Global 21 cm Signal Extraction from Foreground and Instrumental Effects. III. Utilizing Drift-scan Time Dependence and Full Stokes Measurements

Keith Tauscher¹,², David Rapetti¹,³,⁴, and Jack O. Burns¹

¹ Center for Astrophysics and Space Astronomy, Department of Astrophysical and Planetary Science, University of Colorado, Boulder, CO 80309, USA
² Department of Physics, University of Colorado, Boulder, CO 80309, USA
³ NASA Ames Research Center, Moffett Field, CA 94035, USA
⁴ Research Institute for Advanced Computer Science, Universities Space Research Association, Mountain View, CA 94043, USA

Received 2020 February 11; revised 2020 April 21; accepted 2020 April 22; published 2020 July 16

Abstract

When using valid foreground and signal models, the uncertainties on extracted signals in global 21 cm signal experiments depend principally on the overlap between signal and foreground models. In this paper, we investigate two strategies for decreasing this overlap: (i) utilizing time dependence by fitting multiple drift-scan spectra simultaneously, and (ii) measuring all four Stokes parameters instead of only the total power, Stokes $I$. Although measuring polarization requires different instruments than are used in most existing experiments, all existing experiments can utilize drift-scan measurements merely by averaging their data differently. In order to evaluate the increase in constraining power from using these two techniques, we define a method for connecting root-mean-square (rms) uncertainties to probabilistic confidence levels. Employing simulations, we find that fitting only one total power spectrum leads to rms uncertainties at the few mK level, while fitting multiple time-binned, drift-scan spectra yields uncertainties at the $\lesssim$10 mK level. This significant improvement only appears if the spectra are modeled with one set of basis vectors instead of using multiple sets of basis vectors that independently model each spectrum. Assuming that they are simulated accurately, measuring all four Stokes parameters also leads to lower uncertainties. These two strategies can be employed simultaneously, and fitting multiple time bins of all four Stokes parameters yields the most precise measurements of the 21 cm signal, approaching the noise level in the data.

Unified Astronomy Thesaurus concepts: Cosmology (343); Astronomy data analysis (1858); Reionization (1383)

1. Introduction

The hyperfine, spin-flip transition of neutral hydrogen produces radiation of 1420 MHz of frequency in the rest frame, corresponding to a wavelength of 21 cm ([HII] 1970). Although this transition is highly forbidden, with a mean lifetime of around 11 million years (Condon & Ransom 2016, Section 7.8), its emission and absorption are visible from the vast amount of neutral gas in the early universe, redshifted to low frequencies of 10–200 MHz by cosmic expansion (Pritchard & Loeb 2012). It is the only existing direct probe of the neutral hydrogen in the Dark Ages and Cosmic Dawn of the early universe and it could be a powerful tool in the study of the Epoch of Reionization, when the hydrogen in the universe was ionized by light from compact sources like stars and black holes (Furlanetto et al. 2006). Two aspects of this 21 cm signal are currently under study: the power spectrum, where angular variations in the gas evolution manifest (Moraes & Wyithe 2010), and the sky-averaged (global) monopole component, which tracks the average properties of the gas across the universe as a function of cosmic time (Pritchard & Loeb 2010). This paper concerns the latter.

The most difficult analysis task in measuring the global 21 cm signal is separating it from foreground emission from our galaxy that is $\sim 10^{4.6}$ times larger than the signal, which is expected to have an amplitude of a few hundred mK. The foreground emission largely consists of synchrotron radiation, which follows a power law in frequency when the energy of the electrons emitting it follows a power-law distribution (Condon & Ransom 2016, Section 5.2); ergo, it is expected to be very spectrally smooth. However, there are large anisotropies in galactic emission both in magnitude and spectral index, which are averaged together by wide antenna beams that also change in frequency. Due to the corruption caused by this beam averaging, there is no obvious analytical model to use to fit the beam-weighted foreground spectrum, although many have used polynomial-based models (Monsalve et al. 2017; Sathyaranayana Rao et al. 2017; Bowman et al. 2018).

In Paper I of this series (Tauscher et al. 2018), we laid out a procedure for extracting the global signal from foregrounds without assuming a particular foreground model, but instead by simulating the foregrounds many times, with the parameters of these simulations varying between limits corresponding to realistic uncertainties. Using these simulations as a training set of foregrounds, the pipeline performs Singular Value Decomposition (SVD) to extract orthogonal basis vectors with which to fit the foreground. After performing the same process with the (much wider) training set of global signals, we fit the spectral data simultaneously with both SVD models and use the signal basis and the corresponding fit coefficients to construct confidence intervals on the 21 cm signal. The uncertainties on these intervals depend on the noise level of the data and the overlap between foreground and signal basis vectors.

In this paper, we use our pipeline to show that the overlap between foreground and 21 cm signal can be mitigated by utilizing time-dependent drift-scan measurements and observations of the four Stokes parameters describing polarization. The ability to use these extra pieces of data efficiently in constraining the signal is unique to our pipeline. While one can perform inference on drift-scan measurements using a
polynomial-based method, the connection between the foregrounds of the spectra (i.e., the fact that they come from the same beam and sky offset by some angle) cannot be fully accounted for. There is also no clear way to extend polynomial methods to Stokes parameters while utilizing the connection between them to help constrain the signal.

Our method of using SVD—which, for the present purposes, is equivalent to the diagonalization of a covariance matrix—to produce basis vectors is similar to past work performed in 21 cm cosmology. For example, Switzer & Liu (2014) estimated eigenmodes of the foreground frequency covariance matrix from data taken at different times (or equivalently, pointing directions). In our work, the modes are generated from a priori training sets based on previously and independently observed foreground spectra and simulated and/or measured beams instead of the sky-averaged radio spectra themselves. While utilizing the data to find modes is tempting because it relies less on a priori information, it is complicated by the fact that the foreground is never observed without the signal included.

Another method similar to the one discussed here is presented in Vedantham et al. (2014), henceforth denoted as V14, where the authors simulate foreground spectra at different time snapshots, stack the resulting spectra into a matrix, and perform SVD to retrieve eigenmodes. The key difference between this and our method is that V14 only derives modes as a function of frequency by performing SVD on one simulation of spectra from a series of times, whereas our technique utilizes multiple individual simulations, each of which contains spectra for a series of times, to produce modes that differ both as a function of frequency and as a function of time. The correlation of the spectra from time to time is vital to include in the model of the beam-weighted foreground data, as will be shown in Section 6, because even if the frequency modes of V14 can fit the beam-weighted foreground well, using them independently in each time-binned spectrum leads to a large overlap between foreground and signal models, producing large uncertainties. The extremely important role that utilizing spectrum-to-spectrum correlations caused by angular variations in the foreground play in precise measurements of the global 21 cm signal was seen clearly by Liu et al. (2013), who also, similarly to the pipeline first presented in Paper I, provided a generalizable method of producing a generic linear basis for the beam-weighted foreground across different angles through diagonalization of a covariance matrix.

In Paper II (Rapetti et al. 2020), we presented our pipeline’s strategy to translate from spectral constraints to nonlinear signal parameter constraints using a Markov Chain Monte Carlo (MCMC) algorithm, while analytically marginalizing over the same SVD-derived modes for the foreground as used in forming the spectral constraints at each step. This allows us to efficiently explore the MCMC parameter space of the nonlinear signal, fully accounting for complex foreground models from many correlated spectra. The latter is critical to extract the signal at the level required by standard 21 cm models, as we demonstrate in this paper, the third of the series.

In Section 2, we review the pipeline, with a particular focus on how the overlap between signal and foreground generates uncertainties in the signal extraction. In Section 3, we present how we simulate training sets using drift-scan measurements and how they help reduce overlap between foreground and signal. In Section 4, we do the same for measurements of the Stokes parameters by pairs of dipoles. In Section 5, we describe the simulation setup with which we test the benefits of including drift-scan and polarization measurements. In Section 6, we connect rms uncertainties to confidence levels and compare the uncertainties with and without polarization and drift-scan measurements. We conclude in Section 7.

2. Pipeline Review

2.1. Formalism

The basis of our pipeline is the formation of the data vector, \( y \), which contains a large number of individual spectra concatenated,

\[
y = y_{fg} + \Psi_{21} y_{21} + n,
\]

where \( y_{fg} \) and \( y_{21} \) are the true foreground and signal vectors, respectively, \( n \) is a random Gaussian noise vector with covariance \( C \), and \( \Psi_{21} \) is the so-called “signal expansion matrix,” explained further below. Here, \( y \), \( y_{fg} \), and \( n \) are vectors of length \( n_y n_v \), where \( n_v \) is the number of concatenated spectra in the data and \( n_y \) is the number of frequencies in each spectrum. Since the signal is a single spectrum (i.e., a vector of length \( n_v \)), it must be expanded into the full, length-\( n_y n_v \), space of \( y \). Expanding the signal into the dimensions of the full data vector while encoding information on how the data were obtained is the purpose of the signal expansion matrix \( \Psi_{21} \). Because \( \Psi_{21} y_{21} \) must be a length-\( n_y n_v \) vector and \( y_{21} \) is a length-\( n_v \) vector, \( \Psi_{21} \) is an \( n_y n_v \times n_v \) matrix. Examples of the signal expansion matrix in specific circumstances are provided in Sections 3.2, 4.4, and 5.1.\(^5\)

We model the data using weighted combinations of basis vectors contained in matrices denoted \( F_{fg} \) and \( F_{21} \), composed of the singular vectors of the foreground and signal training sets, respectively. These matrices are found via SVD and are normalized such that \( F_{fg}^T C^{-1} F_{fg} = I \) and \( F_{21}^T \Psi_{21}^T C^{-1} \Psi_{21} F_{21} = I \), where \( I \) is the identity matrix. The model of the data is

\[
\mathcal{M}(x_{fg}, x_{21}) = F_{fg} x_{fg} + \Psi_{21} F_{21} x_{21},
\]

where \( x_{fg} \) and \( x_{21} \) are weighting coefficients for the foreground and signal basis vectors, respectively. This is the same as \( \mathcal{M} = G x \) where \( G = [F_{fg} \ \Psi_{21} F_{21}] \) and \( x^T = [x_{fg} \ x_{21}^T] \). The probability distribution of the parameters is then taken to be proportional to the likelihood, given by

\[
\mathcal{L}(x) \propto \exp \left\{ -\frac{1}{2} (y - Gx)^T C^{-1} (y - Gx) \right\}.
\]

This implies that \( x \) is normally distributed with mean \( \xi \) and covariance \( S \), where

\[
S = (G^T C^{-1} G)^{-1} \text{ and } \xi = SG^T C^{-1} y.
\]

We then create signal confidence intervals centered on \( \gamma_{21} \) with a channel covariance \( \Delta_{21} \) given by

\[
\gamma_{21} = F_{21} \Delta_{21},
\]

\(^5\) In some applications, it is useful to define expansion matrices for more components than just the signal. For some examples, see Paper I. In this case, since the beam-weighted foreground training set is made of many sets of spectra covering the whole data space, there is no need for a foreground expansion matrix.
\[ \Delta_{21} = F_{21}S_{21}F_{21}^T, \tag{5b} \]

where \( \xi \) and \( S \) are the parts of \( \xi \) and \( S \) corresponding to the signal parameters. The 1\( \sigma \) root-mean-square (rms) uncertainty on the signal can then be defined as

\[ \text{rms}_{21}^\sigma = \sqrt{\text{Tr}(\Delta_{21}) / n_\nu}. \tag{6} \]

This mathematical formalism is implemented in the \texttt{pylinex} Python code.\(^6\)

### 2.2. Effect of Overlap on Uncertainties

From the reconstruction described by Equations (5), we can define the normalized rms error on the signal as

\[ \text{NRMS}_{21} = \sqrt{\frac{\text{Tr}(C^{-1/2} \Psi_2 \Delta_{21} \Psi_2^T C^{-1/2})}{n_\nu}}, \]

which is essentially the rms of the ratio of the \( 1 \sigma \) uncertainty level to the \( 1 \sigma \) noise level, leading to a unitless summary quantity that is 1 if the \( 1 \sigma \) posterior uncertainty level is the same size as the \( 1 \sigma \) noise level. It is given by

\[ \text{NRMS}_{21} = \sqrt{\frac{\text{Tr}(S_{21}^T F_{21}^T S_{21} F_{21})}{n_\nu}}. \tag{7a} \]

Through block inversion, it is possible to compute that

\[ S_{21} = (I - D^T D)^{-1} \]

where \( D = F_{21} C^{-1/2} \Psi_2^T F_{21} \) is the matrix of overlaps (dot products) between the foreground and signal basis vectors. The trace of \( S_{21} \) is therefore \( \sum_{j=1}^{n_2} \frac{1}{1 - \lambda_j} \), where \( n_2 \) is the number of signal vectors and \( \lambda_j \) are the eigenvalues of \( D^T D \).\(^7\) Thus,

\[ \text{NRMS}_{21} = \sqrt{\frac{1}{n_\nu} \sum_{j=1}^{n_2} \frac{1}{1 - \lambda_j}}. \tag{8} \]

If all foreground and signal basis vectors are orthogonal (i.e., \( D = 0 \), then the eigenvalues are all zero and \( \text{NRMS}_{21} \) reaches its minimum value of \( \sqrt{n_2} / n_\nu \). If, on the other hand, at least one foreground vector can be written as a combination of the signal vectors, or vice versa, then at least one of the eigenvalues of \( D^T D \) is 1 and \( \text{NRMS}_{21} \) diverges to \( \infty \). In general, \( \text{NRMS}_{21} \) lies between these two extremes. While we utilize \( \text{rms}_{21}^\sigma \) to report results in this paper, its normalized version \( \text{NRMS}_{21} \) is useful to illustrate how the overlap between the signal and foreground vectors leads to greater uncertainty. The same effect is present when computing \( \text{rms}_{21}^\sigma \), but this cannot be shown as clearly analytically.

Figure 1 shows a schematic explanation of how noise in data interacts with the overlap between signal and foreground basis vectors, causing the extraction of the signal. In this case, \( \Psi_2 \) is the identity matrix because the foreground and signal exist in the same space. The standard deviations of the one-dimensional confidence intervals on foreground and signal (lengths of blue and green line segments) are projections of the noise (red ellipse) onto the foreground and signal basis in a manner perpendicular to the other basis. In this simple case of two unit vectors, \( \text{NRMS}_{21} \propto \cos \alpha \), where \( \alpha \) is the angle between the unit vectors. So, as in the general case, as the basis vectors get closer to each other (\( \alpha \) gets smaller), the uncertainties grow.

### 3. Time Dependence with Drift-scan

#### 3.1. Drift-scan Formalism

To simulate drift-scan measurements for training sets made to fit data from a ground-based experiment, we compute the boresight direction of a zenith-pointing antenna at a given latitude, longitude, and Local Sidereal Time (LST). Using this direction and the orientation of the antenna with respect to geographic north, we can define a foreground power map that is a function of sky position (given in terms of antenna-based spherical coordinate angles \( \theta \) and \( \phi \), frequency, \( \nu \), and sidereal time, \( t \), as \( T(\theta, \phi, \nu, t) \)). Any real observation will take place over a finite time period, say from \( t_i \) to \( t_f \). The effective foreground seen by the antenna is a smeared version of the foreground created by an integral of \( T \), given by

\[ T_{\text{eff}}(\theta, \phi, \nu, t_i \rightarrow t_f) = \frac{1}{t_f - t_i} \int_{t_i}^{t_f} T(\theta, \phi, \nu, t) \, dt. \tag{9} \]

In practice, we split the time interval into \( n + 1 \) snapshots, so that the integral can be approximated by the following finite Riemann sum:

\[ T_{\text{eff}}(\theta, \phi, \nu, t_i \rightarrow t_f) = \frac{1}{n + 1} \sum_{k=0}^{n} T(\theta, \phi, \nu, t_k). \tag{10} \]

#### 3.2. Drift-scan Expansion Matrix

While the foreground changes as a function of time, the global 21 cm signal exists equally in every spectrum when using a drift-scan measurement strategy. Therefore, if there are \( n_{\text{drift}} \) measured spectra, the drift-scan expansion matrix for the 21 cm signal is

\[ \Psi_{21,\text{drift}}^T = \left[ \begin{array}{c} I \\ \vdots \\ I \end{array} \right]_{n_{\text{drift}} \times T}, \tag{11} \]

where \( I \) is the identity matrix. Because the signal does not change as the foreground changes, drift-scan measurements decrease the similarity between the foreground and signal models.

### 4. Observation of Stokes Parameters

Full Stokes measurements provide another excellent mechanism for reducing overlap between signal and foreground modes because foreground modes appear in all polarization modes whereas the global 21 cm signal appears only in Stokes \( I \) due to its lack of polarization and its isotropy. To include Stokes parameters in any analysis, however, one must first accurately simulate observations including them. This section presents a formalism for simulating full Stokes observations, splitting the results up into two terms: the induced polarization term that comes from projection of unpolarized radiation onto the antenna plane, and the intrinsic polarization term that comes from polarized foreground sources. Simulations using this formalism will later be used to generate training sets with the
Figure 1. Simplified schematic representation of how the overlap between signal and foreground modes increases the uncertainties of both individually separated components with respect to the minimum level determined by the statistical noise. Red circle represents the 2σ noise uncertainty of the data (red vector). Blue and green vectors whose tails sit on the origin represent the signal and foreground basis vectors, respectively. Blue and green intervals demarcated by solid circles are the 1D uncertainties are proportional to $|\text{csc} \alpha|$, where $\alpha$ is the angle between the unit vectors.

The purpose of computing modes that encode correlations between the different Stokes parameters and with which to fit the beam-weighted foreground.

In this section, we outline methods of performing simulations of beam-weighted foreground measurements based on the Jones matrix (see Jones 1941), which was first introduced to describe polarization measurements and coordinate transformations for optical systems, but has since been used in addition for other wavelengths, such as at cosmic microwave background frequencies (see, e.g., O’Dea et al. 2007; Chuss et al. 2012). Data from radio antennas are caused by electric fields, $E_i$, from the sky, which are written in terms of $\theta$ and $\phi$ components, $E_\theta$ and $E_\phi$, i.e., $E_i = E_\theta \hat{\theta} + E_\phi \hat{\phi}$ (note that there is no $\hat{r}$ component of the electric field because the radiation is traveling in the $-\hat{r}$ direction), where $\theta = 0$ is the pointing direction of the antenna. In general, $E_\theta$ and $E_\phi$ are complex, making $E_i$ a complex random vector. The Stokes parameters of the sky radiation, $I_i$, $Q_i$, $U_i$, and $V_i$, which are the real power-unit quantities measuring polarization, are then given by

$$I_i = \langle |E_i|^2 \rangle,$$

$$Q_i = \langle |E_i|^2 - |E_\theta|^2 \rangle,$$

$$U_i = \langle 2 \text{Re}(E_\theta E_\phi^*) \rangle,$$

$$V_i = \langle 2 \text{Im}(E_\theta E_\phi^*) \rangle,$$

where $\langle \ldots \rangle$ denotes the expectation value. This can be written as $P_i = \langle E_i^\dagger \sigma P E_i \rangle$, where $\dagger$ represents the Hermitian transpose and

$$\sigma_I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_Q = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

are the Pauli matrices (Fano 1954). Here, $E_i$ is a function of both sky position and frequency, so $P_i$ is as well.

### 4.1. Sky Polarization

Assuming that there is no coherent radiation coming from the sky, the expectation value of the electric field is zero, $\langle E_i \rangle = 0$. Since $E_i$ is coming from many different electrons (in the case of synchrotron emission) and every phase is equally probable, $E_i$ follows a circularly symmetric complex normal distribution with probability density

$$f(E_i) = \frac{\exp(-E_i^\dagger \Sigma_i^{-1} E_i)}{\pi^{3/2} |\Sigma_i|},$$

where $\Sigma_i = \langle E_i E_i^\dagger \rangle$ is the Hermitian covariance matrix. With this probability density, the expected values of the Stokes parameters are given by

$$P_i = \text{Tr}(\sigma P \Sigma_i).$$

Since the distribution of $E_i$ can represent any elliptical shape around the origin, it can be decomposed into the sum of two independent normally distributed vectors, one with a circular covariance matrix (i.e., proportional to the identity matrix) and another that exists only along a line, specified by a complex vector $v_i$, satisfying $v_i^\dagger v_i = 1$. This means that $\Sigma_i$ can be written as $\Sigma_i = \alpha_s \mathbf{I} + \beta_s v_i v_i^\dagger$, where $\mathbf{I}$ is the $2 \times 2$ identity matrix and $\alpha_s$ and $\beta_s$ are non-negative, so $P_i = \alpha_s \text{Tr}(\sigma P) + \beta_s v_i^\dagger \sigma P v_i$. To

---

8. See Appendix C for an example of how to compute expectation values with this probability density form.
interpret \( \alpha_s \) and \( \beta_s \), we write the expression for the total intensity of the sky radiation, \( I_s \), by plugging in \( \sigma_p = \sigma_0 = I \) and using \( v^T_i v_i = 1 \). We find \( I_s = 2\alpha_s + \beta_s \). Since \( \alpha_s \) is the coefficient in front of the circular covariance matrix, it must involve only unpolarized radiation; ergo, we write \( \alpha_s = [(1 - p_s)/2]I_s \) where \( 0 \leq p_s \leq 1 \) is the polarization fraction of the sky radiation, leaving us with \( \beta_s = p_s I_s \). This means that \( \Sigma_s = [(1 - p_s)/2]I_s \) and \( p_s = \frac{1}{2} I_s \) Tr(\( \sigma_p \)) + \( p_s I_s v^T_i \sigma_p v_i \). (16)

In these expressions, \( I_s \) can be taken from total power maps of the sky at a given frequency. The values of \( p_s \) and \( \psi_s \) can be determined from \( Q_s \), \( U_s \), and \( V_s \) using Equation (16) and noting that \( \text{Tr}(\Sigma_s) = \text{Tr}(\sigma_0) = \text{Tr}(\sigma_I) = 0 \). If there is no circular polarization coming from the sky, \( V_s = 0 \), implying that both components of \( v_i \) have the same phase, meaning that, up to an arbitrary phase, it can be expressed through \( v_i^T = [\cos \psi_i \sin \psi_i] \). Plugging this expression into Equation (16), \( Q_s \) and \( U_s \) can be written in the \( V_s = 0 \) case as

\[
Q_s + iU_s = p_s I_s e^{2i\psi_i}.
\]

Therefore, in this case,

\[
p_s = \left| \frac{Q_s + iU_s}{I_s} \right| \quad \text{and} \quad \psi_s = \frac{1}{2} \arg(Q_s + iU_s).
\]

The random vector \( E_a \) can be written as the sum of two independent random vectors, \( A_s \) with covariance \([(1 - p_s)/2]I_s \) and \( B_s \) with covariance \( p_s I_s v^T_i v_i \). Both \( A_s \) and \( B_s \) contribute to \( I_s \), but only \( B_s \) contributes to \( Q_s \) and \( U_s \).

4.2. Antenna Polarization

The electric fields induced in the antenna can be written as \( E_a = E_x \hat{x} + E_y \hat{y} \) where \( \hat{x} \) and \( \hat{y} \) are the (generally orthogonal) antenna polarization directions. Here, \( E_a \) is derived from \( E_s \) through a matrix known as the Jones matrix, \( J \):

\[
E_a = JE_s,
\]

or, equivalently,

\[
\begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} = \begin{bmatrix} J_{xx} & J_{xy} & J_{xz} \\ J_{yx} & J_{yy} & J_{yz} \\ J_{zx} & J_{zy} & J_{zz} \end{bmatrix} \begin{bmatrix} E_0 \\ E_1 \end{bmatrix}
\]

(20)

We can now solve for the Stokes parameters seen by the antennas by using the complex random vectors \( A_s \) and \( B_s \) in Equation (19), i.e., \( E_0 = J(A_s + B_s) \). This implies

\[
P_a = \langle E^T_a \sigma_p E_a \rangle,
\]

(21a)

\[
= \langle A^T_s J^T \sigma_p J A_s \rangle + \langle B^T_s J^T \sigma_p J B_s \rangle.
\]

(21b)

where the last line follows because \( A_s \) and \( B_s \) are zero-mean and independent. Using the covariances of \( A_s \) and \( B_s \) derived in Section 4.1, this can be written

\[
P_a = \left( \frac{1-p_s}{2} \right) I_s \text{Tr}(J^T \sigma_p J) + p_s I_s v^T_i J^T \sigma_p J v_i.
\]

(22)

As opposed to the sky polarization case, in general, both of these terms contribute to the observed Stokes parameters. The term \( \text{Tr}(J^T \sigma_p J) \) encodes Stokes parameters induced from the unpolarized radiation from the sky, while the second term encodes the effect of polarization intrinsic to the sky, so we term them induced and intrinsic polarization, respectively. In Appendix A, we derive the observed Stokes parameters for the Jones matrix of ideal orthogonal dipoles. Figure 2 shows an intuitive cartoon of the induced polarization component for ideal orthogonal dipoles.

The electric field, \( E_{\text{a,tot}} \), measured by the instrument at each frequency of every spectrum is the sum of the electric fields from all sky positions, \( E_{\text{a,tot}}(\nu) = \int E_a(\nu, \theta, \phi) \, d\Omega \). Since \( E_a(\nu, \theta, \phi) \) is zero-mean with covariance \( \Sigma_a(\nu, \theta, \phi) \) and is independent at each sky position \((\theta, \phi), E_{\text{a,tot}}(\nu) \) is zero-mean with covariance \( \Sigma_{a,tot}(\nu) \) where \( \Sigma_{a,tot}(\nu) = \int \Sigma_a(\nu, \theta, \phi) \, d\Omega \), which implies that the Stokes parameters at each frequency are given by \( P_{a,tot}(\nu) = \int P_a(\nu, \theta, \phi) \, d\Omega \). To calibrate the Stokes
parameters so that antenna temperatures correspond to actual sky brightness temperatures, we consider a case where \( p_s = 0 \)
and \( I_s \) is independent of angle and equal to \( I_0 \). In this case, the calibrated total power, \( I_{a,\text{cal}} \), should be equal to \( I_0 \). By implementing this with a multiplicative factor, we find that
\[
P_{a,\text{cal}}(\nu) = \frac{2P_{a,\text{tot}}(\nu)}{\int \text{Tr}(J^J) \, d\Omega},
\]
where
\[
P_{a,\text{cal}}(\nu) = \int (1 - p_s) I_s \text{Tr}(J^J \sigma_J) \, d\Omega
+ 2 \int p_s I_s \frac{|q_v|^2}{|q|^2} \, d\Omega.
\]
Using this factor, we can also define a calibrated total electric field, \( E_{a,\text{cal}} = \sqrt{2} E_{a,\text{tot}} / \sqrt{\int \text{Tr}(J^J) \, d\Omega} \), and covariance matrix,
\[
\Sigma_{a,\text{cal}} = 2\Sigma_{a,\text{tot}} / \int \text{Tr}(J^J) \, d\Omega.
\]

The Jones matrix-based formalism used here is equivalent to the commonly used Mueller matrix-based formalism. The connection between the Jones and Mueller formalisms is laid out in Appendix B. It is worthwhile to note that the Mueller matrix is proportional to a product including two factors of the Jones matrix, just like both terms in Equation (22) have two factors of \( J \).

4.3. Neglecting Intrinsic Polarization

If the total intensity of the sky, \( I_s \), is known, but intrinsic polarization is neglected in a prediction of the antenna Stokes parameters, then there is an unmodeled residual effect given by
\[
\Delta^{(\text{P})}_{a,\text{cal}} = P_{a,\text{cal}} - \int I_s \text{Tr}(J^J \sigma_J) \, d\Omega.
\]
\[
= \int p_s I_s \frac{|q_v|^2}{|q|^2} \, d\Omega.
\]
If an experiment has only one antenna, then \( J \) becomes a row vector instead of a square matrix. Defining \( q \) as the column vector \( J \), the single antenna power signal \( I_{a,\text{cal}} \), analogous to Equation (23), is given by
\[
I^{\text{1-ant}}_{a,\text{cal}} = \int (1 - p_s) I_s \frac{|q|^2}{|q|^2} \, d\Omega
+ 2 \int p_s I_s \frac{|q_v|^2}{|q|^2} \, d\Omega.
\]
This reflects the fact that there are $4n_{\text{drift}}$ spectra in the data and the signal is in every fourth spectrum (i.e., the Stokes $I$ spectra). In the opposite case, where neither drift-scan nor polarization measurements are used, the signal expansion matrix is simply the identity matrix.

5.2. Beam-weighted Foreground Training Set

In principle, the beam-weighted foreground training set is created from two sources, antenna beam variations and spectral foreground maps. However, in this paper, as in Paper I, we use one foreground map: the map given by Haslam et al. (1982) scaled with a spectral index of $-2.5$, and many beams. The beams are defined using a Jones matrix derived from that of ideal orthogonal dipole antennas (see Appendix A) modulated by an angular Gaussian whose angular scale, $\alpha$, is a function of frequency, $\nu$, allowing for beam chromaticity to be robustly included in the analysis. The full Jones matrices take the form

$$J = \exp\left(-\frac{\theta^2}{4\alpha^2(\nu)^2}\right)\begin{bmatrix} \cos \theta \cos \phi & -\sin \phi \\ \cos \theta \sin \phi & \cos \phi \end{bmatrix}. \quad (30)$$

Since, as mentioned in Section 4.2, the measured Stokes parameters depend on two powers of the Jones matrix, the effective beam (i.e., the Mueller matrix; see Appendix B) is proportional to $\exp\left(-\frac{\theta^2}{2\alpha^2(\nu)^2}\right)$. Thus, the Full Width at Half Maximum (FWHM) is given by $\text{FWHM}(\nu) = \sqrt{8\ln2} \alpha(\nu)$.

We vary FWHM$(\nu)$ between training set elements. For the sake of simplicity, we use FWHM$(\nu)$ curves given by quadratic polynomials in frequency. Instead of choosing the coefficients of each power of frequency independently, we utilize Legendre polynomials for easier control over the magnitude of variations, i.e.,

$$\text{FWHM}(\nu) = \sum_{k=0}^{2} a_k L_k\left(\frac{\nu - \nu_0}{\delta\nu}\right), \quad (31)$$

where $\nu_0 = (\nu_{\text{max}} + \nu_{\text{min}})/2$ is the average frequency, $\delta\nu = (\nu_{\text{max}} - \nu_{\text{min}})/2$ is half the width of the frequency band, and

$$L_0(x) = 1, \quad L_1(x) = x, \quad L_2(x) = \frac{3x^2 - 1}{2}. \quad (32)$$

In our case, where $\nu_{\text{min}} = 40 \text{ MHz}$ and $\nu_{\text{max}} = 120 \text{ MHz}$, $\nu_0 = 80 \text{ MHz}$ and $\delta\nu = 40 \text{ MHz}$. To seed the beam variations in our training set, we draw $a_0$, $a_1$, and $a_2$ from independent normal distributions,

$$a_k \sim \mathcal{N}(\mu_k, \sigma_k^2), \quad (33)$$

with the means and standard deviations $\mu_k$ and $\sigma_k$ given in Table 1. An extra constraint is applied to exclude FWHM$(\nu)$ curves that dip below 20° or rise above 150° in the 40–120 MHz band. The resulting training set of FWHM curves is shown in the left panel of Figure 3.

We simulate observed Stokes parameters with Equation (23) using antenna Jones matrices of the form of Equation (30) with the FWHM functions described above pointing at zenith from

12 Future work will include variations of the foreground map.
first introduced in Tauscher et al. (2018) as
\[
\varepsilon = \sqrt{\frac{1}{n_{\nu}} \sum_{i=1}^{n_{\nu}} \left( \frac{(\gamma_{21} - y_{21})}{(\Delta_{21})_{i}} \right)^2},
\]
where \( y_{21} \) is the input 21 cm signal and \( \gamma_{21} \) and \( \Delta_{21} \) are given in Equations 5(a) and (b). The rms uncertainty of the interval known to include the signal is denoted by \( \text{rms}_{21} \) and is formed by the product of Equations (6) and (35):
\[
\text{rms}_{21} = \varepsilon \text{ rms}^{17}_{21},
\]
\[
= \frac{1}{n_{\nu}} \sum_{i=1}^{n_{\nu}} \left( \frac{(\gamma_{21} - y_{21})}{(\Delta_{21})_{i}} \right)^2.
\]

Using the values of \( \text{rms}_{21} \) for each of the 5000 fits in every case studied, we make a Cumulative Distribution Function (CDF) defined by
\[
\text{CDF}(x) = \text{Pr}[\varepsilon \text{ rms}_{21} < x].
\]

We interpret the values of this CDF as confidence levels for future fits in which \( y_{21} \) is unknown. A CDF for each of our four cases is plotted in Figure 4. Table 2 shows the rms levels corresponding to 68%, 95%, and 99% uncertainties in each case. Clearly, using multiple time bins leads to more robust fits than using a single averaged spectrum, and leveraging all four Stokes parameters yields better fits than using only Stokes I.

Figure 5 shows the confidence level as a function of rms uncertainty for various numbers of LST bins with and without polarization. In both panels, it is clear that it is necessary to include more time bins to achieve reasonable errors, but they eventually saturate at around 5–10 bins due to the size of the beams used in our simulations.

So far in this paper, it has been assumed that the foreground basis vectors exist across all time bins. However, a common method of analysis is to treat every spectrum as independent and model them separately, even though they are being fit simultaneously. Figure 6 shows the effects of this key difference between the two analyses. When each time bin has its own basis vectors, the benefit of using multiple time bins is severely damped. From this, it is clear that, in order to fully benefit from fitting all spectra simultaneously, it is imperative to do so using a single matrix, with the basis vectors spanning all time bins, as opposed to using independent basis vectors in each spectrum.
Conclusions

In this third paper of the series, we have defined a method for converting rms uncertainty to a probabilistic confidence level when using the pipeline we first introduced in Tauscher et al. (2018) (Paper I). We then applied this method to different sets of simulated data representing the global 21 cm signal and foregrounds, with the purpose of testing the benefits of measuring time-binned drift-scan data and Stokes parameters. The largest impact we have found was from the use of drift-scan spectra, which can be done with any global signal experiment. By using the correlations between different time bins and enforcing that the signal must be constant from spectrum to spectrum, we found that fitting multiple time bins instead of only one can decrease uncertainties from the few K level to the few mK level. It is important to note that this large benefit is not seen if using instead the traditional method where spectra are modeled independently, even if they are fit simultaneously.

Measurements of all four Stokes parameters with dual-antenna systems also proved useful in simulations to reduce uncertainties—and when in combination with the drift-scan strategy, can lead to uncertainties approaching the radiometer noise level. However, for both single and dual antenna experiments, extra care must be taken to model the effects of intrinsic sky polarization. If neither of these two independent strategies is used (i.e., if analysis is done with only a single total power spectrum, such as in Bowman et al. (2018)), then the uncertainties are consistently at the few K level.

We thank David Bordenave and Bang Nhan for helpful discussions on polarization measurements. We also thank Neil Bassett and Joshua Hibbard for feedback. D.R. was supported by a National Aeronautics and Space Administration (NASA) Postdoctoral Program Senior Fellowship at the NASA Ames Research Center, administered by the Universities Space Research Association under contract with NASA. This work...
Appendix A

Ideal Dipoles

In this appendix, we consider the Jones matrix of an orthogonal pair of ideal dipoles as given by

$$ J = \begin{bmatrix} \cos \theta \cos \phi - \sin \phi \\ \cos \theta \sin \phi \cos \phi \\ \cos \theta \sin \phi \sin \phi \end{bmatrix}, $$

(A1)

which simply encodes a geometrical projection from electric fields on the celestial sphere to electric fields on the X and Y antennas. This is the matrix from the original Jones (1941) work (see Equation (9) of that paper), generalized to account for radiation coming from directions off the zenith angle (i.e., nonzero \( \theta \)). Since \( J \) is real in this case, \( J^* = J^T \).

A.1. Induced Polarization

The induced portion of measured polarization (first term of Equation (22)) is

$$ (P_a)_{ind} = \left( \frac{1 - p_s}{2} \right) I_i \text{Tr}(J^T \sigma_P J). $$

(A2)

Computing the trace using Equation (A1), we find

$$ (I_a)_{ind} = \left( \frac{1 - p_s}{2} \right) I_i (1 + \cos^2 \theta), $$

(A3a)

$$ (Q_a + iU_a)_{ind} = -\left( \frac{1 - p_s}{2} \right) I_i e^{2i\phi} \sin^2 \theta, $$

(A3b)

$$ (V_a)_{ind} = 0. $$

(A3c)

Equations A3(a), (b), and (c) with \( p_s = 0 \) are the origin of the beam defined in Equation (9) of Tauscher et al. (2018).

A.2. Linear Intrinsic Polarization

The intrinsic portion of measured polarization (second term of Equation (22)) is

$$ (P_a)_{int} = p_s I_i (J_v) S \sigma_P (J_v). $$

(A4)

Using Equation (A1) and the definition of \( v_i \) in terms of \( \psi_i \) (which applies when \( V_i = 0 \)), we find that

$$ J_v = \begin{bmatrix} \cos \theta \cos \phi - \sin \phi \\ \cos \theta \sin \phi \cos \phi \\ \cos \theta \sin \phi \sin \phi \end{bmatrix} \begin{bmatrix} \cos \psi_i \\ \cos \psi_i \sin \psi_i \\ \sin \psi_i \cos \psi_i + \cos \phi \sin \psi_i \end{bmatrix}, $$

(A5a)

$$ = \begin{bmatrix} \cos \theta \cos \phi \cos \psi_i - \sin \phi \sin \psi_i \\ \cos \theta \sin \phi \cos \psi_i + \cos \phi \sin \psi_i \\ \cos \theta \sin \phi \sin \psi_i \end{bmatrix}. $$

(A5b)

This means that

$$ (I_a)_{int} = p_s I_i (1 - \sin^2 \theta \cos^2 \psi_i), $$

(A6a)

$$ (Q_a + iU_a)_{int} = p_s I_i e^{2i\phi} (\cos 2\psi_i - \sin^2 \theta \cos^2 \psi_i) + i \cos \theta \sin 2\psi_i, $$

(A6b)

$$ (V_a)_{int} = 0. $$

(A6c)

A.3. Combined Results

The total power seen by the antennas is given by the sum of Equations A3(a) and A6(a), while the polarization signal seen by the antenna is given by the sum of Equations A3(b) and A6(b). After normalizing so that \( I_a = I_0 \) and \( Q_a - U_a = V_a = 0 \) yield \( I_{a,\text{cal}} = I_0 \) (see Equation (23)), we find that the calibrated antenna temperatures are

$$ I_{a,\text{cal}} = \frac{3}{16\pi} \int I_i [(1 + \cos^2 \theta) - p_s \sin^2 \theta \cos 2\psi_i] d\Omega, $$

(A7a)

$$ Q_{a,\text{cal}} + iU_{a,\text{cal}} = \frac{3}{16\pi} \int I_i e^{2i\phi} \times [(p_s (1 + \cos^2 \theta) \cos 2\psi_i - \sin^2 \theta) + 2ip_s \cos \theta \sin 2\psi_i] d\Omega, $$

(A7b)

$$ V_{a,\text{cal}} = 0. $$

(A7c)

These equations can be generalized to the case where there is intrinsic circular polarization. If it is assumed, as in the simulations of this paper, that no sky sources are intrinsically polarized (\( p_s = 0 \)), then

$$ I_{a,\text{cal}} = \frac{1}{16\pi} \int I_i (1 + \cos^2 \theta) d\Omega \quad \text{and} \quad Q_{a,\text{cal}} + iU_{a,\text{cal}} = -\frac{3}{16\pi} \int I_i e^{2i\phi} \sin^2 \theta d\Omega. $$

Appendix B

Connection to the Mueller Matrix Formalism

Equation (15) states that, in the absence of coherent radiation, the Stokes parameters in a given basis are the trace of the product of the covariance matrix of electric fields in that basis with the Pauli matrices \( P_X = \text{Tr}(\Sigma X \sigma_P) \). Since \( \sigma_P \) form a complete orthogonal basis of \( 2 \times 2 \) Hermitian matrices, subject to the inner product defined by \([A, B] = \text{Tr}(A B)\), we can write

$$ \Sigma_X = \sum_{P \in \{I, Q, U, V\}} \frac{\text{Tr}(\Sigma_X \sigma_P)}{\text{Tr}(\sigma_P^2)} \sigma_P. $$

(B1)

Since \( \text{Tr}(\sigma_P^2) = \text{Tr}(I) = 2 \) and \( \text{Tr}(\Sigma_X \sigma_P) = P_X \), this means that

$$ \Sigma_X = \frac{1}{2} \sum_{P \in \{I, Q, U, V\}} P_X \sigma_P. $$

(B2)

Plugging in \( X = s \), multiplying on the left by \( J \) and on the right by \( J^T \), and noting that \( \Sigma_a = J \Sigma \sigma J^T \), we find

$$ \Sigma_a = \frac{1}{2} \sum_{P \in \{I, Q, U, V\}} P_a J \sigma J^T. $$

(B3)

Writing \( P'_a = \text{Tr}(\Sigma_a \sigma P) \) through \( P'_a = \sum_{P \in \{I, Q, U, V\}} P_a \mathcal{M}_{P \rightarrow P'} \), we can then write

$$ \mathcal{M}_{P \rightarrow P'} = \frac{1}{2} \text{Tr}(J \sigma J^T \sigma P'). $$

(B4)

The Mueller matrix is normalized by the integral over the \( I \rightarrow I_a \) element, \( \mathcal{M}_{a \rightarrow P' a} = \mathcal{M}_{P \rightarrow P' a} \int \mathcal{M}_{I \rightarrow I_a} d\Omega \). This normalized Mueller matrix satisfies

$$ \mathcal{M}_{a \rightarrow P' a}^{(\text{norm})}(\theta, \phi, \nu) = \frac{\text{Tr} \left\{ \left[ J(\theta, \phi, \nu) \right] \sigma_P \left[ J(\theta, \phi, \nu) \right]^* \right\}}{\text{Tr} \left\{ \left[ J(\theta, \phi, \nu) \right]^* \left[ J(\theta, \phi, \nu) \right] \right\}} d\Omega, $$

(B5)
The total calibrated antenna Stokes parameters are given by

\[
P_{a,\text{cal}}(\nu) = \sum_{\nu' \in \{L, Q, U, V\}} \int M_{\nu' \rightarrow \nu}^{(\text{norm})}(\theta, \phi, \nu') P'_a(\theta, \phi, \nu') \, d\Omega.
\]  
(B6)

When assuming that there are no polarized sky sources, as in the simulations of this paper, the Mueller matrix effectively becomes a column vector with elements

\[
M_{\nu' \rightarrow \nu}^{(\text{norm})} = \frac{\text{Tr} \{ [J(\theta, \phi, \nu')] \Psi_{\nu} [J(\theta, \phi, \nu')] \}}{\int \text{Tr} \{ [J(\theta, \phi, \nu')] [J(\theta, \phi, \nu')] \} \, d\Omega},
\]  
(B7)

and the calibrated Stokes parameters can be written

\[
P_{a,\text{cal}}(\nu) = \int M_{\nu' \rightarrow \nu}^{(\text{norm})}(\theta, \phi, \nu) I_\nu(\theta, \phi, \nu) \, d\Omega.
\]  
(B8)

For the orthogonal ideal dipole Jones matrix defined in Appendix A, the full Mueller matrix is given by

\[
M^{(\text{norm})}(\theta, \phi) = \frac{3}{16\pi} \begin{bmatrix}
1 + \cos^2 \theta & -\sin^2 \theta \\
-\sin^2 \theta \cos 2\phi & (1 + \cos^2 \theta) \cos 2\phi - 2 \cos \theta \sin 2\phi \\
-\sin^2 \theta \sin 2\phi & (1 + \cos^2 \theta) \sin 2\phi - 2 \cos \theta \cos 2\phi \\
0 & 2 \cos \theta
\end{bmatrix},
\]  
and the first column is the effective Mueller matrix when \( p_x = 0 \).

**Appendix C**

**Noise on Stokes Parameters**

Denoting the average of a quantity \( X \) over all \( n_s = \Delta \nu \Delta t \) spectra by \( \overline{X} \), Equation (28) is \( P_{a,\text{ave}}(\nu) = \overline{P_{a,\text{cal}}(\nu)} \). The squared noise level on the averaged, measured Stokes parameters is given by

\[
\text{Var}[P_{a,\text{ave}}(\nu)] = \text{Var}\left[ \frac{1}{\Delta \nu \Delta t} \sum_{k=1}^{\Delta \nu \Delta t} P_{a,\text{cal}}^{(k)}(\nu) \right],
\]  
(C1a)

\[
= \frac{\text{Var}[P_{a,\text{cal}}(\nu)]}{\Delta \nu \Delta t},
\]  
(C1b)

\[
= \frac{\langle P_{a,\text{cal}}^{2}(\nu) \rangle - \langle P_{a,\text{cal}}(\nu) \rangle^2}{\Delta \nu \Delta t},
\]  
(C1c)

\[
= \frac{\langle (E_{a,\text{cal}}^{(k)} \Psi_{\nu} E_{a,\text{cal}}^{(k)})^2 \rangle - \langle (E_{a,\text{cal}}^{(k)} \Psi_{\nu} E_{a,\text{cal}}^{(k)}) \rangle^2}{\Delta \nu \Delta t},
\]  
(C1d)

where \( \langle \ldots \rangle \) represents the expectation value and \( \text{Var}[\ldots] \) represents the variance. Because the electric field \( E_{a,\text{cal}}^{(k)} \) follows a complex normal distribution with zero mean and covariance \( \Sigma_{a,\text{cal}}^{(k)} \), the expectation value of an arbitrary function of \( E_{a,\text{cal}}^{(k)} \) is defined as

\[
\langle h(E_{a,\text{cal}}^{(k)}) \rangle = \frac{1}{\pi^2 |\Sigma_{a,\text{cal}}^{(k)}|} \int h(x) e^{-x^* \Sigma_{a,\text{cal}}^{(k)^{-1}} x} \, dx,
\]  
(C2)

where \( x \) is a complex 2D vector. By performing integrals of this form, we can find that

\[
\langle (E_{a,\text{cal}}^{(k)} \Psi_{\nu} E_{a,\text{cal}}^{(k)}) \rangle^2 = \text{Tr}[(\sigma_{\nu} \Sigma_{a,\text{cal}}^{(k)})^2] + (\text{Tr}[(\sigma_{\nu} \Sigma_{a,\text{cal}}^{(k)})])^2.
\]  
(C3b)

Plugging these expressions into Equation (C1d), we can compute that

\[
\text{Var}[P_{a,\text{ave}}(\nu)] = \frac{\text{Tr}[(\sigma_{\nu} \Sigma_{a,\text{cal}}^{(k)})^2]}{\Delta \nu \Delta t}.
\]  
(C4)

Now, we write

\[
\sigma_{\nu} = \begin{bmatrix}
\delta_{\nu L} + \delta_{\nu Q} & \delta_{\nu P} - i \delta_{\nu V} \\
\delta_{\nu P} + i \delta_{\nu V} & \delta_{\nu L} - \delta_{\nu Q}
\end{bmatrix},
\]  
(C5)

where \( \delta_{\nu P'} \equiv \begin{cases} 1 & P = P' \\
0 & P \neq P' \end{cases} \) (Equations (13)). This essentially encodes \( \sigma_{\nu} \) in a single matrix. With this same definition of \( \delta_{\nu P'} \), we can write

\[
P_{a,\text{cal}}^{(k)} = \delta_{\nu L} I_{a,\text{cal}}^{(k)} + \delta_{\nu Q} Q_{a,\text{cal}}^{(k)} + \delta_{\nu P} P_{a,\text{cal}}^{(k)} + \delta_{\nu V} V_{a,\text{cal}}^{(k)}.
\]  
(C6)

Using these definitions of \( \sigma_{\nu} \) and \( P_{a,\text{cal}}^{(k)} \) in Equation (B2), we can write

\[
\Sigma_{a,\text{cal}}^{(k)} = \begin{bmatrix}
I_{a,\text{cal}}^{(k)} + Q_{a,\text{cal}}^{(k)} & U_{a,\text{cal}}^{(k)} - i V_{a,\text{cal}}^{(k)} \\
U_{a,\text{cal}}^{(k)} + i V_{a,\text{cal}}^{(k)} & I_{a,\text{cal}}^{(k)} - Q_{a,\text{cal}}^{(k)}
\end{bmatrix},
\]  
(C7)

Plugging these expressions into Equation (C4) for Stokes \( I \), we compute

\[
\text{Var}[I_{a,\text{ave}}] = \frac{\text{Tr}[(\sigma_{\nu} \Sigma_{a,\text{cal}}^{(k)})^2]}{\Delta \nu \Delta t} = \frac{\text{Tr}[\Sigma_{a,\text{cal}}^{2}]}{\Delta \nu \Delta t},
\]  
(C8a)

\[
\text{Var}[I_{a,\text{ave}}] = \frac{\text{Tr}[\Sigma_{a,\text{cal}}^{2}]}{\Delta \nu \Delta t}.
\]  
(C8b)

\footnote{Note that \( \delta_{\nu P'}^2 = \delta_{\nu P'} \).}
\[
\begin{align*}
\frac{1}{4\Delta \nu \Delta t} \text{Tr} & \left( \left( U_{a,\text{cal}} + Q_{a,\text{cal}} \right)^2 + V_{a,\text{cal}}^2 \right) \\
& \quad - \left( U_{a,\text{cal}} + Q_{a,\text{cal}} - Q_{a,\text{cal}} \right)^2 \\
& \quad - \frac{1}{2} \left( U_{a,\text{cal}} + V_{a,\text{cal}} \right) \\
\end{align*}
\]

(C8c)

\[
\begin{align*}
\frac{1}{2\Delta \nu \Delta t} & \left( \left( U_{a,\text{cal}} + Q_{a,\text{cal}} \right)^2 + \frac{1}{2} \left( U_{a,\text{cal}} + Q_{a,\text{cal}} \right)^2 \\
& \quad + \frac{1}{2} \left( U_{a,\text{cal}} - Q_{a,\text{cal}} \right)^2 \right) \\
\end{align*}
\]

(C8d)

where the off-diagonal elements in Equation C8(c) are left out for clarity. By performing similar calculations for the other Stokes parameters, we find

\[
\begin{align*}
\text{Var}[Q_{a,\text{ave}}] & = \left( Q_{a,\text{cal}}^2 + U_{a,\text{cal}}^2 + V_{a,\text{cal}}^2 \right) \frac{1}{2 \Delta \nu \Delta t}, \\
\text{Var}[U_{a,\text{ave}}] & = \left( U_{a,\text{cal}}^2 + Q_{a,\text{cal}}^2 + V_{a,\text{cal}}^2 \right) \frac{1}{2 \Delta \nu \Delta t}, \\
\text{Var}[V_{a,\text{ave}}] & = \left( V_{a,\text{cal}}^2 + Q_{a,\text{cal}}^2 + U_{a,\text{cal}}^2 \right) \frac{1}{2 \Delta \nu \Delta t},
\end{align*}
\]

(C9a, C9b, C9c)

ORCID iDs

Keith Tauscher @ https://orcid.org/0000-0003-1932-9829
David Rapetti @ https://orcid.org/0000-0003-2196-6675
Jack O. Burns @ https://orcid.org/0000-0002-4468-2117

References

Bowman, J. D., Rogers, A. E. E., Monsalve, R. A., Mozdzen, T. J., & Mahesh, N. 2018, Natur, 555, 67
Chuss, D. T., Wollack, E. J., Pisano, G., et al. 2012, ApOpt, 51, 6824
Condon, J., & Ransom, S. 2016, Essential Radio Astronomy, Vol. 4 (1st ed.; Princeton, NJ: Princeton Univ. Press)
Fano, U. 1954, PhRv, 93, 121
Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, PhR, 433, 181
Haslam, C. G. T., Salter, C. J., Stoffel, H., & Wilson, W. E. 1982, A&AS, 47, 1
Hellwig, H., Vessot, R., Levine, M., et al. 1970, ITIM, 19, 200
Jones, R. C. 1941, JOSA, 31, 488
Liu, A., Pritchard, J. R., Tegmark, M., & Loeb, A. 2013, PhRvD, 87, 043002
Mirocha, J. 2014, MNRAS, 443, 1211
Monsalve, R. A., Rogers, A. E. E., Bowman, J. D., & Mozdzen, T. J. 2017, ApJ, 847, 64
Morales, M. F., & Wyithe, J. S. B. 2010, ARA&A, 48, 127
Nhan, B. D., Bordenave, D. D., Bradley, R. F., et al. 2019, ApJ, 883, 126
O’Dea, D., Challinor, A., & Johnson, B. R. 2007, MNRAS, 376, 1767
Pritchard, J. R., & Loeb, A. 2010, PhRvD, 82, 023006
Pritchard, J. R., & Loeb, A. 2012, RPPh, 75, 086901
Rapetti, D., Tauscher, K., Mirocha, J., & Burns, J. O. 2020, ApJ, 897, 174
Sathyarayana Rao, M., Subrahmanyan, R., Udya Shankar, N., & Chluba, J. 2017, ApJ, 840, 33
Switzer, E. R., & Liu, A. 2014, ApJ, 793, 102
Tauscher, K. 2020, pylinex: Linear and Nonlinear Signal Extraction in Python v1.0, Zenodo, doi:10.5281/zenodo.3661450
Tauscher, K., Rapetti, D., Burns, J. O., & Switzer, E. 2018, ApJ, 853, 187
Vedantham, H. K., Koopmans, L. V. E., de Bruyn, A. G., et al. 2014, MNRAS, 437, 1056