

UC Office of the President

Recent Work

Title

The impact of modelling errors on interferometer calibration for 21 cm power spectra

Permalink

<https://escholarship.org/uc/item/2vw319c7>

Journal

Monthly Notices of the Royal Astronomical Society, 470(2)

ISSN

0035-8711 1365-2966

Authors

Ewall-Wice, Aaron

Dillon, Joshua S

Liu, Adrian

et al.

Publication Date

2017-09-01

DOI

10.1093/mnras/stx1221

Peer reviewed

The impact of modelling errors on interferometer calibration for 21 cm power spectra

Aaron Ewall-Wice,^{1,2★} Joshua S. Dillon,^{3,4} Adrian Liu^{3†} and Jacqueline Hewitt^{1,2}

¹MIT Kavli Institute for Astrophysics and Space Research, Cambridge, MA 02139, USA

²Department of Physics and MIT Kavli Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

³Department of Astronomy and Radio Astronomy Laboratory, UC Berkeley, Berkeley, CA 94720, USA

⁴Berkeley Center for Cosmological Physics, University of California, Berkeley, Berkeley, CA 94720, USA

Accepted 2017 May 16. Received 2017 May 14; in original form 2017 January 10

ABSTRACT

We study the impact of sky-based calibration errors from source mismodelling on 21 cm power spectrum measurements with an interferometer and propose a method for suppressing their effects. While emission from faint sources that are not accounted for in calibration catalogues is believed to be spectrally smooth, deviations of true visibilities from model visibilities are not, due to the inherent chromaticity of the interferometer’s sky response (the ‘wedge’). Thus, unmodelled foregrounds, below the confusion limit of many instruments, introduce frequency structure into gain solutions on the same line-of-sight scales on which we hope to observe the cosmological signal. We derive analytic expressions describing these errors using linearized approximations of the calibration equations and estimate the impact of this bias on measurements of the 21 cm power spectrum during the epoch of reionization. Given our current precision in primary beam and foreground modelling, this noise will significantly impact the sensitivity of existing experiments that rely on sky-based calibration. Our formalism describes the scaling of calibration with array and sky-model parameters and can be used to guide future instrument design and calibration strategy. We find that sky-based calibration that downweights long baselines can eliminate contamination in most of the region outside of the wedge with only a modest increase in instrumental noise.

Key words: instrumentation: interferometers – techniques: interferometric – dark ages, reionization, first stars – radio lines: general.

1 INTRODUCTION

Observations of redshifted 21 cm emission are poised to unveil the properties of the earliest luminous sources in the Universe, their impact on the global state of the intergalactic medium and how they affected the subsequent generations of stars and galaxies (see Furlanetto 2016; McQuinn 2016 for recent reviews).

One approach to detecting the cosmological 21 cm signal is to measure the fluctuations in the brightness temperature that can be mapped tomographically with a radio interferometer. To enhance the significance of a detection, most experiments are attempting to measure the spherically averaged power spectrum of these fluctuations. The mitigation of foregrounds that are four to five orders of magnitude brighter than the signal itself is a central challenge that 21 cm experiments must overcome but is greatly aided by the spectral smoothness of these foregrounds (Di Matteo et al.

2002; Oh & Mack 2003; Morales & Hewitt 2004; Zaldarriaga, Furlanetto & Hernquist 2004). An interferometer measures the brightness distribution on the sky by cross-correlating the outputs from many pairs of antennas. Flat-spectrum radio waves from a single point source, at a given time of observation, appear at a fixed time delay in the correlation between two antennas. Since the delay between two correlated antenna outputs is the Fourier dual to frequency, each fixed-delay source introduces a sinusoidal ripple as a function of frequency with a period that is inversely proportional to the difference of the arrival times of that source at the two correlated antennas. This sinusoid in frequency will correspond to a single comoving cosmological mode. In the absence of reflections, the maximal delay between signals arriving from a source on the sky [corresponding to the maximal line-of-sight (LoS) cosmological Fourier mode that is contaminated] occurs when the source is located along the separation of the antennas, at the horizon. Hence, as viewed by an interferometer, the spectrally smooth foregrounds are naturally contained within a region of Fourier space known as the *wedge* (Datta, Bowman & Carilli 2010; Morales et al. 2012; Parsons et al. 2012b; Vedantham, Udaya Shankar &

* E-mail: aaronew@mit.edu

† Hubble Fellow.

Subrahmanyan 2012; Thyagarajan et al. 2013; Liu, Parsons & Trott 2014a,b), which is given by the horizon delay for each baseline separation and increases with that separation.

It is also possible for the signal chain of the instrument to imprint spectral structure into the measured visibilities. For example, a reflection within the signal path can delay the correlated signal. Hence, longer delays in the signal path contaminate finer frequency scales and are capable of leaking significant power outside of the wedge (Beardsley et al. 2016; Ewall-Wice et al. 2016a,b). Digital artefacts can also introduce fine spectral features such as those introduced by the polyphase filter bank on the Murchison Widefield Array (MWA; Offringa et al. 2016). For each delay that is contaminated by structure in the antenna gains, an attenuated copy of the foregrounds, which are $\sim 10^4$ – 10^5 times larger than the signal, is introduced. Using foreground simulations, Thyagarajan et al. (2016) establish that in order to avoid contaminating the comoving LoS scales of several h^{-1} Mpc or smaller, which are targeted by 21 cm experiments, instrumental chromaticity beyond a 250 ns delay must be suppressed to the ≈ -50 dB level. Thus, (a) the smoothness of the instrumental gain must meet this specification, or (b) calibration methods must be capable of suppressing any instrumental spectral structure to be within these limits.

Interferometric experiments have taken several distinct approaches to calibrating out instrumental spectral structure. Experiments focusing on imaging, such as the MWA (Tingay et al. 2013), the Low-Frequency Array (LOFAR; van Haarlem et al. 2013), early deployments of the Precision Array for Probing the Epoch of Reionization (PAPER; Jacobs et al. 2011, 2013; Kohn et al. 2016) and the Giant Metrewave Radio Telescope (GMRT; Paciga et al. 2013),¹ calibrate their gains on a model of the sky that is usually iteratively improved with self-calibration (where observed sources are fed into an updated sky model that is used to obtain more accurate gain solutions). Pipelines such as the MWA’s real time system (Mitchell et al. 2008), fast holographic deconvolution (Sullivan et al. 2012) and *sageCAL* (Kazemi et al. 2011; Kazemi & Yatawatta 2013; Kazemi, Yatawatta & Zaroubi 2013) rely on the modelling approach that we refer to as sky-based calibration. An alternative route is to constrain the instrumental gains using many redundant measurements of the same visibility (Wieringa 1992; Liu et al. 2010; Zheng et al. 2014). This strategy was implemented by the MIT EoR array (Zheng et al. 2014, 2017a,b), the latest configurations of PAPER (Parsons et al. 2014; Ali et al. 2015) and the Hydrogen Epoch of Reionization Array (HERA; DeBoer et al. 2017; Dillon & Parsons 2016) that is now being commissioned in South Africa.² Finally, *in situ* calibration can be obtained using the injection of known signals (Patra et al. 2017a). The Canadian Hydrogen Intensity Mapping Experiment (Newburgh et al. 2014) is employing a combination of redundant calibration, signal injection and pulsar holography to correct for instrumental gains.

Recent analyses of MWA data, using sky-based calibration, have been contaminated by intrinsic chromaticity in the signal chain at the $\lesssim -20$ dB level (Dillon et al. 2015b; Beardsley et al. 2016; Ewall-Wice et al. 2016b; Jacobs et al. 2016) out to less than a

delay of 2×10^3 ns, arising from a combination of reflections in the beam former to receiver cables and digital artefacts. Increasing the frequency degrees of freedom within sky-based calibration is a potential solution as the gains are permitted to absorb fine-scale instrumental frequency structure at high delays (Offringa et al. 2016) and improvements in features such as cable reflections were noted in power spectra calibrated with additional parameters (Trott et al. 2016).

While calibration solutions with fine-frequency degrees of freedom are able to model the detrimental spectral features in an instrumental bandpass, they are susceptible to absorbing the imperfections in any sky model used for calibration. Naively, errors in a smooth foreground model should not impart spectrally complex errors into a gain solution. However, because every gain participates in many baselines with varying lengths and (due to the wedge) intrinsic chromaticities, calibration can imprint the frequency-dependent errors of the longest baselines in which an antenna participates into its gain solution. The application of this gain solution on the short baselines that the antenna participates in will mix contamination from long to short baselines, potentially contaminating the epoch of reionization (EoR) window. Recent studies by Barry et al. (2016, henceforth B16) and Patil et al. (2016) have demonstrated the existence of these errors in simulations of the special cases of the MWA and LOFAR with specific point source realizations. It has not yet been established how these errors scale with the properties of the instrument and the source catalogue and whether they will pose a fundamental limitation to upcoming 21 cm experiments that expect to rely on sky-based calibration such as the Square Kilometre Array (SKA). Although B16 propose a low-order polynomial-based method to mitigate these effects, it generally relies upon intrinsically spectrally smooth antenna bandpasses, which may not be the case for many interferometers.

In this paper, we employ linearized approximations of the calibration equations developed by Wieringa (1992, hereafter W92) and Liu et al. (2010, hereafter L10) to investigate the amplitude of errors arising from incomplete calibration catalogues. Since these faint unmodelled sources can be described statistically (Liu & Tegmark 2011, 2012; Trott, Wayth & Tingay 2012; Dillon, Liu & Tegmark 2013; Dillon et al. 2014, 2015a; Trott et al. 2016), we will address the ensuing errors as a type of correlated noise that we will hereafter refer to as *modelling noise*. Unlike its thermal counterpart, modelling noise does not integrate down with observing time, biasing any power spectrum estimate. Since interleaved times in this noise are correlated, this bias cannot be eliminated (unlike thermal noise) by the technique of cross-multiplying interleaved time integrations (e.g. Dillon et al. 2014). We will derive equations describing the amplitude of modelling noise and its dependence on the properties of a radio interferometer such as the antenna count, distribution and element size along with the depth of the calibration catalogue. We use these equations to approximate the level of modelling noise in the existing MWA and LOFAR experiments (finding that our analytic results are in broad agreement with the simulation results in B16) along with the expected contamination in the upcoming instruments SKA-1 LOW and HERA. This contamination arises fundamentally from the chromaticity on long baselines; hence, it can be eliminated by downweighting long baselines in calibration, a strategy that we develop and verify in this paper.

We take an analytic approach in order to illuminate the origins of modelling noise in 21 cm power spectrum measurements and guide future array design and calibration strategies. For analytic tractability, we make a number of assumptions, which we attempt to describe clearly in the text, but do not necessarily hold for all

¹ In the GMRT’s case, a single, well-known, pulsar is used while the rest of the sky is eliminated by difference time-steps that correspond to the pulsar’s ‘on’ and ‘off’ states.

² HERA is designed to be fully redundantly calibratable but it is useful to assess the performance of sky-based calibration as an alternative with potentially different systematics. Since redundant calibration does not rely as much on a model of the sky, there exists the possibility of this array outperforming any of the predictions in this paper.

observing scenarios. Thus, our quantitative results should be understood as accurate only to within an order of magnitude, illustrating how modelling noise scales with the properties of the sky catalogue and instrumental parameters. Relaxing the assumptions in this paper for more accurate predictions is the subject of ongoing simulation work.

This paper is organized as follows. In Section 2, we introduce our analytic framework, based on W92 and L10, for describing the impact of calibration modelling errors on the 21 cm power spectrum and discuss its dependence on array and catalogue properties. In Section 3, we apply this formalism to predict the amplitude of calibration errors relative to 21 cm fluctuations in current and upcoming experiments given our current knowledge of foregrounds and primary beams. In Section 4, we explore a strategy for eliminating this noise through inverse baseline-length weighting. We conclude in Section 5.

2 FORMALISM

In this paper, we model gain errors as a statistical noise arising from the myriad of faint unmodelled sources. Such sources are not precisely modelled in calibration.

Since baselines and visibilities are formed from two antennas, we index them with a Greek index and we index antennas with lowercase Latin indices. We will also sometimes explicitly write a baseline index as a 2-tuple of antenna indices (e.g. $\alpha = (i, j)$). We describe the residual, $c_\alpha(v)$, between the true visibility formed from antennas i and j , $v_\alpha^{\text{true}}(v)$, and the model visibility, $y_\alpha(v)$, as a random variable with a mean $\langle c_\alpha(v) \rangle$ and covariance $\mathbf{C}_{\alpha\beta}(v, v') \equiv \langle [c_\alpha(v) - \langle c_\alpha(v) \rangle] [c_\beta(v') - \langle c_\beta(v') \rangle]^* \rangle$. We assume that c_α is composed of the sum of the 21 cm signal, s_α , unmodelled foregrounds, r_α , and a component arising from thermal noise, n_α , whose impact on calibration is explored in Trott & Wayth (2016). The true visibility is the sum between the modelled and unmodelled component,

$$v_\alpha^{\text{true}} = y_\alpha + c_\alpha = y_\alpha + r_\alpha + n_\alpha + s_\alpha. \quad (1)$$

Each unmodelled component is statistically independent, which means that

$$\text{Cov}[c, c^\dagger] \equiv \mathbf{C} = \mathbf{R} + \mathbf{N} + \mathbf{S}, \quad (2)$$

where \dagger denotes the conjugate transpose of a vector or matrix. In Section 2.1, we discuss expressions for the amplitude and frequency coherence of r_α in terms of a parametrized point source population and diffuse galactic emission. This ‘noise’ will be imprinted on the calibration solutions in a way that, for sufficiently small errors, is analytically tractable and can be described using the matrix formalism of W92 and L10, which we overview in Section 2.2. We derive expressions for the impact of these errors on the 21 cm power spectrum in Section 2.3 and the degree to which each visibility covariance contributes in Section 2.4. Using the expressions we derive, we discuss the scaling of modelling noise with the properties of the source catalogue and array in Section 3.4. Since both the 21 cm signal and thermal noise terms are already well considered in the literature (W92; L10; Trott & Wayth 2016), we will focus on the contribution from r_α .

2.1 The statistics of unmodelled source visibilities

Extensive work exists on statistical models of faint point sources in the power spectrum (e.g. Wang et al. 2006; Liu & Tegmark 2011; Trott et al. 2012, 2016; Dillon et al. 2013, 2015b), and we take an

approach similar to these papers and assume that the sources have uniform spectral structure (described by a single power law) that can be factored out of the visibilities and is far less significant than the frequency dependence introduced by the interferometric point spread function (PSF). We now give an overview of our characterization of the unmodelled point sources along with the diffuse emission from the Galaxy. Since residual Galactic emission is, for the most part, uncorrelated with residual point source emission, the covariance of unmodelled emission on each baseline is given by the sum of the covariance of each source,

$$\mathbf{R} = \mathbf{R}^{\text{P}} + \mathbf{R}^{\text{G}}, \quad (3)$$

where \mathbf{R}^{P} is the covariance due to unmodelled point sources and \mathbf{R}^{G} is the covariance of Galactic emission. We now describe our model of the covariances for these two emission sources.

2.1.1 Unmodelled point sources

With the MWA, point sources are completely sampled down to $S_{\text{min}} \approx 50\text{--}80$ mJy (Carroll et al. 2016; Hurley-Walker et al. 2017; Line et al., in preparation) and on LOFAR down to the $S_{\text{min}} \approx 0.1$ mJy level (Williams et al. 2016) within the primary beam. We represent these sources with an achromatic version of the model from Liu & Tegmark (2011). At these faint fluxes, the sources are isotropically distributed and the number of sources with fluxes between S and $S + dS$ within an infinitesimal solid angle $d\Omega$ is well described by a random Poisson process with a power-law mean (Di Matteo et al. 2002),

$$\frac{d^2 N}{dS d\Omega} = k \begin{cases} \left(\frac{S}{S_*}\right)^{-\gamma_1} & S \leq S_* \\ \left(\frac{S}{S_*}\right)^{-\gamma_2} & S > S_* \end{cases}, \quad (4)$$

where $k = 4000 \text{ Jy}^{-1} \text{ sr}^{-1}$, $\gamma_1 = 1.75$, $\gamma_2 = 2.5$ and $S_* = 0.88 \text{ Jy}$.

Consider a visibility, v_α , formed by antennas i and j , which are separated by baseline \mathbf{b}_α . The covariance between two baselines $v_\alpha(v)$ and $v_\beta(v')$ at two frequencies, v and v' , assuming unclustered and flat-spectrum sources is

$$\begin{aligned} \mathbf{R}_{\alpha\beta}^{\text{P}}(v, v') &= \int_0^{S_{\text{min}}} dS \frac{dN}{dS d\Omega} S^2 \int d\Omega |A(\hat{s})|^2 e^{-2\pi i \hat{s} \cdot (\mathbf{b}_\alpha v - \mathbf{b}_\beta v')/c} \\ &= \sigma_r^2(S_{\text{min}}) \int d\Omega |A(\hat{s})|^2 e^{-2\pi i \hat{s} \cdot (\mathbf{b}_\alpha v - \mathbf{b}_\beta v')/c}, \end{aligned} \quad (5)$$

where $A(\hat{s})$ is the primary beam of each antenna that we assume are identical. The Fourier convolution theorem tells us that the last integral in equation (5) is equal to the convolution of the Fourier transform of the beam with itself evaluated at $(\mathbf{b}_\alpha v/c - \mathbf{b}_\beta v'/c)$ in the uv plane. This quantity falls to zero when $|\mathbf{b}_\alpha v/c - \mathbf{b}_\beta v'/c|$ is larger than the diameter of the antenna aperture. Thus, as long as two baselines are separated by a distance greater than the antenna aperture diameter, $R_{\alpha\beta}^{\text{P}} \approx 0$. We may therefore ignore off-diagonal terms in the residual covariance matrix for minimally redundant arrays. It turns out that the diagonal covariance assumption gives similar results, even for maximally redundant arrays (see Appendix A).

2.1.2 Diffuse galactic emission

Diffuse Galactic emission is correlated on large angular scales. We may construct a simple model of this emission using the same

steps we used to obtain equation (5) and assuming that uv power spectrum of the diffuse emission does not evolve significantly over an antenna footprint. Under these assumptions, one can show that the covariance between two visibilities from diffuse emission is

$$\mathbf{R}_{\alpha\beta}^G(v, v') \approx P_G(\mathbf{b}_\alpha/\lambda_0) \int d\Omega |A(\hat{s})|^2 e^{-2\pi i \hat{s} \cdot (\mathbf{b}_\alpha v - \mathbf{b}_\beta v')/c}, \quad (6)$$

where $P_G(\mathbf{u})$ is the power spectrum of diffuse galactic emission in the uv plane and λ_0 is the wavelength of the centre of the interferometer's band. To model $P_G(\mathbf{u})$, we use an empirical power-law fit to the two-dimensional power spectrum of a desourced and destriped (Remazeilles et al. 2015) Galactic emission map (Haslam et al. 1982) centred at RA = 60°, Dec. = −30° and scaled from 408 to 150 MHz using a frequency power law with a spectral index of −0.6 (Rogers & Bowman 2008; Fixsen et al. 2011).³ We find that the angular power spectrum of galactic emission at 150 MHz is well modelled by a power law in $u = |\mathbf{b}|/\lambda_0$, $P_G(\mathbf{u}) = 6 \times 10^{11} u^{-5.7} \text{ Jy}^2 \text{ Sr}^{-1}$. Throughout this paper, we will assume that the model used for calibration completely ignores diffuse emission so that all of $\mathbf{R}_{\alpha\beta}^G$ is included in the covariance of residual visibilities.

2.2 Frequency domain calibration errors

So far, we have a model of the discrepancies between true and modelled interferometer visibilities. Given this model, what are the statistics of the errors in our frequency-dependent gain solutions? Our goal in this subsection is to derive the covariances of errors in gain parameters in terms of the covariances of the visibility residuals discussed in Section 2.1.

We will start by writing down the system of equations that calibration algorithms attempt to solve, and, following W92 and L10, we will reduce this system to a set of linear equations that are valid in the regime of small calibration errors, which is the case for errors generated by the faintest sources on the sky. This approximation holds when the gains are nearly correct after large gain variations are removed by a first iteration of calibration using a reasonably accurate calibration catalogue. Writing down these systems in matrix form, the covariances of the least-squares solutions for these linear systems are readily obtained in the same manner as W92 and L10.

We start by writing down the equations that calibration must solve. In line with the notation of L10, we parametrize the small gain and phase of the i th antenna after rough calibration as the exponent of a complex number,

$$g_i(v) = e^{\eta_i(v) + i\phi_i(v)} \approx 1 + \eta_i(v) + i\phi_i(v), \quad (7)$$

where η_i is the amplitude of the gain and ϕ_i is the phase. In calibration, one attempts to solve the set of equations

$$g_i(v) g_j^*(v) v_{ij}^{\text{true}}(v) = v_{ij}^{\text{meas}}(v), \quad (8)$$

where $v_{ij}^{\text{meas}}(v)$ is the measured visibility. If we divide by y_{ij} on both sides, we have

$$g_i g_j^* \left(1 + \frac{c_{ij}}{y_{ij}} \right) = \frac{v_{ij}^{\text{meas}}}{y_{ij}}. \quad (9)$$

Recall that c_{ij} represents the sum of unmodelled components of a visibility (equation 1) while y_{ij} represents the modelled component. For analytic tractability, we will linearize these equations by

working to first order in c_{ij}/y_{ij} , η_i and ϕ_i . With this approximation, equation (9) becomes

$$\frac{v_{ij}^{\text{meas}}}{y_{ij}} \approx (1 + \eta_i + i\phi_i)(1 + \eta_j - i\phi_j) \left(1 + \frac{c_{ij}}{y_{ij}} \right) \quad (10)$$

$$\approx 1 + \eta_i + \eta_j + i\phi_i - i\phi_j + \frac{c_{ij}}{y_{ij}}. \quad (11)$$

Separating the real and imaginary parts gives us two systems of linear equations:

$$\eta_i + \eta_j + \text{Re} \left(\frac{c_{ij}}{y_{ij}} \right) \approx \text{Re} \frac{v_{ij}^{\text{meas}}}{y_{ij}} - 1 \equiv M_{ij}^\eta \quad (12)$$

and

$$\phi_i - \phi_j + \text{Im} \left(\frac{c_{ij}}{y_{ij}} \right) \approx \text{Im} \frac{v_{ij}^{\text{meas}}}{y_{ij}} \equiv M_{ij}^\phi. \quad (13)$$

Since residual foregrounds may be described statistically, we treat c_{ij} as a noise term in the same way that thermal noise is treated in L10. Unlike thermal noise, which is typically uncorrelated in frequency and ideally has the same variance across baselines, the correlation properties of modelling noise are those of the unmodelled sources discussed in Section 2.1. We can write the system of equations given by equation (12) in matrix form,⁴

$$\begin{pmatrix} M_{12}^\eta \\ M_{23}^\eta \\ \vdots \\ M_{N-1N}^\eta \\ M_{13}^\eta \\ \vdots \\ M_{N-2N}^\eta \\ \vdots \\ M_{1N}^\eta \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 0 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \vdots \\ \eta_N \end{pmatrix} + \text{Re} \left(\frac{\mathbf{c}}{\mathbf{y}} \right), \quad (14)$$

which we write more compactly as $\mathbf{M}^\eta = \mathbf{A}\boldsymbol{\eta} + \text{Re}(\mathbf{c}/\mathbf{y})$. The same can be done for equation (13),

$$\begin{pmatrix} M_{12}^\phi \\ M_{23}^\phi \\ \vdots \\ M_{N-1N}^\phi \\ M_{13}^\phi \\ \vdots \\ M_{N-2N}^\phi \\ \vdots \\ M_{1N}^\phi \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & -1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -1 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & -1 \\ 1 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{pmatrix} + \text{Im} \left(\frac{\mathbf{c}}{\mathbf{y}} \right), \quad (15)$$

⁴ The system of equations used here only attempts to solve for the gains. In redundant calibration, the number of unique true visibilities is reduced to a point where one can also solve for them as well. This leads to different forms for the matrix equations (see W92 and L10 for examples).

³ This power law is for spectral radiance. For brightness temperature, the spectral index is −2.6.

where the last row in the matrix arises from imposing the constraint (L10) that $\sum_j \phi_j = 0$.⁵ We write the imaginary equation as $\mathbf{M}^\phi = \mathbf{B}\boldsymbol{\phi} + \text{Im}(\mathbf{c}/\mathbf{y})$.

Given a model and measurements of \mathbf{M}^ϕ and \mathbf{M}^η , a least-squares estimator that applies weights of \mathbf{W} to each measurement will arrive at solutions for $\boldsymbol{\eta}$ and $\boldsymbol{\phi}$ given by

$$\widehat{\boldsymbol{\eta}} = (\mathbf{A}^\top \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{W} \mathbf{M}^\eta \equiv \boldsymbol{\Lambda} \mathbf{M}^\eta \quad (16)$$

and

$$\widehat{\boldsymbol{\phi}} = (\mathbf{B}^\top \mathbf{W} \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{W} \mathbf{M}^\phi \equiv \boldsymbol{\Psi} \mathbf{M}^\phi. \quad (17)$$

We emphasize that $\widehat{\boldsymbol{\eta}}$ and $\widehat{\boldsymbol{\phi}}$ are estimates of the true values, $\boldsymbol{\eta}$ and $\boldsymbol{\phi}$. The covariance of these estimates,

$$\mathbf{C}_{\boldsymbol{\eta}}(v, v') = \langle \widehat{\boldsymbol{\eta}}(v) \widehat{\boldsymbol{\eta}}^\top(v') \rangle - \langle \widehat{\boldsymbol{\eta}}(v) \rangle \langle \widehat{\boldsymbol{\eta}}^\top(v') \rangle \quad (18)$$

$$\mathbf{C}_{\boldsymbol{\phi}}(v, v') = \langle \widehat{\boldsymbol{\phi}}(v) \widehat{\boldsymbol{\phi}}^\top(v') \rangle - \langle \widehat{\boldsymbol{\phi}}(v) \rangle \langle \widehat{\boldsymbol{\phi}}^\top(v') \rangle, \quad (19)$$

is given by

$$\mathbf{C}_{\boldsymbol{\eta}}(v, v') = \boldsymbol{\Lambda} \text{Cov} \left[\text{Re} \left(\frac{\mathbf{c}}{\mathbf{y}} \right), \text{Re} \left(\frac{\mathbf{c}}{\mathbf{y}} \right)^\top \right] \boldsymbol{\Lambda}^\top \quad (20)$$

$$\mathbf{C}_{\boldsymbol{\phi}}(v, v') = \boldsymbol{\Psi} \text{Cov} \left[\text{Im} \left(\frac{\mathbf{c}}{\mathbf{y}} \right), \text{Im} \left(\frac{\mathbf{c}}{\mathbf{y}} \right)^\top \right] \boldsymbol{\Psi}^\top. \quad (21)$$

Thus, we have arrived at expressions for the covariances of errors in the gain parameters in terms of the covariances of the real and imaginary components of the unmodelled visibilities. Equations (20) and (21) show that the covariance of any given gain solution is the linear combination of the covariances of every visibility in the array. Thus, the application of a gain solution (derived from an incomplete sky model) to a short baseline introduces the fine-frequency errors from long baselines. Our next step is to determine the impact of this leakage on the power spectrum.

2.3 The impact of gain errors on the 21 cm power spectrum

We now propagate the frequency-dependent errors in each gain solution into the delay power spectrum. Calibration gives us an estimate of the gains,

$$\widehat{g}_i = e^{\widehat{\eta}_i + i\widehat{\phi}_i} \approx 1 + \widehat{\eta}_i + i\widehat{\phi}_i. \quad (22)$$

whose deviations from the true gains (η_i and ϕ_i) have covariances given by equations (20) and (21). The corrected, model-subtracted visibilities obtained from calibration are given by

$$\begin{aligned} V_{ij} &= \frac{g_i g_j^*}{\widehat{g}_i \widehat{g}_j^*} (y_{ij} + c_{ij}) - y_{ij} \\ &\approx (y_{ij} + c_{ij}) \\ &\times [1 + (\eta_i - \widehat{\eta}_i) + (\eta_j - \widehat{\eta}_j) + i(\phi_i - \widehat{\phi}_i) - i(\phi_j - \widehat{\phi}_j)] - y_{ij}. \end{aligned} \quad (23)$$

The delay transform (Parsons et al. 2012b) is a popular estimate of the power spectrum in which visibilities are Fourier transformed

⁵ The arbitrary phase reference is often set in sky-based calibration by defining the phases as the differences between each antenna phase and that of an arbitrarily chosen reference antenna. This constraint can be written as $\phi_{\text{ref}} = 0$ and would modify the last row of \mathbf{B} to be zero except for the index of the reference antenna (rather than all ones as we have written it). While choosing the reference antenna form of the phase constraint affects the details of some of the expressions in this paper, it results in the same scaling relationships and has a negligible effect on quantitative results.

from frequency into delay. Delay can be mapped approximately to Fourier modes along the LoS while the uv coordinates of the visibility can be mapped to Fourier modes perpendicular to the LoS. The delay transform is given by

$$\widetilde{V}_{ij}(\tau) = \int d\mathbf{v} e^{2\pi i \mathbf{v} \cdot \boldsymbol{\tau}} V_{ij}(v), \quad (24)$$

which we can apply to the gain-corrected and foreground-subtracted visibility in equation (23). Taking the delay transform of equation (23) and setting $\boldsymbol{\eta}' \equiv \widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}$ and $\boldsymbol{\phi}' \equiv \widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}$, we have

$$\begin{aligned} \widetilde{V}_{ij}(\tau) &\approx -y_{ij} \star (\widetilde{\eta}'_i + \widetilde{\eta}'_j + i\widetilde{\phi}'_i - i\widetilde{\phi}'_j) \\ &\quad + \widetilde{c}_{ij} - \widetilde{c}_{ij} \star (\widetilde{\eta}'_i + \widetilde{\eta}'_j + i\widetilde{\phi}'_i - i\widetilde{\phi}'_j), \end{aligned} \quad (25)$$

where \star denotes a convolution in delay space. For the sake of analytic tractability, we will ignore the chromaticity of y_α and set all $y_\alpha = S_0$, essentially assuming that the modelled visibilities are dominated by a single source near the phase centre that exceeds the flux of all other sources by a factor of several. Even with chromatic y_α , per-frequency inverse-covariance weighting, which multiplies each α th weight by $|y_\alpha|^2$ (L10), removes some of this structure. In Appendix B, we explore the impact of relaxing this assumption and find that our achromatic y_{ij} model predicts the LoS wavenumbers to within ≈ 10 per cent of what we find with chromatic y_{ij} s obtained from a realistic sky model. Still, this dramatic assumption limits the accuracy of our specific quantitative predictions, and we are exploring its impact in full calibration simulations.

The cosmological 21 cm power spectrum, $P(\mathbf{k})$, is well approximated by the mean amplitude square of the delay-transformed visibility multiplied by linear factors given in Parsons et al. (2012a),

$$P(\mathbf{k}) \approx \left(\frac{c^2}{2k_B^2 v_0^2} \right)^2 \frac{X^2(v_0) Y(v_0)}{B_{\text{pp}} \Omega_{\text{pp}}} (|\widetilde{V}(\mathbf{u}, \boldsymbol{\eta})|^2), \quad (26)$$

where $\Omega_{\text{pp}} = \int d\Omega |A(\hat{\mathbf{s}})|^2$ and $B_{\text{pp}} = \int d\mathbf{f} |B(v)|^2$ are respectively the integrals of the squares of the beam and bandpass, v_0 is the centre frequency of the observation, k_B is the Boltzmann constant, and (X, Y) are multiplicative factors converting between native interferometry coordinates and comoving cosmological coordinates, $2\pi(u, v, \eta) = (Xk_x, Xk_y, Yk_z)$. In line with Parsons et al. (2012a), $\boldsymbol{\eta}$ is used to denote the Fourier dual to frequency at fixed $|\mathbf{u}|$ and $\boldsymbol{\tau}$ to denote the frequency Fourier transform of a visibility that integrates over a slanted line in uv space. While this slanted integral introduces non-negligible mode mixing (namely the wedge), it is a decent approximation for the range of η s probed by current- and next-generation experiments.

Therefore, we can estimate the power spectrum from calibration modelling errors by cross-multiplying \widetilde{V}_α with its complex conjugate. If we denote the expectation value

$$\langle \widetilde{V}_\alpha(\boldsymbol{\tau}) \widetilde{V}_\beta^*(\boldsymbol{\tau}') \rangle \equiv \mathbf{P}_{\alpha\beta}(\boldsymbol{\tau}, \boldsymbol{\tau}'), \quad (27)$$

the bias from visibility residuals is equal to $\mathbf{P}_{\alpha\beta}(\boldsymbol{\tau}, \boldsymbol{\tau}')$ multiplied by the constant prefactors in equation (26). While we only need $\mathbf{P}_{\alpha\beta}(\boldsymbol{\tau}, \boldsymbol{\tau}')$ for the bias, we will need off-diagonal terms to calculate the variances of binned and averaged power spectrum estimates. We first write down $\mathbf{P}_{\alpha\beta}(\boldsymbol{\tau}, \boldsymbol{\tau}')$ to second order in \mathbf{c}/\mathbf{y} with baseline α formed from antennas i and j and baseline β formed from antennas m and n ,

$$\begin{aligned} \langle \widetilde{V}_\alpha(\boldsymbol{\tau}) \widetilde{V}_\beta^*(\boldsymbol{\tau}') \rangle &\equiv \mathbf{P}_{\alpha\beta} \\ &\approx S_0^2 [\langle \widetilde{\eta}'_i \widetilde{\eta}'_m^* \rangle + \langle \widetilde{\eta}'_j \widetilde{\eta}'_n^* \rangle + \langle \widetilde{\eta}'_i \widetilde{\eta}'_m^* \rangle + \langle \widetilde{\eta}'_j \widetilde{\eta}'_n^* \rangle \\ &\quad + \langle \widetilde{\phi}'_i \widetilde{\phi}'_m^* \rangle - \langle \widetilde{\phi}'_i \widetilde{\phi}'_n^* \rangle - \langle \widetilde{\phi}'_j \widetilde{\phi}'_m^* \rangle + \langle \widetilde{\phi}'_j \widetilde{\phi}'_n^* \rangle \end{aligned}$$

$$\begin{aligned}
 & -i(\tilde{\eta}'_i \tilde{\phi}'_m) + i(\tilde{\eta}'_i \tilde{\phi}'_n) - i(\tilde{\eta}'_j \tilde{\phi}'_m) + i(\tilde{\eta}'_j \tilde{\phi}'_n) \\
 & -i(\tilde{\phi}'_i \tilde{\eta}'_m) + i(\tilde{\phi}'_i \tilde{\eta}'_n) - i(\tilde{\phi}'_j \tilde{\eta}'_m) - i(\tilde{\phi}'_j \tilde{\eta}'_n) \\
 & -S_0[(\tilde{\eta}'_i \tilde{c}'_\beta) - (\tilde{\eta}'_j \tilde{c}'_\beta) - i(\tilde{\phi}'_i \tilde{c}'_\beta) + i(\tilde{\phi}'_j \tilde{c}'_\beta) \\
 & - \langle \tilde{c}'_\alpha \tilde{\eta}'_m \rangle - \langle \tilde{c}'_\alpha \tilde{\eta}'_n \rangle + i(\tilde{c}'_\alpha \tilde{\phi}'_m) - i(\tilde{c}'_\alpha \tilde{\phi}'_n)] \\
 & + \langle \tilde{c}'_\alpha \tilde{c}'_\beta \rangle.
 \end{aligned} \tag{28}$$

For the sake of space, we do not explicitly write τ or τ' in every term but understand that every complex conjugated term in each product is a function of τ' and every non-conjugated term is a function of τ .

Equation (28) involves six types of terms: those involving cross-multiples of $\tilde{\eta}'$, cross-multiples of $\tilde{\phi}'$, cross-multiples of $\tilde{\eta}'$ and $\tilde{\phi}'$, products between $\tilde{\eta}$ and $\tilde{\phi}$ with \tilde{c} , and finally the covariances of the residuals themselves. In Appendix C, we obtain approximate expressions for each of the first five terms when the baseline separation is longer than the antenna diameter,

$$\begin{aligned}
 \langle \tilde{\eta}'_i(\tau) \tilde{\eta}'_m(\tau') \rangle & \approx \int dv dv' e^{2\pi i(v\tau - v'\tau')} [\mathbf{C}_\eta(v, v')]_{im} \\
 & \equiv \frac{S_0^{-2}}{2} [\mathbf{\Lambda} \tilde{\mathbf{C}}(\tau, \tau') \mathbf{\Lambda}^\top]_{im} \\
 & = \frac{S_0^{-2}}{2} \Lambda_{i\gamma} \Lambda_{\delta m}^\top \tilde{\mathbf{C}}^{\gamma\delta}(\tau, \tau')
 \end{aligned} \tag{29}$$

$$\begin{aligned}
 \langle \tilde{\phi}'_i(\tau) \tilde{\phi}'_m(\tau') \rangle & \approx \int dv dv' e^{2\pi i(v\tau - v'\tau')} [\mathbf{C}_\phi(v, v')]_{im} \\
 & = \frac{S_0^{-2}}{2} [\mathbf{\Psi} \tilde{\mathbf{C}}(\tau, \tau') \mathbf{\Psi}^\top]_{im} \\
 & = \frac{S_0^{-2}}{2} \Psi_{i\gamma} \Psi_{\delta m}^\top \tilde{\mathbf{C}}^{\gamma\delta}(\tau, \tau')
 \end{aligned} \tag{30}$$

$$\begin{aligned}
 \langle \tilde{c}'_\alpha(\tau) \tilde{\eta}'_m(\tau') \rangle & \approx \frac{1}{2} \Lambda_{m\gamma} \int dv dv' e^{2\pi i(v\tau - v'\tau')} [\mathbf{C}(v, v')]_{\alpha\gamma} \\
 & = \frac{1}{2} \Lambda_{i\gamma} \tilde{\mathbf{C}}^{\gamma\alpha}(\tau, \tau')
 \end{aligned} \tag{31}$$

$$\begin{aligned}
 \langle \tilde{c}'_\alpha(\tau) \tilde{\phi}'_m(\tau') \rangle & \approx \frac{i}{2} \Psi_{m\gamma} \int dv dv' e^{2\pi i(v\tau - v'\tau')} [\mathbf{C}(v, v')]_{\alpha\gamma} \\
 & = \frac{i}{2} \Psi_{m\gamma} \tilde{\mathbf{C}}^{\gamma\alpha}(\tau, \tau')
 \end{aligned} \tag{32}$$

$$\langle \tilde{\eta}'_i(\tau) \tilde{\phi}'_m(\tau') \rangle \approx 0, \tag{33}$$

where we used Einstein notation with repeated raised and lowered indices to denote summation and have defined $\tilde{\mathbf{C}}(\tau, \tau')$ as the delay transform of the \mathbf{C} matrix. $\tilde{\mathbf{C}}_{\alpha\beta}(\tau, \tau') \equiv \int dv dv' e^{-2\pi i(\tau v - \tau' v')} \mathbf{C}_{\alpha\beta}(v, v')$. We also denote the delay transform of the \mathbf{N} , \mathbf{R} and \mathbf{S} matrices in a similar way as $\tilde{\mathbf{N}}$, $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{S}}$. The final term in equation (28) is simply the covariance matrix of the delay-transformed residual visibilities, $\tilde{\mathbf{C}}_{\alpha\beta}(\tau, \tau')$. Using the above identities, we may write equation (28) as

$$\begin{aligned}
 \mathbf{P}_{\alpha\beta}(\tau, \tau') & = \frac{1}{2} [\Lambda_{i\gamma} \Lambda_{\delta m}^\top + \Lambda_{i\gamma} \Lambda_{\delta n}^\top + \Lambda_{j\gamma} \Lambda_{\delta m}^\top + \Lambda_{j\gamma} \Lambda_{\delta n}^\top \\
 & + \Psi_{i\gamma} \Psi_{\delta m}^\top - \Psi_{i\gamma} \Psi_{\delta n}^\top - \Psi_{j\gamma} \Psi_{\delta m}^\top + \Psi_{j\gamma} \Psi_{\delta n}^\top] \tilde{\mathbf{C}}^{\gamma\delta}(\tau, \tau') \\
 & - \frac{1}{2} (\Lambda_{i\gamma} + \Lambda_{j\gamma} - \Psi_{i\gamma} + \Psi_{j\gamma}) \tilde{\mathbf{C}}^{\gamma\beta}(\tau, \tau') \\
 & - \frac{1}{2} (\Lambda_{m\gamma} + \Lambda_{n\gamma} + \Psi_{m\gamma} - \Psi_{n\gamma}) \tilde{\mathbf{C}}^{\gamma\alpha}(\tau, \tau') \\
 & + \tilde{\mathbf{C}}_{\alpha\beta}(\tau, \tau').
 \end{aligned} \tag{34}$$

The power spectrum bias in delay-transformed estimates is given by $\mathbf{P}_{\alpha\alpha}(\tau, \tau)$ ($i = m$ and $j = n$),

$$\begin{aligned}
 \mathbf{P}_{\alpha\alpha}(\tau, \tau) & = \frac{1}{2} (\Lambda_{i\gamma} \Lambda_{\delta i}^\top + 2\Lambda_{i\gamma} \Lambda_{\delta j}^\top + \Lambda_{j\gamma} \Lambda_{\delta j}^\top) \tilde{\mathbf{C}}^{\gamma\delta}(\tau, \tau) \\
 & + \frac{1}{2} (\Psi_{i\gamma} \Psi_{\delta i}^\top - 2\Psi_{i\gamma} \Psi_{\delta j}^\top + \Psi_{j\gamma} \Psi_{\delta j}^\top) \tilde{\mathbf{C}}^{\gamma\delta}(\tau, \tau) \\
 & - (\Lambda_{i\gamma} + \Lambda_{j\gamma}) \tilde{\mathbf{C}}^{\gamma\alpha}(\tau, \tau) + \tilde{\mathbf{C}}_{\alpha\alpha}(\tau, \tau).
 \end{aligned} \tag{35}$$

Equations (35) and (34) show how calibration leaks unmodelled structure in every visibility, including the highly chromatic ones, into the power spectrum of otherwise smooth short baselines. The last term in equation (35) is the power spectrum of unmodelled foregrounds, noise and the signal itself. Recall that since foregrounds are naturally contained within the horizon delay of b_α , it does not contribute power into the EoR window. The sums in the first two lines, on the other hand, mix the chromaticity of foregrounds on all baselines into the delay power spectrum of the α th visibility. Baselines that are longer than b_α contribute emission at delays below their individual horizon delays that are greater than the horizon delay of b_α , allowing for contamination of the EoR window.

Typically, an estimate of $\mathbf{P}_{\alpha\alpha}$ is obtained by cross-multiplying integration over independent time intervals and since noise within each interval is independent, $\mathbf{N}_{\alpha\beta}(v, t; v', t') = 0$ (Dillon et al. 2014) and we can ignore the thermal noise contribution to the bias given by equation (34). However, a subtlety introduced by calibration errors is that if calibration solutions for the cross-multiplied visibilities are not derived from complementary time intervals, there will still exist a thermal noise bias arising from all but the last term in equation (34). This is the case in Dillon et al. (2015b), Beardsley et al. (2016) and Ewall-Wice et al. (2016b) where 0.5 s time-steps are used for interleaving visibilities but 112 s non-interleaved time-steps are used for calibration. This bias can also survive cross-multiplying different redundant measurements of the same visibility as is done with PAPER (Parsons et al. 2014; Ali et al. 2015).

2.4 Calibration bias for a simplified model

How do the covariances between pairs of visibilities contribute to the final power spectrum? We showed in Section 2.1 that \mathbf{R} , the covariance matrix of unmodelled foreground visibilities, is well approximated as a diagonal in minimally redundant arrays. The same is true for thermal noise that arises from independent fluctuations at each antenna and the 21 cm signal. Thus, the first two lines of equation (35) are formed from the weighted sum of the variance of the $N_{\text{ant}}(N_{\text{ant}} - 1)/2$ visibilities (where N_{ant} is the number of antenna elements) with $\gamma = \delta$ where the weight of each variance is given by $\Lambda_{i\gamma} \Lambda_{\gamma j}^\top$ and $\Psi_{i\gamma} \Psi_{\gamma j}^\top$. These values depend crucially on our choice of visibility weighting, \mathbf{W} , but it is highly instructive to examine the case where \mathbf{W} is equal to the identity. In Appendix D, we use matrix algebra to show that in the case of \mathbf{W} equal to the identity,

$$\Lambda_{i\gamma} = \begin{cases} \frac{1}{N_{\text{ant}} - 1} & i \in \gamma \\ -1 & i \notin \gamma \end{cases} \tag{36}$$

and

$$\Psi_{i\gamma} = \begin{cases} \frac{1}{N_{\text{ant}}} & \gamma = (i, \cdot) \\ -\frac{1}{N_{\text{ant}}} & \gamma = (\cdot, i) \\ 0 & i \notin \gamma \end{cases} \tag{37}$$

where $\gamma = (i, \cdot)$ denotes any visibility with i as the non-conjugated antenna and $\gamma = (\cdot, i)$ any baseline with i as the conjugated antenna. Equation (36) makes intuitive sense if we recall that $\Lambda_{i\gamma}$ is the linear weight multiplied by each M_γ^η that is summed to form the i th gain solution. Inspecting equation (12), we see that summing all $N_{\text{ant}} - 1$ $\Lambda_{i\gamma} M_\gamma^\eta$ that antenna i participates in gives us

$$\sum_{\gamma \ni i} \Lambda_{i\gamma} M_\gamma^\eta = \eta_i + \frac{1}{N_{\text{ant}} - 1} \sum_{k \neq i} \eta_k. \quad (38)$$

To remove the extra sum, and isolate η_i , we must subtract the sum all M^η that do not include the i th antenna, divided by $N_{\text{ant}} - 1$. For each $k \neq i$, there are $N_{\text{ant}} - 2$ baselines that involve k but not i , so we must also divide each term by $N_{\text{ant}} - 2$. This gives us the weights for baselines not involving the i th antenna in equation (36). We can apply similar logic to equation (37) by inspecting equations (13) and (37).

The weight with which the covariance between each pair of measurements contributes to the total covariance of $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\phi}}$ is just the product of the weight with which each measurement is linearly summed,

$$\Lambda_{i\gamma} \Lambda_{\delta j}^\top = \begin{cases} \frac{1}{(N_{\text{ant}} - 1)^2} & i \in \gamma \text{ and } j \in \delta \\ \frac{-1}{(N_{\text{ant}} - 2)(N_{\text{ant}} - 1)^2} & i \in \gamma \text{ and } j \notin \delta \\ \frac{-1}{(N_{\text{ant}} - 2)(N_{\text{ant}} - 1)^2} & i \notin \gamma \text{ and } j \in \delta \\ \frac{1}{(N_{\text{ant}} - 1)^2(N_{\text{ant}} - 2)^2} & i \notin \gamma \text{ and } j \notin \delta \end{cases} \quad (39)$$

and

$$\Psi_{i\gamma} \Psi_{\delta j}^\top = \begin{cases} \frac{1}{N_{\text{ant}}^2} & \gamma = (i, \cdot), \delta = (j, \cdot) \text{ or } \gamma = (\cdot, i), \delta = (\cdot, j) \\ -\frac{1}{N_{\text{ant}}^2} & \gamma = (i, \cdot), \delta = (\cdot, j) \text{ or } \gamma = (\cdot, i), \delta = (j, \cdot) \\ 0 & \text{otherwise.} \end{cases} \quad (40)$$

For non-redundant arrays, $\mathbf{C}(v, v')$ is diagonal and we can focus on $\gamma = \delta$ terms. From equations (39) and (40), we see that when $i = j$, each visibility variance is weighted by $\sim N_{\text{ant}}^{-2}$ when $i \in \gamma$ and at most by N_{ant}^{-4} when $i \notin \gamma$. Since there are $\sim N_{\text{ant}}$ visibilities with antenna i and $\sim N_{\text{ant}}^2$ visibilities without antenna i , $i = j$ terms contributing to $\mathbf{P}_{\alpha\alpha}$ are given roughly by the average of visibility variances not involving i divided by $\sim N_{\text{ant}}^{-2}$ plus the average of visibility variances involving antenna i divided by $\sim N_{\text{ant}}$,

$$\begin{aligned} \Lambda_{i\gamma} \Lambda_{\delta i}^\top \tilde{\mathbf{C}}^{\gamma\delta}(\tau, \tau) &\approx \frac{1}{N_{\text{ant}}^2} \langle \tilde{\mathbf{C}}_{\delta\delta}(\tau, \tau) \rangle_{i \notin \delta} + \frac{1}{N_{\text{ant}}} \langle \tilde{\mathbf{C}}_{\delta\delta}(\tau, \tau) \rangle_{i \in \delta} \\ &\approx \frac{1}{N_{\text{ant}}} \langle \tilde{\mathbf{C}}_{\delta\delta}(\tau, \tau) \rangle_{i \in \delta}, \end{aligned} \quad (41)$$

where the $\langle \rangle_{i \in \delta}$ indicates an average over the set of baselines that include antenna i and $\langle \rangle_{i \notin \delta}$ denotes an average over baselines that are not formed using antenna i . The same equation holds for the Ψ sums. Considering how elements of $\mathbf{C}(v, v')$ scale with N_{ant} , we can see the average of visibilities involving antenna i dominate $\mathbf{P}_{\alpha\alpha}$ by a factor of N_{ant} .

For $i \neq j$, there is exactly one visibility that involves both antennas and will be weighted at most by N_{ant}^{-2} . The $\sim N_{\text{ant}}$ visibilities formed from i XOR j are weighted by N_{ant}^{-3} and the $\sim N_{\text{ant}}^2$ visibilities that involve neither i nor j are weighted by N_{ant}^{-4} . Thus, all terms with $i \neq j$ in equation (35) give contributions of the order of the average of the visibility variances divided by N_{ant}^2 .

It follows that if $\mathbf{C}(v, v')$ is diagonal for all v, v' and \mathbf{W} is equal to the identity, $i = j$ sums in equation (35) dominate by $\sim N_{\text{ant}}$ and are well approximated by the average visibility variance involving antenna i or j divided by N_{ant} . The overall level of foreground contamination from calibration errors therefore goes as N_{ant}^{-1} with the details of the extent in delay contamination depending on the antenna distribution and primary beam. Replacing each $i = j$ sum in equation (35) with an average over visibility covariances involving i and j and ignoring $i \neq j$ sums, we arrive at an approximate formula that can be readily used to estimate $\mathbf{P}_{\alpha\alpha}$,

$$\begin{aligned} \mathbf{P}_{\alpha\alpha}(\tau, \tau) &\approx \frac{1}{N_{\text{ant}}} \left[\langle \tilde{\mathbf{C}}_{\delta\delta}(\tau, \tau) \rangle_{i \in \delta} + \langle \tilde{\mathbf{C}}_{\delta\delta}(\tau, \tau) \rangle_{j \in \delta} \right] \\ &\quad + \tilde{\mathbf{C}}_{\alpha\alpha}(\tau, \tau). \end{aligned} \quad (42)$$

The two assumptions going into this formula are that for each v and v' , $\mathbf{C}(v, v')$ is diagonal (minimal redundancy) and that \mathbf{W} is set equal to unity. Equation (42) illustrates how the bias of a power spectrum estimate depends on both the covariance of the individual baseline (the second term) and the covariances of the baselines that share common antennas. In other words, the measurement of the power spectrum for a given baseline and delay depends on both the residual foregrounds, noise and signal at that baseline and delay and, suppressed by a factor of N_{ant} , that of all other baselines at that delay that share an antenna with it.

3 MODELLING NOISE IN EXISTING ARRAYS

Having developed our formalism in Section 2, we may obtain order-of-magnitude estimates for the visibility modelling noise using equation (34) for four existing or planned arrays: LOFAR, MWA, HERA and the re-baselined SKA-LOW. We discuss our models of each instrument (Section 3.1). We then determine the level of modelling noise in Section 3.2 and the impact of beam modelling errors in Section 3.3. Equation (42) can be used to provide us with some intuition for how the properties of the noise scale with those of the array and catalogue. In Section 3.4, we discuss these scalings and to what extent they may be used to reduce the amplitude of modelling noise. In each simulation, we assume that the foreground model, used for calibration and subtraction, contains point sources modelled down to some minimal flux level S_{min} and that the true sky contains both the diffuse emission and all point sources.

3.1 Instrumental models

For all arrays, we consider an Airy beam for an aperture with diameter d_{ant} ,

$$A(\hat{s}) = A(\theta) = \left(2 \frac{J_1(\pi d_{\text{ant}} \cos \theta / \lambda_0)}{\pi d_{\text{ant}} \cos \theta \lambda_0} \right)^2, \quad (43)$$

where θ is the arc length from the beam pointing centre. An Airy beam has the virtue of a simple analytic expression that, unlike a Gaussian beam, exhibits realistic side-lobe structure, which in turn affects foreground contamination near the edge of the wedge (Thyagarajan et al. 2015; Pober et al. 2016).

Strictly speaking, the primary beam evolves with frequency; however, we find in numerical calculations that allowing for this variation has a negligible impact on our results. In order to expedite the computation of calibration noise, especially for the arrays with large numbers of antennas, such as HERA and the SKA-1, we also assume that \mathbf{R} is diagonal. This is clearly not the case for the highly redundant HERA layout but we find (Appendix A) that this only

impacts the amplitude of the modelling noise by a factor of order unity and has a negligible impact on which modes are contaminated. In all arrays, we assume that every baseline is given equal weighting of unity. Note that for the SKA and LOFAR, we do not explicitly include outrigger antennas in our model of calibration. Our models for each individual instrument are as follows.

(i) *The Murchison Widefield Array*. For the MWA, we use the 128-tile layout described in Beardsley et al. (2012) and Tingay et al. (2013). Antennas are modelled as 4 m diameter circular apertures. We assume a flux limit of 86 mJy, which is the limit for the array’s naturally weighted PSF at 150 MHz and similar to limits obtained in Carroll et al. (2016). Other analyses have obtained complete samples down to 35–50 mJy (Hurley-Walker et al. 2017) but this order unity change in S_{\min} does not significantly impact the modelling noise level that scales as $\sim S_{\min}^{1.25}$ (see Section 3.4). The deeper TIFR GMRT Sky Survey (TGSS) covers a significant portion of the MWA’s field of view and is complete down to 10 mJy (Intema et al. 2017). We therefore also consider an optimistic scenario where a deep TGSS catalogue is used to calibrate the instrument.

(ii) *The Low-Frequency Array*. We model LOFAR as the 48-element high band core described in van Haarlem et al. (2013), with 30 m diameter circular stations. The confusion limit for the naturally weighted core is ≈ 35 mJy at 150 MHz. However, the use of LOFAR’s extended baselines measures source catalogues that are complete down to $S_{\min} \approx 0.1$ mJy (Williams et al. 2016). While the Williams et al. (2016) survey is over a $\approx 4^\circ$ field of view, the catalogue we consider here covers the entire sky. Such a catalogue would involve accurately characterizing ~ 27 million sources and may not happen before the SKA but we consider it as a very optimistic bracket on LOFAR’s performance.

(iii) *The Hydrogen Epoch of Reionization Array*. For our model of HERA, we use the 331-element hexagonally packed core described in Pober et al. (2014). Each element is modelled as a 14 m diameter circular aperture. HERA is designed to be calibrated redundantly (Dillon & Parsons 2016); hence, the power spectrum estimates it obtains will not directly be affected by the modelling errors we explore in this paper. We choose to include HERA in order to assess the performance of compact cores in sky-based calibration and to explore sky-based calibration as an alternative to redundant calibration. HERA’s confusion limit is $S_{\min} \approx 11$ Jy. However, the Dec. $\approx -30^\circ$ stripe that it will scan is also covered by the TGSS survey that is complete down to ≈ 10 mJy. We therefore also consider an optimistic scenario in which the TGSS catalogue is used for calibration.

(iv) *The Square Kilometre Array*. We investigate the level of modelling noise in the SKA-1 LOW design proposed in Dewdney (2013) but scaled down to half of the described collecting area due to recent re-baselining. The array consists of 497, 30 m stations with a number density distributed as a Gaussian in radius where 75 per cent of antennas fall within 1 km of the centre, corresponding to a standard deviation of $\sigma_{\text{ant}} \sim 600$ m. The confusion limit of the SKA’s core is ≈ 27 mJy; however, the inclusion of extended baselines out to ≈ 100 km will bring the confusion limit at 150 MHz to ≈ 0.1 mJy (Prandoni & Seymour 2015), which we also consider as an optimistic case.

3.2 Modelling results

In Fig. 1, we plot cylindrically binned and averaged delay-transformed power spectra of residual visibilities from unmodelled foregrounds, calculated using equation (35) for the MWA, LOFAR,

HERA and SKA-1. We explore two different S_{\min} values for each array. As we might expect, the majority of residual power is contained within the wedge, arising from the last term in equation (35). This term is the power spectrum of the unmodelled residual sources and would exist in the absence of calibration errors. For the MWA, which has a smaller aperture, hence a wider primary beam, the wedge of unmodelled sources extends to larger k_{\parallel} values. Beyond the wedge extends the power spectrum of calibration errors that exist at the level of $10^6\text{--}10^8 h^3 \text{Mpc}^{-3} \text{mK}^2$, one to two orders of magnitude greater than the 21 cm signal. For the MWA, the level of contamination inside of the EoR window is within an order of magnitude of the simulated errors encountered in B16 who consider a calibration catalogue that is incomplete to a similar depth. It is apparent that the calibration errors experience a sharp cutoff at the k_{\parallel} corresponding to the delay of the edge of the main lobe on the longest baselines of the array. A vertical stripe of additional contamination appears in the LOFAR plot at $k_{\perp} \approx 0.6 h \text{Mpc}^{-1}$, which corresponds to separation scale for the high band (HBA) antenna pairs. Since even the longest outriggers participate in a short baseline with this length, more significant contamination is introduced at the corresponding Fourier mode.

We also estimate the region of k -space in which the 21 cm signal will be accessible by computing the ratio between the 2d power spectrum of residual visibilities and a representative signal computed using 21cmFAST⁶ (Mesinger & Furlanetto 2007; Mesinger, Furlanetto & Cen 2011). The reionization parameters are set to $T_{\text{vir}}^{\text{min}} = 2 \times 10^4 \text{K}$, $\zeta = 20$ and $R_{\text{mfp}} = 15 \text{Mpc}$, yielding a redshift of 50 per cent reionization of ≈ 8.5 . For fiducial catalogue limits, we see that the entire EoR window is unusable for LOFAR and the MWA while the SKA is only able to detect the signal at large $k \gtrsim 0.4 h \text{Mpc}^{-1}$. If LOFAR and the SKA use their extended baselines to obtain deep source catalogues and calibrate on these catalogues *with only their core antennas*, they will be able to isolate modelling errors to be contained primarily within the wedge. By calibrating on a deep 10 mJy catalogue such as the TGSS, HERA can rely on traditional sky-based calibration as a potential alternative to its primary redundant strategy.

3.3 The impact of primary beam modelling errors

So far, we have assumed that the antenna primary beam is known perfectly. Here we examine the impact of an imperfectly modelled beam on calibration noise. It is worth noting that while we only examine the impact of beam errors on calibration, errors in beam modelling can affect other aspects of the analysis (for example, the power spectrum normalization in equation 26). Recent *in situ* measurements with Orbcomm satellites (Neben et al. 2015, 2016) indicate that electromagnetic modelling of instrumental primary beams may only be accurate to the 1 per cent level within the central lobe and only to the 10 per cent level within the side lobes. Even if a complete model of the sky exists, systematic errors in the apparent flux of these sources will cause calibration errors similar to those encountered in Section 3.1. We describe beam modelling errors as an angle-dependent function, $D(\hat{s})$, that is added to the known component of the beam $B(\hat{s})$,

$$A(\hat{s}) = B(\hat{s}) + D(\hat{s}). \quad (44)$$

For the purposes of this section only, we take the optimistic case that we have a perfect external catalogue and that all calibration

⁶ http://homepage.sns.it/mesinger/DexM__21cmFAST.html

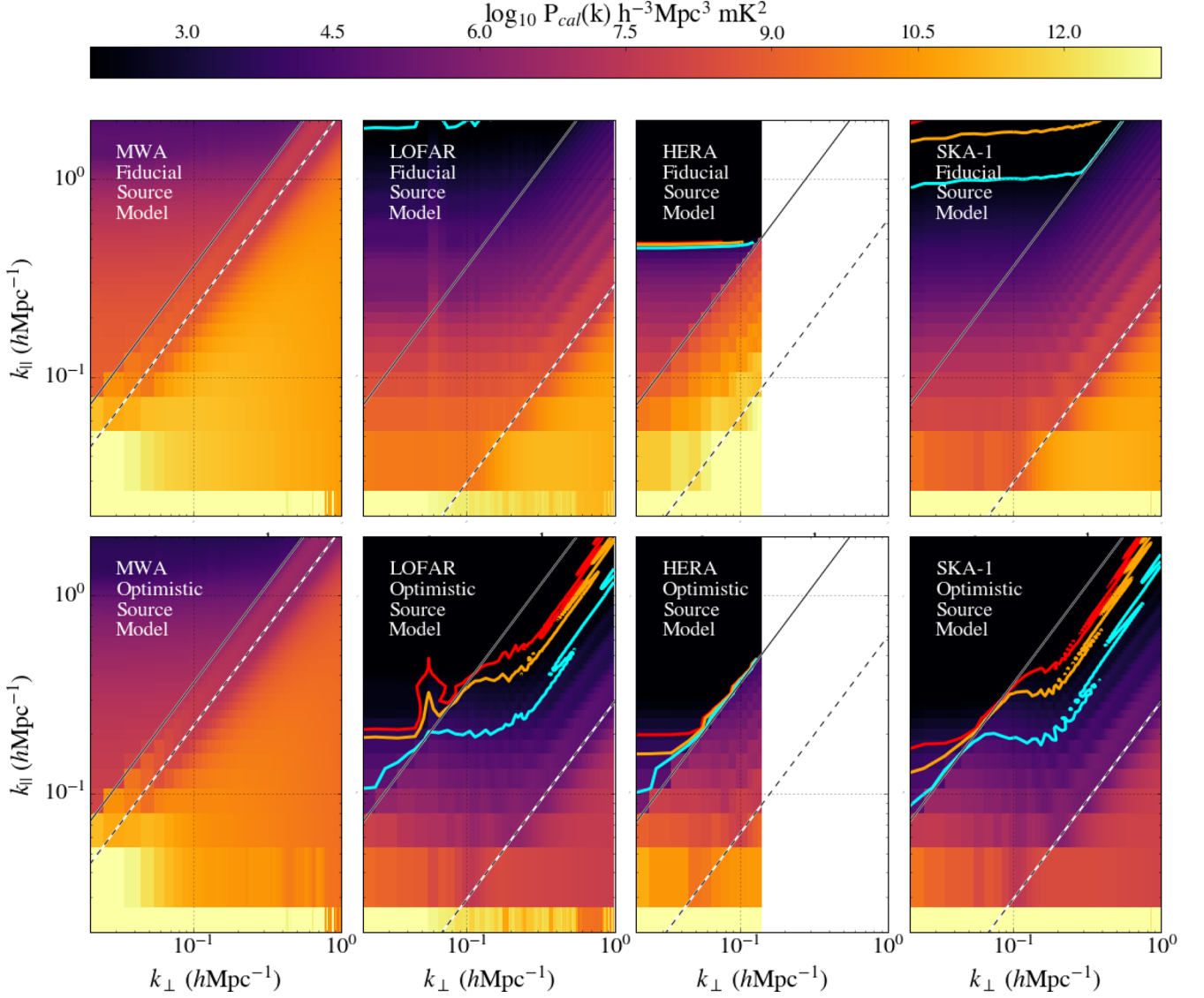


Figure 1. Top: the power spectrum residuals computed using equation (35) for the MWA, LOFAR, HERA and the SKA-1 LOW designs with sources modelled down to the array confusion limit. Unmodelled foregrounds are contained within the wedge, which is demarked by dashed black lines at the first primary beam null and solid black lines at the horizon. The calibration errors introduced by these foregrounds bleed out of the wedge into the EoR window. The narrower central lobes (larger stations) employed by LOFAR and the SKA help to significantly reduce the leakage at large k_{\parallel} that exists for the MWA. Contours where the signal, from a 21cmFAST simulation, is equal to 1, 5 and 10 times the calibration noise are indicated by cyan, orange and red lines, respectively. Bottom: the same as the top for optimistic scenarios. The optimistic scenario for LOFAR and the SKA involves complete modelling of point sources down to 0.1 mJy using additional long baselines. For HERA and the MWA, the optimistic scenario assumes that the 10 mJy source catalogue from the TGSS is used for calibration. If long baselines can faithfully model the sources down to 0.1 mJy, modelling noise does not appear to limit LOFAR and the SKA. Sky-based calibration with HERA is improved significantly by using a deep source catalogue from a complementary array. The vertical stripe in the LOFAR figure at $k_{\perp} \sim 0.6 h \text{Mpc}^{-1}$ arises from the arrangement of the HBA antennas in short spaced pairs so that even the outrigger antennas, which are heavily contaminated, participate in a single short baseline.

modelling error comes from an incorrect model of the primary beam. A true visibility in the presence of these errors is

$$\begin{aligned} v_{\alpha}^{\text{true}} &= \int d\Omega [B(\hat{s}) + D(\hat{s})] I(\hat{s}) e^{-2\pi i b_{\alpha} v/c} \\ &= y_{\alpha} + \int d\Omega D(\hat{s}) I(\hat{s}) e^{-2\pi i b_{\alpha} v/c}. \end{aligned} \quad (45)$$

Our new calibration residual, r_{α} , takes on the form

$$r_{\alpha} \rightarrow \int d\Omega D(\hat{s}) I(\hat{s}) e^{-2\pi i b_{\alpha} v/c}. \quad (46)$$

This leads to a new form of $R_{\alpha\beta}$ as well:

$$R_{\alpha\beta} \rightarrow \int_0^{S_{\text{max}}} dS \frac{d^2 N}{dS d\Omega} S^2 \int d\Omega |D(\hat{s})|^2 e^{-2\pi i (b_{\alpha} v/c - b_{\beta} v'/c)}, \quad (47)$$

where S_{max} is the flux of the highest flux source in the field of view that is obtained by setting the number of sources with intrinsic flux greater than S_{max} equal to unity,

$$S_{\text{max}} = S_* \left[k \int d\Omega A(\hat{s}) \right]^{1/\gamma}. \quad (48)$$

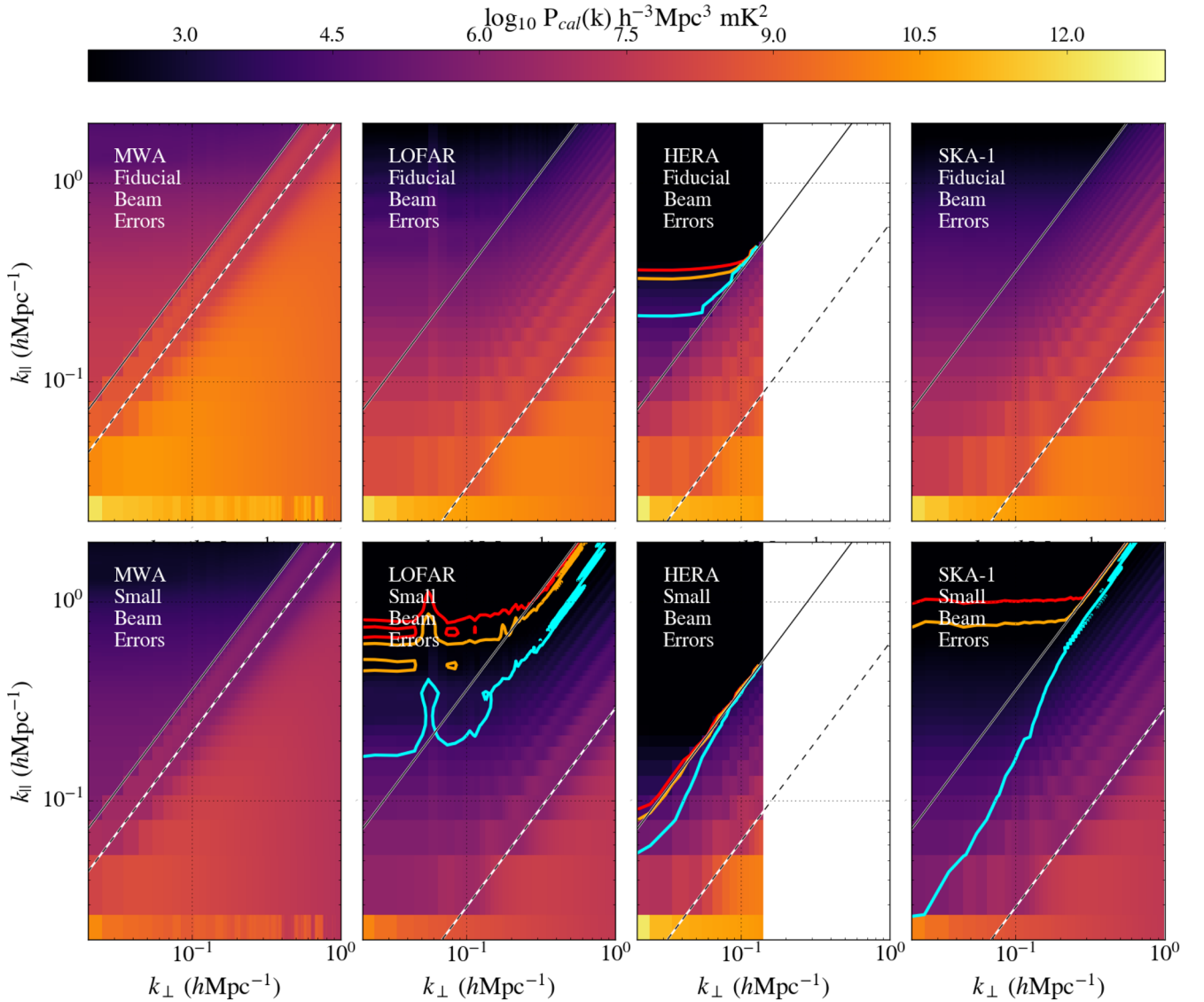


Figure 2. Top: same as Fig. 1 except now we consider a perfect calibration catalogue with calibration errors arising only from mismodelling the primary beam at the 1 per cent level in the main lobe and the 10 per cent level in the side lobes. With the exception of HERA, foreground residuals arising from primary beam modelling errors dominate the signal in the entire EoR window. Bottom: the same as above, but with a uniform modelling accuracy of 1 per cent. Much of the EoR window is still contaminated for LOFAR and the MWA while significant bias exists in much of the EoR window for the SKA.

Since the literature typically reports fractional errors in beam modelling, we describe $D(\hat{s})$ as the true beam $A(\hat{s})$ multiplied by a fractional error function $D(\hat{s}) = f(\hat{s})A(\hat{s})$, where we parametrize $f(\hat{s})$ as the following piecewise function,

$$f(\hat{s}) = \begin{cases} A [1 - (1 - e_z) \exp(-\cos^2 \theta / 2\sigma_e^2)] & |\cos \theta| < s_1 \\ A [1 - (1 - e_z) \exp(-s_1^2 / 2\sigma_e^2)] & |\cos \theta| \geq s_1, \end{cases} \quad (49)$$

where Ae_z is the fractional beam modelling error at the pointing centre, s_1 is the angular distance of the pointing centre to the first side lobe and A, σ_e may be adjusted to give different fractional modelling errors in the side lobes. This function allows us to assign an arbitrary modelling uncertainty to the zenith and side lobes. We compute the level of beam modelling noise in the 21 cm power spectrum for two different scenarios, one in which the beam is known to 1 per cent at zenith and 10 per cent in the side lobes, which is consistent with the precision reported in Neben et al. (2015). We

also consider a scenario in which an order-of-magnitude improvement in beam modelling has been achieved and the beam is known to 1 per cent in both the side lobes and the main lobe, which is the target precision for in-development drone experiments (Jacobs et al. 2017). We note that our model describes beam modelling errors that are completely correlated between antennas. It is possible that uncorrelated errors (which we might expect to arise from imperfections in the construction of each station) will integrate down differently from the modelling errors we consider here.

Plotting the resulting residual power spectra from equation (34) in Fig. 2, we see that with the current precision of primary beam models, the calibration noise masks the power spectrum across all of k -space for the MWA, LOFAR and the SKA. Even with an order-of-magnitude improvement in our modelling, a significant foreground bias $\gtrsim 20$ per cent of the signal amplitude will be present in measurements by LOFAR and SKA-1. Thus, even under the optimistic foreground modelling scenarios considered above, foreground errors will still contaminate the EoR window unless significant

Table 1. Inspection of the equations in Section 2 yields a number of analytic and qualitative relationships between the properties of an array and modelling catalogue.

Strategy	Impact on error $P(k)$
Reduce S_{\min}	Reduce amplitude as $S_{\min}^{3-\gamma}$
Reduce the standard deviation of antenna positions, σ_{ant}	(a) Increase amplitude as $\sigma_a^{2(3-\gamma)/(1-\gamma)}$ (b) Reduce maximum $k_{\parallel}^{\min} \sim \sigma_a$ of errors
Increase N_{ant}	(a) Reduce amplitude as N_{ant}^{-1} (b) Decrease S_{\min}^a
Increase aperture diameter, d_{ant}	(a) Reduces $k_{\parallel}^{\min} \sim d_{\text{ant}}^{-1}$ (b) Requires larger σ_{ant} leading to larger k_{\parallel}^{\min}

^aDepends on the distribution of additional antennas.

improvements in beam modelling are made. HERA’s compact layout limits the impact of beam errors to small delays so that a significant portion of the EoR window is accessible even in the fiducial beam modelling scenario.

3.4 The dependence of modelling noise on array and catalogue properties

We can use the equations developed in Section 2.3 to determine the impact of array configuration and catalogue depth on the power spectrum bias $\mathbf{P}_{\alpha\alpha}$. Since we are interested in the contribution from modelling errors that, unlike thermal noise, do not average down with integration time, we will let $\mathbf{C} = \mathbf{R}$ in equations (28)–(34). From equation (42), we list the effects of changing various parameters in the instrument design and source catalogue in Table 1. There are a number of adjustments in the array layout that can be made to reduce the amplitude of the errors. Several of these adjustments have multiple effects that work against each other.

3.4.1 Catalogue depth

Ignoring diffuse emission, the power spectrum of modelling errors in equation (42) is proportional to $\sigma_r^2 = \int_0^{S_{\min}} S^2 S^{-\gamma} dS \propto S_{\min}^{3-\gamma}$. With the power-law index of 1.75 for faint source populations, the noise level will scale as $S_{\min}^{1.25}$. Hence, clearing a contaminated region requires improvements in catalogue depth on the same order of magnitude as the ratio of the bias to the expected signal.

3.4.2 Time averaging

If the instrumental gains are stable in time, modelling noise can be suppressed by averaging over local sidereal times (LST). We investigate the level of suppression that is possible for *non-fringe-stopped* baselines using multi-field averaging by calculating the temporal coherence of the modelling noise over some time interval, Δt . After a time Δt has passed, the primary beam of the instrument that had a gain of $A(\hat{s})$ towards the brightness field at $I(\hat{s})$ at time t will now have a gain of $A(\hat{s})$ towards $I(\hat{s} + \Delta\hat{s})$ at time $t + \Delta t$,

$$\begin{aligned} \mathbf{R}_{\alpha\alpha}(v, t; v, t + \Delta t) &= \left(\int d\Omega d\Omega' e^{-2\pi i v b_{\alpha}(\hat{s}-\hat{s}')/c} \right) \\ &\quad \times \text{Cov}[I_r(\hat{s}), I_r^*(\hat{s}' + \Delta\hat{s})] \\ &= \left(\sigma_r^2 (S_{\min}) e^{-2\pi i v b_{\alpha} \Delta\hat{s}/c} \right) \\ &\quad \times \int d\Omega A(\hat{s}) A^*(\hat{s} - \Delta\hat{s}). \end{aligned} \quad (50)$$

When $\Delta\hat{s}$ is larger than the extent of the beam on the sky, the integral in equation (50) is close to zero. Hence, a baseline is temporally

coherent with itself when $\Delta\hat{s}$ is small enough that its fields of view at the different times overlap. If the gains are stable over time, one can calibrate on multiple fields and reduce the power spectrum of calibration modelling errors by a factor of N_p , where N_p is the number of non-overlapping pointings. More significant suppression can arise from the oscillating term in equation (50), which arises from our assumption that the sky has moved by $\Delta\hat{s}$ and would not appear in the covariance between non-fringe-stopped baselines. Averaging this oscillatory term over multiple LSTs can potentially lead to a significant reduction in the amplitude of modelling noise and is the subject of future work.

3.4.3 Array configuration

There are three primary ways of changing the array configuration to affect modelling errors.

(i) *Antenna distribution.* Reducing the length of baselines involved in calibration reduces the chromaticity of gain errors and thus the smallest Fourier mode, k_{\parallel}^{\min} , that is not dominated by modelling noise. On the other hand, the array PSF, and hence the minimal flux that an array can model for self-calibration, is also set by its compactness. If the antennas are distributed as a Gaussian with standard deviation σ_a , then the naturally weighted PSF can be approximated by a Gaussian with standard deviation, $\sigma_p = \lambda_0/(2\pi\sigma_a)$. Condon (1974) determines that the confusion limit of an array, S_{\min} , depends on the PSF as $\sigma_p^{2/(\gamma-1)} \propto \sigma_a^{2/(1-\gamma)}$. Since the amplitude of the calibration noise is proportional to $S_{\min}^{3-\gamma}$, the overall normalization of calibration noise will scale with the standard deviation of the antenna distribution as $\sigma_a^{2(3-\gamma)/(1-\gamma)} \sim \sigma_a^{-3.33}$. At a glance, this is a very steep change in amplitude that might counteract the decrease in chromaticity. However, will find below that the impact of chromaticity is much more important.

(ii) *Antenna count.* Increasing the number of antennas will cause the amplitude of the modelling noise power spectrum to reduce as $\sim N_{\text{ant}}$ but larger numbers of antennas will also force the array to be less compact, potentially increasing k_{\parallel}^{\min} while driving down the confusion limit.

(iii) *Antenna size.* Increasing the size of each antenna reduces the primary beamwidth and hence the contamination from foregrounds at delays near the horizon but also drives up the minimal baseline size.

The scaling of the noise with the array characteristics listed above can be illuminated with some further simplifying assumptions. In particular, if all of the stations have Gaussian beams with angular standard deviations of $\sigma_b \approx \epsilon\lambda/d_{\text{ant}}$, where d_{ant} is the antenna diameter and $\epsilon \approx 0.45$ and that the antennas are Gaussian distributed with a standard deviation of σ_{ant} , equation (42) allows us to derive a closed-form prediction of the minimal k_{\parallel} in such an interferometer that is not contaminated by foregrounds (Appendix E),

$$\begin{aligned} k_{\parallel}^{\min} &\approx 1.24 h \text{ Mpc}^{-1} \sqrt{\frac{1+z}{10}} \left(\frac{\sigma_{\text{ant}}}{1 \text{ km}} \right) \left(\frac{d_{\text{ant}}}{10 \text{ m}} \right)^{-1} \\ &\quad \times \left[1 + 0.35 \log \left(\frac{1+z}{10} \right) - 0.04 \log \left(\frac{\Omega_m}{0.27} \right) \right. \\ &\quad + 0.1 \log \left(\frac{S_{\min}}{10 \text{ mJy}} \right) - 0.08 \log \left(\frac{P_{21}}{10^4 \text{ mK}^2 h^{-3} \text{ Mpc}^3} \right) \\ &\quad - 0.08 \log \left(\frac{N_{\text{ant}}}{100} \right) - 0.08 \log \left(\frac{\sigma_{\text{ant}}}{1 \text{ km}} \right) \\ &\quad \left. + \log \left(\frac{d_{\text{ant}}}{10 \text{ m}} \right) \right], \end{aligned} \quad (51)$$

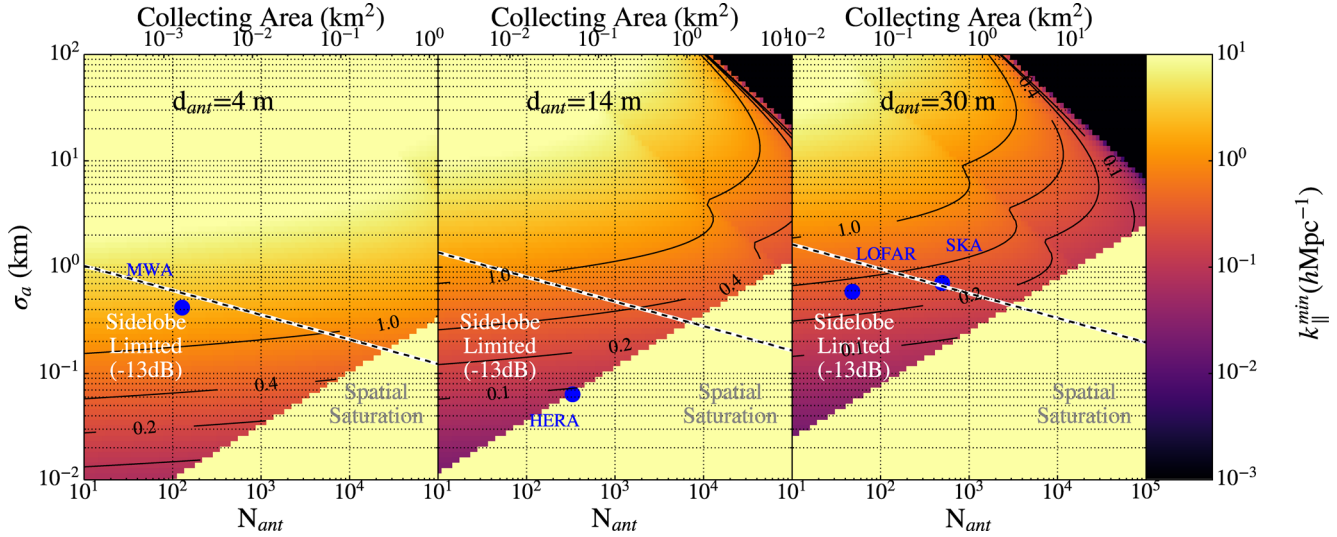


Figure 3. Using equation (51), we show the smallest k_{\parallel}^{\min} that is not dominated by modelling noise errors for arrays with Gaussian primary beams, a random circular Gaussian distribution of N_{ant} antennas with standard deviation σ_{ant} , each with diameter d_{ant} . The area below the white dashed line is where side lobes render the assumption of Gaussian beams inaccurate. Beige regions on the lower-right corner of each plot denote unphysically high packing densities. We see that for all existing instrument designs, calibration noise extends to large k_{\parallel} values that will reduce their sensitivity to the 21 cm signal. HERA will benefit greatly from the fact that it can be calibrated redundantly with minimal reliance on a sky model.

where P_{21} is the amplitude of the 21 cm power spectrum. This formula can be used to get a quick order-of-magnitude sense as to whether a mode will be accessible to an instrument; however, it is very optimistic in that it assumes a Gaussian primary beam. While it also strictly assumes that the antennas are distributed as a Gaussian, we have found that it holds to 10 percent accuracy for non-Gaussian arrays (such as LOFAR and the MWA) as well.

From equation (51), we see that the extent of modelling noise contamination depends primarily on $\sigma_{\text{ant}}/d_{\text{ant}}$ while other quantities, such as N_{ant} and S_{min} , contained within a logarithm have a much weaker impact on k_{\parallel}^{\min} . This proportionality makes sense intuitively since larger apertures have smaller primary beams, suppressing emission at large zenith angles and larger delay. In closely packed arrays, the $\sigma_{\text{ant}}/d_{\text{ant}}$ proportionality can be saturated so that the σ_r^2 inside of the logarithm will matter. While this equation ignores the existence of side lobes, it gives us an order-of-magnitude estimate of how modelling noise scales with array properties. In Fig. 3, the k_{\parallel}^{\min} values predicted from the naturally weighted confusion limits of various planned arrays exceed $k_{\parallel}^{\min} \gtrsim 0.2 h \text{ Mpc}^{-1}$, including for the SKA-1 core. Since interferometers such as the SKA and LOFAR focus most of their sensitivity at small k_{\parallel} values, their ability to detect the 21 cm signal will be heavily impacted by foreground modelling errors (Pober et al. 2014).

We can get a conservative sense for how side lobes extend k_{\parallel}^{\min} beyond the values predicted in equation (51) by setting the amplitude of the modelling noise at zero delay, multiplied by square of the side-lobe amplitudes (for an Airy beam, equal to -13 dB) equal to the 21 cm signal (see Appendix E). We denote the region of instrumental parameter space that is affected by side lobes in Fig. 3 with a grey overlay. Since all planned instruments fall within this region, the k_{\parallel}^{\min} predictions in this figure are actually optimistic. For these arrays, a more detailed calculation of equation (34) with realistic side lobes is necessary. We found in Section 3.2, with more realistic side lobes considered, that the k_{\parallel}^{\min} obtained is indeed significantly larger than predicted by equation (51).

4 ELIMINATING MODELLING NOISE WITH BASELINE WEIGHTING

While optimistic scenarios in foreground characterization may be precise enough to suppress calibration modelling noise below the 21 cm signal, elimination of this contamination will also require beam characterization that is beyond the current state of the art. Enabling a power spectrum detection in existing sky-based calibrated experiments calls for an alternate strategy. Redundant calibration is one existing, and so-far successful alternative though it can only be applied to regularly spaced arrays. Though redundant calibration does not rely on a detailed sky model, it is possible that antenna-to-antenna beam variations and position errors can violate the assumption of redundancy and introduce chromatic artefacts that are similar to the ones we have found for sky-based calibration, a potential shortcoming that is being investigated. One approach is to ensure that the instrument contains no structure in region of k -space relevant for 21 cm studies, allowing for smooth fits that do not contaminate the EoR window (B16). This is one of the approaches being adopted by HERA (Ewall-Wice et al. 2016a; Neben et al. 2016; Thyagarajan et al. 2016; Patra et al. 2017b) and an upgrade to the MWA. In this section, we explore an alternative strategy that can be used even when the bandpass is not already intrinsically smooth. By exponentially suppressing long baselines, sky-based calibration is able to remove fine-frequency structure while avoiding contamination within the EoR window.

Supra-horizon contamination from calibration noise arises from the inclusion of longer baselines in calibrating gain solutions that are applied to short baselines, leaking power from large to small k_{\perp} . One way of mitigating this source of contamination is to weight the visibilities contributing to each gain solution in a way that dramatically upweights short baselines over long ones. This can be accomplished by choosing an appropriate \mathbf{W} matrix in equations (16) and (17). In Section 4.1, we explore the efficacy of using a specific form of baseline weighting to eliminate modelling noise. The use of non-unity weights will result in an increase in thermal noise that we discuss in Section 4.2.

4.1 Gaussian weighting for sky-based calibration

We explore the performance of a \mathbf{W} matrix that downweights long baselines with the functional form

$$\mathbf{W}_{\alpha\beta} = \begin{cases} \exp\left(-\frac{b_\alpha^2}{2\sigma_w^2}\right) & \alpha = \beta \\ 0 & \alpha \neq \beta. \end{cases} \quad (52)$$

This function can result in weights that vary over a range beyond what is allowed for by numerical precision. We note that generally the off-diagonal elements of the weighting matrix can mix different baselines (which might be desired if we wished to suppress or emphasize features that covary between baselines). Our choice of a diagonal matrix corresponds to simply multiplying each visibility by a different weighting factor with which they will contribute to the sum of squares that is being minimized in determining $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\phi}}$. In order to avoid poorly conditioned matrices, a regularization term is also added equal to the identity multiplied by 10^{-6} , which is large enough to avoid numerical precision errors, but also small enough such that the weights on long baselines are negligible compared to the short ones (and below the dynamic range between foregrounds and signal). With this weighting, core antennas participating in many short baselines will have their gain solutions dominated by relatively achromatic core visibilities. Meanwhile, outrigger antennas that participate in only long baselines will derive their solutions from many baselines with similarly small weights. In both cases, a normalization step of $(\mathbf{AWA}^T)^{-1}$ corrects for the fact that these weights do not sum to unity. Thus, long and short baselines are both effectively calibrated using the Gaussian weighting scheme while the leakage of chromatic errors on long baselines into gain solutions being applied to short baselines is stymied.

We calculate $\mathbf{P}_{\alpha\alpha}$ given by equation (34) for the arrays considered in this paper with different values of σ_w . For LOFAR and the SKA, we use $\sigma_w = 100$ m. For the MWA and HERA, whose cores are especially compact and have larger fields of view than LOFAR and the SKA, we apply more aggressive weighting with $\sigma_w = 50$ m. We compare the cylindrically binned and averaged results in the middle row of Fig. 4 with cylindrical power spectra with \mathbf{W} equal to the identity (top row) and find that most of the EoR window is now free of foreground contamination with the power spectrum accessible at $k_{\parallel} \gtrsim 0.1 h \text{ Mpc}^{-1}$ for most arrays.

However, stripes of foreground contamination still extend into the EoR window at distinct k_{\perp} values in the MWA, LOFAR and to a lesser degree for the SKA. Isolating these baselines in the uv plane, we find that this contamination arises from antennas that are associated with less than two baselines that receive significant weighting. Define $n_{\text{eff}}(i)$ for the i th antenna to be equal to the sum of the weights of all visibilities that include this antenna divided by their maximum value. As far as calibration is concerned, $n_{\text{eff}}(i)$ describes the effective number of baselines that an antenna participates in. If the number of effective baselines that are used to derive gain solutions is too small, the system is underconstrained and a degeneracy exists between possible solutions for the antenna gains. Two antennas have two gains to solve for, but only one visibility between them. The estimator is forced to break these degeneracies by upweighting the contribution from the long baselines. An example where n_{eff} is smaller than 2, in our Gaussian weighting scheme, would be for an antenna that is extremely far away from all but one other antenna. Only a single baseline associated with this antenna has significant weight while the rest are downweighted to zero.

We identify these problematic baselines by calculating $n_{\text{eff}}(i)$ for each antenna. We then flag and exclude from the fit the highest

weighted visibilities on all antennas with $n_{\text{eff}} \leq 2$ until all n_{eff} are greater than 2. Flagging these visibilities leads to a loss in ≈ 6 per cent of visibilities for LOFAR, 1.3 per cent for the MWA, 0.1 per cent for the SKA and no visibilities for HERA. The high n_{eff} for HERA antennas is something we would expect given its compact configuration (every antenna has many short baselines associated with it). Similarly, the SKA core we model is a compact Gaussian with few isolated antennas. LOFAR, on the other hand, has antennas that are arranged in pairs that are separated by short distances so that all of the isolated outriggers have a single short baseline associated with them (which results in the vertical stripe at $k_{\perp} \approx 0.4 h \text{ Mpc}^{-1}$ in the second row of Fig. 4). The MWA lacks these pairs, and as a result has fewer low- n_{eff} antennas that tend to lie in the transition between its compact core and extended outriggers. We show cylindrically binned power spectra formed from the delay-transformed residuals of unflagged visibilities in the bottom row of Fig. 4, finding that upon flagging this small population of problematic baselines, the EoR window is almost entirely clear above $0.1 h \text{ Mpc}^{-1}$ for all arrays studied. We also show the delay-transformed power spectrum estimates of visibilities contaminated by primary beam modelling errors of 1 per cent at zenith and 10 per cent in the side lobes with and without Gaussian visibility weighting applied in the calibration solutions (Fig. 5). With Gaussian weighting, we are also able to mitigate contamination with the current level of primary beam modelling errors.

4.2 The impact of inverse baseline weighting on power spectrum sensitivity

For an interferometer with identical antenna elements, the thermal noise level on every baseline is the same and \mathbf{N} is proportional to the identity matrix. For the point source approximation of the modelled foregrounds, the optimal weighting minimizing the errors due to thermal noise in each gain solution is therefore also the identity matrix. Because of its departure from identity weights, the Gaussian weighting that we proposed in the previous section has the effect of increasing thermal noise uncertainties in both the gains and the final power spectrum estimate. In order to see how Gaussian weighting increases the variance due to thermal noise in the gain solutions, one can consider the fact that the variance of the gain solutions goes as N_{ant}^{-1} (equation 42). For a particular antenna gain, Gaussian weighting reduces the effective number of visibilities whose noises are averaged over in each gain solution so that the variance of the antenna gain is now $\sim n_{\text{eff}}^{-1}$ rather than N_{ant}^{-1} . In the weighting schemes employed in Section 4.1, n_{eff} goes down by a factor of order 1–10, remaining between 10 and 100 for LOFAR and the MWA.

Assuming Gaussian errors, the covariance between the square of two delay-transformed visibilities is given by

$$\begin{aligned} \sigma_{\alpha\beta}^2 &= \langle (|\tilde{V}_\alpha(\tau)|^2 |\tilde{V}_\beta(\tau)|^2) \rangle - \langle (|\tilde{V}_\alpha(\tau)|^2) \rangle \langle (|\tilde{V}_\beta(\tau)|^2) \rangle \\ &= (\mathbf{P}_{\alpha\beta}^N + \mathbf{P}_{\alpha\beta}^R + \mathbf{P}_{\alpha\beta}^S)^2, \end{aligned} \quad (53)$$

where $\mathbf{P}_{\alpha\beta}^N \equiv \langle \tilde{V}_\alpha^N(\tau) \tilde{V}_\beta^{N*}(\tau) \rangle$ is the covariance matrix of the thermal noise component of delay-transformed visibilities and $\mathbf{P}_{\alpha\beta}^{R(S)}$ are the covariances of the delay-transformed residual foreground (signal) visibilities. While the residual foreground component can contribute significantly, it is only of concern in the regions of k -space where the amplitude of the foreground modelling noise is comparable to or greater than the level of the 21 cm signal. Since we are interested in how the thermal noise increases

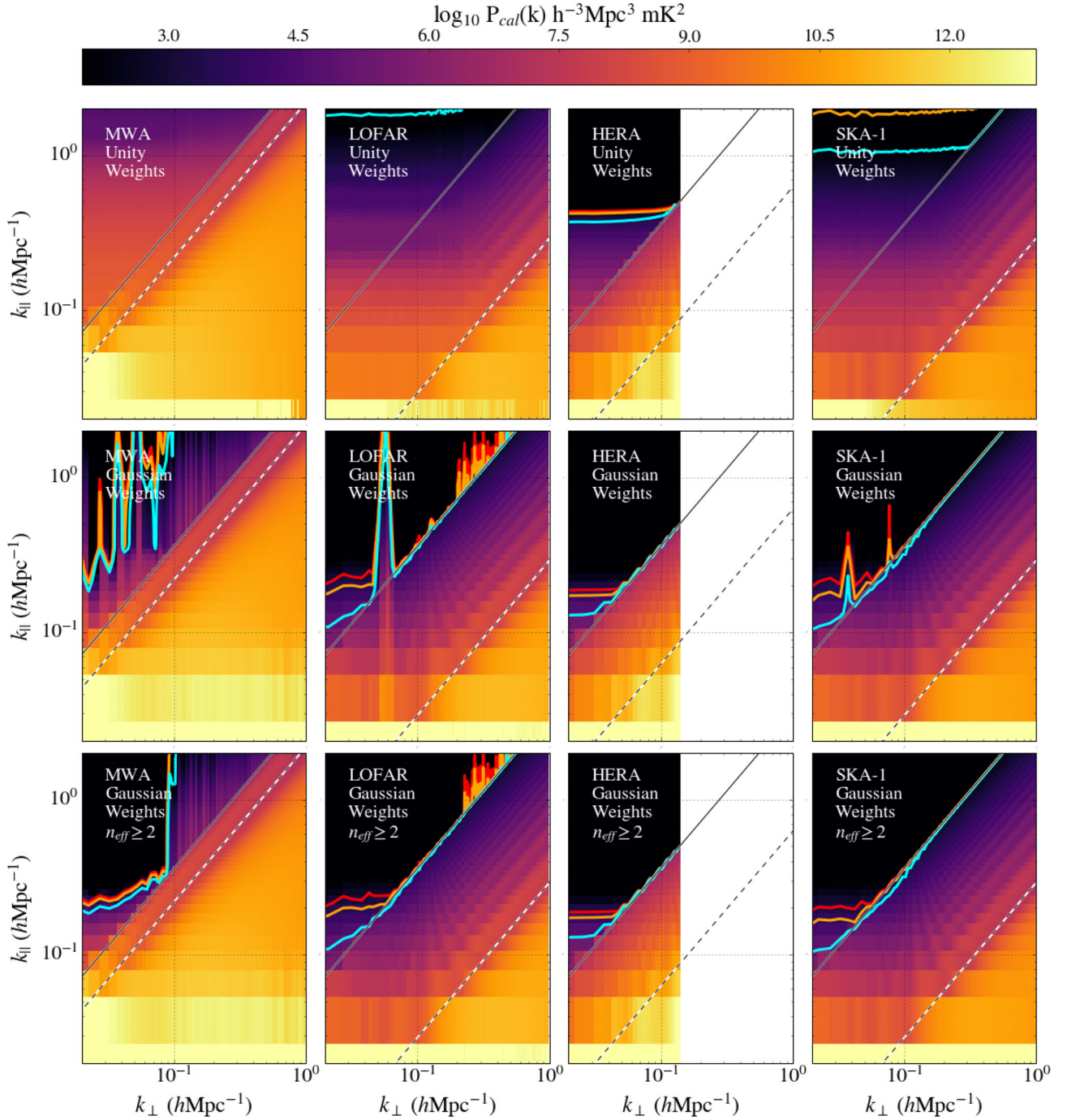


Figure 4. Top: residual power spectra with each visibility weighted equally in determining the calibration solutions (\mathbf{W} set to the identity matrix). Middle row: the same but now weighting visibilities with a Gaussian function of baseline length (equation 52). Much of the EoR window is cleared of contamination from calibration residuals. However, pronounced stripes of contamination still exist, especially for LOFAR and the MWA. These stripes arise from short baselines formed from antennas involved in no other short baselines. In order to solve for both antenna gains, they must use information from long baselines, resulting in significant chromaticity on the few short baselines to which the problematic antenna gains are applied. Bottom: flagging visibilities after calibration until all gains participate in $n_{\text{eff}} \geq 2$ baselines, we find the EoR window free of these stripes. To reiterate, solid lines demarcate regions where the fiducial EoR signal is 1, 5 or 10 times the power of the calibration modelling error. The dashed diagonal line indicates the location of the wedge associated with the first null of the primary beam; the solid line indicates the horizon wedge.

in the region of k -space where we have reduced foreground bias to well below the signal level, we will focus our attention on the thermal noise component and ignore the sample variance from modelling noise and signal for the remainder of this discussion.

We may compute $\mathbf{P}_{\alpha\beta}^N$ using equation (34) with $\mathbf{C} \rightarrow \mathbf{N}$. Typically, thermal noise is uncorrelated between baselines so $\tilde{\mathbf{N}}^{\alpha\beta}$ is diagonal. In the absence of calibration errors, the covariance between the squares of different delay-transformed visibility products arising from thermal noise would therefore also be zero. The

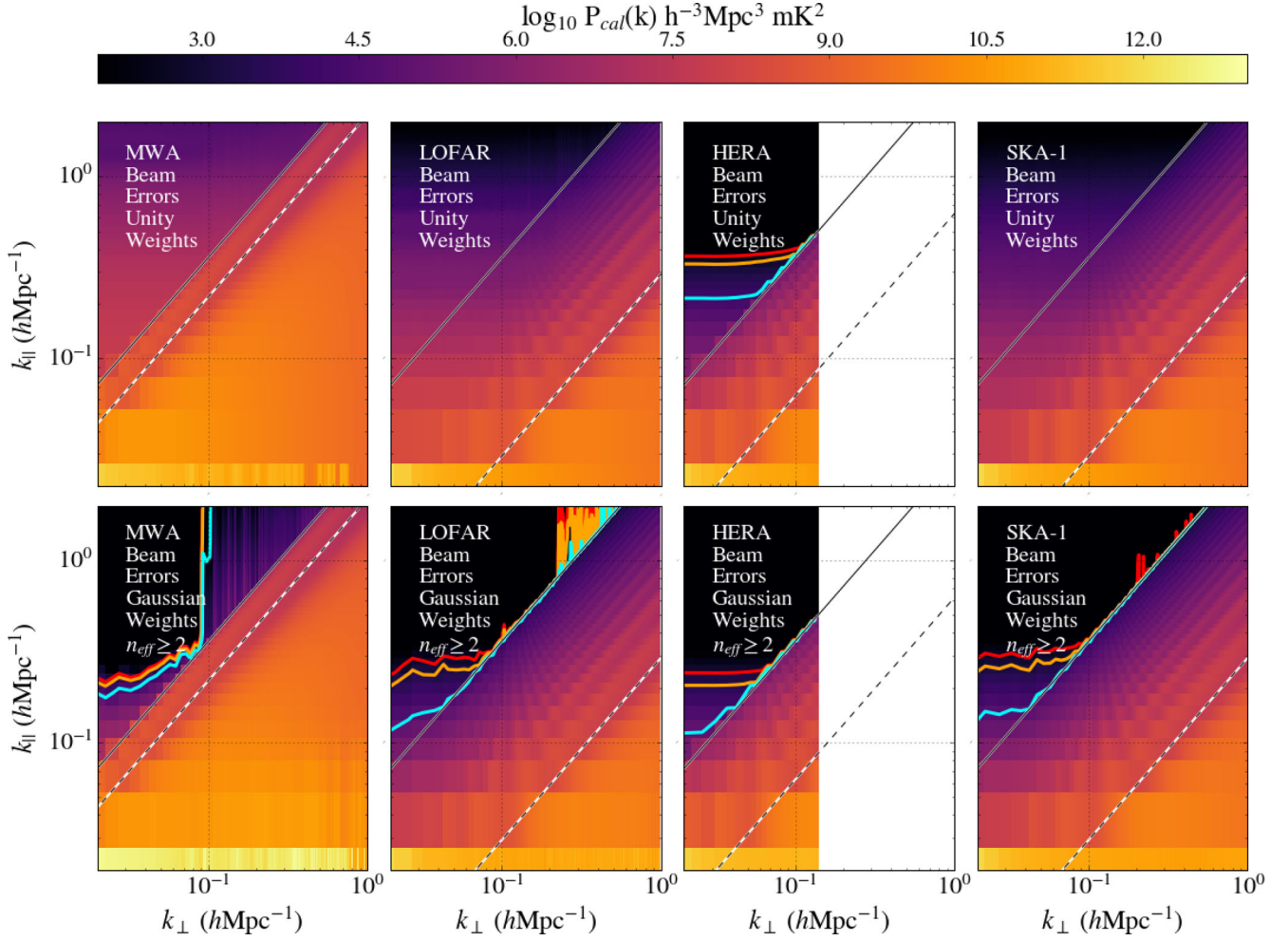


Figure 5. Top: cylindrically binned power spectra of calibration errors due to beam modelling errors at the level of 1 per cent in the main lobe and 10 per cent in the side lobes (equation 49). Each visibility has been weighted equally in determining the calibration solutions. Bottom row: the level of cylindrically binned power spectrum residuals from the primary beam modelling errors in the top row but now with calibration solutions derived from visibilities that are weighted with equation (52). Short baselines contributing to antennas with $n_{\text{eff}} \leq 2$ have also been flagged from the calibration fit. Weighting with a Gaussian is capable of removing calibration modelling errors due to beam mismodelling at the level that we see in today’s experiments (Neben et al. 2015).

presence of calibration errors introduces additional components to the thermal noise in all but the last term of equation (34) that are correlated from baseline to baseline. For identity weights and a diagonal noise covariance, the off-diagonal terms in $\mathbf{P}_{\alpha\beta}$ go roughly as $\sim N_{\text{ant}}^{-2} \sim N_{\text{vis}}^{-1}$ compared to the diagonal terms (which have the order unity contribution that does not arise from calibration). Thus, for $\alpha \neq \beta$, $\sigma_{\alpha\beta}^2 \sim N_{\text{vis}}^{-2} \sigma_{\alpha\alpha}^2$ and has, so far, been ignored in other sensitivity calculations (e.g. McQuinn et al. 2006; Parsons et al. 2012a; Beardsley et al. 2013; Pober et al. 2014).

In order to obtain enough sensitivity for a detection, interferometry experiments are expected to perform spherical binning and averaging in k -space to obtain power spectrum estimates, \hat{p}_A , whose covariance we denote as Σ_{AA} (denoting band powers with uppercase Latin subscripts). The variance of a binned and averaged power spectrum estimate with identity weights is given by

$$\begin{aligned} \Sigma_{AA} &= N_A^{-2} \left(\sum_{\alpha \in A} \sigma_{\alpha\alpha}^2 + \sum_{\alpha \in A} \left[\sum_{\beta \in A: \alpha \neq \beta} \sigma_{\alpha\beta}^2 \right] \right) \\ &\sim N_A^{-2} \left(\sum_{\alpha \in A} \sigma_{\alpha\alpha}^2 + \frac{N_A}{N_{\text{vis}}^2} \sum_{\alpha \in A} \sigma_{\alpha\alpha}^2 \right). \end{aligned} \quad (54)$$

Thus, the contribution to Σ_{AA} from off-diagonal elements of the noise covariance is subdominant to the contribution from diagonal elements as $\sim N_A/N_{\text{vis}}^2$, where N_A is the number of visibilities averaged within the A th bin.

Non-uniform weighting in calibration decreases the effective number of visibilities in calibration, increasing the off-diagonal terms in equation (34). This in turn leads to an increase in the overall error bar on each spherically binned and averaged power spectrum estimate. We compute the degree to which Gaussian weighting degrades sensitivity to the spherically binned power spectrum by comparing Σ_{AA} for both uniform and Gaussian weighting within a single LST. While calibration correlates the noise on different squared visibilities in the same power spectrum bin, we can minimize the extra error by inverse-covariance weighting them before averaging.

We perform this averaging and report how the Gaussian down-weighting of long baselines affects the thermal noise on the final power spectrum estimate in Fig. 6. Because the covariance matrices for HERA and the SKA are very large and would require significant computation to invert, we only perform this calculation for LOFAR and the MWA. We also assume that each power spectrum estimate only incorporates visibilities outside of the wedge. The proportion

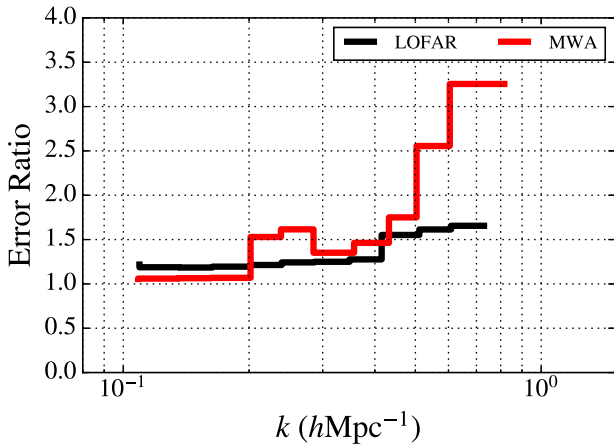


Figure 6. The ratio between thermal noise errors on a spherically averaged power spectrum estimate with Gaussian weighting and uniform weighting of visibilities in calibration. We show this ratio for LOFAR where $\sigma_w = 100$ m and the MWA with $\sigma_w = 50$ m. In both cases, the reduction in sensitivity to the power spectrum is by a factor less than 2 for small k where the interferometers have maximum sensitivity. Under the Gaussian weighting scheme, antennas with fewer short baselines have increased thermal noise in their gains. Increasingly large k -bins include larger numbers of visibilities formed from antennas with fewer short baselines (small n_{eff}) that have large increases in their thermal noise, leading to a trend of increasing sensitivity loss with increasing k . Since the MWA has a narrower weighting function, with $\sigma_w = 50$ m, this increase occurs faster than for LOFAR.

of long baselines that tend to be formed from antennas with smaller n_{eff} increases with each k -bin. Hence, the decrease in sensitivity increases with k . Since the MWA weighting function is more compact, with $\sigma_w = 50$ m, the increase in the error ratio goes faster than for LOFAR that has a wider weighting function with $\sigma_w = 100$ m. Within the region that instruments are expected to be sensitive to the 21 cm signal, the error bars only go up by less than two. Gaussian weighting increases the thermal noise in the power spectrum measurement, but only by a level similar to intrinsic thermal noise that would be present even if calibration were perfect. Gaussian weighting can therefore allow us to circumvent the problem of foreground modelling noise in calibration while only sacrificing a small amount of sensitivity to the 21 cm power spectrum.

While baseline-dependent weighting is able to clear the EoR window, it does not necessarily allow any instrument to work within the wedge. Figs B1 and 2 show that this would still require superb foreground models accurate to the 0.1 mJy level and modelling of the primary beam to the 10^{-3} level in the main lobe and 10^{-2} level in the side lobes. Until these milestones are achieved, extended arrays will suffer a disproportionate reduction in delay power spectrum sensitivity relative to compact arrays like HERA (Poher et al. 2014).

5 CONCLUSIONS

In this work, we derived expressions for the amplitude of the power spectrum bias arising from the imprint of foreground modelling errors on calibration. These expressions assumed that calibration errors are small enough such that their solutions are obtained through a linear set of equations, which is the case in the final stages of iterative, sky-based calibration schemes when the errors in the foreground model are small. Using these equations, we are able to explain the amplitude of the biases that have been simulated for the special cases of the MWA (B16) and LOFAR (Patil et al. 2016) and to predict the amplitude of modelling noise in the power spectrum

for the SKA-1 and HERA (which does not actually rely on this approach). We performed this analysis in a variety of foreground and beam modelling scenarios. We also use our formalism to determine the dependence of modelling noise on the parameters of the array and the accuracy of the calibration catalogue. These results do not apply to the redundant calibration strategies used by HERA and PAPER, although errors introduced by deviations from redundancy still have the potential to contaminate the window in a similar way. Our analysis also reveals that noise bias exists in current power spectrum estimates where separate calibration solutions are not obtained for interleaved data sets. Whether this bias limits 21 cm experiments requires further analysis but it can easily be avoided by obtaining independent calibration solutions for cross-multiplied data.

This paper is aimed to illuminate the source of calibration errors within the EoR window. In order to make our analysis analytically tractable, we employed a number of assumptions. These include assuming that the array is minimally redundant so that we can ignore off-diagonal elements of the visibility covariance matrix, and that the sources themselves are flat spectrum. A more significant assumption that will not hold in many observing scenarios is that we ignored the chromaticity of modelled foregrounds, which holds approximately when the modelled fluxes are dominated by a source at the phase centre that exceeds the flux of the next brightest source by a factor of a few. We also assumed that our instruments had Airy beams, that sources could be characterized down to a fixed flux level across the entire sky, ignored ionospheric effects and polarization (Sault, Hamaker & Bregman 1996; Jelić et al. 2010; Moore et al. 2013, 2017; Asad et al. 2015; Kohn et al. 2016) which is especially severe on the large spatial scales (Lenc et al. 2016) that we suggest should be relied upon in calibration strategies. Hence, specific quantitative predictions in this paper should be regarded as accurate to within an order of magnitude and on the optimistic side. In validating the design of future instruments, full end-to-end simulations should be employed, though this is left to future work.

Our calculations indicate that for current catalogue limits presented in Carroll et al. (2016), Hurley-Walker et al. (2017) and Williams et al. (2016) both the MWA and LOFAR will observe an EoR window that is heavily contaminated by chromatic calibration errors due to unmodelled sources. Since the chromaticity of these errors increases with the length of baselines involved in calibration, removing inner baselines from calibration, as is required to avoid signal loss with direction-dependent calibration (Patil et al. 2016), will only exacerbate these chromatic errors and is probably the source of the systematics floor observed by LOFAR in Patil et al. (2017, these authors note that calibration errors as a likely culprit but not that the use of long baselines is exacerbating the problem). Our analytic treatment suggests that instead sky-based experiments should use their short baselines to calibrate power spectrum data that may preclude the direction-dependent approach to avoid signal loss and will likely require more accurate models of diffuse emission. LOFAR may also be able to reduce the amplitude of calibration errors below the power spectrum, at large spatial scales, by averaging over multiple fields of view (if its gains are temporally stable) and/or by building a source catalogue complete down to ≈ 100 μ Jy across the entire sky. Even if such a catalogue is constructed, beam modelling precision will also need to be improved by an order of magnitude over what has been achieved in the literature. The large field of view on the MWA decreases the number of fields that can be averaged over and increases the k_{\parallel} values contaminated by modelling errors, making the path to remove this noise with extant methods considerably more difficult than for LOFAR.

Our analysis motivates a potential solution to the problem of modelling noise in sky-based calibration. Since contamination within the EoR window arises from the coupling of long baseline errors into the calibration solutions on short ones, our proposed strategy is to downweight the contribution of long baselines to the gain solutions that are applied to short baselines. The linear least-squares estimator formalism employed in this paper provides a natural framework for incorporating such weights. Experimenting with a Gaussian weighting scheme, we find that downweighting long baselines should allow for both existing and future arrays to correct fine-frequency band-pass structures without introducing chromatic sky-modelling errors. While such weighting will increase the level of thermal noise present in calibration solutions, we find that this noise increase will only result in power spectrum error bars that are ≈ 1 – 1.5 times larger than the case where all visibilities are weighted identically. This method prevents calibration errors from limiting the foreground avoidance approach, which seeks to detect the 21 cm signal within the EoR window and thus requires the calibrated instrumental response to be spectrally smooth. This method is not sufficient to enable *foregrounds subtraction*, accessing the signal inside of the wedge. Working within the delay wedge will require significant improvements in foreground and primary beam modelling.

ACKNOWLEDGEMENTS

We would like to thank Danny Jacobs, Nichole Barry, Miguel Morales, Jonathan Pober, Bryna Hazelton, Cathryn Trott and Aaron Parsons for helpful discussions. AEW acknowledges support from an NSF Graduate Research Fellowship under Grant No. 1122374. JSD acknowledges support from a Berkeley Center for Cosmological Physics Fellowship. AL acknowledges support for this work by NASA through Hubble Fellowship grant #HST-HF2-51363.001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555. A portion of this work was performed at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1066293.

REFERENCES

- Ali Z. S. et al., 2015, *ApJ*, 809, 61
 Asad K. M. B. et al., 2015, *MNRAS*, 451, 3709
 Barry N., Hazelton B., Sullivan I., Morales M. F., Pober J. C., 2016, *MNRAS*, 461, 3135 (B16)
 Beardsley A. P. et al., 2012, *MNRAS*, 425, 1781
 Beardsley A. P. et al., 2013, *MNRAS*, 429, L5
 Beardsley A. P. et al., 2016, *ApJ*, 833, 102
 Carroll P. A. et al., 2016, *MNRAS*, 461, 4151
 Condon J. J., 1974, *ApJ*, 188, 279
 Datta A., Bowman J. D., Carilli C. L., 2010, *ApJ*, 724, 526
 DeBoer D. et al., 2017, *PASP*, 129, 045001
 Dewdney P., 2013, System Baseline Design. Available at: http://www.skatelescope.org/wp-content/uploads/2013/03/SKA-TELSKO-DD-001-1_BaselineDesign1.pdf, p. 1
 Di Matteo T., Perna R., Abel T., Rees M. J., 2002, *ApJ*, 564, 576
 Dillon J. S., Parsons A. R., 2016, *ApJ*, 826, 181
 Dillon J. S., Liu A., Tegmark M., 2013, *Phys. Rev. D*, 87, 043005
 Dillon J. S. et al., 2014, *Phys. Rev. D*, 89, 023002
 Dillon J. S. et al., 2015a, *Phys. Rev. D*, 91, 023002
 Dillon J. S. et al., 2015b, *Phys. Rev. D*, 91, 123011
 Ewall-Wice A. et al., 2016a, *ApJ*, 831, 196
 Ewall-Wice A. et al., 2016b, *MNRAS*, 460, 4320
 Fixsen D. J. et al., 2011, *ApJ*, 734, 5
 Furlanetto S. R., 2016, in Mesinger A., ed., *Astrophysics and Space Science Library*, Vol. 423, Understanding the Epoch of Cosmic Reionization. Springer, Berlin, p. 247
 Haslam C. G. T., Salter C. J., Stoffel H., Wilson W. E., 1982, *A&AS*, 47, 1
 Hurlay-Walker N. et al., 2017, *MNRAS*, 464, 1146
 Intema H. T., Jagannathan P., Mooley K. P., Frail D. A., 2017, *A&A*, 598, A78
 Jacobs D. C. et al., 2011, *ApJ*, 734, L34
 Jacobs D. C. et al., 2013, *ApJ*, 776, 108
 Jacobs D. C., Burba J., Bowman J., Neben A. R., Stinnett B., Turner L., 2017, *PASP*, 129, 035002
 Jacobs D. C. et al., 2016, *ApJ*, 825, 114
 Jelić V., Zaroubi S., Labropoulos P., Bernardi G., de Bruyn A. G., Koopmans L. V. E., 2010, *MNRAS*, 409, 1647
 Kazemi S., Yatawatta S., 2013, *MNRAS*, 435, 597
 Kazemi S., Yatawatta S., Zaroubi S., Lampropoulos P., de Bruyn A. G., Koopmans L. V. E., Noordam J., 2011, *MNRAS*, 414, 1656
 Kazemi S., Yatawatta S., Zaroubi S., 2013, *MNRAS*, 430, 1457
 Kohn S. A. et al., 2016, *ApJ*, 823, 88
 Lenc E. et al., 2016, *ApJ*, 830, 38
 Liu A., Tegmark M., 2011, *Phys. Rev. D*, 83, 103006
 Liu A., Tegmark M., 2012, *MNRAS*, 419, 3491
 Liu A., Tegmark M., Morrison S., Lutmirski A., Zaldarriaga M., 2010, *MNRAS*, 408, 1029 (L10)
 Liu A., Parsons A. R., Trott C. M., 2014a, *Phys. Rev. D*, 90, 023018
 Liu A., Parsons A. R., Trott C. M., 2014b, *Phys. Rev. D*, 90, 023019
 McQuinn M., 2016, *ARA&A*, 54, 313
 McQuinn M., Zahn O., Zaldarriaga M., Hernquist L., Furlanetto S. R., 2006, *ApJ*, 653, 815
 Mesinger A., Furlanetto S., 2007, *ApJ*, 669, 663
 Mesinger A., Furlanetto S., Cen R., 2011, *MNRAS*, 411, 955
 Mitchell D. A., Greenhill L. J., Wayth R. B., Sault R. J., Lonsdale C. J., Cappallo R. J., Morales M. F., Ord S. M., 2008, *IEEE J. Sel. Top. Signal Process.*, 2, 707
 Moore D. F., Aguirre J. E., Parsons A. R., Jacobs D. C., Pober J. C., 2013, *ApJ*, 769, 154
 Moore D. F. et al., 2017, *ApJ*, 836, 154
 Morales M. F., Hewitt J., 2004, *ApJ*, 615, 7
 Morales M. F., Hazelton B., Sullivan I., Beardsley A., 2012, *ApJ*, 752, 137
 Neben A. R. et al., 2015, *Radio Sci.*, 50, 614
 Neben A. R. et al., 2016, *ApJ*, 826, 199
 Newburgh L. B. et al., 2014, *Proc. SPIE*, 9145 91454V
 Offringa A. R. et al., 2016, *MNRAS*, 458, 1057
 Oh S. P., Mack K. J., 2003, *MNRAS*, 346, 871
 Paciga G. et al., 2013, *MNRAS*, 433, 639
 Parsons A., Pober J., McQuinn M., Jacobs D., Aguirre J., 2012a, *ApJ*, 753, 81
 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., 2014, *ApJ*, 788, 106
 Patil A. H. et al., 2016, *MNRAS*, 463, 4317
 Patil A. H. et al., 2017, *ApJ*, 838, 65
 Patra N., Bray J., Roberts P., Ekers R., 2017a, *Exp. Astron.*, 43, 119
 Patra N. et al., 2017b, preprint ([arXiv:1701.03209](https://arxiv.org/abs/1701.03209))
 Pober J. C. et al., 2014, *ApJ*, 782, 66
 Pober J. C. et al., 2016, *ApJ*, 819, 8
 Prandoni I., Seymour N., 2015, *Proc. Sci.*, Revealing the Physics and Evolution of Galaxies and Galaxy Clusters with SKA Continuum Surveys. SISSA, Trieste, PoS(AASKA14)067
 Remazeilles M., Dickinson C., Banday A. J., Bigot-Sazy M.-A., Ghosh T., 2015, *MNRAS*, 451, 4311
 Rogers A. E. E., Bowman J. D., 2008, *AJ*, 136, 641
 Sault R. J., Hamaker J. P., Bregman J. D., 1996, *A&AS*, 117, 149
 Sullivan I. S. et al., 2012, *ApJ*, 759, 17
 Thyagarajan N. et al., 2013, *ApJ*, 776, 6
 Thyagarajan N. et al., 2015, *ApJ*, 804, 14
 Thyagarajan N., Parsons A., DeBoer D., Bowman J., Ewall-Wice A., Neben A., Patra N., 2016, *ApJ*, 825, 9
 Tingay S. J. et al., 2013, *PASA*, 30, 7
 Trott C. M., Wayth R. B., 2016, *PASA*, 33, e019
 Trott C. M., Wayth R. B., Tingay S. J., 2012, *ApJ*, 757, 101
 Trott C. M. et al., 2016, *ApJ*, 818, 139

van Haarlem M. P. et al., 2013, *A&A*, 556, A2
 Vedantham H., Udaya Shankar N., Subrahmanyan R., 2012, *ApJ*, 745, 176
 Wang X., Tegmark M., Santos M. G., Knox L., 2006, *ApJ*, 650, 529
 Wieringa M. H., 1992, *Exp. Astron.*, 2, 203 (W92)
 Williams W. L. et al., 2016, *MNRAS*, 460, 2385
 Zaldarriaga M., Furlanetto S. R., Hernquist L., 2004, *ApJ*, 608, 622
 Zheng H. et al., 2014, *MNRAS*, 445, 1084
 Zheng H. et al., 2017a, *MNRAS*, 464, 3486
 Zheng H. et al., 2017b, *MNRAS*, 465, 2901

APPENDIX A: THE IMPACT OF REDUNDANCY

Throughout this paper, we ignored the impact of redundancy between visibilities, letting $\mathbf{R}_{\alpha\beta}$ be diagonal when calculating modelling noise. However, redundancy is significant in highly compact arrays, such as HERA. Here we argue that the impact of redundancy on the modelling noise levels, calculated in this work, is to multiply the overall noise level by a factor of order unity, which only has a small effect on the extent of contaminated modes in k -space. We also verify this argument with a numerical calculation.

When \mathbf{C} is diagonal, the sum in equation (35) is only over terms with $\gamma = \delta$. The existence of redundant baselines introduces non-negligible off-diagonal terms in the visibility covariance matrix \mathbf{R} . For each $\gamma = \delta$ term in the non-redundant sum, we can consider the additional summands, with $\gamma \neq \delta$ that are introduced for each ii/jj term and ij/ji term. We start with ii/ji .

For a fixed baseline γ that involves antenna i , there will be at most $\sim N_{\text{ant}}$ additional baselines that are redundant with γ and do not involve gain i . From equations (39) and (40), the weighting of the covariance between two different baselines in which only one involves antenna i goes as $\sim N_{\text{ant}}^{-3}$. Thus, the presence of redundancy adds no more than $\sim N_{\text{ant}}$ terms involving antenna i but not antenna j and vice versa, for each ii summand in equation (35). Multiplying this overall factor of N_{ant}^{-2} by N_{ant} to account for the N_{ant} different ii sums leads to a contribution to the noise amplitude of the order of $\sim N_{\text{ant}}^{-1}$, similar to the level of the noise without redundancy. As a result, redundancy changes the modelling error amplitude by a factor of order unity in the diagonal terms. Next, we consider the ij terms in equation (35).

For a given γ and $i \neq j$, there will be at most $\sim N_{\text{ant}}$ redundant baselines that do not involve the i th or j th gains, causing the weighting of each unique variance term to go as N_{ant}^{-3} rather than N_{ant}^{-4} in the non-redundant case. Since there are $\sim N_{\text{ant}}^2$ unique baselines that do not involve i or j , the overall sum of these terms goes as N_{ant}^{-1} . As a result, the ij terms in equation (35) will have a similar magnitude as the ii/jj terms but the overall impact on the amplitude of the modelling noise described in equation (35) still changes the amplitude by a factor of order unity.

We confirm these arguments with a numerical comparison between the amplitude of the modelling noise with and without redundancy taken into account for two redundant arrays of 91 and 331 hexagonally packed 14 m apertures and S_{min} equal to the naturally weighted confusion limit. We compute the off-diagonal elements of \mathbf{R} by numerically computing the beam integral in equation (5) for all $\mathbf{R}_{\alpha\beta}$ with Airy beams and perform the full matrix inversions prescribed in equations (16) and (17). We compare our results to the same calculation where all off-diagonal elements of \mathbf{R} are set to zero (Fig. A1) and find that the difference in amplitude is essentially a factor of order unity, leading to a negligible increase in the effective $k_{\parallel}^{\text{min}}$. This calculation confirms our argument for HERA-scale arrays.

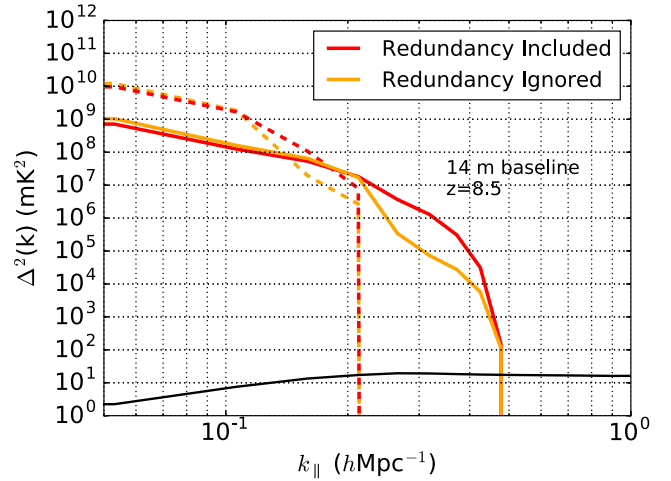


Figure A1. We compare the amplitude of modelling noise on a short baseline when \mathbf{R} is assumed to be diagonal (orange line) and the off-diagonal terms of \mathbf{R} are explicitly included for a 91-element (dashed lines) and 331-element (solid lines) hexagonally packed array of 14 m apertures. We find that even in a maximally redundant array, the effect of redundancy is to change the overall amplitude of the modelling noise by a factor of a few. This only has a small impact ($\lesssim 10$ per cent) on $k_{\parallel}^{\text{min}}$, the smallest k_{\parallel} where the 21 cm signal (black line) dominates over the modelling noise, as computed from equation (51) that ignores the effect of redundancy.

APPENDIX B: THE POINT SOURCE APPROXIMATION FOR MODELLED FOREGROUNDS

For analytic tractability, we assumed that the modelled component of foregrounds was well characterized by a flat-spectrum point source at zenith, whose visibilities are achromatic. Throughout the paper, the rest of the unmodelled foregrounds considered in our calculations were not assumed to be a single point source and are characterized by chromatic visibilities (see Sections 2.1.1 and 2.1.2). In this appendix, we explore the consequences of relaxing this assumption.

A significant consequence of the foregrounds not being dominated by a single point source at the phase centre is that for a fixed frequency, y_{α} 's amplitude will vary significantly from baseline to baseline, often approaching zero where source fringes destructively interfere. As a result, $\text{Cov}[\frac{c_{\alpha}}{y_{\alpha}}, \frac{c_{\alpha}^*}{y_{\alpha}^*}]$ can vary rapidly in frequency where y_{α} approaches zero. Thus, any weighting scheme that does not take these nulls into account will experience calibration error chromaticity in large excess of what we have found so far.

Instead, it is typical for calibration solutions to be obtained for each frequency through inverse-covariance weighting. Since the thermal noise covariance matrix is usually proportional to the identity, per-frequency inverse-covariance weights are proportional to $|y_{\alpha}|^2$. Under this scheme, we may employ a weights matrix that is frequency dependent,

$$\mathbf{W}_{\alpha\alpha} \rightarrow \mathbf{W}'_{\alpha\alpha} = \mathbf{W}_{\alpha\alpha} |y_{\alpha}(\nu)|^2, \quad (\text{B1})$$

which leads Λ and Ψ to be frequency dependent as well and we can no longer separate them from $\tilde{\mathbf{C}}$ in the delay transform. The delay-transformed visibility in equation (25) becomes

$$\tilde{V}_{\alpha} \approx \int d\nu e^{2\pi i \nu \tau} \left[y_{\alpha} \left(\tilde{\eta}'_i + \tilde{\eta}'_j + i\tilde{\phi}'_i - i\tilde{\phi}'_j \right) + c_{\alpha} \right], \quad (\text{B2})$$

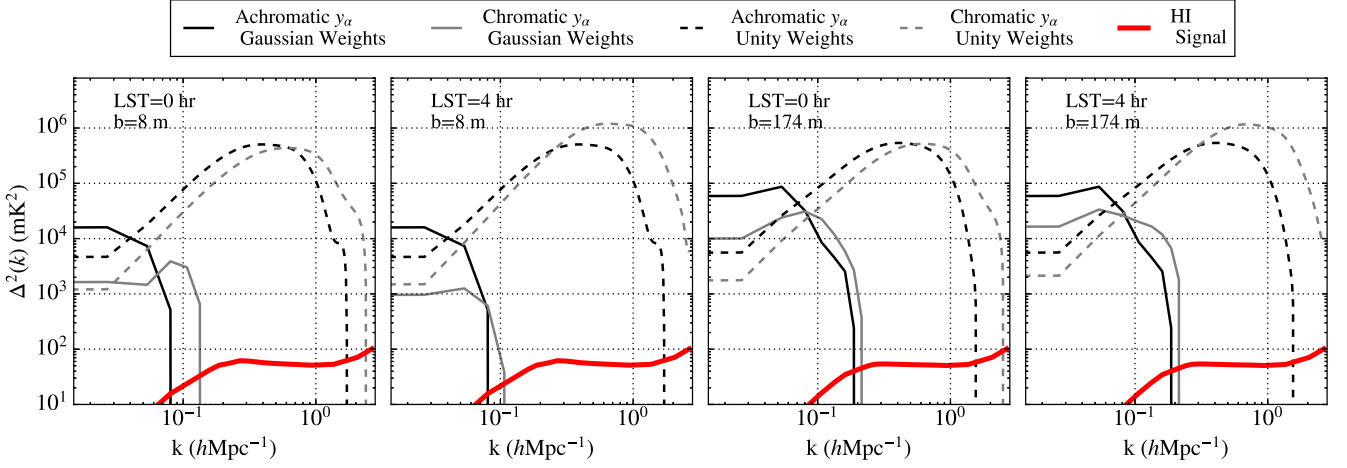


Figure B1. Comparisons between calibration modelling noise with realistic modelled foregrounds (grey lines) and the point source foregrounds used throughout this paper (black lines) with (solid) and without (dashed) Gaussian baseline weighting for two different LSTs and baseline lengths on the MWA from 10 MHz noise equivalent bandwidth centred at 150 MHz. The red line denotes the amplitude of the H I power spectrum generated with 21CMFAST. We find that fully modelled foregrounds change the overall amplitude of the calibration noise since the amplitude of a particular modelled visibility does not necessarily equal the amplitudes of every other modelled visibility. Chromatic y_α s also introduce some additional spectral structure that results in a larger width of calibration errors in k_\parallel . The overall impact on the LoS mode where modelling noise bias falls below the 21 cm signal is only of the order of 10 per cent with Gaussian weighting.

where every term, including y_α , is a function of frequency. The expectation value for the delay-transformed product of \tilde{V}_α with its complex conjugate to second order in c/y (equation 28) is now

$$\begin{aligned}
 \mathbf{P}_{\alpha\beta} = & \int dv dv' e^{2\pi i \tau(v-v')} [y_\alpha y_\beta^* \langle \eta'_i \eta_\ell'^* \rangle + y_\alpha y_\beta^* \langle \eta'_j \eta_m'^* \rangle \\
 & + y_\alpha y_\beta^* \langle \eta'_j \eta_\ell'^* \rangle + y_\alpha y_\beta^* \langle \eta'_i \eta_m'^* \rangle \\
 & + y_\alpha y_\beta^* \langle \phi'_i \phi_\ell'^* \rangle + y_\alpha y_\beta^* \langle \phi'_j \phi_m'^* \rangle \\
 & - i y_\alpha y_\beta^* \langle \phi'_i \phi_m'^* \rangle - i y_\alpha y_\beta^* \langle \phi'_j \phi_\ell'^* \rangle \\
 & + y_\alpha \langle \eta'_i c_\beta^* \rangle + y_\alpha \langle \eta'_j c_\beta^* \rangle \\
 & + i y_\alpha \langle \phi'_i c_\beta^* \rangle - i y_\alpha \langle \phi'_j c_\beta^* \rangle \\
 & + y_\beta^* \langle \eta_\ell^* c_\alpha \rangle + y_\beta^* \langle \eta_m^* c_\alpha \rangle \\
 & - i y_\beta^* \langle \phi_\ell^* c_\alpha \rangle + i y_\beta^* \langle \phi_m^* c_\alpha \rangle \\
 & + \langle c_\alpha c_\beta^* \rangle], \tag{B3}
 \end{aligned}$$

where every complex conjugated quantity is a function of ν' and every non-conjugated quantity is a function of ν . Since the weight and design matrices are no longer frequency independent, second-order moments cannot be separated into frequency-dependent and frequency-independent components as we did with the point source approximation. In order to compute $\mathbf{P}_{\alpha\beta}$, we must calculate all second-order moments with a given source model and design matrix and take the Fourier transforms.

For realistic y_α , we use simulations of point source foregrounds obtained from the PRISM software package (Thyagarajan et al. 2015) for the MWA-128T array layout with antennas modelled as 4 m diameter dishes. For each 100 kHz channel over a 20 MHz band, we use a weights matrix $W'_{\alpha\alpha}(\nu) = W_{\alpha\alpha} |y_\alpha(\nu)|^2$ and compute the two-dimensional Fourier transform in equation (B3) to obtain $\mathbf{P}_{\alpha\alpha}$ for several baselines. We take the MWA to be pointing at a declination equal to its latitude of $-26^\circ 701$ (Tingay et al. 2013) at LST=0 and 4 h. We run two different simulations, one in which $\mathbf{W}_{\alpha\alpha}$ is set to unity (and the weights matrix $\mathbf{W}'_{\alpha\alpha} = \mathbf{W}_{\alpha\alpha} |y_\alpha|^2$) and the other where $\mathbf{W}_{\alpha\alpha}$ is given by equation (52) with $\sigma_w = 50$ m.

In Fig. B1, we show the amplitude of calibration modelling noise on a short (8 m) and long (174 m) MWA baseline with and without

fully modelled foregrounds for several different LSTs. In all cases, we see that the fully modelled foregrounds extend the width of the foreground noise to larger k_\parallel , something we would expect to occur with the additional spectral structure they introduce. In addition, the amplitude of the modelling noise is modified since the multiplication by the modelled foregrounds on a particular baseline, $y_\alpha y_\beta^*$ (equation B3), does not necessarily cancel out the modelled foregrounds in the numerator of each summed $\langle c_\gamma c_\delta^* \rangle / (y_\gamma y_\delta^*)$ as they do when y_α is constant. Despite these differences, we find that over the range of LSTs and baselines studied, the overall impact on the minimal LoS wavenumber of modes that can be observed at a particular uv point is only of the order of ≈ 10 per cent. Thus, the approximation of the modelled foregrounds as a point source at the phase centre does not have a significant impact on a range of modes that are masked by foreground modelling errors.

APPENDIX C: EXPRESSIONS FOR SECOND MOMENTS OF DELAY-TRANSFORMED CALIBRATION ERRORS

In this appendix, we derive the approximate expressions for the second moments that we use to go from equation (28) to equation (34). To derive equations (29) through (33), we first note that

$$\begin{aligned}
 \text{Cov}[\text{Re}(\mathbf{c}), \text{Re}(\mathbf{c})^\top] & \approx \frac{1}{2} \mathbf{C}, \\
 \text{Cov}[\text{Im}(\mathbf{c}), \text{Im}(\mathbf{c})^\top] & \approx \frac{1}{2} \mathbf{C}. \tag{C1}
 \end{aligned}$$

This assertion is true for thermal noise, \mathbf{n} , since both the real and imaginary components of the thermal noise are given by identical, zero-mean normal distributions. We need to only show that this assertion holds for the unmodelled foregrounds \mathbf{r} . We start by writing

$$\begin{aligned}
 & \text{Cov}[\text{Re}(\mathbf{r}), \text{Re}(\mathbf{r})^\top]_{\alpha\beta} \\
 & \propto \int d\Omega \cos\left(\frac{2\pi\nu \mathbf{b}_\alpha \cdot \hat{\delta}}{c}\right) \cos\left(\frac{2\pi\nu' \mathbf{b}_\beta \cdot \hat{\delta}}{c}\right) |A(\hat{\delta})|^2
 \end{aligned}$$

$$= \frac{1}{4} \int d\Omega [e^{2\pi i(v\mathbf{b}_\alpha + v'\mathbf{b}_\beta)\cdot\hat{s}/c} + e^{-2\pi i(v\mathbf{b}_\alpha + v'\mathbf{b}_\beta)\cdot\hat{s}/c}] A(\hat{s})]^2 \\ + \frac{1}{4} \int d\Omega [e^{2\pi i(v\mathbf{b}_\alpha - v'\mathbf{b}_\beta)\cdot\hat{s}/c} + e^{-2\pi i(v\mathbf{b}_\alpha - v'\mathbf{b}_\beta)\cdot\hat{s}/c}] A(\hat{s})]^2, \quad (\text{C2})$$

where we dropped the multiplicative instrument-independent terms in the covariance (equations 5 and 6) in favour of a proportionality sign. All of the terms in equation (C2) will integrate to zero unless $|\mathbf{b}_\alpha \pm \mathbf{b}_\beta| \lesssim d_{\text{ant}}$ (less than one fringe fits within the primary beam main lobe), which is only true if $\mathbf{b}_\alpha \approx \pm \mathbf{b}_\beta$, where the negative case causes exponential terms in the first line of equation (C2) to be non-zero and the positive case causes the second line to be non-zero. We may choose baseline indexing such that we never have $\mathbf{b}_\alpha \approx -\mathbf{b}_\beta$, by having antenna numbers increase with increasing E-W and than N-S position. With this indexing,

$$\text{Cov}[\text{Re}(r_\alpha), \text{Re}(r_\beta)] \\ \propto \frac{1}{4} \int d\Omega [e^{2\pi i(v\mathbf{b}_\alpha - v'\mathbf{b}_\beta)\cdot\hat{s}/c} + e^{-2\pi i(v\mathbf{b}_\alpha - v'\mathbf{b}_\beta)\cdot\hat{s}/c}] A(\hat{s})]^2 \\ = \frac{1}{4} [\text{Cov}(r_\alpha, r_\beta^*) + \text{Cov}(r_\alpha^*, r_\beta)]. \quad (\text{C3})$$

For beams that are symmetric around the phase centre, $\text{Cov}[r_\alpha, r_\beta^*]$ is real and $\text{Cov}[r_\alpha, r_\beta^*] = \text{Cov}[r_\alpha^*, r_\beta]$, proving our assertion that

$$\text{Cov}[\text{Re}(r_\alpha), \text{Re}(r_\beta)] = \frac{1}{2} \mathbf{R}_{\alpha\beta}. \quad (\text{C4})$$

A very similar set of steps with identical assumptions yields

$$\text{Cov}[\text{Im}(r_\alpha), \text{Im}(r_\beta)] = \frac{1}{2} \mathbf{R}_{\alpha\beta}. \quad (\text{C5})$$

Next, we show that

$$\mathbf{C}_{\alpha\beta} \gg \langle c_\alpha \rangle \langle c_\beta^* \rangle. \quad (\text{C6})$$

This can be seen by writing the product of the averages

$$\langle c_\alpha \rangle \langle c_\beta^* \rangle = \int d\Omega e^{-2\pi i v \mathbf{b}_\alpha \cdot \hat{s} / c} A(\hat{s}) \langle I(\hat{s}) \rangle \\ \times \int d\Omega' e^{2\pi i v' \mathbf{b}_\beta \cdot \hat{s}' / c} A(\hat{s}') \langle I(\hat{s}') \rangle \quad (\text{C7})$$

both integrate to zero when $b_{\alpha/\beta} \gtrsim d_{\text{ant}}$ and $\langle I(\hat{s}) \rangle$ is smooth as a function of position (which is typically true of foreground residuals and signal).

We can now show derivations for equations (29) through (33).

C1 Derivation of equation (29)

We start on the left-hand side with

$$\langle \tilde{\eta}_i \tilde{\eta}_j^* \rangle = \int dv dv' e^{-2\pi i \tau (v-v')\tau} [\langle \mathbf{C}_\eta \rangle_{ij} + \langle \eta_i \rangle \langle \eta_j^* \rangle] \\ \approx S_0^{-2} \Lambda_{i\gamma} \Lambda_{\delta j}^\top \int dv dv' e^{-2\pi i \tau (v-v')\tau} \text{Cov}[\text{Re}(c^\gamma), \text{Re}(c^\delta)] \\ = \frac{S_0^{-2}}{2} \Lambda_{i\gamma} \Lambda_{\delta j}^\top \int dv dv' e^{-2\pi i \tau (v-v')\tau} \mathbf{C}^{\gamma\delta}(v, v') \\ = \frac{S_0^{-2}}{2} \Lambda_{i\gamma} \Lambda_{\delta j}^\top \tilde{\mathbf{C}}^{\gamma\delta}. \quad (\text{C8})$$

In going from the first to the second line, we threw away the product of the means (equation C6). Going from the second to the third line, we used equation (C4).

C2 Derivation of equation (30)

Following the same procedure for equation (29),

$$\langle \phi_i \tilde{\phi}_j^* \rangle = \int dv dv' e^{-2\pi i \tau (v-v')\tau} [\langle \mathbf{C}_\phi \rangle_{ij} + \langle \phi_i \rangle \langle \phi_j^* \rangle] \\ \approx S_0^{-2} \Psi_{i\gamma} \Psi_{\delta j}^\top \int dv dv' e^{-2\pi i \tau (v-v')\tau} \text{Cov}[\text{Im}(c^\gamma), \text{Im}(c^\delta)] \\ = \frac{S_0^{-2}}{2} \Psi_{i\gamma} \Psi_{\delta j}^\top \int dv dv' e^{-2\pi i \tau (v-v')\tau} \mathbf{C}^{\gamma\delta}(v, v') \\ = \frac{S_0^{-2}}{2} \Psi_{i\gamma} \Psi_{\delta j}^\top \tilde{\mathbf{C}}^{\gamma\delta}. \quad (\text{C9})$$

C3 Derivation of equation (31)

Starting with the left-hand side of equation (31),

$$\langle \tilde{c}_\alpha \tilde{\eta}_i^* \rangle = \frac{S_0^{-1}}{2} \int dv dv' e^{-2\pi i (v-v')\tau} \Lambda_{i\gamma} \langle c_\alpha (c^\gamma + c^{\gamma*}) \rangle \\ \approx \frac{S_0^{-1}}{2} \Lambda_{i\gamma} \int dv dv' e^{-2\pi i (v-v')\tau} \mathbf{C}_\alpha^\gamma(v, v') \\ = \frac{S_0^{-1}}{2} \Lambda_{i\gamma} \tilde{\mathbf{C}}_\alpha^\gamma. \quad (\text{C10})$$

To go from the first to second line here, we used the fact that $\langle c_\alpha c^\gamma \rangle \propto \int d\Omega e^{-2\pi i \hat{s} \cdot (\mathbf{b}_\alpha v + \mathbf{b}_\beta v') / c} |A(\hat{s})|^2$, which, as discussed above, integrates to zero for $b_{\alpha/\beta} \gtrsim d_{\text{ant}}$.

C4 Derivation of equation (32)

Following the same steps used for equation (31),

$$\langle \tilde{c}_\alpha \tilde{\phi}_i^* \rangle = \frac{S_0^{-1}}{2i} \int dv dv' e^{-2\pi i (v-v')\tau} \Psi_{i\gamma} \langle c_\alpha (c^\gamma - c^{\gamma*}) \rangle \\ \approx \frac{i S_0^{-1}}{2} \Psi_{i\gamma} \int dv dv' e^{-2\pi i (v-v')\tau} \mathbf{C}_\alpha^\gamma(v, v') \\ = S_0^{-1} \frac{i}{2} \Psi_{i\gamma} \tilde{\mathbf{C}}_\alpha^\gamma. \quad (\text{C11})$$

C5 Derivation of equation (33)

We may show this last identity by expanding the real and imaginary components of c ,

$$\langle \tilde{\eta}_i \tilde{\phi}_j^* \rangle = S_0^{-2} \int dv dv' e^{2\pi i (v-v')\tau} \Lambda_{i\gamma} \Psi_{\delta j}^\top (\text{Re}(c^\gamma) \text{Im}(c^\delta)^*) \\ = \frac{S_0^{-2}}{4i} \int dv dv' e^{2\pi i (v-v')\tau} \Lambda_{i\gamma} \Psi_{\delta j}^\top \langle (c^\gamma + c^{\gamma*})(c^\delta - c^{\delta*}) \rangle \\ \approx 0. \quad (\text{C12})$$

We obtain the last line approximately equal to zero due to the fact that $\langle (c^\gamma + c^{\gamma*})(c^\delta - c^{\delta*}) \rangle = \langle c^\gamma c^\delta \rangle + \langle c^{\gamma*} c^{\delta*} \rangle + \langle c^\gamma c^{\delta*} \rangle - \langle c^{\gamma*} c^\delta \rangle$. The first two terms evaluate to zero since they involve integrals over $e^{\pm 2\pi i \hat{s} \cdot (\mathbf{b}_\gamma v + \mathbf{b}_\delta v')}$ and the last two terms are equal to each other so they subtract to give 0.

APPENDIX D: COMPONENTS OF Λ AND Ψ FOR NON-REDUNDANT, UNIFORMLY WEIGHTED CALIBRATION SOLUTIONS

In this appendix, we derive equations (36) and (37) that are valid when the weights matrix is equal to unity. While of limited applicability, they provide us with insight into the scaling of modelling

noise with properties of the source catalogue and array and allow us to identify the degree to which any visibility covariance contributes to the covariances of gain solutions.

D1 Equation (36)

We wish to evaluate

$$\Lambda_{i\gamma} = [(\mathbf{AA}^\top)^{-1}\mathbf{A}^\top]_{i\gamma}. \quad (\text{D1})$$

We start with $(\mathbf{AA}^\top)_{ij}$. Evaluating this matrix product for a non-redundant array is straightforward since each element is given by the dot product of the i th column of \mathbf{A} with the j th column. Since a given column is equal to unity at the indices of visibilities in which that antenna participates and zero otherwise, the dot product of columns is equal to $N_{\text{ant}} - 1$ if $i = j$ and equal to unity if $i \neq j$,

$$\mathbf{AA}^\top = \begin{pmatrix} N_{\text{ant}} - 1 & 1 & 1 & \dots & 1 \\ 1 & N_{\text{ant}} - 1 & 1 & \dots & 1 \\ 1 & 1 & N_{\text{ant}} - 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & N_{\text{ant}} - 1 \end{pmatrix}. \quad (\text{D2})$$

\mathbf{AA}^\top can be decomposed into the sum of a diagonal matrix and a matrix formed from an outer product,

$$\mathbf{AA}^\top = \begin{pmatrix} N_{\text{ant}} - 2 & 0 & 0 & \dots & 0 \\ 0 & N_{\text{ant}} - 2 & 0 & \dots & 0 \\ 0 & 0 & N_{\text{ant}} - 2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & N_{\text{ant}} - 2 \end{pmatrix} + \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix}, \quad (\text{D3})$$

and thus can be inverted using the Sherman–Morrison formula,

$$(\mathbf{AA}^\top)^{-1}_{ij} = \begin{cases} \frac{1}{N_{\text{ant}} - 2} \left[1 - \frac{1}{2(N_{\text{ant}} - 1)} \right] & i = j \\ \frac{-1}{2(N_{\text{ant}} - 1)(N_{\text{ant}} - 2)} & i \neq j. \end{cases} \quad (\text{D4})$$

We can now evaluate $(\mathbf{AA}^\top)^{-1}_{ij}(\mathbf{A}^\top)^j_\gamma$, which is the sum of the entries in the i th row of $(\mathbf{AA}^\top)^{-1}$ that correspond to antennas that participate in the γ th baseline. If $i \in \gamma$, we add an entry in $(\mathbf{AA}^\top)^{-1}$ where $i = j$ to an entry where $i \neq j$. If $i \notin \gamma$, we add two entries in $(\mathbf{AA}^\top)^{-1}$ where $i \neq j$. For these two cases, we get

$$\Lambda_{i\gamma} = [(\mathbf{AA}^\top)^{-1}\mathbf{A}^\top]_{i\gamma} = \begin{cases} \frac{1}{N_{\text{ant}} - 1} & i \in \gamma \\ \frac{-1}{(N_{\text{ant}} - 1)(N_{\text{ant}} - 2)} & i \notin \gamma, \end{cases} \quad (\text{D5})$$

which completes the proof.

D2 Equation (37)

We begin evaluating $\Psi_{i\gamma} = (\mathbf{BB}^\top)^{-1}\mathbf{B}^\top$ with the product $(\mathbf{BB}^\top)_{ij}$, which is the dot product of the i th column of \mathbf{B} with the j th column. Each i th column contains $N_{\text{ant}} - 1$ non-zero rows that are 1 when the i th antenna is the non-conjugated participant in the baseline and -1 when the antenna is the conjugated participant. The last row of \mathbf{B} is composed entirely of ones. Thus, the dot product of any column with itself is N_{ant} and the dot product of a column with any other is equal to zero. Thus,

$$(\mathbf{BB}^\top)_{ij} = \begin{cases} N_{\text{ant}} & i = j \\ 0 & i \neq j, \end{cases} \quad (\text{D6})$$

whose inverse is trivial. $(\mathbf{BB}^\top)^{-1}(\mathbf{B}^\top)^j_\gamma$ is the sum of each element of the i th row of $(\mathbf{BB}^\top)^{-1}$ that participates in the γ th visibility. Since $(\mathbf{BB}^\top)^{-1}$ is diagonal, this sum is only non-zero when $i = j$. If i is the non-conjugated antenna in the visibility, then \mathbf{B}^\top multiplies by 1 and if i is the conjugated antenna in γ , \mathbf{B}^\top multiplies by -1 . We obtain

$$\Psi_{i\gamma} = \begin{cases} \frac{1}{N_{\text{ant}}} & \gamma = (i, \cdot) \\ -\frac{1}{N_{\text{ant}}} & \gamma = (\cdot, i) \\ 0 & i \notin \gamma \end{cases}, \quad (\text{D7})$$

completing our proof.

APPENDIX E: A SIMPLIFIED EXPRESSION FOR MINIMAL ACCESSIBLE LOS MODES

In this appendix, we derive equation (51) from equation (42) with the additional assumptions that the array has a Gaussian beam with standard deviation $\sigma_b = \epsilon \lambda_0 / d_{\text{ant}}$ and that its antennas are arranged in a Gaussian configuration with standard deviation σ_{ant} . We also ignore the contribution from thermal noise, which integrates down with time, and assume that $\mathbf{C} = \mathbf{R}$. We first compute $\tilde{\mathbf{C}}_{\delta\delta}(\tau, \tau)$ for a Gaussian beam

$$\begin{aligned} \tilde{\mathbf{C}}_{\delta\delta}(\tau, \tau) &= \sigma_r^2 \int d\nu d\nu' \int d\Omega e^{2\pi i \tau (v - v')} e^{-2\pi i b_\delta \cdot \hat{s} (v - v') / c} |A(\hat{s})|^2 \\ &\approx \sigma_r^2 \int d\nu d\nu' \int d\Omega e^{2\pi i \tau (v - v')} e^{-2\pi i b_\delta \cdot \hat{s} (v - v') / c} e^{-s^2 / 2\sigma_b^2} \\ &\approx \sigma_r^2 B \int d\Delta v \int d\Omega e^{2\pi i \tau \Delta v} e^{-2\pi i b_\delta \cdot \hat{s} \Delta v / c} e^{-s^2 / \sigma_b^2} \\ &\approx \sigma_r^2 B \frac{c}{b_\delta} \sqrt{\pi} \sigma_b \exp\left(-\frac{c^2 \tau^2}{b_\delta^2 \sigma_b^2}\right). \end{aligned} \quad (\text{E1})$$

To derive the last line, we used the flat-sky approximation, letting the angular integral run over infinity. We also approximate the bandwidth as infinite. Thus,

$$(\tilde{\mathbf{C}}_{\delta\delta})_{i \in \delta} \approx \sigma_r^2 c B \sqrt{\pi} \sigma_b \left\langle \frac{1}{b_\delta} \exp\left(-\frac{c^2 \tau^2}{b_\delta^2 \sigma_b^2}\right) \right\rangle_{i \in \delta}. \quad (\text{E2})$$

Since the chromaticity increases monotonically with increasing baseline length and the antennas with the largest numbers of short baselines are at the centre of the array, the minimal $k_{\parallel}^{\text{min}}$ accessible by an interferometer will occur on a short baseline formed from two antennas near the core of the array. With the core antenna positions equal to $r_i \approx 0$ so that $b_\delta = |\mathbf{r}_i - \mathbf{r}_k| \approx r_k$, the average of a function

of the length of baselines that a core antenna participates in is equal to the average of that function over antenna positions

$$\langle \tilde{\mathbf{C}}_{\delta\delta} \rangle_{i \in \delta} \approx \sigma_r^2 c B \sqrt{\pi} \sigma_b \left\langle \frac{1}{r_k} \exp \left(-\frac{c^2 \tau^2}{r_k^2 \sigma_b^2} \right) \right\rangle_{r_k}. \quad (\text{E3})$$

We can compute this average analytically if the antennas are distributed as a Gaussian with standard deviation σ_{ant} ,

$$\begin{aligned} \left\langle \frac{1}{r_k} \exp \left(-\frac{c^2 \tau^2}{r_k^2 \sigma_b^2} \right) \right\rangle_{r_k} &= \frac{1}{2\pi\sigma_{\text{ant}}^2} \int d^2 r_k r_k^{-1} e^{-\frac{c^2 \tau^2}{r_k^2 \sigma_b^2} - \frac{r_k^2}{2\sigma_{\text{ant}}^2}} \\ &= \sqrt{\frac{\pi}{2}} \frac{1}{\sigma_{\text{ant}}} \exp \left(-\frac{\sqrt{2} c \tau}{\sigma_b \sigma_{\text{ant}}} \right). \end{aligned} \quad (\text{E4})$$

It follows that for i and j antennas close to the core, the averages over baselines evaluate to

$$\langle \tilde{\mathbf{C}}_{\delta\delta} \rangle_{i \in \delta} \approx \langle \tilde{\mathbf{C}}_{\delta\delta} \rangle_{j \in \delta} \approx \sigma_r^2 c B \frac{\pi}{\sqrt{2}} \frac{\sigma_b}{\sigma_{\text{ant}}} \exp \left(-\frac{\sqrt{2} c \tau}{\sigma_b \sigma_{\text{ant}}} \right). \quad (\text{E5})$$

Thus, for two core antennas with Gaussian beams in a Gaussian antenna distribution, the contamination from calibration errors in equation (42) reduces to

$$\mathbf{P}_{\alpha\alpha} - \tilde{\mathbf{C}}_{\alpha\alpha} \approx 2\sigma_r^2 B \frac{c}{N_{\text{ant}}} \frac{\pi}{\sqrt{2}} \frac{\sigma_b}{\sigma_{\text{ant}}} \exp \left(-\frac{\sqrt{2} c \tau}{\sigma_b \sigma_{\text{ant}}} \right). \quad (\text{E6})$$

The minimal delay where the signal can be measured, τ_{min} , is set by where the calibration noise passes below the signal. Thus, we obtain τ_{min} by setting equation (E6), multiplied by the prefactors in equation (26) that convert from $\text{Jy}^2 \text{Hz}^2$ to $\text{mK}^2 h^{-3} \text{Mpc}^3$ equal to the 21 cm power spectrum,

$$\begin{aligned} P_{21} &\approx \left(\frac{\lambda_0^2}{2k_{\text{B}}} \right)^2 \frac{X^2 Y}{B_{\text{pp}} \Omega_{\text{pp}}} \left(\mathbf{P}_{\alpha\alpha}(\tau_{\text{min}}, \tau_{\text{min}}) - \tilde{\mathbf{C}}_{\alpha\alpha}(\tau_{\text{min}}, \tau_{\text{min}}) \right) \\ &\approx \left(\frac{\lambda_0^2}{2k_{\text{B}}} \right)^2 \frac{X^2 Y}{B_{\text{pp}} \Omega_{\text{pp}}} \sqrt{2} \pi \sigma_r^2 B \frac{c}{N_{\text{ant}}} \frac{\sigma_b}{\sigma_{\text{ant}}} \exp \left(-\frac{\sqrt{2} c \tau_{\text{min}}}{\sigma_b \sigma_{\text{ant}}} \right), \end{aligned} \quad (\text{E7})$$

and invert it,

$$\begin{aligned} \tau_{\text{min}} &\approx \frac{\sigma_b \sigma_{\text{ant}}}{\sqrt{2} c} \log \left(\frac{\lambda_0^4 X^2 Y \sqrt{2} \pi \sigma_r^2 c \sigma_b}{4k_{\text{B}}^2 P_{21} \Omega_{\text{pp}} B_{\text{pp}} N_{\text{ant}} \sigma_{\text{ant}}} \right) \\ &= \frac{\epsilon}{\sqrt{2} \nu_0} \frac{\sigma_{\text{ant}}}{d_{\text{ant}}} \log \left(\frac{\lambda_0^4 X^2 Y \sigma_r^2 B \nu_0}{B_{\text{pp}} 2 \sqrt{2} k_{\text{B}}^2 N_{\text{ant}} \epsilon P_{21} \sigma_{\text{ant}}} \right). \end{aligned} \quad (\text{E8})$$

Using the fact that $k_{\parallel}^{\text{min}} = 2\pi\tau_{\text{min}}/Y$, we arrive at equation (51),

$$k_{\parallel}^{\text{min}} = \frac{\epsilon \sqrt{2} \pi}{Y \nu_0} \frac{\sigma_{\text{ant}}}{d_{\text{ant}}} \log \left(\frac{\lambda_0^4 X^2 Y \sigma_r^2 B \nu_0}{B_{\text{pp}} 2 \sqrt{2} k_{\text{B}}^2 N_{\text{ant}} \epsilon P_{21} \sigma_{\text{ant}}} \right). \quad (\text{E9})$$

Checking this approximate expression against direct calculation for arrays with Gaussian beams using equation (34) yields an accuracy of ≈ 10 per cent, even in arrays that are not strictly Gaussian such as the MWA, LOFAR and HERA where σ_{ant} is the standard deviation of the non-Gaussian antenna distribution.

The primary shortcoming of equation (E9) is that it assumes a Gaussian primary beam that only accounts for the delay at which the contamination from the main lobe falls beneath the signal. Since side lobes can easily enter at the $\gtrsim 5$ per cent level, it is possible for them to contaminate the EoR window at much larger k_{\parallel} than the $k_{\parallel}^{\text{min}}$ predicted in equation (E9). While the contribution of side lobes for different baselines will fall at different delays and will not add coherently when averaging over the antenna distribution, we can assume that they add directly to obtain a conservative upper bound on when their contribution will affect $k_{\parallel}^{\text{min}}$. If the side lobes added directly in the antenna average, then their contribution to the amplitude of foreground residuals would be of the order of f_{sl}^2 the level of the foreground residuals at zero delay, where f_{sl} is the ratio between the gain of the side lobe and the gain at boresight. A conservative estimate of when side lobes are at the level of the 21 cm signal can be obtained by setting the right-hand side of equation (E7) at zero delay multiplied f_{sl}^2 equal to the 21 cm power spectrum,

$$\begin{aligned} P_{21} &\approx f_{\text{sl}}^2 \left(\frac{\lambda_0^2}{2k_{\text{B}}} \right)^2 \frac{X^2 Y}{B_{\text{pp}} \Omega_{\text{pp}}} \sqrt{2} \pi \sigma_r^2 B \frac{c}{N_{\text{ant}}} \frac{\sigma_b}{\sigma_{\text{ant}}} \\ &\approx f_{\text{sl}}^2 \left(\frac{\lambda_0^2}{2k_{\text{B}}} \right)^2 \frac{X^2 Y}{B_{\text{pp}} \pi} \sqrt{2} \pi \sigma_r^2 B \frac{c}{N_{\text{ant}}} \frac{d_{\text{ant}}}{\epsilon \sigma_{\text{ant}} \lambda_0}. \end{aligned} \quad (\text{E10})$$

We use this condition to denote the white dashed region of parameter space in Fig. 3 where side lobes may render the predictions of equation (51) inaccurate.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.