A Bayesian approach to high fidelity interferometric calibration II: demonstration with simulated data

Peter H. Sims,[1,2]⋆ Jonathan C. Pober,[3] and Jonathan L. Sievers[1,2]

[1]*McGill Space Institute, McGill University, 3550 University Street, Montreal, QC H3A 2A7, Canada*
[2]*Department of Physics, McGill University, 3600 University Street, Montreal, QC H3A 2T8, Canada*
[3]*Department of Physics, Brown University, Providence, RI 02912, USA*

29th June 2022

ABSTRACT

In a companion paper, we presented BayesCal, a mathematical formalism for mitigating sky-model incompleteness in interferometric calibration. In this paper, we demonstrate the use of BayesCal to calibrate the degenerate gain parameters of full-Stokes simulated observations with a HERA-like hexagonal close-packed redundant array, for three assumed levels of completeness of the a priori known component of the calibration sky model. We compare the BayesCal calibration solutions to those recovered by calibrating the degenerate gain parameters with only the a priori known component of the calibration sky model both with and without imposing physically motivated priors on the gain amplitude solutions and for two choices of baseline length range over which to calibrate. We find that BayesCal provides calibration solutions with up to four orders of magnitude lower power in spurious gain amplitude fluctuations than the calibration solutions derived for the same data set with the alternate approaches, and between $\sim 10^7$ and $\sim 10^{10}$ times smaller than in the mean degenerate gain amplitude on the full range of spectral scales accessible in the data. Additionally, we find that in the scenarios modelled only BayesCal has sufficiently high fidelity calibration solutions for unbiased recovery of the 21 cm power spectrum on large spectral scales ($k_\parallel \lesssim 0.15 \, h\mathrm{Mpc}^{-1}$). In all other cases, in the completeness regimes studied, those scales are contaminated.

Key words: methods: data analysis – methods: statistical – dark ages, reionization, first stars – cosmology: observations

1 INTRODUCTION

High fidelity radio interferometric data calibration that minimises spurious spectral structure in the calibrated data is essential in astrophysical applications, such as 21 cm cosmology, which rely on knowledge of the relative spectral smoothness of distinct astrophysical emission components to separate the signal of interest from foreground emission (see e.g. Liu & Shaw 2020 for a review of signal separation techniques in 21 cm data analysis).

Existing approaches to determining the antenna dependent complex gains during interferometric calibration rely on either (i) fitting a forward model of the sky to the data, or (ii) the presence of redundancy in the baseline layout of the array to solve for the majority of the calibration degrees of freedom, followed by fitting a forward model of the sky to the data to solve for the remaining calibration terms. However, both approaches produce biased results if the sky model used to calibrate the data is incomplete (e.g. Barry, et al. 2016 and Ewall-Wice, et al. 2017, and Byrne, et al. 2019, respectively) and, without mitigation, sky-model-incompleteness-

sourced biases in the gain parameters are sufficiently large to produce calibration-error-induced foreground systematics at a level that can overwhelm the high-redshift 21 cm signal and preclude unbiased recovery of modes of the 21 cm power spectrum.

In a companion paper (hereafter, paper I), we present BayesCal, a Bayesian framework for interferometric calibration designed to recover high fidelity gain solutions by mitigating sky-model incompleteness. BayesCal addresses the sky-model incompleteness problem by supplementing the a priori known component of the forward model of the sky with a statistical model for the a priori unknown component of the sky brightness distribution associated with (i) sources below the completeness level of catalogues in the field of interest and (ii) with the residual diffuse emission deriving from the difference between the true brightness distribution of diffuse emission on the sky and our model for it.

The BayesCal framework presented in Paper I addresses the calibration of visibilities corresponding to a single instrumental polarisation, deriving from full-Stokes sky emission, while accounting for the dominant component of sky-model incompleteness associated with missing and mis-modelled emission in Stokes *I*. The BayesCal calibration model is comprised of (i) a simulated visib-

---

⋆ E-mail: peter.sims@mail.mcgill.ca

ility model, V^{sim}, which describes the expected contribution to the observed visibilities associated with a priori known sky emission (this model component is standard in sky-referenced calibration) and (ii) a fitted visibility model, V^{fit}, that models the contribution to the observed visibilities from emission not included in V^{sim} due to the incompleteness and uncertainties associated with our a priori knowledge of the brightness distribution of the sky. When calibrating data with BAYESCAL, one jointly fits for instrumental gains, constrained by available priors on their spectral structure, and the most probable sky-model parameter values for V^{fit}, constrained by a prior on their total power.

In this paper, we demonstrate the extent to which the BAYES-CAL framework mitigates these sky-model-incompleteness-sourced biases in the gain parameters for varying levels of completeness of the simulated calibration sky model, in the context of absolute calibration of visibilities deriving from simulated interferometric observations, accounting for both diffuse and point source emission, with a HERA-like hexagonal close-packed redundant array. Additionally, we compare the fidelity of BAYESCAL calibration solutions to those recovered with two alternate absolute calibration algorithms:

(i) standard[1] absolute calibration with an incomplete sky model,

(ii) absolute calibration with an incomplete sky model when the spectral structure in the gain solutions is constrained by a physically motivated prior.

The first of these approaches provides a fiducial calibration fidelity that calibration solutions derived with BAYESCAL can be compared against. The second represents an intermediate step between standard absolute calibration and BAYESCAL and illustrates the limited efficacy of constraining the gain solutions with a physical prior[2] when using an incomplete calibration model.

Through this comparison, we will illustrate that BAYESCAL enables recovery of significantly higher fidelity calibration solutions, with in an improvement of 1–4 orders of magnitude additional suppression of power in spurious gain amplitude fluctuations in the calibrated data on 3, 4.5 and 9 MHz spectral scales relevant to constraining the astrophysics of CD and the EoR with 21 cm cosmology ($k_\parallel \simeq 0.17, 0.11$ and $0.06\ h\mathrm{Mpc}^{-1}$, respectively, for redshifted 21 cm emission in the frequency range of our simulated observed data) relative to the alternate calibration approaches.

In the context of 21 cm cosmology applications, we will show that, in the scenarios modelled, only BAYESCAL enables recovery of the 21 cm signal at $k_\parallel \lesssim 0.15\ h\mathrm{Mpc}^{-1}$. For all the other techniques in all completeness regimes, that scale is contaminated.

---

[1] Here, we define standard absolute calibration as absolute calibration of redundantly calibrated visibilities without imposing informative priors on the degenerate gain parameters (as considered in Byrne, et al. 2019). We defer to future work, a comparison to versions of absolute calibration with spectrally smooth gain solutions enforced through the gain parametrisation (e.g. Barry, et al. 2016) or filtering of the gain solutions (e.g. Kern et al. 2020).

[2] Here, by physical prior, we mean a soft prior with finite uncertainty. In contrast, applying a hard (zero-uncertainty) prior on the gain solutions, if accurate, can limit sky-model-incompleteness-sourced bias introduced into the calibration solutions on particular spectral scales (e.g. Barry, et al. 2016; Kern et al. 2020). However, this approach implies hard (zero-uncertainty) knowledge about structure in the gains which, if imperfect, will bias covariant propagation of uncertainties on the calibration parameters through to the calibrated data and create an additional avenue for introducing spurious spectral structure into the gain solutions, and correspondingly into the calibrated data, if the gains are overly constrained and are unable to fit real instrumental structure (see paper I for details).

A number of ongoing and upcoming interferometric 21 cm experiments have redundancy in their array configurations and, thus, can take advantage of the BAYESCAL approach we demonstrate here. These include the Canadian Hydrogen Intensity Mapping Experiment (CHIME; Newburgh, et al. 2014), the Hydrogen Epoch of Reionization Array (HERA; DeBoer et al. 2017), the Hydrogen Intensity and Real-time Analysis eXperiment (HIRAX; Newburgh et al. 2016), Phase II of the Murchison Widefield Array (MWA; Wayth et al. 2018) and Tianlai (Li et al. 2020; Wu et al. 2021). For the simulated observations calibrated in this paper we use a HERA-like hexagonal close-packed redundant array.

The generalisation of BAYESCAL to non-redundant and semi-redundant interferometric array configurations discussed in paper I is applicable to 21 cm experiments including the Giant Metrewave Radio Telescope (GMRT; Paciga et al. 2013), the LOw Frequency ARray (LOFAR; van Haarlem et al. 2013) and the Square Kilometre Array (SKA; Mellema et al. 2013). The application of BAYESCAL in this regime will be illustrated in future work.

The remainder of the paper is organised as follows. In Section 2, we briefly summarize the results of Paper I relevant to estimation of absolute calibration solutions for redundantly calibrated visibilities using standard absolute calibration absolute calibration supplemented by priors on the gain solutions and BAYESCAL. In Section 3, we describe the instrument model we use when constructing the simulated observed visibility data which we use to demonstrate the BAYESCAL algorithm. In Section 4, we describe the sky models used in the simulated observations. In Section 5, we demonstrate the result of applying BAYESCAL to derive absolute calibration solutions for the simulated observed data from a HERA-like hexagonal close-packed redundant array and compare them to standard calibration solutions for the same data sets. We discuss our conclusions in Section 6.

2   SUMMARY OF PAPER I

In this section, we summarize the results of Paper I relevant to estimation of absolute calibration solutions for redundantly calibrated visibilities with the three calibration approaches considered in Section 5. Section 2.1 introduces per-instrumental-polarisation direction-independent interferometric calibration. Sections 2.2 and 2.3 briefly describes the relative and absolute calibration steps that the calibration of redundant array can be subdivided into and the impact of low-level non-redundancy in the baselines of an assumed redundant array. Section 2.4 describes the Gaussian likelihood for standard absolute calibration of redundantly calibrated visibilities, which we maximise to derive 'absolute calibration solutions'. Building on this, Section 2.5 describes the posterior probability distribution for absolute calibration of redundantly calibrated visibilities supplemented by priors on the spectral structure of the degenerate amplitude parameters, which we maximise to derive 'absolute calibration plus gain prior solutions'. Section 2.6 describes the posterior probability distribution for absolute calibration of redundantly calibrated visibilities with BAYESCAL, which we sample from to derive 'BAYESCAL solutions'. Finally, Section 2.7 describes our model for the temporal evolution of the gain solutions.

2.1   Calibration formalism

The polarised interferometric visibilities corresponding to the cross-correlation of the voltages v_p^a and v_q^b induced in feeds a and b, respectively, of two dual-polarisation antennas, p and q, by the electric

field due to a continuous sky brightness distribution, expressed in an equatorial basis, can be written in the frame of the instrument, in which the brightness distribution on the celestial sphere is rotating overhead, using the Radio Interferometer Measurement Equation (RIME; Hamaker, Bregman, & Sault 1996; Smirnov 2011; hereafter, HBS96 and Sm11, respectively) as,

$$V_{pq}^{ab}(\boldsymbol{u}_{pq}, \nu, t) = n(\nu, t) + (g_p^a(\nu, t))(g_q^b(\nu, t))^*$$
$$\times \iint\limits_{lm} \frac{\mathrm{d}l\,\mathrm{d}m}{n} \left[P_{ab,I}(\boldsymbol{l}, \nu) I(\boldsymbol{l}, \nu, t) + P_{ab,Q}(\boldsymbol{l}, \nu) Q(\boldsymbol{l}, \nu, t) \right.$$
$$\left. + P_{ab,U}(\boldsymbol{l}, \nu) U(\boldsymbol{l}, \nu, t) + P_{ab,V}(\boldsymbol{l}, \nu) V(\boldsymbol{l}, \nu, t) \right] \mathrm{e}^{-2\pi i \boldsymbol{u}_{pq}^T(\nu)\boldsymbol{l}(\nu)} . \tag{1}$$

Here, $\boldsymbol{u}_{pq} = \boldsymbol{b}_{pq}/\lambda = (u_{pq}, v_{pq}, w_{pq})$ are time-stationary uvw-coordinates of the baseline in the frame of the instrument; $n(\nu, t)$ describes the noise on $V_{pq}^{ab}(\boldsymbol{u}_{pq}, \nu, t)$; $P_{ab,j}(\boldsymbol{l}, \nu)$, where $j \in [I, Q, U, V]$, are the elements of the polarised primary beam matrix that map the polarised sky expressed in an equatorial basis in terms of Stokes parameters I, Q, U, and V onto the apparent sky viewed in a topocentric basis of the antennas, in which the basis vectors are aligned with feeds a and b (see Section 3.3); g_p and g_q are complex direction-independent gains of antennas p and q and we have assumed that the antenna configuration is sufficiently compact and the antennas have a sufficiently narrow FoV that direction-dependent sky-based signal propagation effects can be neglected.

Going forward, we consider antennas p and q each with north-south and east-west oriented orthogonal dual feeds, n and e, respectively. The four cross-correlations of the voltages v_p^n, v_p^e, v_q^n and v_q^e induced in the feeds of antennas p and q yield the interferometric visibilites $V_{pq}^{nn} = \langle (v_p^n)(v_q^n)^* \rangle$, $V_{pq}^{ne} = \langle (v_p^n)(v_q^e)^* \rangle$, $V_{pq}^{en} = \langle (v_p^e)(v_q^n)^* \rangle$ and $V_{pq}^{ee} = \langle (v_p^e)(v_q^e)^* \rangle$, where, $\langle . \rangle$ denotes an average over a small interval in time and frequency. In this work, we demonstrate calibration in the context of V_{pq}^{nn} visibilities; thus, all calculated visibilities refer to correlations of pairs of north-south oriented feeds so we omit the 'n' superscripts for brevity. However, each of the remaining three instrumental visibility correlations can be calibrated in an equivalent manner upon substitution of the corresponding terms of the polarised primary beam matrix and complex gains for the visibity correlation to be calibrated.

In the following sections, it will be helpful to write a discretised version of Equation 1 for a set of baselines as,

$$V^{\mathrm{obs}} = \boldsymbol{n} + \mathbf{G}V^{\mathrm{true}} , \tag{2}$$

where V^{obs}, $\boldsymbol{n}$ and V^{true} are given by the concatenation over baselines of the observed visibilities, noise on the data and 'true visibilities' (corresponding to the discretised version of the integral in Equation 1), respectively, each vectorised over a discrete set of frequencies and times. $\mathbf{G}$ is a diagonal matrix encoding the antenna, frequency and time dependent instrumental gains, and, for a single frequency and time, it has elements $G_{ij} = \delta_{ij} g_p g_q^*$, with p and q the antennas associated with the ith visibility, such that V^{obs} has elements,

$$V_i^{\mathrm{obs}} = n_{pq} + g_p g_q^* V_i^{\mathrm{true}} . \tag{3}$$

Here, V_i^{obs} and V_i^{true} are the ith observed visibility and 'true' visibility, respectively, and the 'true' visibility is the cross-correlation of the signals from antennas p and q at time t and frequency ν that would have been measured if the direction-independent instrumental gains were unity; i is an index over visibilities which runs

over the $n_t n_\nu N_{\mathrm{ant}}(N_{\mathrm{ant}} - 1)/2$ cross-correlations between the signals from the N_{ant} antennas in the array, at the n_t time integrations and n_ν frequency channels in the data set to be calibrated.

In an array with a redundant antenna layout, such that it measures visibility data derived from the cross-correlation of the voltage responses from sets of antennas with identical baseline separation vectors and beam patterns, it is advantageous to reparametrise the antenna gain parameters as $g_p = h_p f(A, \phi_l, \phi_m)$ where h_p is the relative gain of the antenna and $f(A, \phi_l, \phi_m)$ is a function of degenerate gain amplitude, A, and tip-tilt phases, ϕ_l and ϕ_m, respectively. In this case, we can rewrite the instrumental gain matrix as,

$$\mathbf{G} = \mathbf{H}\mathbf{D} , \tag{4}$$

where the relative gain matrix $\mathbf{H}$ is a diagonal matrix encoding the antenna, frequency and time dependent relative antenna gains, and, for a single frequency and time, it has elements $H_{ij} = \delta_{ij} h_p h_q^*$, with p and q the antennas associated with the ith visibility, and the degenerate gain matrix $\mathbf{D}$ is a diagonal matrix encoding the degeneracy function $f(A, \phi_l, \phi_m)$ (see Section 2.2 for details).

Defining an interferometric fringe matrix with elements[3] $\mathbf{F}_{\mathrm{fr},jk} = \gamma \mathrm{e}^{-2\pi i(\boldsymbol{u} \otimes \boldsymbol{l} + \boldsymbol{v} \otimes \boldsymbol{m} + \boldsymbol{w} \otimes \boldsymbol{n})_{jk}}$, which encodes the transformation from the image domain to the uvw-coordinates sampled by the interferometric array, where γ is a normalisation factor (see Section 3.1, and paper I for details), and polarised primary beam matrices, $\mathbf{P}_{\mathrm{nn},I}$, $\mathbf{P}_{\mathrm{nn},Q}$, $\mathbf{P}_{\mathrm{nn},U}$ and $\mathbf{P}_{\mathrm{nn},V}$, which encode the polarisation, frequency and direction-dependent primary beam profiles, we can rewrite V^{true} as[4],

$$V^{\mathrm{true}} = \mathbf{F}_{\mathrm{fr}} [\mathbf{P}_{\mathrm{nn},I} T_I + \mathbf{P}_{\mathrm{nn},Q} T_Q + \mathbf{P}_{\mathrm{nn},U} T_U + \mathbf{P}_{\mathrm{nn},V} T_V] . \tag{5}$$

We can rewrite Equation 3 using Equations 4 and 5 as,

$$V^{\mathrm{obs}} = \boldsymbol{n} + \mathbf{H}\mathbf{D}\mathbf{F}_{\mathrm{fr}} [\mathbf{P}_{\mathrm{nn},I} T_I + \mathbf{P}_{\mathrm{nn},Q} T_Q + \mathbf{P}_{\mathrm{nn},U} T_U + \mathbf{P}_{\mathrm{nn},V} T_V] . \tag{6}$$

The data calibration problem requires solving for gain parameters h_p^{m} and $f(A, \phi_l, \phi_m)$, given constraints deriving from array redundancy and sky modelling. Approaches to achieving this are summarised in Sections 2.2 – 2.6.

2.2 Redundant calibration

When an array has a redundant antenna layout, such that it measures visibility data derived from the cross-correlation of the voltage responses from sets of antennas with identical baseline separation vectors and beam patterns, the fact that the true visibilities, V^{true}, that would be measured if all instrument gains were unity are identical on redundant baselines enables one to constrain the relative gain parameters of the array independently of the sky model (e.g. Wieringa 1992; Liu, et al. 2010). This enables redundant arrays to be fully calibrated via a two-stage process:

(i) *'relative' calibration*, in which baseline redundancy is used to derive redundant gain parameters, h_p^{m}, that are equal to the general direction-independent gain parameter, g_p^{m}, associated with antenna p, to within a frequency, time and polarisation dependent complex degeneracy factor. This complex degeneracy factor is described by

---

[3] To eliminate any ambiguity between the earlier use of i as a counting index and its use here as an imaginary number we advise the reader that i in all instances of the variable combination $2\pi i$ in this paper is the imaginary unit.

[4] The specific form that the component matrices in Equation 5 take for the sky and instrument models considered in this paper will be discussed in Sections 3 and 4.

a set of degenerate calibration parameters that cannot be solved for using baseline redundancy.

(ii) *'absolute' calibration*, in which the degenerate calibration parameters associated with relative calibration are solved for with reference to a sky model.

For further details of relative calibration, see e.g. Liu, et al. 2010 or paper I. Here, we will assume that this first stage has been completed and, by fixing the gain amplitude of a reference antenna and the gain phases of two reference antennas to an arbitrary choice of values (e.g. Li, et al. 2018; Dillon, et al. 2020), accurate solutions have been obtained relative to these reference parameters for the redundant calibration parameters of the antennas of the array, $\hat{h}_p^{\mathrm{m}}$, encoded in a diagonal relative calibration matrix, $\hat{\mathbf{H}}^{\mathrm{m}}$.

To fully calibrate the array, one must subsequently solve for the degenerate gain parameters A, ϕ_l, ϕ_m corresponding to the per-frequency, time and polarisation dependent difference between the arbitrary choice of reference parameters used during relative calibration and the true values of these parameters[5] in the data set under consideration.

2.3 Imperfect array redundancy

For a truly redundant array, relative calibration recovers unbiased relative gain solutions. If there are non-redundancies in an array that is assumed to be redundant, relative calibration will introduce additional biases in the gain solutions (e.g. Joseph, Trott & Wayth 2018). However, in Choudhuri, Bull, & Garsden (2021) it is found that primary beam non-redundancies, of a variety of forms, yield a low level of additional spectral structure in the gain solutions. Additionally, in Orosz et al. (2019) it is shown that spurious spectral structure in recovered relative gain solutions introduced through antenna-position and primary beam non-redundancy can be largely mitigated by relative upweighting of short baselines during calibration. These biases can also be mitigated by treating nominally redundant baselines as correlated rather than identical to within the instrumental noise (e.g. Sievers 2017; Byrne et al. 2021).

In contrast, in Byrne, et al. (2019) it is demonstrated that even in the absence of relative calibration errors, sky-model-incompleteness-sourced biases in the gain parameters solved-for with absolute calibration are sufficiently large to produce calibration-error-induced foreground systematics at a level that can overwhelm the high-redshift 21 cm signal and prevent a detection of of the 21 cm power spectrum on a range of spatial scales. In this work we demonstrate the effectiveness of BAYESCAL for resolving the sky-model incompleteness problem in the context of calibration of redundant interferometric arrays. The application of BAYESCAL to resolving the sky-model incompleteness problem in the context of a partially- or non-redundant array or when calibrating with a sky model alone, is discussed in paper I. We defer application of BAYESCAL in this context to future work.

2.4 Standard absolute calibration

To derive absolute calibration solutions one can define an absolutely calibrated model for the data as,

$$V^{\mathrm{model}} = \hat{\mathbf{H}}^{\mathrm{m}} \mathbf{D}^{\mathrm{m}} V^{\mathrm{sim}} \, . \tag{7}$$

Here, $\mathbf{D}^{\mathrm{m}}$ is our model degenerate gain matrix, which is a diagonal matrix that for a single time integration and frequency channel has elements $D_{ij} = \delta_{ij} A \mathrm{e}^{i(\boldsymbol{b}_i^T \boldsymbol{\Phi})}$; $\hat{\mathbf{H}}^{\mathrm{m}}$ is the diagonal matrix containing the maximum likelihood redundant gain parameters, $\hat{h}_p^{\mathrm{m}}$, and the elements of the model vector are given by,

$$V_{\alpha,j}^{\mathrm{model}} = \hat{h}_p^{\mathrm{m}} \hat{h}_q^{\mathrm{m}*} A \mathrm{e}^{i(\boldsymbol{b}_i^T \boldsymbol{\Phi})} V_{\alpha}^{\mathrm{sim}} \, , \tag{8}$$

where $V_{\alpha}^{\mathrm{sim}}$ is a simulated model for the true visibility on the αth unique baseline,

$$\boldsymbol{V}^{\mathrm{sim}} = \mathbf{F}_{\mathrm{fr}} [\mathbf{P}_{\mathrm{nn},I} \boldsymbol{T}_I^{\mathrm{sim}} + \mathbf{P}_{\mathrm{nn},Q} \boldsymbol{T}_Q^{\mathrm{sim}} + \mathbf{P}_{\mathrm{nn},U} \boldsymbol{T}_U^{\mathrm{sim}} + \mathbf{P}_{\mathrm{nn},V} \boldsymbol{T}_V^{\mathrm{sim}}] \, . \tag{9}$$

Here $\boldsymbol{T}_j^{\mathrm{sim}}$, with $j \in [I, Q, U, V]$, are our simulated sky models for emission in Stokes I, Q, U and V (see Section 4 for details).

A Gaussian likelihood for our full model of the data as a function of amplitude and tip-tilt gain phase parameters, A, $\boldsymbol{\Phi}_l$ and $\boldsymbol{\Phi}_m$, encoded in the degeneracy calibration matrix, $\mathbf{D}^{\mathrm{m}}$, is thus given by,

$$\mathrm{Pr}(V^{\mathrm{obs}} \mid A, \boldsymbol{\Phi}_l, \boldsymbol{\Phi}_m) = \frac{1}{\pi^{N_{\mathrm{vis}}} \det(\mathbf{N})}$$
$$\times \exp\left[-\left(V^{\mathrm{obs}} - \hat{\mathbf{H}}^{\mathrm{m}} \mathbf{D}^{\mathrm{m}} V^{\mathrm{sim}}\right)^{\dagger} \mathbf{N}^{-1} \left(V^{\mathrm{obs}} - \hat{\mathbf{H}}^{\mathrm{m}} \mathbf{D}^{\mathrm{m}} V^{\mathrm{sim}}\right) \right] \, . \tag{10}$$

In Section 5, when using Equation 10 to solve for the maximum likelihood values of A, $\boldsymbol{\Phi}_l$, $\boldsymbol{\Phi}_m$ for a given set of visibilities, we use a Generalized Simulated Annealing algorithm (Xiang et al. 1997), as implemented in scipy[6].

Since A, $\boldsymbol{\Phi}_l$, $\boldsymbol{\Phi}_m$ are all frequency dependent, their maximum likelihood solutions will still absorb chromatic errors in V^{sim} relative to V^{true} as spurious spectral structure. Within this framework, Byrne, et al. (2019) have shown that model incompleteness can couple a prohibitive level of spectral structure into the gain solutions for 21 cm cosmology applications. We also find this to be the case for the sky and instrument model considered here, when calibrating simulated observed data using redundant calibration (see Section 5 for details).

2.5 Absolute calibration with gain priors

In addition to using a sky model to constrain the spectral structure of the antenna gain solutions, as described in Section 2.4, structure present in the gains can also be estimated with alternate techniques, such as electric and electromagnetic co-simulations of the receiver system (e.g. Fagnoni et al. 2021) and reflectometry measurements of the feed-dish system (e.g. Patra, et al. 2018) that are independent of an astrophysical sky model.

In general, to leverage this a priori knowledge, one can incorporate these constraints as spectral priors on the gain solutions. For absolute calibration of a redundantly calibrated data set, this corresponds to placing priors on the spectral dependence of the redundant gain degeneracy function $f_{\mathrm{vis}}(A, \boldsymbol{\Phi}_l, \boldsymbol{\Phi}_m)$. In this work, we focus on incorporating information about the power in the amplitude fluctuations of the redundant gain degeneracy function.

The Fourier domain of the degenerate gain amplitudes provides a natural space to encode priors on the power in these gain fluctuations as a function of frequency since, if power in the degenerate

---

[5] The degenerate calibration amplitude A, and tip-tilt phases, ϕ_l, ϕ_m, determine the per-frequency, time and polarisation dependent flux-scale and pointing center of the data, respectively.

[6] https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.dual_annealing.html

amplitudes is limited to the spectral scales that are Nyquist sampled in the data set being calibrated[7], the diagonal elements of the covariance matrix of the Fourier modes encode the power in the fluctuations as a function of spatial scale. Although here we do not impose an informative prior on the distribution of the tip-tilt phases, we, nevertheless, follow Paper I and for mathematical convenience apply to them the same reparametrisation. We, therefore, rewrite our degeneracy function parameters $A(\nu)$, $\boldsymbol{\Phi}_l(\nu)$ and $\boldsymbol{\Phi}_m(\nu)$ using a new set of parameters that describe the amplitudes of their Fourier decompositions, $A_{\mathrm{F}}(\eta)$, $\boldsymbol{\Phi}_{l,\mathrm{F}}(\eta)$ and $\boldsymbol{\Phi}_{m,\mathrm{F}}(\eta)$, with,

$$A = \mathsf{F}^{-1} A_{\mathrm{F}} \tag{11}$$

$$\boldsymbol{\Phi}_l = \mathsf{F}^{-1} \boldsymbol{\Phi}_{l,\mathrm{F}}$$

$$\boldsymbol{\Phi}_m = \mathsf{F}^{-1} \boldsymbol{\Phi}_{m,\mathrm{F}} \,.$$

Here, F^{-1} is a one dimensional inverse Fourier transform matrix mapping from the 'halfcomplex' representation of the η-domain of the gain parameters to their frequency-domain representations, and A_{F}, $\boldsymbol{\Phi}_{l,\mathrm{F}}$ and $\boldsymbol{\Phi}_{m,\mathrm{F}}$ are $N_\nu \times 1$ real column vectors encoding the 'halfcomplex' representation of Fourier degenerate gain amplitude, tip- and tilt-phase parameters, respectively (see paper I for details).

We assume a uniform prior on the parameters defining the Fourier decomposition of the tip-tilt phases such that $\mathrm{Pr}(\boldsymbol{\Phi}_{l,\mathrm{F}}) = \mathrm{Pr}(\boldsymbol{\Phi}_{m,\mathrm{F}}) = 1$ and impose a Gaussian prior encoding the power in amplitude fluctuations of the redundant gain degeneracy function as,

$$\mathrm{Pr}(A_{\mathrm{F}} \mid \sigma^2_{A_{\mathrm{F}}}) \propto \frac{1}{\sqrt{\det(\boldsymbol{\Sigma}_{A_{\mathrm{F}}})}}$$
$$\times \exp\left[-\frac{1}{2}(A_{\mathrm{F}} - \bar{A}_{\mathrm{F}})^T \boldsymbol{\Sigma}^{-1}_{A_{\mathrm{F}}}(A_{\mathrm{F}} - \bar{A}_{\mathrm{F}})\right] \,. \tag{12}$$

Here, $\bar{A}_{\mathrm{F}}$ is a vector of amplitudes defining the expectation value of the Fourier decomposition of A, $\boldsymbol{\sigma}_{A_{\mathrm{F}}}$ is the vector of expected variances of A_{F} and $\boldsymbol{\Sigma}_{A_{\mathrm{F}}}$ is a diagonal matrix with elements $\Sigma_{A_{\mathrm{F}},ij} = \delta_{ij}\sigma^2_{A_{\mathrm{F}}}$.

In this work, we assume that the mean amplitude of the gain solution is known a priori, from e.g. electromagnetic co-simulations of the receiver system and reflectometry measurements of the feed-dish system, with a one-sigma fractional uncertainty of 10%. We construct our simulated observed data such that it is described by a true redundant gain degeneracy function with a frequency-dependent bandpass amplitude, $A^{\mathrm{true}} = \mathsf{F}A^{\mathrm{true}}_{\mathrm{F}}$ (Figure 1, right), and, motivated by simulations of the first generation HERA feed (Fagnoni et al. 2021), we draw $A^{\mathrm{true}}_{\mathrm{F}}$ from a Gaussian distribution with an exponentially decreasing spectral-scale dependent standard deviation (Figure 1, left), which we assume is known a priori, from a combination of electric and electromagnetic co-simulations of the receiver system and reflectometry measurements of the feed-dish system, for the purposes of applying the prior defined in Equation 12.

Using Bayes' theorem, we can thus derive the posterior probability distribution for the degenerate gain parameters conditioned

on the raw data and the power in fluctuations in the degenerate gain amplitude as (see Paper I for details),

$$\mathrm{Pr}(A_{\mathrm{F}}, \boldsymbol{\Phi}_{l,\mathrm{F}}, \boldsymbol{\Phi}_{m,\mathrm{F}} \mid V^{\mathrm{obs}}, \sigma^2_{A_{\mathrm{F}}}) \propto \mathrm{Pr}(V^{\mathrm{obs}} \mid A_{\mathrm{F}}, \boldsymbol{\Phi}_{l,\mathrm{F}}, \boldsymbol{\Phi}_{m,\mathrm{F}})$$
$$\times \mathrm{Pr}(A_{\mathrm{F}} \mid \sigma^2_{A_{\mathrm{F}}}) \, \mathrm{Pr}(A_{\mathrm{F}}) \, \mathrm{Pr}(\boldsymbol{\Phi}_{l,\mathrm{F}}) \, \mathrm{Pr}(\boldsymbol{\Phi}_{m,\mathrm{F}}) \,, \tag{13}$$

where $\mathrm{Pr}(V^{\mathrm{obs}} \mid A_{\mathrm{F}}, \boldsymbol{\Phi}_{l,\mathrm{F}}, \boldsymbol{\Phi}_{m,\mathrm{F}})$ is given by Equation 10 reparametrised in terms of the Fourier gain parameters of Equation 11. Here, we assume $\sigma^2_{A_{\mathrm{F}}}$ is known a priori. If the uncertainties associated with estimates of $\sigma^2_{A_{\mathrm{F}}}$ are significant, one can, instead, sample from $\boldsymbol{\sigma}_{A_{\mathrm{F}}}$ jointly with A_{F}, $\boldsymbol{\Phi}_{l,\mathrm{F}}$, $\boldsymbol{\Phi}_{m,\mathrm{F}}$ and $\boldsymbol{\varepsilon}$ and fit for the most probable degenerate gain parameters and power in the degenerate gain amplitude given the data (see Paper I for details).

2.6 Absolute calibration using BAYESCAL

To extend the approach to absolute calibration of redundantly calibrated visibilities described in Section 2.5 to explicitly account for sky-model incompleteness, we subdivide our sky model into two components,

(i) a simulated component, V^{sim}, derived using our a priori knowledge of the brightness distribution of known Galactic emission and extragalactic sources in the region of sky under observation and

(ii) a fitted component, $V^{\mathrm{fit}}(\boldsymbol{\varepsilon})$, with $\boldsymbol{\varepsilon}$ a set of image-domain parameters of the fitted model, that accounts for the contribution to V^{true} missing in V^{sim}.

We write our model visibilites as the sum of these components,

$$V^{\mathrm{model}} = \hat{\mathsf{H}}^{\mathrm{m}} \mathsf{D}^{\mathrm{m}}(V^{\mathrm{sim}} + V^{\mathrm{fit}}(\boldsymbol{\varepsilon})) \,. \tag{14}$$

Following paper I, we write our fitted visibility model over all LSTs of our calibration data set as,

$$V^{\mathrm{fit}} = \mathsf{F}_{\mathrm{fr}} \mathsf{P}_{\mathrm{nn},I} \mathsf{S} \mathsf{C} \boldsymbol{\varepsilon} \,. \tag{15}$$

Here,

- F_{fr} and $\mathsf{P}_{\mathrm{nn},I}$ are as defined in Equation 5.
- S encodes our spectral model for the fitted component of the emission. In Section 5, we use BAYESCAL to calibrate simulated data on sub-30 m HERA baselines[8] (see Section 3 for details of the instrument model). In this baseline range, Galactic diffuse synchroton emission (GDSE) accounts for a large fraction of the power in the visibilities and, owing to the level of uncertainty associated with models for this emission, it is also the dominant source of model incompleteness. As such, we use a power law model for the spectral structure of the fitted emission in this regime with power law index $\beta_{\mathrm{m}} = \langle \beta \rangle_{\mathrm{GDSE}}$, where $\langle \beta \rangle_{\mathrm{GDSE}}$ is the mean spectral index of the a priori known component of the GDSE brightness temperature distribution. In general, the Bayesian evidence can be used to select an optimal spectral model from a set of models for the data or, for this fixed spectral model, to determine the optimal choice of

---

[7] If, additionally, there is power in the degenerate amplitudes on spectral scales larger than the bandwidth of the calibration data set, the degenerate gain parameters can instead be jointly modelled by a set of Fourier modes on Nyquist sampled scales and, for example, a quadratic large spectral scale model that prevents distortion of the expected power spectrum due to spectral leakage (see Sims et al. 2019 for an example of this approach applied to large spectral scale modelling in 21 cm power spectrum estimation).

[8] While there is no fundamental limitation restricting the use of BAYESCAL to calibration of visibilities on short baselines, calibrating on short baselines has the benefit of reducing the spatial resolution necessary for V^{fit} to accurately model the incomplete component of the data and, thus, decreasing the computation time required for the calibration. Additionally, 21 cm cosmology experiments are generally most sensitive on large angular scales; thus, if sky-model incompleteness can be mitigated by BAYESCAL, barring potential complications such as increased cross-talk between antennas, short baselines constitute a preferred subset if one wishes to use the same data during calibration and data analysis.
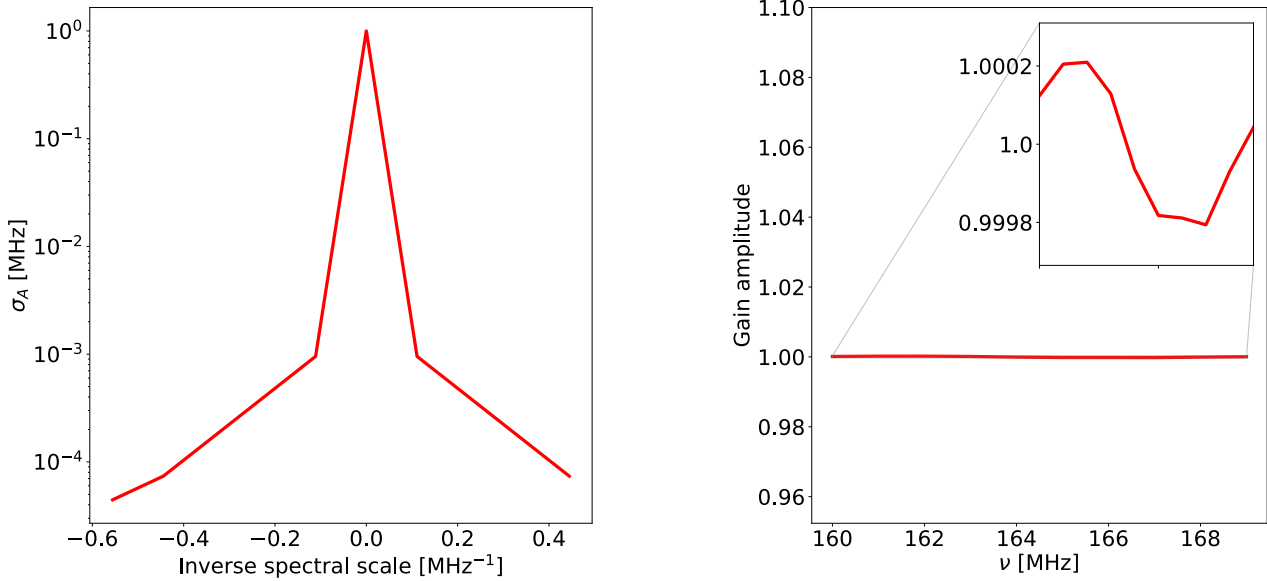
Figure 1. [Left] Prior on the standard deviation of the amplitudes of the basis vectors comprising the Fourier decomposition of the amplitude parameters of the redundant gain degeneracy function, σ_A. [Right] Degenerate gain amplitude realisation used in the construction of the simulated observed visibilites calibrated in Section 5.

model spectral index. However, in Section 5, we will show that this a priori choice of fixed spectral model is already sufficiently robust for high fidelity recovery of calibration solutions with the BAYES-CAL calibration framework, for the observational setup described in Section 3, so we do not consider such a model selection analysis further here.

- ε are our full set of sky-model amplitude parameters for describing the Stokes I brightness temperature of the component of emission in V^{true} that is unaccounted for in V^{sim}. We define these parameters on a HEALPIX grid (Górski, et al. 2005) at a reference frequency, ν_0.
- **C** is a matrix that maps between the parameters of our fitted calibration model, ε, and $\varepsilon_{\text{concat}}$, where $\varepsilon_{\text{concat}}$ is the vectorisation of the concatenation of model maps over the LSTs in the data set to be calibrated. The N_{pix} pixels of the HEALPIX grid cover the full sky; however, if the primary beam of a visibility downweights emission at large zenith angles, a subset of the pixels within the observer's horizon may be sufficient to model the visibilities. To take advantage of this and reduce the computational cost of the calibration, when applying BAYESCAL, we define our set of parameters, ε, such that they represent the amplitudes in the $N_{\text{pix,s}}$ length subset of pixels defined by the union of pixels that fall within a fixed zenith angle, θ_{cut}, such that, within this region, the primary beam lies above a given threshold weight, at each central local sidereal time (LST) of the integrations over which the calibration solutions are being jointly estimated. Commonly, one wishes to calibrate interferometric data comprised of a number of shortly spaced successive time integrations, in which case, the region of sky contributing to successive integrations is strongly correlated in time. ε represents the amplitudes of fixed LST-independent regions of the sky; however, **S** and $\mathbf{F}_{\text{fr}}\mathbf{P}_{\text{nn},I}$ operate on vectorisations of the concatenation, over the LSTs in the data set to be calibrated, of model maps ($\varepsilon_{\text{concat}}$) and model image cubes ($\mathbf{S}\varepsilon_{\text{concat}}$), respectively. **C** is defined such that $\varepsilon_{\text{concat}} = \mathbf{C}\varepsilon$ (see paper I for details).

2.6.1 *BAYESCAL likelihood*

Our likelihood at this stage can be written as (see paper I),

$$\Pr(V^{\text{obs}} \mid A, \Phi_l, \Phi_m, \varepsilon) =$$

$$\frac{1}{\pi^{N_{\text{vis}}}\det(\mathbf{N})} \exp\left[-\left(V^{\text{obs}} - \hat{\mathbf{H}}^{\text{m}}\mathbf{D}^{\text{m}}(V^{\text{sim}} + V^{\text{fit}})\right)^{\dagger}\right.$$

$$\left.\mathbf{N}^{-1}\left(V^{\text{obs}} - \hat{\mathbf{H}}^{\text{m}}\mathbf{D}^{\text{m}}(V^{\text{sim}} + V^{\text{fit}})\right)\right] . \quad (16)$$

Here, **N** is the expected covariance matrix of the raw visibilities and can be estimated by,

$$\mathbf{N} = (\mathbf{HD})^{\dagger}\mathbf{N}'(\mathbf{HD}) , \quad (17)$$

where $\mathbf{N}'$ is the expected covariance matrix of the calibrated data. In this paper, we assume the noise is uncorrelated between baselines and thus **N** has elements,

$$N'_{ij} = \left\langle n_i n_j^* \right\rangle = \delta_{ij}\sigma_j^2 . \quad (18)$$

Here, $\langle .. \rangle$ represents the expectation value and σ_j is the rms value of the noise term for visibility j, which, for a pair of identical antennas individually experiencing equal system noise, is given by (e.g. Taylor, Carilli, & Perley 1999),

$$\sigma = \frac{1}{\eta_s\eta_a} \frac{2k_B T_{\text{sys}}}{A\sqrt{2\Delta\nu\tau}} , \quad (19)$$

where $k_B = 1.3806 \times 10^{-23}$ JK^{-1} is Boltzmann's constant, T_{sys} is the system noise temperature, A is the antenna area, $\Delta\nu$ is the channel width, τ is the integration time and η_s and η_a are the system and antenna efficiencies, respectively.

In Section 2.2, the gain solutions at each frequency can be solved for independently. The introduction in BAYESCAL of the fitted frequency dependent visibility model, V^{fit}, as well as the inclusion of spectrally dependent gain priors in both the absolute calibration

with gain priors and BayesCal approaches, means that in each of these cases the instrumental gains are correlated in frequency and, correspondingly, the gain parameters of each channel of the data set must be fit for jointly. Thus, for both of these cases, we redefine our model degenerate gain parameter matrix such that for a single time integration, $\mathbf{D}^{\mathrm{m}}$ is a block diagonal matrix where each block, $\mathbf{D}_r^{\mathrm{m}}$, is a diagonal matrix with elements $D_{r,ij} = \delta_{ij} A_r e^{ib_i^T \boldsymbol{\Phi}_r}$ and the subscript r runs over the N_ν channels of the data set. We also redefine $\hat{\mathbf{H}}^{\mathrm{m}}$ in an analogous manner to now encode the maximum likelihood redundant gain parameters for each channel of the data set.

2.6.2 Fitted calibration model power priors

Paper I details how one can place informative priors on the power in the fitted component of the BayesCal calibration model, parametrised in terms of:

(i) a model for the spatial power spectrum of $\boldsymbol{\varepsilon}$, or
(ii) a model for the covariance matrix of $\boldsymbol{\varepsilon}$.

These can be used to eliminate the degeneracies between $\boldsymbol{\varepsilon}$ and instrumental calibration parameters that are present when $\boldsymbol{\varepsilon}$ is unconstrained. A diagonal model for the covariance matrix of $\boldsymbol{\varepsilon}$ encoding the expected power in the image domain attributable to emission missing in the simulated sky model from which V^{sim} derives provides a minimally informative prior that is sufficient for this goal. For the incomplete sky models considered in this paper (see Section 4) and on sub-30 m baselines (which are the baseline-lengths included in data sets on which we demonstrate absolute calibration with BayesCal in Section 5), the total sky-model error is dominated by the power in unmodelled or mismodelled diffuse emission that derives from the discrepancy between our model for the full-Stokes diffuse emission and the true full-Stokes brightness distribution. In this paper, we assume a uniform uncertainty on the diffuse emission in the field being calibrated and approximate the covariance matrix of $\boldsymbol{\varepsilon}$ as diagonal, with elements,

$$\Sigma_{lm,\delta T,ii} = \sigma_\varepsilon^2 \ . \tag{20}$$

Here, σ_ε is the expected RMS error on the sky model used to construct V^{sim}, which we approximate as $\sigma_\varepsilon = f(\sum(T_i - \overline{T})^2/N_\varepsilon)^{1/2}$, where f and $(\sum(T_i - \overline{T})^2/N_\varepsilon)^{1/2}$ are the fractional uncertainty (see Section 4) and RMS of our simulated calibration sky model, respectively. We can correspondingly write our power spectral prior as,

$$\Pr(\boldsymbol{\varepsilon} \mid \sigma_\varepsilon^2) \ \propto \ \frac{1}{\sqrt{\det(\Sigma_{lm,\delta T})}} \exp\left[-\frac{1}{2}\,\boldsymbol{\varepsilon}^T \Sigma_{lm,\delta T}^{-1} \boldsymbol{\varepsilon}\right] \ . \tag{21}$$

Incorporating this prior on the power in the parameters of our fitted visbility model and not assuming any a priori correlation between A, $\boldsymbol{\Phi}_l$, $\boldsymbol{\Phi}_m$, $\boldsymbol{\varepsilon}$, we can write the joint probability density of our calibration parameters and our image-space model coefficients conditioned on the raw data and the power in $\boldsymbol{\varepsilon}$ as,

$$\Pr(A, \boldsymbol{\Phi}_l, \boldsymbol{\Phi}_m, \boldsymbol{\varepsilon} \mid V^{\mathrm{obs}}, \sigma_\varepsilon^2) \ \propto \ \Pr(V^{\mathrm{obs}} \mid A, \boldsymbol{\Phi}_l, \boldsymbol{\Phi}_m, \boldsymbol{\varepsilon})$$
$$\times \Pr(\boldsymbol{\varepsilon} \mid \sigma_\varepsilon^2) \Pr(A) \Pr(\boldsymbol{\Phi}_l) \Pr(\boldsymbol{\Phi}_m) \ . \tag{22}$$

As discussed in paper I, in a scenario where the uncertainty on f is assumed to be significant, one could jointly sample from a parametrised model for σ_ε^2, in addition to $\boldsymbol{\varepsilon}$ and the calibration parameters. Here, we calculate σ_ε^2 as described below Equation 20; however, we have confirmed that the maximum a posteriori calibration parameters recovered in Section 5 are relatively insensitive to

this assumption, with errors of up to 20% on the value of f used to define the power spectral prior altering the level of spurious spectral fluctuations in the derived calibration solutions by less than a factor of two.

2.6.3 Analytic marginalisation over the sky-model parameters

Using Bayes' theorem, we can write the posterior probability distribution of the degenerate gain parameters and sky-model parameters conditioned on the raw data, the power in spectral fluctuations in the degenerate gain amplitude (Equation 12) and the power in V^{fit} (Equation 21) as,

$$\Pr(A_{\mathrm{F}}, \boldsymbol{\Phi}_{l,\mathrm{F}}, \boldsymbol{\Phi}_{m,\mathrm{F}}, \boldsymbol{\varepsilon} \mid V^{\mathrm{obs}}, \sigma_\varepsilon^2, \sigma_{A_{\mathrm{F}}}^2) \ \propto$$
$$\Pr(V^{\mathrm{obs}} \mid A_{\mathrm{F}}, \boldsymbol{\Phi}_{l,\mathrm{F}}, \boldsymbol{\Phi}_{m,\mathrm{F}}, \boldsymbol{\varepsilon}) \Pr(\boldsymbol{\varepsilon} \mid \sigma_\varepsilon^2)$$
$$\times \Pr(A_{\mathrm{F}} \mid \sigma_{A_{\mathrm{F}}}^2) \Pr(\boldsymbol{\Phi}_{l,\mathrm{F}}) \Pr(\boldsymbol{\Phi}_{m,\mathrm{F}}) \ , \tag{23}$$

In principle, one could sample from the posterior probability distribution given in Equation 23. However, since we are interested specifically in the calibration parameters, we can analytically marginalise over the sky-model parameters and directly sample from the far lower dimensional space of the posterior probability distribution for the calibration parameters. This gives our marginal posterior probability distribution, $\Pr(A_{\mathrm{F}}, \boldsymbol{\Phi}_{l,\mathrm{F}}, \boldsymbol{\Phi}_{m,\mathrm{F}} \mid V^{\mathrm{obs}}, \sigma_\varepsilon^2, \sigma_{A_{\mathrm{F}}}^2)$. For a detailed derivation of this distribution see paper I. Here, we quote the solution for the marginalised distribution for the case of uniform priors on the degenerate tip-tilt phases,

$$\Pr(A_{\mathrm{F}}, \boldsymbol{\Phi}_{l,\mathrm{F}}, \boldsymbol{\Phi}_{m,\mathrm{F}} \mid V^{\mathrm{obs}}, \sigma_\varepsilon^2, \sigma_{A_{\mathrm{F}}}^2) \propto$$
$$\frac{\det(\mathbf{N})^{-1}}{\sqrt{\det(\boldsymbol{\Upsilon}) \det(\Sigma_{lm,\delta T}) \det(\Sigma_{A_{\mathrm{F}}})}}$$
$$\times \exp\left[-(\delta V^{\mathrm{obs}})^\dagger \mathbf{N}^{-1}(\delta V^{\mathrm{obs}}) + (\overline{\delta V}^{\mathrm{obs}})^T \boldsymbol{\Upsilon}^{-1}(\overline{\delta V}^{\mathrm{obs}})\right.$$
$$\left. -\frac{1}{2}(A_{\mathrm{F}} - \bar{A}_{\mathrm{F}})^T \Sigma_{A_{\mathrm{F}}}^{-1}(A_{\mathrm{F}} - \bar{A}_{\mathrm{F}})\right]. \tag{24}$$

Here, $\overline{\delta V}^{\mathrm{obs}} = \boldsymbol{\Lambda}^\dagger \mathbf{N}^{-1} \delta V^{\mathrm{obs}}$ represents the weighted gridded projection of δV^{obs} on the parameter space of the fitted visibility model, $\delta V^{\mathrm{obs}} = V^{\mathrm{obs}} - \hat{\mathbf{H}}^{\mathrm{m}} \mathbf{D}^{\mathrm{m}} V^{\mathrm{sim}}$ is the residual between the visibilities and the fixed simulated component of our calibration model and $\boldsymbol{\Upsilon} \equiv \boldsymbol{\Lambda}^\dagger \mathbf{N}^{-1} \boldsymbol{\Lambda} + \Sigma_{lm,\delta T}^{-1}$ is the covariance matrix of $\overline{\delta V}^{\mathrm{obs}}$, with $\boldsymbol{\Lambda} \equiv \hat{\mathbf{H}}^{\mathrm{m}} \mathbf{D}^{\mathrm{m}} \mathbf{F}_{\mathrm{fr}} \mathbf{PSC}$ the system matrix mapping from our image domain parameters to the fitted model visibilities.

The dense matrix inversion of $\boldsymbol{\Upsilon}$ and the calculation of the determinants are the most computationally expensive elements in Equation 24. However, terms in the exponent without a calibration parameter dependence can be precomputed prior to sampling, the determinant prefactors that are independent of the calibration parameters can be neglected for the purposes of parameter estimation and $\det(\mathbf{N})$ is diagonal and thus can be calculated inexpensively as the product of the diagonal elements. The remaining diagonal matrix inversions can also be calculated trivially as the inverse of the matrix diagonals.

When calibrating simulated observed data using the BayesCal framework in Section 5, we compute $\boldsymbol{\Upsilon}^{-1}$ via Cholesky decomposition and we use Equation 24 to sample directly from the marginalised posterior for the degenerate gain Fourier parameters, $\Pr(A_{\mathrm{F}}, \boldsymbol{\Phi}_{l,\mathrm{F}}, \boldsymbol{\Phi}_{m,\mathrm{F}} \mid V^{\mathrm{obs}}, \sigma_\varepsilon^2, \sigma_{A_{\mathrm{F}}}^2)$, using nested sampling as implemented by the PolyChord algorithm (Handley, Hobson & Lasenby 2015a,b).

2.7 Temporal model

Simultaneous calibration of multiple time integrations, such that the calibration solutions are described by a given temporal model, provides a more stringent constraint on the calibration parameters, in general, relative to calibrating each integration independently, and allows one to take advantage of a priori knowledge of the expected temporal evolution of the calibration parameters. When calibrating with BAYESCAL, it additionally enables one to better constrain the parameters of the fitted sky model.

In Section 5, we will use absolute calibration, absolute calibration with gain priors and BAYESCAL to calibrate five minute data sets comprised of ten snapshot observations (see Section 3). To achieve this, we define a degenerate gain parameter matrix, $\mathbf{D}^m$, that encodes the degenerate gain parameters for all time integrations in the data set being calibrated. In this case, $\mathbf{D}^m$ is a block diagonal matrix where each block, $\mathbf{D}_s^m$, is itself a block diagonal matrix comprised of blocks, $\mathbf{D}_r^m$, which are diagonal matrices with elements $D_{s,r,ij} = \delta_{ij} A_{s,r} e^{i b_i^T \Phi_{s,r}}$, where the subscripts r and s run over the N_ν channels and N_t integrations in the data set, respectively (see paper I for details). In an analogous manner, $\hat{\mathbf{H}}^m$ is redefined to comprise N_t block diagonal matrices with the sth block diagonal encoding the maximum likelihood redundant gain parameters for the sth integration in the data set; we correspondingly redefine our total visibility data as $V^{obs} = [(V_{t_0}^{fit})^T, (V_{t_1}^{fit})^T, \ldots (V_{t_n}^{fit})^T]^T$, where the visibility data is comprised of n time integrations and $V_{t_i}^{fit}$ is a vector of the raw visibilites corresponding to the ith time integration. V^{sim} and V^{fit} are redefined in an analogous manner to V^{obs}. Substituting these redefinitions, the likelihoods and posteriors described in Sections 2.2 – 2.6 can be applied, as written, to simultaneously calibrate multiple-integration data sets.

In this paper, we will assume that the instrument gains are stable over the five minute duration of the data to be calibrated, use time-stationary gains in the simulated observed data and use a time-stationary gain model when calibrating the data. On longer time scales, variations in the environmental conditions of the receiver and signal chain, such as changing ambient temperature, will impart temporal fluctuations in the antenna gains on the temporal scales of those variations. To address this, a model can be introduced for the temporal evolution of the calibration solutions, the parameters of which can be constrained by their own priors, and the optimal temporal model for the data can be determined using Bayesian model selection (see paper I for details).

3 INSTRUMENT MODEL

To explore the effectiveness of the BAYESCAL framework for interferometric calibration derived in the preceding section, in Section 5 we demonstrate its applications to the absolute calibration of a simulated observed data set. In Section 4, we will define the diffuse and point source components of the complete and incomplete sky models used to construct our simulated observed visibility data set, V^{obs}, and calibration visibility model, V^{sim}, respectively. But first, here, we describe the instrument model we use to construct V^{obs} and V^{sim}, as well as the instrumental forward model component of our fitted sky model, V^{fit}.
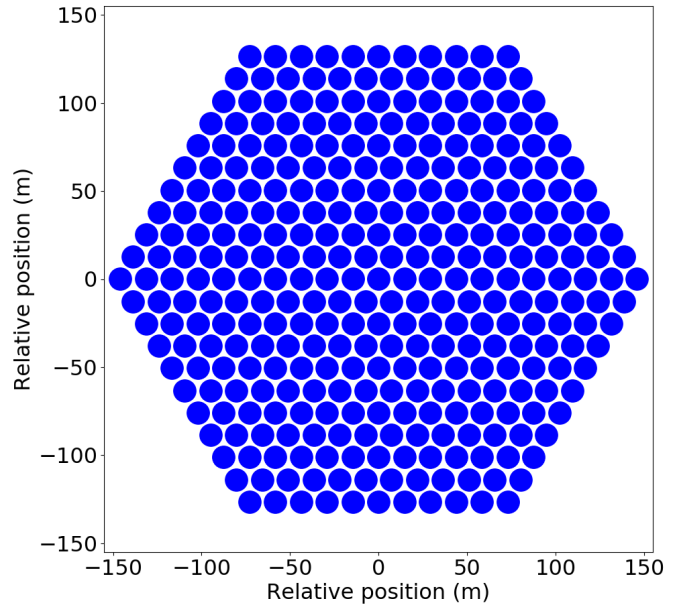

Figure 2. Simulated hexagonal close-packed antenna configuration with 331 antennas. Antenna positions plotted relative to $30°721S$ $21°411E$ at the SKA-South Africa site, Karoo desert.

3.1 Interferometric array

For our instrumental model, we use a 331 antenna close-packed hexagonal array configuration[9] with 14.6 m antenna spacing (see Figure 2), situated at the geographic location of HERA (DeBoer et al. 2017). We simulate a five minute observation, commencing at 02:52:00 UTC on the 1st of January 2020 and comprised of ten snapshot observations[10]. We select this commencement time so that a cold region of the Galaxy, lying away from the Galactic plane (zenith pointing RA = $164°42$, Dec = $-30°61$ at the start of the observation; see Figure 3, left) is passing through zenith[11]. The choice of $n_t = 10$ relatively long 30 s integration times maximises the calibration signal to noise, while minimising the computational expense associated with joint estimation of the calibration solutions over multiple integrations[12].

In general, astrophysical emission from all parts of the sky where the primary beam is non-zero contributes to the measured visibilities. Since the beam is non-zero for zenith angles above the horizon, ideally, one would include all discretised emission elements with zenith angle coordinates $\theta(t) \lesssim 90°0$ when calculating

---

[9] In the built HERA antenna layout, the hexagonal grid is fractured into three sectors and incorporates outrigger antennas for improved imaging capabilities (Dillon & Parsons 2016). The 331 antenna close-packed hexagonal array configuration that we use here is a simplified version of that array layout but preserves the feature relevant to calibration comparison carried out in Section 5.

[10] Here, we model the visibilities based on a snapshot at the center of the time step and add noise in the uv-domain appropriate for the stated integration time.

[11] This is a preferred choice of field for 21 cm cosmology because it reduces the dynamic range between the foreground emission and 21 cm signal, mitigating the level of foreground systematics that will be introduced to the calibrated data for a given level of spurious spectral structure in the calibration solutions.

[12] Nevertheless, we emphasise that the calibration formulation presented here can be applied to an arbitrary number of integrations, each with arbitrarily short integration time, subject to available computational resources.
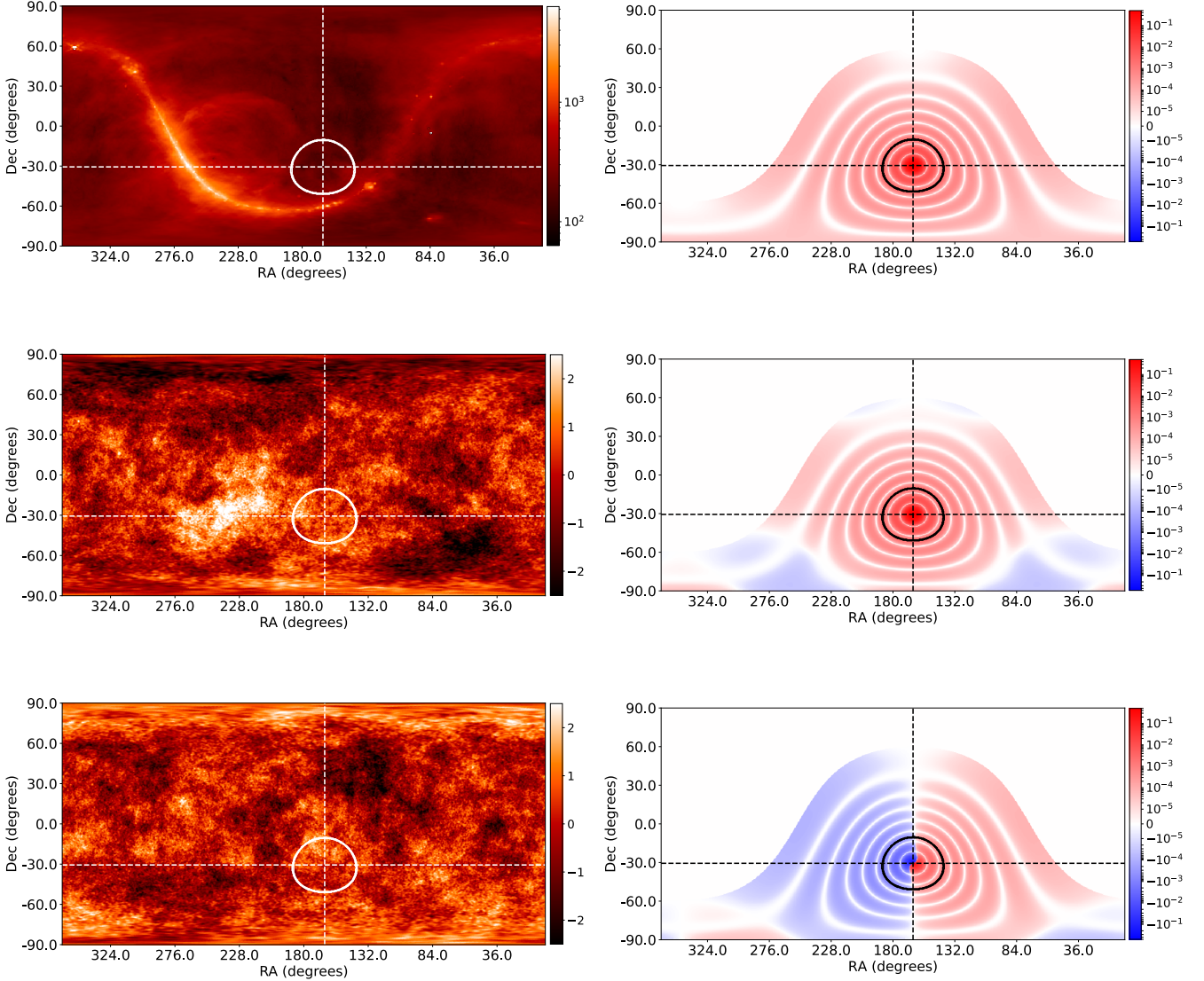
Figure 3. Diffuse emission brightness temperature in Kelvin (left) in Stokes I (top), Q (middle) and U (bottom) and polarised primary beam patterns (right), encoding the projections of Stokes I (top), Q (middle) and U (bottom) emission onto south-north oriented feeds of the antenna in the array, calculated from Equation 32 via Equation 28, at the longitude and latitude of our simulated array and LST of the start of our simulated observed data set. In each subplot, the intersections of the dashed white (left) and black (right) lines is at zenith for the simulated array in the first integration of the simulated observed data and the solid white (left) and black (right) circles show the perimeter of the 40°.0-diameter FoV used in the simulated observations.

$V(\mathbf{u}_{pq}, \nu, t)$. However, because the computational expense associated with calculating Υ^{-1} in Equation 24 scales approximately as N_{pix}^3, where N_{pix} is the number of sky-model parameters associated with V^{fit}, we limit our model to a restricted FoV centered on the RA and Dec of zenith at the longitude and latitude of our simulated array at each LST of our simulated observations. In the simulations considered in Section 5, we use a 40°.0-diameter FoV at each LST, which is approximately double the $\sim 20°.0$ diameter of the first null of our beam model (see Section 3.3). We have confirmed that repeating the analyses for FoVs smaller than the fiducial 40°.0 considered here does not affect the conclusions of this paper. We plan to explore computationally efficient means to extend the FoV of the BayesCal model in future work.

For this fixed FoV, at a fixed LST, t, we define $\mathbf{F}_{\mathrm{fr},t}$ as the non-

uniform discrete Fourier transform, to the uvw-domain coordinates sampled by our array, of our image-space flux-density distribution defined at a discrete set of lmn coordinates that describe either the pixel centers of our image cube, for our diffuse emission simulation, or the coordinates of the catalogued point sources, for our EGS simulation. As such, $\mathbf{F}_{\mathrm{fr},t}$ has elements,

$$F_{\mathrm{fr},t,ij} = \gamma \exp(-2\pi i(u_i l_j + v_i m_j + w_i n_j)) , \qquad (25)$$

where γ is a unit conversion factor which, (i) for our image domain brightness temperature diffuse emission simulation (see Section 4.1) in units of K, takes the form $\gamma = 2 \times 10^{26} \nu^2 k_{\mathrm{B}} \Delta\Omega / c^2$, with $\Delta\Omega$ the image domain pixel area and, (ii) for our EGS simulations (see Section 4.3) with source flux-densities in Jy takes a value of $\gamma = 1$. In both cases, we derive visibilities in units of Jy.

To derive lmn coordinates at a given LST, we follow the ap-

proach in HEALVIS (Lanman & Kern 2019; Lanman & Pober 2019). We begin by defining a 40°.0 FoV about the (RA,Dec)-coordinates of the LST-dependent zenith pointing at the longitude and latitude of the simulated observatory. We derive the *lmn* coordinates as,

$$l = \sin(\phi)\sin(\theta) \tag{26}$$
$$m = \cos(\phi)\sin(\theta)$$
$$n = \cos(\theta),$$

where θ and ϕ are the zenith angle and azimuth coordinates of the emission sources (point source coordinates or HEALPIX pixel centers in the EGS and diffuse emission components of our simulations, respectively) in our selected FoV. For each emission source, we calculate the zenith angle and azimuth coordinates as,

$$\theta = \arccos(\hat{s} \cdot \hat{e}_u) \tag{27}$$
$$\phi = \arctan(s_e, s_n).$$

Here, $\hat{s} = [s_n, s_e, s_u]^T$, $\hat{e}_n$, $\hat{e}_e$ and $\hat{e}_u$ are unit vectors pointing from the antenna to the (RA, Dec) coordinates of the emission source and north, east, and up relative to the antenna in the local tangent plane coordinate system, respectively.

The frequency dependent *uvw* coordinates of the visibilities sampled by the ith baseline in the array are given by $\boldsymbol{b}_i/\lambda$, where $\boldsymbol{b}_i$ is the ith baseline vector and λ is the central wavelength of the radiation measured by a given channel. When constructing our simulated observations to be calibrated in Section 5, we consider two cases:

(i) *The 14.5 − 290* m *(7.73 − 154.66 λ at 160* MHz*) 'long baseline' data set* – which includes visibilities calculated on the full range of baseline lengths in our simulated array and results in 630 uniquely sampled *uvw*-coordinates per channel.

(ii) *The 14.5 − 29* m *(7.73 − 15.47 λ at 160* MHz*) 'short baseline' data set* – which includes visibilities calculated on the three shortest baseline lengths in our simulated array and results in 9 uniquely sampled *uvw*-coordinates per channel.

There are several reasons why calibrating the short baseline subset of the visibilities is of interest. For the array layout considered here, these baselines have the highest redundancy and highest power spectrum sensitivities. This makes them a promising subset for 21 cm analyses that seek to constrain the redshifted 21 cm power spectrum on a per-baseline basis (e.g. Parsons et al. 2012). If one intends to use such a baseline subset for science analysis, the simplest approach to ensuring internal consistency of the calibration and power spectrum analysis is to derive calibration solutions for the data one intends to use from the corresponding raw data[13] (rather than from a superset or alternate baseline range subset of the data from the array).

Additionally, short baselines have intrinsically lower chromaticity, meaning that, for a fixed calibration model completeness as a function of baseline length, more accurate calibration solutions will be derived by up-weighting the calibration constraints on short baselines (e.g. Ewall-Wice, et al. 2017; Orosz et al. 2019).

Finally, in Bayesian approaches to recovering the three dimensional power spectrum of the redshifted 21 cm emission from interferometric data by jointly fitting for the 21 cm signal and foregrounds (see e.g. Sims et al. 2019), the computational efficiency

Table 1. Instrumental and observational parameters.

Parameter	Description	Value
B	Central frequency range	$160 - 169$ MHz
N_ν	Number of channels	10
$\Delta\nu$	Channel width	1 MHz
τ	Integration time	30 s
η_s	System efficiency	1
η_a	Antenna efficiency	1
A	Antenna area	150 m^2

of calculating the power spectrum posterior scales with area of the model *uv*-plane cubed. Thus, here, analysing the high sensitivity short baseline subset of visibilities in a HERA-like array configuration also represents a good compromise between data sensitivity and computation time.

Despite these considerations, we retain the long baseline data set as a reference and additional point of comparison when calibrating simulated observed data with either standard absolute calibration or absolute calibration with gain priors, since the computation time required to calibrate the long baseline data sets with these approaches scales simply with the data volume[14].

In both cases, we assume that the antennas lie in a plane at fixed height, such that $w = 0$, and we average over the visibilites in redundant baseline groups for the purposes of absolute calibration in Section 5.

3.2   Visibility noise

The noise level on a visibility resulting from a pair of identical antennas individually experiencing equal system noise is given by Equation 19. We list the values of the parameters used in our instrument simulation in Table 1. For these parameter values and assuming a constant system noise temperature of $T_{\text{sys}} = 200$ K across our 9 MHz bandwidth, Equation 19 yields a visibility noise $\sigma_V = 0.5$ Jy per integration prior to redundant baseline averaging, which we add independently to the real and imaginary components of each of the sampled visibilities.

3.3   Primary beam model

The linearly polarised primary beam response of the baseline between two antennas p and q to polarised sky emission parametrised in terms of the Stokes coherency vector is given by (see paper I for details),

$$\mathbf{P}_{pq} = \mathbf{E}_{pq}\mathbf{T}^{\text{S}}. \tag{28}$$

Here,

$$\mathbf{E}_{pq} = \mathbf{E}_p \otimes \mathbf{E}_q^*, \tag{29}$$

is a 4×4 matrix encoding the polarisation-, frequency- and direction-dependent primary beam response of the baseline to the autocorrelation of the vector electric field associated with the sky brightness distribution; $\mathbf{E}_p$ and $\mathbf{E}_q$ are the voltage beams of antennas p and q,

---

[13] A data set that is calibrated using only a subset of the data is subject to 'leverage' which manifests itself as excess noise in the component of the data omitted during calibration. See Patil et al. (2016, 2017) for a description of leverage in interferometric calibration.

[14] In contrast to using standard absolute calibration or absolute calibration with gain priors, the long baseline data set additionally necessitates a higher resolution fitted image-domain sky model when using BAYESCAL, increasing the inversion time of Υ in proportion to $N_{\text{pix,s}}^3$ (see Section 2.6).

respectively; $\otimes$ denotes a Kronecker product and the Stokes transformation matrix,

$$\mathbf{T}^S = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & i \\ 0 & 0 & 1 & i \\ 1 & -1 & 0 & 0 \end{pmatrix} \tag{30}$$

encodes the mapping between the interferometric coherency vector for a baseline comprised of linearly polarised antennas,

$$\mathbf{C} = \begin{pmatrix} I + Q \\ U + iV \\ U - iV \\ I - Q \end{pmatrix}, $$

and the Stokes coherency vector,

$$\mathbf{C}^S = \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix}, $$

such that $\mathbf{C} = \mathbf{T}^S \mathbf{C}^S$.

In general, $\mathbf{E}_P$ can be derived from electromagnetic simulations of the response of the antenna in an array, or measurements of the antenna response via holographic beam forming can be used to estimate the primary beam response associated with the baselines of an interferometric array.

However, for simplicity, in this work, we consider an analytic, zenith pointing, polarised Airy beam primary beam response of an array comprised of identical linearly polarised antennas with west-east and south-north oriented feeds with no polarisation leakage[15]. In an $(\hat{\mathbf{e}}_n, \hat{\mathbf{e}}_e)$ two-component vector basis of an east-north-up (ENU) coordinate system, this corresponds to a polarised voltage pattern of the form,

$$\mathbf{E}(\theta_j, \phi_j) = \frac{2J_1(ka\sin(\theta_j))}{ka\sin(\theta_j)} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} . \tag{31}$$

Here, J_1 is the Bessel function of the first kind of order one, θ_j and ϕ_j are the zenith angle and azimuth coordinates of the discretised emission element j (i.e. the sources in our EGS simulation and HEALPix pixels in our diffuse emission simulations), $k = 2\pi/\lambda$ is the wavenumber, with λ the wavelength of the emission, and a the aperture radius. In the simulations in this paper, we use an aperture radius $a = 7.25$ m, such that Equation 31 is a reasonable first-order approximation to the voltage beam of a 14.5 m diameter fixed dish antenna used by HERA.

While a vector basis tied to the orientations of the antenna feeds provides the simplest vector space for defining the polarised voltage beam, a sky-centric basis tied to the celestial sphere provides a more natural vector space in which to define the polarised sky models from which V^{true}, V^{sim} and V^{fit} are derived. In this work, we use an equatorial basis for this purpose. To evaluate Equation 6

---

[15] Here, we mean polarisation leakage to refer to voltage excited in nominally south-north and west-east oriented feeds by the east and north components of the vector electric field at the antenna (written in an ENU basis) E_e and E_n, respectively, as well as any coupling between the voltages along the signal chains between the feeds and the correlator. In contrast, we do account for instrumentally induced polarisation due to the different voltage beams of the antenna (see Equation 32). This polarisation manifests, for example, as a non-zero contribution of linearly polarised sky emission to pseudo-stokes I visibilities and, correspondingly, to an image reconstructed from these visibilities.

one must transition the beam and the sky to a mutual vector basis. To this end, we re-write Equation 31 with respect to an equatorial basis as,

$$\mathbf{E}(\theta_j, \phi_j) = \frac{2J_1(ka\sin(\theta_j))}{ka\sin(\theta_j)} \mathbf{T}_{\text{ENU−equatorial}} , \tag{32}$$

where $\mathbf{T}_{\text{ENU−equatorial}}$ is the transition matrix between the $(\hat{\mathbf{e}}_n, \hat{\mathbf{e}}_e)$ and $(\hat{\mathbf{e}}_{\text{Dec}}, \hat{\mathbf{e}}_{\text{RA}})$ vector bases,

$$\mathbf{T}_{\text{ENU−equatorial}} = \begin{pmatrix} \hat{\mathbf{e}}_n \cdot \hat{\mathbf{e}}_{\text{Dec}} & \hat{\mathbf{e}}_n \cdot \hat{\mathbf{e}}_{\text{RA}} \\ \hat{\mathbf{e}}_e \cdot \hat{\mathbf{e}}_{\text{Dec}} & \hat{\mathbf{e}}_e \cdot \hat{\mathbf{e}}_{\text{RA}} \end{pmatrix} . \tag{33}$$

The analytic form that Equation 33 takes, as a function of antenna location, hour and parallactic angle, is derived in A5.

Using Equation 28, our polarised primary beam matrix takes the form,

$$\mathbf{P}(\theta_j, \phi_j) = \left[\frac{2J_1(ka\sin(\theta_j))}{ka\sin(\theta_j)} \right]^2 \bar{\mathbf{T}} . \tag{34}$$

Here, $\bar{\mathbf{T}} = (\mathbf{T}_{\text{ENU−equatorial}} \otimes \mathbf{T}^*_{\text{ENU−equatorial}}) \mathbf{T}^S$, with $\mathbf{T}^S$ the Stokes transformation matrix given in Equation 30, and $\mathbf{P}(\theta_j, \phi_j)$ is a 4×4 matrix with elements,

$$\mathbf{P} = \begin{pmatrix} P_{\text{nn},I} & P_{\text{nn},Q} & P_{\text{nn},U} & P_{\text{nn},V} \\ P_{\text{ne},I} & P_{\text{ne},Q} & P_{\text{ne},U} & P_{\text{ne},V} \\ P_{\text{en},I} & P_{\text{en},Q} & P_{\text{en},U} & P_{\text{en},V} \\ P_{\text{ee},I} & P_{\text{ee},Q} & P_{\text{ee},U} & P_{\text{ee},V} \end{pmatrix} . \tag{35}$$

Figure 3, right, shows the projections on the sky of the $P_{\text{nn},I}$, $P_{\text{nn},Q}$ and $P_{\text{nn},U}$ elements of the polarised primary beam matrix, relevant to the V^{nn} visibilities calibrated in Section 5, calculated from Equation 34, at the longitude and latitude of our simulated array and LST of the start of our simulated observed data set.

3.4 Degenerate gain parameters

In addition to the array layout, instrument noise and primary beam models described in Sections 3.1 – 3.3, another component of the instrument model that must be defined is the antenna gains. As described in Section 2.2, for a redundant array, the instrument gains can be divided into redundant gains, solved for with relative calibration, and a smaller set of degenerate gain parameters that must be solved for with reference to a sky model in order to absolutely calibrate the visibilities. When calibrating the simulated observed data sets in Section 5, we will assume that the maximum likelihood redundant gain parameters have been determined, leaving only the degenerate gain parameters to be solved for. Correspondingly, when performing absolute calibration, we set the relative gain solutions in the calibration model equal to the expectation value of the solutions that would be obtained for these parameters with relative calibration (matching their values in the simulated data) and solve for the remaining gain parameters corresponding to the degenerate gain degrees of freedom that cannot be solved for with baseline-redundancy-based constraints.

For accurate relative calibration, as assumed here, the absolute calibration solutions are independent of the specific choice of relative calibration parameters. Therefore, for simplicity we set them to unity in both the simulated data and the relative calibration component of the calibration model. The degenerate gain amplitude parameters used to construct the simulated observed visibilites that are calibrated in Section 5 are described in Section 2.5 and shown in Figure 1, right, and we define the true degenerate tip-tilt phase parameters used to construct the simulated observed visibilities to

be zero in all channels of the data set. Accurate calibration in Section 5 corresponds to accurate recovery of these degenerate gain parameters by the calibration algorithm.

4 SKY MODEL

In this section, we define the diffuse and point source components of the full sky model used to construct V^{obs} and the incomplete versions of that sky model (assuming uncertainty on the diffuse emission model and a flux density limited catalogue comprising a subset of the full point source model) used to construct V^{sim}. We then apply the instrument models defined in Section 3 to those sky models to generate V^{obs} and V^{sim}.

4.1 Sky brightness temperature distribution

At the radio frequencies relevant to 21 cm cosmology observations ($\sim 10^2$–10^3 MHz), the time and frequency dependent polarised sky brightness temperature distribution observed by the interferometer, C, is dominated by Stokes I, which, in turn, is comprised of three major emission components: (i) GDSE radiated through the acceleration of high energy cosmic-ray electrons in the magnetic field of the Galaxy, (ii) synchrotron emission from extragalactic sources (EGS), and (iii) free-free emission from within the Galaxy.

Of these components, the power in GDSE and free–free emission as a function of angular scale on the sky is reasonably approximated by inverse power laws. Since the angular scale probed by a baseline is inversely proportional to its length, GDSE and free-free emission contribute most significantly to the visibilities on short baselines. Depending on the field observed, GDSE can be the dominant astrophyical contributor to the measured visibilities on baseline lengths up to several tens of meters at ~ 100 MHz (see e.g. Sims et al. 2016). The contribution from free-free emission is subdominant to GDSE on all angular scales in the relevant frequency range and is expected to account for approximately 1% of the sky temperature at 150 MHz (e.g. Shaver, et al. 1999). The EGS two-dimensional spatial power spectrum is approximately flat, subject to perturbations due to source clustering, and is the dominant contributor to the measured visibilities on baseline lengths $\gtrsim$ 10s of meters, at $\sim 10^2$ MHz, in regions excluding the brightest point sources on the sky.

Total intensity, Stokes I, emission makes up the dominant fraction of the sky temperature from these emission sources; however, polarised emission arises generically in processes involving magnetised plasmas, and the linear polarisation fraction of diffuse emission has been measured at levels between $\sim 1-5\%$ at 154 MHz (Lenc, et al. 2016) and $\sim 15-20\%$ at 408 MHz (Mathewson & Milne 1965).

To construct V^{obs}, we separate the emission components described above into two categories based on whether we parametrise our simulation of their contribution to V^{obs} in terms of a brightness distribution on a regular model grid or in terms of a set of source flux-densities defined at coordinates given by source surveys. In the first category, we include diffuse emission comprised of GDSE, diffuse free-free emission, and a diffuse sea of unresolved point sources below the completeness limit of cataloged sources in the region of sky under observation; in the second, we include emission from catalogued point sources.

4.2 Diffuse emission

4.2.1 Stokes I

We define our diffuse emission sky on a HEALPIX grid (Górski, et al. 2005). For our assumed a priori known component of the Stokes I diffuse emission, we use the 2016 version of the Global Sky Model (GSM; Zheng, et al. 2017), parametrised in terms of sky brightness temperature, T, as implemented in the PYGSM package (Danny C. Price 2016). Here, T is related to the spectral brightness distribution via the Rayleigh–Jeans law,

$$T(\nu) = \frac{c^2 B}{2\nu^2 k_{\mathrm{B}}}, \qquad (36)$$

with c and k_{B} the speed of light and Boltzmann constant, respectively.

When constructing our simulations, we use a spatial resolution parametrised by the HEALPIX N_{Nside} parameter, which is greater than or equal to double the spatial scale probed by the longest baselines in the data set to be calibrated[16], and consider a 10 channel data set with 1 MHz channel widths and central frequencies spanning the range $160 \leq \nu \leq 169$ MHz. This results in an image cube, T_{GSM}, comprised of N_ν full-sky GSM maps evaluated at the central frequencies of our simulated channels, each containing $N_{\mathrm{pix}} = 12N_{\mathrm{Nside}}^2$ pixels.

In order to construct a complete Stokes I simulation, in addition to estimating the Stokes I total intensity from the GSM model, one must account for the expected discrepancy between the GSM model for Stokes I emission and the true brightness distribution of Stokes I emission on the sky associated with the non-zero uncertainty level on the GSM model. To account for this, we construct our Stokes I sky simulation as the sum of T_{GSM} and an additional image cube, $T_{\delta I}$, where $T_{\delta I}$ is a model for the difference between the GSM model for Stokes I emission and the true brightness distribution of Stokes I emission that will be present in observational data. To construct $T_{\delta I}$, we assume that its intensity and power law index can be modelled by Gaussian random fields (GRFs). We begin by deriving a sky brightness temperature realisation for the central frequency of the cube, $T_{\delta I}(\nu_{\mathrm{c}})$, using the SYNFAST method of the HEALPY[17] package (Zonca et al. 2019), as a GRF realisation with an angular power spectrum proportional to that of $T_{\mathrm{GSM}}(\nu_{\mathrm{c}})$, as estimated with HEALPY ANAFAST. We scale the absolute power in $T_{\delta I}(\nu_{\mathrm{c}})$ such that $\mathrm{RMS}(T_{\delta I}(\nu_{\mathrm{c}})) = f * \mathrm{RMS}(T_{\mathrm{GSM}}(\nu_{\mathrm{c}}))$. Here, f is the estimate of the average fractional error in T_{GSM}, at frequency ν_{c}, associated with using the imperfect total intensity GSM map as a model of the Stokes I brightness temperature distribution. Given the uncertainty in the true value of f, we consider two values:

(i) $f = 0.5$, which we consider to be a conservative upper limit on the fractional error in approximating the true Stokes I brightness temperature distribution of diffuse emission with the imperfect Stokes I GSM sky model in a relatively cold region of sky, lying out of the Galactic plane, in the 160–169 MHz spectral range of our simulated spectral band. We use this value when constructing the $T_{\delta I}(\nu_{\mathrm{c}})$ contribution to the simulated true visibilities, for our high and moderate model incompleteness calibration scenarios (see Section 5 for more details).

---

[16] This requires $N_{\mathrm{Nside}} \geq 32$ for the $14.5 - 29$ m data set and $N_{\mathrm{Nside}} \geq 256$ for the $14.5 - 290$ m data set.
[17] https://healpy.readthedocs.io/en/latest/index.html

(ii) $f = 0.25$, which is a more modest estimate[18] of the fractional error in approximating the true Stokes I intensity distribution with the imperfect Stokes I GSM sky model. We use this value when constructing the $T_{\delta I}(\nu_c)$ contribution to our simulated true visibilities in our low model incompleteness calibration scenario.

An interferometer is insensitive to the frequency dependent sky monopole, therefore interferometric data calibration is correspondingly insensitive to the mean of $T_{\delta I}(\nu_c)$, except insofar as it influences the sky noise. Here, we set the mean of the GRF to zero. The foregrounds are detectable at a high signal-to-noise level in the instrumental array and observing setups that we will consider in this paper, and, as such, we expect the impact of this approximation to be small. However, in an observing regime where this was not the case, the approach taken here can be adapted to account for the impact of the mean sky temperature of the a priori unknown component of the emission on the noise level in the data (as well as other sources of uncertainty on the noise level) by jointly estimating the noise in the data with the other parameters of the analysis (see Sims et al. 2019 and Sims & Pober 2019) for applications of this approach in the context of power spectral estimation using interferometric data).

We model the spectral structure of $T_{\delta I}$ with a spatially dependent Gaussian spectral index distribution, $\beta(l, m, n)$, with mean and standard deviation consistent with measurements of the spectral index distribution of Galactic emission between 90 and 190 MHz with EDGES (Mozdzen et al. 2017): $\langle \beta \rangle_{\mathrm{GDSE}} = -2.6$ and $\sigma_{\beta_{\mathrm{GDSE}}} = 0.02$, respectively. Incorporating this spectral dependence, we construct our full brightness temperature image cube for this emission component as,

$$T_{\delta I}(l, m, n, \nu) = T_{\delta I}(l, m, n, \nu_c) \left(\frac{\nu}{\nu_c} \right)^{\beta(l,m,n)} . \quad (37)$$

4.2.2 Stokes Q, U and V

We follow an equivalent procedure to that described in Section 4.2.1 to construct our assumed a priori known components and complete Stokes Q and U diffuse emission skies. However, in place of the GSM model, we construct simulated sky brightness temperature maps for both our assumed a priori known and incomplete components of the Stokes Q and U diffuse emission. Specifically, we assume that the intensity of Stokes Q and U diffuse emission can be spatially modelled as a Gaussian random field (GRF). We construct GRF realisations of the emission intensity fields, $T_Q(l, m, n, \nu_c)$ and $T_U(l, m, n, \nu_c)$, respectively, with equal angular power spectra; in both cases, they are proportional to the angular power spectrum of our Stokes I model, scaled such that the linear polarisation fraction of the known component of the emission, I_L/I (with $I_L^2 = Q^2 + U^2$), is approximately 3%, consistent with measurements in the frequency range of our simulations by Lenc, et al. (2016). The central channels of the Stokes I, Q, and U maps, $T_{\mathrm{GSM}}(l, m, n, \nu_c)$, $T_Q(l, m, n, \nu_c)$, $T_U(l, m, n, \nu_c)$, respectively, used for constructing the full array simulated visibilities are shown in Figure 3, left. We construct full Stokes Q and U image cubes, $T_Q(l, m, n, \nu)$ and $T_U(l, m, n, \nu)$, respectively, following an

analogous spectral extrapolation to that described by Equation 37, now applied to $T_Q(l, m, n, \nu_c)$ and $T_U(l, m, n, \nu_c)$. Finally, we assume negligible astrophysical emission in Stokes V ($V/I \ll 1$).

To construct $T_{\delta Q}$ and $T_{\delta U}$ we use the same approach as that described for deriving $T_{\delta I}$ in Section 4.2.1 and, for simplicity, assume the same fractional uncertainties, in a given calibration model completeness scenario, for all three Stokes parameters.

4.2.3 Complete simulated observed data and incomplete simulated calibration model

From our diffuse emission cubes, we construct simulated true visibilities corresponding to a single time integration as,

$$V_{\mathrm{diffuse}, t_i}^{\mathrm{true}} = \mathbf{F}_{\mathrm{fr}, t_i} [\mathbf{P}_{\mathrm{nn}, I, t_i} (T_{\mathrm{GSM}, t_i} + T_{\delta I, t_i})$$
$$+ \mathbf{P}_{\mathrm{nn}, Q, t_i} (T_{Q, t_i} + T_{\delta Q, t_i}) + \mathbf{P}_{\mathrm{nn}, U, t_i} (T_{U, t_i} + T_{\delta U, t_i})] , \quad (38)$$

where, as described in Section 3.1, we limit our simulation to emission within a $40°.0$ field encompassing the highest beam weighted fraction of the sky emission and centered on the RA and Dec of zenith at the longitude and latitude of our simulated array, at each LST of our simulated observations. T_{GSM, t_i}, $T_{\delta I, t_i}$, T_{Q, t_i}, $T_{\delta Q, t_i}$, T_{U, t_i} and $T_{\delta U, t_i}$ are vectorisations of these subsets of our polarised image cubes, each of length $N_{\mathrm{pix,s}} \times N_\nu$, with $N_{\mathrm{pix,s}}$ the number of pixels at the resolution of the simulation, per frequency channel, falling within our $40°.0$ subset fields.

Similarly, we construct simulated model visibilities, corresponding to a single time integration, but now exclusively from our assumed a priori known diffuse emission cubes, as,

$$V_{\mathrm{diffuse}, t_i}^{\mathrm{sim}} = \mathbf{F}_{\mathrm{fr}, t_i} [\mathbf{P}_{\mathrm{nn}, I, t_i} T_{\mathrm{GSM}, t_i} + \mathbf{P}_{\mathrm{nn}, Q, t_i} T_{Q, t_i} + \mathbf{P}_{\mathrm{nn}, U, t_i} T_{U, t_i}] . \quad (39)$$

To derive versions of $V_{\mathrm{diffuse}}^{\mathrm{true}}$ and $V_{\mathrm{diffuse}}^{\mathrm{sim}}$ incorporating all time integrations of our observation, we simply substitute instances of the T_{GSM, t_i} and $T_{\delta \mathrm{diff}, t_i}$ image vectors for single time integrations with their concatenation over all time integrations and substitute instances of $\mathbf{F}_{\mathrm{fr}, t_i}$ and $\mathbf{P}_{t_i}$ with an analogous pair of block diagonal matrices defined such that each block is given by $\mathbf{F}_{\mathrm{fr}, t_i}$ or $\mathbf{P}_{t_i}$ at a single time integration. When implementing these matrices, as well as matrices associated with the instrument model and gain parameters, we use the SCIPY sparse matrix linear algebra package[19] for all diagonal and block diagonal matrices, to minimise their memory footprints[20].

4.3 Point source emission

We use the sources catalogued in the Murchison Widefield Array GLEAM survey[21] (Wayth et al. 2015; Hurley-Walker et al. 2017) as the basis of our point source emission simulation.

When constructing our point source emission simulation, we use GLEAM integrated fluxes and their associated uncertainties over the 95–103, 147–154 and 185–193 MHz spectral ranges, which we denote here as $S_{\mathrm{int},99}$, $S_{\mathrm{int},151}$, and $S_{\mathrm{int},189}$ and $\sigma_{S_{99}}$, $\sigma_{S_{151}}$, and

---

[18] Work on an improved version of the GSM that will provide an estimate of the uncertainty on the model at a given interpolation frequency indicates that uncertainty on the improved GSM map is $\sim 10\%$ for emission in the frequency range considered here (Adrian Liu, private communication). Assuming comparable errors on the Zheng, et al. (2017) GSM map, the uncertainties on the Stokes I diffuse model considered here are conservative.

[19] https://docs.scipy.org/doc/scipy/reference/sparse.html
[20] For the baseline ranges and instrument model considered here, use of sparse matrices reduces the memory footprint of the calibration by 3–6 orders of magnitude.
[21] https://vizier.u-strasbg.fr/viz-bin/VizieR-3?-source=VIII/100/gleamegc

$\sigma_{S_{189}}$, respectively. We discard any sources that do not have positive, greater than two sigma detections[22] in $S_{\mathrm{int},99}$, $S_{\mathrm{int},151}$ and $S_{\mathrm{int},189}$.

We assume that the spectra of the catalogued sources can be approximated as a power law of the form: $S \propto \nu^{-\alpha}$, and we calculate the 99–189 MHz source spectral indices as,

$$\alpha_{99-189} = -\log(S_{\mathrm{int},189}/S_{\mathrm{int},99})/\log(189/99) . \quad (40)$$

We subsequently define the integrated flux densities of the point sources in our EGS simulation, over the 1 MHz spectral channels in the 160–169 MHz spectral range of our simulated spectral band, as,

$$S(\nu) = S_{\mathrm{int},151} \left(\frac{\nu}{151\,\mathrm{MHz}} \right)^{-\alpha_{99-189}} . \quad (41)$$

When we construct $V^{\mathrm{true}}_{\mathrm{EGS},t_i}$, we include sources that meet the criteria described above and that fall within a $40°.0$ field (see Section 3.1) encompassing the highest beam weighted fraction of the sky emission and centered on the RA and Dec of zenith at the longitude and latitude of our simulated array at time t_i,

$$V^{\mathrm{true}}_{\mathrm{EGS},t_i} = \mathbf{F}_{\mathrm{fr},t_i} \mathbf{P}_{\mathrm{nn},I,t_i} S_{\mathrm{EGS},t_i} . \quad (42)$$

Here, S_{EGS,t_i} is the vectorisation of the flux densities of all sources meeting these criteria.

In a practical calibration scenario, to maximise the completeness of the calibration model, one would incorporate all catalogued sources in $V^{\mathrm{sim}}_{\mathrm{EGS},t_i}$. In this case, the model EGS visibilities would be incomplete due to the missing contribution from sources below the completeness limit of the catalogue. For the simulations considered here, to avoid needing to use a hypothetical catalogue of sources below the completeness limit of GLEAM to construct $V^{\mathrm{true}}_{\mathrm{EGS},t_i}$ and then neglecting these additional sources when constructing an incomplete $V^{\mathrm{sim}}_{\mathrm{EGS},t_i}$, we instead take the approach used in e.g. Barry, et al. (2016) and use GLEAM sources to construct $V^{\mathrm{true}}_{\mathrm{EGS},t_i}$ and a subset of GLEAM sources to construct $V^{\mathrm{sim}}_{\mathrm{EGS},t_i}$.

When constructing $V^{\mathrm{sim}}_{\mathrm{EGS},t_i}$ from a subset of the sources included in $V^{\mathrm{true}}_{\mathrm{EGS},t_i}$, in addition to the cuts associated with $V^{\mathrm{true}}_{\mathrm{EGS},t_i}$, we impose a flux density cut of the form: $S_{\mathrm{int},151} > S_{\mathrm{min}}$. We consider two cases for this minimum flux density cut,

(i) $S_{\mathrm{min}} = 1$ Jy, which we associate with our high foregrounds incompleteness model, and

(ii) $S_{\mathrm{min}} = 100$ mJy, which we associate with our moderate and low foregrounds incompleteness models.

In both cases, we construct our EGS simulated visibilities as,

$$V^{\mathrm{sim}}_{\mathrm{EGS},t_i} = \mathbf{F}_{\mathrm{fr},t_i} \mathbf{P}_{\mathrm{nn},I,t_i} S_{\mathrm{EGS,s},t_i} , \quad (43)$$

where $S_{\mathrm{EGS,s},t_i}$ is the vectorisation of the flux densities of the subset of sources meeting the selection criteria for case (i) or (ii), depending on the model incompleteness of the calibration scenario considered. We assume that the polarised point sources contribution to the visibilities is small relative to unpolarised sources (e.g. Riseley et al. 2020) and, thus, do not include equivalent polarised source terms in Equation 43.

We generalise Equation 43 to multiple time integrations in an analogous manner to the multi-integration diffuse model in Section 4.2. To derive versions of $V^{\mathrm{true}}_{\mathrm{EGS}}$ and $V^{\mathrm{sim}}_{\mathrm{EGS}}$ incorporating all time integrations of our observation, we simply substitute instances

---

of the S_{EGS,t_i} and $S_{\mathrm{EGS,s},t_i}$ source flux density vectors for single time integrations with their concatenation over all time integrations and substitute point source model instances of $\mathbf{F}_{\mathrm{fr},t_i}$ and $\mathbf{P}_{\mathrm{nn},I,t_i}$ with an analogous pair of block diagonal matrices defined such that each block is given by $\mathbf{F}_{\mathrm{fr},t_i}$ or $\mathbf{P}_{\mathrm{nn},I,t_i}$ at the ith single time integration.

Over the observing duration of the data sets calibrated in Section 5, for the two point source model completeness cases described above, $V^{\mathrm{sim}}_{\mathrm{EGS}}$ includes approximately 940 and 11,000, respectively, of the 14,300 sources that fall within the union of the $40°.0$ fields centered on zenith at the 10 LSTs of our 5 minute data set. This corresponds to the inclusion of 41% and 95% of the total flux in $V^{\mathrm{true}}_{\mathrm{EGS}}$, respectively.

5 RESULTS

In order to test the efficacy of the BAYESCAL calibration methodology, in this section, we demonstrate its application to simulated observed data, constructed as described in Sections 3 and 4, and compare the inferred gains to those recovered using standard absolute gain calibration of redundantly calibrated visibilities (Section 2.4) and absolute gain calibration supplemented with a prior on the level of power in spectral fluctuations in the degenerate gain amplitude (Section 2.5).

For a detailed description of the mathematical formalism underpinning the three approaches, we refer the reader to paper I, or, alternatively, for a summary, to Section 2.6 of this paper. However, in short, absolute gain calibration supplemented with a prior on the level of power in spectral fluctuations in the degenerate gain amplitude can be viewed as an extension of standard absolute calibration of redundantly calibrated visibilities that incorporates physically motivated informative priors on the spectral structure of the degenerate gain amplitude solutions, sourced from external constraints such as electric and electromagnetic co-simulations of the receiver system and reflectometry measurements of the feed-dish system that are independent of an astrophysical sky model. In both standard absolute calibration or absolute gain calibration with gain priors the calibration visibility model is impacted by incompleteness and uncertainties associated with our a priori knowledge of the brightness distribution of the sky. In contrast, BAYESCAL, which in turn can be viewed as an extension to the absolute gain calibration with gain priors framework, explicitly incorporates a fitted visibility model, V^{fit}, for the component of the true observed visibilites that is unaccounted for in the simulated calibration visibility model, V^{sim}, used in the first two calibration approaches. In the limit that the spectral and instrument forward model used in V^{sim} and V^{fit} is accurate, the joint use of the two models allows one to construct a complete calibration visibility model.

For all three calibration methodologies, we demonstrate the effect on the recovered calibration solutions of three levels of incompleteness of the a priori known component of the calibration model, V^{sim}, corresponding to different levels of uncertainty on the distribution of low frequency radio emission in the field to be calibrated: high, medium and low. The uncertainties of the diffuse emission models and minimum flux densities of the point sources included in these models are summarised in Table 2. The number of sky-model parameters, $N_{\mathrm{pix,s}}$, associated with V^{fit} scales quadratically with the angular resolution of the model, which, in turn, is proportional to the maximum baseline length of the data set to be calibrated. The computational complexity associated with inverting Υ scales as $N^3_{\mathrm{pix,s}}$. For this reason, for each level of simulated

---

[22] For the data calibrated in Section 5, this amounts to discarding $\sim 35\%$ of the sources.

calibration sky-model incompleteness, we apply relative calibration plus BAYESCAL (Section 2.6) exclusively to the $14.5 - 29$ m data set. In contrast, the computational complexity associated with relative plus absolute calibration, both with (Section 2.5) and without (Section 2.4) placing a prior on the gain solutions, scales linearly with data volume. Thus, we derive calibration solutions for both the $14.5 - 29$ m and $14.5 - 290$ m data sets with these methodologies. For all calibration types, we assume the calibration solutions are stable in time over the five minute duration of the data set (see Section 2.7).

For all data calibration in this section, when using either absolute calibration or BAYESCAL, we assume the interferometric array is redundant and that the maximum likelihood redundant gain parameters have been determined, leaving only the degenerate gain parameters to be solved for (see Section 3.4).

5.1 Comparison between BAYESCAL and absolute calibration of redundantly calibrated visibilities

Figure 4 shows recovered calibration solutions in the low, medium and high simulated visibility model incompleteness regimes, from left to right. In each case, five sets of calibration solutions are displayed: BAYESCAL solutions for the $14.5 - 29$ m data set (purple lines), absolute calibration solutions for the $14.5 - 29$ m and $14.5 - 290$ m data sets without gain priors (blue and red lines, respectively) and absolute calibration solutions for the $14.5 - 29$ m and $14.5 - 290$ m data sets with gain priors (cyan and orange lines, respectively).

The top panel of each of the sub-figures, displays:

(i) the ratio of the input degenerate calibration amplitudes and the mean of the functional posterior probability distribution of the degenerate gain amplitudes, $\left\langle \Pr(A_{\text{true}}/A \mid A_{\text{F}}, V^{\text{obs}}) \right\rangle$, recovered when using BAYESCAL to fit the sum of a simulated model for the known flux contributions from diffuse emission and point sources and a statistical model for the a priori unknown component of sky emission, jointly with instrumental gain parameters, to visibilities with $14.5–29$ m baselines (solid purple lines),

(ii) the ratio of the input degenerate calibration amplitudes and the maximum a posteriori degenerate gain amplitudes, $A_{\text{true}}/\hat{A}_1$, when fitting an incomplete simulation of the known flux contributions from diffuse and point sources, to $14.5–29$ m baselines or $14.5–290$ m baselines, while imposing priors on the amplitude of spectral fluctuations in the gain amplitude solutions (solid cyan and orange lines, respectively), and

(iii) the ratio of the input degenerate calibration amplitudes and the maximum likelihood degenerate gain amplitudes, $A_{\text{true}}/\hat{A}_2$, when only fitting an incomplete simulation of the known flux contributions from diffuse and point sources to $14.5–29$ m baselines (solid blue lines) or $14.5–290$ m baselines (solid red lines).

Going forward, we use $\hat{A}$ as a stand-in variable for both the maximum a posteriori and maximum likelihood degenerate gain amplitudes, but, in each case, whether one or both variables are being referred to and, if only one, which one is being referred to will be described explicitly or will be clear from context. The dashed black lines represent the unitary value of this ratio that would be recovered by hypothetical data calibration in the limit that the signal-to-noise ratio on the data tends to infinity and the calibration model is perfect.

The bottom panel of each of the sub-figures, displays:

(i) the mean of the functional posterior probability distribution of the degenerate gain tip-tilt phases recovered with

BAYESCAL: $\left\langle \Pr(\boldsymbol{\Phi}_l \mid \boldsymbol{\Phi}_{l,\text{F}}, V^{\text{obs}}) \right\rangle$ (dot-dashed purple lines) and $\left\langle \Pr(\boldsymbol{\Phi}_m \mid \boldsymbol{\Phi}_{m,\text{F}}, V^{\text{obs}}) \right\rangle$ (dotted purple lines), respectively,

(ii) the maximum a posteriori degenerate tip and tilt phases, $\hat{\boldsymbol{\Phi}}_{l,1}$ (dot-dashed cyan and orange lines on $14.5–29$ m and $14.5–290$ m baselines, respectively) and $\hat{\boldsymbol{\Phi}}_{m,1}$ (dotted cyan and orange lines on $14.5–29$ m and $14.5–290$ m baselines, respectively), recovered with absolute calibration while imposing priors on the amplitude of spectral fluctuations in the gain amplitude solutions,

(iii) the maximum likelihood degenerate tip and tilt phases, $\hat{\boldsymbol{\Phi}}_{l,2}$ (dot-dashed blue and red lines on $14.5–29$ m and $14.5–290$ m baselines, respectively) and $\hat{\boldsymbol{\Phi}}_{m,2}$ (dotted blue and red lines on $14.5–29$ m and $14.5–290$ m baselines, respectively), recovered with absolute calibration alone.

As with the degenerate gain amplitudes, going forward, we use $\hat{\boldsymbol{\Phi}}_l$ and $\hat{\boldsymbol{\Phi}}_m$ as stand-in variables for both the maximum a posteriori and maximum likelihood degenerate tip-tilt phases, but, in each case, whether one or both variables are being referred to and, if only one, which one is being referred to will be described explicitly or will be clear from context. The black dashed line shows the null input value of the tip-tilt phases in the simulated observed data.

When comparing the recovered calibration solutions shown in Figure 4, it is useful to separate errors relative to the true calibration parameters associated with the simulated observed data set into two classes: (i) the mean bias in the recovered parameter estimates and (ii) the level of spurious spectral fluctuations imparted to the data by imperfect estimation of the gain parameters. In the first case, the average bias in the redundant gain amplitude parameters acts as a multiplicative scale factor to the data. As such, it is comparatively benign in 21 cm cosmology analysis, because it does not complicate the spectral separation of the foregrounds and the 21 cm signal in the data[23]. In contrast, in the second case, spurious spectral fluctuations imparted to the data by imperfect estimation of the redundant gain amplitude parameters introduce foreground systematics that will bias estimation of the 21 cm signal on the spectral scales of those fluctuations; thus, mitigating spurious spectral fluctuations in the calibrated data is of paramount importance.

5.1.1 Bias in gain parameter estimates

The mean biases in the ratio of the recovered and true degenerate gain amplitude solutions and those in the recovered estimates of the tip-tilt phases, relative to the null tip-tilt phases of the input simulated observed data, are summarised in Table 3, for each level of calibration model incompleteness, baseline length range, and for both BAYESCAL and absolute calibration. The average absolute biases when using absolute calibration both with and without imposing an informative gain amplitude prior are comparable in all cases; thus, in this subsection references to absolute calibration should be read as encompassing both cases.

The largest average absolute biases in the degenerate tip-tilt

---

[23] The average bias in the redundant gain amplitude parameters is less likely to entirely preclude recovery of the 21 cm signal relative to spurious spectral fluctuations imparted to the data by imperfect estimation of the redundant gain amplitude parameters. Nevertheless, we note that the largest average bias in the redundant gain amplitude parameters is $\sim 6\%$, which, following calibration, corresponds to misestimation of the amplitude of the power spectrum at a $\sim 12\%$ level. Thus, the average bias in the redundant gain amplitude can also make a contribution to the power spectrum error budget at a level that is significant for precision cosmology applications.

Table 2. Simulated calibration model, V^{sim}, completeness levels.

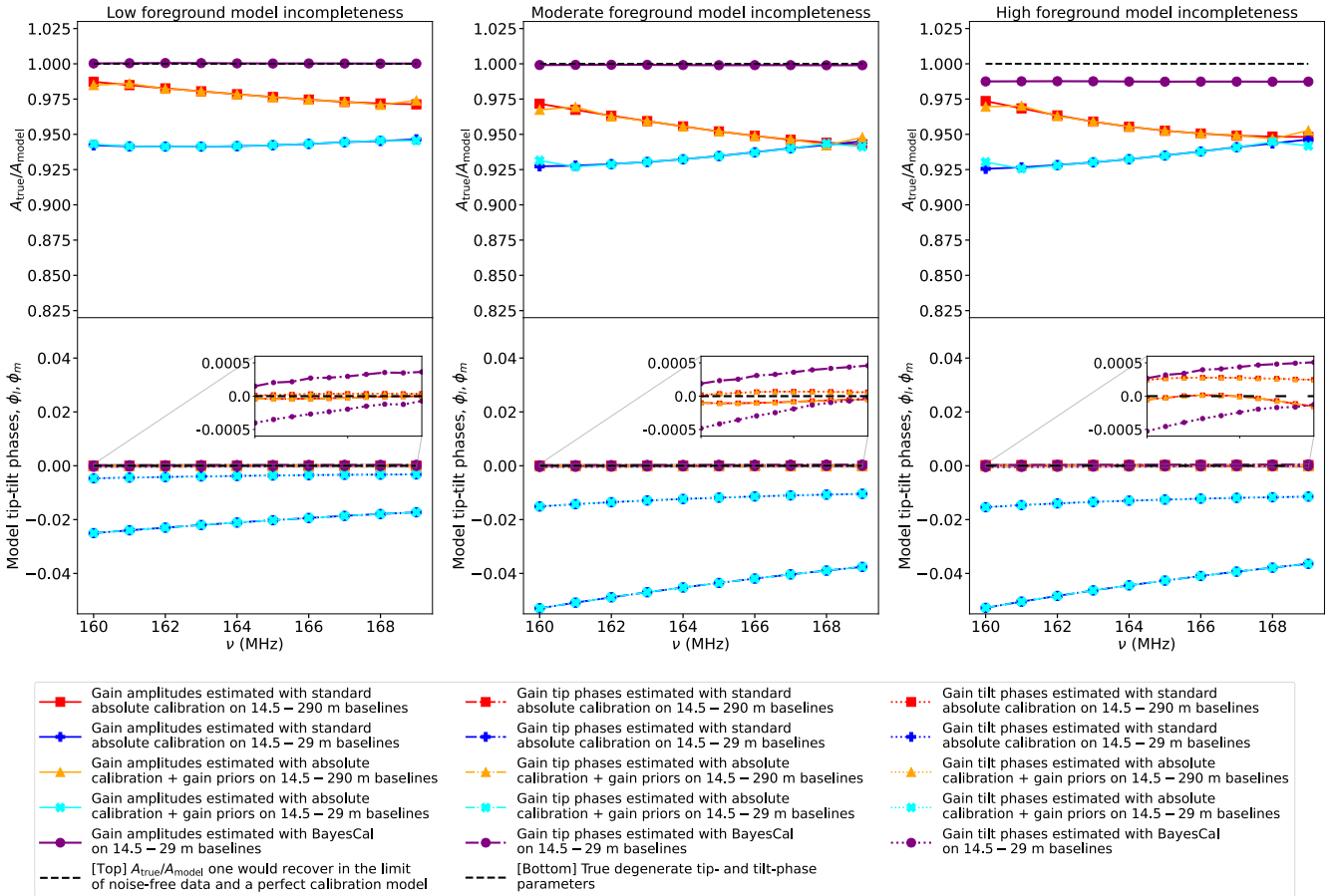| Model name	Fractional error on diffuse emission in V^{sim}	S_{min} (Jy) of sources included in V^{sim}
low incompleteness	0.25	0.1
moderate incompleteness	0.5	0.1
high incompleteness	0.5	1

Figure 4. Calibration solutions derived in three regimes of incompleteness of the simulated component of the calibration model (see Table 2): low (left), moderate (middle) and high incompleteness (right). [Top] Ratio of the input degenerate calibration amplitudes and i) the mean BAYESCAL estimates (see main text for details) of the degenerate calibration amplitudes (purple lines), ii) the absolute calibration solutions for the degenerate calibration amplitudes when calibrating data on 14.5–29 m baselines or 14.5–290 m baselines (blue and red lines, respectively), or iii) the absolute calibration with gain priors solutions for the degenerate calibration amplitudes when calibrating data on 14.5–29 m baselines or 14.5–290 m baselines (cyan and orange lines, respectively). [Bottom] The BAYESCAL solutions for the degenerate tip-tilt phases (dot-dashed and dotted purple lines, respectively) and the maximum a posteriori and maximum likelihood tip-tilt phases recovered with absolute calibration of visibilities, with and without imposition of informative gain priors, on 14.5–29 m baselines (cyan and blue dot-dashed and dotted lines, respectively) and 14.5–290 m baselines (orange and red dot-dashed and dotted red lines, respectively). The zoomed inset axes display the low-level structure in the BAYESCAL and maximum a posteriori and maximum likelihood tip-tilt phases recovered with absolute calibration of visibilities, with and without imposition of informative gain priors, on 14.5–290 m baselines. We note that the maximum a posteriori and maximum likelihood parameters recovered with absolute calibration of visibilities, with and without imposition of informative gain priors, have only minor differences in their large spectral scale power and biases and thus red and orange lines and cyan and blue lines, respectively, overlay each other to a significant degree in both the top and bottom sub-plots. The dashed black lines in the top sub-plots show the true unitary value of the ratio of the true and fitted degenerate gain amplitudes, which would be recovered by hypothetical data calibration if the data was noise-free and the calibration model was perfect. In the bottom sub-plots, the dashed black lines display the true null value of the input degenerate tip-tilt phase solutions in the simulated observed data.

phase parameters are found in estimates recovered with standard absolute calibration of the 14.5 − 29 m data set, in which case the average absolute tip-tilt phase offset ranges from a few times 10^{-3} rad, at best, for the degenerate tilt phase, paired with a larger $\sim 2 \times 10^{-2}$ rad offset in average absolute tip phase, when the low

incompleteness simulated calibration visibility models is used in the calibration, to greater than 10^{-2} rad in both tip and tilt phase when either moderate or high incompleteness simulated calibration visibility models are used. Relative to the degenerate tip-tilt phase solutions recovered for the short-baseline data set using absolute

Table 3. Average absolute bias in gain parameter estimates, expressed as fractional errors, $1 - \langle |x_i/x_{i,\mathrm{ideal}}| \rangle$, for the ratio of the true and recovered degenerate gain amplitude parameters and absolute errors, $\langle |x_i - x_{i,\mathrm{ideal}}| \rangle$, for the tip and tilt gain phases. In both cases, x_i corresponds to the ith parameter in the parameter column and $x_{i,\mathrm{ideal}}$ corresponds to the value of x_i that would be obtained in the limit that one calibrated noiseless data with a perfect calibration model.

Parameter	Calibration method	Baseline length range (m)	Calibration model incompleteness	Average absolute bias
$A_{\mathrm{true}}/\hat{A}$	Absolute calibration	14.5–290	low	2.19%
			moderate	4.49%
			high	4.32%
		14.5–29	low	5.70%
			moderate	6.54%
			high	6.54%
$\langle \mathrm{Pr}(A_{\mathrm{true}}/A \mid A_{\mathrm{F}}, V^{\mathrm{obs}}) \rangle$	BAYESCAL	14.5–29	low	0.03%
			moderate	0.09%
			high	1.25%
$\hat{\Phi}_l$	Absolute calibration	14.5–290	low	2.4×10^{-5} rad
			moderate	8.3×10^{-5} rad
			high	4.8×10^{-5} rad
		14.5–29	low	2.1×10^{-2} rad
			moderate	4.5×10^{-2} rad
			high	4.4×10^{-2} rad
$\langle \mathrm{Pr}(\Phi_l \mid \Phi_{l,\mathrm{F}}, V^{\mathrm{obs}}) \rangle$	BAYESCAL	14.5–29	low	2.8×10^{-4} rad
			moderate	3.4×10^{-4} rad
			high	4.1×10^{-4} rad
$\hat{\Phi}_m$	Absolute calibration	14.5–290	low	3.0×10^{-5} rad
			moderate	5.5×10^{-5} rad
			high	2.6×10^{-4} rad
		14.5–29	low	3.8×10^{-3} rad
			moderate	1.2×10^{-2} rad
			high	1.3×10^{-2} rad
$\langle \mathrm{Pr}(\Phi_m \mid \Phi_{m,\mathrm{F}}, V^{\mathrm{obs}}) \rangle$	BAYESCAL	14.5–29	low	2.2×10^{-4} rad
			moderate	2.3×10^{-4} rad
			high	2.9×10^{-4} rad

calibration, both with and without an informative prior on the degenerate gain amplitudes, the high resolution information available in the $14.5-290$ m data set results in a significant reduction in biases in the average absolute degenerate tip-tilt phase offsets recovered with the same calibration approaches.

In general, a less complete calibration model will result in larger tip-tilt phase biases[24]. In fitting the calibration model to the data, one expects bias in the recovered degenerate gain solutions to be a function of signal-to-noise weighted calibration model completeness, which is itself a function of baseline length. The visibility phase bias ($b \cdot (\Phi - \Phi_{\mathrm{true}})$) imparted to a model visibility by a given degenerate tip-tilt phase vector bias, ($\Phi - \Phi_{\mathrm{true}}$), is linearly dependent on the baseline length on which the visibility is measured. The greater sensitivity of long baselines to degenerate tip-tilt phase bias thus results in reduced tip-tilt phase bias, for a given calibration model completeness, when the data set includes longer baselines. Additionally, for the data sets calibrated here, the increased completeness of the model on long baselines due to the

relatively more complete point source model, when compared with the diffuse model, and the larger extent to which the diffuse emission is resolved out on long baselines of the $14.5 - 290$ m data set further act to reduce the average absolute biases in the degenerate tip-tilt phase parameters, relative to those recovered from the $14.5 - 29$ m data set.

The level of average absolute bias in the degenerate tip-tilt phase offsets recovered when calibrating the $14.5 - 29$ m data set with BAYESCAL is 1–2 orders of magnitude lower than in those recovered with absolute calibration, with or without the imposition of an informative degenerate gain amplitude prior, when performed on the same data set in all three calibration model incompleteness regimes, and is within an order of magnitude of the level of average absolute bias in the degenerate tip-tilt phase offsets achieved with absolute calibration on $14.5 - 290$ m. This improvement with BAYESCAL results from the greater completeness of the combined $V^{\mathrm{sim}} + V^{\mathrm{fit}}$ BAYESCAL calibration model, with sufficiently high completeness achieved to compensate for the loss of resolution associated with calibrating a short-baseline data set.

The level of the detrimental effect one can expect biases in the tip-tilt phases to have on 21 cm power spectrum estimation is dependent on the extent to which visibility phase information is used in the power spectrum estimator. Existing proposals for the analysis of calibrated data sets for 21 cm cosmology applications can be classified into two distinct families, based on whether they estimate the power spectrum of the measured or reconstructed sky (Morales et al. 2019). When the power spectrum of the measured sky is estimated, for example, via the delay transform of the visibilites

[24] On average, a less complete calibration model will result in larger gain errors. However, owing to an interferometric array's incomplete sampling of the uv-plane and constructive and destructive interference of sky emission in the visibilities, in addition to the level of calibration model completeness, the gain error is dependent on (i) the spatial distribution of the observed emission and calibration sky model and (ii) the array layout and null space of the instrument model. As a result of these additional dependencies, gain error is not guaranteed to be a monotonic function of calibration model completeness.

along the frequency axis (e.g. Parsons et al. 2012), the phase of the visibilities is discarded and, thus, the recovered signal estimates are robust to biases in the degenerate tip-tilt phase parameters. In contrast, in approaches that forward model or reconstruct the signal (e.g. Sims et al. 2019 and Barry et al. 2019, respectively), bias, and especially spurious spectral fluctuations, in the degenerate tip-tilt phase solutions have the potential to introduce foreground systematics into the recovered 21 cm signal estimates by shifting apparent source positions as a function of frequency and, in so doing, imparting non-smooth spectral structure along any given line of sight.

In the context of recovered sky power spectrum estimators, if one were to use absolute calibration of the short-baseline data sets, such as those considered here, for 21 cm cosmology applications, the biases in the degenerate tip-tilt phase parameters are as large as several degrees, with gradients of order a degree over the 9 MHz bandwidth of the data set. Such large offsets and gradients will bias the reconstructed sky model and have the potential to introduce additional spectral structure in recovered sky power spectrum estimators that will bias their power spectrum estimates if not accounted for. In contrast, the greatly reduced degenerate tip-tilt phase biases found with BayesCal correspond to shifts in source positions of order an arcminute when the high and moderate incompleteness simulated calibration visibility models are used in the calibration and sub-arcminute shifts in source positions for the low incompleteness simulated calibration visibility model. In both cases, in the frequency range relevant to 21 cosmology during CD and the EoR ($\nu \lesssim 200$ MHz), these shifts are 1–2 orders of magnitude smaller than the angular scales probed by the 10s of metre baselines on which interferometric experiments are generally attempting to extract the 21 cm signal. The amplitude of fluctuations in the phase parameters is comparable or smaller than this level. Thus, it is unclear how challenging such fluctuations will be for analyses in this category. Resolving this question is likely to require investigation on a per-analysis basis.

The maximum bias in $A_{\text{true}}/\hat{A}$ occurs when calibrating the $14.5 - 29$ m data set using absolute calibration, with a bias of $\sim 6.5\%$ when calibrating data with the high or moderate incompleteness calibration models. This drops to $\sim 5.7\%$ with the low incompleteness simulated visibility model. The completeness of the $14.5 - 290$ m visibility model varies with baseline length and is dominated by the incompleteness of the diffuse and point source visibility model on short and long baselines, where the respective emission types make more significant contributions to the measured visibilities. The bias in values of $A_{\text{true}}/\hat{A}$ when calibrating the $14.5 - 290$ m data set using absolute calibration is $\sim 4\%$ in the high and moderate incompleteness regime and drops to $\sim 2\%$ in the low incompleteness regime. The lower bias in the values of $A_{\text{true}}/\hat{A}$ when calibrating the $14.5 - 290$ m data set relative to the $14.5 - 29$ m data set can be understood in terms of the higher completeness of the point source relative to diffuse components of the simulated calibration sky models. As a result, the bias in $A_{\text{true}}/\hat{A}$ for the $14.5 - 290$ m data set will be a weighted average of the larger and smaller biases associated with the different baseline lengths of the data set.

When calibrating either the $14.5 - 29$ m or $14.5 - 290$ m data sets with absolute calibration using the high or moderate incompleteness simulated calibration sky models, both when calibrating with or without an informative prior on the degenerate gain amplitudes, the average absolute bias in $A_{\text{true}}/\hat{A}$ is comparable. The only difference between the high or moderate incompleteness simulated calibration sky models is the reduction in the minimum flux-density

(from 1 to 0.1 Jy at 151 MHz) of GLEAM sources included in V^{sim}. Thus, the similarity between the average absolute biases in $A_{\text{true}}/\hat{A}$ in these two cases implies the limited importance, with respect to reducing bias, of the increased point source completeness resulting from the additional sources included in the model when there is comparatively large incompleteness in the diffuse emission component of the sky model used for V^{sim}.

The $\sim 1.25\%$ bias in $\left\langle \text{Pr}(A_{\text{true}}/A \mid A_{\text{F}}, V^{\text{obs}}) \right\rangle$ when calibrating the $14.5 - 29$ m data set with BayesCal using the high incompleteness simulated calibration sky model and the sub-0.1% biases when calibrating the $14.5 - 29$ m data set with BayesCal using the moderate and low incompleteness simulated calibration sky models are lower than the corresponding biases found when calibrating the same data set with absolute calibration by factors of $\sim 5, 73$ and 190 in the high, moderate and low completeness regimes, respectively. Relative to the biases associated with the average absolute values of the degenerate gain amplitudes recovered using absolute calibration of the $14.5 - 290$ m data set, the biases in $\left\langle \text{Pr}(A_{\text{true}}/A \mid A_{\text{F}}, V^{\text{obs}}) \right\rangle$ when calibrating the $14.5 - 29$ m data set with BayesCal are also lower by factors of $\sim 4, 50$ and 73, in the high, moderate and low completeness regimes, respectively.

Nevertheless, while important for accurate data normalisation and error propagation, low level bias in the redundant gain amplitude parameters is still comparatively benign in 21 cm cosmology applications, since it does not complicate spectral separation of the foregrounds and the 21 cm signal in the data. In contrast, spurious spectral fluctuations imparted to the data by imperfect estimation of the redundant gain amplitude parameters must be suppressed to a high degree in order to recover the 21 cm signal in both measured and reconstructed sky-based approaches to 21 cm signal estimation.

5.1.2 Spurious spectral structure

When the ratio of the mean gain solution to the amplitude of spurious spectral fluctuations in the gains, on a given spectral scale, is equal to the dynamic range between the foregrounds and the 21 cm signal, the foreground systematic introduced will be at the level of the 21 cm signal on that scale. Thus, in order to recover unbiased estimates of the 21 cm signal, at a given spectral scale, one requires the ratio between the mean degenerate gain amplitude and the amplitude of spurious spectral fluctuations present in the data following calibration to be much smaller than the dynamic range between the foregrounds and the 21 cm signal at that scale. For the field calibrated here, and assuming a white noise-like 21 cm signal with an RMS amplitude[25] of 10 mK, this corresponds to a requirement that the amplitude of spurious spectral fluctuations in the ratio of the estimated and true degenerate gain amplitudes be smaller than 1 part in $\sim 10^{3.5}$. For an instrument with a mean gain amplitude of unity and a 1 MHz channel width, as considered here, this equates to a requirement that the power in these fluctuations be much smaller than $\sim 10^{-7}$ MHz^2.

To determine whether this threshold is met, we define $R \equiv A_{\text{true}}/A_{\text{model}}$, where A_{model} is the mean of the functional posterior probability distribution, maximum a posteriori, or maximum

---

[25] We use an amplitude of 10 mK for our reference white noise-like 21 cm signal to facilitate simple estimation of the corresponding ratio for alternate choices of EoR model. More stringent spectral calibration requirements will be needed for fainter EoR signals, with less power than this on a given spectral scale, and vice versa for brighter EoR signals.
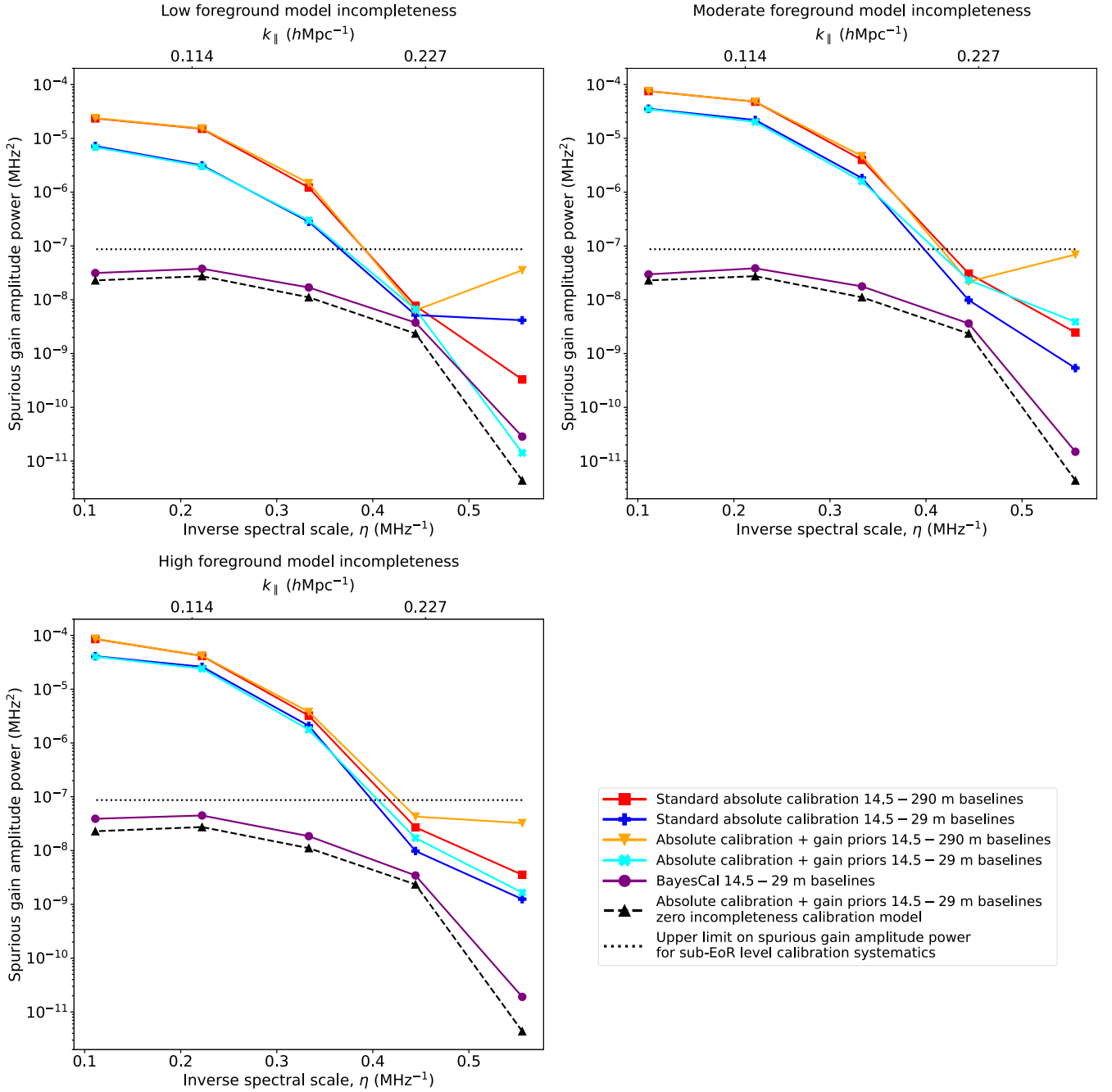
Figure 5. Power spectra, $P_A(\eta)$, of $R = A_{\text{true}}/A_{\text{model}}$ characterising the level of discrepant spectral structure imparted to the data as a result of imperfect calibration, derived in three regimes of incompleteness of the simulated component of the calibration model (see Table 2): low (top left), moderate (top right) and high incompleteness (bottom left). In all cases, $P_A(\eta)$ is calculated via Equation 44. The degenerate gain amplitude is real and therefore has a symmetric power spectrum; here, we plot half of the power spectrum for positive η. The line colours, denoting the calibration approach used and the baseline lengths it is applied to are labelled in the legend (see main text for details). The dotted black lines show the level at which spurious spectral fluctuations imparted to the data by imperfect calibration will introduce foreground systematics equal in amplitude, at a given spectral scale, to a white noise proxy for the 21 cm signal with an RMS amplitude of 10 mK. High fidelity calibration solutions with spurious gain amplitude power below the level of the dotted black line are necessary for unbiased recovery of the power spectrum of the 21 cm signal proxy from interferometric data. The dashed black lines show $P_A(\eta)$ for the maximum a posteriori degenerate gain amplitude solutions that would hypothetically be recovered using absolute calibration with gain priors in the limit of a perfect zero-imcompleteness calibration model, in which the non-zero spurious gain amplitude power derives purely from the noise in the data. This also corresponds to the BayesCal solutions that would be recovered in the limit of a perfect fitted calibration model component of the BayesCal data model for the contribution to the data from incomplete and imperfectly modelled emission on the sky.

likelihood parameter estimates of the degenerate gain amplitudes when discussing the calibration solutions derived with BAYESCAL and absolute calibration with and without imposing gain priors, respectively. The level of spurious spectral fluctuations that will be imparted to the calibrated data due to imperfect calibration, in the range of cases considered, can be assessed by calculating the power spectrum, $P_A(\eta)$, of fluctuations in the ratio of the recovered and true redundant gain amplitude parameters,

$$\left\langle \widetilde{\delta RW}(\eta)\widetilde{\delta RW}^*(\eta') \right\rangle (\Delta \nu)^2 \equiv \delta_K(\eta - \eta')P_A(\eta) . \tag{44}$$

Here, η measures inverse spectral scale[26]. $\widetilde{\delta RW}(\eta)$ is the discrete Fourier transform of δRW with $\delta R = (R - \overline{R})$, where $\overline{R}$ is the mean of R, and W is a Blackmann–Harris window function applied to down-weight spectral leakage of power from large spectral scale fluctuations in R. $\Delta \nu$ is the channel width in MHz and δ_K is the Kronecker delta function.

In Figure 5, we show $P_A(\eta)$, estimated using Equation 44, for our five sets of calibration solutions: BAYESCAL solutions for the $14.5 - 29$ m data set (purple lines), absolute calibration solutions with gain priors for the $14.5 - 29$ m and $14.5 - 290$ m data sets (cyan lines and orange lines, respectively) and absolute calibration solutions without gain priors for the $14.5 - 29$ m and $14.5 - 290$ m data sets (blue lines and red lines, respectively), derived in the low, medium and high simulated visibility model incompleteness regimes (top left, top right and bottom left, respectively). The dotted black lines show the level at which spurious spectral fluctuations imparted to the data by imperfect calibration will introduce foreground systematics equal in amplitude, at a given spectral scale, to a white noise-like 21 cm signal with an RMS amplitude of 10 mK. The dashed black lines shows $P_A(\eta)$ for the maximum a posteriori degenerate gain amplitude solutions that would hypothetically be recovered using absolute calibration with gain priors[27] in the limit of a perfect zero-incompleteness calibration model, in which the non-zero spurious gain amplitude power derives purely from the noise in the data. This also corresponds to the BAYESCAL solutions that would be recovered in the limit that the V^{fit} component of the BAYESCAL data model perfectly modelled the contribution to the data from incomplete and imperfectly modelled emission on the sky missing in the standard a priori known component of the

data model, V^{sim}, in the high, moderate and low calibration model incompleteness scenarios.

Both standard absolute calibration and absolute calibration with informative priors on the gain solutions result in spurious spectral fluctuations in R at a level which will result in foreground systematics exceeding the amplitude of our fiducial 21 cm signal on 3, 4.5 and 9 MHz spectral scales (corresponding to $k_\parallel \simeq 0.17, 0.11$ and 0.06 $h\mathrm{Mpc}^{-1}$, respectively, for redshifted 21 cm emission in the frequency range of our simulated observed data)[28] in the high, moderate and low calibration model incompleteness simulated visibility models when calibrating data sets comprised of $14.5 - 29$ m baselines or data sets comprised of $14.5 - 290$ m baselines. Calibration of the data with either of these approaches precludes separation of the 21 signal from the foregrounds, in the calibrated data set, on these spectral scales, by spectral means.

In contrast, significantly higher fidelity solutions are recovered when using BAYESCAL. Specifically, we find that spurious spectral fluctuations in R when using BAYESCAL are sufficiently low for foreground systematics that would result from applying spectral scale-based signal separation analysis techniques used in 21 cosmology to be below the level of the fiducial 21 cm signal on all spectral scales accessible in the data set and for all levels of simulated visibility model completeness considered. Comparable performance within a factor of a few is achieved in all three regimes of incompleteness of the a priori known component of the calibration sky model. In each case, the BAYESCAL calibration solutions reduce spurious redundant gain amplitude fluctuations to a few parts in 10^4 of the amplitude of simulated redundant gain parameters on 9 MHz spectral scales, with continuing reduction down to a few parts in 10^6 on 1.8 MHz spectral scales.

Comparing the BAYESCAL calibration solutions to those that would hypothetically be recovered using absolute calibration with gain priors in the limit of a perfect zero-imcompleteness calibration model in which the non-zero spurious gain amplitude power derives purely from the noise in the data, it can be seen that BAYESCAL enables recovery of calibration solutions similar in fidelity to those recovered in this perfect calibration model scenario. In all cases, the BAYESCAL solutions mitigate spurious gain amplitude power to within a factor of two of the complete zero-error calibration model solutions on all but the smallest (1.8 MHz) Nyquist sampled spectral scale in the calibration data. On the 1.8 MHz spectral scale, the difference in fidelity of the BAYESCAL and perfect calibration model solutions increases to a factor of ten in power in spurious spectral structure in the calibration solutions; nevertheless, the absolute fidelity of the BAYESCAL solutions on this spectral scale is sufficiently high for foreground systematics introduced by imperfections in the calibration solutions on this spectral scale to be negligible relative to the power in the fiducial EoR model on the same scale.

In the context of 21 cm cosmology applications, these results show that, in the scenarios modelled, only BAYESCAL enables recovery of the 21 cm at $k_\parallel \lesssim 0.15$ $h\mathrm{Mpc}^{-1}$. For all the other techniques in all completeness regimes, that scale is contaminated. More generally, when calibrating using BAYESCAL, the level of spurious spectral structure in the calibration solutions implies that the power in foreground systematics that would result from 21 cm signal separation of the foregrounds by spectral means in the presence of spurious spectral fluctuations in R are a factor of a few below the fiducial

---

[26] When plotting the power spectra, we also display the values of the inverse radial distance parameter corresponding to the plotted η-modes relevant to redshifted 21-cm emission in the frequency range of the calibrated data: $k_\parallel \approx \dfrac{2\pi \eta H_0 f_{21} E(z)}{c(1+z)^2}$. Here, $f_{21} \simeq 1420$ MHz is the rest frame frequency of the 21-cm line, c is the speed of light in a vacuum, $H_0 = h100$ km s^{-1} Mpc^{-1} is the present day value of the Hubble constant, where we take $h = 0.6766$, and $E(z) = \sqrt{\Omega_M(1+z)^3 + \Omega_\Lambda}$ is the dimensionless Hubble parameter, with $\Omega_M = 0.3111$ and $\Omega_\Lambda = 0.6889$ the present day matter and dark energy density in units of the critical density, respectively. We use the best fitting values from the Planck Collaboration et al. (2020) TT,TE,EE+lowE+lensing+BAO analysis for h, Ω_M and Ω_Λ.

[27] Since in the limit of zero incompleteness or error in the a priori known component of the calibration model the corresponding zero-power prior on the fitted component of the BAYESCAL calibration model suppresses to zero its contribution to the total calibration model, in this limit, BAYESCAL simplifies to the absolute calibration with gain priors algorithm described in Section 2.5. Thus, equivalent solutions would be derived in this limit using absolute calibration with gain priors or BAYESCAL. However, for brevity and to avoid confusion when discussing these solutions with those recovered in the realistic high, moderate and low calibration model incompleteness scenarios, we use the former to refer to the solutions in this zero-error and zero-incompleteness limit going forward.

[28] A 1 MHz spectral scale corresponds to a $k_\parallel \approx 0.5$ $h\mathrm{Mpc}^{-1}$ redshifted 21 cm inverse spatial scale at the center of the 160–169 MHz spectral band of the data sets being calibrated.

21 cm signal level on 9 MHz spectral scales ($k_\parallel \simeq 0.06\ h\mathrm{Mpc}^{-1}$), and drop to close to 4 orders of magnitude below the power in the fiducial 21 cm signal by 1.8 MHz scales ($k_\parallel \simeq 0.28\ h\mathrm{Mpc}^{-1}$).

5.1.3 *The impact of gain priors*

The generic effect of imposing informative gain priors when using either absolute calibration or BayesCal (i.e. joint estimation of absolute calibration solutions and a sky model) is the up-weighting of gain solutions consistent with one's prior knowledge of the spectral structure in the instrumental gains. However, when an incomplete calibration model is used, imposing informative gain priors will, in general, worsen the fit of the calibration model to the visibility data (since the fit of the model to the data alone is optimised in the absence of informative priors). Comparing the level of spurious spectral structure in degenerate gain amplitude parameters recovered using absolute calibration with and without informative priors on the gain solutions demonstrates how application of gain priors without explicitly addressing the incompleteness of the simulated calibration sky model is not guaranteed to be advantageous on all spectral scales. Rather, it leads to a redistribution of spurious spectral structure with, in some cases, minor improvements on large spectral scales (where the maximum likelihood solutions are least consistent with the prior and true gain solutions and, thus, where the impact of the prior is largest) coming at the cost of less accurate solutions on smaller spectral scales.

Ultimately, the impact of applying gain priors when using an incomplete sky model is a complex interplay between the incompleteness of the specific flux distribution of calibration model of the sky observed in the data being calibrated and the instrumental chromaticity of the baselines being calibrated. In contrast, imposing gain priors while jointly fitting for gain parameters and a model for the contribution to the visibilities from the component of the flux distribution absent in one's simulated calibration sky model, as done with BayesCal, mitigates this problem and optimises for calibration models in which both the recovered gains reflect their priors and the product of the gain parameters and the sum of the fitted and simulated components of the calibration model accurately describes the data.

5.2 BayesCal functional posterior probability distributions

In Section 5.1, we showed the frequency dependent means of the functional posterior probability distributions of the parameters of the BayesCal solutions, in order to simplify comparison with the maximum a posteriori and maximum likelihood solutions recovered with absolute calibration of visibilities, with and without imposition of informative gain priors. However, there is significantly greater information content in the functional posterior probability distributions of the parameters than in the mean of the distribution alone.

Estimation of the functional posterior probability distributions allows one to go beyond a single estimate of the calibrated visibility data, and, to instead derive the probability distribution of the data conditional on the calibration solutions. This provides a direct means of robustly propagating the statistical uncertainties in the calibration solutions through to the estimates of the 21 cm signal, which, in turn, is important if one wishes to avoid underestimation of the uncertainties on 21 cm estimates and mitigate the risk of either a false detection of the signal or of inflating the significance of a signal detection.

In Figure 6, we show contour plots[29] of the functional posterior probability distributions for the ratio of the input degenerate calibration amplitude and the model degenerate calibration amplitude ($\Pr(A/A_\mathrm{true} \mid A_\mathrm{F}, V^\mathrm{obs})$, top), tip-phase ($\Pr(\Phi_l \mid \Phi_{l,\mathrm{F}}, V^\mathrm{obs})$; center) and tilt-phase ($\Pr(\Phi_m \mid \Phi_{m,\mathrm{F}}, V^\mathrm{obs})$; bottom) recovered when using the low, moderate and high simulated visibility models in the left, center and right panels, respectively.

For all levels of sky-model incompleteness considered, we find that estimates of the degenerate gain amplitude and tilt phase solutions recovered with BayesCal calibration are consistent with the true calibration solutions to within the 3-σ limits of their conditional probability distributions. The tip phase solutions recovered with BayesCal are consistent to within the 3-σ limits of their conditional probability distributions in the lower half of the band and marginally inconsistent in the upper half of the band. This reflects the fact that V^fit provides a good but not perfect model of the contribution to the observed visibilities from emission that is missing or imperfectly modelled in the a priori known component of the calibration model. However, as noted in Section 5.1.1, the absolute bias in the degenerate tip phase solutions is small relative to the angular scales on which interferometers designed to detect the power spectrum of redshifted 21 cm emission are maximised. Thus, we expect the small absolute level of the bias to limit its impact on 21 cm signal recovery in reconstructed sky based frameworks. Furthermore, it has no impact on measured sky-based power spectrum estimation based frameworks, such as delay-spectrum power spectrum estimation, that do not use the phase information in the data.

We also note that in addition to the means of the conditional probability distributions for the ratio of the input degenerate calibration amplitude and the recovered degenerate calibration amplitude being spectrally smooth, as shown in Figure 5, the iso-probability-contours of the distributions are also spectrally smooth, and it is the variance in the mean amplitudes of the samples, rather than their shape, which dominates the calibration uncertainties. This derives from the individual parameter samples from the posterior probability distributions of the Fourier mode parametrisations of the degenerate gain amplitudes corresponding to individual model gain solutions with spectral structure that is consistent with the true instrumental gains in the data.

This is favourable for 21 cm cosmology applications, because, when the BayesCal calibration uncertainties are propagated through to the signal estimates, not only will the mean calibrated data solutions have high spectral fidelity, but the individual samples from the conditional probability distribution of the calibrated data set will as well. This will result in reduced uncertainties on estimates of the 21 cm signal recovered from the data, relative to a situation where the mean of the distribution is flat but other iso-probability-contours have spectral structure.

Nevertheless, we note that the optimality, with respect to bias minimisation, of propagating the samples from the posteriors for the degenerate gain parameters through to the target statistic one wishes to estimate from the data depends on the convexity of the function that takes one from the raw data to that statistic. Jensen's inequality states that for a convex function (such as the power spectrum of the 21 cm emission), $f(x)$ on $\mathbb{R}_X$, where X is a discrete random

---

[29] We derive the contour plots of the functional posterior probability distributions from the posteriors of the Fourier mode parametrisations of the degenerate gain parameters using the FGIVENX software package (Handley 2018).
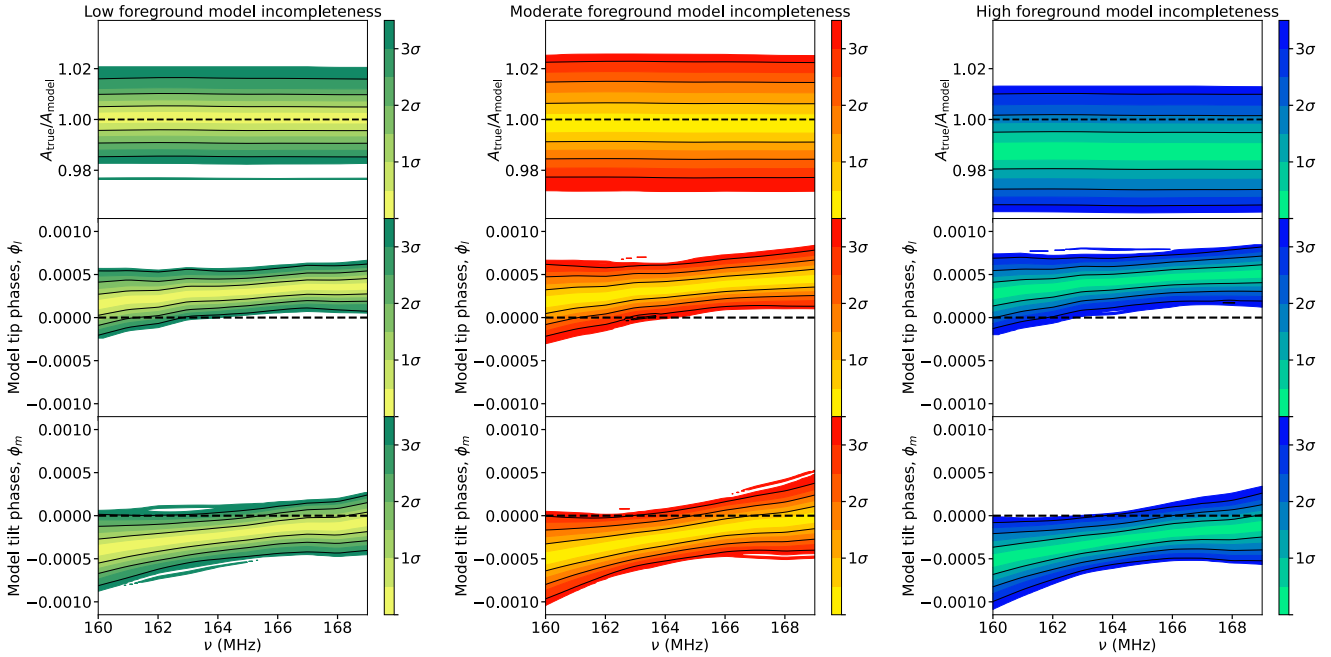
Figure 6. Contour plots of the functional posterior probability distributions for the ratio of the input degenerate calibration amplitude and the model degenerate calibration amplitude ($\Pr(A/A_{\text{true}} \mid A_{\text{F}}, V^{\text{obs}})$; top), tip-phase ($\Pr(\Phi_l \mid \Phi_{l,\text{F}})$, center) and tilt-phase ($\Pr(\Phi_l \mid \Phi_{l,\text{F}})$, bottom), when fitting the sum of a simulated model for the known flux contributions from diffuse and point sources and a statistical model for the unknown component of sky emission, jointly with the degenerate gain solutions when the simulated sky model has low, moderate and high incompleteness from left to right, respectively. The dashed black lines show the input values of the parameters.

variable, assuming $E[f(X)]$ and $f(E[X])$ are finite, we have,

$$E[f(X)] \geq f(E[X]) .\qquad (45)$$

This implies that bias in the mean of the power spectra of samples from the probability distribution of the data conditional on the calibration solutions will be equal to or exceed the bias in the power spectrum of the data derived from the means of the posteriors on the degenerate gain parameters. This derives from random errors in samples from the posteriors on the degenerate gain adding up constructively in the former case and averaging down in the latter. Thus, averaging the functional posteriors prior to forming the power spectrum[30] will minimise imperfect-calibration-derived spurious spectral structure in the final power spectral estimates.

6   SUMMARY & CONCLUSIONS

BAYESCAL, a new Bayesian framework for interferometric calibration designed to recover high fidelity gain solutions by mitigating sky-model incompleteness, was introduced in paper I of this series. The BAYESCAL calibration model is comprised of (i) a simulated visibility model, V^{sim}, which describes the expected contribution to the observed visibilities associated with a priori known sky emission (this model component is standard in sky-referenced calibration) and (ii) a fitted visibility model, V^{fit}, that models the contribution to the observed visibilities from emission not included in V^{sim} due to the incompleteness and uncertainties associated with our a priori

knowledge of the brightness distribution of the sky. When calibrating data with BAYESCAL, one jointly fits for instrumental gains, constrained by available priors on their spectral structure, and the most probable sky-model parameter values for V^{fit}, constrained by a prior on their total power.

BAYESCAL provides a powerful new tool for high fidelity interferometric calibration that is particularly well suited for application to analyses for which either the accuracy or spectral fidelity of the instrumental calibration is important to the science goal. Estimation of a faint 21 cm signal in the presence of bright spectrally smooth foregrounds is an example of such an analysis.

Here, we have demonstrated the application of BAYESCAL in the context of absolute calibration of data from redundant interferometric arrays, in which relative gains between antennas can be solved for by using sky-model-independent constraints based on the internal consistency of the redundant groups of baselines. To assess the efficacy of the BAYESCAL calibration methodology, we have applied it to simulated observations of full-Stokes sky emission in the 160–169 MHz spectral band, with a HERA-like hexagonal close-packed redundant array, for three levels of completeness of the a priori known component of the calibration model (corresponding to three levels of uncertainty and completeness of our model for the distribution of diffuse low frequency radio emission and point sources in the data to be calibrated).

We have shown that BAYESCAL enables recovery of significantly higher fidelity calibration solutions than standard absolute calibration or absolute calibration with physical gain priors. Relative to these approaches, we find the largest improvements on the largest spectral scales. On small and intermediate spectral scales, power in spurious gain amplitude fluctuations is comparable to or lower that these alternate approaches. On larger spectral scales (3, 4.5 and 9 MHz spectral scales, corresponding to $k_{\parallel} \simeq 0.17, 0.11$

---

[30] In practice, this can be done by averaging the samples from the posteriors on the degenerate gain parameters, as in Section 5.1, or the conditional posteriors at any subsequent stage of data processing before non-linear operations equivalent to convex functions on the data are applied.

and 0.06 $h\mathrm{Mpc}^{-1}$, respectively, for redshifted 21 cm emission in the frequency range of our simulated observed data calibrated here) power in spurious gain amplitude fluctuations is suppressed by an additional 1–4 orders of magnitude, relative to those approaches.

In the context of 21 cm cosmology applications, our results show that, in the scenarios modelled, only BAYESCAL yields sufficiently high fidelity calibration solutions for recovery of the 21 cm signal at $k_\parallel \lesssim 0.15\ h\mathrm{Mpc}^{-1}$. For all the other techniques in all completeness regimes, that scale is contaminated. The overall fidelity of the calibration solutions recovered with BAYESCAL is such that foreground systematics imparted by imperfections in the calibration solutions are below a white noise proxy for the 21 cm signal with an RMS amplitude of 10 mK, on all spectral scales in the calibrated data set, ranging from a factor of a few below the fiducial 21 cm signal level on 9 MHz spectral scales ($k_\parallel \simeq 0.06\ h\mathrm{Mpc}^{-1}$) to close to 4 orders of magnitude below the power in the fiducial 21 cm signal on 1.8 MHz scales ($k_\parallel \simeq 0.28\ h\mathrm{Mpc}^{-1}$).

DATA AVAILABILITY

The data from this study will be shared on reasonable request to the corresponding author.

References

Barry N., Hazelton B., Sullivan I., Morales M. F., Pober J. C., 2016, MNRAS, 461, 3135

Barry N., Beardsley A. P., Byrne R., Hazelton B., Morales M. F., Pober J. C., Sullivan I., 2019, PASA, 36, e026

Byrne R., et al., 2019, ApJ, 875, 70

Byrne R., Morales M. F., Hazelton B. J., Wilensky M., 2021, MNRAS, 503, 2457. doi:10.1093/mnras/stab647

Choudhuri S., Bull P., Garsden H., 2021, MNRAS, 506, 2066. doi:10.1093/mnras/stab1795

Danny C. Price, 2016, ascl.soft, ascl:1603.013

DeBoer D. R., et al., 2017, PASP, 129, 045001

Dillon J. S., Parsons A. R., 2016, ApJ, 826, 181

Dillon J. S., et al., 2020, arXiv, arXiv:2003.08399

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

Fagnoni N., de Lera Acedo E., DeBoer D. R., Abdurashidova Z., Aguirre J. E., Alexander P., Ali Z. S., et al., 2021, MNRAS, 500, 1232. doi:10.1093/mnras/staa3268

Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759

Hamaker J. P., Bregman J. D., Sault R. J., 1996, A&AS, 117, 137

Handley W., 2018, JOSS, 3, 849

Handley W. J., Hobson M. P., Lasenby A. N., 2015a, MNRAS, 453, 4384

Handley W. J., Hobson M. P., Lasenby A. N., 2015b, MNRAS, 450, L61

Hurley-Walker N., et al., 2017, MNRAS, 464, 1146

Joseph R. C., Trott C. M., Wayth R. B., 2018, AJ, 156, 285

Kern N. S., Dillon J. S., Parsons A. R., Carilli C. L., Bernardi G., Abdurashidova Z., Aguirre J. E., et al., 2020, ApJ, 890, 122. doi:10.3847/1538-4357/ab67bc

Kohn S. A., Aguirre J. E., La Plante P., Billings T. S., Chichura P. M., Fortino A. F., Igarashi A. S., et al., 2019, ApJ, 882, 58. doi:10.3847/1538-4357/ab2f72

Lanman A. E., Kern N., 2019, ascl.soft, ascl:1907.002

Lanman A. E., Pober J. C., 2019, MNRAS, 487, 5840. doi:10.1093/mnras/stz1639

Lenc E., et al., 2016, ApJ, 830, 38

Li J., Zuo S., Wu F., Wang Y., Zhang J., Sun S., Xu Y., et al., 2020, SCPMA, 63, 129862. doi:10.1007/s11433-020-1594-8

Li W., et al., 2018, ApJ, 863, 170

Liu A., Shaw J. R., 2020, PASP, 132, 062001. doi:10.1088/1538-3873/ab5bfd

Liu A., Tegmark M., Morrison S., Lutomirski A., Zaldarriaga M., 2010, MNRAS, 408, 1029

Martinot Z. E., Aguirre J. E., Kohn S. A., Washington I. Q., 2018, ApJ, 869, 79. doi:10.3847/1538-4357/aaeac6

Mathewson D. S., Milne D. K., 1965, AuJPh, 18, 635

Mellema G. et al., 2013, Exp. Astron., 36, 235

Morales M. F., Beardsley A., Pober J., Barry N., Hazelton B., Jacobs D., Sullivan I., 2019, MNRAS, 483, 2207

Mozdzen T. J., Bowman J. D., Monsalve R. A., Rogers A. E. E., 2017, MNRAS, 464, 4995

Newburgh L. B., Bandura K., Bucher M. A., Chang T.-C., Chiang H. C., Cliche J. F., Davé R., et al., 2016, SPIE, 9906, 99065X

Newburgh L. B., et al., 2014, SPIE, 9145, 91454V, SPIE.9145

Orosz N., Dillon J. S., Ewall-Wice A., Parsons A. R., Thyagarajan N., 2019, MNRAS, 487, 537

Paciga G. et al., 2013, MNRAS, 433, 639

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

Patil A. H., Yatawatta S., Zaroubi S., Koopmans L. V. E., de Bruyn A. G., Jelić V., Ciardi B., et al., 2016, MNRAS, 463, 4317

Patil A. H., Yatawatta S., Koopmans L. V. E., de Bruyn A. G., Brentjens M. A., Zaroubi S., Asad K. M. B., et al., 2017, ApJ, 838, 65

Patra N., et al., 2018, ExA, 45, 177

Planck Collaboration, Aghanim N., Akrami Y., Ashdown M., Aumont J., Baccigalupi C., Ballardini M., et al., 2020, A&A, 641, A6. doi:10.1051/0004-6361/201833910

Riseley C. J., Galvin T. J., Sobey C., Vernstrom T., White S. V., Zhang X., Gaensler B. M., et al., 2020, PASA, 37, e029. doi:10.1017/pasa.2020.20

Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, A&A, 345, 380

Sievers J. L., 2017, arXiv, arXiv:1701.01860

Sims P. H., Pober J. C., 2019, MNRAS, 488, 2904

Sims P. H., Lentati L., Alexander P., Carilli C. L., 2016, MNRAS, 462, 3069

Sims P. H., Lentati L., Pober J. C., Carilli C., Hobson M. P., Alexander P., Sutter P. M., 2019, MNRAS, 484, 4152

Smirnov O. M., 2011, A&A, 527, AA106

Taylor G. B., Carilli C. L., Perley R. A., 1999, ASPC, 180

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

Wayth R. B., et al., 2015, PASA, 32, e025

Wayth R. B., Tingay S. J., Trott C. M., Emrich D., Johnston-Hollitt M., McKinley B., Gaensler B. M., et al., 2018, PASA, 35, 33

Wieringa M. H., 1992, ExA, 2, 203

Wu F., Li J., Zuo S., Chen X., Das S., Marriner J. P., Oxholm T. M., et al., 2021, MNRAS, 506, 3455. doi:10.1093/mnras/stab1802

Xiang Y., Sun D. Y., Fan W., Gong X. G., 1997, Physics Letters A, 233, 216-220

Zheng H., et al., 2017, MNRAS, 464, 3486

Zonca A., Singer L., Lenz D., Reinecke M., Rosset C., Hivon E., Gorski K., 2019, JOSS, 4, 1298. doi:10.21105/joss.01298

APPENDIX A: THE TRANSITION MATRIX BETWEEN ENU AND EQUATORIAL VECTOR BASES

Careful choice of vector bases enables significant simplification of the component matrices of the RIME. A vector basis tied to the orientations of the antenna feeds provides the simplest vector space for defining the polarised voltage beam and a sky-centric basis tied to the celestial sphere provides a more natural vector space in which to define one's polarised sky model. However, to evaluate Equation 1, one must transition the beam and the sky to a mutual vector basis.

A1   Two dimensional vector field basis transformations

Suppose $A = \{\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\alpha}}_2\}$ and $B = \{\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2\}$ are two ordered bases for a 2-dimensional vector space V and $\boldsymbol{v}_A = (v_{\alpha_1}, v_{\alpha_2})$ and $\boldsymbol{v}_B = (v_{\beta_1}, v_{\beta_2})$ are the coordinate 2-tuples of a vector $\boldsymbol{v}$ in V with respect to basis A and B, respectively. The components of the coordinate 2-tuples of $\boldsymbol{v}$ are given by the projection of $\boldsymbol{v}$ onto the associated basis vectors[31],

$$v_{\alpha_1} = \boldsymbol{v} \cdot \hat{\boldsymbol{\alpha}}_1 \tag{A1}$$
$$v_{\alpha_2} = \boldsymbol{v} \cdot \hat{\boldsymbol{\alpha}}_2$$
$$v_{\beta_1} = \boldsymbol{v} \cdot \hat{\boldsymbol{\beta}}_1$$
$$v_{\beta_2} = \boldsymbol{v} \cdot \hat{\boldsymbol{\beta}}_2 \, .$$

The components of the coordinate 2-tuple in basis A can be calculated from those in basis B via (e.g. Martinot et al. 2018),

$$v_{\alpha_1} = \boldsymbol{v} \cdot \hat{\boldsymbol{\alpha}}_1 = (v_{\beta_1}\hat{\boldsymbol{\beta}}_1 + v_{\beta_2}\hat{\boldsymbol{\beta}}_2) \cdot \hat{\boldsymbol{\alpha}}_1 \tag{A2}$$
$$v_{\alpha_2} = \boldsymbol{v} \cdot \hat{\boldsymbol{\alpha}}_2 = (v_{\beta_1}\hat{\boldsymbol{\beta}}_1 + v_{\beta_2}\hat{\boldsymbol{\beta}}_2) \cdot \hat{\boldsymbol{\alpha}}_2$$

Thus, the transition matrix encoding the coordinate transformation between the coordinate 2-tuple in basis A and B is given by,

$$\mathbf{T}_{\hat{\alpha}-\hat{\beta}} = \begin{pmatrix} \hat{\boldsymbol{\alpha}}_1 \cdot \hat{\boldsymbol{\beta}}_1 & \hat{\boldsymbol{\alpha}}_1 \cdot \hat{\boldsymbol{\beta}}_2 \\ \hat{\boldsymbol{\alpha}}_2 \cdot \hat{\boldsymbol{\beta}}_1 & \hat{\boldsymbol{\alpha}}_2 \cdot \hat{\boldsymbol{\beta}}_2 \end{pmatrix} \, , \tag{A3}$$

such that,

$$\begin{pmatrix} v_{\alpha_1} \\ v_{\alpha_2} \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{\alpha}}_1 \cdot \hat{\boldsymbol{\beta}}_1 & \hat{\boldsymbol{\alpha}}_1 \cdot \hat{\boldsymbol{\beta}}_2 \\ \hat{\boldsymbol{\alpha}}_2 \cdot \hat{\boldsymbol{\beta}}_1 & \hat{\boldsymbol{\alpha}}_2 \cdot \hat{\boldsymbol{\beta}}_2 \end{pmatrix} \begin{pmatrix} v_{\beta_1} \\ v_{\beta_2} \end{pmatrix} \, . \tag{A4}$$

Furthermore, Equation A3 is trivially extended to the case that the basis vectors of bases *A* and *B* are, themselves, a function of direction on the sky, $\hat{\boldsymbol{s}}$, as in Equation 33, in which case we have,

$$\mathbf{T}_{\hat{\alpha}-\hat{\beta}}(\hat{\boldsymbol{s}}) = \begin{pmatrix} \hat{\boldsymbol{\alpha}}_1(\hat{\boldsymbol{s}}) \cdot \hat{\boldsymbol{\beta}}_1(\hat{\boldsymbol{s}}) & \hat{\boldsymbol{\alpha}}_1(\hat{\boldsymbol{s}}) \cdot \hat{\boldsymbol{\beta}}_2(\hat{\boldsymbol{s}}) \\ \hat{\boldsymbol{\alpha}}_2(\hat{\boldsymbol{s}}) \cdot \hat{\boldsymbol{\beta}}_1(\hat{\boldsymbol{s}}) & \hat{\boldsymbol{\alpha}}_2(\hat{\boldsymbol{s}}) \cdot \hat{\boldsymbol{\beta}}_2(\hat{\boldsymbol{s}}) \end{pmatrix} \, . \tag{A5}$$

A2   Geometry of antenna and sky coordinate frames

The transition matrix, $\mathbf{T}_{\mathrm{ENU-equatorial}}$, encodes the transform of the components of the electric vector field between an RA, Dec basis in an equatorial coordinate system[32] and an east-north-up (ENU) topocentric coordinate basis. To bridge the gap between the equatorial and the ENU coordinate systems and simplify the derivation of $\mathbf{T}_{\mathrm{ENU-equatorial}}$, we introduce a third coordinate system, zaaz[33].

---

[31] This is true independently of the coordinate system used to describe the components of $\boldsymbol{v}$.

[32] We use the RA, Dec coordinates of the standard equatorial spherical coordinate system, as well as the radial coordinate axis increasing from the origin of the coordinate system.

[33] We use the zenith angle and azimuth coordinates of a standard zaaz coordinate system plus a radial coordinate pointing between the observer and zenith.

Since transformations between the equatorial, zaaz and ENU bases are linear, $\mathbf{T}_{\mathrm{ENU-equatorial}}$ is the product of two intermediate transition matrices, $\mathbf{T}_{\mathrm{ENU-zaaz}}$ and $\mathbf{T}_{\mathrm{zaaz-equatorial}}$, between the respective coordinate bases,

$$\mathbf{T}_{\mathrm{ENU-equatorial}} = \mathbf{T}_{\mathrm{ENU-zaaz}} \mathbf{T}_{\mathrm{zaaz-equatorial}} \, , \tag{A6}$$

where $\mathbf{T}_{\mathrm{ENU-zaaz}}$ and $\mathbf{T}_{\mathrm{zaaz-equatorial}}$ are transition matrices between the Cartesian ENU basis and the corresponding topocentric spherical zaaz coordinate basis, and between the zaaz and equatorial coordinate bases, respectively. Furthermore, as a transition matrix between the angular basis vectors of two spherical coordinate bases differing only in the poles of their coordinate axes, $\mathbf{T}_{\mathrm{zaaz-equatorial}}$ is fully characterised by the relative rotation angle of the respective angular basis vectors (see Section A4).

The coordinate axis definitions and basis vectors relevant to the three coordinate frames used in Equation A6 are illustrated in Figure A1. ENU is a topocentric coordinate system with $\hat{\boldsymbol{e}}_{\mathrm{n}}$ and $\hat{\boldsymbol{e}}_{\mathrm{e}}$ basis vectors lying in the local tangent plane of the observer in the directions of north and east, respectively, and $\hat{\boldsymbol{e}}_{\mathrm{u}}$ points from the observer towards zenith. The zaaz and equatorial coordinate frames are defined on the celestial sphere (i.e. r = constant). Zenith angle ($0 \geq za \geq \pi$ rad) increases from zero at the observer's zenith to π rad at the observer's nadir. Azimuth ($0 \geq az \geq 2\pi$ rad) increases in the direction of east from zero at $\hat{\boldsymbol{e}}_{\mathrm{n}}$. The LST dependence of zaaz coordinates of fixed sources on the celestial sphere follows from the rotation, with the siderial motion of the observer, of the pole of the angular coordinates on the celestial sphere. RA ($-\pi \leq$ RA $\leq \pi$ rad) increases eastward from zero at vernal equinox (VE), where VE is the direction of the ascending node of the intersection of the ecliptic and the celestial equator. Declination ($-\pi/2 \leq$ Dec $\leq \pi/2$ rad) increases from zero on the celestial equator to $\pi/2$ rad at NCP. Equatorial coordinates, and, correspondingly, the equatorial vector basis, of a given reference epoch (B1950 or J2000, for example), are stationary on the celestial sphere[34].

A3 $\mathbf{T}_{\mathrm{ENU-zaaz}}$

Before deriving the basis vectors of the ENU and zaaz coordinate systems, we note two physical constraints on the electric and voltage vector fields:

(i) The electric field lies purely in the plane perpendicular to its Poynting vector, thus, in an antenna-centric spherical basis, such as zaaz or equatorial, the radial component of the electric field incident on the antenna is identically zero.

(ii) A dual-feed antenna is sensitive to the amplitude of the electric field vector in the plane defined by the orientation of the feeds. Without loss of generality, we will assume this is the N-E plane. Thus, in an ENU basis, the $\hat{\boldsymbol{e}}_{\mathrm{u}}$ component of the voltage vector is identically zero.

Here, for generality, we will work with the three dimensional basis vectors at the outset and then apply the above constraints when deriving the transition matrices for the system of interest.

ENU and zaaz coordinates are related as follows (see Fig-

---

[34] Use of a reference epoch eliminates the time dependence of the coordinate system that is otherwise present due to slow time-dependence of the direction of VE resulting from the precession and nutation of Earth's rotation axis.
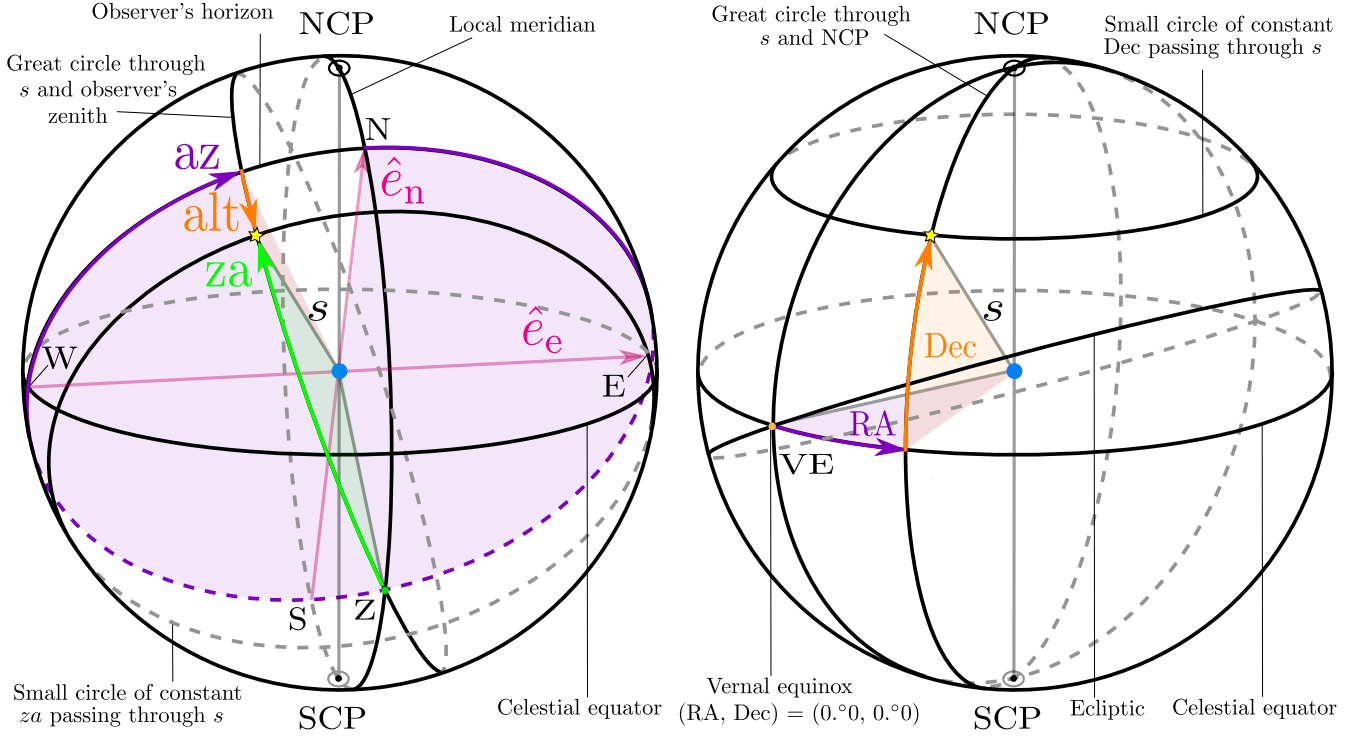
Figure A1. [Left] Azimuth (az), zenith angle (za) and altitude (alt) coordinates of a source on the celestial sphere at position s. The zenith pointing, Z, and the corresponding plane perpendicular to it and tangent to Earth at the observer, the intersection of which with the celestial sphere describes the 'observer's horizon', are shown for an observer at a latitude of $-30°.0$. The directions of the geographic cardinal points, in the tangent plane to Earth, at the observer's location, are given by pointings from the origin of the tangent plane to its intersections, on the observer's horizon, with the local meridian and celestial equator. Azimuth, $0 \leq az < 2\pi$ rad, is measured from north (N) towards east (E) in the tangent plane to the Earth at the observer's location. Zenith angle, $0 \leq az < \pi$ rad, increases from zero at the observer's zenith to $180°.0$ at the observer's nadir. The east and north basis vectors of an east-north-up (ENU) coordinate system at the location of the observer are labelled as $\hat{e}_e$ and $\hat{e}_n$, respectively. The up-axis, and corresponding basis vector, is not labelled but increases in the zenith pointing direction from zero at the origin of the observer's tangent plane. [Right] Right ascension (RA) and declination (Dec) coordinates of a source on the celestial sphere at the position s. Right ascension, $0 \leq RA < 2\pi$ rad, increases eastward from zero at vernal equinox (VE), where VE is defined as the direction of the ascending node of the intersection of the ecliptic and the celestial equator. Declination, $-\pi/2 \leq Dec \leq \pi/2$ rad, increases from zero on the celestial equator to $\pi/2$ rad at NCP.

ure A1, left),

$$E = r \sin(za) \sin(az) \tag{A7}$$
$$N = r \sin(za) \cos(az)$$
$$U = r \cos(za) \,,$$

$$r = \sqrt{E^2 + N^2 + U^2} \tag{A8}$$
$$\cos(za) = \frac{U}{\sqrt{E^2 + N^2 + U^2}}$$
$$\tan(az) = \frac{E}{N} \,.$$

The ENU basis vectors $\hat{e}_e$, $\hat{e}_n$ and $\hat{e}_u$ are unit vectors in the local east, north and up directions, respectively, and, in an ENU coordinate frame, they have components[35],

$$\hat{e}_n = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \ \hat{e}_e = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \ \hat{e}_u = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} . \tag{A9}$$

The components of the zaaz unit vectors in the ENU coordinate frame are given by,

$$\hat{e}_r = \frac{1}{|\nabla r|} \nabla r \tag{A10}$$

$$\hat{e}_{za} = \frac{1}{|\nabla za|} \nabla za$$

$$\hat{e}_{az} = \frac{1}{|\nabla az|} \nabla az \,,$$

[35] Following Kohn et al. (2019), we use a left-handed ENU coordinate basis.

where ∇ is the gradient operator. From Equation A7 and Equation A10 we have,

$$\hat{\boldsymbol{e}}_r = \frac{E}{\sqrt{E^2+N^2+U^2}}\,\hat{\boldsymbol{e}}_e + \frac{N}{\sqrt{E^2+N^2+U^2}}\,\hat{\boldsymbol{e}}_n + \frac{U}{\sqrt{E^2+N^2+U^2}}\,\hat{\boldsymbol{e}}_u$$
(A11)

$$\hat{\boldsymbol{e}}_{za} = \frac{EU}{\sqrt{E^2+N^2}\sqrt{E^2+N^2+U^2}}\,\hat{\boldsymbol{e}}_e + \frac{NU}{\sqrt{E^2+N^2}\sqrt{E^2+N^2+U^2}}\,\hat{\boldsymbol{e}}_n$$
$$- \frac{\sqrt{E^2+N^2}}{\sqrt{E^2+N^2+U^2}}\,\hat{\boldsymbol{e}}_u$$

$$\hat{\boldsymbol{e}}_{az} = \frac{N}{\sqrt{E^2+N^2}}\,\hat{\boldsymbol{e}}_e - \frac{E}{\sqrt{E^2+N^2}}\,\hat{\boldsymbol{e}}_n \,,$$

or, equivalently, in terms of r, za and az coordinates,

$$\hat{\boldsymbol{e}}_r = \sin(za)\sin(az)\,\hat{\boldsymbol{e}}_e + \sin(za)\cos(az)\,\hat{\boldsymbol{e}}_n + \cos(za)\,\hat{\boldsymbol{e}}_u \quad \text{(A12)}$$

$$\hat{\boldsymbol{e}}_{za} = \sin(az)\cos(za)\,\hat{\boldsymbol{e}}_e + \cos(az)\cos(za)\,\hat{\boldsymbol{e}}_n + \sin(za)\,\hat{\boldsymbol{e}}_u$$

$$\hat{\boldsymbol{e}}_{az} = \cos(az)\,\hat{\boldsymbol{e}}_e - \sin(az)\,\hat{\boldsymbol{e}}_n \,.$$

From conditions 1 and 2 above, the vector electric field at the antenna is non-zero exclusively in the $\hat{\boldsymbol{e}}_{za}$ and $\hat{\boldsymbol{e}}_{az}$ basis vector directions and the component of the electric field vector that will induce a voltage response in the antenna feeds is exclusively in the $\hat{\boldsymbol{e}}_e$, $\hat{\boldsymbol{e}}_n$ basis vector directions. At this stage, we, therefore, restrict calculation of the transition matrix to the projections of basis vectors with these non-zero components. Using Equation A9 and Equation A12, this yields the transition matrix encoding the conversion of the non-zero and measured components of the electric field vector between zaaz and ENU coordinates,

$$\mathbf{T}_{\mathrm{ENU-zaaz}} = \begin{pmatrix} \hat{\boldsymbol{e}}_n \cdot \hat{\boldsymbol{e}}_{za} & \hat{\boldsymbol{e}}_n \cdot \hat{\boldsymbol{e}}_{az} \\ \hat{\boldsymbol{e}}_e \cdot \hat{\boldsymbol{e}}_{za} & \hat{\boldsymbol{e}}_e \cdot \hat{\boldsymbol{e}}_{az} \end{pmatrix} = \begin{pmatrix} \cos(az)\cos(za) & -\sin(az) \\ \sin(az)\cos(za) & \cos(az) \end{pmatrix} \,.$$
(A13)

A4 $\mathbf{T}_{\mathrm{zaaz-equatorial}}$

Since the angular coordinates of the equatorial and zaaz coordinate systems differ only in the locations of their poles, the conversion between their angular bases is given by the $\hat{s}$-dependent coordinate transform defined by the rotation between the axes of the respective coordinate systems in the tangent plane to the sphere (see Figure A2). Thus, from Figure A2, right, we have,

$$\hat{\boldsymbol{e}}_{za} \cdot \hat{\boldsymbol{e}}_{\mathrm{Dec}} = \cos(\pi - \psi_p) \tag{A14}$$

$$\hat{\boldsymbol{e}}_{za} \cdot \hat{\boldsymbol{e}}_{\mathrm{RA}} = \cos(3\pi/2 - \psi_p)$$

$$\hat{\boldsymbol{e}}_{az} \cdot \hat{\boldsymbol{e}}_{\mathrm{Dec}} = \cos(\psi_p - \pi/2)$$

$$\hat{\boldsymbol{e}}_{az} \cdot \hat{\boldsymbol{e}}_{\mathrm{RA}} = \cos(\pi - \psi_p) \,.$$

Here, the parallactic angle, $\psi_p(\hat{s})$, is the angle between the hour circle through point s and the great circle through s and zenith, with positive parallactic angle measured eastward from NCP (or equivalently, measured clockwise from $\hat{\boldsymbol{e}}_{\mathrm{Dec}}$ in the tangent plane to the celestial sphere at point s). For an antenna located at a latitude λ, the parallactic angle of a source in the direction $\hat{s}$, located at an hour angle h, and declination δ, is given by (e.g. Taylor, Carilli, & Perley 1999),

$$\tan(\psi_p) = \frac{\cos(\lambda)\sin(h)}{\sin(\lambda)\cos(\delta) - \cos(\lambda)\sin(\delta)\cos(h)} \,. \tag{A15}$$

Substituting the projections in Equation A14 into Equation A3 and simplifying, we have,

$$\mathbf{T}_{\mathrm{zaaz-equatorial}} = \begin{pmatrix} \hat{\boldsymbol{e}}_{za} \cdot \hat{\boldsymbol{e}}_{\mathrm{Dec}} & \hat{\boldsymbol{e}}_{za} \cdot \hat{\boldsymbol{e}}_{\mathrm{RA}} \\ \hat{\boldsymbol{e}}_{az} \cdot \hat{\boldsymbol{e}}_{\mathrm{Dec}} & \hat{\boldsymbol{e}}_{az} \cdot \hat{\boldsymbol{e}}_{\mathrm{RA}} \end{pmatrix} = -\begin{pmatrix} \cos(\psi_p) & \sin(\psi_p) \\ -\sin(\psi_p) & \cos(\psi_p) \end{pmatrix} \,.$$

(A16)

A5 $\mathbf{T}_{\mathrm{ENU-equatorial}}$

Finally, substituting Equation A13 and Equation A16 into Equation A6, we arrive at the transition matrix between equatorial and ENU coordinates,
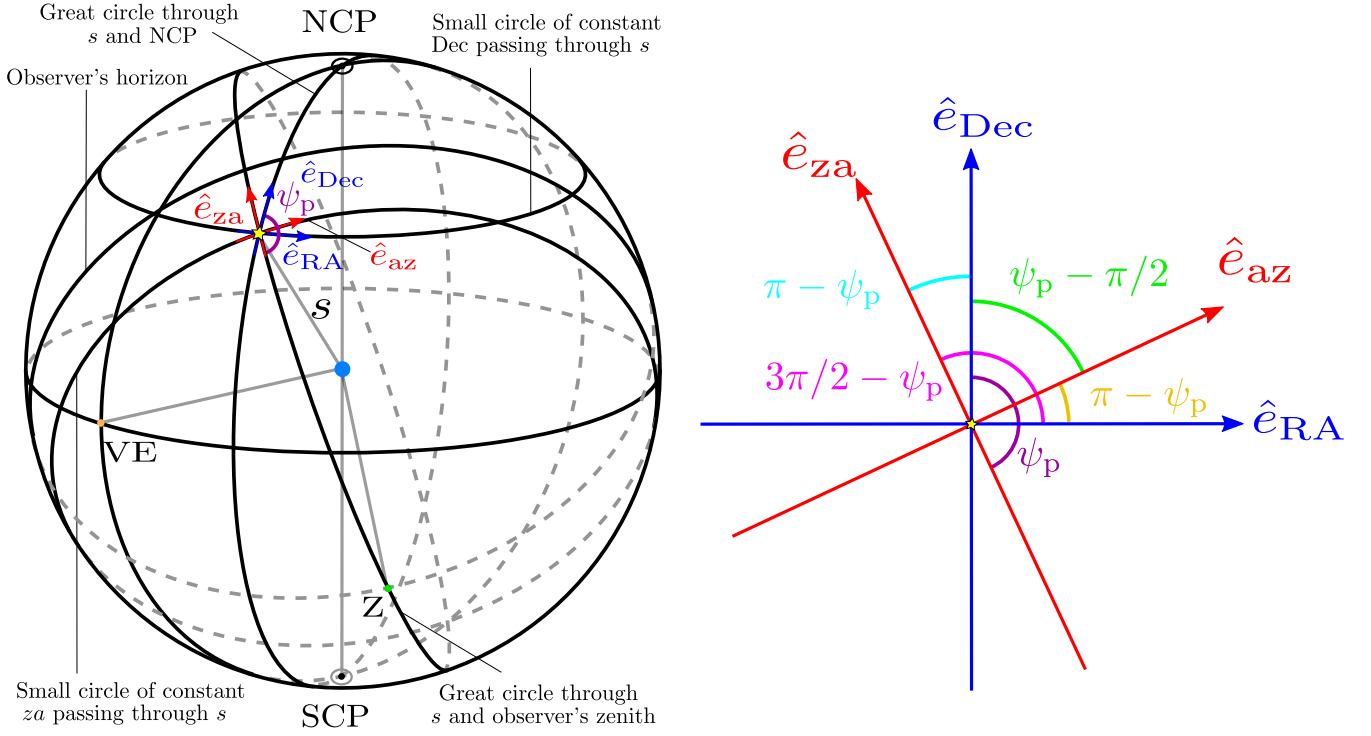
Figure A2. [Left] Basis vectors of right ascension ($\hat{e}_{Ra}$) and declination ($\hat{e}_{Dec}$), corresponding to an equatorial coordinate system, and zenith angle ($\hat{e}_{za}$) and azimuth ($\hat{e}_{az}$), corresponding to a zaaz coordinate system, are labelled. All four lie in the tangent plane to the celestial sphere at position s. $\hat{e}_{Dec}(\hat{s})$ and $\hat{e}_{Ra}(\hat{s})$ point towards the NCP and eastward along the small circle defined by the declination of the source (or great circle for a source on the celestial equator), respectively. $\hat{e}_{za}(\hat{s})$ and $\hat{e}_{az}(\hat{s})$ point away from zenith along the great circle passing though Z and s and from north towards east along the small circle of constant za passing through s, respectively. The parallactic angle, ψ_p, between the hour circle through point s and the great circle through s and zenith ($\angle ZsNCP$) is labelled. Positive parallactic angle is measured eastward from NCP. [Right] A face-on view of the tangent plane to the celestial sphere at point s. Basis vectors of right ascension and declination and zenith angle and azimuth are shown as red and blue arrows, respectively. The parallactic angle, ψ_p, is the angle at the origin subtended between $\hat{e}_{Dec}$ and -$\hat{e}_{za}$, with positive ψ_p measured clockwise from $\hat{e}_{Dec}$. Both it and the angles between basis vectors, relevant to the $\mathbf{T}_{\text{zaaz−equatorial}}$ transition matrix, are labelled.

$$\mathsf{T}_{\text{ENU-equatorial}} = \begin{pmatrix} -\sin(az)\sin(\psi_{\text{p}}) - \cos(az)\cos(za)\cos(\psi_{\text{p}}) & \sin(az)\cos(\psi_{\text{p}}) - \cos(az)\cos(za)\sin(\psi_{\text{p}}) \\ \cos(az)\sin(\psi_{\text{p}}) - \sin(az)\cos(za)\cos(\psi_{\text{p}}) & -\cos(az)\cos(\psi_{\text{p}}) - \sin(az)\cos(za)\sin(\psi_{\text{p}}) \end{pmatrix} . \tag{A17}$$