Global 21 cm Signal Extraction from Foreground and Instrumental Effects. II. Efficient and Self-consistent Technique for Constraining Nonlinear Signal Models

David Rapetti1,2,3, Keith Tauscher1,4, Jordan Mirocha5,6, 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 NASA Ames Research Center, Moffett Field, CA 94035, USA
3 Research Institute for Advanced Computer Science, Universities Space Research Association, Mountain View, CA 94043, USA
4 Department of Physics, University of Colorado, Boulder, CO 80309, USA
5 McGill University Department of Physics & McGill Space Institute, 3600 Rue University, Montréal, QC, H3A 2T8, Canada

Received 2019 December 2; revised 2020 June 1; accepted 2020 June 8; published 2020 July 16

Abstract

We present the completion of a data analysis pipeline that self-consistently separates global 21 cm signals from large systematics using a pattern recognition technique. This pipeline will be used for both ground and space-based hydrogen cosmology instruments. In the first paper of this series, we obtain optimal basis vectors from signal and foreground training sets to linearly fit both components with the minimal number of terms that best extracts the signal given its overlap with the foreground. In this second paper, we utilize the spectral constraints derived in the first paper to calculate the full posterior probability distribution of any signal parameter space of choice. The spectral fit provides the starting point for a Markov Chain Monte Carlo (MCMC) engine that samples the signal without traversing the foreground parameter space. At each MCMC step, we marginalize over the weights of all linear foreground modes and suppress those with unimportant variations by applying priors gleaned from the training set. This method drastically reduces the number of MCMC parameters, augmenting the efficiency of exploration, circumvents the need for selecting a minimal number of foreground modes, and allows the complexity of the foreground model to be greatly increased to simultaneously describe many observed spectra without requiring extra MCMC parameters. Using two nonlinear signal models, one based on the Experiment to Detect the Global Epoch-of-Reionization Signature (EDGES) observations and the other on phenomenological frequencies and temperatures of theoretically expected extrema, we demonstrate the success of this methodology by recovering the input parameters from multiple randomly simulated signals at low radio frequencies (10–200 MHz), while rigorously accounting for realistically modeled beam-weighted foregrounds.

Unified Astronomy Thesaurus concepts: Astrostatistics techniques (1886); Computational methods (1965); Reionization (1383)

1. Introduction

Measurements of the sky-averaged (global) 21 cm signal can be utilized to trace the thermal history of the early universe. This permits us to investigate both (1) astrophysical properties of the first populations of stars, galaxies, and black holes during Cosmic Dawn—when these first luminous objects formed—and the overall Epoch of Reionization (EoR) driven by energetic photons emitted by those objects, which ultimately ionized the primordial neutral hydrogen (HI), extinguishing its 21 cm spin-flip signal; and (2) the underlying cosmological model, including potential exotic phenomena—such as dark matter decay, annihilation, and interaction with baryons—affecting the cosmic mean temperature during those epochs and particularly the end of the preceding era, the Dark Ages, before astrophysical sources existed.

Recent results from the Experiment to Detect the Global EoR Signature (EDGES; Bowman et al. 2018a) show a 78 MHz absorption trough located within the frequency range expected for Cosmic Dawn. However, the amplitude of this trough is about 2–3 times larger than the maximum depth expected from adiabatic cooling due to the cosmic expansion in the concordance cosmological constant plus cold dark matter model. This has generated numerous attempts to explain such an anomaly via excess cooling from nonstandard physics, including dark matter particles scattering off baryons (Barkana 2018; Barkana et al. 2018; Berlin et al. 2018; Fialkov et al. 2018; Loeb & Muñoz 2018). Other possibilities include modifications of the cosmic radio background due to, for example, Population III objects or primordial black holes (Ewall-Wice et al. 2018, 2020; Feng & Holder 2018; Fialkov & Barkana 2019; Mebane et al. 2020).

Unaccounted systematics could alternatively resolve the current discrepancy between observations and theoretical modeling. Using the processed data set released by the EDGES collaboration, Bradley et al. (2019) showed that the EDGES result could be explained by resonances due to a ground plane artifact, instead of by a signal from the sky. Due to its relevance, this potential systematic is under further investigation by its proposers and the EDGES collaboration. In addition, other concerns about systematics can also be found in the recent literature (Hills et al. 2018; Draine & Miralda-Escudé 2018; Singh & Subrahmanyan 2019; Spinelli et al. 2019; Sims & Pober 2020).

Importantly, other contemporary global 21 cm experiments are working toward verifying these results: Shaped Antenna measurement of the background RAdio Spectrum (SARAS; Patra et al. 2013; Singh et al. 2017), Sonda Cosmológica de las Islas para la Detección de Hidrógeno Neutro (SCI-HI; Voytek

5 CITI National Fellow.

7 http://loco.lab.asu.edu/edges/edges-data-release/
et al. 2014), Zero-spacing Interferometer Measurements of the Background Radio Spectrum (ZEBRA; Mahesh et al. 2014), Large-aperture Experiment to detect the Dark Ages (LEDA; Bernardi et al. 2015, 2016; Price et al. 2018), Broadband Instrument for Global HydrOgen ReioNisation Signal (BIG-HORNS; Sokolowski et al. 2015), Probing Radio Intensity at high-Z from Marion (PRlzM; Philip et al. 2019), Radio Experiment for the Analysis of Cosmic Hydrogen (REACH; de Lera Acedo 2019), and the Cosmic Twilight Polarimeter (CTP; Nhan et al. 2017, 2019).

Given the impact of the EDGES results and the increasing efforts of the community to verify them, it is key to also probe the higher-redshift, purely cosmological Dark Ages absorption trough of the global 21 cm signal. For this, a space-based mission such as the Dark Ages Polarimeter PathfindER (DAPPER), which is able to collect data at low radio frequencies (~17–38 MHz) in the absence of Earth’s ionosphere, and to do so with minimal terrestrial radio frequency interference (RFI) in the shadow of the Moon (Burns et al. 2017, 2019; Bassett et al. 2019; Burns 2020), will be crucial.

Within this context, we are developing a new, flexible, end-to-end, data analysis pipeline that can be used to both optimize global 21 cm experiments and analyze their observations. This pipeline differs from previous techniques that utilize generic polynomials and fit a single averaged spectrum (e.g., Singh et al. 2017; Bowman et al. 2018a). In the first paper of this series (Tauscher et al. 2018, hereafter Paper I), we analytically calculated constraints on the spectral shapes of simulated 21 cm signals embedded in $10^4$–$10^6$ times larger foregrounds by applying a novel technique combining pattern recognition and information criteria (IC). Furthermore, this work included an innovative experimental design based on rotation-induced foreground polarization that we will continue to employ here in Paper II and discuss further in Paper III (Tauscher et al. 2020a).

In this second paper of the series, we present how we transform the spectral constraints derived in the first step of the pipeline into constraints on nonlinear signal parameters of interest via a Bayesian Markov Chain Monte Carlo (MCMC) analysis. In contrast with previous global 21 cm signal studies, we implement a simultaneous nonlinear fit of signal and foreground by marginalizing over the singular value decomposition (SVD) foreground parameters at each step instead of including them in the parameter space explored by the MCMC. This marginalization, which drastically improves MCMC efficiency, is performed analytically thanks to the foreground model being linear. Including only the signal parameters in the MCMC allows us to combine multiple data spectra, utilize correlations between them, and make fits more reliable.

With our pipeline, we show how an MCMC analysis can efficiently find input values of global 21 cm signal parameters in the presence of large beam-weighted foregrounds without a priori knowledge on the region of parameter space in which these values reside. The MCMC is initialized using a mean and covariance derived from a Fisher matrix–based procedure that converts the spectral signal constraints of Paper I into estimates of signal parameters for any given model. In Section 2, we sketch the three main components of the full data analysis pipeline: signal extraction, conversion between parameter spaces, and marginal Bayesian inference. Section 3 contains the motivations behind the signal models that we select as examples to test the pipeline under different frequency-dependent forms. Motivated by the EDGES results (Bowman et al. 2018a), we employ two signal models that allow for departures from standard 21 cm shapes. One is a flattened Gaussian (FG) as used by the EDGES team to fit their recent observations (Section 3.1), and the other is a parametric model motivated by theoretical thermal history milestones commonly referred as turning points (TPs) (Section 3.2). We describe the construction of realistically simulated data from our input signals and foreground modeling in Section 4 and present test cases for both signal models in Section 5. We finally discuss further progress to pursue and summarize in Sections 6 and 7.

The Python code, known as pylinex, underpinning our analyses in Papers I and II is publicly available. This software is particularly useful to fit measurements for which no analytical modeling for the signal and/or the systematics is known and a large covariance between them is present.

### 2. SVD/MCMC 21 cm Pipeline Analysis

Using simulated data, in Paper I we demonstrated that we can extract a wide variety of 21 cm signals from large foregrounds by fitting weighting coefficients of SVD eigen-modes derived from two separate training sets, one for the signal and the other for the foreground, and by selecting the number of modes for each set via the deviance information criterion (DIC). We review this procedure in Section 2.1.

The next steps of the pipeline are to (1) transform the constraints from the SVD linear fit into a physically motivated signal parameter space and (2) perform a nonlinear MCMC fit of this signal space by marginalizing over the SVD foreground parameter space. We describe these methods in Sections 2.2 and 2.3, respectively.

#### 2.1. First Step: Signal Extraction with pylinex

As a brief summary of Paper I, we remind the reader that we introduced pylinex as a generic scheme to efficiently and rapidly separate an arbitrary number of distinct sources of information (“data components,” in the terminology of Paper I), intrinsically mixed by the experiment (in a fashion described by “expansion matrices” in Paper I), plus random noise, whose properties can be described by either an analytic/numerical framework (such as theoretical modeling of the global signal), simulations (e.g., radio antenna beam patterns), or lab/sky measurements (e.g., receiver calibrations/foreground observations).

We tested this separation capability by building realistic data sets combining two of these components—21 cm signals generated with ares and Gaussian beam-weighted foregrounds—with...
statistical noise whose level is given by the radiometer equation. A natural extension of this initial exercise is to incorporate additional systematics into the foreground training set, such as, for instance, from a receiver (Paper IV of this series, in preparation).

Using pylinex, after defining linear models based on the SVD eigenbases for the different data components and combining them into a single linear model, we solve for the coefficients in that model $\xi$ and their covariance matrix $S$ using

$$ S = (G^T C^{-1} G)^{-1} \text{ and } \xi = SG^T C^{-1} y, \quad (1) $$

where $y$ is the data vector, $C$ is the covariance matrix of the noise distribution, and $G$ is a matrix containing basis vectors as columns. The basis vectors in $G$ have been “expanded” from the ones provided by SVD to account for the fact that the data components often exist in spaces smaller than the full data. For instance, since the signal appears identically in every fourth spectrum (i.e., all Stokes $I$ spectra), the realization of the model representing the signal’s effect on the data is $G_{21} x_{21} = \Psi_{21} F_{21} x_{21}$, where $x_{21}$ is the signal parameter vector, $F_{21}$ is the matrix of basis vectors obtained from the SVD eigenbasis of the signal training set, and $\Psi_{21}$ is the expansion matrix given by $\Psi_{21} = [I \ 0 \ 0 \ 0 \ I \ 0 \ 0 \ 0 \ \ldots]$, instead of just $F_{21} x_{21}$, which represents a single signal spectrum.\footnote{More information on the expansion matrix $\Psi$ is given in Paper I, and detailed cases are examined in Paper III.} From $\xi$ and $S$, we analytically calculate the maximum-likelihood estimate $\gamma_k$ of each data component $y_k$ (for the analysis of this paper, $k$ enumerates the signal and beam-weighted foreground), its channel covariance, $\Delta_k$, and its averaged 1$\sigma$ rms error, $\text{RMS}_k$, through

$$ \gamma_k = F_k \xi_k, \quad (2a) $$

$$ \Delta_k = F_k S_{kk} F_k^T, \quad (2b) $$

$$ \text{RMS}_k = \sqrt{\frac{\text{Tr}(\Delta_k)}{n_k}}, \quad (2c) $$

where $\xi_k$ is the portion of the parameter mean $\xi$ containing parameters modeling $y_k$, $S_{kk}$ is the diagonal block of the parameter covariance matrix $S$ corresponding to those parameters, and $n_k$ is the number of data channels in the $F_k$ basis.

2.2. Second Step: Transforming to Physical Parameters

The first step after obtaining SVD signal parameter distributions from pylinex is to approximately transform them into the chosen space of physically motivated signal parameters by searching for the best fit in the target parameter space. Naïvely, one might attempt to do this by fitting the pylinex-outputted signal band in frequency space and minimizing a likelihood such as

$$ \mathcal{L}_{\text{LSF-naïve}}(\theta_{21}) \propto \exp \left\{ -\frac{1}{2} \delta^T \Delta_{21}^{-1} \delta \right\}, \quad (3) $$

where $\delta = \gamma_{21} - \mathcal{M}_{21}(\theta_{21})$, $\gamma_{21}$ is given by Equation (2a), $\mathcal{M}_{21}(\theta_{21})$ is the physical signal model evaluated at the signal parameter vector $\theta_{21}$, and $\Delta_{21}$ is the diagonal matrix whose elements are the variances of the frequency channels under the pylinex fit, as defined in Equation (2b). Minimizing this likelihood corresponds to directly fitting the bands shown in Figure 7 of Paper I. However, for our purpose, this fit is insufficient to start our MCMC sampler for two reasons:

1. It does not use all information from pylinex due to the fact that the covariance matrix $\Delta_{21}$ only accounts for channel variances. Attempting to account for channel covariances makes $\Delta_{21}$ singular since there are more frequency channels than there are SVD signal modes.

2. A single parameter vector, such as the one found by the least square fit, cannot be used to initialize multiple MCMC chains.

2.2.1. Transforming Signal into SVD Parameters

The first issue above can be solved by performing the least square fit in SVD coefficient space instead of frequency space. This is performed by redefining the likelihood function being minimized as

$$ \mathcal{L}_{\text{LSF}}(\theta_{21}) \propto \exp \left\{ -\frac{1}{2} \delta^T S_{21}^{-1} \delta \right\}, \quad (4) $$

where $\delta = \gamma_{21} - \mathcal{M}_{21}(\theta_{21})$ now represents the displacement of the physical signal associated with $\theta_{21}$ transformed into SVD coefficient space from the mean $\xi_{21}$ of the SVD coefficient distribution with respect to its covariance $\Delta_{21}$. The matrix $\Phi_{21}$, which transforms a signal in frequency space $x_{21}$ to the SVD coefficient vector $x_{21}$, minimizes the weighted least squares residual $\langle x_{21} - x_{21} \rangle C^{-1} (x_{21} - x_{21} - T_{21})$, where $x_{21}$ is the matrix with signal basis vectors as its columns and $C$ is the full data noise covariance matrix, is given by

$$ \Phi_{21} = (F_{21}^T C_{21}^{-1} F_{21})^{-1} F_{21}^T C_{21}^{-1}. \quad (5) $$

Note that $\delta = \Phi_{21} \delta$. Unlike minimizing the likelihood defined in Equation (3), minimizing the likelihood in Equation (4) builds in all SVD covariances consistently and concisely. We denote the parameter vector that minimizes the likelihood of Equation (4) as $\theta_{21}$.

2.2.2. Fisher Matrix Formalism

To solve the second issue in Section 2.2, we estimate the signal parameter covariances by inverting the Fisher information matrix of $\mathcal{L}_{\text{LSF}}$ (Equation (4)), that is,

$$ \Lambda^{(0)}_{ij} \equiv \text{Cov}[(\theta_{21}), (\theta_{21})], \quad (6a) $$

$$ \approx [\langle D^T S_{21}^{-1} D \rangle^{-1}]_{ij}, \quad (6b) $$

where $\Lambda^{(0)}$ is the initial covariance matrix estimate and

$$ D \equiv \Phi_{21} \frac{\partial \mathcal{M}_{21}}{\partial \theta_{21}} \bigg|_{\theta_{21} = \hat{\theta}_{21}}. \quad (7) $$

In cases where the gradient $\frac{\partial \mathcal{M}_{21}}{\partial \theta_{21}}$ is not implemented or is difficult to compute analytically, it is estimated numerically from appropriately chosen finite steps.

Under the Fisher approximation, our initial estimate of the signal distribution in physical parameter space is a multivariate
Gaussian given by
\[ \theta_{21} \sim \mathcal{N}(\overline{\theta}_{21}, \Lambda^{(0)}). \] (8)

### 2.3. Third Step: MCMC Fit

The fit performed through the methods in Section 2.2 does not include the SVD foreground parameters, whose distribution must be considered when exploring the distribution of signal parameters. As long as the signal model is nonlinear, this must be calculated with numerical sampling. For this purpose, we have implemented a custom MCMC sampler in the \texttt{pylinex} code, based upon a Metropolis–Hastings (MH) algorithm (Gelman et al. 2013). While MH samplers are simple to implement, they have a disadvantage in that a significant amount of information must be supplied up front. Specifically, one must generate not only a probability density function (PDF) to sample, but also a proposal distribution, which determines the probability density of a chain moving from a starting to an ending point, and a distribution from which to draw initial points for individual, independent MCMC chains. We solve these drawbacks as follows in Sections 2.3.1, 2.3.2, 2.3.3, and 2.3.4.13

#### 2.3.1. Probability Density Function

A natural choice of the PDF to explore is the likelihood function multiplied by priors,
\[ p(\theta_{fg}, \theta_{21}) = \frac{\pi_X(\theta_{fg}) \pi_{21}(\theta_{21}) \mathcal{L}(\theta_{fg}, \theta_{21})}{Z}. \] (9)

In this equation, \( \pi_X(\theta_k) \) are the priors on the \( X \) parameters, \( Z \) is a \( \theta \)-independent constant, and the likelihood function is
\[ \mathcal{L}(\theta_{fg}, \theta_{21}) = |2\pi C|^{-1/2} \times \exp \left\{ -\frac{1}{2} [r(\theta_{fg}, \theta_{21})]^T C^{-1} [r(\theta_{fg}, \theta_{21})] \right\} , \] (10)

where \( r(\theta_{fg}, \theta_{21}) = y - M_{fg}(\theta_{fg}) - M_{21}(\theta_{21}) \) is the residual of the model of the data vector \( y \) evaluated at \( \theta_{fg} \) and \( \theta_{21} \), written with the foreground and 21 cm components separated.

The PDF given by Equation (9) is the joint density of all parameters, but we aim at only exploring numerically the signal parameters, \( \theta_{21} \). To do so, we modify the density to be the marginal signal parameter distribution by integrating over \( \theta_{fg} \), yielding
\[ p(\theta_{21}) = \frac{\pi_{21}(\theta_{21})}{Z} \int \mathcal{L}(\theta_{fg}, \theta_{21}) \pi_{fg}(\theta_{fg}) d\theta_{fg}. \] (11)

We define \( \mathcal{P}(\theta_{fg}, \theta_{21}) \equiv \mathcal{L}(\theta_{fg}, \theta_{21}) \pi_{fg}(\theta_{fg}) \) for convenience, which is, up to a multiplicative constant, equal to the conditional posterior PDF of the foreground parameters when the signal parameters are \( \theta_{21} \). Therefore, up to such a multiplicative constant, the integral is equal to the conditional Bayesian evidence of the foreground model when the signal parameters are fixed to \( \theta_{21} \). When the foreground priors are Gaussian, as we take them to be, \( \mathcal{P} \) is Gaussian in \( \theta_{fg} \) and the integral is equal to \(|2\pi \Sigma_{p}(\theta_{21})|^{1/2} \mathcal{P}_{\max}(\theta_{21})\), where \( \mathcal{P}_{\max}(\theta_{21}) \) is the maximum value of \( \mathcal{P}(\theta_{fg}, \theta_{21}) \) with \( \theta_{21} \) fixed and \( \Sigma_{p}(\theta_{21}) \) is the covariance of the Gaussian form of \( \mathcal{P}(\theta_{fg}, \theta_{21}) \) with \( \theta_{21} \) fixed.14

**Sampling this distribution instead of the PDF in Equation (9) is much faster because the dimension of the explored space is greatly reduced, and it uses the knowledge that the conditional distribution of the foreground is Gaussian by analytically marginalizing over the weights of the used foreground SVD eigenmodes at each MCMC step instead of numerically exploring them. This marginalization technique is exact, as opposed to the standard approximation performed by not including some of the numerically sampled parameters to produce constraints marginalized over them.**

#### 2.3.2. Initial Distribution of MCMC Iterates

The initial guess distribution for the signal parameters is given by the output of step 2 of the pipeline (Equation (8)), which is normal with mean \( \overline{\theta}_{21} \) and covariance \( \Lambda^{(0)} \).

#### 2.3.3. Proposal Distribution

One of the most critical inputs to an MH MCMC sampler is its proposal distribution, the distribution from which it draws new points at which to evaluate the probability density in Equation (12). If the variances (diagonal components of the proposal distribution covariance) are too narrow, nearly all steps will be accepted, but the sampler will not move efficiently through the parameter space. If they are too broad, nearly all steps will be rejected because the sampler will attempt to move too far in parameter space.

The off-diagonal components of the covariance are also important. For constant variances, excluding the off-diagonal covariances leads to a 1σ interval whose hypervolume is \( \frac{|\det(C_{\text{full}})|}{|\det(C_{\text{diag}})|} \) times larger than the same interval when the full covariance is used, which, in most cases, leads to a similar situation as mentioned above in the case where the variances are too broad.

The covariance matrix of the initial Gaussian proposal distribution for the parameters is equal to \( \Lambda^{(0)} / c(\alpha) \), where \( \Lambda^{(0)} \) is the covariance matrix of the initial distribution of MCMC iterates and \( c(\alpha) \), as defined in Appendix D, is a proportionality constant meant to achieve an acceptance fraction \( \alpha \). Therefore, the probability density of proposing a jump from \( \theta_{21} \) to \( \theta_{21}^{(i)} = \theta_{21} + \xi \) is
\[ p(\theta_{21}^{(i)} \rightarrow \theta_{21} + \xi) = \sqrt{c(\alpha)} \|2\pi \Lambda^{(0)}\|^{-1/2} \times \exp \left\{ -\frac{c(\alpha)}{2} \xi^T (\Lambda^{(0)})^{-1} \xi \right\}. \] (13)

#### 2.3.4. Updating and Acceptance Rate

In order to increase the efficiency of the MCMC search when using the basic MH algorithm, we schedule updates of the proposal based on the given distributions of all the MCMC
chains up that time. At the $k$th update, the covariance of recently visited points is computed and denoted $A^{(k)}$. Then the proposal matrix is updated to $A^{(k)}/\alpha$ where, once again, $\alpha$ is the desired acceptance fraction, leading the jumping probability to be equal to that shown in Equation (13) with $A^{(0)}$ replaced by $A^{(k)}$.

### 2.4. Foreground Priors

In our MCMC fit, we use a very large number of SVD foreground terms. This is sensible because we use the foreground training set to seed prior information. We fit each curve in the training set with the linear model, and a Gaussian approximation of the resulting eigenmode coefficients is computed.\(^\text{15}\) Then we use Gaussian distributions with the means and variances of the mode weights obtained from these fits as priors. While using the foreground modes themselves relies on the training set variations being similar in form to the data, using these priors amounts to the assumption that the magnitude of the data variations is similar to the magnitudes found in the training set.

### 2.5. Pipeline Overview

We summarize the pipeline elements and actions in the diagram of Figure 1. The blue arrows and boxes represent the linear fitting procedure described in Paper I (and reviewed in Section 2.1), after obtaining a combined, linear signal and foreground model using training sets (yellow boxes), SVD (light gray boxes), and an information criterion such as the DIC (dark gray box). On the other hand, the red boxes represent the MCMC fitting presented in this paper with analytical marginalization over SVD foreground eigenmodes at each step of the MCMC parameter search for a nonlinear signal model of choice.

Once realistic training sets are employed (including another for a receiver; Paper IV), this pipeline can be utilized to analyze observational data by inputting calibrated data from an instrument, instead of a simulated observation, in the green box on the right-hand side.

### 3. Nonlinear 21 cm Signal Models

For testing purposes, we will examine two physically motivated models. One based on EDGES observations (Section 3.1) and the other on key physical processes theoretically predicted to govern the time evolution of the global 21 cm signal (Section 3.2).

#### 3.1. Flattened Gaussian Model

First we will demonstrate our pipeline using an analytical model that was recently fitted to EDGES data by Bowman et al. (2018a). This is an FG model with four parameters: the amplitude $A$, center frequency $\nu_0$, FWHM $w$, and flattening $\tau$. In terms of these parameters, the 21 cm signal is modeled as

$$T_{21}(\nu) = A \frac{1 - e^{-\tau w}}{1 - e^{-\tau}}$$

where

$$B = \left(\frac{\nu - \nu_0}{w/2}\right)^2 \ln \left[ -\frac{1}{\tau} \ln \left( \frac{1 + e^{-\tau}}{2} \right) \right].$$

This is a phenomenological model with no physical motivation beyond representing an absorption trough. It was adopted by the EDGES collaboration for its ability to significantly reduce the rms of the residuals when fitting their data (Bowman et al. 2018a). For these fits, they used foreground models based on polynomial expansions around the dominant power-law behavior. Two of their models, however, were loosely inspired by ionospheric effects, but the parameters obtained were clearly unphysical as pointed out by Hills et al. (2018) (see also the EDGES reply in Bowman et al. 2018b).

Despite the shortcomings of the FG model, its simplicity makes it a useful initial example to exercise our pipeline. The left panel of Figure 2 shows how $\nu_0$, $w$, and $A$ shift and scale the model, as well as the effect of the flattening parameter $\tau$, which continuously modulates the shape of the signal between a Gaussian ($\tau \rightarrow 0$) and a square pulse ($\tau \rightarrow \infty$).
Turnings point model. Second, we parameterize the global 21 cm signal based on physically motivated extrema in its spectral shape, known as turning points (Pritchard & Loeb 2010; Harker et al. 2016). These are milestones in the cosmic history of the hydrogen gas.

After recombination decoupled the gas from photon temperature, the 21 cm spin temperature coupled to that of the gas. Since the gas cooled faster than the cosmic microwave background (CMB), the signal, which is the 21 cm brightness temperature relative to the CMB, goes into absorption. When the coupling of the 21 cm brightness to the gas temperature became ineffective compared with the coupling to the CMB because of the low gas density, the 21 cm temperature recoupled to that of the CMB, causing the signal to turn around and creating an absorption trough with a minimum typically labeled TP A. At TP B, the first stars turn on. Via the Wouthuysen–Field effect (Wouthuysen 1952; Field 1958), the Lyα radiation from the first stars recoupled the 21 cm transition to the temperature of the gas, which had continued cooling with respect to the CMB, triggering another absorption trough. From its minimum, TP C, the signal rises back due to the first stars and black holes significantly heating the gas. At TP D, the reionization of the gas begins extinguishing the signal down to its disappearance at TP E.

The free parameters of the model are the frequencies and brightness temperatures of the TPs, except for the temperature of E, which is fixed to zero. The model is a cubic spline between the TPs. In order to force these points to be extrema (i.e., have derivative zero), each TP uses two spline knots placed at the same temperature and 20 kHz apart symmetrically around the TP frequency given by the parameters. In addition to TPs A–E, there are two knots placed at 0 K and 10 ± 10 kHz. The model always evaluates to 0 K at frequencies above that of TP E. The right panel of Figure 2 shows schematically the relative locations and allowed ranges (red rectangles) for the TP modeling that we use here.

### 3.2. Turning Point Model

Second, we parameterize the global 21 cm signal based on physically motivated extrema in its spectral shape, known as turning points (Pritchard & Loeb 2010; Harker et al. 2016). These are milestones in the cosmic history of the hydrogen gas.

After recombination decoupled the gas from photon temperature, the 21 cm spin temperature coupled to that of the gas. Since the gas cooled faster than the cosmic microwave background (CMB), the signal, which is the 21 cm brightness temperature relative to the CMB, goes into absorption. When the coupling of the 21 cm brightness to the gas temperature became ineffective compared with the coupling to the CMB because of the low gas density, the 21 cm temperature recoupled to that of the CMB, causing the signal to turn around and creating an absorption trough with a minimum typically labeled TP A. At TP B, the first stars turn on. Via the Wouthuysen–Field effect (Wouthuysen 1952; Field 1958), the Lyα radiation from the first stars recoupled the 21 cm transition to the temperature of the gas, which had continued cooling with respect to the CMB, triggering another absorption trough. From its minimum, TP C, the signal rises back due to the first stars and black holes significantly heating the gas. At TP D, the reionization of the gas begins extinguishing the signal down to its disappearance at TP E.

The free parameters of the model are the frequencies and brightness temperatures of the TPs, except for the temperature of E, which is fixed to zero. The model is a cubic spline between the TPs. In order to force these points to be extrema

### Table 1

| Symbol | Parameter | Units | Distribution |
|--------|-----------|-------|--------------|
| A      | Amplitude | K     | Unif(−1, −0.1) |
| ν₀     | Center    | MHz   | Unif(60, 90)  |
| w      | FWHM     | MHz   | Unif(1, 30)   |
| τ      | Flattening| N/A   | Exp(1)       |

### Table 2

| Symbol | Parameter | Units | Distribution |
|--------|-----------|-------|--------------|
| νₐ     | A frequency | MHz | Unif(10, 26) |
| Tₐ     | A temperature | mK | Unif(−100, −10) |
| ν₉     | B frequency | MHz | Unif(30, 80) |
| T₉     | B temperature | mK | Unif(−5, 0) |
| ν₆     | C frequency | MHz | Unif(60, 120) |
| T₆     | C temperature | mK | Unif(−350, −100) |
| ν₉     | D frequency | MHz | Unif(100, 150) |
| T₉     | D temperature | mK | Unif(0, 25) |
| ν₉     | E frequency | MHz | Unif(100, 200) |

Note. The frequencies of adjacent TPs are also constrained to differ by at least 10 MHz.

(i.e., have derivative zero), each TP uses two spline knots placed at the same temperature and 20 kHz apart symmetrically around the TP frequency given by the parameters. In addition to TPs A–E, there are two knots placed at 0 K and 10 ± 10 kHz. The model always evaluates to 0 K at frequencies above that of TP E. The right panel of Figure 2 shows schematically the relative locations and allowed ranges (red rectangles) for the TP modeling that we use here.

### 4. Simulated Data

#### 4.1. Signal Training Sets

Signal training sets are formed with the models described above and are seeded with distributions of the underlying parameters. The distributions of the FG parameters are shown in Table 1. The training set consists of 10⁶ curves made from parameters drawn from these distributions. The distributions of the TP parameters are shown in Table 2. The TP training set consists of 10⁵ curves drawn from these distributions. Figure 3 shows samples of the training sets for both models in frequency space.
4.2. Foreground Modeling

The simulated foregrounds are constructed in the same manner as in Paper I, as briefly described here. For simplicity, we include only beam-weighted foreground emission, ignoring other systematics such as human activity–generated RFI, refraction, absorption and emission due to Earth’s ionosphere, and receiver gain and noise temperature variations (the latter will be discussed in Paper IV). The experiment simulated here is most analogous to a pair of antennas orbiting the Moon and taking data above the far side, where the ionospheric effects and RFI need not be addressed (Burns et al. 2017, 2019; Burns 2020).

The simulated data products, \( y \), of all fits in this paper are concatenations of 96 brightness temperature spectra, which include four Stokes parameters, \( I, Q, U, \) and \( V \), at \( N_\nu = 6 \) different rotation angles, \( \chi \), about the antenna boresight for \( N_n = 4 \) different antenna pointing directions, \( n \). The antenna simulated is a dual-dipole system modulated by angular Gaussian profiles with spectrally varying FWHM. The sky brightness temperature in the simulations is given by the observed 408 MHz Haslam map (Haslam et al. 1982) with each pixel scaled by the power law \([\nu/(408 \text{ MHz})]^{-2.5} \). Paper III of this series will explore how measuring Stokes parameters at multiple rotation angles and antenna pointings makes fits more rigorous and decreases errors. In addition to beam variations, upcoming work will include multiple sky models derived from observations, improving the accuracy of the uncertainties toward realistic analyses.

The variance of the noise added to the data in each fit, \( \sigma^2 \), is constant across the different Stokes parameters and is related to the total power (Stokes \( I \)) brightness temperature, \( T_{bI} \), through the radiometer equation,

\[
\sigma^2(\nu, \chi, n) = \frac{N_\nu N_n}{\Delta \nu \Delta t} [T_b(\nu, \chi, n)]^2, \tag{15}
\]

with a frequency channel width \( \Delta \nu \) of 1 MHz and a total integration time \( \Delta t \) of 800 hr. The data are split into five different components—one for the 21 cm signal and one for the beam-weighted foregrounds (which are correlated across boresight angles and frequency) of each pointing, \( n \). The signal is the same across all \( N_n = 4 \) pointings, while the foregrounds for each pointing only affect the data from that pointing.

5. Results

In Section 5.1, we examine the ability of the pipeline to utilize the SVD spectral constraints on the signal, retrieved without a priori knowledge, to inform our MCMC on the starting location and covariance proposal. Given a large parameter space such as ours, this initial information is critical for an efficient MCMC search. We then test two key elements of this analysis in Section 5.2. These are the number of linear SVD foreground terms that will be marginalized over the signal parameter space explored by the MCMC algorithm, and the robustness of this selection given our use of foreground priors (see Section 2.4 and Appendix C). We end the section with an exploration of the constraining power on the Cosmic Dawn and Dark Ages troughs and how these constraints are affected by statistical noise and systematics confusion (Section 5.3).

5.1. Signal Extraction and MCMC Fits

We present results for both models of Section 3 using various random cases for each. In Section 5.1.2, we discuss the signal reconstructions of these two sets of cases, achieved first by “linear signal extraction” (initial step of the pipeline as presented in Paper I and summarized in Section 2.1) and second by “nonlinear marginal Bayesian inference” (third step as presented in Section 2.3). As shown in Figures 5 and 6, the excellent matches between the results obtained by each of the two methods serve as proof of concepts for these two novel techniques: (1) pattern recognition based on SVD+IC for the signal extraction and (2) fitting that marginalizes over linear foreground parameters during an MCMC exploration of a nonlinear signal space. The latter ensures a simultaneous and self-consistent fit of the systematics parameters (in these examples, foreground).

In Section 5.1.4, for each model used we present the constraints inferred on their signal parameters and how we recover their input values in a statistically consistent manner for all nine random cases fitted, validating the robustness of the pipeline.

5.1.1. Linear Systematic Errors

As described in Paper I, for the linear signal fits it is critical to calibrate the confidence levels of individual cases, as presented below in Section 5.1.2, using a set of simulated curves (5000 in our cases) from the corresponding training set.
This calibration is necessary because the SVD linear models of the foreground and signal do not exactly fit their corresponding training set curves—even though the combined linear model might fit the added curves better, due to overlap between the two components. The unmodeled parts of both components caused by those biases create systematic uncertainty on top of the statistical noise. Figure 4 presents the confidence level calibration for both signal models, FG (blue curve) and TP (orange curve). In addition to the number of parameters selected by the DIC (see Section 5.1.3), the overlap between the SVD signal and foreground modes, which are independently obtained from each training set, is also key in determining the size of the errors. Figure 4 shows that the foreground model has a notably larger overlap with the FG model than with the TP model. This accentuates the difference between the linear and MCMC fits for each model, as seen when comparing Figures 5 and 6.

The ideal scenario is having training sets with SVD bases that are orthogonal. Keeping this in mind when designing an experiment and forming training sets is important. Our use of induced polarization pursues this goal by adding data components to the foreground training set that minimize the overlap with the signal.

### 5.1.2. Signal Reconstructions

Figure 5 shows the success of the pipeline for the FG model, while Figure 6 shows the same for the TP model. In each figure, the blue (red) regions indicate linear (MCMC) constraints on the signals in frequency-brightness temperature space.

For the FG model (Figure 5), the MCMC constraints are clearly tighter than those obtained from the linear fits. However, this difference varies considerably among the cases displayed, being most extreme in case number 3, FG3, where the signal is relatively well localized in frequency, as it is also

**Figure 4.** Histogram showing cumulative distributions of the signal bias statistic $\varepsilon$ defined in Paper I (and here recalled in Equation (A1)) for the FG (blue curve) and TP (orange) models. These curves represent the fraction of simulations with $\varepsilon$ less than the value given on the x-axis. The $\varepsilon$ values where the histograms cross 95% indicate the $\sigma$ level of the 95% confidence intervals. The crossing of this value for the two models is very different: for the TP model the sigma level ($\sim 2.5\sigma$) is relatively close to that of the reference $\chi^2(1)$ distribution ($2\sigma$), whereas the FG crosses at the $\sim 8.75\sigma$ level. As in Paper I, these distributions were computed by performing 5000 simulations of linear fits with each of the training sets.

**Figure 5.** Pipeline constraints in frequency space for the FG model. The blue intervals correspond to the 95% intervals from the linear fit, which uses SVD modes to represent the signal in addition to the foreground, while the red intervals correspond to the 95% confidence intervals from the MCMC fit, which uses the full nonlinear signal model and SVD foreground modes. For the linear fit, the 95% confidence intervals correspond to $8.75\sigma$ (see Figure 4). The instances of the FG model are denoted as FG1, FG2, FG3, FG4, and FG5 in the text and in Tables 3 and 4.

in cases FG1 and FG5, but closer to the edge of the frequency band than these two cases. On the other hand, signals FG2 and FG4 are wider, particularly FG4, for which the difference in constraining power between the linear and MCMC fits is the smallest, and FG2 is also flatter (i.e., has a larger value of $\tau$) than the rest.

For the TP model (Figure 6), the four random cases tested also present overall tighter reconstructions for the MCMC fits at the end of the pipeline, but the differences between the MCMC and linear fits are smaller than for most of the FG cases (FG1–3, FG5). All of the TP model cases and FG4 span wide fractions of the frequency band, leading to larger MCMC uncertainties than the other signals.

In comparison with FG, the absorption features in the TP cases cover more similar frequency ranges between each other because they are built based on the thermal history milestones described in Section 3.2. This physically motivated similarity between spectral shapes can be seen in the training set sample of the TP model shown in the right panel of Figure 3, particularly compared with that of the left panel of this figure for the FG model.

#### 5.1.3. Numbers of SVD Eigenmodes

An estimate of the number of SVD eigenmodes needed for each training set (foreground and signal) is shown in Figure B1 of Appendix B. For the FG signal model, fitting all curves of the signal training set below the noise level requires a large number of SVD terms (see the corresponding calculation in Appendix B), of the order of 40 (top, left panel), while for the TP signal model all curves in the training set can be fitted with significantly fewer SVD terms, about 20 (top, right panel).

Given that for each signal model fit, the foreground training set used is the same within the frequency range in common (50–100 MHz), the significant differences in shape among the curves shown in the left panel of Figure 3 (e.g., central frequency, width, and flattening factor of the troughs) require a relatively large number of terms to describe the entire FG
model training set. This will generally imply larger frequency band errors when reconstructing individual signals, in particular for curves with little overlap with the rest of the training set, such as FG3 at the edge of the frequency range in Figure 5.

Comparatively, the TP model presents less variation and more overlap between curves (right panel of Figure 3) and thus requires fewer SVD modes to describe the corresponding training set. The bottom panels of Figure B1 show the numbers of terms (7 and 12) required to fit the foreground training set curves corresponding to each signal model. In this case, the larger frequency range covered in the TP model (10–200 MHz) is bound to increase the number of SVD terms needed to fit all curves below the noise level.

Note, however, that the number of parameters shown in Figure B1 is calculated for the overall training sets. For the individual cases displayed in Figures 5 and 6 for each model, Table 3 shows the number of terms chosen by the DIC and the corresponding rms uncertainties when simultaneously fitting signal and foreground. The individual cases follow the same pattern as is seen in Figure B1, where TP model instances require more terms for foreground and signal than those for the FG model.

5.1.4. Recovering Input Parameters

The MCMC recovered one-dimensional (1D) posterior probability distributions for the FG and TP model parameters are shown in gray in the diagonal plots of Figures 7 and 8, respectively, where the red dashed lines mark the input parameters. The contours in the two-dimensional (2D) off-diagonal plots in these figures show the 68% (green) and 95%
(blue) confidence levels and present the covariances between these parameters as found by the MCMC calculation. In red contours, for comparison, we show 95% confidence level contours for Fisher matrix–derived covariances (see Section 2.2.2) from the statistical noise of the radiometer equation.

We find that our pipeline successfully obtains constraints consistent with the input values for both models within the noise level (red contours), and it does so efficiently by rapidly reaching both the targeted acceptance rate (25%) for the MCMC sampler (see Section 2.3.4 and Appendix D) and ultimately a high level of convergence for all parameters, as determined by the commonly employed Gelman–Rubin test (Gelman & Rubin 1992).

Figure 7 shows constraints for two cases, FG4 (left panel) and FG2 (right) from Figure 5, which are representative of two types of results. For case FG2 (right), the 95% confidence level MCMC constraints (blue contours) are much tighter than those obtained with the linear fit (see Figure 5), reaching down almost to the noise level (red contours). On the other hand, the MCMC constraints (blue) for case FG4 (left) are noticeably larger than those for FG2. Note also that the nonlinear constraints of FG4 (blue) are significantly larger than the corresponding Fisher matrix contours (red) derived only from the 800 hr of integration, that is, the noise level. This indicates that for FG4 the overlap between signal and foreground is larger than for FG2.

The difference between the actual constraints in blue contours, and those in red for the noise level reference, corresponds to the effect of simultaneously fitting the signal together with the foreground. This is necessary to fit the data consistently, properly accounting for systematic uncertainties on the signal caused by the combination of random noise and large foreground systematics. Thus, by comparing cases FG4 and FG2, it is clear that FG4 suffers from larger systematic errors due to its larger overlap with the foreground training set. Figure E1 also shows this difference in the form of the correlations between all the parameters, obtained as described in Appendix E, for the same two cases, FG4 and FG2, as explicit examples of two largely different levels of overlap between the signal and foreground.

Figure 8 shows the constraints on the TP model parameters for case TP1. Note that there are large differences between systematic (blue contours at the 95% confidence level) and statistical (red) errors among some of the parameters and that some parameters (e.g., $\nu_A$, $\nu_E$, and $T_B$) hit edges of their prior space.

As a summary of all MCMC fits performed for the FG and TP cases presented in Figures 5 and 6, Tables 4 and 5 show 99% confidence intervals on their respective parameters.

### 5.2. Number of Marginalized Foreground Parameters

As an important test to be passed by our new methodology, in Figure 9 we show the change in constraints when we vary the number of SVD foreground terms marginalized over in the MCMC analysis of TP model case 3 (TP3). As described in Section 5.1.2, Table 3 indicates that this specific realization of TP3 requires 24 foreground terms. Based on this reference value, we calculate constraints on the TP model parameters when using three different numbers of foreground terms: 10, 25, and 40. Given the reference value, we expect to have highly biased means and spurious uncertainties for the 10 terms case, in contrast with those of the 25 and 40 terms cases. This is actually shown in Figure 9 for the 95% confidence level constraints on A, B, C, and D in frequency-temperature space. In addition, the

![Figure 7. 1D and 2D MCMC posterior distributions for the FG parameters, with red dashed lines marking the input parameters. The left (right) triangle plot shows constraints for the signal case FG4 (FG2) in Figure 5 and Table 4. In the 2D plots, the green and blue contours show 68% and 95% confidence intervals, respectively, and the red contours represent the 95% confidence regions obtained from a Fisher matrix covariance of these parameters for the statistical, radiometer noise. These Fisher matrix estimates assume that systematics are subtracted out perfectly, as if the signal was being observed in isolation with only noise obscuring it. This is necessary to properly accounting for systematic uncertainties on the signal caused by the combination of random noise and large foreground systematics. Thus, by comparing cases FG4 and FG2, it is clear that FG4 suffers from larger systematic errors due to its larger overlap with the foreground training set. Figure E1 also shows this difference in the form of the correlations between all the parameters, obtained as described in Appendix E, for the same two cases, FG4 and FG2, as explicit examples of two largely different levels of overlap between the signal and foreground. As a summary of all MCMC fits performed for the FG and TP cases presented in Figures 5 and 6, Tables 4 and 5 show 99% confidence intervals on their respective parameters.](image)
uncertainties of the cases with 25 and 40 terms are remarkably similar between each other despite the large difference in number of parameters. This indicates the success of our technique of employing training set priors (see Section 2.4 and Appendix C) to avoid the contributions of SVD foreground modes with SVD importances below the noise level. Thus, our technique only requires selecting a number of foreground terms large enough above that found by the linear fit to provide unbiased, accurate, and robust parameter measurements.

5.3. Constraints versus Integration Time

In this section, we utilize our new pipeline to run fits for TP model case TP1 when evenly increasing the integration time by factors of 5 from 800 (see the top, left panel of Figure 6) to 4000, 20,000, and 100,000 hr. These times provide the spectral noise profiles shown in the left panel of Figure 10, while the right panel shows the corresponding increases in constraining power on the spectral shape of the signal. These results show that up to the highest integration time used, the constraints are not limited by systematics, which in this case is the overlap of the signal with the foreground.

The 2D constraints at the 95% confidence level on each pair of frequency and temperature parameters for TPs A, B, C, and D are shown in Figure 11. Consistent with the right panel of Figure 10, these constraints are increasingly tighter with longer integration times, and, as discussed above, no systematic floor is found up to $10^5$ hr. Note that for each of the four runs the same noise seed was used to have an identical noise shape with

Figure 8. Same as Figure 7 but for the TP case shown in the upper left panel of Figure 6 (TP1). Green and blue contours in the 2D plots show 68% and 95% confidence intervals, while the red ellipses show Fisher matrix–estimated 95% confidence intervals, which assume only statistical noise. All intervals are for 800 hr of integration. Some parameters, such as the temperature of TP B, which is only allowed to vary from −5 to 0 mK, are not constrained within the prior volume, while others, such as the temperature of TP C, are constrained.
Table 4
MCMC-derived 99% Confidence Intervals on the FG Parameters

| Par. | Input | Recovered |
|------|-------|-----------|
| FG1  |       |           |
| FG2  |       |           |
| FG3  |       |           |
| FG4  |       |           |
| FG5  |       |           |

Note. All fits done with 800 hr of integration. Spectral constraints of these FG cases are shown in Figure 5. FG4 (FG2) corresponds to the triangle plot in the left (right) panel of Figure 7.

Table 5
MCMC-derived 99% Confidence Intervals on the TP Parameters

| Parameter | TP1 | TP2 | TP3 | TP4 |
|-----------|-----|-----|-----|-----|
| νₐ (MHz)  | 24  | 21  | 12  | 18  | 24  |
| Tₛ (mK)  | -56 | -74 | -52 | -34 | -15 |
| νᵦ (MHz) | 76.51 | 76.63 | 50.07 | 42.49 | 43.91 |
| Tₖ (mK)  | -1.5 | -2.5 | -0.5 | -0.4 | -4.0 |
| νₐ (MHz) | 107.819 | 107.819 | 106.638 | 61.83 | 74.99 |
| Tₕ (mK)  | -111.9 | -112.1 | -231.8 | -159.2 | -184.1 |
| νₐ (MHz) | 127.625 | 127.625 | 131.934 | 128.56 | 133.02 |
| Tₕ (MHz) | 18.4 | 18.0 | 6.3 | 14.2 | 12.2 |
| νₐ (MHz) | 193.63 | 195.69 | 197 | 153.980 | 186 |

Note. All fits performed with 800 hr of integration. These parameter constraints are subject to the priors given in Table 2, except for that on νₐ, which was allowed to uniformly vary from (1, 30) for extra variability in the MCMC search. Spectral constraints on these signal cases are shown in Figure 6. TP1 corresponds to the triangle plot in Figure 8.

the magnitude scaled as $\frac{1}{\sqrt{T}}$. This ensures that the constraints for these four runs are exactly comparable in terms of mean and covariance shape and only the size changes as a function of integration time.

This exercise exemplifies a straightforward application of our pipeline to simulate experimental setups. Given a training set for each of the data components, the pipeline can establish whether a certain amount of integration time reaches or not the systematic floor for the modeling used. In our idealized example, we learn that we could significantly tighten constraints on the Dark Ages trough, with a foreground level considerably higher than for Cosmic Dawn, by, for instance, adding single dual antennas (assuming that they are similar enough to allow comparable calibrations) to efficiently increase the integration time. In upcoming studies, including additional data components such as an instrument (Paper IV) will continue this line of research and provide further applications for the utilization of our pipeline.

6. Discussion

It is important to emphasize that, even though forecasting applications such as the one described in the previous section are valuable products of this work, our pipeline is primarily aimed at analyzing observations from current and upcoming global 21 cm experiments. In particular, it has been designed to analyze full polarization data from CTP (Nhan et al. 2019) and DAPPER (Burns et al. 2019; Burns 2020). Both of these experiments are being developed to benefit from the use of all four Stokes parameters plus multiple spectra from different rotation angles or drift-scan time bins.

As shown in Paper III, the combination of these two techniques, polarization and drift-scan, significantly helps in separating the signal from the foreground. The reasons are, respectively, because only the foreground is polarized and changes in the sky as a function of sidereal time. In Paper III, we find that in order to obtain constraints at the level required to detect the predicted global 21 cm signal it is necessary to utilize multiple spectra. In two additional papers, we also discuss, in one, the implication of these findings for the single averaged spectrum of the EDGES analysis (Tauscher et al. 2020b) and, in the other, the importance of employing our methodology for a comprehensive beam-weighted foreground study (Hibbard et al. 2020, submitted).

As already demonstrated in Paper I, and now here in Paper II (see the previous section), our pipeline straightforwardly accounts for many correlated spectra, which, for example, in the simulations discussed above are typically captured by of the order of 80 foreground parameters. Since our analysis allows us to analytically marginalize over a large number of SVD linear parameters, they negligibly contribute to the overall computational time budget. This also implies that we can efficiently add essentially arbitrarily large numbers of systematic parameters to model complex beam-weighted foregrounds or calibrate the receiver (see Paper IV), using well-characterized training sets from which to obtain as many SVD linear parameters as needed. Note, however, that if for some of these systematics a reliable nonlinear model is available, the corresponding
parameters can instead be added as MCMC parameters together with those of the nonlinear signal model.

As an integral part of our pipeline, we are also developing rigorous statistical tests to determine when the training sets used in an observational analysis are not accurate enough to model the systematics and simultaneously fit the signal (Bassett et al., in preparation). In this case, a standard goodness-of-fit test of the overall model would not be sufficient due to the potential interchange between the beam-weighted foreground and the signal, which could bias the signal while still providing an acceptable fit to the combined model. The new tests would then indicate the need to improve the training sets used in our modeling before a subsequent reanalysis, in order to avoid biased measurements.

7. Conclusions

This is the second paper in a series presenting a complete analysis pipeline for global 21 cm observations that accounts for each of the data components simultaneously and self-consistently. In this work, we have advanced the analysis by incorporating the ability to not only separate the signal from foregrounds, as described in the first paper, but also start a marginal MCMC exploration of signal model parameters based on the mean and covariance of the spectral signal constraint derived analytically in the first step.

In each step of the MCMC search over the signal parameter space, we utilize the linear, optimal description of the foreground obtained in the initial SVD calculation to marginalize over its SVD eigenmodes analytically, instead of exploring them numerically with the MCMC. As is to be expected, this greatly improves the efficiency of the MCMC. We also implement the use of priors derived from the foreground training set to ensure that the variations that are deemed unimportant by the SVD analysis do not unduly affect our results. This allows us to select a number of SVD foreground eigenmodes that is large enough to avoid biases without unnecessarily increasing the uncertainties.
We demonstrate that this technique successfully recovers input parameters for two analytical models. The first is inspired by the recent results from the EDGES collaboration, where an FG shape was employed to fit the observations, together with various foreground models (Bowman et al. 2018a). The other signal model builds upon a well-known simplified theoretical description of the global 21 cm spectral form based on predicted extrema (TPs) caused by cosmological and astrophysical phenomena (Pritchard & Loeb 2010) during the end of the Dark Ages, Cosmic Dawn, and the EoR. For the purpose of testing the pipeline, both of these models are allowed to vary beyond the adiabatic cooling limit of the standard model, as suggested by the EDGES results.

Using a particular case for the TP model, we test that when varying the number of foreground parameters with respect to the reference value selected by the DIC in the linear fit our technique does behave as intended. That is, when not using a sufficient number of terms we predictably obtain large biases and spuriously tight constraints, and when correctly using a number close to or larger than that from the DIC we are able to

![Figure 10.](image)

*Figure 10. Left: 1σ noise levels for the integration times used in the application example of Section 5.3. The red rectangles encompass the allowed ranges of frequencies and absolute values of temperatures for TPs A and C, representing the Dark Ages and Cosmic Dawn troughs, respectively. The black stars indicate the input values of the frequencies and temperatures of these TPs for signal TP1. The uncertainties given by the curves correspond only to the noise levels and do not account for uncertainties deriving from the overlap between foreground and signal. Right: full (statistical plus systematic) uncertainties of TP1 in frequency space for four evenly increasing integration times up from our reference value of 800 hr (see the top, left panel of Figure 6), with the same random seed for noise generation to ensure comparability.*

![Figure 11.](image)

*Figure 11. 95% confidence level constraints on the frequencies and temperatures of TPs A–D vs. integration time. The same random seed is used for each case, meaning that the noise has the same shape but a different magnitude, given by the integration time. The input signal for these fits is TP1, as in Figure 8. The colors for the integration times match those of Figure 10.*
reproduce the input values with uncertainties that properly include the statistical noise plus a systematic error accounting for the overlap between the signal and foreground modeling. The latter varies for individual cases and training sets and is crucially captured by our self-consistent analysis.

It is also worth noting that if choosing a number of terms much larger than that from the DIC, the errors do not undesirably increase thanks to our use of priors derived from the foreground training set. These incorporate the knowledge on the importance of each mode of variation, providing automatic downweighting of irrelevant modes.

As an initial application, we then employ our newly verified pipeline to examine how increasing the integration time affects the constraints on a TP model case. Our framework allows us to straightforwardly model the foreground and beam realistically by building informed training sets, as to be presented in currently ongoing work, as well as, in the same manner, to incorporate a full receiver model (see upcoming Paper IV of this series for such an analysis). However, for the idealized foreground and beam used here as test examples, we find no systematic floor on constraining the Dark Ages and Cosmic Dawn troughs when increasing up to a factor of 125 our reference modest time of 800 hr. This simple exercise serves as an instance of the opportunities opening to the hydrogen cosmology community in utilizing our publicly available, statistically robust pipeline to rigorously analyze both simulations, in preparation of experiments, and actual observations.

Critically, our analysis technique fully accounts for covariances and systematic uncertainties as encoded in detailed, readily changeable training sets for each of the components forming a given set of sky-averaged 21 cm measurements. This also implies that no analytical modeling is required. In its absence, additional measurements and/or simulations can be employed to construct training sets.

Due to the fact that foreground parameters are analytically marginalized instead of being numerically sampled by the MCMC engine, a large number of foreground parameters can be added without loss of efficiency, allowing for many unaveraged spectra to be processed by the pipeline with negligible added computational complexity. Furthermore, it is useful to note that our methodology can be directly adapted to any type of data and is especially beneficial when covariances between signal and systematics are relatively large.

Planned observations with EDGES, CTP, SARAS, LEDA, PRizM, REACH, and other experiments from the ground, as well as with DAPPER from lunar orbit, should greatly benefit from the framework described here, which pioneers combining linear pattern recognition with nonlinear Bayesian statistics.

We thank Eric Switzer for useful discussions at the beginning of this project, and for the seed idea of marginalizing over foreground parameters. D.R. was supported for part of the time by a NASA Postdoctoral Program Senior Fellowship at NASA’s Ames Research Center, administered by the Universities Space Research Association under contract with the National Aeronautics and Space Administration (NASA). This work was also supported by NASA under award number NNA16BD14C for NASA Academic Mission Services. The theoretical work was partially funded by NASA ATP grant NNX15AK80G. J.M. acknowledges support from a CITA National Fellowship. This work is directly supported by the NASA Solar System Exploration Virtual Institute cooperative agreement 80ARCO17M0006. Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center.

Appendix A
Calibrating Confidence Intervals for Linear Fits

As described in Paper I, the confidence intervals from our linear fits do not match the common 68%-95%-99.7% rule for 1, 2, and 3 sigma implied by the $\chi$ distribution. Instead, these confidence intervals depend on the training sets and the selected numbers of terms. In a nutshell, we perform fits with many foreground, signal, and noise realizations. For each fit, we calculate the signal bias statistic defined as

$$v = \sqrt{\frac{1}{N_{\text{channels}}} \sum_{k=1}^{N_{\text{training}}} \frac{(y_{21} - \gamma_{21})C^T_k}{(\Delta_{21})_{kk}}}$$

where $y_{21}$ is the input signal and $\gamma_{21}$ and $\Delta_{21}$ are the mean and 1$\sigma$ channel covariance calculated in Equations (2a) and (2b).

Figure 4 shows the cumulative distribution functions of 5000 fits for the FG (blue curve) and TP (orange curve) models and indicates that the 95% confidence intervals for the FG (TP) fits correspond to 8.75$\sigma$ (2.5$\sigma$). This difference indicates that it is much easier to linearly separate the beam-weighted foreground model from the TP model than from the FG model.

Appendix B
Training Set Eigenvalue Spectra

The eigenmodes of each training set, $B$, are given by its singular value decomposition, $B = U \Sigma V^T$, where $U^T \Sigma^{-1} U = I$ ($C$ is the noise covariance matrix), $V^T V = I$, and $\Sigma$ has the same shape as $B (N_{\text{channels}} \times N_{\text{curves}})$, but all off-diagonal elements are zero and the diagonal elements are nonnegative and decreasing. The matrix $F_{N_{\text{modes}}}$, whose columns form the basis vectors of our model of the component described by the training set when we choose $N_{\text{modes}}$ modes, consists of the first $N_{\text{modes}}$ columns of $U$. The rms error in number of noise levels when $F_{N_{\text{modes}}}$ is used to fit the curves of the training set is

$$\text{RMS}_{N_{\text{modes}}} = \sqrt{\frac{1}{N_{\text{channels}} N_{\text{curves}}} \text{Tr} \left[ (I - F_{N_{\text{modes}}} F_{N_{\text{modes}}} C^{-1}) B^T C^{-1} (I - F_{N_{\text{modes}}} F_{N_{\text{modes}}} C^{-1}) B \right]}$$

from the framework described here, which pioneers combining linear pattern recognition with nonlinear Bayesian statistics.

$$= \frac{1}{N_{\text{channels}}} \frac{\text{min}(N_{\text{channels}}, N_{\text{curves}})}{\sum_{n=N_{\text{modes}}+1}^\infty \sigma_n^2},$$

where $\sigma_n = \Sigma_{nn}$. Figure B1 shows the importance spectrum of the FG and TP models training sets using the RMS$_{N_{\text{modes}}}$ metric, assuming 800 hr of integration.
Appendix C
Prior distributions inferred from SVD Matrices

This appendix concerns prior distributions inferred from training sets. As such, all fits discussed here refer to only one data component, taken here to be the foreground. If we choose orthonormal basis vectors for our foreground model (using $f_k \cdot f_k = \delta_{kk}$ as the orthonormality condition, where $f_k$ is the $k$th basis vector and $C$ is the noise covariance matrix), then the distribution of weights on the $n$th mode $f_n$ when fitting a training set $B$ is given by (see, e.g., Equation (5))

$$x_n = f_n^T C^{-1} B.$$  \hfill (C3)

Here we assume, as above (see Appendix B), that $B = U \Sigma V^T$, where $U^T U = I$, $V^T V = I$, and $\Sigma$ is the same shape as $B$ but all off-diagonal elements are zero and the diagonal elements are nonnegative and decreasing. Therefore,

$$x_n = f_n^T C^{-1} U \Sigma V^T.$$  \hfill (C4)

Since we choose our basis vectors $\{f_i\}$ via the SVD of $B$, $f_n$ is the $n$th column of $U$. Because the columns of $U$ are orthonormal (automatically satisfying our orthonormality condition from earlier), $(f_n^T C^{-1} U)_{ij} = \delta_{ij}$, where $\delta_{ij}$ is 1 if $i = j$ and 0 otherwise. Writing $\Sigma_{jj} = \sigma_j^2$ to account for its diagonal nature and defining $v_j$ as the $j$th column of $V$, it can be seen that $x_n = \sigma_n v_n$. Hence, if we define $j$ as a vector with the same dimension as $x_n$ (the number of training set vectors, $N$) whose elements are all 1, and note that $|v_n|^2 = 1$ since $V^T V = I$, then the mean and covariance of the $n$th mode weight can be written

$$E[x_n] = \frac{\sigma_n}{N} j^T v_n,$$  \hfill (C5a)

$$\text{Var}[x_n] = \frac{\sigma_n^2}{N} \left[ 1 - \frac{(j^T v_n)^2}{N} \right].$$  \hfill (C5b)

This information from the training set is used to seed Gaussian priors that allow us to suppress variations in unimportant modes (Section 2.4). While in principle covariances of the mode weights in the training set could be added, this could lead to numerical issues when inverting the covariance matrix. In addition, it is conservative to use only the variances in the priors.
Appendix D

Choosing a Proposal Covariance Matrix from an Estimated Covariance Matrix

In Sections 2.3.3 and 2.3.4 we use a function \( c(\alpha) \) that is the constant of proportionality between the estimated covariance matrix of a distribution and the optimal proposal covariance matrix with which to explore that distribution given a desired acceptance fraction of proposals, \( \alpha \). Assuming the distribution to explore is Gaussian with mean \( \mathbf{x} \) and covariance \( \Lambda \), and the proposal distribution is also Gaussian, the acceptance fraction when jumping from \( \mathbf{x} \) if the proposal distribution has covariance \( \Lambda/c \) is equal to

\[
\alpha = \left(1 + \frac{1}{c}\right)^{-N/2},
\]

where \( N \) is the dimension of the Gaussians. Solving for \( c \) as a function of \( \alpha \), this is

\[
c(\alpha) = \frac{1}{\alpha^{-2/N} - 1}.
\]

Thus, the Gaussian distribution that leads to an acceptance fraction of \( \alpha \) when exploring a distribution with covariance \( \Lambda \) from its mean is \( \Lambda/c(\alpha) \).

Appendix E

Full Covariance Matrix of Signal and Foreground Parameters

In this paper, we have outlined a method of exploring only signal parameters in an MCMC exploration while analytically integrating over the foreground parameters. Because the foreground parameters are not numerically explored, the covariance of the foreground parameters with themselves and with the signal parameters cannot be immediately estimated from the MCMC chain of signal parameters alone. However, in order to compute the posterior at the \( n \)th iteration, we find the mean and covariance of the foreground parameters, \( \theta_{fg} \), under the condition that the signal parameters, \( \theta_{21} \), are equal to their values at the \( n \)th iteration, \( \theta_{21}^{(n)} \). We define \( \mu_n \) and \( \Gamma_n \) to be the conditional mean and covariance, respectively, that is,

\[
\mu_n = \text{E}[\theta_{fg}|\theta_{21} = \theta_{21}^{(n)}], \quad \Gamma_n = \text{Cov}[\theta_{fg}|\theta_{21} = \theta_{21}^{(n)}],
\]

where \( \text{E}[\ldots] \) and \( \text{Cov}[\ldots] \) represent the expectation value and covariance respectively and \( \theta_{fg}|\theta_{21} = \theta_{21}^{(n)} \) should be read as \( \theta_{fg} \) under the condition that \( \theta_{21} = \theta_{21}^{(n)} \). By denoting the average of a given quantity \( A_n \) over an MCMC chain as \( \langle A_n \rangle \), we can write the full covariance of foreground and signal parameters,

\[
\Lambda \equiv \text{Cov}\left\{\left[\theta_{21}\right]\right\},
\]

in the following form,

\[
\Lambda = \begin{bmatrix}
\text{Cov}[\theta_{21}, \theta_{21}] & \text{Cov}[\theta_{21}, \theta_{fg}] \\
\text{Cov}[\theta_{fg}, \theta_{21}] & \text{Cov}[\theta_{fg}, \theta_{fg}]
\end{bmatrix}
\]

The correlation matrix can then be computed as

\[
\Xi \equiv [\text{diag}(\Lambda)]^{1/2} \Lambda [\text{diag}(\Lambda)]^{-1/2},
\]

where \( \text{diag}(\Lambda) \) refers to the matrix with the same diagonal elements as \( \Lambda \) but with no off-diagonal elements. While the signal-foreground covariance is robustly and exactly accounted for in our method of marginalizing over the foreground parameters analytically at each MCMC iteration, Equations (E10) and (E11) allow one to employ the MCMC chains to specifically calculate and examine the full parameter covariance and correlation matrix, respectively. Two specific examples of correlation matrices are shown in Figure E1.

\[\text{Note that, in our case, since the conditional foreground distribution is normal due to the fact that the noise is Gaussian and the foreground model is linear, } \mu_n \text{ and } \Gamma_n \text{ completely describe the distribution of } \theta_{fg} \text{ when } \theta_{21} \text{ is fixed to } \theta_{21}^{(n)}. \text{ However, even if this would not be true, Equations (E10) and (E11) hold.} \]
Figure E1. Correlation matrices for the MCMC fits of the FG cases FG4 (left panel) and FG2 (right), as examples of two significantly different instances of this model. The panels of this figure match the cases shown in the panels of Figure 7, correspondingly showing stronger correlations for FG4 than for FG2. Note also the large difference between the number of signal (4) and foreground (40) for each of the four antenna pointings, with a total of 160 parameters. For these cases, therefore, our marginal MCMC technique permitted us to drastically reduce the number of numerically explored parameters from 164 to 4. It is also important to point out that most of the foreground parameters (above a certain mode order) have negligible correlations (in gray), becoming thereby irrelevant thanks to the downweighting power of the foreground priors.

References

Barkana, R. 2018, Natur, 555, 71
Barkana, R., Outmezguine, N. J., Redigol, D., & Volansky, T. 2018, PhRvD, 98, 103005
Bernardi, G., McQuinn, M., & Greenhill, L. J. 2015, ApJ, 799, 90
Bernardi, G., Zwart, J. T. L., Price, D., et al. 2016, MNRAS, 461, 2847
Bassett, N., Rapetti, D., Burns, J., & Tauscher, K. 2019, AdSpR, in press
Bradley, R. F., Tauscher, K., Rapetti, D., & Burns, J. O. 2019, ApJ, 874, 153
Burns, J. O., Bradley, R., Tauscher, K., et al. 2017, ApJ, 844, 33
de Lera Acedo, E. 2019, Proc. Int. Conf. on Electromagnetics in Advanced Applications (Piscataway, NJ: IEEE) 0626, doi:10.1109/ICEAA.2019.8879199
Draine, B. T., & Miralda-Escudé, J. 2018, ApJL, 858, L10
Ewall-Wice, A., Chang, T.-C., Lazio, J., et al. 2018, ApJ, 868, 63
Ewall-Wice, A., Chang, T.-C., & Lazio, T. J. W. 2020, MNRAS, 492, 6086
Feng, C., & Holder, G. 2018, ApJL, 858, L17
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Fialkov, A., & Barkana, R. 2019, MNRAS, 486, 1763
Fialkov, A., Barkana, R., & Cohen, A. 2018, PhRvL, 121, 011101

Field, G. B. 1958, PIRE, 46, 240
Gelman, A., Carlin, J. B., Stern, H. S., et al. 2013, Bayesian Data Analysis (3rd ed.; Boca Raton, FL: CRC Press), doi:10.1201/b16018
Gelman, A., & Rubin, D. B. 1992, StaSc, 7, 457
Handley, W. J., Hobson, M. F., & Lasenby, A. N. 2015, MNRAS, 453, 3484
Harker, G. J. A., Mirocha, J., Burns, J. O., & Pritchard, J. R. 2016, MNRAS, 455, 3829
Haslam, C. G. T., Salter, C. J., Stoffel, H., & Wilson, W. E. 1982, A&AS, 47, 1
Hills, R., Kulkarni, G., Meerburg, P. D., & Puchwein, E. 2018, Natur, 564, E32
Loeb, A., & Muñoz, J. B. 2018, PhyOJ, 11, 69
Mahesh, N., Subrahmanyan, R., Udaya Shankar, N., & Raghunathan, A. 2014, arXiv:1406.2585
Mebane, R. H., Mirocha, J., & Furlanetto, S. R. 2020, MNRAS, 493, 1217
Mirocha, J. 2014, MNRAS, 443, 1211
Mirocha, J., & Furlanetto, S. R. 2019, MNRAS, 483, 1980
Mirocha, J., Skory, S., Burns, J. O., & Wise, J. H. 2012, ApJ, 756, 94
Nhan, B. D., Bordeneuve, D. D., Bradley, R. F., et al. 2019, ApJ, 883, 126
Nhan, B. D., Bradley, R. E., & Burns, J. O. 2017, ApJ, 836, 90
Patra, N., Subrahmanyan, R., Raghunathan, A., & Udaya Shankar, N. 2013, ExA, 36, 319
Philip, L., Abdurashidova, Z., Chiang, H. C., et al. 2019, JAI, 8, 1950004
Price, D. C., Greenhill, L. J., Fialkov, A., et al. 2018, MNRAS, 478, 4193
Pritchard, J. R., & Loeb, A. 2010, PhRvD, 82, 023006
Sims, P. H., & Pober, J. C. 2020, MNRAS, 492, 22
Singh, S., & Subrahmanyan, R. 2019, ApJL, 880, 26
Singh, S., Subrahmanyan, R., Udaya Shankar, N., et al. 2017, ApJL, 845, L12
Sokolowski, M., Tremblay, S. E., Wayth, R. B., et al. 2015, PASA, 32, e004
Spindler, M., Bernardi, G., & Santos, M. G. 2019, MNRAS, 489, 4007
Tauscher, K. 2020, pylinex: linear and nonlinear signal extraction in Python 0.626, doi:10.1109/ICEAA.2019.8879199
Wouthuysen, S. A. 1952, AJ, 57, 31

ORCID iDs

David Rapetti @ https://orcid.org/0000-0003-2196-6675
Keith Tauscher @ https://orcid.org/0000-0003-1932-9829
Jordan Mirocha @ https://orcid.org/0000-0002-8802-5581
Jack O. Burns @ https://orcid.org/0000-0002-4468-2117

The Astrophysical Journal, 897:174 (18pp), 2020 July 10

Rapetti et al.