dillon[1,*], Pat Klima[11], David H. E. MacMahon[2], David F. Moore[5], Chuneeta D. Nunhokee[8], William P. Walbrugh[7], Andre Walker[7]

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

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

[3]School of Earth and Space Exploration, Arizona State U., Tempe AZ

[4]Berkeley Center for Cosmological Physics, Berkeley, CA

[5]Dept. of Physics and Astronomy, U. Penn., Philadelphia PA

[6]Dept. of Physics, Brown University, Providence RI

[7]Square Kilometer Array, S. Africa, Cape Town South Africa

[8]Dept. of Physics and Electronics, Rhodes University, South Africa

[9]INAF-Instituto di Radioastronomia, Bologna Italy

[10]Dept. of Electrical and Computer Engineering, U. Virginia, Charlottesville VA

[11]National Radio Astronomy Obs., Charlottesville VA

[12]Dept. of Astronomy, U. Virginia, Charlottesville VA

[13]National Radio Astronomy Obs., Socorro NM

[14]Cavendish Lab., Cambridge UK

ABSTRACT

The Epoch of Reionization (EoR) is an uncharted era in our Universe's history during which the first stars and galaxies led to the ionization of neutral hydrogen in the intergalactic medium. There are many experiments investigating the EoR by tracing the 21 cm line of neutral hydrogen. Because this signal is very faint and difficult to isolate, it is crucial to develop analysis techniques that maximize sensitivity and suppress contaminants in data. It is also imperative to understand the trade-offs between different analysis methods and their effects on power spectrum estimates. Specifically, with a new generation of instruments and a statistical power spectrum detection in our foreseeable future, it has become increasingly important to understand how certain analysis choices can lead to the loss of the EoR signal. In this paper, we focus on signal loss associated with power spectrum estimation. We describe the origin of this loss using both toy models and data taken by the 64-element configuration of the Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER). In particular, we highlight how detailed investigations of signal loss have led to a revised, higher 21 cm power spectrum upper limit from PAPER-64. Additionally, we summarize errors associated with power spectrum error estimation that were previously unaccounted for. While we only focus on a subset of PAPER-64 data in this paper, revised power spectrum limits from the PAPER experiment are presented in a forthcoming paper by Kolopanis et al. (*in prep.*) and supersede results from previously published PAPER analyses.

1. INTRODUCTION

By about one billion years after the Big Bang ($z \sim 6$), the first stars and galaxies are thought to have ionized all the neutral hydrogen that dominated the baryonic matter content in the Universe. This transition period, during which the first luminous structures formed from gravitational collapse and began to emit intense radiation that ionized the cold neutral gas into a plasma, is known as the Epoch of Reionization (EoR). The EoR is a relatively unexplored era in our Universe's history, which spans the birth of the first stars to the full reionization of the intergalactic medium (IGM). This epoch encodes important information regarding the nature of the first galaxies and the processes of structure formation. Direct measurements of the EoR would unlock

◇ccheng@berkeley.edu

†Hubble Fellow

*NSF AAPF Fellow

powerful characteristics about the IGM, revealing connections between the matter distribution exhibited via cosmic microwave background (CMB) studies and the highly structured web of galaxies we observe today (for a review, see Barkana & Loeb (2001), Furlanetto et al. (2006) and Loeb & Furlanetto (2013)).

One promising technique to probe the EoR is to target the 21 cm wavelength signal that is emitted and absorbed by neutral hydrogen via its spin-flip transition (Furlanetto et al. 2006; Barkana & Loeb 2008; Morales & Wyithe 2010; Pritchard & Loeb 2010; Pritchard & Loeb 2012). This technique is powerful because it can be observed both spatially and as a function of redshift — that is, the wavelength of the signal reaching our telescopes can be directly mapped to a distance from where the emission originated before stretching out as it traveled through expanding space. Hence, 21 cm tomography offers a unique window into both the spatial and temporal evolution of ionization, temperature, and density fluctuations.

In addition to the first tentative detection of our *Cosmic Dawn* (pre-reionization era) made by the Experiment to Detect the Global EoR Signature (EDGES; Bowman et al. 2018; Bowman & Rogers 2010), there are several radio telescope experiments that have succeeded in using the 21 cm signal from hydrogen to place constraints on the brightness of the signal. Examples of experiments investigating the mean brightness temperature of the 21 cm signal relative to the CMB are the Large Aperture Experiment to Detect the Dark Ages (LEDA; Bernardi et al. 2016), the Dark Ages Radio Explorer (DARE; Burns et al. 2012), the Sonda Cosmológica de las Islas para la Detección de Hidrógeno NeutroSciHi (SCI-HI; Voytek et al. 2014), the Broadband Instrument for Global HydrOgen ReioNisation Signal (BIGHORNS; Sokolowski et al. 2015), and the Shaped Antenna measurement of the background RAdio Spectrum (SARAS; Patra et al. 2015). Radio interferometers which seek to measure statistical power spectra include the Giant Metre-wave Radio Telescope (GMRT; Paciga et al. 2013a), the LOw Frequency ARray (LOFAR; van Haarlem et al. 2013), the Murchison Widefield Array (MWA; Tingay et al. 2013), the 21 Centimeter Array (21CMA; Peterson 2004; Wu 2009), the Square Kilometre Array (SKA; Koopmans et al. 2015), and PAPER (Parsons et al. 2010). The Hydrogen Epoch of Reionization Array (HERA), which is currently being built, is a next-generation instrument that aims to combine lessons learned from previous experiments and has a forecasted sensitivity capable of a high-significance power spectrum detection with an eventual 350 elements using current analysis techniques (Pober et al. 2014; Liu & Parsons 2016; Dillon & Parsons 2016; DeBoer et al. 2017).

The major challenge that faces all 21 cm experiments is isolating a small signal that is buried underneath foregrounds and instrumental systematics that are, when combined, four to five orders of magnitude brighter (e.g.,

Santos et al. 2005; Ali et al. 2008; de Oliveira-Costa et al. 2008; Jelić et al. 2008; Bernardi et al. 2009, 2010; Ghosh et al. 2011; Pober et al. 2013; Bernardi et al. 2013; Dillon et al. 2014; Kohn et al. 2016). A clean measurement therefore requires an intimate understanding of how data analysis choices, which are often tailored to maximize sensitivity and minimize contaminants, affect power spectrum results. More specifically, it is imperative to develop techniques that ensure the accurate extraction and recovery of the EoR signal, despite the analysis method chosen and how much loss (of both contaminants and the EoR signal) accompanies the method. In this paper, we specifically discuss signal loss — the loss of the *cosmological* signal — associated with power spectrum estimation. This is an issue that is essential to investigate for a robust 21 cm power spectrum analysis and one that has motivated a revised PAPER analysis. We first approach this topic from a broad perspective, and then perform a detailed case study using data from the 64-element configuration of PAPER. In this study we use a subset of PAPER-64 data to illustrate our revised analysis methods, while a related paper, Kolopanis et al. (*in prep.*), builds off of the methods in this paper to present revised PAPER-64 results for multiple redshifts and baseline types.

Finally, we also highlight several additional errors made in previous PAPER analyses, including those related to bootstrapping and error estimation. This paper adds to the growing foundations of lessons which have been documented, for example, in Paciga et al. (2013b), Patil et al. (2016), and Jacobs et al. (2016), by the GMRT, LOFAR, and MWA projects respectively. These lessons are imperative as the community as a whole moves towards higher sensitivities and potential EoR detections.

This paper is organized into three main sections. In Section 2 we use toy models to develop intuition about signal loss, its origin, and its subtleties. In Section 3, we present a case study using data from the PAPER-64 array, highlighting key changes from the signal loss methods used in the published result in Ali et al. (2015), henceforth known as A15, which previously underestimated loss. In Section 4, we summarize additional lessons learned since A15 that have shaped our revised analysis. Finally, we conclude in Section 5.

2. SIGNAL LOSS TOY MODELS

Signal loss refers to attenuation of the target cosmological signal in a power spectrum estimate. Certain analysis techniques can cause this loss, and if the amount of loss is not quantified accurately, it could lead to false non-detections and overly aggressive upper limits. Determining whether an analysis pipeline is lossy, and estimating the amount of loss if so, has subtle challenges but is necessary to ensure the accuracy of any result.

One type of signal loss can occur when weighting data by itself. Broadly speaking, a dataset can be weighted to emphasize certain features and minimize others. For

example, one flavor of weighting employed by previous PAPER analyses is inverse covariance weighting in frequency, which is a generalized version of inverse variance weighting that also takes into account frequency correlations (Liu & Tegmark 2011; Dillon et al. 2013; Liu et al. 2014a; Liu et al. 2014b; Dillon et al. 2014; Dillon et al. 2015). Using such a technique enables the down-weighting of contaminant modes that obey a different covariance structure from that of cosmological modes. However, a challenge of inverse covariance weighting is in estimating a covariance matrix that is closest to the true covariance of the data; the discrepancy between the two, as we will see, can have large impacts on signal loss. In this paper we focus specifically on loss associated with the use of an empirically estimated covariance matrix with the "optimal quadratic estimator" formalism. This loss was significantly underestimated in the A15 analysis and is the main reason motivating a revised power spectrum result.

2.1. The Quadratic Estimator Method

We begin with an overview of the quadratic estimator (QE) formalism used for power spectrum estimation. The goal of power spectrum analysis is to produce an unbiased estimator of the EoR power spectrum in the presence of both noise and foreground emission. Prior to power spectrum estimation, the data will often have been prepared to have minimal foregrounds by some method of subtraction, so this foreground emission may appear either directly (because it was not subtracted) or as a residual of some subtraction process not in the power spectrum domain. If an accurate estimate of the total covariance of the data is known, including both the desired signal and any contaminants, then the "optimal quadratic estimator" formalism provides a method of producing a minimum variance, unbiased estimator of the desired signal, as shown in Liu & Tegmark (2011), Dillon et al. (2013), Liu et al. (2014a), Liu et al. (2014b), Trott et al. (2012), Dillon et al. (2014), Dillon et al. (2015), Switzer et al. (2015), and Trott et al. (2016).

Suppose that the measured visibilities for a single baseline in Jy are arranged as a data vector, $\mathbf{x}$. It has length $N_t N_f$, where N_t is the number of time integrations and N_f is the number of frequency channels. The covariance of the data is given by

$$\mathbf{C} \equiv \langle \mathbf{x}\mathbf{x}^\dagger \rangle = \mathbf{S} + \mathbf{U} \qquad (1)$$

where the average over an ensemble of data realizations produces the true covariance, and we further assume it may be written as the sum of the desired cosmological signal $\mathbf{S}$ and other terms $\mathbf{U}$.

We are interested in estimating the three-dimensional power spectrum of the EoR. Visibilities are measurements of the Fourier transform of the sky along two spatial dimensions (using the flat-sky approximation), and since we are interested in three-dimensional Fourier modes we only need to take one Fourier transform of our visibilities along the line-of-sight dimension. We consider band powers P^α of the power spectrum of $\mathbf{x}$ over

some range in cosmological $\mathbf{k}$, where α indexes a waveband in $k_\parallel$ (a cosmological wavenumber $k_\parallel$ is the Fourier dual to frequency under the delay approximation (Parsons et al. 2012b), which is a good approximation for the short baselines that PAPER analyzes). The fundamental dependence of the covariance on the power spectrum band powers P^α is encoded as

$$\mathbf{S} = \sum_\alpha P^\alpha \frac{\partial \mathbf{C}}{\partial P^\alpha} \equiv \sum_\alpha P^\alpha \mathbf{Q}^\alpha \qquad (2)$$

where we define $\frac{\partial \mathbf{C}}{\partial P^\alpha} \equiv \mathbf{Q}^\alpha$. In other words, $\mathbf{Q}$ describes the response of the covariance to a change in the power spectrum, relating a quadratic statistic of the data (the covariance) to a quadratic statistic in Fourier-space (the power spectrum).

The optimal quadratic estimator prescription is then to compute

$$\widehat{P}^\alpha = \sum_\beta (\mathbf{F}^{-1})^{\alpha\beta} (\widehat{q}^\beta - \widehat{b}^\beta) \qquad (3)$$

where $\mathbf{F}$ is the Fisher matrix (which determines errors on the power spectrum estimate)

$$F^{\alpha\beta} \equiv \frac{1}{2} \mathrm{tr}\left(\mathbf{C}^{-1}\mathbf{Q}^\alpha \mathbf{C}^{-1}\mathbf{Q}^\beta \right), \qquad (4)$$

$\widehat{q}$ is the un-normalized power spectrum estimate

$$\widehat{q}^\alpha = \frac{1}{2}\mathbf{x}^\dagger \mathbf{C}^{-1}\mathbf{Q}^\alpha \mathbf{C}^{-1}\mathbf{x}, \qquad (5)$$

and $\widehat{b}$ is the additive bias

$$\widehat{b}^\alpha = \frac{1}{2}\,\mathrm{tr}\left(\mathbf{U}\mathbf{C}^{-1}\mathbf{Q}^\alpha \mathbf{C}^{-1} \right). \qquad (6)$$

The power spectrum estimator in Equation (3) is the minimum variance (smallest error bar) estimate of the power spectrum subject to the constraint that it is also unbiased; that is, the ensemble average of the estimator is equal to its true value

$$\langle \widehat{P}^\alpha \rangle = P^\alpha \qquad (7)$$

(Tegmark 1997; Bond et al. 1998).

Intuitively, the estimator must be capable of "suppressing" or "removing" the effects of contaminates in order to obtain an unbiased estimate of the power spectrum. By construction, the subtraction of the residual foreground and noise bias accomplishes this, removing any additive bias. However, the $\mathbf{C}^{-1}$ piece of Equation (5) also has the effect of suppressing residual foregrounds and noise, in both the additive bias and any contributions the residuals may have to the variance. The way in which the optimal estimator accomplishes this is illustrated with a toy model in Appendix A. In this appendix, we show that the effect of the weighting in Equation 5 is to project out the modes of $\mathbf{U}$ with a different covariance structure than $\mathbf{S}$ in the power spectrum estimate, and the effect of Equation 6 is to subtract out the remaining bias. Similar effects for a realistic model of

the EoR and foregrounds are shown in Liu & Tegmark (2011).

The toy model in Appendix A also illustrates that if the covariance structure of the contaminants is sufficiently different from the desired power spectrum, then the linear bias term may be expected to be quite small, and it is only necessary to know $\mathbf{C}$ and $\mathbf{Q}^\alpha$, but not $\mathbf{U}$. Since the foregrounds are expected to be strongly correlated between frequencies whereas the EoR is not, we expect different covariance structures and therefore a small linear bias. Moreover, because the linear bias is always positive and there is no multiplicative bias, the quadratic-only term will always produce an estimate which is *high* relative to the true value, and which can conservatively be interpreted as an upper limit. These considerations, and the difficulty of obtaining an estimate for $\mathbf{U}$, motivate the neglect of the linear bias in the rest of this analysis.

Motivated by the desire to retain the advantageous behavior of suppressing contributions of $\mathbf{U}$ to estimates of the EoR power spectrum, we note that is possible to define a modified version of the quadratic estimator where Equation (5) is replaced by

$$\widehat{q}^\alpha = \frac{1}{2}\mathbf{x}^\dagger \mathbf{R}\mathbf{Q}^\alpha \mathbf{R}\mathbf{x} \tag{8}$$

where $\mathbf{R}$ is a weighting matrix chosen by the data analyst. For example, inverse covariance weighting (the optimal form of QE) would set $\mathbf{R} \equiv \mathbf{C}^{-1}$ and a uniform-weighted case would use $\mathbf{R} \equiv \mathbf{I}$, the identity matrix. Again, the matrix $\mathbf{Q}^\alpha$ encodes the dependence of the covariance on the power spectrum but in practice also also does other things, including implementing a transform of the frequency domain visibilities to $\mathbf{k}$-space, taking into account cosmological scalings, and converting the visibilities from Jansky to Kelvin.

With an appropriate normalization matrix $\mathbf{M}$, the quantity

$$\widehat{\mathbf{P}} = \mathbf{M}\widehat{\mathbf{q}} \tag{9}$$

is a sensible *estimate* of the *true* power spectrum $\mathbf{P}$.

To ensure that $\mathbf{M}$ correctly normalizes our power spectrum, one may take the expectation value of Equation (9) to obtain

$$\langle \widehat{P}^\alpha \rangle = \frac{1}{2}\sum_{\beta\gamma} M^{\alpha\gamma}\mathrm{tr}\left(\mathbf{R}\mathbf{Q}^\gamma \mathbf{R}\mathbf{Q}^\beta\right)P^\beta + \frac{1}{2}\sum_\gamma \mathrm{tr}\left(\mathbf{U}\mathbf{R}\mathbf{Q}^\gamma \mathbf{R}\right)$$
$$\equiv \sum_\beta W^{\alpha\beta}P^\beta + \frac{1}{2}\sum_\gamma \mathrm{tr}\left(\mathbf{U}\mathbf{R}\mathbf{Q}^\gamma \mathbf{R}\right), \tag{10}$$

where $W^{\alpha\beta}$ are elements of a window function matrix. Considering the first term of this expression (again, we are assuming that the linear bias term is significantly suppressed; and if this is not the case, we are simply assuming that we are setting a conservative upper limit), if $\mathbf{W}$ ends up being the identity matrix for our choices of $\mathbf{R}$ and $\mathbf{M}$, then we recover Equation (7) for the first term, and we have an estimator that has no multiplicative matrix bias. However, Equation (7) is a rather restrictive condition, and it is possible to violate it and still have a sensible (and correctly normalized) power spectrum estimate. In particular, as long as the rows of $\mathbf{W}$ sum to unity, our power spectrum will be correctly normalized. Beyond this, the data analyst has a choice for $\mathbf{M}$, and for simplicity throughout this paper we choose $\mathbf{M}$ to be diagonal. In a preview of what is to come, we also stress that the derivation that leads to Equation (10) assumes that $\mathbf{R}$ and $\mathbf{x}$ are not correlated. If this assumption is violated, a simple application of the (now incorrect) formulae in this section can result in an improperly normalized power spectrum estimator that does not conserve power, i.e., one that has signal loss.

Given the advantages of inverse covariance weighting, a question arises of how one goes about estimating $\mathbf{C}$. One method is to empirically derive it from the data $\mathbf{x}$ itself. Similar types of weightings that are based on variance information in data are done in Chang et al. (2010) and Switzer et al. (2015). In previous PAPER analyses, one time-averages the data to obtain:

$$\widehat{\mathbf{C}} \equiv \langle \mathbf{x}\mathbf{x}^\dagger \rangle_t \approx \langle \mathbf{x}\mathbf{x}^\dagger \rangle, \tag{11}$$

assuming $\langle \mathbf{x} \rangle_t = 0$ (a reasonable assumption since fringes average to zero over a sufficient amount of time), where $\langle \rangle_t$ denotes a finite average over time. The weighting matrix for our empirically estimated inverse covariance weighting is then $\mathbf{R} \equiv \widehat{\mathbf{C}}^{-1}$, where we use a hat symbol to distinguish the empirical covariance from the true covariance $\mathbf{C}$.

In the next three sections, we use toy models to investigate the effects of weighting matrices on signal loss by experimenting with different matrices $\mathbf{R}$ and examining their impact on the resulting power spectrum estimates $\widehat{\mathbf{P}}$. Our goal in experimenting with weighting is to suppress foregrounds and investigate EoR losses associated with it. We note that we purposely take a thorough and pedagogical approach to describing the toy model examples given in the next few sections. The specifics of how signal loss appears in PAPER's analysis is later described in Section 3.

As a brief preview, we summarize our findings in the following sections here:

- If the covariance matrix is estimated from the data, a strong correlation between the estimated modes and the data will in general produce an estimate of the signal power spectrum which is strongly biased *low* relative to the true value. In this context, this is what we call "signal loss" (Section 2.2).

- The effect of the bias is worsened when the number of independent samples used to estimate the covariance matrix is reduced (Section 2.3).

- The rate at which empirical eigenvectors converge to their true forms depends on the sample variance

in the empirical estimate and the shape of the empirical eigenspectrum. In general, larger sample variances lead to more loss (Section 2.3).

- Knowing these things, there are some simple ways of altering the empirical covariance matrix to decouple it from the data and produce unbiased power spectrum estimates (Section 2.4).

2.2. *Empirical Inverse Covariance Weighting*

Using a toy model, we will now build intuition into how weighting by the inverse of the empirically estimated covariance, $\widehat{\mathbf{C}}^{-1}$, can give rise to signal loss. We construct a simple dataset that contains visibility data with 100 time integrations and 20 frequency channels. This model represents realistic dimensions of about an hour of PAPER data which might be used for a power spectrum analysis. For PAPER-64 (both the A15 analysis and our new analysis) we use ~ 8 hours of data (with channel widths of 0.5 MHz and integration times of 43 seconds), but here we scale it down with no loss of generality.

We create mock visibilities, $\mathbf{x}$, and assume a non-tracking, drift-scan observation. Hence, flat spectrum sources (away from zenith) lead to measured visibilities which oscillate in time and frequency. We therefore form a mock visibility measurement of a bright foreground signal, $\mathbf{x}_{\mathrm{FG}}$, as a complex sinusoid that varies smoothly in time and frequency, a simplistic but realistic representation of a single bright source. We also create a mock visibility measurement of an EoR signal $\mathbf{x}_{\mathrm{EoR}}$ as a complex, Gaussian random signal. A more realistic EoR signal would have a sloped power spectrum in $p(k)$ (instead of flat, as in the case of white noise), which could be simulated by introducing frequency correlations into the mock EoR signal. However, here we treat all k's separately, so a simplistic white noise approximation can be used. Our combined data vector is then $\mathbf{x} = \mathbf{x}_{\mathrm{FG}} + \mathbf{x}_{\mathrm{EoR}}$, to which we apply different weighting schemes throughout Section 2. The three data components are shown in Figure 1.

We compute the power spectrum of our toy model dataset $\mathbf{x}$ using Equations 8 and 9, with $\mathbf{R} \equiv \widehat{\mathbf{C}}^{-1}$. Figure 2 shows the estimated covariances of our toy model datasets along with the $\widehat{\mathbf{C}}^{-1}$ weighted data. The foreground sinusoid is clearly visible in $\widehat{\mathbf{C}}_{\mathrm{FG}}$. The power spectrum result is shown in green in the left plot of Figure 3. Also plotted in the figure are the uniform-weighted ($\mathbf{R} \equiv \mathbf{I}$) power spectrum of the individual components $\mathbf{x}_{\mathrm{FG}}$ (blue) and $\mathbf{x}_{\mathrm{EoR}}$ (red). As shown, our $\widehat{\mathbf{C}}^{-1}$ weighted result successfully suppresses foregrounds, demonstrated in Figure 3 by the missing foreground peak in the weighted power spectrum estimate (green). It is also evident that our result fails to recover the EoR signal — it exhibits the correct shape, but the amplitude level is slightly low. It is this behavior which we describe as signal loss.
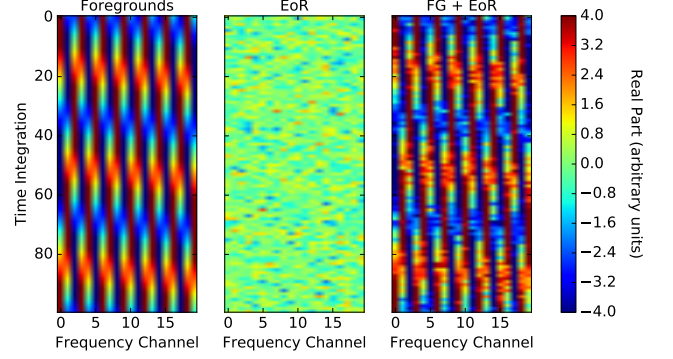

Figure 1. Our toy model dataset to which we apply different weighting schemes to in order to investigate signal loss. We model a mock foreground-only visibility with a sinusoid signal that varies smoothly in time and frequency. We model a mock visibility of an EoR signal as a random Gaussian signal. We add the two together to form $\mathbf{x} = \mathbf{x}_{\mathrm{FG}} + \mathbf{x}_{\mathrm{EoR}}$. Real parts are shown here.
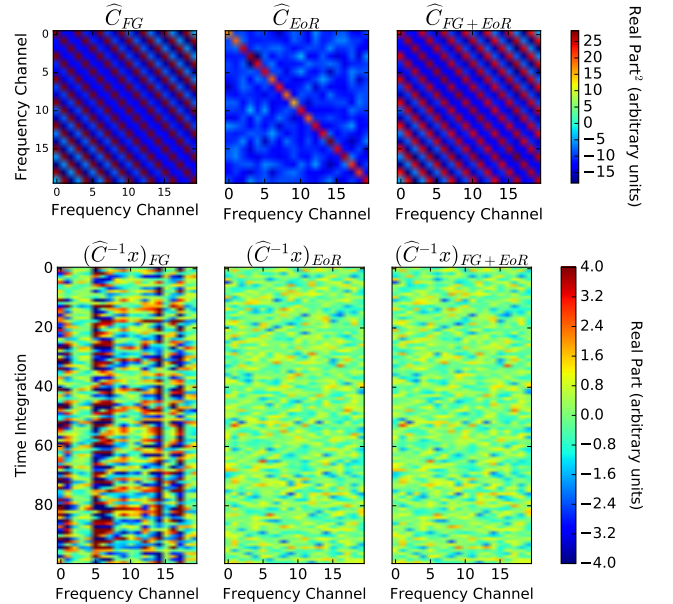
Figure 2. The estimated covariance matrices (top row) and inverse covariance-weighted data (bottom row) for FG only (left), EoR only (middle), and FG + EoR (right). Real parts are shown here.

As discussed in Section 2.1, this behavior is *not* expected in the case that we were to use a true $\mathbf{C}^{-1}$ weighting. Rather, we would obtain the behavior shown in the toy model in Appendix A[1], with suppression of the foreground mode resulting in a nearly unbiased estimate of the power spectrum. The key difference is that since $\widehat{\mathbf{C}}$ is estimated from the data, its eigenvectors and eigenvalues are strongly coupled to the particular data realization that was used to compute it, and this coupling

---

[1] Note that there the true covariance matrix is also the sum of a diagonal portion describing the signal, and a single mode describing the contaminant (similar to Figure 2).
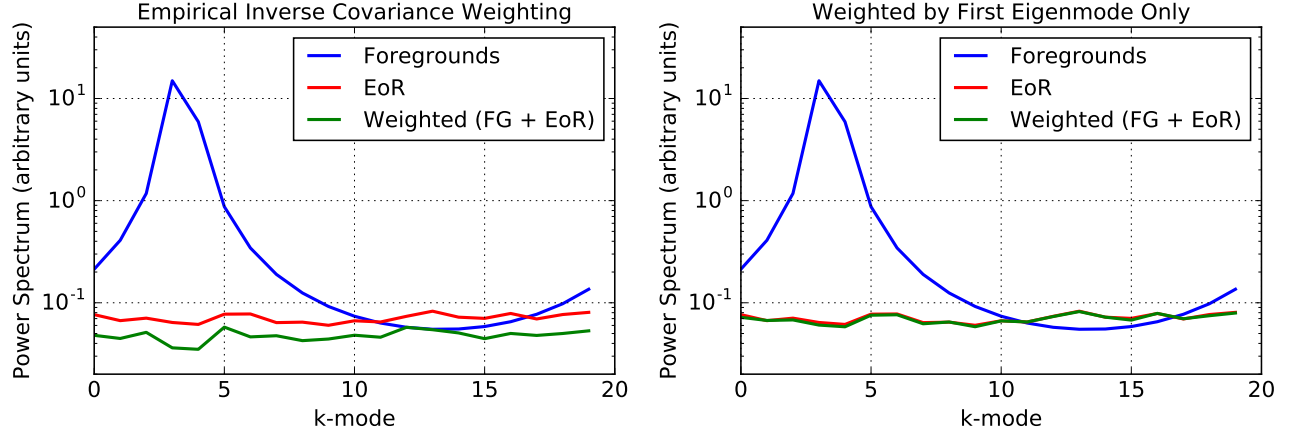
Figure 3. Resulting power spectrum estimates for the toy model simulation described in Section 2.2 — foregrounds only (blue), EoR only (red), and the weighted FG + EoR dataset (green). The power spectrum of the foregrounds peaks at a k-mode based on the frequency of the sinusoid used to create the mock FG signal. In the two panels, we compare using empirically estimated inverse covariance weighting where $\mathbf{C}$ is derived from the data (left), and projecting out the zeroth eigenmode only (right). In the former case, signal loss arises from the coupling of the eigenmodes of $\widehat{\mathbf{C}}$ to the data. There is negligible signal loss when all eigenmodes besides the foreground one are no longer correlated with the data.

leads to loss.

For the case of an eigenmode which can be safely assumed to be predominantly a foreground, its presence in the true covariance matrix will result in the desired suppression via a kind of projection (Equations A3 and A4); whether or not it is strongly correlated with the the actual data vector is irrelevant. However, in the case of an empirically estimated covariance matrix, the eigenmodes of $\widehat{\mathbf{C}}_{\mathrm{EoR}}$ will both be incorrect and can be correlated with the data. If these incorrect eigenmodes are not correlated with the data, it will lead to non-minimum variance estimates but will not produce the suppression of the power spectrum amplitude as seen in the left plot of Figure 3. As described in Section 3.1 and shown mathematically in Appendix B, however, if $\widehat{\mathbf{C}}_{\mathrm{EoR}}$ *is* correlated with the data vector $\mathbf{x}$, there is a kind of projection of power in the *non*-foreground modes from the resulting power spectrum estimate, thus producing an estimate that is biased low. In short, *if the covariance is computed from the data itself, it carries the risk of overfitting information in the data and introducing a multiplicative bias (per k) to estimates of the signal.*

The danger of an empirically estimated covariance matrix comes mostly from not being able to describe the EoR-dominated eigenmodes of $\mathbf{C}$ accurately, for which the EoR signal is brighter than foregrounds. In such a case, the coupling between these modes to the data realization leads to the overfitting and subtraction of the EoR signal. More specifically, the coupling between the estimated covariance and the data is anti-correlated in nature (which is explained in more detail in Section 3.1 and Appendix B), which leads to loss. Mis-estimating $\mathbf{C}$ for EoR-dominated eigenmodes is therefore more harmful than for FG-dominated modes, and since the lowest-valued eigenmodes of an eigenspectrum are typically EoR-dominated, using this part of the spectrum for weighting is most dangerous.

Armed with this information, we can tweak the covariance in a simple way to suppress foregrounds and yield minimal signal loss. Recall that our toy model foreground can be perfectly described by a single eigenmode. Using the full dataset's (foreground plus EoR signal) empirical covariance, we can project out the zeroth eigenmode and then take the remaining covariance to be the identity matrix. This decouples the covariance from the data for the EoR modes. The resulting power spectrum estimate for this case is shown in the right plot of Figure 3. In this case we recover the EoR signal, demonstrating that if we can disentangle the foreground-dominated modes and EoR-dominated modes, we can suppress foregrounds with negligible signal loss.

Altering $\widehat{\mathbf{C}}$ as such is one specific example of a regularization method for this toy model, in which we are changing $\widehat{\mathbf{C}}$ in a way that reduces its coupling to the data realization. There are several other simple ways to regularize $\widehat{\mathbf{C}}$, and we will discuss some in Section 2.4.

2.3. *Fringe-Rate Filtering*

We have shown how signal loss can arise due to the coupling of EoR-dominated eigenmodes to the data. We will next show how this effect is exacerbated by reducing the total number of independent samples in a dataset.

A fringe-rate filter is an analysis technique designed to maximize sensitivity by integrating in time (Parsons et al. 2016). Rather than a traditional box-car average in time, a time domain filter can be designed to up-weight temporal modes consistent with the sidereal motion on the sky, while down-weighting modes that are noise-like.

Because fringe-rate filtering is analogous to averaging in time, it comes at the cost of reducing the total number of independent samples in the data. With fewer independent modes, it becomes more difficult for the empirical covariance to estimate the true covariance matrix of the fringe-rate filtered data. We can quantify this

effect by evaluating a convergence metric $\varepsilon(\widehat{\mathbf{C}})$ for the empirical covariance, which we define as

$$\varepsilon(\widehat{\mathbf{C}}) \equiv \sqrt{\frac{\sum_{ij}(\widehat{C}_{ij} - C_{ij})^2}{\sum_{ij} C_{ij}^2}}, \qquad (12)$$

where $\mathbf{C}$ is the true covariance matrix. To compute this metric, we draw different numbers of realizations (different draws of Gaussian noise) of our toy model EoR measurement, $\mathbf{x}_{\mathrm{EoR}}$, and take their ensemble average. We then compare this to the "true" covariance, which in our simulation is set to be the empirical covariance after a large number (500) of realizations. As shown in Figure 4, we perform this computation for a range of total independent ensemble realizations (horizontal axis) and number of independent samples in the data following time-averaging, or "fringe-rate filtering" (different colors). With more independent time samples (i.e., more realizations) in the data, one converges to the true fringe-rate filtered covariance more quickly.

The situation here with using a finite number of time samples to estimate our covariance is analogous to a problem faced in galaxy surveys, where the non-linear covariance of the matter power spectrum is estimated using a large — but finite — number of expensive simulations. There, the limited number of independent simulations results in inaccuracies in estimated covariance matrices (Dodelson & Schneider 2013; Taylor & Joachimi 2014), which in turn result in biases in the final parameter constraints (Hartlap et al. 2007). In our case, the empirically estimated covariances are used for estimating the power spectrum, and as we discussed in the previous section (and will argue more thoroughly in Section 3.1 and Appendix B), couplings between these covariances and the data can lead to power spectrum estimates that are biased *low*—which is precisely signal loss. In future work, it will be fruitful to investigate whether advanced techniques from the galaxy survey literature for estimating accurate covariance matrices can be successfully adapted for 21 cm cosmology. These techniques include the imposition of sparsity priors (Padmanabhan et al. 2016), the fitting of theoretically motivated parametric forms (Pearson & Samushia 2016), covariance tapering (Paz & Sánchez 2015), marginalization over the true covariance (Sellentin & Heavens 2016), and shrinkage methods (Pope & Szapudi 2008; Joachimi 2017).

The overall convergence of the covariance is important, but also noteworthy is the fact that different eigenvectors converge to their true forms at different rates. This is illustrated by Figure 5, which shows the convergence of eigenvectors in an empirical estimate of a covariance matrix. For this particular toy model, we construct a covariance whose true form combines the same mock foreground from the previous toy models with an EoR component that is modeled as a diagonal matrix with eigenvalues spanning one order of magnitude (more specifically, we construct the EoR covariance as a diagonal matrix in the Fourier domain, where the signal is ex-

pected to be uncorrelated; its Fourier transform is then the true covariance of the EoR in the frequency domain, or $\mathbf{C}_{\mathrm{EoR}}$). For different numbers of realizations, we draw random EoR signals that are consistent with $\mathbf{C}_{\mathrm{EoR}}$, add them to the mock foreground data, and compute the combined empirical covariance by averaging over the realizations. The eigenvectors of this empirical covariance are then compared to the true eigenvectors $\widehat{\mathbf{v}}$, where we use as a convergence metric $\varepsilon(\widehat{\mathbf{v}})$, defined as:

$$\varepsilon(\widehat{\mathbf{v}}) \equiv \sqrt{\sum_{i}^{N_f} |\mathbf{v} - \widehat{\mathbf{v}}|_i^2}, \qquad (13)$$

where N_f is the number of frequencies (20) in the mock data. The eigenmode convergence curves in Figure 5 are ranked ordered by eigenvalue, such that "Eigenmode #0" illustrates the convergence of the eigenvector with the largest eigenvalue, "Eigenmode #1" for the second largest eigenvalue, and so on. We see that the zeroth eigenmode — the mode describing the foreground signal — is quickest to converge.

Our numerical test reveals that the convergence rates of empirical eigenvectors is related to the sample variance in our empirical estimate. In general, computing an empirical covariance from a finite ensemble average means that the empirical eigenmodes have sample variances. Consider first a limiting case where all eigenvalues are equal. In such a scenario, any linear combination of eigenvectors is also an eigenvector, and thus there is no sensible way to define the convergence of eigenvectors. In our current test, aside from the zeroth mode, the eigenvalues have similar values but are not precisely equal. Hence, there is a well-defined set of eigenvectors to converge to. However, due to the sample variance of our empirical covariance estimate, there may be accidental degeneracies between modes, where some modes are mixing and swapping with others. Therefore, the steeper an eigenspectrum, the easier it is for the eigenmodes to decouple from each other and approach their true forms. A particularly drastic example of this can be seen in the behavior of mode 0 (the foreground mode), whose eigenvalue differs enough from the others that it is able to converge reasonably quickly despite substantial sample variance in our empirical covariance estimate. To break degeneracies in the remaining modes, however, requires many more realizations.

While the connection between the rate of convergence of an empirical eigenvector with the sample variance of an eigenspectrum is interesting, it is also important to note that regardless of convergence rate, any mode that is coupled to the data is susceptible to signal loss. The true eigenvectors are not correlated with the data realizations; thus, if our empirical eigenvectors are converged fully, there will not be any signal loss. However, an unconverged eigenvector estimate will retain some memory of the data realizations used in its generation, leading to signal loss.

In the toy models throughout Section 2, we exploit
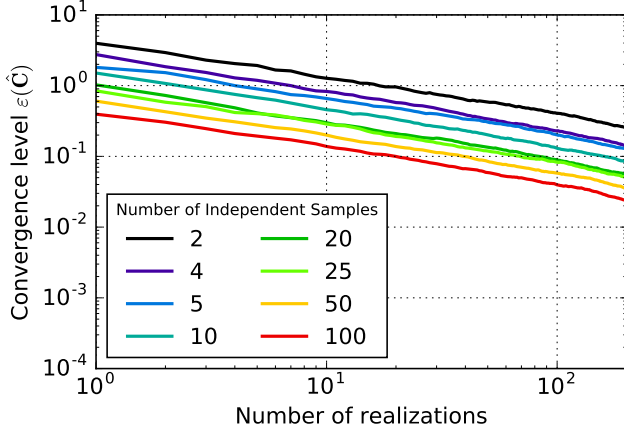
Figure 4. The convergence level, as defined by Equation (12), of empirically estimated covariances of mock EoR signals with different numbers of independent samples. In red, the mock EoR signal is comprised entirely of independent samples. Subsequent colors show time-averaged signals. As the number of realizations increases, we see that the empirical covariances approach the true covariances. With more independent samples, the quicker an empirical covariance converges (i.e., the quicker it decouples from the data), and the less signal loss we would expect to result.

the fact that the strongest eigenmode (highest eigenvalue mode) is dominated by foregrounds in order to purposely incur signal loss for that mode. Even for the case of real PAPER data (Section 3), we make the assumption that the strongest eigenmodes are likely the most contaminated by foregrounds. However, in general, foregrounds need not be restricted to the strongest eigenmodes, and as we have seen, it is really the degeneracies between modes that determines how quickly they converge, and hence how much signal loss can result.

With Figures 4 and 5 establishing the connection between convergence rates (of empirical covariances and eigenvectors) and number of realizations, we now turn back to our original toy model used in Section 2.2, which is comprised of a mock foreground and mock EoR signal. We mimic a fringe-rate filter by averaging every four time integrations of our toy model dataset together, yielding 25 independent samples in time (Figure 6). We choose these numbers so that the total number of independent samples is similar to the number of frequency channels — hence our matrices will be full rank. We use this "fringe-rate filtered" mock data for the remainder of Section 2.

The power spectrum results for this model are shown in Figure 7, and as expected there is a much larger amount of signal loss for this time-averaged dataset since we do a worse job estimating the true covariance. In addition, as a result of having fewer independent samples, we obtain an estimate with more scatter. This is evident by noticing that the green curve in Figure 7 fails to trace the shape of the uniform-weighted EoR power spectrum (red). It is worth noting that a hybrid method, where the empirical covariance is calculated using the original dataset (not fringe-rate filtered) but applied to the fil-
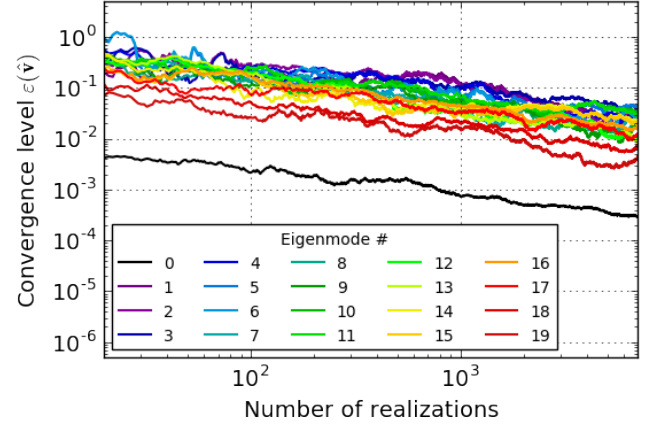
Figure 5. The convergence level, as defined by Equation (13), of empirically estimated eigenvectors for different numbers of mock data realizations. The colors span from the 0th eigenmode (has the highest eigenvalue) to the 19th eigenmode (has the lowest eigenvalue), where they are ordered by eigenvalue in descending order. This figure shows that the zeroth eigenmode converges the quickest, implying that eigenvectors with eigenvalues that are substantially different than the rest (the FG-dominated mode has a much higher eigenvalue than the EoR modes) are able to converge to the true eigenvectors the quickest. On the other hand, eigenmodes 1-19 have similar eigenvalues and are slower to converge because of degeneracies between them.
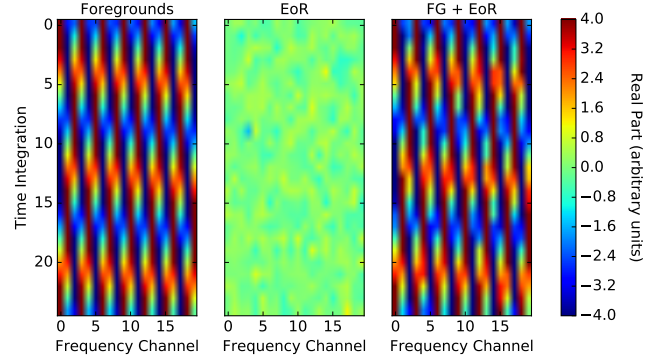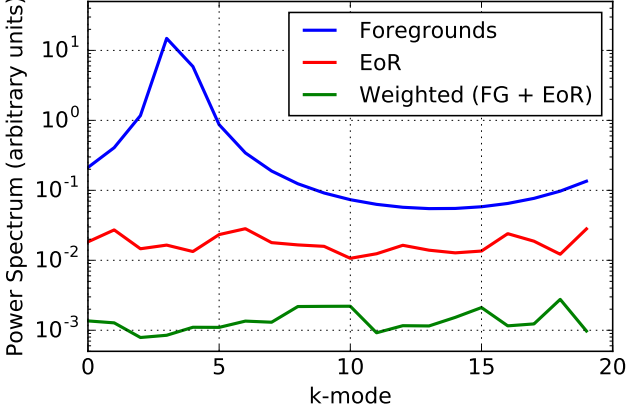

Figure 6. Our "fringe-rate filtered" (time-averaged) toy model dataset. We average every four samples together, yielding 25 independent samples in time. Real parts are shown here.

tered dataset, results in less loss than what is exhibited in Figure 7 (though still lossy). This approach is worth exploring in future work; however in this paper we focus solely on ways to calculate and minimize loss associated with weighting a dataset based on itself (rather than weights associated with a dataset produced by a separate analysis).

Using our toy model, we have seen that a sensitivity-driven analysis technique like fringe-rate filtering has trade-offs of signal loss and noisier estimates when using data-estimated covariance matrices. Longer integrations increase sensitivity but reduce the number of independent samples, resulting in eigenmodes correlated with the data that can overfit signal greatly. We note that a fringe-rate filter does have a range of benefits, many

Figure 7. Resulting power spectrum estimate for the "fringe-rate filtered" (time-averaged) toy model simulation — foregrounds only (blue), EoR only (red), and the weighted FG + EoR dataset (green). We use empirically estimated inverse covariance weighting where $\mathbf{C}$ is computed from the data. There is a larger amount of signal loss than for the non-averaged data, a consequence of weighting by eigenmodes that are more strongly coupled to the data due to there being fewer independent modes in the data.

described in Parsons et al. (2016), so it can still be advantageous to use one despite the trade-offs.

2.4. *Other Weighting Options*

In Section 2.2 we showed one example of how altering $\widehat{\mathbf{C}}$ can make the difference between nearly zero and some signal loss. We will now use our toy model to describe several other ways to tailor $\widehat{\mathbf{C}}$ in order to minimize signal loss. We choose four independent regularization methods to highlight in this section, which have been chosen due to their simplicity in implementation and straightforward interpretations. We illustrate the resulting power spectra for the different cases in Figure 8. These examples are not meant to be taken as suggested analysis methods but rather as illustrative cases.

As a first test, we model the covariance matrix of EoR as a proof of concept that if perfect models are known, signal loss can be avoided. We know that our simulated EoR signal should have a covariance matrix that mimics the identity matrix, with its variance encoded along the diagonal. We model $\mathbf{C}_{\mathrm{EoR}}$ as such (i.e., the identity), instead of computing it based on $\mathbf{x}_{\mathrm{EoR}}$ itself. Next, we add $\mathbf{C}_{\mathrm{EoR}} + \widehat{\mathbf{C}}_{\mathrm{FG}}$ (where $\widehat{\mathbf{C}}_{\mathrm{FG}} = \langle \mathbf{x}_{\mathrm{FG}} \mathbf{x}_{\mathrm{FG}}^{\dagger} \rangle_t$) to obtain a final $\widehat{\mathbf{C}}_{\mathrm{reg}}$ (regularized empirical covariance matrix) to use in weighting. In Figure 8 (upper left), we see that there is negligible signal loss. This is because by modeling $\mathbf{C}_{\mathrm{EoR}}$, we avoid overfitting EoR fluctuations in the data that our model doesn't know about (but, an empirically derived $\widehat{\mathbf{C}}_{\mathrm{EoR}}$ would know about the fluctuations). In practice such a weighting option is not feasible, as it is difficult to model $\mathbf{C}_{\mathrm{EoR}}$, and $\widehat{\mathbf{C}}_{\mathrm{FG}}$ is unknown because we do not know how to separate out the foregrounds from the EoR in our data.

The second panel (top right) in Figure 8 uses a regularization method of setting $\widehat{\mathbf{C}}_{\mathrm{reg}} \equiv \widehat{\mathbf{C}} + \gamma\mathbf{I}$, where $\gamma = 5$

(an arbitrary strength of $\mathbf{I}$ for the purpose of this toy model). By adding the identity matrix, element-wise, we are weighting the diagonal elements of the estimated covariance matrix more heavily than those off-diagonal. Since the identity component does not know anything about the data realization, it alters the covariance to be less coupled to the data and there is no loss.

The third panel (bottom left) in Figure 8 minimizes signal loss by only using the first three eigenmodes of the estimated covariance. Recalling that our toy model foregrounds can be described entirely by the zeroth eigenmode, this method intentionally projects out the highest-valued modes only by replacing all but the three highest weights in the eigenspectrum with 1's (equal weights). Again, avoiding the overfitting of EoR-dominated modes which are coupled to the data results in negligible signal loss. While this case is illuminating for the toy model, in practice it is not obvious which eigenmodes are foreground or EoR dominated (and they could be mixed as well), so determining which subset of modes to down-weight is not trivial. We experiment with this idea using PAPER data in Section 3.3.

The last regularization scheme we are highlighting here is setting $\widehat{\mathbf{C}}_{\mathrm{reg}} \equiv \widehat{\mathbf{C}} \circ \mathbf{I}$ (element-wise multiplication), or inverse variance weighting (i.e., keeping only the diagonal elements of $\widehat{\mathbf{C}}$). In the bottom right panel of Figure 8, we see that this method does not down-weight the foregrounds at all — this regularization altered $\widehat{\mathbf{C}}$ in a way where it is no longer coupled to *any* of the empirically estimated eigenmodes, including the FG-dominated one. To understand this, we recall that our foregrounds are spread out in frequency and therefore have non-negligible frequency-frequency correlations. Multiplying by the identity matrix, element-wise, results in a diagonal matrix, meaning we do not have any correlation information. Because of this, we do a poor job suppressing the foreground. But because we decoupled the whole eigenspectrum from the data, we also avoid signal loss. Although this method did not successfully recover the EoR signal for this particular simulation, it is important that we show that there are many options for estimating a covariance matrix, and some may down-weight certain eigenmodes more effectively than others based on the spectral nature of the components in a dataset.

In summary, we have shown how signal loss is caused by weighting a dataset by itself, and in particular how estimated covariances can overfit EoR modes when they are coupled to data and not converged to their true forms. We have also seen that there are trade-offs between a chosen weighting method, its foreground-removal effectiveness, the number of independent samples in a dataset, and the amount of resulting signal loss.

3. SIGNAL LOSS IN PAPER-64

We now turn to a detailed signal loss investigation using a subset of the PAPER-64 dataset from A15. In
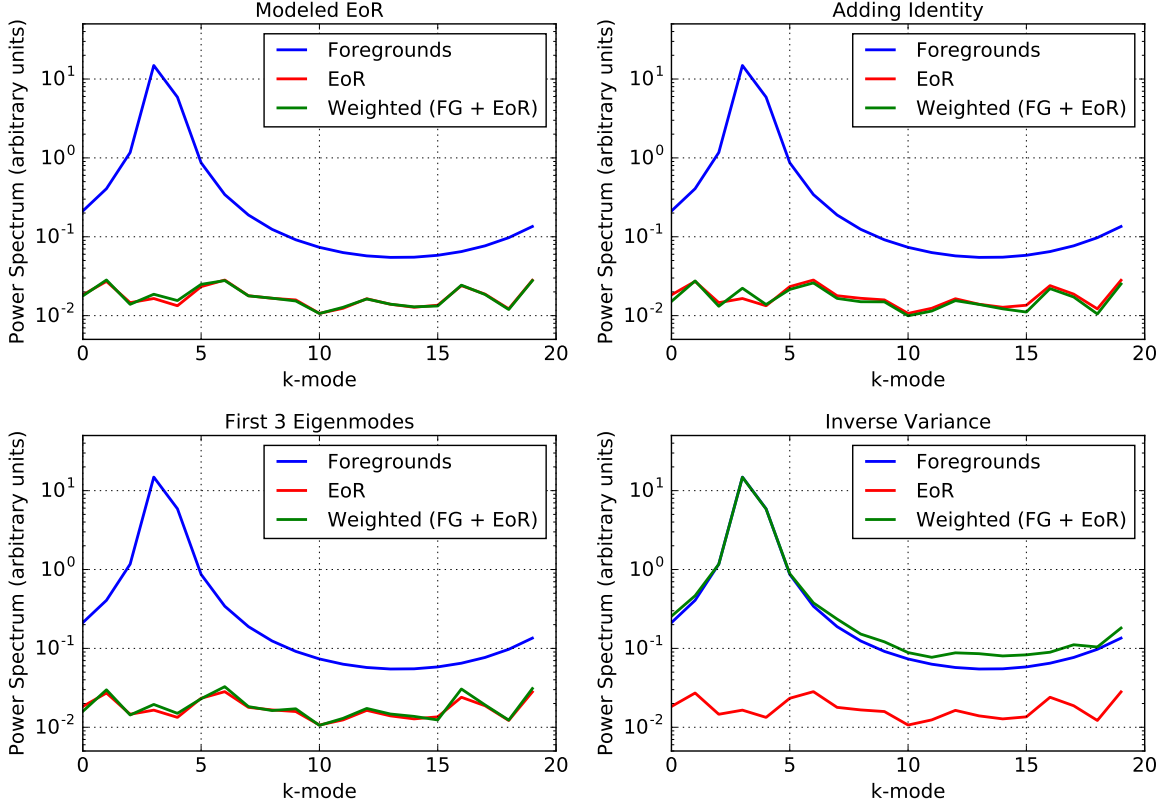
Figure 8. Resulting power spectra estimates for our "fringe-rate filtered" (time-averaged) toy model simulation — foregrounds only (blue), EoR only (red), and the weighted FG + EoR dataset (green). We show four alternate weighting options that each minimize signal loss, including modeling the covariance matrix of EoR (upper left), regularizing $\widehat{\mathbf{C}}$ by adding an identity matrix to it (upper right), using only the first three eigenmodes of $\widehat{\mathbf{C}}$ (lower left), and keeping only the diagonal elements of $\widehat{\mathbf{C}}$ (lower right). The first case (upper left) is not feasible in practice since we do not know $\mathbf{C}_{\mathrm{FG}}$ and $\mathbf{C}_{\mathrm{EoR}}$ like we do in the toy model.

the previous section we showed how signal loss arises when weighting data with empirically estimated covariances; in this section we highlight how the amount of this loss was underestimated in the previous analysis. Additionally, we illustrate how we have revised our analysis pipeline in light of our growing understandings.

As a brief review, PAPER is a dedicated 21 cm experiment located in the Karoo Desert in South Africa. The PAPER-64 configuration consists of 64 dual-polarization drift-scan elements that are arranged in a grid layout. For our case study, we focus solely on Stokes I estimated data (Moore et al. 2013) from PAPER's 30 m East/West baselines (Figure 9). All data is compressed, calibrated (using self-calibration and redundant calibration), delay-filtered (to remove foregrounds inside the wedge), LST-binned, and fringe-rate filtered. For detailed information about the backend system of PAPER-64, its observations, and data reduction pipeline, we refer the reader to Parsons et al. (2010) and A15. We note that all data processing steps are identical to those in A15 until after the LST-binning step in Figure 3 of A15.

The previously best published 21 cm upper limit result from A15 placed a 2σ upper limit on $\Delta^2(k)$, defined as

$$\Delta^2(k) = \frac{k^3}{2\pi^2}\,\hat{P}(k), \qquad (14)$$
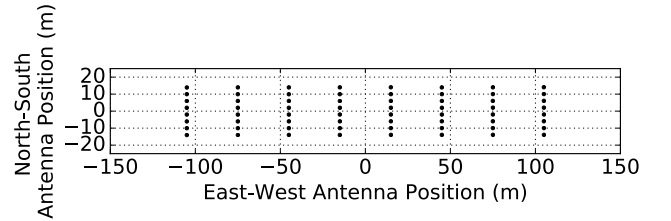
Figure 9. The PAPER-64 antenna layout. We use only 10 of the 30 m East/West baselines for the analysis in this paper (i.e., a subset of the shortest horizontal spacings).

of $(22.4\ \mathrm{mK})^2$ in the range $0.15 < k < 0.5\,h\ \mathrm{Mpc}^{-1}$ at $z = 8.4$. The need to revise this limit stems mostly from previously underestimated signal loss, which we address in this section.

For the analysis in this paper, we use 8.1 hours of LST, namely an RA range of 0.5-8.6 hours (A15 uses a slightly longer RA range of 0-8.6 hours; we found that some early LSTs were more severely foreground contaminated). We also use only 10 baselines, a subset of the 51 total East/West baselines used in A15, in order to illustrate our revised methods. All power spectrum results are produced for a center frequency of 151 MHz using a width of 10 MHz (20 channels), identical to the analysis
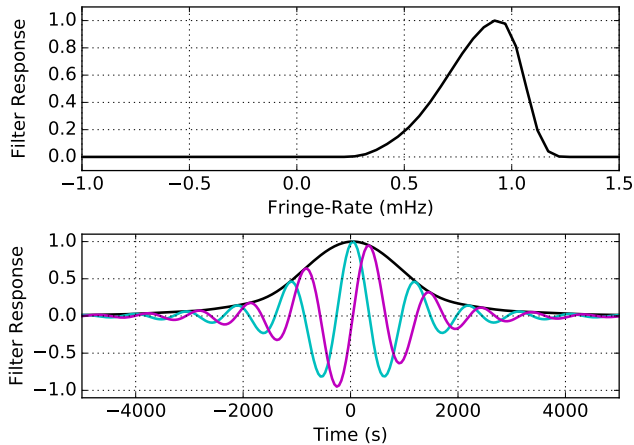
Figure 10. Top: the normalized optimal power-spectrum sensitivity weighting in fringe-rate space for our fiducial baseline and Stokes I polarization beam. Bottom: the time domain convolution kernel corresponding to the top panel. Real and imaginary components are illustrated in cyan and magenta, respectively, with the absolute amplitude in black. The fringe-rate filter acts as an integration in time, increasing sensitivity but reducing the number of independent samples in the dataset.

in A15. In the case study in this paper, we only use one baseline type instead of the three as in A15, but Kolopanis et al. (*in prep.*) uses the full dataset presented in A15 to revise the result and place limits on the EoR at multiple redshifts (using a straightforward and not lossy approach to avoid many of the issues that will be made clear later on).

The most significant changes from A15 occur in our revised power spectrum analysis, which is explained in the rest of this paper, but we also note that the applied fringe-rate filter is also slightly different. In A15, the applied filter was not equivalent to the optimal fringe-rate filter (which is designed to maximize power spectrum sensitivity). Instead, the optimal filter was degraded slightly by widening it in fringe-rate space. This was chosen in order to increase the number of independent modes and reduce signal loss associated with the quadratic estimator, though as we will explain in the next section, this signal loss was still underestimated. With the development of a new, robust method for assessing signal loss, we choose to use the optimal filter in order to maximize sensitivity. This filter is computed for a fiducial 30 m baseline at 150 MHz, the center frequency in our band. The filter in both the fringe-rate domain and time domain is shown in Figure 10.

Finally, we emphasize that the discussion that follows is solely focused on signal loss associated with empirical covariance weighting. As mentioned in Section 2, there are several other ways the PAPER analysis pipeline can lead to loss, including through calibration and delay-filtering. We refer the reader to Parsons et al. (2010) and A15 for discussions about these other types of signal loss, noting here that they are much less significant than the loss focused on in this paper.

We present our PAPER-64 signal loss investigation

in three parts. We first give an overview of our signal injection framework which is used to estimate loss. In this framework (and as in A15), we inject simulated cosmological signals into our data and test the recovery of those signals (an approach also taken by Masui et al. (2013)). As we will see, correlations between the injected signals and the data are significant complicating factors which were previously not taken into account. Next, we describe our methodology in practice and detail how we map our simulations into a posterior for the EoR signal. Finally, we build off of the previous section by experimenting with different regularization schemes on PAPER data in order to minimize loss. Throughout each section, we also highlight major differences from the signal loss computation used in A15.

3.1. *Signal Loss Methodology*

In short, our method for estimating signal loss consists of adding an EoR-like signal into visibility data and then measuring how much of this injected signal would be detectable given any attenuation of this signal by the (lossy) data analysis pipeline. To capture the full statistical likelihood of signal loss, one requires a quick way to generate many realizations of simulated 21 cm signal visibilities. Here we use the same method as in A15, where mock Gaussian noise visibilities (mock EoR signals) are filtered in time using an optimal fringe-rate filter to retain only "sky-like" modes. Since the optimal filter has a shape that matches the rate of the sidereal motion of the sky, this transforms the Gaussian noise into a measurement that PAPER could make. This signal is then added to the visibility data.[2]

Mathematically, suppose that **e** is the mock injected EoR signal (at some amplitude level). We do not know the true EoR signal contained within our visibility data, **x**, so **e** takes on the role of the true EoR signal (for which we measure its loss). Furthermore, one can make the assumption that the true EoR signal is small within our measured data, so the data vector **x** itself is representative of mostly contaminants. Using this assumption, the sum of **x** and **e**, defined as **r**:

$$\mathbf{r} = \mathbf{x} + \mathbf{e}, \tag{15}$$

can be thought of as the sum of contaminants plus EoR. The quantity **r** then becomes the dataset for which we are measuring how much loss of **e** there is due to our power spectrum pipeline.

---

[2] One specific change from A15 is that we add this simulated signal - which has been fringe-rate filtered once already in order to transform it into a "sky-like" signal - into the analysis pipeline before a fringe-rate filter is applied to the data (i.e., prior to the analysis step of fringe-rate filtering). Previously, the addition was done after the fringe-rate filter analysis step. This change results in an increased estimate of signal loss, likely due to the use of the fringe-rate filter as a simulator. However, this pipeline difference, while significant, is not the dominant reason why signal loss was underestimated in A15 (the dominant reason is explained in the main text in Section 3.1).

We are interested in quantifying how much variance in $\mathbf{e}$ is lost after weighting $\mathbf{r}$ and estimating the power spectrum according to QE formalism. We investigate this by comparing two quantities we call the input power spectrum and output power spectrum: $\widehat{P}_{\mathrm{in}}$ and $\widehat{P}_{\mathrm{out}}$, estimated using QE as

$$\widehat{P}_{\mathrm{in}}^{\alpha} \equiv \mathrm{M}_{\mathrm{in}}^{\alpha} \mathbf{e}^{\dagger} \mathbf{I} \mathbf{Q}^{\alpha} \mathbf{I} \mathbf{e} \tag{16}$$

and

$$\begin{aligned}
\widehat{P}_{\mathrm{out}}^{\alpha} &\equiv \widehat{\mathbf{P}}_{r}^{\alpha} \\
&= \mathrm{M}_{r}^{\alpha} \mathbf{r}^{\dagger} \mathbf{R}_{r} \mathbf{Q}^{\alpha} \mathbf{R}_{r} \mathbf{r},
\end{aligned} \tag{17}$$

where, for illustrative purposes and notational simplicity, we have written these equations with scalar normalizations M, even though for our numerical results we choose a diagonal matrix normalization using $\mathbf{M}$ as in Equation (9).

The quantity $\widehat{P}_{\mathrm{in}}$, defined by Equation (16), is a uniformly weighted estimator of the power spectrum of $\mathbf{e}$. It can be considered the power spectrum of this particular realization of the EoR; alternatively, it can be viewed as the true power spectrum of the injected signal up to cosmic variance fluctuations. The role of $\widehat{P}_{\mathrm{in}}$ in our analysis is to serve as a reference for the power spectrum that would be measured if there were no signal loss or other systematics. The input power spectrum is then to be compared to $\widehat{P}_{\mathrm{out}}$, which approximates the (lossy) power spectrum estimate that is output by our analysis pipeline prior to any signal loss adjustments.

Under this injection framework, we can begin to see explicitly why there can be large signal loss. Expanding out Equation (17), $\widehat{P}_{\mathrm{out}}$ becomes:

$$\begin{aligned}
\widehat{P}_{\mathrm{out}}^{\alpha} &= \mathrm{M}_{r}^{\alpha} (\mathbf{x}+\mathbf{e})^{\dagger} \mathbf{R}_{r} \mathbf{Q}^{\alpha} \mathbf{R}_{r} (\mathbf{x}+\mathbf{e}) \\
&= \mathrm{M}_{a}^{\alpha} \mathbf{x}^{\dagger} \mathbf{R}_{r} \mathbf{Q}^{\alpha} \mathbf{R}_{r} \mathbf{x} + \mathrm{M}_{b}^{\alpha} \mathbf{e}^{\dagger} \mathbf{R}_{r} \mathbf{Q}^{\alpha} \mathbf{R}_{r} \mathbf{e} \\
&\quad + \mathrm{M}_{c}^{\alpha} \mathbf{x}^{\dagger} \mathbf{R}_{r} \mathbf{Q}^{\alpha} \mathbf{R}_{r} \mathbf{e} + \mathrm{M}_{d}^{\alpha} \mathbf{e}^{\dagger} \mathbf{R}_{r} \mathbf{Q}^{\alpha} \mathbf{R}_{r} \mathbf{x}.
\end{aligned} \tag{18}$$

Assuming $\mathbf{R}_{r}$ is symmetric, the two cross-terms (terms with one copy of $\mathbf{e}$ and one copy of $\mathbf{x}$) can be summed together as:

$$\begin{aligned}
\widehat{P}_{\mathrm{out}}^{\alpha} &= \mathrm{M}_{a}^{\alpha} \mathbf{x}^{\dagger} \mathbf{R}_{r} \mathbf{Q}^{\alpha} \mathbf{R}_{r} \mathbf{x} + \mathrm{M}_{b}^{\alpha} \mathbf{e}^{\dagger} \mathbf{R}_{r} \mathbf{Q}^{\alpha} \mathbf{R}_{r} \mathbf{e} \\
&\quad + 2 \mathrm{M}_{c}^{\alpha} \mathbf{x}^{\dagger} \mathbf{R}_{r} \mathbf{Q}^{\alpha} \mathbf{R}_{r} \mathbf{e}.
\end{aligned} \tag{19}$$

One of the key takeaways of this section is that the A15 analysis estimated signal loss by comparing *only* the signal-only term (second term in Equation (19)) with $\widehat{P}_{\mathrm{in}}$, whereas in fact the cross-term (third term in Equation (19)) can substantially lower $\widehat{P}_{\mathrm{out}}$. In order to investigate the effect of each of these terms on signal loss, all three components are plotted in Figure 11 for two cases: empirically estimated inverse covariance weighting ($\mathbf{R}_{r} \equiv \widehat{\mathbf{C}}_{r}^{-1}$) and uniform weighting ($\mathbf{R}_{r} \equiv \mathbf{I}$). We will now go into further detail and examine the behavior of this equation in three different regimes of the injected

signal - very weak (left ends of the P_{in} axes in Figure 11), very strong (right ends), and in between (middle portions).

Small injection: In this regime, the cross-terms (red) behave as noise averaged over a finite number of samples. Output values are Gaussian distributed around zero, spanning a range of values set by the injection level. This is because $\widehat{\mathbf{R}}_{r}$ is dominated by the data $\mathbf{x}$, avoiding correlations with $\mathbf{e}$ that can lead to solely negative power (explained further below). In fact, for the uniformly weighted case, the cross-term $\mathrm{M}_{x}^{\alpha} \mathbf{x}^{\dagger} \mathbf{I} \mathbf{Q}^{\alpha} \mathbf{I} \mathbf{e}$ is well modeled as a symmetric distribution with zero mean and width $\sqrt{\widehat{\mathbf{P}}_{e}} \sqrt{\widehat{\mathbf{P}}_{x}}$. We also note that in this regime, $\widehat{\mathbf{P}}_{r}$ (black) approaches the data-only power spectrum value (gray) as expected.

Large injection: When the injected signal is much larger than the measured power spectrum, the data-only components can be neglected as they are many orders of magnitude smaller. We include a description of this regime for completeness in our discussion, but note that the upper limits that we compute are typically not determined by simulations in this regime (i.e., in using an empirical weighting scheme we've assumed the data to be dominated by foregrounds rather than the cosmological signal). However, it is useful as a check of our system in a relatively simple case. As we can see from Figure 11, the cross-terms (red) are small in comparison to the signal-only term (green). Here only does the signal-only term used in A15 dominate the total power output. We again see that, in the empirical inverse covariance weighted case, the cross-terms behave as noise (positive and negative fluctuations around zero mean). This is for the same reason as at small injections — here $\widehat{\mathbf{C}}_{r}$ is dominated by the signal $\mathbf{e}$. The cross-correlation can again be modeled as a symmetric distribution of zero mean and width $\sqrt{\widehat{\mathbf{P}}_{e}} \sqrt{\widehat{\mathbf{P}}_{x}}$.

In between: When the injected signal is of a similar amplitude to the data by itself, the situation becomes less straightforward. We see that the weighted injected power spectrum component mirrors the input power indicating little loss (i.e., the green curve follows the dotted black line), eventually departing from unity when the injected amplitude is well above the level of the data power spectrum. However, in this regime the cross-term (red) has nearly the same amplitude, but with a negative sign. As explained below, this negativity is the result of cross-correlating inverse covariance weighted terms. This negative component drives down the $\widehat{P}_{\mathrm{out}}$ estimator (black). Again, we emphasize that in A15, signal loss was computed by only looking at the second term in Equation (19) (green), which incorrectly implies no loss at the data-only power spectrum level. Ignoring the effect of the negative power from the cross-terms is the main reason for underestimating power spectrum limits in A15.

The source of the strong negative cross-term is not immediately obvious, however it is an explainable effect.
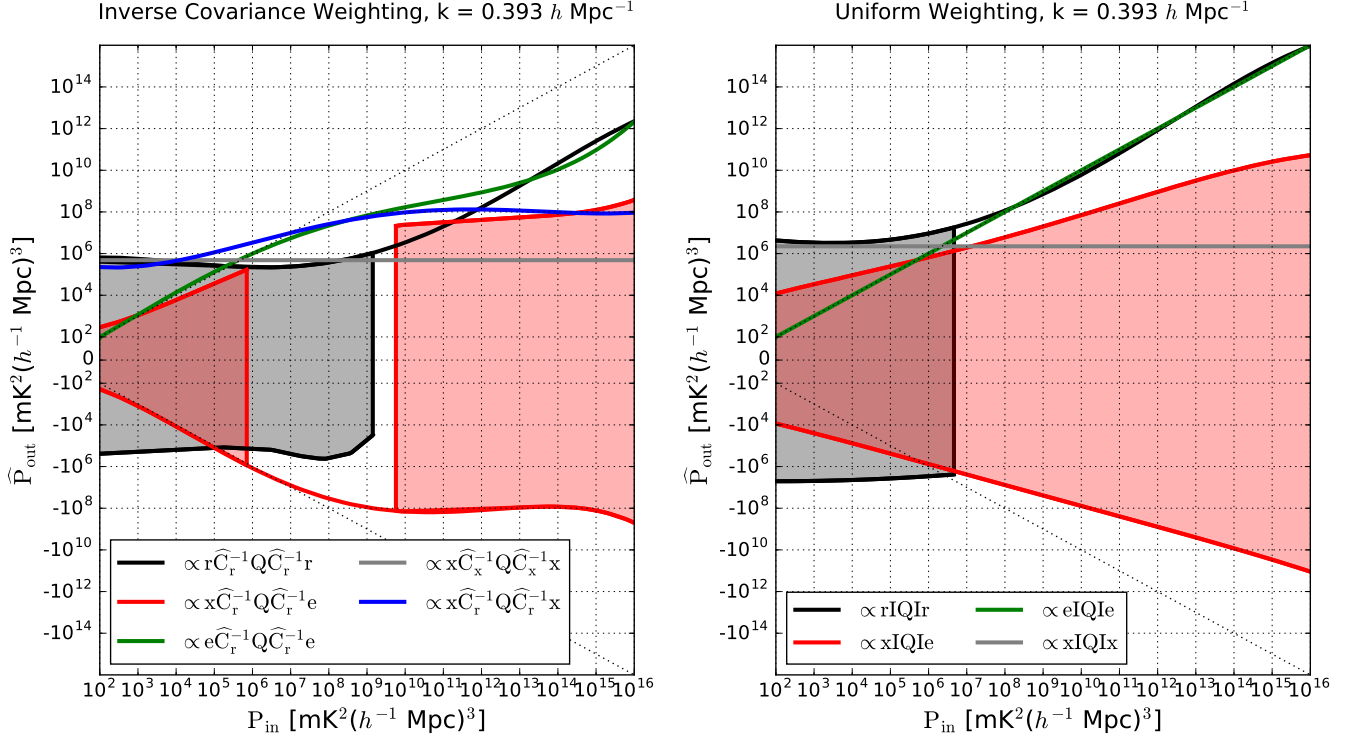
Figure 11. Illustration of the power spectrum amplitude of five different power spectrum terms, each a function of visibility data ($\mathbf{x}$), simulated injected EoR signal ($\mathbf{e}$), or both ($\mathbf{r}$). This figure shows how these quantities behave as the power level of the injected EoR signal increases (along the x-axis). The details of the simulation used to generate the figure is explained in Section 3.2; here we sample a larger P_{in} range and fit smooth polynomials to our data points to make an illustrative example. We emphasize that the output power spectrum in black ($\widehat{P}_{\mathrm{out}} = \widehat{\mathbf{P}}_r$) approximates the (lossy) power spectrum estimate that is output by our analysis pipeline prior to any signal loss adjustments. Roughly speaking, it can be compared to the input signal level (P_{in}) to estimate the amount of signal loss. Left: Empirical inverse covariance weighting is used in power spectrum estimation, as done in A15. The dotted diagonal black line indicates perfect 1:1 input-to-output mapping (no signal loss). The gray horizontal line is the power spectrum value of data alone, $\widehat{\mathbf{P}}_x$ (it does not depend on injected power). The green signal-signal component is the term used in A15 to estimate signal loss. It is significantly higher than $\widehat{\mathbf{P}}_r$ (black) when the cross-terms (red) are large and negative (black = green + red + blue). In the regime where cross-correlations between signal and data are not dominant (small and large P_{in}), the cross-terms have a noise-like term with width $\sqrt{\widehat{\mathbf{P}}_e}\sqrt{\widehat{\mathbf{P}}_x}$. However, at power levels comparable to the data (the middle region), the cross-terms can produce large, negative estimates due to couplings between $\mathbf{x}$ and $\mathbf{e}$ which affect $\widehat{\mathbf{C}}_r$. This causes the difference between the green curve (which exhibits negligible loss at the data-only power spectrum value) and the black curve (which exhibits ~ 4 orders of magnitude of loss). Right: The same power spectrum terms illustrated for the uniform weighted case.

When $\mathbf{R}_r$ is taken to be $\widehat{\mathbf{C}}_r^{-1}$, the third term of Equation (19) is a cross-correlation between $\widehat{\mathbf{C}}_r^{-1}\mathbf{x}$ and $\widehat{\mathbf{C}}_r^{-1}\mathbf{e}$. As shown in Switzer et al. (2015), this cross-correlation term is non-zero, and in fact negative in expectation. This negative cross-term power arises from a coupling between the inverse of $\widehat{\mathbf{C}}_r$ and $\mathbf{x}$. Intuitively, we can see this by expanding the empirical covariance of $\mathbf{r} = \mathbf{x} + \mathbf{e}$:

$$\widehat{\mathbf{C}}_r = \langle \mathbf{rr}^\dagger \rangle_t$$
$$= \langle \mathbf{xx}^\dagger \rangle_t + \langle \mathbf{xe}^\dagger \rangle_t + \langle \mathbf{ex}^\dagger \rangle_t + \langle \mathbf{ee}^\dagger \rangle_t, \quad (20)$$

where we can neglect the first term because $\mathbf{x}$ is small (i.e., the large negative cross-term power in the left panel of Figure 11 occurs when the injected amplitude surpasses the level of the data-only power spectrum). Without loss of generality, we will assume an eigenbasis of $\mathbf{e}$,

so that $\langle \mathbf{ee}^\dagger \rangle_t$ is diagonal. The middle two terms, however, can have power in their off-diagonal terms due to the fact that, when averaging over a finite ensemble, $\langle \mathbf{xe}^\dagger \rangle_t$ is not zero. As shown in Appendix C of Parsons et al. (2014), to leading order the inversion of a diagonal-dominant matrix like $\widehat{\mathbf{C}}_r$ (from $\langle \mathbf{ee}^\dagger \rangle_t$) with smaller off-diagonal terms results in a new diagonal-dominant matrix with negative off-diagonal terms. These off-diagonal terms depend on both $\mathbf{x}$ and $\mathbf{e}$. Then, when $\widehat{\mathbf{C}}_r^{-1}$ is multiplied into $\mathbf{x}$, the result is a vector that is similar to $\mathbf{x}$ but contains a residual correlation to $\mathbf{e}$ from the off-diagonal components of $\widehat{\mathbf{C}}_r^{-1}$. The correlation is negative because the product $\widehat{\mathbf{C}}_r^{-1}\mathbf{x}$ effectively squares the $\mathbf{x}$-dependence of the off-diagonal terms in $\widehat{\mathbf{C}}_r^{-1}$ while retaining the negative sign that arose from the inversion of a diagonal-dominant matrix.

In general: Another way to phrase the shortcoming of the empirical inverse covariance estimator (which is also discussed in Appendix B) is that it is not properly normalized. Signal loss due to couplings between the data and its weightings arise because our unnormalized quadratic estimator from Equation (8) ceases to be a quadratic quantity, and instead contains higher order powers of the data. However, the normalization matrix $\mathbf{M}$ is derived assuming that the unnormalized estimator is quadratic in the data. The power spectrum estimate will therefore be incorrectly normalized, which manifests as signal loss. We leave a full analytic solution for $\mathbf{M}$ for future work, since our simulations already capture the full phenomenology of signal loss and have the added benefit of being more easily generalizable in the face of non-Gaussian systematics.

3.2. *Signal Loss in Practice*

We now shift our attention towards computing upper limits on the EoR signal for the fringe-rate filtered PAPER-64 dataset in a way that accounts for signal loss. While our methodology outlined below is independent of weighting scheme, here we demonstrate the computation using empirically estimated inverse covariance weighting ($\mathbf{R} \equiv \widehat{\mathbf{C}}^{-1}$), the weighting scheme used in A15 which leads to substantial loss.

One issue to address is how one incorporates the randomness of $\widehat{P}_{\rm out}$ into our signal loss corrections. A different realization of the mock EoR signal is injected with each bootstrap run, causing the output to vary in three ways — there is noise variation from the bootstraps, there is cosmic variation from generating multiple realizations of the mock EoR signal, and there is a variation caused by whether the injected signal looks more or less "like" the data (i.e., how much coupling there is, which affects how much loss results).

For each injection level, the true $P_{\rm in}$ is simply the average of our bootstrapped estimates $\widehat{P}_{\rm in}$, since $\widehat{P}_{{\rm in},\alpha}$ is by construction an unbiased estimator. Phrased in the context of Bayes' rule, we wish to find the posterior probability distribution $p(P_{\rm in}|\widehat{P}_{\rm out})$, which is the probability of $P_{\rm in}$ given the uncorrected/measured power spectrum estimate $\widehat{P}_{\rm out}$. Bayes' rule relates the posterior, which we don't know, to the likelihood, which we can forward model. In other words,

$$p(P_{\rm in}|\widehat{P}_{\rm out}) \propto \mathcal{L}(\widehat{P}_{\rm out}|P_{\rm in})\,p(P_{\rm in}), \qquad (21)$$

where $\mathcal{L}$ is the likelihood function defined as the distribution of data plus signal injection ($\widehat{P}_{\rm out}$) given the injection $P_{\rm in}$. We construct this distribution by fixing $P_{\rm in}$ and simulating our analysis pipeline for many realizations of the injected EoR signal consistent with this power spectrum. The resulting distribution is normalized such that the sum over $\widehat{P}_{\rm out}$ is unity, and the whole process is then repeated for a different value of $P_{\rm in}$.

The implementation details of the injection process require some more detailed explanation. In our code,

we add a new realization of EoR to each independent bootstrap of data (see Section 4.1 for a description of PAPER's bootstrapping routine) with the goal of simultaneously capturing cosmic variance, noise variance, and signal loss. To limit computing time we perform 20 realizations of each $P_{\rm in}$ level. We also run 50 total EoR injection levels, yielding $P_{\rm in}$ values that range from $\sim 10^5 \, {\rm mK}^2 \ (h^{-1} \ {\rm Mpc})^3$ to $\sim 10^{11} \, {\rm mK}^2 \ (h^{-1} \ {\rm Mpc})^3$, resulting in a total of 1000 data points on our $P_{\rm in}$ vs. $\widehat{P}_{\rm out}$ grid.

Going forward, we treat every k-value separately in order to determine an upper limit on the EoR signal per k. We bin our simulation outputs along the $P_{\rm in}$ axis (one bin per injection level) and, since they are well-approximated by a Gaussian distribution in our numerical results, we smooth the distribution of $\widehat{P}_{\rm out}$ values by fitting Gaussians for each bin based on its mean and variance (and normalize them). Stitching all of them together results in a 2-dimensional transfer function — the likelihood function in Bayes' rule, namely $\mathcal{L}(\widehat{P}_{\rm out}|P_{\rm in})$. We then have a choice for our prior, $p(P_{\rm in})$, and we choose to invoke a Jeffreys prior (Jaynes 1968) because it is a true uninformative prior. For a derivation and more details about the Jeffreys prior used in our analysis, see Appendix C.

Finally, our transfer functions are shown in Figure 12 for both the weighted (left) and unweighted (right) cases. Our bootstrapped power spectrum outputs are shown as black points and the colored heat-map overlaid on top is the likelihood function modified by our prior. Although we only show figures for one k-value, we note that the shape of the transfer curve is similar for all k's. We then invoke Bayes' interpretation and reinterpret it as the posterior $p(P_{\rm in}|\widehat{P}_{\rm out})$ where we recall that $\widehat{P}_{\rm out}$ represents a (lossy) power spectrum. To do this we make a horizontal cut across at the data value $\widehat{\mathbf{P}}_x$ (setting $\widehat{P}_{\rm out} = \widehat{\mathbf{P}}_x$), shown by the gray solid line, to yield a posterior distribution for the signal. We normalize this final distribution and compute the 95% confidence interval (an upper limit on EoR).

By-eye inspection of the transfer function in Figure 12 gives a sense of what the signal loss result should be. The power spectrum value of our data, $\widehat{\mathbf{P}}_x$ is marked by the solid gray horizontal lines. From the left plot (empirically estimated inverse covariance weighting), one can eyeball that a data value of $10^5 \, {\rm mK}^2 \ (h^{-1} \ {\rm Mpc})^3$, for example, would map approximately to an upper limit of $\sim 10^9 \, {\rm mK}^2 \ (h^{-1} \ {\rm Mpc})^3$, implying a signal loss factor of $\sim 10^4$.

The loss-corrected power spectrum limit for empirically estimated inverse covariance weighted PAPER-64 data is shown in Figure 13 (solid red), which we can compare to the original lossy result (dashed red). Post-signal loss estimation, the power spectrum limits are higher than both the theoretical noise level (green) and uniform-weighted power spectrum (which is shown three ways: black and gray points are positive and negative power spectrum values, respectively, with 2σ error bars
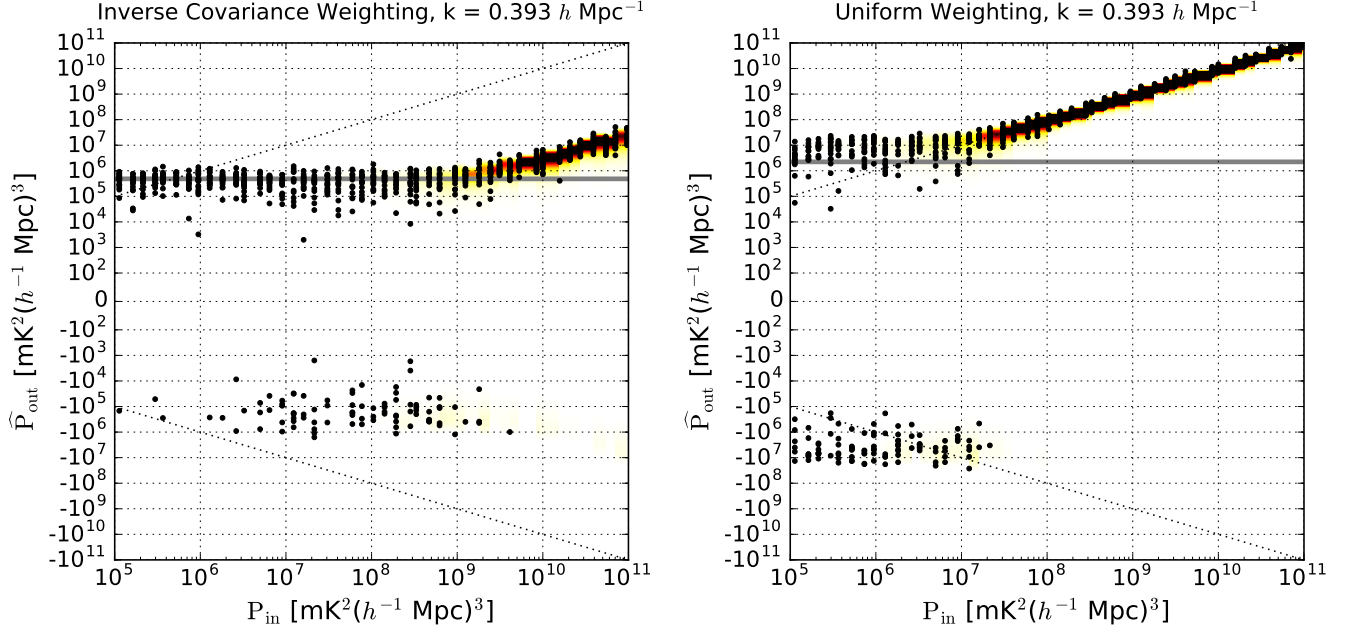
Figure 12. Signal loss transfer functions showing the relationship of P_{in} and $\widehat{P}_{out}$, as defined by Equations (16) and (17). Power spectra values (black points) are generated for 20 realizations of **e** per signal injection level. Since our $\widehat{P}_{out}$ values are well-approximated by a Gaussian distribution, we fit Gaussians to each injection level based on the mean and variance of the simulation outputs. This entire likelihood function is then multiplied by a Jeffreys prior for $p(P_{in})$, with the final result shown as the colored heat-maps on top of the points. Two cases are displayed: empirically estimated inverse covariance weighted PAPER-64 data (left) and uniform-weighted data (right). The dotted black diagonal lines mark a perfect unity mapping, and the solid gray horizontal line denotes the power spectrum value of the data $\widehat{\mathbf{P}}_x$, from which a posterior distribution for the signal is extracted. From these plots, it is clear that the weighted case results in ~ 4 orders of magnitude of signal loss at the data-only power spectrum value, whereas the uniform-weighted case does not exhibit loss. The general shape of these transfer functions are also shown by the black curves in Figure 11 for comparison.

from bootstrapping, the solid blue is the upper limit on the EoR signal using the full signal injection framework, and the shaded gray is the power spectrum values with thermal noise errors). We elaborate on this point in the next section, as well as investigate alternate weighting schemes to inverse covariance weighting, with the goal of finding one that balances the aggressiveness of down-weighting contaminants and minimizing the loss of the EoR signal.

3.3. *Minimizing Signal Loss*

With a signal loss formalism established, we now have the capability of experimenting with different weighting options for **R**. Our goal here is to choose a weighting method that successfully down-weights foregrounds and systematics in our data without generating large amounts of signal loss as we have seen with the inverse covariance estimator. We have found that the balance between the two is a delicate one and requires a careful understanding and altering of empirical covariances.

We saw in Section 2.4 how limiting the number of down-weighted eigenmodes (i.e., flattening out part of the eigenspectrum and effectively decoupling the lowest-valued eigenmodes, which are typically EoR-dominated, from the data) can help minimize signal loss. We experiment with this idea on PAPER-64 data, dial-

ing the number of modes that are down-weighted from zero (which is equivalent to identity-weighting, or the uniform-weighted case) to 21 (which is the full inverse covariance estimator). The power spectrum results for one k-value, both before and after signal loss estimation, are shown in the top panel in Figure 14. We see that the amount of signal loss increases as weighting becomes more aggressive (dashed red). In other words, more EoR-dominated fluctuations are being overfit and subtracted as more modes are down-weighted. We also find that the power spectrum upper limit, post-signal loss estimation, increases with the number of down-weighted modes (solid red). The more modes we use in down-weighting, the stronger the coupling between the weighting and the data, and the greater the error we have in estimating the power spectrum. Switzer et al. (2013) took a similar approach in determining the optimal number of modes to down-weight in GBT data, finding similar trends and noting that removing too few modes is limited by residual foregrounds and removing too many modes is limited by large error bars and signal loss.

Optimistically, we expect there to be a "sweet spot" as we dial our regularization knob; a level of regularization where weighting is beneficial compared to uniform weighting (blue). In other words, we would like a weight-
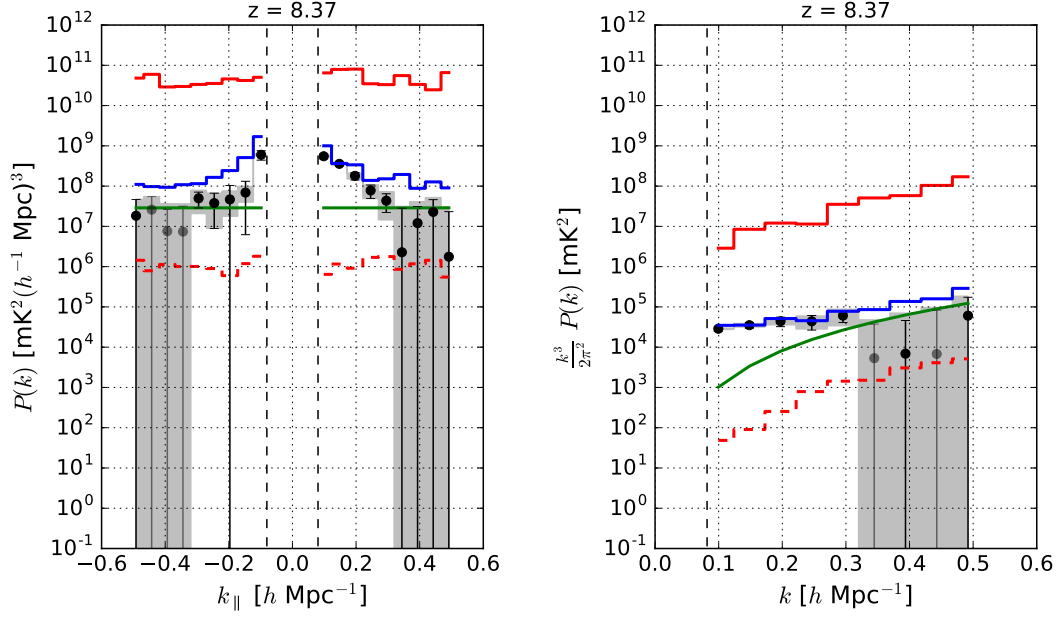
Figure 13. A power spectrum of a subset of PAPER-64 data illustrating the use of empirical inverse covariance weighting. The solid red curve is the 2σ upper limit on the EoR signal estimated from our signal injection framework using empirical inverse covariance weighting. Shown for comparison is the lossy limit prior to signal loss estimation (dashed red). The theoretical 2σ thermal noise level prediction based on observational parameters is in green, whose calculation is detailed in Section 4.2. Additionally, the power spectrum result for the uniform weighted case is shown in three different ways: power spectrum values (black and gray points as positive and negative values, respectively, with 2σ error bars from bootstrapping), the 2σ upper limit on the EoR signal using our full signal injection framework (solid blue), and the measured power spectrum values with 2σ thermal noise errors (gray shaded regions). The vertical dashed black lines signify the horizon limit for this analysis using 30 m baselines. In this example, we see that the lossy power spectrum limit is ~ 4 orders of magnitude too low when using empirical inverse covariance weighting.

ing scheme that down-weights eigenmodes that predominantly describe foreground modes, but not EoR modes. We see in Figure 14 that this occurs roughly when only the $\sim$ 2-3 highest-valued eigenmodes are down-weighted and the rest are given equal weights (though for the case shown, weighting does not actually outperform uniform weighting). For a similar discussion on projecting out modes (zeroing out eigenmodes, rather than just ignoring their relative weightings as we do in this study), see Switzer et al. (2013).

We also saw in Section 2.4 how adding the identity matrix to the empirical covariance can minimize signal loss. We experiment with this idea as well, shown in the bottom panel of Figure 14. The dashed red and solid red lines represent power spectrum limits pre and post-signal loss estimation, respectively, as a function of the strength of $\mathbf{I}$ that is added to $\widehat{\mathbf{C}}$, quantified as a percentage of $\mathrm{Tr}(\widehat{\mathbf{C}})\mathbf{I}$ added to $\widehat{\mathbf{C}}$. We parameterize this "regularization strength" parameter as γ, namely $\widehat{\mathbf{C}} \equiv \widehat{\mathbf{C}} + \gamma \mathrm{Tr}(\widehat{\mathbf{C}})\mathbf{I}$. From this plot we see that only a small percentage of $\mathrm{Tr}(\widehat{\mathbf{C}})$ is needed to significantly reduce loss. We expect that as the strength of $\mathbf{I}$ is increased (going to the left), both the red curves will approach the uniform-weighted case. We also notice that the post-signal loss limit hovers around the uniform-weighted limit for a large range of regularization strengths and while an overall trend from high-to-

low signal loss is seen as the strength increases, there does not appear to be a clear "minimum" that produces the least loss.

In addition to our thermal noise prediction (green) and uniform-weighted power spectrum limit (blue), one additional horizontal line is shown in Figure 14 in both panels and represents a third regularization technique. This line (black) denotes the power spectrum value, post-signal loss estimation, for inverse variance weighting (multiplying an identity matrix element-wise to $\widehat{\mathbf{C}}$). This result is single-valued and not a function of the horizontal axis. We see that all three regularization schemes shown (solid red top panel, solid red bottom panel, black) perform similarly at their best (i.e., when $\sim$ 2-3 eigenmodes are down-weighted in the case of the top panel's solid red curve). However, for the remainder of this paper, we choose to use the weighting option of $\widehat{\mathbf{C}} + 0.09 \mathrm{Tr}(\widehat{\mathbf{C}})\mathbf{I}$, or $\gamma = 0.09$, which we will denote as $\widehat{\mathbf{C}}_{\mathrm{eff}}$. We choose this weighting scheme merely as a simple example of regularizing PAPER-64 covariances, noting that the power spectrum upper limit remains roughly constant for a broad range of values of γ.

It is important to note that our signal injection methodology for assessing loss makes the assumption that we know the true signal's strength and structure. Realistically, these details about the EoR signal are un-
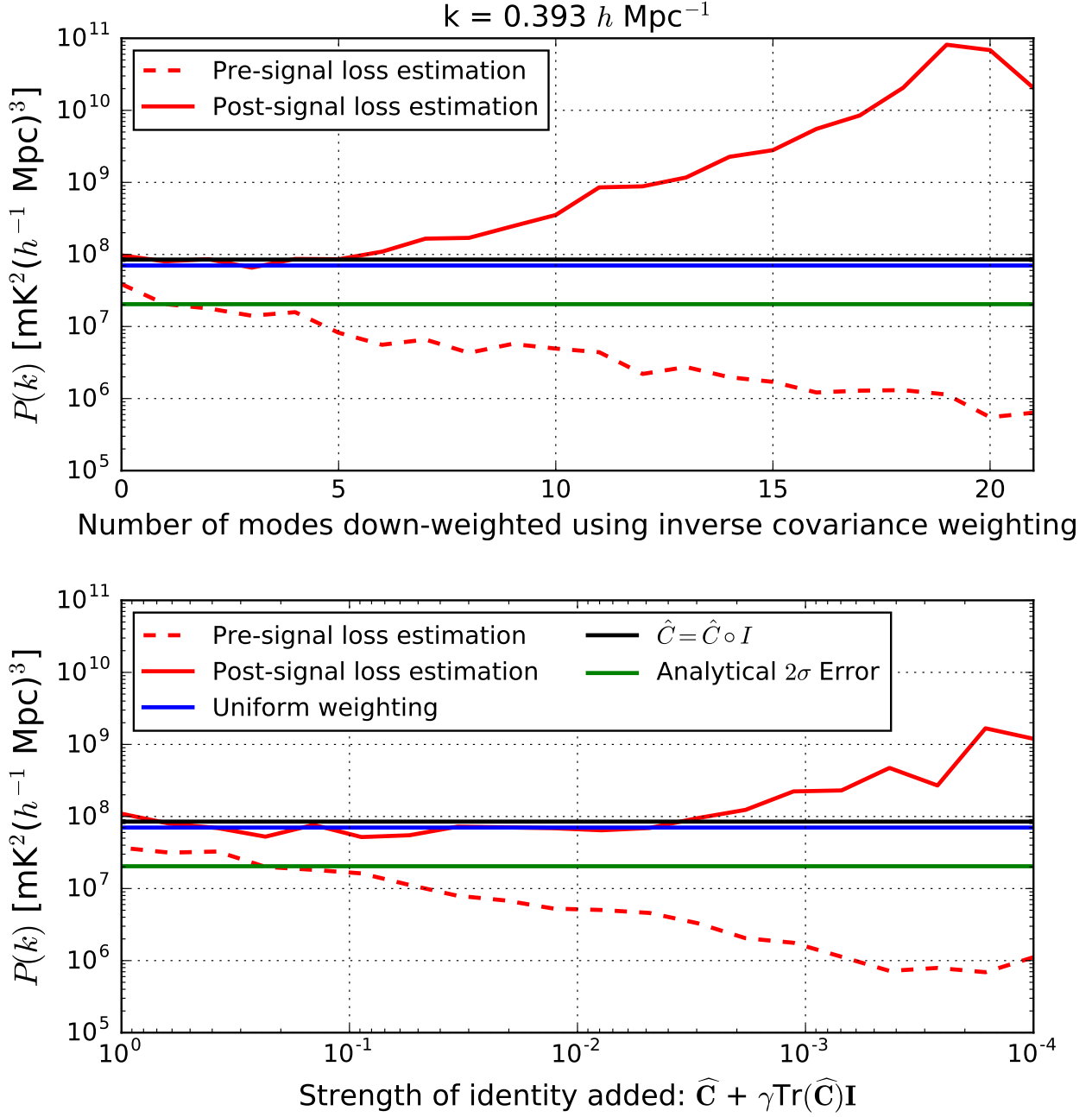
Figure 14. Power spectra 2σ upper limits for $k = 0.393\,h\ \mathrm{Mpc}^{-1}$ for fringe-rate filtered PAPER-64 data. Top: Values are shown before (dashed red) and after (solid red) signal loss estimation via our signal injection framework as a function of number of eigenmodes of $\widehat{\mathbf{C}}$ that are down-weighted. This regularization knob is tuned from 0 modes on the left (i.e., unweighted) to 21 modes on the right (i.e., the full inverse covariance estimator). ~ 4 orders of magnitude of signal loss results when using empirically estimated inverse covariance weighting. Bottom: Power spectrum upper limits before (dashed red) and after (solid red) signal loss estimation as a function of identity added to the empirical covariance. This regularization knob is tuned from $\gamma = 10^{-4}$ on the right (i.e., very little regularization) to $\gamma = 1$ on the left (see main text for the definition of γ). Also plotted in both panels for comparison are 2σ power spectrum upper limits for the uniform-weighted case (blue) and inverse variance weighted case (black); both are after signal loss estimation. Finally, a theoretical prediction for noise (2σ error) is plotted as green. In the PAPER-64 analysis in this paper, we choose to use a regularization scheme of $\widehat{\mathbf{C}}_{\mathrm{eff}} \equiv 0.09\,\mathrm{Tr}(\widehat{\mathbf{C}})\mathbf{I} + \widehat{\mathbf{C}}$ ($\gamma = 0.09$) as a simple example of regularization that minimizes loss, and note that the power spectrum limits using this type of regularization are roughly constant across a large range of values of γ.

known, and therefore our signal loss framework is subject to bias. Therefore, while this paper employs this methodology as an example of one way of estimating loss, Kolopanis et al. (*in prep.*) use uniform weightings in order to produce more trustworthy, straightforward power spectrum limits that do not suffer from loss.

The power spectrum result for our subset of PAPER-64 data (using only one baseline separation type, 10 baselines, and $\widehat{\mathbf{C}}_{\mathrm{eff}}$) using the analysis presented in this paper is shown in Figure 15. Again, the solid red curve represents our upper limit on the EoR signal using the full signal injection framework. The uniform weighted case is shown as the black and gray points, which correspond to positive and negative power spectrum values respectively (with 2σ errors bars from bootstrapping). It is also shown as an upper limit using the signal injection framework (solid blue), which is interestingly larger than the errors computed from bootstrapping, likely because the full injection framework takes into account additional sample variance whereas the bootstrapped errors do not. Finally, the gray shaded regions combine the measured uniform weighted power spectrum values with thermal noise errors. We show this power spectrum result as one example of how a simple regularization of an empirical covariance matrix can minimize signal loss, though we also note that this weighting does not produce more stringent limits than the uniform weighted case, thus further motivating uniform-weighting for Kolopanis et al. (*in prep.*).

In this section we have shown three simple ways of regularizing $\widehat{\mathbf{C}}$ to minimize signal loss using PAPER-64 data. There are many other weighting schemes that we leave for consideration in future work. For example, one could estimate $\widehat{\mathbf{C}}$ using information from different subsets of baselines. For redundant arrays this might mean calculating $\widehat{\mathbf{C}}$ from a different but similar baseline type, such as the $\sim 30\,\mathrm{m}$ diagonal PAPER baselines (instead of the horizontal E/W ones). Alternatively, covariances could be estimated from all baselines except the two being cross-multiplied when forming a power spectrum estimate. This method was used in Parsons et al. (2014) (a similar method was also used in Dillon et al. (2015)) in order to avoid suppressing the 21 cm signal, and it is worth noting that the PAPER-32 results are likely less impacted from the issue of signal loss underestimation because of this very reason (however, they are affected by the error estimation issues described in Section 4.2, so we also regard those results as suspect and superseded by those of Kolopanis et al. (*in prep.*)).

Another possible way to regularize $\widehat{\mathbf{C}}$ is to use information from different ranges of LST. For example, one could calculate $\widehat{\mathbf{C}}$ with data from LSTs where foregrounds are stronger (earlier or later LSTs than the "foreground-quiet" range typically used in forming power spectra) — doing so may yield a better description of the foregrounds that we desire to down-weight, especially if residual foreground chromaticity is instrumental in origin and stable in time. Fundamentally, each of these examples are similar in that they rely on a computation of $\widehat{\mathbf{C}}$ from data that is similar but not exactly the same as the data that is being down-weighted. Ideally this would be effective in down-weighting shared contaminants yet avoid signal loss from overfitting EoR modes in the power spectrum dataset itself.

In Section 3, we have detailed several aspects of signal loss in PAPER-64: how the loss arises, how it can be estimated from an injection framework, and ways it can be minimized. We again emphasize that these lessons learned about signal loss are largely responsible for shaping our revised analysis of PAPER data. In the remainder of this paper, we will transition to other aspects of our analysis that have been revised since A15.

4. ADDITIONAL PAPER-64 REVISIONS

Underestimated signal loss is the main reason for the revision of the power spectrum limits from A15. It is interesting to note that — had all the other aspects of the original analysis been correct — the underestimated limits may have been more easily caught. Unfortunately, two related power spectrum components, namely the error bars on the power spectrum data points and the theoretical noise prediction, were also calculated incorrectly.

In this section, we summarize multiple inconsistencies and errors that have been found since the previous analysis in terms of error estimation. We first describe updated methods regarding bootstrapping, which determines the error bars on our limits. We then highlight an updated calculation for the theoretical noise sensitivity of PAPER-64 and illustrate how our revised calculation has been verified through simulations.

4.1. *Bootstrapping*

Broadly speaking, we desire robust methods for determining accurate confidence intervals for our measurements. For PAPER's analysis, we choose a data-driven method of error estimation, computing error bars that have been derived from the inherent variance of our measurements. A common technique used to do this is bootstrapping, which we first define below and then discuss its application to PAPER.

Bootstrapping uses sampling with replacement to estimate a posterior distribution. For example, bootstrap measurements (of power spectra, for example) can be made from different random samples of data. Each of these bootstraps is a different realization drawn from some underlying distribution, and realizations are correlated with each other to a degree set by the fraction of sampled points that are held in common between them. Through the process of re-sampling and averaging along different axes of a dataset, such as along baselines or times, we can estimate error bars for our results which represent the underlying distribution of values that are allowed by our measurements (Efron & Tibshirani 1994; Andrae 2010).
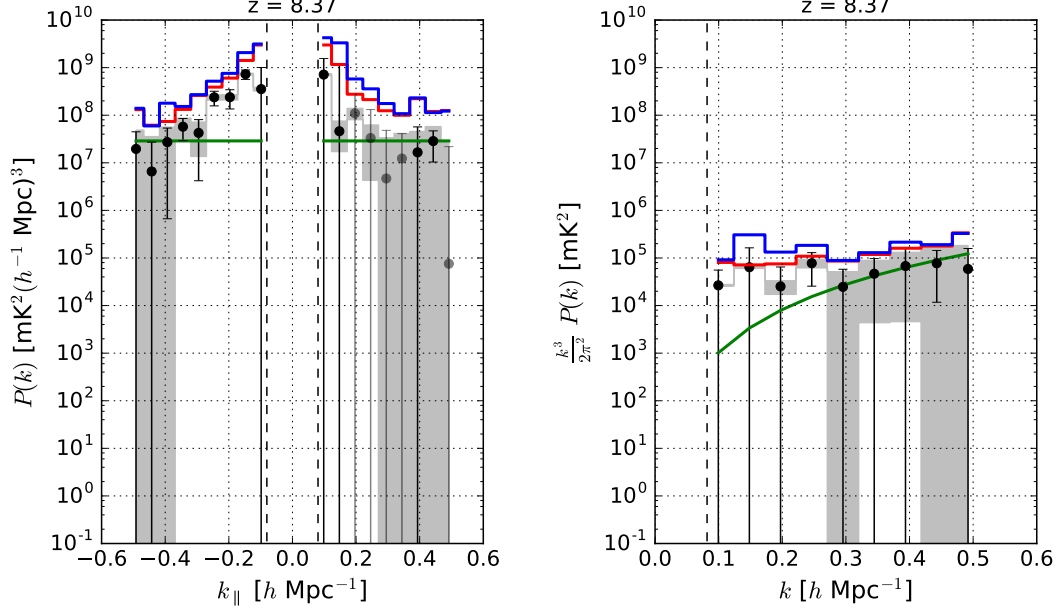
Figure 15. A power spectrum of a subset of PAPER-64 data illustrating the use of $\widehat{\mathbf{C}}_{\mathrm{eff}}$ to minimize signal loss. The solid red curve is the 2σ upper limit on the EoR signal estimated from our signal injection framework. The theoretical 2σ thermal noise level prediction based on observational parameters is in green. Additionally, the power spectrum result for the uniform weighted case is shown in three different ways: power spectrum values (black and gray points as positive and negative values, respectively, with 2σ error bars from bootstrapping), the 2σ upper limit on the EoR signal using our full signal injection framework (solid blue), and the measured power spectrum values with 2σ thermal noise errors (gray shaded regions). The vertical dashed black lines signify the horizon limit for this analysis using 30 m baselines. This power spectrum result does not use the full dataset's sensitivity as in A15 and Kolopanis et al. (*in prep.*), though we include all analysis changes which have mostly stemmed from revisions regarding signal loss, bootstrapping, and the theoretical error computation. We see that the regularization scheme used here produces limits similar to the unweighted limits.

One major caveat of bootstrapping arises when working with correlated data. If, for example, a dataset has many repeated values inside it, this would be reflected in each bootstrap. The same value would be present multiple times within a bootstrap and also be present between bootstraps, purely because it has a more likely chance of being drawn if there are repeats of itself. Therefore, bootstrapping correlated data results in a smaller variation between bootstraps, and hence, underestimates errors.

This is the precisely how errors were underestimated in PAPER-64. Because of fringe-rate filtering, which averages data in time to increase sensitivity, PAPER-64 data is correlated along the time axis. Hence, there are fewer independent samples after filtering, thus decreasing the variance of the bootstraps.

More specifically, the PAPER-64 pipeline outputs 20 bootstraps (over baselines), each a 2-dimensional power spectrum that is a function of k and time. In A15, a second round of bootstrapping occurred over the time axis, and a total of 400 bootstraps were created in this step, each comprised of randomly selected values sampled with replacement (i.e., each of these bootstraps contained the same number of values as the number of time integrations, which, at ~ 700, greatly exceeds the approximate number of independent samples after fringe-rate filtering). Means were then taken of the val-

ues in each bootstrap. Finally, power spectrum limits were computed by taking the mean and standard deviation over all the bootstraps. We emphasize again that in this previous analysis, the number of elements sampled per bootstrap greatly exceeded the number of independent LST samples, underestimating errors. A random draw of 700 measurements from this dataset has many repeated values, and the variance between hundreds of these random samples is smaller than the true underlying variance of the data.

Given our new understanding of the sensitivity of bootstraps to the number of elements sampled, we have removed the second bootstrapping step along time entirely and now simply bootstrap over the baseline axis. Power spectrum 2σ errors (computed from bootstrap variances) with this bootstrapping change for a fringe-rate filtered noise simulation are shown in Figure 16. The estimates are uniformly weighted in order to disentangle the effects of bootstrapping from signal loss. As shown in the figure, when more elements are drawn for each bootstrap than the number of independent samples (by over-sampling elements along the time axis), repeated values begin to crop up and the apparent variation between bootstraps drops, resulting in limits (gray) below the predicted noise level (green). Using the revised bootstrapping method, where bootstrapping only occurs over the baseline axis, the limits (black) are

shown to agree with the analytic prediction for noise. While Figure 16 implies that errors are underestimated by a factor of ~ 5 in mK^2 for the noise simulation (whose creation details are outlined in the next section), in practice this factor is lower for the case of real data (a factor of ~ 3 in mK^2 instead), possibly due to the data being less correlated in time than the fringe-rate filtered noise in the simulation.

In addition to learning how sample independence affects bootstrapped errors, we have made three additional changes to our bootstrapping procedure since A15, summarized here:

- A second change to our bootstrapping procedure is that we now bootstrap over baseline cross-products, instead of the baselines themselves. In the previous analysis, baselines were bootstrapped prior to forming cross power spectra, and using this particular ordering of operations (bootstrapping, then cross-multiplication) yields variances that have been found to disagree with predicted errors from bootstrapping using simulations. On the contrary, bootstrapping over cross power spectra ensures that we are estimating the variance of our quantity of interest (i.e., the power spectrum). This change, while fundamental in retaining the integrity of the bootstrapping method in general, alters the resulting power spectrum errors by factors of < 2 in practice.

- In A15, individual baselines were divided into five independent groups, where no baselines were repeated in each group. Then, baselines within each group were averaged together, and the groups were cross-multiplied to form power spectra. This grouping method was used to reduce computational time, however upon closer examination it has been found that the initial grouping introduces an element of randomness into the final measurements — more specifically, the power spectrum value fluctuates depending on how baselines are assigned into their initial groups. Our new approach removes this element of randomness at the cost of computational expense, as we now perform all baseline cross-products.

- Finally, the last change from the A15 method is that our power spectrum points (previously computed as the mean of all bootstraps), are now computed as the power spectrum estimate resulting from not bootstrapping at all. More specifically, we compute one estimate without sampling, and this estimate is propagated through our signal loss computation (this estimate is $\widehat{\mathbf{P}}_x$). The difference between taking the mean of the bootstrapped values and using the estimate from the no-bootstrapping case is small, but doing the latter ensures that we are forming results that reflect the estimate preferred by all our data.
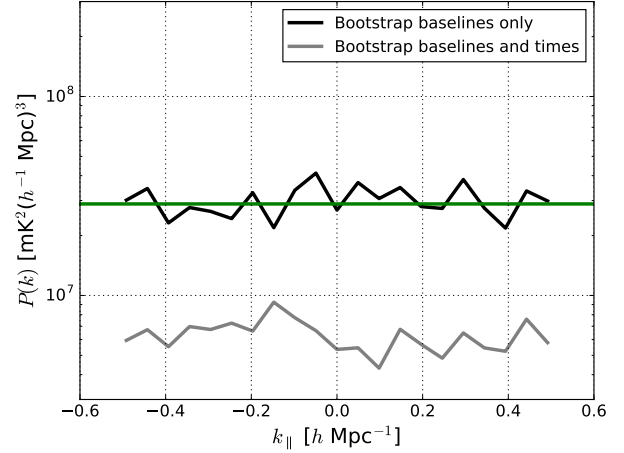
Figure 16. 2σ power spectrum errors (from bootstrap variances) for a noise simulation (computed via Equation (27)) using PAPER-64 observing parameters) using two different bootstrapping methods. The noise is fringe-rate filtered and a weighting matrix of $\mathbf{I}$ (uniform-weighted) is used in order to disentangle the effects of bootstrapping from signal loss. The bootstrapping method used in A15 is shown in gray, where bootstrapping occurs along both the baseline and time axes. This underestimates errors by sampling more values than independent ones in the dataset (fringe-rate filtering reduces the number of independent samples along time). We use the method illustrated by the black curve in our updated analysis, where bootstrapping only occurs along the baseline axis. We find that these revised limits agree with the 2σ analytic prediction for noise (green).

In summary, we have learned several lessons regarding bootstrapping and have revised our analysis procedure in order to determine error bars that correctly reflect the variance in our power spectrum estimates. Bootstrapping can be an effective and straightforward way to estimate errors of a dataset, however, bootstrapping as a means of estimating power spectrum errors from real fringe-rate filtered data requires knowledge of the number of independent samples, which is not always a trivial task. We have thus avoided this issue by removing one of our bootstrap axes, as well as updated several other details of our procedure to ensure accurate re-sampling and error estimation.

4.2. Theoretical Error Estimation

One useful way of cross-checking measured power spectrum values and errors is to compute a theoretical estimation of thermal noise based on observational parameters. Although a theoretical model often differs from true errors, it is helpful to understand the ideal case and the factors that affect its sensitivity. Upon re-analysis of PAPER-64, we have discovered that this estimate was also underestimated in previous analyses.

To compute our theoretical noise estimate, we use an analytic sensitivity calculation. Through detailed studies using several independently generated noise simulations, what we found was that our simulations all agreed but were discrepant with the previous calculations. The analytic calculation is only an approxima-

tion and attempts to combine a large number of pieces of information in an approximate way; however, when re-considering some of the approximations, the differences were large enough (factors of 10 in some cases) to warrant a careful investigation. What follows here is an accounting of the differences which have been discovered. We note that our theoretical error estimate, which is plotted as the solid green curve in many of the previous power spectrum plots, is computed with these changes accounted for.

The noise prediction $n(k)$ (Parsons et al. 2012a; Pober et al. 2013) for a power spectral analysis of interferometric 21 cm data, in temperature-units, is:

$$N(k) = \frac{X^2 Y \Omega_{\mathrm{eff}} T_{\mathrm{sys}}^2}{\sqrt{2 N_{\mathrm{lst}} N_{\mathrm{seps}}} \, t_{\mathrm{int}} N_{\mathrm{days}} N_{\mathrm{bls}} N_{\mathrm{pols}}}. \tag{22}$$

We will now explain each factor in Equation (22) and highlight key differences from the numbers used in A15.

- $X^2 Y$: Conversion factors from observing coordinates (angles on the sky and frequency) to cosmological coordinates (co-moving distances). For $z = 8.4$, $X^2 Y = 5 \times 10^{11} \, h^{-3} \, \mathrm{Mpc}^3 \, \mathrm{str}^{-1} \, \mathrm{GHz}^{-1}$.

- Ω_{eff}: The effective primary beam area in steradians (Parsons et al. 2010; Pober et al. 2012). The effective beam area changes with the application of a fringe-rate filter, since different parts of the beam are up-weighted and down-weighted. Using numbers from Table 1 in Parsons et al. (2016), $\Omega_{\mathrm{eff}} = 0.74^2/0.24$ for an optimal fringe-rate filter and the PAPER primary beam.

- T_{sys}: The system temperature is set by:

$$T_{\mathrm{sys}} = 180 \left(\frac{\nu}{0.18} \right)^{-2.55} + T_{\mathrm{rcvr}}, \tag{23}$$

where ν are frequencies in GHz (Thompson et al. 2001). We use a receiver temperature of 144 K, yielding $T_{\mathrm{sys}} = 431$ K at 150 MHz. This is lower than the T_{sys} of 500 K used in A15 because of several small miscalculation errors that were identified[3].

- $\sqrt{2}$: This factor in the denominator of the sensitivity equation comes from taking the real part of the power spectrum estimates after cross-multiplying two independent visibility measurements. In A15, a factor of 2 was mistakenly used.

- N_{lst}: The number of independent LST bins that go into a power spectrum estimation. The sensitivity scales as the square root because we integrate incoherently over time. For PAPER-64, $N_{\mathrm{lst}} = 8$.

---

[3] For example, there was a missing a square root in going from a variance to a standard deviation.

- N_{seps}: The number of baseline separation types (where baselines of a unique separation type have the same orientation and length) averaged incoherently in a final power spectrum estimate. For the analysis in this paper, we only use one type of baseline (PAPER's 30 m East/West baselines). However, both the updated limits in Kolopanis et al. (*in prep.*) and the sensitivity prediction in Figure 17 use three separation types ($N_{\mathrm{seps}} = 3$) to match A15.

- t_{int}: Length of an independent integration of the data. It is crucial to adapt this number if filtering is applied along the time axis (i.e., a fringe-rate filter). We compute the effective integration time of our fringe-rate filtered data by scaling the original integration time t_i using the following:

$$t_{\mathrm{int}} = t_i \frac{\int 1 \, df}{\int w^2(f) \, df}, \tag{24}$$

where $t_i = 43$ seconds, t_{int} is the fringe-rate filtered integration time, w is the fringe-rate profile, and the integral is taken over all fringe-rates. For PAPER-64, this number is $t_{\mathrm{int}} = 3857$ s.

- N_{days}: The total number of days of data analyzed. In A15, this number was set to 135. However, because we divide our data in half (to form "even" and "odd" datasets, or $N_{\mathrm{datasets}} = 2$), this number should reflect the number of days in each individual dataset instead of the total. Additionally, this number should be adjusted to reflect the actual number of cross-multiplications that occur between datasets ("even" with "odd" and "odd" with "even", but not "odd" with "odd" or "even" with "even" in order to avoid noise biases). Finally, because our LST coverage is not 100% complete (it doesn't overlap for every single day), we incorporate a root-mean-square statistic in computing a realistic value of N_{days}. Our expression therefore becomes:

$$N_{\mathrm{days}} = \sqrt{\langle N_i^2 \rangle} \sqrt{(N_{\mathrm{datasets}}^2 - N_{\mathrm{datasets}})} \tag{25}$$

where i indexes LST and frequency channel over all datasets (Jacobs et al. 2015). For PAPER-64, our revised estimate of N_{days} is ~ 47 days.

- N_{bls}: The number of baselines contributing to the sensitivity of a power spectrum estimate. In A15, this number was the total number of 30 m East/West baselines used in the analysis. However, using the total number of baselines ($N_{\mathrm{bls\_total}} = 51$) neglects the fact that the A15 analysis averages baselines into groups for computational speed-up when cross-multiplying data. Our revised estimate for the parameter is:

$$N_{\mathrm{bls}} = \frac{N_{\mathrm{bls\_total}}}{N_{\mathrm{gps}}} \sqrt{\frac{N_{\mathrm{gps}}^2 - N_{\mathrm{gps}}}{2}}, \tag{26}$$
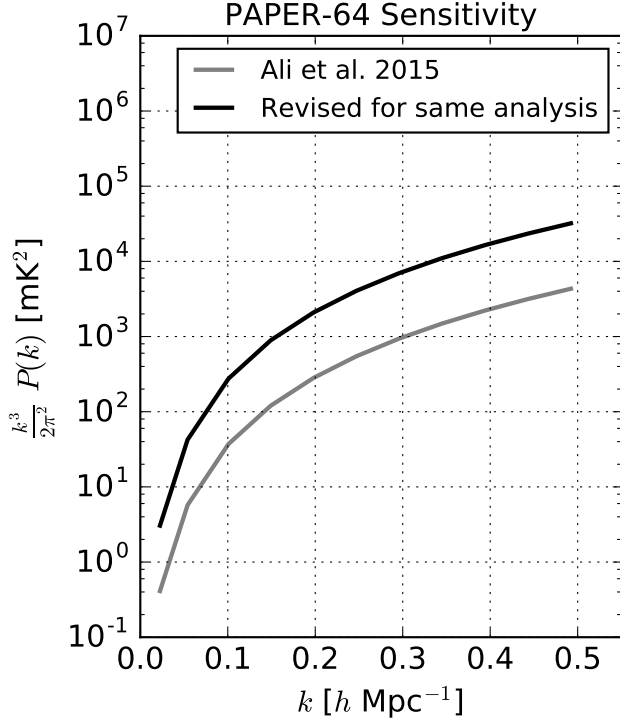
Figure 17. An updated prediction for the thermal noise level of PAPER-64 data (black) is shown in comparison to previously published sensitivity limits (gray), both computed for the parameters and methods used in A15. Major factors that contribute to the discrepancy are Ω_{eff}, N_{days} and N_{bls}, as in Equation (22) and described in Section 4.2, which when combined decreases our sensitivity (higher noise floor) by a factor of ~ 7 in mK^2.

where, in the A15 analysis, $N_{\mathrm{gps}} = 5$. Each baseline group averages down linearly as the number of baselines entering the group $(N_{\mathrm{bls\_total}}/N_{\mathrm{gps}})$ and then as the square root of the number of cross-multiplied pairs $\left(\sqrt{\frac{N_{\mathrm{gps}}^2 - N_{\mathrm{gps}}}{2}}\right)$. A revised A15 analysis should therefore use $N_{\mathrm{bls}} \sim 32$ instead of 51, and this change is taken into account in Figure 17. However, the analysis in this paper and in Kolopanis et al. (*in prep.*) no longer averages baselines into groups $(N_{\mathrm{gps}} = 1)$. For the subset of data presented in this paper, $N_{\mathrm{bls}} = 10$.

- N_{pols}: The number of polarizations averaged together. For the case of Stokes I, $N_{\mathrm{pols}} = 2$.

An additional factor of $\sqrt{2}$ is gained in sensitivity when folding together positive and negative k's to form $\Delta^2(k)$.

Our revised sensitivity estimate for the A15 analysis of PAPER-64 is shown in Figure 17. Together, the revised parameters yield a decrease in sensitivity (higher noise floor) by a factor of ~ 7 in mK^2.

To verify our thermal noise prediction, we form power spectra estimates using a pure noise simulation. We create Gaussian random noise assuming a constant T_{rcvr} (translated into T_{sys} via Equation (23)) but accounting

for the true N_{days} as determined by LST sampling counts for each time and frequency in the LST-binned data. We convert T_{sys} into a root-mean-square variance statistic using:

$$T_{\mathrm{rms}} = \frac{T_{\mathrm{sys}}}{\sqrt{\Delta\nu\Delta t N_{\mathrm{days}} N_{\mathrm{pols}}}}, \qquad (27)$$

where $\Delta\nu$ is the channel spacing, Δt is the integration time, N_{days} is the number of daily counts for a particular time and frequency that went into our LST-binned set, and N_{pols} is the number of polarizations (2 for Stokes I). This temperature sets the variance of the Gaussian random noise.
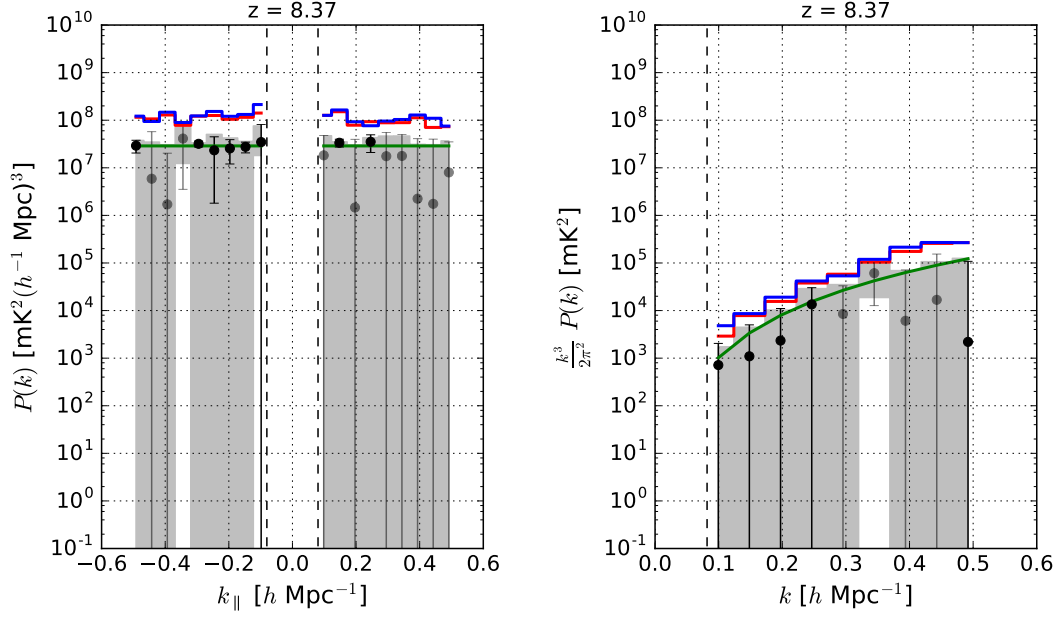
Power spectrum results for the noise simulation, which uses our full power spectrum pipeline, are shown in Figure 18. We highlight that the bootstrapped data (black and gray points, with 2σ error bars) and thermal noise prediction (solid green) show good agreement, as bootstrapping provides an accurate estimate of the noise variance. However, the limits from the full signal loss framework (weighted and unweighted in red and blue, respectively) are inflated, likely due to the additional inclusion of sample variance that comes from the EoR simulations. While the noise simulation provides an important indicator about the accuracy of our theoretical noise calculation, we note that the calculation did not take into account additional sources of error associated with earlier analysis steps (for example, Trott & Wayth (2017) show how calibration specifically can add errors to visibilities). We therefore interpret our noise prediction as the sensitivity floor for our measurements.

5. CONCLUSION

Although current 21 cm published power spectrum upper limits lie several orders of magnitude above predicted EoR levels, ongoing analyses of deeper sensitivity datasets from PAPER, MWA, and LOFAR, as well as next generation instruments like HERA, are expected to continue to push towards EoR sensitivities. As the field progresses towards a detection, we have shown that it is crucial for future analyses to have a rigorous understanding of signal loss in an analysis pipeline and be able to accurately and robustly calculate both power spectrum and theoretical errors.

In particular, in this paper we have investigated the subtleties and tradeoffs of common 21 cm power spectrum techniques on signal loss and error estimation, which can be summarized as follows:

- Substantial signal loss can result when weighting data using empirically estimated covariances due to couplings with the data realizations (Section 2). Loss of the 21 cm signal is especially significant the fewer number of independent modes that exist in the data. Hence, there exists a trade-off between sensitivity driven time-averaging techniques such as fringe-rate filtering and signal loss when using empirically estimated covariances.

Figure 18. The power spectrum for a noise simulation that mimics the noise level of a subset of PAPER-64 data, where the solid red curve is the 2σ upper limit on the EoR signal estimated from our signal injection framework using $\widehat{\mathbf{C}}_{\mathrm{eff}}$. The theoretical 2σ thermal noise level prediction based on observational parameters (calculated by Equation (22)) is in green. Additionally, the power spectrum result for the uniform weighted case is shown in three different ways: power spectrum values (black and gray points as positive and negative values, respectively, with 2σ error bars from bootstrapping), the 2σ upper limit on the EoR signal using our full signal injection framework (solid blue), and the measured power spectrum values with 2σ thermal noise errors (gray shaded regions). The vertical dashed black lines signify the horizon limit for this analysis using 30 m baselines. We highlight that the bootstrapped data points and thermal noise prediction show good agreement, while the limits from the full injection framework (red and blue) are inflated due to the additional inclusion of sample variance that comes from the injection simulations.

- Signal injection and recovery simulations can be used to quantify signal loss (Section 3.1). However, a signal-only simulation (i.e., comparing a uniformly weighted vs. weighted power spectrum of EoR only) can underestimate loss by failing to account for correlations between the data and signal which can be large and negative.

- Errors that are estimated via bootstrapping can be underestimated if samples in the dataset are significantly correlated (Section 4.1). However, if the number of independent samples in a dataset is well-determined, bootstrapping is a simple and accurate way of estimating errors.

As a consequence of our investigations, we have also used a subset of PAPER-64 data to make a new power spectrum analysis. This serves as an illustrative example of using a signal injection framework, correctly computing errors via bootstrapping, and accurately estimating thermal noise. Our revised PAPER-64 limits are presented in Kolopanis et al. (*in prep.*), which supersede all previously published PAPER limits. Because of the many challenges associated with signal loss and its estimation as described in this paper, Kolopanis et al. (*in prep.*) use a straightforward power spectrum estimation approach that is not lossy. However, the main reasons for a previously underestimated limit and ways in which

our new analysis differs can still be summarized by the following:

- Signal loss, previously found to be < 2% in A15, was underestimated by a factor of >1000 for the case of empirically estimated inverse covariance weighting. Using a regularized covariance weighting method can minimize loss (Section 3.3), however, because a regularized weighting method is not as aggressive as the former, it produces limits that are still higher than the lossy empirical inverse covariance limits. Underestimated signal loss therefore represents the bulk of our revision.

- Power spectrum errors, originally computed by bootstrapping, were underestimated for the data by a factor of ~ 2 in mK due to oversampling data whose effective number of independent samples was reduced from fringe-rate filtering (Section 4.1). Several other errors were also found regarding error estimation, though with smaller effects.

- Several factors used in an analytic expression to predict the noise-level in PAPER-64 data were revised, yielding a decrease in predicted sensitivity level by a factor of ~ 3 in mK (Section 4.2). We note that our sensitivity prediction is revised by a factor less than our overall power spectrum result,

implying that if taken at face value, the theoretical prediction for noise in A15 was too high for its data points.

The future of 21 cm cosmology is exciting, as new experiments have sensitivities that expect to reach and surpass EoR levels, improved foreground mitigation and removal strategies are being developed, and simulations are being designed to better understand instruments. On the power spectrum analysis side, robust signal loss simulations and precise error calculations will play critical roles in accurate 21 cm results. With strong foundations being established now, it is safe to say that we can expect to learn much about reionization and our early Universe in the coming years.

APPENDIX

A. A TOY MODEL FOR INVERSE COVARIANCE WEIGHTING

In this Appendix, we focus on the optimal quadratic estimator, Equation (3), and mathematically illustrate its role in estimating the power spectrum of EoR. While there exists detailed literature about quadratic estimators in general (e.g., Liu & Tegmark 2011; Trott et al. 2012; Liu et al. 2014b; Dillon et al. 2014), here we focus on two simple cases in order to outline one situation where the estimator successfully suppresses contamination and one where it does not. By describing these two cases, we hope to clarify and motivate the desire to use OQE while also understanding its limitations.

We specifically choose toy models where the data covariance is diagonal, as indeed we expect the EoR signal to be. We assume we have N data points Δ_i which are the sum of a desired signal σ_i and an undesired contaminant v_i

$$\Delta_i = \sigma_i + v_i \tag{A1}$$

with

$$\langle \sigma_i \rangle = 0; \ \langle \sigma_i^2 \rangle = s; \text{ and } \langle \boldsymbol{\sigma}\boldsymbol{\sigma}^T \rangle = s\mathbf{I}_{N \times N} \equiv \mathbf{S}, \tag{A2}$$

where we wish to estimate s. The contaminant in this first case has a similar structure (as the EoR) for its covariance, and is assumed uncorrelated with the signal

$$\langle v_i \rangle = 0; \ \langle v_i^2 \rangle = u; \ \langle \boldsymbol{v}\boldsymbol{v}^T \rangle = u\mathbf{I}_{N \times N} \equiv \mathbf{U}; \text{and } \langle \sigma_i v_j \rangle = 0. \tag{A3}$$

With the covariance matrix given by $\mathbf{C} = \mathbf{S} + \mathbf{U}$, the estimator for s using only the quadratic part of Equation 3 is

$$\hat{s} = \frac{\boldsymbol{\Delta}^T \boldsymbol{\Delta}}{N} \tag{A4}$$

and its expectation is

$$\langle \hat{s} \rangle = s + u. \tag{A5}$$

Thus, *when the covariance structure of the contaminant is identical to the signal* ($\frac{\partial \mathbf{S}}{\partial s} = \frac{\partial \mathbf{U}}{\partial u} = \frac{\partial \mathbf{C}}{\partial s}$), the information available to the quadratic portion of the estimator to distinguish between the two is degenerate, and knowledge only of $\mathbf{C}$ and $\frac{\partial \mathbf{C}}{\partial s}$ is inadequate. In order to obtain an unbiased estimate of s, one must also use knowledge of $\mathbf{U}$. Indeed, computing the linear bias from Equation (6), one finds $b = u$.

Now consider a case, chosen to be very similar to the toy model in 2.2, in which the data again have an additive contaminant, now given by

$$\Delta_i = \sigma_i + vm_i \tag{A6}$$

where the properties of σ_i are as before, but now v is a random variable and m_i is a fixed function of i with

$$\langle v \rangle = 0; \ \langle v^2 \rangle = u; \ \langle \boldsymbol{v}\boldsymbol{v}^T \rangle = u\mathbf{m}\mathbf{m}^T \equiv \mathbf{U}; \ \mathbf{m}^T\mathbf{m} = 1; \text{ and } \langle \sigma_i v \rangle = 0. \tag{A7}$$

Here $\mathbf{m}$ represents a mode which is correlated across many data points (i.e., we are assuming $\mathbf{U}$ need *not* be diagonal), with amplitude given by v. The normalization of $\mathbf{m}$ is a matter of convention, and can be absorbed in the variance u; the choice above will be convenient for understanding the limiting case $u \gg s$.

We can calculate the quadratic portion of the estimator explicitly by using the Sherman-Morrison identity to invert the covariance matrix. Defining

$$\xi \equiv \frac{u/s}{1 + u/s}, \tag{A8}$$

we have

$$\mathbf{C}^{-1} = \frac{1}{s}\left(\mathbf{I} - \xi\mathbf{m}\mathbf{m}^T\right) \tag{A9}$$

and

$$\hat{s} = \frac{\boldsymbol{\Delta}^T(\mathbf{I} + (\xi^2 - 2\xi)\mathbf{m}\mathbf{m}^T)\boldsymbol{\Delta}}{N + \xi^2 - 2\xi} \tag{A10}$$

with expectation

$$\langle\hat{s}\rangle = s + \frac{1 - 2\xi + \xi^2}{N + -2\xi + \xi^2}u. \tag{A11}$$

It is worth observing immediately that there is no multiplicative bias on s, and that the additive bias is strictly $< u/N$.

An instructive limit is $u \gg s$, $\xi \to 1$, in which case the virtue of weighting by $\mathbf{C}^{-1}$ becomes clearer, as it becomes

$$\mathbf{C}^{-1} = \frac{1}{s}\left(\mathbf{I} - \mathbf{m}\mathbf{m}^T\right) \tag{A12}$$

where $\mathbf{I} - \mathbf{m}\mathbf{m}^T$ is the projection operator, projecting out $\mathbf{m}$ from any vector it acts on, and further, the linear bias tends to 0 as $\xi \to 1$ (i.e., the projection is "perfect" and not "undone" by the Fisher matrix normalization).

This is the ideal case for the inverse covariance weighting performed in the PAPER analysis, where removal of contamination with a known covariance can be suppressed by a kind of projection of the offending modes. But even in this case, it is worth pointing out that the estimator still has a linear bias for finite u. We have also assumed that the contaminating mode $\mathbf{m}$ is known perfectly; the next appendix takes up the case where the modes are estimated from the data.

B. A TOY MODEL FOR SIGNAL LOSS

In this Appendix, we examine a toy model for signal loss. Our goal is to derive an analytic formula for power spectrum signal loss. While this model does not apply generally to all the scenarios presented in this paper, it provides some analytic intuition for how the coupling between data and an empirical covariance can result in signal loss.

The minimum-variance quadratic estimator $\widehat{P}^\alpha$ for the αth bandpower of the power spectrum is given by

$$\widehat{P}^\alpha = \frac{1}{2\mathbf{F}^{\alpha\alpha}}\mathbf{x}^t\mathbf{C}^{-1}\mathbf{Q}^\alpha\mathbf{C}^{-1}\mathbf{x}, \tag{B13}$$

where

$$F^{\alpha\alpha} \equiv \frac{1}{2}\mathrm{tr}\left(\mathbf{C}^{-1}\mathbf{Q}^\alpha\mathbf{C}^{-1}\mathbf{Q}^\alpha\right) \tag{B14}$$

is the αth diagonal element of the Fisher matrix. For this section only, with no loss of generality, we assume that the data $\mathbf{x}$ are real. We also assume for simplicity that $\mathbf{x}$ is the data from a single instant in time, so that it is of length N_f, where N_f is the number of frequency channels.

In our case, we do not have *a priori* knowledge of the covariance matrix. Thus, we deviate from the true minimum-variance quadratic estimator and replace $\mathbf{C}$ with $\widehat{\mathbf{C}}$, its data-derived approximation. Our estimator then becomes

$$\widehat{P}^\alpha_{\mathrm{loss}} = \frac{1}{2\widehat{\mathbf{F}}^{\alpha\alpha}}\mathbf{x}^t\widehat{\mathbf{C}}^{-1}\mathbf{Q}^\alpha\widehat{\mathbf{C}}^{-1}\mathbf{x}, \tag{B15}$$

where

$$\widehat{F}^{\alpha\alpha} \equiv \frac{1}{2}\mathrm{tr}\left(\widehat{\mathbf{C}}^{-1}\mathbf{Q}^\alpha\widehat{\mathbf{C}}^{-1}\mathbf{Q}^\alpha\right), \tag{B16}$$

with the label "loss" to foreshadow the fact that this will be an estimator with signal loss (i.e., a multiplicative bias of less than unity). We will now provide an explicit demonstration of this by modeling the estimated covariance as

$$\widehat{\mathbf{C}} = (1 - \eta)\mathbf{C} + \eta\mathbf{x}\mathbf{x}^t, \tag{B17}$$

where η is a parameter quantifying our success at estimating the true covariance matrix. If $\eta = 0$, our covariance estimate has perfectly modeled the true covariance and $\widehat{\mathbf{C}} = \mathbf{C}$. On the other hand, if $\eta = 1$, then our covariance estimate is based purely on the one realization of the covariance that is our actual data, and we would expect a high level of overfitting and signal loss.

Our strategy for computing the signal loss will be to insert Equation (B17) into Equation (B15) and to express the resulting estimator $\widehat{P}_{\text{loss}}^\alpha$ in terms of $\widehat{P}^\alpha$. We begin by expressing $\widehat{\mathbf{C}}^{-1}$ in terms of $\mathbf{C}^{-1}$ using the Woodbury identity so that

$$\widehat{\mathbf{C}}^{-1} = \frac{\mathbf{C}^{-1}}{1 - \eta} \left[\mathbf{I} - \frac{\eta \mathbf{x} \mathbf{x}^t \mathbf{C}^{-1}}{1 + \eta(g-1)} \right], \tag{B18}$$

where we have defined $g \equiv \mathbf{x}^t \mathbf{C}^{-1} \mathbf{x}$. Inserting this into our Fisher estimate we have

$$\widehat{F}^{\alpha\alpha} = \frac{F^{\alpha\alpha}}{(1-\eta)^2} \left[1 - \frac{\eta}{1 + \eta(g-1)} \frac{h^{\alpha\alpha}}{F^{\alpha\alpha}} + \frac{1}{2} \left(\frac{\eta}{1 + \eta(g-1)} \right)^2 \frac{(h^\alpha)^2}{F^{\alpha\alpha}} \right], \tag{B19}$$

where $h^\alpha \equiv \mathbf{x}^t \mathbf{C}^{-1} \mathbf{Q}^\alpha \mathbf{C}^{-1} \mathbf{x}$ and $h^{\alpha\alpha} \equiv \mathbf{x}^t \mathbf{C}^{-1} \mathbf{Q}^\alpha \mathbf{C}^{-1} \mathbf{Q}^\alpha \mathbf{C}^{-1} \mathbf{x}$. Note that g, h^α, and $h^{\alpha\alpha}$ are all random variables, since they depend on $\mathbf{x}$. Inserting these expressions into our estimator gives

$$\widehat{P}_{\text{loss}}^\alpha = \frac{1}{2} \frac{h^\alpha}{F^{\alpha\alpha}} \left[1 - \frac{\eta g}{1 + \eta(g-1)} \right]^2 \left[1 - \frac{\eta}{1 + \eta(g-1)} \frac{h^{\alpha\alpha}}{F^{\alpha\alpha}} + \frac{1}{2} \left(\frac{\eta}{1 + \eta(g-1)} \right)^2 \frac{(h^\alpha)^2}{F^{\alpha\alpha}} \right]^{-1}. \tag{B20}$$

Both for the purposes of analytical tractability and to provide intuition, we expand this expression to leading order in η. This approximates the limiting case where the covariance $\widehat{\mathbf{C}}$ is close to the ideal and the lossy covariance is a small perturbation. The result is

$$\widehat{P}_{\text{loss}}^\alpha \approx \frac{1}{2} \frac{h^\alpha}{F^{\alpha\alpha}} \left[1 - \eta \left(g - \frac{h^{\alpha\alpha}}{F^{\alpha\alpha}} \right) \right]. \tag{B21}$$

Taking the ensemble average of both sides and noting that the true power spectrum P^α is equal to $\langle h^\alpha \rangle / 2F^{\alpha\alpha}$, we obtain

$$\langle \widehat{P}_{\text{loss}}^\alpha \rangle \approx (1 - \eta N_f) P^\alpha + 4\eta \frac{\text{tr}(\mathbf{C}^{-1} \mathbf{Q}^\alpha \mathbf{C}^{-1} \mathbf{Q}^\alpha \mathbf{C}^{-1} \mathbf{Q}^\alpha)}{[\text{tr}(\mathbf{C}^{-1} \mathbf{Q}^\alpha \mathbf{C}^{-1} \mathbf{Q}^\alpha)]^2} \approx (1 - \eta N_f) P^\alpha, \tag{B22}$$

where recall that N_f is the length of $\mathbf{x}$, or the number of frequency channels. In the last step we dropped the final term, since it scales as ηP^α (without the factor of N) and is therefore typically small compared to the terms that have been retained.

Recalling that P^α is the *true* power spectrum, one sees that when the covariance in the optimal quadratic estimator is naively replaced by an empirical covariance, the resulting power spectrum estimate is biased low, i.e., there is signal loss. This occurs because of couplings between $\widehat{\mathbf{C}}$ and $\mathbf{x}$, which formally means that what was originally a quadratic estimator is no longer quadratic, but contains higher-order correlations. This violates the assumptions implicit in the derivation of $F^{\alpha\alpha}$ as the normalization factor for converting unnormalized bandpowers $\frac{1}{2} \mathbf{x}^t \mathbf{C}^{-1} \mathbf{Q}^\alpha \mathbf{C}^{-1} \mathbf{x}$ into properly normalized power spectrum estimates, where the unnormalized bandpowers are assumed to be two-point (i.e., quadratic) statistics (Liu & Tegmark 2011). The result is an improperly normalized—and thus lossy—power spectrum estimate.

C. DERIVATION OF THE JEFFREYS PRIOR

The Jeffreys prior is an objective, non-informative prior distribution for a parameter space using Bayesian probability (Jaynes 1968). For the signal injection framework outlined in Section 3.2, we wish to compute the prior $p(P_{\text{in}})$, or the probability density of the power spectrum of the EoR signal.
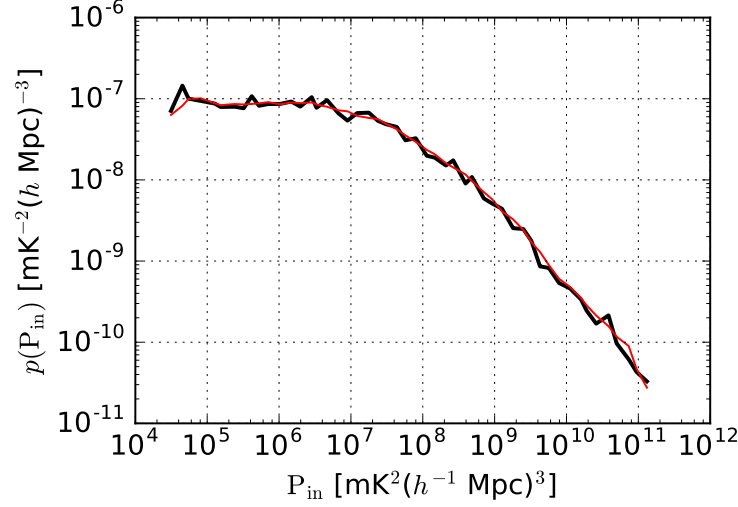
The Jeffreys prior is defined as:

$$p(P_{\text{in}}) \propto \sqrt{\left\langle \left(\frac{\partial \mathcal{L}}{\partial P_{\text{in}}} \right)^2 \right\rangle}, \tag{C23}$$

where

$$\mathcal{L} = \ln p(\widehat{P}_{\text{out}} | P_{\text{in}}), \tag{C24}$$

recalling that in our framework P_{in} is the power spectrum of the EoR signal (uniformly weighted), and $\widehat{P}_{\text{out}}$ is the weighted output power spectrum of the data plus EoR.

Figure C1. An example of the typical Jeffreys prior shape for the PAPER-64 analysis as computed by Equation (C26) (black). We smooth the prior using a sliding boxcar average over every 5 injection levels (red). Most noticeably, the Jeffreys prior is not constant with $P_{\rm in}$, meaning a uniform prior would be an informative prior.

For a single injection amplitude, our bootstrapped $\widehat{P}_{\rm out}$ values are well-approximated by a Gaussian distribution. Simplifying our notation so that $x = P_{\rm in}$ and $y = \widehat{P}_{\rm out}$:

$$p(y|x) = \frac{1}{\sigma(x)\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{y-\bar{y}(x)}{\sigma}\right)^2}, \tag{C25}$$

where σ is the standard deviation of $\widehat{P}_{\rm out}$ and $\bar{y}$ is the mean of $\widehat{P}_{\rm out}$, and they are both functions of $P_{\rm in}$. Combining Equations (C23), (C24), and (C25) the Jeffreys prior simplifies to become:

$$p(x) \propto \sqrt{\frac{1}{\sigma^2}\left(2\left(\frac{\partial\sigma}{\partial x}\right)^2 + \left(\frac{\partial\bar{y}}{\partial x}\right)^2\right)}. \tag{C26}$$

When we simulate our full injection framework as in Section 3.2, we sample 50 $P_{\rm in}$ values that range from $\sim 10^5\,{\rm mK}^2\,(h^{-1}\,{\rm Mpc})^3$ to $\sim 10^{11}\,{\rm mK}^2\,(h^{-1}\,{\rm Mpc})^3$, and we note that the prior is set to zero outside those regions. For the injections that we do sample, we can simply fit analytic functions to the mean and standard deviations of $\widehat{P}_{\rm out}$ ($\bar{y}$ and σ) as functions of $P_{\rm in}$.

We show the typical shape of the Jeffreys prior used in our analysis in Figure C1, as computed by Equation (C26). Most noticeably, it is not constant with $P_{\rm in}$, meaning a uniform prior, which is often used for simplicity, is informative in our application. Therefore, due to its objective nature we choose to use a Jeffreys prior in our analysis, multiplying our likelihood functions by Equation (C26) before computing posterior distributions.

REFERENCES

Ade, P., et al. 2008, Astrophysical Journal, 674, 22

Ade, P. A. R., et al. 2017, PhRvD, 96, 102003

Ali, S. S., Bharadwaj, S., & Chengalur, J. N. 2008, MNRAS, 385, 2166

Ali, Z. S., et al. 2015, ApJ, 809, 61

Andrae, R. 2010, ArXiv e-prints

Araujo, D., et al. 2012, The Astrophysical Journal, 760, 145

Barkana, R., & Loeb, A. 2001, PhR, 349, 125

—. 2008, Monthly Notices of the Royal Astronomical Society, 384, 1069

Beardsley, A. P., et al. 2016, The Astrophysical Journal, 833, 102

Bernardi, G., et al. 2009, A&A, 500, 965

—. 2010, A&A, 522, A67+

Bernardi, G., et al. 2013, The Astrophysical Journal, 771, 105

Bernardi, G., et al. 2016, MNRAS, 461, 2847

BICEP2 Collaboration et al. 2016, ApJ, 833, 228

Bischoff, C., et al. 2011, The Astrophysical Journal, 741, 111

Bond, J. R., Jaffe, A. H., & Knox, L. 1998, PhRvD, 57, 2117

Bowman, J. D., & Rogers, A. E. E. 2010, Nature, 468, 796

Bowman, J. D., Rogers, A. E. E., Monsalve, R. A., Mozdzen, T. J., & Mahesh, N. 2018, Nature, 555, 67

Burns, J. O., et al. 2012, Advances in Space Research, 49, 433

Chang, T.-C., Pen, U.-L., Bandura, K., & Peterson, J. B. 2010, Nature, 466, 463

Chapman, E., et al. 2012, Monthly Notices of the Royal Astronomical Society, 423, 2518

Chiang, H. C., et al. 2010, The Astrophysical Journal, 711, 1123

Crites, A. T., et al. 2015, The Astrophysical Journal, 805, 36

Das, S., et al. 2011a, Physical Review Letters, 107, 021301

—. 2011b, ApJ, 729, 62

Datta, A., Bowman, J. D., & Carilli, C. L. 2010, The Astrophysical Journal, 724, 526

de Oliveira-Costa, A., Tegmark, M., Gaensler, B. M., Jonas, J., Landecker, T. L., & Reich, P. 2008, MNRAS, 388, 247

DeBoer, D. R., et al. 2017, Publications of the Astronomical Society of the Pacific, 129, 045001

Dillon, J. S., Liu, A., & Tegmark, M. 2013, PhRvD, 87, 043005

Dillon, J. S., & Parsons, A. R. 2016, The Astrophysical Journal, 826, 181

Dillon, J. S., et al. 2014, Phys. Rev. D, 89, 023002

—. 2015, Phys. Rev. D, 91, 123011

Dodelson, S., & Schneider, M. D. 2013, PhRvD, 88, 063537

Efron, B., & Tibshirani, R. 1994, An Introduction to the Bootstrap, Chapman & Hall/CRC Monographs on Statistics & Applied Probability (Taylor & Francis)

Ewall-Wice, A., Dillon, J. S., Liu, A., & Hewitt, J. 2017, MNRAS, 470, 1849

Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, PhR, 433, 181

Ghosh, A., Bharadwaj, S., Ali, S. S., & Chengalur, J. N. 2011, MNRAS, 418, 2584

Hartlap, J., Simon, P., & Schneider, P. 2007, A&A, 464, 399

Jacobs, D. C., et al. 2015, ApJ, 801, 51

Jacobs, D. C., et al. 2016, ApJ, 825, 114

Jaynes, E. 1968, IEEE Transactions on Systems Science and Cybernetics, 4, 227

Jelić, V., et al. 2008, MNRAS, 389, 1319

Joachimi, B. 2017, MNRAS, 466, L83

Keating, G. K., Marrone, D. P., Bower, G. C., Leitch, E., Carlstrom, J. E., & DeBoer, D. R. 2016, The Astrophysical Journal, 830, 34

Kerrigan, J., et al. 2018, ArXiv e-prints

Kohn, S. A., et al. 2016, ApJ, 823, 88

Koopmans, L., et al. 2015, Advancing Astrophysics with the Square Kilometre Array (AASKA14), 1

Liu, A., & Parsons, A. R. 2016, Monthly Notices of the Royal Astronomical Society, 457, 1864

Liu, A., Parsons, A. R., & Trott, C. M. 2014a, PhRvD, 90, 023018

—. 2014b, PhRvD, 90, 023019

Liu, A., & Tegmark, M. 2011, Phys. Rev. D, 83, 103006

Loeb, A., & Furlanetto, S. 2013, The First Galaxies in the Universe (Princeton University Press)

Masui, K. W., et al. 2013, ApJL, 763, L20

Moore, D. F., Aguirre, J. E., Parsons, A. R., Jacobs, D. C., & Pober, J. C. 2013, The Astrophysical Journal, 769, 154

Morales, M. F., & Wyithe, J. S. B. 2010, ARA&A, 48, 127

Paciga, G., et al. 2013a, MNRAS

—. 2013b, MNRAS, 433, 639

Padmanabhan, N., White, M., Zhou, H. H., & O'Connell, R. 2016, MNRAS, 460, 1567

Parsons, A., Pober, J., McQuinn, M., Jacobs, D., & Aguirre, J. 2012a, ApJ, 753, 81

Parsons, A. R., Liu, A., Ali, Z. S., & Cheng, C. 2016, ApJ, 820, 51

Parsons, A. R., Pober, J. C., Aguirre, J. E., Carilli, C. L., Jacobs, D. C., & Moore, D. F. 2012b, ApJ, 756, 165

Parsons, A. R., et al. 2010, AJ, 139, 1468

—. 2014, ApJ, 788, 106

Patil, A. H., et al. 2016, MNRAS, 463, 4317

—. 2017, ApJ, 838, 65

Patra, N., Subrahmanyan, R., Sethi, S., Udaya Shankar, N., & Raghunathan, A. 2015, ApJ, 801, 138

Paz, D. J., & Sánchez, A. G. 2015, MNRAS, 454, 4326

Pearson, D. W., & Samushia, L. 2016, MNRAS, 457, 993

Peterson, U.-L. P. X.-P. W. J. 2004, ArXiv Astrophysics e-prints

Petrovic, N., & Oh, S. P. 2011, MNRAS, 413, 2103

Pober, J. C., Greig, B., & Mesinger, A. 2016a, MNRAS, 463, L56

Pober, J. C., et al. 2012, AJ, 143, 53

Pober, J. C., et al. 2013, The Astrophysical Journal Letters, 768, L36

Pober, J. C., et al. 2013, AJ, 145, 65

—. 2014, ApJ, 782, 66

—. 2016b, ApJ, 819, 8

Pope, A. C., & Szapudi, I. 2008, MNRAS, 389, 766

Pritchard, J. R., & Loeb, A. 2010, PhRvD, 82, 023006

—. 2012, Reports on Progress in Physics, 75, 086901

Quenouille, M. H. 1949, Ann. Math. Statist., 20, 355

Santos, M. G., Cooray, A., & Knox, L. 2005, ApJ, 625, 575

Sellentin, E., & Heavens, A. F. 2016, MNRAS, 456, L132

Sherwin, B. D., et al. 2017, Phys. Rev. D, 95, 123529

Sokolowski, M., et al. 2015, PASA, 32, e004

Switzer, E. R., Chang, T.-C., Masui, K. W., Pen, U.-L., & Voytek, T. C. 2015, ApJ, 815, 51

Switzer, E. R., et al. 2013, MNRAS, 434, L46

Taylor, A., & Joachimi, B. 2014, MNRAS, 442, 2728

Tegmark, M. 1997, PhRvD, 55, 5895

Thompson, A. R., Moran, J. M., & Swenson, Jr., G. W. 2001, Interferometry and Synthesis in Radio Astronomy, 2nd Edition

Thyagarajan, N., et al. 2013, ApJ, 776, 6

Tingay, S. J., et al. 2013, PASA, 30, 7

Trott, C. M., & Wayth, R. B. 2017, Publications of the Astronomical Society of Australia, 34, e061

Trott, C. M., Wayth, R. B., & Tingay, S. J. 2012, ApJ, 757, 101

Trott, C. M., et al. 2016, The Astrophysical Journal, 818, 139

Tukey. 1958, Ann. Math. Statist., 29, 614

van Haarlem, M. P., et al. 2013, A&A, 556, A2

Vedantham, H., Shankar, N. U., & Subrahmanyan, R. 2012, The Astrophysical Journal, 745, 176

Voytek, T. C., Natarajan, A., Jáuregui García, J. M., Peterson, J. B., & López-Cruz, O. 2014, ApJL, 782, L9

Wang, J., et al. 2013, The Astrophysical Journal, 763, 90

Wolz, L., Abdalla, F. B., Blake, C., Shaw, J. R., Chapman, E., & Rawlings, S. 2014, Monthly Notices of the Royal Astronomical Society, 441, 3271

Wu, X. 2009, in Bulletin of the American Astronomical Society, Vol. 41, American Astronomical Society Meeting Abstracts #213, 474