



Detection of the Cosmological 21 cm Signal in Autocorrelation at $z \sim 1$ with the Canadian Hydrogen Intensity Mapping Experiment

The CHIME Collaboration, Mandana Amiri¹ , Kevin Bandura^{2,3} , Arnab Chakraborty^{4,5} , Jean-François Cliche^{4,5} , Matt Dobbs^{4,5} , Simon Foreman⁶ , Liam Gray¹ , Mark Halpern¹ , Alex S Hill^{7,8} , Gary Hinshaw¹ , Carolin Höfer^{1,9} , Albin Joseph⁶ , Nolan Kruger⁶ , T. L. Landecker⁸ , Rik van Lieshout^{10,11} , Joshua MacEachern⁸ , Kiyoshi W. Masui^{12,13} , Juan Mena-Parra^{14,15} , Kyle Miller⁴ , Nikola Milutinovic¹ , Arash Mirhosseini¹ , Laura Newburgh¹⁶ , Anna Ordog^{7,8,17} , Ue-Li Pen^{10,14,18,19,20} , Tristan Pinsonneault-Marotte^{21,22} , Alex Reda¹⁶ , Andre Renard¹⁴ , Kana Sakaguri¹⁶ , J. Richard Shaw¹ , Shabbir Shaikh⁶ , Seth R. Siegel^{4,5,20,23} , Saurabh Singh²⁴ , David Spear^{1,8} , Yukari Uchibori¹ , Keith Vanderlinde^{14,15} , Haochen Wang^{12,13} , Donald V. Wiebe¹ , and Dallas Wulf^{4,5}

¹ Department of Physics and Astronomy, University of British Columbia, Vancouver, BC, Canada

² Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV, USA

³ Center for Gravitational Waves and Cosmology, West Virginia University, Morgantown, WV, USA

⁴ Department of Physics, McGill University, Montréal, QC, Canada; arnab.chakraborty2@mail.mcgill.ca, Seth.Siegel@skao.int

⁵ Trotter Space Institute, McGill University, Montréal, QC, Canada

⁶ Department of Physics, Arizona State University, Tempe, AZ, USA

⁷ Department of Mathematics, Physics, & Statistics, University of British Columbia-Okanagan, Kelowna, BC, Canada

⁸ Dominion Radio Astrophysical Observatory, Herzberg Astronomy & Astrophysics Research Centre, National Research Council Canada, Penticton, BC, Canada

⁹ Kapteyn Astronomical Institute, University of Groningen, Groningen, The Netherlands

¹⁰ Canadian Institute for Theoretical Astrophysics, Toronto, ON, Canada

¹¹ ASTRON, The Netherlands Institute for Radio Astronomy, Dwingeloo, The Netherlands

¹² MIT Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA, USA

¹³ Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA

¹⁴ Dunlap Institute for Astronomy and Astrophysics, University of Toronto, Toronto, ON, Canada

¹⁵ David A. Dunlap Department of Astronomy & Astrophysics, University of Toronto, Toronto, ON, Canada

¹⁶ Department of Physics, Yale University, New Haven, CT, USA

¹⁷ Department of Physics & Astronomy, University of Western Ontario, London, ON, Canada

¹⁸ Institute of Astronomy and Astrophysics, Academia Sinica, Taipei, Taiwan

¹⁹ Canadian Institute for Advanced Research, Toronto, ON, Canada

²⁰ Perimeter Institute for Theoretical Physics, Waterloo, ON, Canada

²¹ Kavli Institute for Particle Astrophysics and Cosmology, Stanford, CA, USA

²² SLAC National Accelerator Laboratory, Menlo Park, CA, USA

²³ SKA Observatory, Science Operations Centre, Kensington, WA, Australia

²⁴ Raman Research Institute, Sadashivanagar, Bengaluru, India

Received 2025 November 24; revised 2026 June 30; accepted 2026 August 9; published 2026 September 28

Abstract

We present the first detection of the cosmological 21 cm intensity mapping signal in autocorrelation at $z \sim 1$ with the Canadian Hydrogen Intensity Mapping Experiment (CHIME). Using 94 nights of observation, we have measured the 21 cm auto power spectrum over a frequency range of 608.2–707.8 MHz ($z = 1.34$ – 1.01) at $0.4 \text{ h Mpc}^{-1} \lesssim k \lesssim 1.5 \text{ h Mpc}^{-1}$, with a detection significance of 12.4σ . Our analysis employs significant improvements to the CHIME data processing pipeline compared to previous work, including novel radio frequency interference detection and masking algorithms, achromatic beamforming techniques, and foreground filtering before time averaging to minimize spectral leakage. We establish the robustness and reliability of our detection through a comprehensive suite of validation tests. We also measure the 21 cm signal in two independent sub-bands centered at $z \sim 1.08$ and $z \sim 1.24$ with detection significance of 8.6σ and 9.1σ , respectively. We briefly discuss the theoretical interpretation of these measurements in terms of a power spectrum model, deferring the details to a companion paper. This auto power spectrum detection demonstrates CHIME’s capability to probe large-scale structure through 21 cm intensity mapping without reliance on external galaxy surveys.

Unified Astronomy Thesaurus concepts: Cosmology (343); Large-scale structure of the universe (902); H I line emission (690); Radio astronomy (1338); Astronomy data analysis (1858)

1. Introduction

The technique of *21 cm intensity mapping* seeks to measure the clustering of atomic hydrogen (H I) over a wide span of redshift and cosmic volume. Hyperfine transitions in hydrogen atoms are

associated with photons having a rest wavelength of 21.106 cm (corresponding to 1420.406 MHz in frequency), and the cosmic abundance of H I is sufficiently high that radio telescopes can make large-scale maps of the emission or absorption of these photons. The fluctuations in these maps can provide a wealth of information about the behavior of H I in galaxies and the intergalactic medium, along with the background cosmological model.

21 cm intensity mapping can probe the cosmic “Dark Ages” prior to the formation of the first luminous objects, and

instrument concepts for making observations at these high redshifts are currently being explored (e.g., S. D. Bale et al. 2023; X. Chen et al. 2024; R. S. Polidan et al. 2024; C. D. Brinkerink et al. 2025). Intensity mapping observations during the cosmic dawn and epoch of reionization can provide insights into the nature of the ionizing sources and the details of how reionization occurred, and steady progress toward a detection is being made by several teams (e.g., Z. Abdurashidova et al. 2022; E. Ceccotti et al. 2025; F. G. Mertens et al. 2025; C. M. Trott et al. 2025).

In the post-reionization era, most of the H I mass is associated with galaxies, and therefore 21 cm intensity maps in this epoch primarily probe the distribution of these galaxies. As a result, intensity mapping surveys at these redshifts can provide large-scale structure measurements that are complementary to those from optical galaxy surveys, due to the significant differences in observing technologies and observational systematics at optical and radio frequencies. Furthermore, because intensity mapping directly measures the distribution of H I that fuels star formation in galaxies, it can shed light on how galaxies and their environments evolve through cosmic time.

On the other hand, 21 cm intensity mapping observations have proven to be very challenging in practice. At all relevant observing frequencies, Galactic and extragalactic radio foregrounds are several orders of magnitude larger than the cosmological signal. This large dynamic range sets stringent requirements on instrument characterization, removal of radio frequency interference (RFI), mitigation of systematics, removal of foregrounds, and overall robustness of data processing methods (for recent reviews, see, e.g., Cosmic Visions 21 cm Collaboration et al. 2018; Z. Ahmed et al. 2019, and A. Liu & J. R. Shaw 2020).

Most 21 cm intensity mapping analyses thus far have taken the form of cross correlations with previous galaxy survey data. After the initial detections using HIPASS (U.-L. Pen et al. 2009) and the Green Bank Telescope (T.-C. Chang et al. 2010), several others have followed, at redshifts between 0.05 and 2.3 (K. W. Masui et al. 2013; C. J. Anderson et al. 2018; D. Tramonte & Y.-Z. Ma 2020; L.-C. Li et al. 2021; L. Wolz et al. 2022; CHIME Collaboration et al. 2023; S. Cunningham et al. 2023; CHIME Collaboration et al. 2024a; I. P. Carucci et al. 2025; Z. Chen et al. 2025). In addition, S. Paul et al. (2026) have claimed a detection of the 21 cm auto power spectrum at $z = 0.32$ and 0.44 at 8.0σ and 11.5σ , respectively, over the wavenumber range of $0.3 \text{ Mpc}^{-1} \lesssim k < 8 \text{ Mpc}^{-1}$, using data collected by the MeerKAT telescope.

In this paper, we present a measurement of the 21 cm auto power spectrum at a mean redshift of 1.16, using data from the Canadian Hydrogen Intensity Mapping Experiment (CHIME). CHIME is an interferometric radio telescope located at the Dominion Radio Astrophysical Observatory in British Columbia, Canada. CHIME uses four cylindrical reflectors, each 20 m wide and 100 m long, to focus incoming radiation onto 256 dual-polarized feeds located along the focal lines of each cylinder. CHIME observes in the 400–800 MHz band and is pursuing several science goals, including intensity mapping, fast radio bursts (CHIME/FRB Collaboration et al. 2018), pulsars (CHIME/FRB Collaboration et al. 2021), Galactic magnetism (N. Mohammed et al. 2024), and 21 cm absorption systems (CHIME Collaboration et al. 2026b). CHIME

Collaboration et al. (2022a) describes the telescope design and subsystems in more detail.

In CHIME Collaboration et al. (2023), 102 nights of CHIME data were used to perform spectral stacking of radio sky maps on luminous red galaxies (LRGs), emission line galaxies (ELGs), and quasars from the extended Baryon Oscillation Spectroscopic Survey (eBOSS), yielding detections with SNs of 7.1, 5.7, and 11.1. In CHIME Collaboration et al. (2024a), a reprocessed version of the same dataset was used to measure the cross correlation with Ly α forest measurements from eBOSS, at a mean redshift of 2.3 and with 9σ significance. Since those publications, we have made numerous improvements in our data processing pipeline, including the identification of RFI and the mitigation of other sources of spurious spectral structure.

These improvements have enabled CHIME’s first measurement of the 21 cm intensity mapping signal that does not rely on external large-scale structure data. Using 94 nights of data, we have measured the 21 cm auto power spectrum over a range of 608.2–707.8 MHz ($z = 1.34\text{--}1.01$) at $0.4 h \text{ Mpc}^{-1} \lesssim k \lesssim 1.5 h \text{ Mpc}^{-1}$, with a signal-to-noise ratio (S/N) of 12.4.

We have verified that the measurements are consistent across several splits of the data and variations of our analysis choices. We have also separately measured the power spectrum in each of two halves of this frequency band, with an S/N = 9.1 for 608.2 to 658.2 MHz and an S/N = 8.6 for 658.2 to 707.8 MHz.

Our companion paper (CHIME Collaboration et al. 2026a) describes our interpretation of these measurements in terms of the large-scale clustering of H I, but we briefly state our approach here. We use a power spectrum model that depends on the mean density of H I, $\Omega_{\text{HI}}(z)$, and the linear bias of H I, $b_{\text{HI}}(z)$, along with parameters that describe the shape of the power spectrum at the (nonlinear) scales of our measurement. We constrain a degenerate combination of Ω_{HI} and b_{HI} , defined as $\mathcal{A}_{\text{HI}}^2 = 10^6 \Omega_{\text{HI}}^2 (b_{\text{HI}}^2 + \langle f\mu^2 \rangle)^2$, at the mean redshift of our measurement, where $\langle f\mu^2 \rangle$ is a factor capturing the angular average of the Kaiser redshift-space distortion term over the wavenumbers we have measured. In addition, we compare the measured 21 cm power spectrum with that predicted by the distribution of H I in the IllustrisTNG simulations (J. P. Naiman et al. 2018; D. Nelson et al. 2018; F. Marinacci et al. 2018; A. Pillepich et al. 2018; V. Springel et al. 2018).

This paper is organized as follows:

1. In Section 2, we introduce our conventions for coordinates and visibilities, and describe the dataset used in this work.
2. In Section 3, we discuss the data processing pipeline used in this work, focusing on improvements over the pipelines used in previous CHIME cross-correlation studies (CHIME Collaboration et al. 2023, 2024a). In particular, we highlight our methods for RFI flagging (Section 3.1 and Section 3.4), foreground removal (Section 3.2), correcting for residual bandpass-calibration errors (Section 3.3), time averaging (Section 3.5), mapmaking (Section 3.6), and forming pseudo-Stokes maps (Section 3.7). We also describe the identification of several new candidate H I 21 cm absorption systems, and how they are masked in this analysis (Section 3.8).
3. In Section 4, we repeat the spectral stacking analysis of CHIME Collaboration et al. (2023) using eBOSS quasars

and CHIME data processed as in Section 3, in order to validate our pipeline.

4. In Section 5, we describe technical aspects of our power spectrum estimation methodology: delay spectrum estimation (Section 5.1), spatial masking (Section 5.2), binning and normalization (Section 5.3), and noise covariance estimation (Section 5.4).
5. In Section 6, we present our main results: the power spectrum measurement in our chosen frequency band and two sub-bands (Section 6.1). In Section 6.2, we briefly describe the theoretical model we use for interpreting the measurement, and in Section 6.3 we assess the detection significance.
6. In Section 7, we describe our suite of validation tests, which demonstrate that our measurement is free of significant contamination from various systematics.
7. In Section 8, we conclude and discuss the context of our measurements.

Several appendices provide further details about our algorithms for RFI detection (Appendix A), masking (Appendix B), delay spectrum estimation (Appendix C), and estimating detection significance (Appendix D).

For computations requiring a cosmological model, we use cosmological parameters from the final Planck data release (specifically, the “TT,TE,EE+lowE+lensing+BAO” parameters from Table 2 of Planck Collaboration et al. 2020).

2. Instrument and Observations

A full description of the CHIME instrument and its real-time pipeline is given in CHIME Collaboration et al. (2022a), and the offline processing used in earlier cosmology work is described in CHIME Collaboration et al. (2023). Here we introduce the array geometry, coordinate conventions, and visibility notation that will be used in this work, along with the subset of CHIME observations included in this analysis.

2.1. Array Layout

CHIME consists of four cylindrical reflectors separated in the east–west (EW) direction by $\Delta_{\text{EW}} = 22$ m, each equipped with 256 dual-polarization feeds along its north–south (NS) focal line with spacing of $\Delta_{\text{NS}} = 0.3048$ m. The center of the array is located at geodetic latitude and longitude

$$(\Lambda, \Phi) = (49.320\,709, -119.623\,677).$$

To describe baselines and sky directions, we work on the plane tangent to Earth at this location and use a local Cartesian basis $(\hat{x}, \hat{y}, \hat{z})$, with $\hat{x}$ pointing due east on the tangent plane, $\hat{y}$ pointing due north, and $\hat{z}$ pointing to the local zenith.

Indexing the designed CHIME baseline grid by an integer pair (e, n) , we write

$$\mathbf{b}_{en}^{(0)} = e \Delta_{\text{EW}} \hat{x} + n \Delta_{\text{NS}} \hat{y}, \quad (1)$$

with $e \in [-3, 3]$ and $n \in [-255, 255]$ for the nominal array.²⁵ As discussed in CHIME Collaboration et al. (2023), early analyses modeled small departures from this designed layout

as a rotation of the cylinder frame with respect to true north. Subsequent solar (CHIME Collaboration et al. 2022b) and holographic (CHIME Collaboration et al. 2024b) calibration show that these departures are more simply described as a systematic shift in the NS position of the feeds with the cylinder. We therefore use an effective baseline

$$\mathbf{b}_{en} = e \Delta_{\text{EW}} \hat{x} + (n \Delta_{\text{NS}} + \kappa e \Delta_{\text{EW}}) \hat{y}, \quad (2)$$

where $\kappa \simeq -0.00124$ is a small, empirically determined slope giving the NS shift per unit EW separation.

2.2. Coordinate Conventions

We represent sky positions by the direction cosines of the unit sky vector $\hat{\mathbf{n}}$ in the telescope frame $(\hat{x}, \hat{y}, \hat{z})$. These are conventionally written (l, m, n) , but to prevent confusion with the harmonic index m used later, we write

$$(x, y, z) \equiv \hat{\mathbf{n}} \cdot (\hat{x}, \hat{y}, \hat{z}),$$

and refer to (x, y, z) as the telescope coordinates.

We denote the local Earth Rotation Angle (ERA) by ϕ . This quantity replaces local sidereal time and is equivalent to the current R.A. of the local meridian in the Celestial Intermediate Reference System (CIRS).

For a point source at R.A. α and decl. θ , the hour angle (HA) is given by $\text{HA} = \phi - \alpha$, and the telescope frame coordinates are

$$x = -\cos \theta \sin \text{HA}, \quad (3)$$

$$y = \cos \Lambda \sin \theta - \sin \Lambda \cos \theta \cos \text{HA}, \quad (4)$$

$$z = \sin \Lambda \sin \theta + \cos \Lambda \cos \theta \cos \text{HA}, \quad (5)$$

which satisfy $x^2 + y^2 + z^2 = 1$.

2.3. Visibility Data Product

The CHIME real-time pipeline computes visibilities between all pairs of the 2048 digitized inputs for 1024 frequency channels spanning 400 to 800 MHz, integrates them to a cadence of 9.9404 s, applies complex gain calibration, and averages redundant copies of each baseline to reduce the data rate by a factor of ~ 300 . The resulting archived visibilities are denoted V_{pfent} . Here $p \in \{\text{XX}, \text{XY}, \text{YX}, \text{YY}\}$ is the polarization, f indexes frequency channels (with center frequency ν_f), (e, n) label the EW and NS baseline indices on the CHIME grid, and t indexes the time. In this work, we consider only the co-polar visibilities $p \in \{\text{XX}, \text{YY}\}$. Conjugate symmetry then implies

$$V_{pf,-e,-n,t} = (V_{pfent})^*.$$

The measured visibilities contain contributions from sky emission, thermal noise, and contamination such as RFI. The Stokes I contribution from the sky to observed polarization p is given by

$$\begin{aligned} \mathcal{V}_{pfent} = & \int d\phi' \int d\theta' \cos \theta' I(\nu_f, \theta', \phi') \\ & \times |A_p(\nu_f, \theta', \phi_t - \phi')|^2 \\ & \times \exp \{j2\pi\nu_f \mathbf{b}_{en} \cdot \hat{\mathbf{n}}(\theta', \phi_t - \phi')/c\}, \end{aligned} \quad (6)$$

where I is the sky intensity, A_p is the effective voltage primary beam pattern (assumed common to feeds of the same polarization), c is the speed of light, and ϕ_t is the local ERA at time t .

²⁵ Note that in CHIME Collaboration et al. (2023), we used (x, y) instead of (e, n) as baseline indices, while in this work, we instead use (x, y) for the telescope coordinates described in Section 2.2. In CHIME Collaboration et al. (2023), we also used (d_x, d_y) instead of $(\Delta_{\text{EW}}, \Delta_{\text{NS}})$ for cylinder and feed spacings.

For each coarse 9.9404 s integration, the real-time pipeline also constructs an empirical variance estimate for each visibility from even–odd differences of 30 ms sub-integrations. On this timescale, the sky contribution is approximately constant and cancels in the difference (see Appendix A.1). This variance estimate is propagated through the gain-calibration and redundant-baseline averaging operations that are applied to the visibilities, yielding the archived visibility-level variance estimates σ_{pft}^2 . Deviations of σ_{pft}^2 from the radiometer expectation indicate excess variance in the data and are used to identify RFI-contaminated times and frequencies, which are masked from further processing. After this masking, σ_{pft}^2 closely follows the radiometer expectation and provides the visibility-level thermal noise variance estimate used throughout the subsequent analysis.

2.4. Observations

In this work, we analyze 102 sidereal days drawn from the period between 2019 January 1 and 2019 November 5. These 102 days are the same subset identified and used in CHIME Collaboration et al. (2023), where they were selected on the basis of the data quality metrics described in that work.

Because auto power spectrum estimation is more sensitive to residual systematics than the cross-correlation analysis presented in CHIME Collaboration et al. (2023), we revalidate each of the 102 sidereal days using several independent diagnostics, including the delay-domain power spectrum of the visibilities, χ^2 statistics after RFI masking (see Appendix A.3), the total fraction of data masked, and sky maps from individual days at a few representative frequencies. This screening identifies 8 days that fail one or more of these checks, and we exclude them. Our final dataset therefore consists of 94 well-characterized sidereal days from 2019. Note that we use only the portions of each sidereal day when the Sun is below the horizon to avoid contamination from solar emission.

CHIME has been operating continuously since 2019 and now has nearly 7 yr of archival data. Here we focus on this 2019 subset in order to establish and validate the methodology. Analysis of the full archival dataset will be presented in future work.

3. Data Processing Pipeline

In this section, we describe the modifications to the offline data processing pipeline of the CHIME Collaboration et al. (2023) that enable the detection of the 21 cm auto power spectrum. For reference, Table 1 summarizes the principal pipeline choices applied in this analysis, including data selections, masks, filters, weights, corrections, and mode exclusions described throughout this section and in subsequent sections.

3.1. RFI Detection and Masking

We have made substantial improvements to the detection and masking of human-made RFI. The offline algorithms described in CHIME Collaboration et al. (2023) are no longer used; instead, we have developed three new methods, which differ primarily in how they subtract the background sky signal in order to isolate RFI, and in how they combine information across baselines in order to enhance sensitivity to weak interference that would not be detectable in individual visibilities. A detailed description of each algorithm is presented in Appendix A; here we summarize the main features.

3.1.1. Radiometer Test Method

The first method flags frequencies and times where the measured variance of the visibilities deviates from that expected for Gaussian, uncorrelated noise, and excludes them from further analysis.

As discussed in Section 2.3, the real-time pipeline measures the variance of each visibility from even–odd differences of 30 ms subintegrations within each 9.9404 s integration. This timescale is chosen so that the sky contribution is approximately constant and cancels in the even–odd difference, leaving only noise and RFI. The expected variance is computed from the autocorrelations of the two feeds and the number of integrated samples using the radiometer equation. Both the measured and expected variances are then averaged over all baselines within a given polarization product to improve sensitivity to low-level deviations. The square root of the ratio of these baseline-averaged values defines the radiometer test statistic ϱ_{pft} for each polarization product p , frequency channel f , and coarse-resolution time t (see Appendix A.1 for details). Under the Gaussian, uncorrelated noise assumption, ϱ_{pft} has expectation value equal to unity; deviations from this expectation indicate inconsistency with the noise model and are used to flag contaminated data.

Outliers in ϱ_{pft} are identified iteratively as a function of frequency and time for each polarization. A slowly varying background level of ϱ_{pft} is estimated using a robust two-dimensional (2D) median filter and subtracted to remove small, systematic variations with frequency and time. The local scatter of the residuals is estimated with a median absolute deviation (MAD) filter. RFI events are identified using a combination of MAD thresholding for isolated outliers and a custom implementation of the `SumThreshold` algorithm (A. R. Offringa et al. 2010), which is effective at masking events that are extended in time or frequency. The mask, background, and scatter estimates are updated over a fixed number of iterations with progressively lower thresholds to improve sensitivity to low-level contamination. Finally, the polarization-dependent masks are combined with a logical OR across polarizations to produce a single time-frequency mask, Q_{ft}^{rad} , where $Q_{ft}^{\text{rad}} = 0$ indicates good data and $Q_{ft}^{\text{rad}} = 1$ indicates masked data.

We adopt an aggressive masking strategy and prefer to discard marginally contaminated data rather than risk residual RFI. However, we monitor the mask as a function of local ERA to ensure that flagging is not systematically correlated with the sky signal, thereby introducing holes in the sky coverage when averaging over many sidereal days. A more detailed optimization of thresholds is left for future work.

3.1.2. Fringe-rate Filtering Method

The second method targets transient RFI by exploiting the fact that emission from steady celestial sources fringes at a well-defined rate set by Earth’s rotation and the array geometry. For a given frequency, there is a *maximum fringe rate* set by the longest EW baseline and a source at the celestial equator. Any power at higher fringe rates cannot arise from steady signals from the sky and must correspond to transient signals, whether astrophysical or human-made (A. R. Parsons & D. C. Backer 2009; A. R. Offringa et al. 2012a; J. R. Shaw et al. 2014). We apply a high-pass filter in time with a cutoff set by this maximum fringe rate to suppress the celestial

Table 1
Summary of Principal Pipeline Choices

Data Selection					
Pipeline Stage	Operation	Key Parameters	Primary Intent	Fraction Cut	Section
Observation selection	Remove 8 of 102 candidate sidereal days	χ^2 statistics; masked fraction; sky maps; delay power spectrum	Remove sidereal days that fail visual quality assessment	7.8% of candidate sidereal days excluded	Section 2.4
Frequency band selection	Restrict to sub-band	608–708 MHz	Select band with minimal RFI; prevent out-of-band RFI from coupling through delay filter	75% of 400–800 MHz band excluded	Section 3.2.1
Field selection	Restrict angular footprint	$\phi \in [110^\circ, 265^\circ]$, $y \in [-0.19, 0.19]$	Select near-zenith region free from NS aliasing and away from the Galactic plane	94.7% of full sky excluded	Section 3.6
Time-domain selection	Restrict to nighttime data under nominal observing conditions	Sun below horizon; valid calibration; no recent rainfall	Remove solar contamination and poor-quality observing periods	~70% of time excluded	Section 3.5.1
RFI Masking					
Pipeline Stage	Operation	Key Parameters	Primary Intent	Fraction Cut	Section
Static frequency mask	Exclude known bad channels	Fixed channel list	Prevent known persistent contamination from entering later analysis stages	3.5% of frequency channels	Section 3.1.4
Radiometer test method	Identify excess variance relative to radiometer equation	30 ms variance estimate; summed over baselines; iterative background subtraction and thresholding (MAD+SumThreshold)	Remove samples with non-Gaussian variance on short timescales	29.2%	Section 3.1.1 Appendix A.1
Fringe-rate filtering method	Identify outliers after applying high-pass fringe-rate filter to beamformed data	Stokes I FFT map; cutoff set by maximum celestial fringe rate; hysteresis thresholding; 1% flagged-beam threshold	Remove spatially localized contamination with non-celestial time variability	21.1%	Section 3.1.2 Appendix A.2
Spectral filtering method	Identify outliers after applying high-pass delay filter to RFI-masked visibilities	DAYENU with $\tau_{\text{cut}} = 400$ ns; χ^2 per dof calculated over baselines; 5σ threshold	Remove residual spectrally structured contamination	2.2%	Section 3.1.3 Appendix A.3
Final mask assembly	Combine RFI masks	Logical OR of static, radiometer, fringe rate, and spectral masks	Remove contaminated samples identified by any of the above methods	38.7%	Section 3.1.4 Table A1
Foreground and Systematics Mitigation (Before Sidereal Stacking)					
Pipeline Stage	Operation	Key Parameters	Primary Intent	Fraction Cut	Section
NS baseline selection	Restrict NS baseline range	Maximum NS baseline: 73 m	Produce an achromatic and temporally stable NS synthesized beam	1% of NS baselines	Section 3.2.2
NS beamforming	Apply frequency-dependent NS baseline window	Triangular window with $n_{\text{max}} = 239$ and $\nu_{\text{min}} = 608.2$ MHz (see Equation (12))	Match the NS synthesized beam across frequency	...	Section 3.2.2
Time-integration selection	Reject 9.9404 s integrations with excessive missing or masked subintegrations	Maximum missing or masked fraction: 2%	Prevent foreground leakage from averaging over the rotating sky with frequency-dependent sampling	10.2% of time integrations	Section 3.2.3
Foreground and Systematics Mitigation (Before Sidereal Stacking)					
Pipeline Stage	Operation	Key Parameters	Primary Intent	Fraction Cut	Section

Table 1
(Continued)

Data Selection					
Pipeline Stage	Operation	Key Parameters	Primary Intent	Fraction Cut	Section
Delay filtering	Apply DAYENU high-pass delay filter	$\tau_{\text{cut}} = 200 \text{ ns}$; $\epsilon = 10^{-12}$	Suppress spectrally smooth foregrounds before time averaging	$k_{\parallel} < 0.25 \text{ h Mpc}^{-1}$	Section 3.2.3
HyFoReS bandpass correction	Estimate and subtract bandpass-induced foreground residuals	Sky-averaged leakage coefficients per day, frequency, polarization, and EW baseline	Reduce foreground leakage caused by residual bandpass errors	...	Section 3.3
Residual RFI masking	Identify outliers in foreground-filtered, hybrid-beamformed visibilities	χ^2 calculated over EW baselines and $ y < 0.40$; iterative background subtraction and thresholding (MAD+SumThreshold)	Mask residual RFI surviving all prior stages	3.7%	Section 3.4
Noise cross talk removal	Estimate and subtract additive noise crosstalk	Weighted median over 1 hr nighttime reference ϕ range	Remove quasi-stationary additive instrumental noise crosstalk	...	Section 3.5.2
Foreground and Systematics Mitigation (After Sidereal Stacking)					
Pipeline Stage	Operation	Key Parameters	Primary Intent	Fraction Cut	Section
EW baseline weighting	Exclude intracylinder baselines and redundancy weight intercyylinder baselines	$W_e^{\text{EW}} = [0.0, 0.5, 0.333, 0.166]$ for $e = [0, 1, 2, 3]$	Reduce impact of noise crosstalk and diffuse Galactic emission; downweight noisier EW baselines	$k_{\perp} < 0.08 \text{ h Mpc}^{-1}$	Section 3.6
m -mode weighting	Apply instrument response to m -mode transformed data	Model for instrument response	Apply matched filter for sky signal; downweight poorly measured modes	...	Section 3.6
Absorber mask	Apply 3D taper around candidate 21 cm absorbers	40 candidates with an $S/N > 7$; taper with $w_y = 0.018$, $w_{\phi} = 4.8$, $w_{\nu} = 5.1 \text{ MHz}$	Remove narrowband absorption features	1.35% of voxels	Section 3.8 Appendix B
Spatial mask (transit)	Apply elliptical taper around bright source transits	Sources $> 10 \text{ Jy}$ at 600 MHz; taper with $w_y = 0.018$, $w_{\phi} = 4.8$	Remove residual main lobe foreground leakage	1% of selected field	Section 5.2 Appendix B
Spatial mask (track)	Apply taper around full tracks of brightest sources	Sources $> 60 \text{ Jy}$ at 600 MHz; same taper widths as transit mask	Remove foreground leakage through far sidelobes	27% of selected field	Section 5.2 Appendix B
Spatial mask (Galactic plane)	Apply taper around full track of inner Galactic plane	Selected inner Galactic plane source tracks	Remove Galactic emission entering through far sidelobes	7% of selected field	Section 5.2 App. Appendix B
Delay mode exclusion	Exclude low-delay modes from power spectrum estimation	Minimum τ : 280 ns	Avoid modes attenuated by the foreground filter transition region	$k_{\parallel} < 0.35 \text{ h Mpc}^{-1}$	Section 5.3

Note. Rows are listed in approximate processing order. Each row corresponds to a distinct stage of the analysis. The “Fraction Cut” column gives the relevant fraction removed; the range of modes made inaccessible; or ... where no direct cut is applied. For RFI masks, the quoted values give the fraction of time-frequency samples removed. Values in the 608–708 MHz analysis band are quoted unless otherwise noted.

Rows follow the same conventions as in the first part of the table. Values in the 608–708 MHz analysis band are quoted unless otherwise noted.

contribution. The residual contains only transient signals, which we interpret as RFI. We mask data with high-pass residuals above a threshold.

Importantly, this method is applied independently of the radiometer mask described in the previous section; no prior masking is applied at this stage. To reduce data volume, the XX and YY polarization pairs are summed for each baseline.

A high-pass filter is then applied along the time axis by subtracting a flat-top convolved copy of the data. The filtered data are beamformed to enhance sensitivity to localized RFI, producing a (frequency, beam, time) cube. For each frequency and beam, the local scatter is estimated over time using an MAD filter and the data are normalized by this scatter. Hysteresis thresholding is then applied in (beam, time) space: a

pixel is flagged if it exceeds a high threshold, or if it exceeds a lower threshold and is connected to higher-valued pixels. This approach captures the typical triangular time profiles of transient RFI, where samples at either edge of an event have lower power than the peak, while limiting false positives. A sample is flagged if the fraction of beams identified as outliers exceeds 1% for that frequency and time, yielding the final mask.

Among other types of transient RFI, this method is particularly effective at identifying distant broadcast television stations briefly reflecting off objects in the sky, such as airplanes, satellites, or meteor ionization trails. These signals often appear approximately Gaussian in time and can evade the radiometer test described in the previous section.

A minimal secondary stage targets persistent narrowband RFI by exploiting a complementary principle. On long EW baselines, sources fringe rapidly unless they are near the north celestial pole. Applying a low-pass filter in time therefore suppresses the sky contribution in these long EW baselines while isolating persistent sources of RFI. In practice, we apply this stage to baselines formed from feeds separated by two or three cylinders in the EW direction. A 30 minute low-pass filter is applied in time, and the visibilities are then averaged over these baselines to improve sensitivity to faint signals. A median-filtered version of the data along frequency is subtracted to isolate narrowband structure. The `SumThreshold` algorithm is applied in time to detect outliers, and the resulting mask is expanded using a scale-invariant rank operator (A. R. Offringa et al. 2012b). While this stage typically flags a small fraction of data, it is effective at capturing weak, narrowband, persistent RFI that is not detected by other methods. The masks from the transient and persistent stages are combined to form a single time-frequency mask, $Q_{ft}^{\text{t-filt}}$.

3.1.3. Spectral Filtering Method

The delay domain, obtained by Fourier transforming the visibility data along the frequency axis, provides a natural basis for separating the sky signal from many types of RFI (A. R. Thompson et al. 2001; A. R. Parsons & D. C. Backer 2009). We use this to generate our third RFI filter. The sky signal is dominated by astrophysical foreground emission, which is spectrally smooth and therefore confined to low-delay modes, whereas RFI typically occupies narrow spectral bands with sharp band edges and thus has power at high delay. Applying an aggressive high-pass filter in the delay domain removes the spectrally smooth foreground emission and can be used to isolate residual contamination.

For each time and polarization, a high-pass delay filter is constructed using the `DAYENU` method (A. Ewall-Wice et al. 2021), which accounts for frequencies masked by the previous two methods and suppresses delays below the geometric horizon limit plus a buffer. The filter is applied to the visibilities, while the noise variances are propagated through the filter to account for correlations introduced by filtering. A $\chi^2/\text{degrees of freedom (dof)}$ test statistic, $\tilde{\chi}_{ft}^2$, is formed for each frequency and time by taking the sum across baselines of the squared magnitude of the filtered visibilities, weighted by the inverse noise variance, and then dividing by the number of baselines. Under the assumption of thermal noise, this statistic has unit mean and a known variance that decreases with the

number of baselines, making it a sensitive metric for detecting residual contamination.

Outliers are flagged in two stages. First, for each frequency channel we take the median $\tilde{\chi}^2$ over the sidereal day and, if it is an outlier, we mask that channel at all times. Second, for each frequency channel we form the time series of $\tilde{\chi}^2$, compute a slowly varying background in time, and mask only those time samples that depart significantly from this background. This procedure is not applied during daytime, bright source transits, nor around the transit of the pulsar B0329+54, where the statistic is biased by leakage of signal from the astrophysical source. The resulting mask, $Q_{ft}^{\text{t-filt}}$, identifies residual high-delay RFI not captured by the other methods.

3.1.4. Final Mask Assembly

The masks obtained from the three methods discussed above are combined using a logical OR to produce the global time-frequency mask,

$$Q_{ft} = Q_{ft}^{\text{rad}} \vee Q_{ft}^{\text{t-filt}} \vee Q_{ft}^{\text{f-filt}}, \quad (7)$$

where $Q_{ft} = 0$ for unmasked (good) data and $Q_{ft} = 1$ for masked or missing data. We also define the complementary “good data” flag

$$G_{ft}^{(0)} \equiv 1 - Q_{ft}, \quad (8)$$

so that $G_{ft}^{(0)} = 1$ for unmasked (good) data and $G_{ft}^{(0)} = 0$ for masked or missing data. The superscript “(0)” indicates that this is the initial version of the good data flag. This serves as the starting point for downstream stages, which may incorporate additional criteria and define updated versions used in subsequent analysis. The fraction of data in the analysis band removed by each RFI mask and by their combination is provided in Table 1.

3.2. Foreground Removal

We adopt a foreground removal strategy that exploits the distinct spectral behavior of the cosmological 21 cm signal and the astrophysical foregrounds. At the radio frequencies considered, foreground emission is primarily Galactic and extragalactic synchrotron radiation, which in total intensity (Stokes I) varies smoothly with frequency (T. Di Matteo et al. 2002; S. P. Oh & K. J. Mack 2003; A. de Oliveira-Costa et al. 2008). In contrast, the 21 cm signal exhibits structure across a broad range of spectral scales. This is because the observed frequency of the 21 cm line corresponds to the redshift of the emitting or absorbing hydrogen gas, which in turn sets its approximate comoving distance along the line of sight. The three-dimensional (3D) spatial distribution of neutral hydrogen across cosmological volumes therefore manifests as frequency-dependent fluctuations in the observed 21 cm signal, with the scale and amplitude of these fluctuations governed by the line-of-sight power spectrum of H I. By applying a high-pass filter in frequency, we suppress the spectrally smooth foregrounds while retaining the higher line-of-sight ($k_{\parallel}$) modes of the 21 cm signal. While this comes at the cost of discarding the low- $k_{\parallel}$ modes that are most sensitive to cosmological parameters, it provides a simple framework for isolating the relatively uncontaminated regions of Fourier space where the 21 cm signal is expected to dominate.

In practice, the success of this strategy depends critically on how much additional frequency structure is imprinted on otherwise smooth foregrounds by the telescope and the data processing pipeline. Since the foregrounds are several orders of magnitude brighter than the expected 21 cm signal, even weak spectral dependence in the instrument or analysis can cause significant leakage, overwhelming the faint cosmological signal.

Instrumental chromaticity in an interferometer arises from several well-understood physical mechanisms. First, the signal from a source in a given sky direction $\hat{n}$ reaches the two antennas of a baseline $\mathbf{b}$ with a geometric delay $\tau = \mathbf{b} \cdot \hat{n}/c$. When these signals are correlated, the contribution of that source to the visibility exhibits a sinusoidal fringe pattern as a function of frequency, with a period inversely proportional to this delay (see Equation (6)). Longer baselines produce higher delay (i.e., more rapidly varying) fringes, and the maximum delay is set by a source on the horizon aligned with the baseline vector. This geometric effect confines smooth-spectrum foreground emission to a characteristic region in Fourier space known as the *foreground wedge* (M. F. Morales et al. 2012; A. R. Parsons et al. 2012; A. Liu et al. 2014), whose extent is determined by the horizon delay limit for each baseline.

Additional sources of chromaticity include mutual coupling between nearby antenna elements, which modifies the effective primary beam pattern in a frequency-dependent way (A. T. Josaitis et al. 2022), and reflections in the analog signal chain, which introduce sinusoidal ripples in the bandpass (N. S. Kern et al. 2019). While these effects are not negligible, they are all bounded by the physical layout and electrical length scales of the telescope. As such, contamination from instrumental chromaticity is confined to delay space, typically within the wedge and a modest buffer region.

By contrast, spectral artifacts imprinted on the foregrounds during data processing are not so constrained and can easily contaminate all Fourier modes along the line of sight. Minimizing analysis-induced spectral structure is therefore critical for isolating the cosmological signal. In the remainder of this section, we describe the design choices in our pipeline aimed at suppressing spectral features and ensuring that the high-delay power is dominated by 21 cm fluctuations rather than analysis artifacts.

3.2.1. Sub-band Isolation

We restrict attention to a relatively RFI-free narrow band of data between 608.2 and 707.8 MHz, which we process separately from the rest of the data. This choice serves several purposes. First, it prevents artifacts from possibly contaminated regions of the spectrum from coupling into the frequencies of interest via the foreground filter. Second, when beamforming, we intentionally degrade our maps to a common spatial resolution to enforce spectral smoothness (see Section 3.2.2), so we prefer to limit the loss of resolution to the portion of the band directly relevant for this analysis. Third, the portion of the sky that is free from geometric aliasing in the NS direction decreases with increasing frequency, so treating the band in contiguous, limited segments ensures that each segment is analyzed with a consistent spatial footprint. The remaining portions of the 400–800 MHz band will be addressed in future work.

3.2.2. Beamforming in the NS Direction

We beamform over NS baselines to a uniform grid in the telescope y coordinate, indexed by $d \in [0, \dots, N_y - 1]$, with $N_y = 1217$ points equally spaced between -1 and $+1$:

$$\check{V}_{pfedt} = \sum_n W_{fn}^{\text{NS}} V_{pfent} \exp \{-j2\pi\nu_f y_d \mathbf{b}_{en} \cdot \hat{\mathbf{y}}/c\}. \quad (9)$$

We refer to $\check{V}_{pfedt}$ as the hybrid-beamformed visibilities, since they have been beamformed in the NS direction but not in the EW direction. W_{fn}^{NS} is a window function, taken to be the same for all polarizations, EW baseline separations, and times. The window function is normalized separately at each frequency to ensure that the synthesized beam will have a unit peak response:

$$W_{fn}^{\text{NS}} \equiv \frac{w_{fn}^{\text{NS}}}{\sum_n w_{fn}^{\text{NS}}}, \quad (10)$$

where w_{fn}^{NS} is the unnormalized window function. We propagate the noise variance through the beamforming operation

$$\check{\sigma}_{pfet}^2 = \sum_n (W_{fn}^{\text{NS}})^2 \sigma_{pfent}^2, \quad (11)$$

where σ_{pfent}^2 is the variance of the visibility estimated by differencing even and odd time samples at 30 ms cadence. Here we assume noise that is uncorrelated across baselines, such that the d -dependent geometric phase does not change the beamformed variance, and the same $\check{\sigma}_{pfet}^2$ is assigned to all grid points d . We will return to this assumption in Section 5.4.

Because sightlines near zenith are well sampled by the array in the NS direction, emission in the $|y| \lesssim 0.4$ region relevant to this work does not suffer from NS aliasing. By applying a fixed taper W_{fn}^{NS} across the NS baselines, we form synthesized beams in y with controlled sidelobes, and the long NS extent of the array gives these beams narrow main lobes. In effect, the beamforming step projects the visibilities onto sky directions labeled by telescope y_d : signals arriving from y_d add coherently (their NS geometric delay is compensated), while signals from other y are significantly attenuated by the synthesized beam. In the original visibility domain, the foreground delay extent grows with NS baseline length; after projection onto the y_d beams, this dependence is removed, so a single, y -independent foreground filter is adequate. Directions with $|y| \gtrsim 0.4$ are not well localized in this band and can exhibit frequency-dependent aliasing; we exclude these directions from power spectrum estimation for this initial analysis.

This localization will later enable targeted masking of bright sources. Because foreground leakage is primarily multiplicative in sky intensity, the dominant contamination comes from a small number of very bright sources as they transit. By masking the corresponding y_d beams at the relevant times we reduce leakage to high delays and limit contamination of the 21 cm signal.

Because the EW baselines provide only sparse coverage of the corresponding spatial Fourier modes, we cannot form EW beams with localization comparable to the NS case, so sky emission at large telescope x is observed with a large EW geometric delay that we do not correct at this stage. However, sky emission at large telescope x is already significantly attenuated by the primary beam response (e.g., CHIME

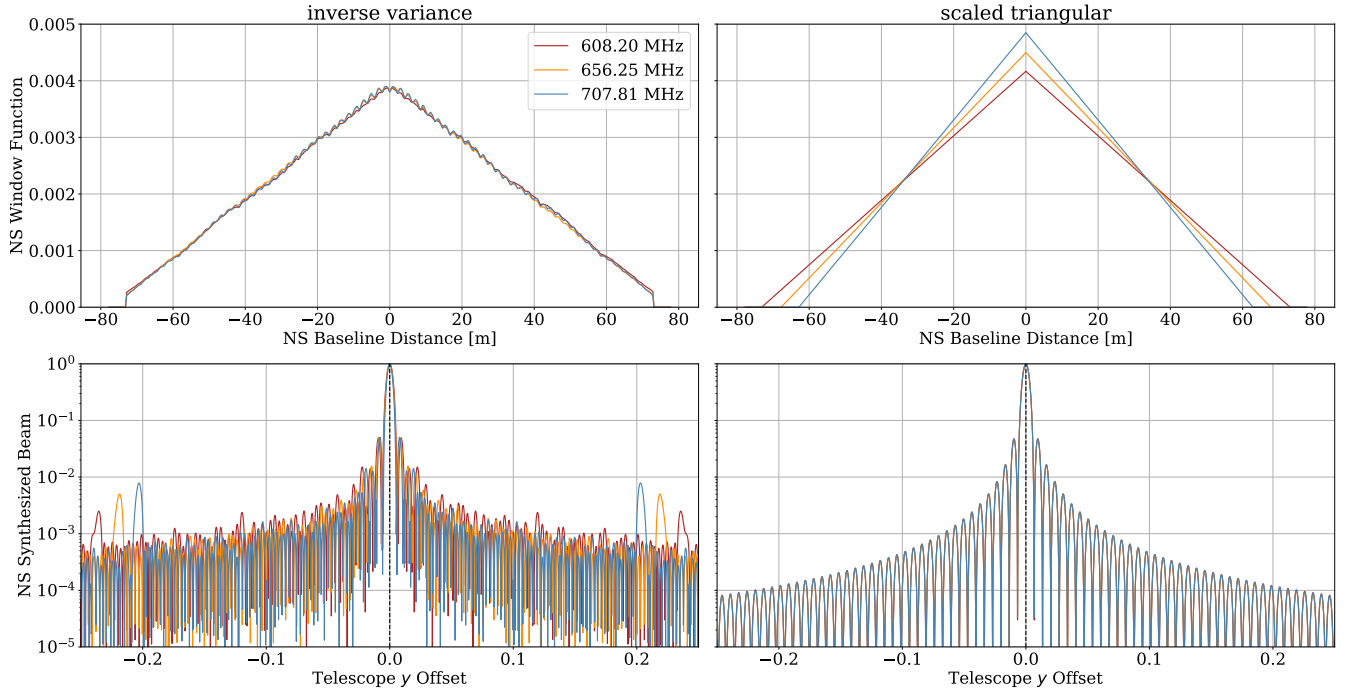


Figure 1. Top: normalized window function, W^{NS} vs. NS baseline distance, shown at three representative frequencies near the bottom, middle, and top of the band. The left column shows the window function used in CHIME Collaboration et al. (2023), which is derived from the inverse-variance weights, while the right column shows the scaled triangular window function adopted in this work. Bottom: Fourier transforms of the normalized window functions, illustrating the expected synthesized beams as a function of offset in the telescope y coordinate. The inverse-variance weights produce a beam that narrows with increasing frequency and exhibits large frequency-dependent sidelobes, in part due to ripple-like structure from noise crosstalk. By contrast, the triangular window is scaled with frequency to maintain an identical synthesized beam across the band.

Collaboration et al. 2022b, 2024b). Additional suppression can be obtained with a fringe-rate filter and by masking the very brightest sources in the sidelobes, as discussed below.

Excluding the longest NS baselines. Although CHIME’s maximum NS baseline is 77.7 m, we exclude baselines longer than 73 m from the previously discussed beamforming stage. These baselines are few in number and primarily connect feeds at the cylinder ends, which have distinct primary beam patterns relative to the interior feeds. Including them would risk distorting the synthesized beam and compromising localization. Moreover, because of their low redundancy, these baselines are vulnerable to disappearing entirely when feeds are masked in the real-time pipeline, which would induce time-dependent variations in the synthesized beam. While our foreground filtering procedure mitigates sensitivity to such variations, we prefer to maintain a synthesized beam that is stable and straightforward to model. Excluding the longest baselines achieves this with negligible loss of sensitivity.

Applying achromatic beamforming weights. We depart from the inverse-variance weighting used in CHIME Collaboration et al. (2023) and instead use a frequency-dependent triangular window versus NS baseline length that yields an achromatic synthesized beam. These two options are compared in Figure 1. With inverse-variance weights, the synthesized beam naturally narrows with increasing frequency due to the wavelength dependence of the Fourier transform. The weights can also inherit fine-scale spectral structure from noise crosstalk and residual RFI, imprinting additional artificial chromaticity on the synthesized beam. Our triangular window

function is defined as

$$w_{fn}^{\text{NS}} = \begin{cases} 1 - |\rho_{fn}|, & |\rho_{fn}| \leq 1 \\ 0, & |\rho_{fn}| > 1 \end{cases} \quad (12)$$

with

$$\rho_{fn} \equiv \frac{n}{n_{\text{max}}} \frac{\nu_f}{\nu_{\text{min}}}. \quad (13)$$

Here $n_{\text{max}} = 239$ is the maximum NS baseline index, and $\nu_{\text{min}} = 608.2$ MHz is the minimum frequency considered. Equation (12) produces a frequency-independent synthesized beam that approximately matches the inverse-variance scheme at the lowest frequency, but progressively upweights short baselines relative to long baselines at higher frequency to maintain achromaticity. The resulting synthesized response in the y -direction (see, e.g., K. W. Masui et al. 2019) is

$$B_{\text{synth}}^{\text{NS}}(y - y_d) \approx \frac{\sin^2(\pi \nu_{\text{min}} n_{\text{max}} \Delta_{\text{NS}}(y - y_d)/c)}{n_{\text{max}}^2 \sin^2(\pi \nu_{\text{min}} \Delta_{\text{NS}}(y - y_d)/c)}, \quad (14)$$

which has an FWHM of 0.006 in telescope y coordinates, corresponding to roughly $20'$ at zenith. Crucially, the synthesized response described by Equation (14) does not vary with frequency. This approach sacrifices a small amount of point-source sensitivity in favor of enforcing an achromatic synthesized beam and avoiding the introduction of artificial spectral structure into the data prior to foreground filtering.

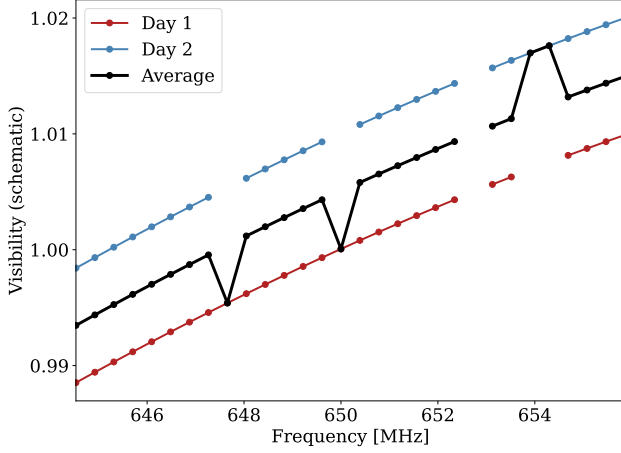


Figure 2. Schematic illustration of how spectrally smooth gain variations can couple to the RFI mask, resulting in leakage of foreground power into high-delay modes. The red and blue curves show the real component of a visibility as a function of frequency over a 10 MHz sub-band for the same local ERA bin on two different days. They differ by an overall 1% multiplicative factor that is approximately constant across this sub-band, representative of gain variations expected in CHIME. Gaps correspond to channels flagged by the RFI mask. The black curve shows the average of the 2 days, which is no longer spectrally smooth despite each day being smooth individually. This motivates our choice to apply foreground filtering on individual days prior to averaging.

3.2.3. Foreground Filtering Before Time Averaging

In addition to the improved RFI masking described in Section 3.1, our second major processing change is to apply a foreground filter directly to the 9.9404 s time series visibilities output by the correlator, prior to any additional time averaging. Frequency-dependent RFI masks used during time averaging couple to temporal variation in the bright, spectrally smooth foregrounds, imprinting spurious high-delay power (see Figure 2 for an illustration of this effect). Because these foregrounds are orders of magnitude brighter than the 21 cm signal, even small changes over the averaging interval can bias 21 cm power spectrum estimates. This temporal variation in the foregrounds is caused by Earth’s rotation when rebinning the data to a fixed grid in local ERA, and by day-to-day variability in the instrument response during sidereal day averaging. To control this effect, we first discard any 9.9404 s integration for which the real-time system reports more than 2% of samples missing or masked, and then apply a per-integration foreground filter along the frequency axis, constructed using the contemporaneous RFI mask. After filtering, the data are integrated in time further, first rebinning to a common local ERA grid and then averaging over sidereal days. Throughout, we also accumulate the sidereal day-averaged filter and frequency–frequency noise covariance, which are used to propagate the filtering to 21 cm simulations and to estimate the noise, respectively.

To filter foregrounds, we use the DAYENU algorithm (A. Ewall-Wice et al. 2021) with an implementation based on that described in CHIME Collaboration et al. (2023), applied to the NS beamformed visibilities $\check{V}_{pfedt}$ introduced in Section 3.2.2. In this space, where the NS interferometric delays have been removed, and for the region of the sky considered here, the foregrounds should be confined to delays below $\tau \approx 200$ ns, beyond which the delay spectrum becomes noise-like (see Figure 10 in CHIME Collaboration et al. 2023). We thus adopt a simple model for the foreground plus noise covariance designed to deweight the foreground contaminated

delays:

$$\tilde{C}(\tau, \tau') = \begin{cases} \epsilon^{-1} \frac{1}{2\tau_{\text{cut}}} \delta^D(\tau - \tau'), & |\tau| \leq \tau_{\text{cut}}, \\ \Delta_\nu \delta^D(\tau - \tau'), & |\tau| > \tau_{\text{cut}}, \end{cases} \quad (15)$$

where $\delta^D(\tau - \tau')$ is the Dirac delta function, $\Delta_\nu = 390.625$ kHz is the channel width, $\tau_{\text{cut}} = 200$ ns is the delay cutoff, and $\epsilon = 10^{-12}$ is the assumed ratio of noise to foreground variance. Fourier transforming to frequency space yields an analytic expression for the frequency–frequency covariance,

$$[C]_{ff'} = \epsilon^{-1} \text{sinc}[2\pi\tau_{\text{cut}}(\nu_f - \nu_{f'})] + \delta_{ff'} \quad (16)$$

where here $\delta_{ff'}$ is the Kronecker delta function. At each time index t , we apply the frequency-dependent RFI mask by forming the diagonal masking operator $[G_t]_{ff'} = \delta_{ff'} G_{ff}^{(0)}$, which zeros flagged frequencies. We then compute the Moore–Penrose pseudoinverse to obtain the inverse-covariance filter

$$H_t = (G_t C G_t)^+, \quad (17)$$

where $^+$ denotes the pseudoinverse. Foreground removal is achieved by weighting the visibilities with this inverse covariance, i.e.,

$$\check{V}_{pfedt}^{\text{hpf}} = \sum_{f'} H_{ff'} \check{V}_{pf'et}. \quad (18)$$

Thus, the suppression of smooth foregrounds is cast as an inverse-covariance weighting problem under this simple foreground covariance model.

The filter H_t depends only on the RFI mask and the assumed covariance, and is identical for all polarizations, EW baselines, and telescope y coordinates at a given time. Unlike the implementation in CHIME Collaboration et al. (2023), we compute a distinct filter for each 9.9404 s integration and apply it before rebinning and averaging over sidereal days, mitigating spectral leakage from time variability in the visibilities coupling to the RFI mask.

The frequency–frequency noise covariance at time index t is propagated through the same operator:

$$\Sigma_{petff'} = \sum_{f''} H_{tf''} \check{\sigma}_{pf''et}^2 H_{f''f'}, \quad (19)$$

where $\check{\sigma}_{pf''et}^2$ is the beamformed noise variance given by Equation (11). Both the filter and the frequency–frequency noise covariance are retained and propagated through the pipeline. These are later used to apply consistent filtering to simulated 21 cm signals and to generate noise realizations, both of which are essential for interpreting the measured power spectrum.

3.3. Estimating and Correcting Residual Bandpass Errors with HyFoReS

After applying the foreground filter, we observe residual leakage of bright point sources in the hybrid-beamformed visibilities. This leakage is spatially coincident with the brightest sources on the sky and is consistent with structure in the instrumental bandpass that couples with foreground emission to evade the delay filter. To quantify this effect, we construct a model for how a point source j would appear in the *unfiltered* hybrid-beamformed visibilities, evaluated at the

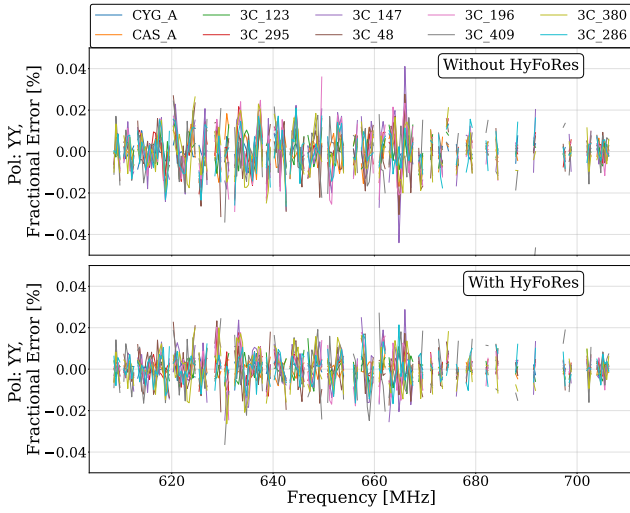


Figure 3. Bandpass leakage coefficients (fractional units) for the YY polarization, 22 m EW baseline as a function of frequency, measured from 10 bright point sources. Top panel: Leakage coefficients estimated from the sidereal day stack of foreground-filtered hybrid visibilities using the point-source fitting procedure of Equation (22). Bottom panel: leakage coefficients estimated from the corresponding stack when HyFoReS is applied to the per-day foreground-filtered hybrid visibilities before sidereal day averaging. HyFoReS reduces the amplitude of the bandpass leakage coefficients across most of the band.

decl. grid point d_j nearest to the source. For polarization p , frequency channel f , and EW baseline index e , the template is

$$\Psi_{pfe}(\theta_j, \phi - \alpha_j) = S(\nu_f) B_{pfe}(\theta_j, \phi - \alpha_j), \quad (20)$$

where (α_j, θ_j) are the R.A. and decl. of the source in CIRS coordinates, S is a model of the source flux density taken from an external catalog of extragalactic point sources, `specfind v3` (Y. Stein et al. 2021), interpolated into the CHIME band, and B is the beam transfer function,

$$B_{pfe}(\theta_j, \phi - \alpha_j) = |A_{pf}(\theta_j, \phi - \alpha_j)|^2 \times e^{-j2\pi\nu_f e \Delta_{EW}(\theta_j, \phi - \alpha_j)/c}, \quad (21)$$

where A_{pf} is our model for the average primary beam response evaluated at frequency channel f and for polarization p , and the exponential factor accounts for the phase introduced by the geometric delay along the projected EW baseline.

We fit this template to the *filtered* hybrid visibilities $\check{V}^{\text{hpf}}$ to solve for a multiplicative leakage coefficient $\delta g^{(j)}$ for each polarization, frequency, and EW baseline:

$$\widehat{\delta g}_{pfe}^{(j)} = \frac{\sum_t G_{ft}^{(0)} \Psi_{pfe}^*(\theta_j, \phi_t - \alpha_j) \check{V}_{pfedt}^{\text{hpf}}}{\sum_t G_{ft}^{(0)} |\Psi_{pfe}(\theta_j, \phi_t - \alpha_j)|^2}, \quad (22)$$

where $G_{ft}^{(0)}$ is 1 for valid samples and 0 for missing or masked samples (see Equation (8)), and the sum runs over a time window centered on the source's transit and spanning the extent of the primary beam at that decl. In the absence of instrumental systematics, the delay filter would remove the point-source contribution entirely, and $\widehat{\delta g}$ would be consistent with noise. In practice, residual high-delay structure in the bandpass gain couples with bright foreground flux and leaks through the delay cut, producing a measurable $\widehat{\delta g}$ for the brightest sources (see Figure 3, top panel).

We find that the frequency dependence of the multiplicative leakage coefficients $\widehat{\delta g}^{(s)}$ is largely common across different sources j , suggesting that a single correction per polarization, frequency channel, and baseline can account for most of the leakage. We interpret this shared frequency dependence as arising from an effective bandpass error introduced during daily real-time calibration. Instead of estimating these coefficients from only a limited set of bright sources, we use the hybrid foreground residual subtraction (HyFoReS) algorithm to estimate the leakage coefficients by summing over the full sky (H. Wang et al. 2022, 2025, 2026).

HyFoReS is a class of foreground subtraction algorithms designed to remove residual foregrounds in linearly filtered 21 cm data. The key assumption underlying HyFoReS is that foreground residuals in the filtered data are correlated with the true foreground emission. By cross-correlating the filtered data with a suitable foreground template, one can estimate both the foreground residuals and any parameterizable systematics (such as bandpass-related leakage) that produce them.

A natural choice for the foreground template would be the prefiltered hybrid visibilities $\check{V}$, since they are dominated by foregrounds. However, the prefiltered data also contain 21 cm signal and noise components that can correlate with the filtered data and bias the estimated leakage coefficients. To avoid this, we construct the foreground template by subtracting the filtered visibilities from the prefiltered visibilities, $\check{V}^{\text{lpf}} = \check{V} - \check{V}^{\text{hpf}}$, thereby removing the cosmological signal and noise contributions at high delay. In addition, noise crosstalk is subtracted from both the foreground template and the filtered data using the procedure described in Section 3.5.2.

HyFoReS then estimates the leakage coefficients by cross-correlating the filtered data with the template, effectively averaging over all sky positions:

$$\widehat{\delta g}_{pfe} = \frac{\sum_d \sum_t G_{ft}^{(0)} \check{V}_{pfedt}^{\text{lpf},*} \check{V}_{pfedt}^{\text{hpf}}}{\sum_d \sum_t G_{ft}^{(0)} \check{V}_{pfedt}^{\text{lpf},*} \check{V}_{pfedt}^{\text{lpf}}}. \quad (23)$$

These coefficients are estimated independently for each sidereal day. The averages are taken over nighttime data on that day, excluding telescope y coordinates that experience significant NS aliasing ($|y| \gtrsim 0.4$). Within the nighttime data, we use uniform weighting, but exclude all samples that are masked in either time or frequency. The resulting $\widehat{\delta g}_{pfe}$ thus represents a daily, sky-averaged estimate of the residual bandpass error for each polarization, frequency channel, and EW baseline.

The estimated leakage coefficients are then used to subtract bandpass-induced foreground residuals from the filtered visibilities:

$$\check{V}_{pfedt}^{\text{cal}} = \check{V}_{pfedt}^{\text{hpf}} - \sum_{f'} H_{ff'} \widehat{\delta g}_{pf'e} \check{V}_{pf'edt}, \quad (24)$$

where $H_{ff'}$ is the delay filter operator introduced in Equation (17) and $\check{V}_{pfedt}^{\text{cal}}$ denotes the bandpass-calibrated, foreground-filtered hybrid-beamformed visibilities.

To assess the impact of HyFoReS, we re-estimate the leakage coefficients from the bandpass-calibrated hybrid visibilities following the point-source fitting procedure described above. The results are shown in the bottom panel of Figure 3, demonstrating that a single sky-averaged correction

provides a modest reduction in bandpass-induced foreground leakage.

3.4. Residual RFI Detection and Masking

After applying the pipeline steps described above, we observe residual contamination in the foreground-filtered data with time-frequency morphology consistent with RFI. This is most apparent when we form a χ^2 -like metric by summing the squared magnitude of the foreground-filtered, hybrid-beamformed visibilities with inverse-variance noise weighting over telescope y coordinates. To address this, we apply a final masking stage that follows the overall logic of the *spectral filtering method* in Appendix A.3, but is modified to operate directly on the hybrid-beamformed visibilities used for power spectrum estimation. For this stage, we also adopt the more advanced outlier identification algorithm originally developed for the *radiometer test method*, as described in Appendix A.1.

We operate directly on the main data product, $\tilde{V}_{pft}^{\text{cal}}$, which employs a less aggressive delay cut (200 ns) compared to that used in the previous masking stage (400 ns), as discussed in Section 3.2.3 and Appendix A.3. The χ^2 test statistic is evaluated by summing over intercylinder EW baselines and telescope y coordinates near zenith ($|y| \lesssim 0.40$), consistent with the subset of the data that will be used for power spectrum estimation. Because the noise is correlated as a function of telescope y , we calibrate the background level and scatter of the test statistic using robust, median-based estimators rather than assuming idealized Gaussian statistics.

Masking is performed using a multi-pass iterative procedure identical to that in Appendix A.1, but with three modifications: the threshold $\eta^{(k)}$ is lowered by a factor of 1.5 at each iteration from an initial value of $\eta^{(0)} = 15.1875$ to a final value of $\eta^{(4)} = 3$; the maximum convolution kernel length in the SumThreshold algorithm is set to $l_{\text{max}} = 16$ samples; and the targeted flagging of digital TV channels is omitted. The resulting mask is denoted $Q_{pft}^{\text{f-filt-2}}$.

Unlike earlier stages, the mask is allowed to vary with polarization, minimizing unnecessary data loss. The resulting mask is incorporated into the global good data flag,

$$G_{pft}^{(1)} = G_{ft}^{(0)} \times (1 - Q_{pft}^{\text{f-filt-2}}), \quad (25)$$

and is used in all downstream analyses.

3.5. Time Averaging

The next step is to further average the bandpass-calibrated, foreground-filtered hybrid-beamformed visibilities given by Equation (24), which are sampled at the native 9.9404 s cadence, over longer timescales in order to reduce thermal noise. This is done in four stages. First, we apply a time-domain data quality mask that excludes known contaminated intervals, including all periods when the Sun is above the horizon. Second, we subtract a daily estimate of the noise crosstalk. Third, we rebin each sidereal day onto a fixed grid in local ERA. Finally, we average over sidereal days. Because the bright, spectrally smooth foregrounds have already been filtered at high time resolution using a high-pass delay filter, these subsequent averaging steps no longer risk coupling temporal variations in the foregrounds into high-delay modes. This allows us to adopt a simpler averaging scheme than in CHIME Collaboration et al. (2023), as described below.

3.5.1. Daytime and Data Quality Masking

We construct two masks that identify time intervals to be excluded from the analysis. The daytime mask Q_t^{day} takes the value 1 for time samples when the Sun is above the horizon and 0 otherwise. Similarly, the data quality mask Q_t^{bad} takes the value 1 during intervals of poor data quality and 0 otherwise. These include intervals associated with correlator or acquisition restarts, rainfall, anomalous autocorrelation behavior, or poor complex gain calibration. Both masks are obtained following the procedures described in Section 3.2.7 of CHIME Collaboration et al. (2023). In contrast to that analysis, however, we do not mask times when the Moon is transiting, as it remains at telescope y coordinates outside the range considered in this work. We combine these time-domain masks with the current good data flag $G_{pft}^{(1)}$ (defined in Equation (25)) to obtain

$$G_{pft}^{(2)} = G_{pft}^{(1)} \times (1 - Q_t^{\text{day}}) \times (1 - Q_t^{\text{bad}}) \quad (26)$$

so that $G_{pft}^{(2)} = 1$ only for nighttime samples that are not flagged as RFI-contaminated by any of our four tests and that pass all other time-domain quality cuts, and $G_{pft}^{(2)} = 0$ otherwise.

3.5.2. Noise Crosstalk Removal

CHIME visibilities contain a quasi-stationary additive component arising from instrumental noise crosstalk. As discussed in Section 3.3.4 of CHIME Collaboration et al. (2023), this crosstalk exhibits a complex spatial and spectral structure, but its temporal variability is relatively modest. In particular, the rms intraday variations are typically $\lesssim 5\%$ of the mean level for zero-cylinder separations, increasing gradually at larger separations, while the interday variations are only slightly larger. This low level of temporal variation motivates a simplified treatment of the crosstalk in our analysis.

Physically, crosstalk arises when receiver noise from one feed is broadcast and subsequently picked up by neighboring feeds, producing an additive offset in their cross correlation. The frequency dependence of this offset is set by a small number of discrete delays corresponding to different propagation paths between the feeds, sometimes including one or more reflections off the cylinder surface.

In our analysis, we represent the crosstalk as a single additive term per sidereal day, with full dependence on polarization, frequency, EW baseline, and telescope y coordinate. Unlike CHIME Collaboration et al. (2023), where crosstalk estimation and subtraction were performed directly on the visibilities prior to foreground filtering, we estimate and subtract our crosstalk estimate in hybrid-beamformed visibility space *after* the foreground filter has been applied. This ensures that the subtraction targets only the residual crosstalk at high delay and is not biased by the bulk of the bright sky signal, which has already been removed by the filter.

We estimate the additive crosstalk component during nighttime to avoid bias from the Sun. However, the portion of the sky that transits at night changes over the course of the year (see Figure 4 of CHIME Collaboration et al. 2023). Since any constant component of the sky within the estimation interval will be absorbed into the crosstalk estimate and subtracted, it is important to apply this procedure consistently throughout the year.

To accomplish this, we divide the year into four quarters, each covering a successive 3 month period. For each quarter,

we assign a fixed reference range in ϕ that will be used to estimate and subtract the additive crosstalk component. These ranges are denoted $\mathcal{Q}_1$ – $\mathcal{Q}_4$ and are each one hour wide. They are chosen to satisfy two key criteria: (1) they must transit during the nighttime throughout the associated quarter, and (2) they must avoid bright sources such as Cygnus A and Cassiopeia A. Specifically,

$$\begin{aligned}\mathcal{Q}_1 : 150^\circ &\leq \phi < 165^\circ \\ \mathcal{Q}_2 : 240^\circ &\leq \phi < 255^\circ \\ \mathcal{Q}_3 : 315^\circ &\leq \phi < 330^\circ \\ \mathcal{Q}_4 : 15^\circ &\leq \phi < 30^\circ.\end{aligned}$$

These ranges differ slightly from those used in CHIME Collaboration et al. (2023), where $\mathcal{Q}_1$ and $\mathcal{Q}_4$ were defined as 165° – 180° and 45° – 60° respectively; these modifications were made to avoid antipodal transits of Cassiopeia A and other bright sources.

For each sidereal day, we identify the time intervals during which the relevant reference range in ϕ is transiting overhead and compute, for every polarization, frequency, EW baseline, and telescope y coordinate, the *weighted median* of the hybrid-beamformed visibilities over this interval. The weights are given by the flag dataset $G_{pft}^{(2)}$. The weighted median is applied separately to the real and imaginary components of the visibilities.

The crosstalk-subtracted visibilities are thus given by

$$\begin{aligned}\check{V}_{pfdt}^{\text{xtalk}} &= \check{V}_{pfdt}^{\text{cal}} - \text{Med}_{t \in \mathcal{Q}_q} \left(\text{Re} \{ \check{V}_{pfdt}^{\text{cal}} \} \right) \\ &\quad - j \text{Med}_{t \in \mathcal{Q}_q} \left(\text{Im} \{ \check{V}_{pfdt}^{\text{cal}} \} \right),\end{aligned}\quad (27)$$

where $\text{Med}_{t \in \mathcal{Q}_q}(\cdot)$ denotes the weighted median evaluated over times t within the reference range $\mathcal{Q}_q$ appropriate for quarter q . This procedure ensures that all visibilities within a quarter are referenced to the same local ERA range.

If less than 30% of the time samples in the reference range $\mathcal{Q}_q$ are marked as good for a given frequency and polarization, the crosstalk estimate is considered unreliable, and the corresponding data are masked entirely. This is implemented by updating the good data flag as

$$G_{pft}^{(3)} = G_{pft}^{(2)} \times \Theta \left(\sum_{t \in \mathcal{Q}_q} G_{pft}^{(2)} - 0.3 \sum_{t \in \mathcal{Q}_q} 1 \right), \quad (28)$$

where $\Theta(u)$ is the Heaviside step function, equal to 1 when $u > 0$ and 0 otherwise.

3.5.3. Sidereal Rebinning

We rebin the hybrid-beamformed visibilities onto a fixed grid in ϕ (local ERA). We refer to the interval between successive zero-crossings of ϕ as a sidereal day. The rebinning is accomplished with a simple fractional-overlap averaging procedure. This approach was motivated by issues we encountered when using the Wiener filter-based regridding from CHIME Collaboration et al. (2023), which has a tendency to introduce baseline-dependent artifacts at the edges of RFI-masked regions and is sensitive to the choice of sky covariance. The rebinning method avoids both issues. However, because the RFI mask is frequency dependent, the number of native time samples contributing to a given ϕ -bin

varies with frequency. Since the sky signal evolves across the width of a ϕ bin due to Earth’s rotation, this frequency dependence converts intra-bin temporal variation into small-scale spectral structure, which leaks power to high delays upon Fourier transforming along frequency. We therefore apply a foreground filter prior to rebinning to suppress the slowly evolving component within each bin, as discussed in Section 3.2.3.

Let t index the raw time samples of width $\Delta_t = 9.9404$ s, r index the local ERA grid samples, and s index the sidereal day. For each time sample t , we define the integration as $[\phi_t^-, \phi_t^+]$, where $\phi_t^\pm$ are determined from the known mapping between time and local ERA. For each grid point r , we define the corresponding local ERA bin interval $[\phi_r^-, \phi_r^+]$ with $\phi_r^\pm \equiv \phi_r \pm \frac{\Delta_\phi}{2}$. The grid spacing is $\Delta_\phi = 5.27'$, a factor of roughly 2 larger than the raw integration length, yielding $N_\phi = 4096$ grid points over 360° .

The overlap length between the time sample interval t and the local ERA bin interval r is

$$\ell_{rt} \equiv \max \left[0, \min \{ \phi_t^+, \phi_r^+ \} - \max \{ \phi_t^-, \phi_r^- \} \right]. \quad (29)$$

We then define the fractional-overlap kernel

$$K_{rt} \equiv \frac{\ell_{rt}}{\phi_t^+ - \phi_t^-}, \quad 0 \leq K_{rt} \leq 1, \quad (30)$$

which equals 1 if the time interval is fully contained within the local ERA bin, 0 if there is no overlap, and otherwise the appropriate fractional value.

The effective number of valid raw time samples in each local ERA bin is

$$\mathcal{N}_{pfrs} = \sum_{t \in T_s} K_{rt} G_{pft}^{(3)}, \quad (31)$$

where $G_{pft}^{(3)}$ is the flag indicating good (1) or bad (0) data, and T_s denotes the set of all raw time samples whose integration interval overlaps with the local ERA interval corresponding to sidereal day s . For bins with $\mathcal{N}_{pfrs} = 0$, all rebinned quantities defined below are set to zero.

The rebinned visibilities are computed as

$$\check{V}_{pfdrs}^{\text{rebin}} = \frac{1}{\mathcal{N}_{pfrs}} \sum_{t \in T_s} K_{rt} G_{pft}^{(3)} \check{V}_{pfdt}^{\text{xtalk}}. \quad (32)$$

The filter is propagated through the same linear operation,

$$H_{rsff'}^{\text{rebin}} = \frac{1}{\mathcal{N}_{pfrs}} \sum_{t \in T_s} K_{rt} G_{pft}^{(3)} H_{tff'}^{\text{rebin}}, \quad (33)$$

as is the noise covariance,

$$\Sigma_{persff'}^{\text{rebin}} = \frac{1}{\mathcal{N}_{pfrs} \mathcal{N}_{pf'rs}} \sum_{t \in T_s} K_{rt}^2 G_{pft}^{(3)} G_{pf't}^{(3)} \Sigma_{petff'}, \quad (34)$$

which is assumed to be diagonal in r .

3.5.4. Sidereal Day Averaging

Finally, we average the hybrid-beamformed visibilities over 94 sidereal days acquired between 2019 January 1 and 2019 November 5 (see Section 2.4). Each visibility is weighted by $\mathcal{N}$, the number of raw integrations contributing to that local ERA bin, corresponding to an effectively *uniform weighting*

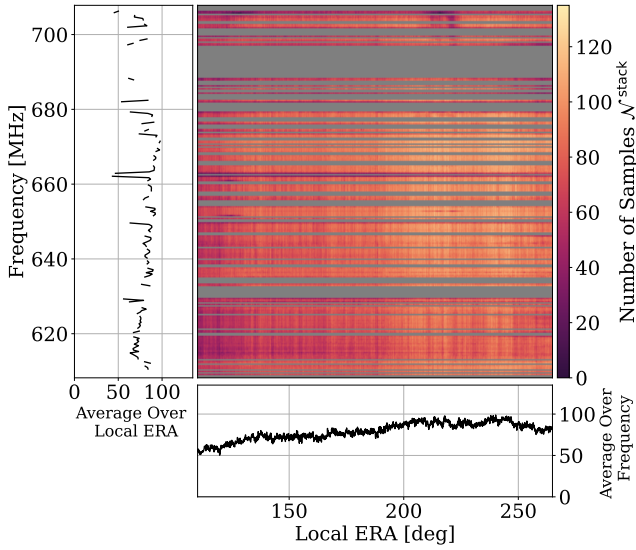


Figure 4. Number of 9.9404 s integrations contributing to each local ERA—frequency bin in the stack over all 94 days of data. The center panel shows the 2D distribution as a function of local ERA and frequency, restricted to the local ERA range defining the field used in this work. Frequency channels that are entirely masked are shown in gray. The left panel shows the average over local ERA, while the bottom panel shows the average over frequency, excluding masked channels from the average. Summing the bottom panel over local ERA bins yields a total integration time of 385 hr on this field.

scheme. As discussed in Section 3.2.3, the data have been foreground filtered on a per-day basis. The filter will strongly attenuate frequencies near large gaps of missing frequencies in order to suppress foregrounds to the required level. This attenuation affects both the 21 cm signal and the noise, but an inverse-variance weighting scheme would only reflect the noise suppression, effectively upweighting these heavily attenuated data points. Uniform weighting avoids this imbalance.

Hybrid-beamformed visibilities are stacked over sidereal days via

$$\check{V}_{pfedr}^{\text{stack}} = \frac{1}{\mathcal{N}_{pfr}^{\text{stack}}} \sum_s \mathcal{N}_{pfrs} \check{V}_{pfedr}^{\text{rebin}} \quad (35)$$

where s indexes sidereal days, and the sum runs over all days that pass the data quality tests described in Section 3.3.1 of CHIME Collaboration et al. (2023). Here $\mathcal{N}^{\text{stack}}$ is the total number of raw integrations and is given by

$$\mathcal{N}_{pfr}^{\text{stack}} = \sum_s \mathcal{N}_{pfrs}. \quad (36)$$

This quantity is shown in Figure 4. The same procedure is applied to the filter

$$H_{rff'}^{\text{stack}} = \frac{1}{\mathcal{N}_{pfr}^{\text{stack}}} \sum_s \mathcal{N}_{pfrs} H_{rff'}^{\text{rebin}} \quad (37)$$

and to the noise covariance

$$\Sigma_{perff'}^{\text{stack}} = \frac{1}{\mathcal{N}_{pfr}^{\text{stack}} \mathcal{N}_{pfr'}^{\text{stack}}} \sum_s \mathcal{N}_{pfrs} \mathcal{N}_{pfr's} \Sigma_{perff'}^{\text{rebin}}. \quad (38)$$

In the staged averaging described below, these expressions are applied repeatedly to the appropriate subset of days.

To construct the final sidereal day average, we apply the stacking rules in Equation (35)–(38) in several stages. First,

the valid sidereal days are sorted by date, and within each quarter of the year, divided into two partitions that we label “even” and “odd” by assigning consecutive days alternately to each partition. This procedure yields two stacks with approximately equal numbers of days while keeping the temporal separation between days in each stack small. Within each partition and quarter, visibilities are then averaged over sidereal days using Equation (35). Next, the quarterly averages for the first, second, and third quarters are aligned to a common reference by subtracting the median value over Q_2 (as described in Section 3.5.2). These three quarterly averages are then combined into a single stack, which we reference to Q_4 before averaging with the fourth-quarter stack. This staged referencing ensures that all visibilities are placed on a common sky-relative reference while maximizing the amount of data available to determine that reference level. Frequencies for which more than 30% of the time samples in the relevant reference interval are missing are masked entirely. The end result is a sidereal day stack for each of the even and odd partitions.

Figure 5 presents the stacked hybrid-beamformed visibility $\check{V}_{pfedr}^{\text{stack}}$ for a representative polarization and EW baseline, evaluated at the decl. nearest to 3C 295, the brightest source in our field. We show both the full sidereal day average (middle panel) and the difference of the even and odd partitions (bottom panel). Given CHIME’s night-to-night gain stability at the $\sim 1\%$ level (CHIME Collaboration et al. 2022a), foreground power should remain coherent across days and therefore cancel in the difference, leaving predominantly thermal noise. The close agreement in morphology between the average and even-minus-odd maps, together with their comparable variances (within 5%), indicates that any residual foreground leakage in this data product is subdominant to the noise after per-day foreground filtering and bandpass correction—even in the vicinity of the brightest source.

3.6. Mapmaking

The next stage of the pipeline constructs sky maps from roughly year-long stacks of hybrid-beamformed visibilities for the even and odd partitions. Since the NS direction has already been beamformed, mapmaking requires only the combination of measurements from the different EW cylinder separations. We perform this combination in the m -mode domain (J. R. Shaw et al. 2014, 2015)—the Fourier conjugate of local ERA—where each EW baseline is sensitive to a narrow, largely disjoint set of azimuthal modes m . Our procedure follows the general framework described in Section 4.3 of CHIME Collaboration et al. (2023), with several modifications detailed below.

We first mask frequency channels without full sidereal day coverage for both polarizations to avoid leakage in the subsequent Fourier transform over local ERA due to missing data. This is achieved by constructing a final good frequency flag,

$$G_f^{\text{stack}} = \prod_{p,r} \Theta(\mathcal{N}_{pfr}^{\text{stack}}), \quad (39)$$

which equals 1 if and only if every (p, r) bin at frequency f contains at least one unflagged sample in the stack, and zero otherwise. Frequency channels with $G_f^{\text{stack}} = 0$ are excluded from the subsequent analysis.

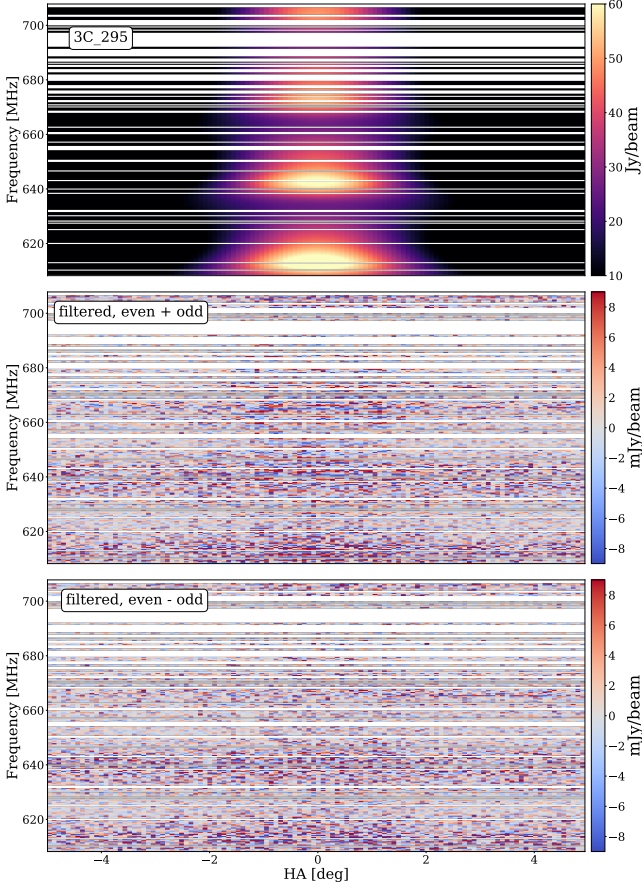


Figure 5. Top: hybrid-beamformed visibility as a function of frequency and time for the XX polarization and 22 m EW baseline, evaluated at the telescope y coordinate closest to that of 3C 295 (the brightest radio source in the field considered in this analysis) at transit. The time axis is expressed in degrees of 3C 295 hr angle, and the color scale shows the real component of the visibility after fringing. The top panel shows the average of unfiltered data over the full 94 day dataset, with features corresponding to the intrinsic spectrum of 3C 295 modulated by the primary beam response. Middle: same quantity after applying the HyFoReS bandpass correction and a time-dependent high-pass filter along the frequency axis to the raw time series prior to averaging over sidereal days. This is the primary dataset used in our analysis, in which the foregrounds are suppressed by a factor of $\gtrsim 10^4$. Bottom: difference between the even and odd sidereal day stacks, which cancels foreground power that is stable across sidereal days. The close agreement with the middle panel demonstrates that foreground leakage is subdominant compared to thermal noise at this stage of the data processing. The frequency dependence of the noise level, including the prominent ~ 30 MHz ripple, reflects the primary beam response at the transit of Cygnus A, which is imprinted onto the data through the complex gain calibration procedure.

For each polarization p , frequency channel f , EW baseline index e , decl. grid point d , and Fourier mode m , we take a discrete Fourier transform over local ERA of both the stacked hybrid-beamformed visibilities and the beam transfer function:

$$\check{V}_{pfedm}^{m\text{-mode}} = \sum_r \check{V}_{pfedr}^{\text{stack}} e^{-jm\phi_r}, \quad (40)$$

$$B_{pfedm}^{m\text{-mode}} = \sum_r B_{pfe}(\theta_d, \phi_r) e^{-jm\phi_r}, \quad (41)$$

where ϕ_r refers to the local ERA grid points and $B_{pfe}(\theta_d, \phi_r)$ is our model for the beam transfer function (see Equation (21)), evaluated at declinations and hour angles given by the (decl., local ERA) grid of the hybrid-beamformed visibilities.

The hybrid-beamformed visibilities from different EW baselines are then combined in m -mode space to form maps:

$$M_{pfdm}^{m\text{-mode}} = \sum_e W_e^{\text{EW}} (B_{pfedm}^{m\text{-mode}})^* \check{V}_{pfedm}^{m\text{-mode}}, \quad (42)$$

where $W_e^{\text{EW}} = [0.0, 0.5, 0.333, 0.166]$ for $e = [0, 1, 2, 3]$, which masks intracylinder baselines and weights the inter-cylinder baselines according to their redundancy. The intracylinder (i.e., $e = 0$) baselines are excluded in this analysis because they exhibit both enhanced noise crosstalk and a much stronger response to diffuse Galactic emission than the inter-cylinder baselines. Masking them results in maps in which the foregrounds are dominated by compact extragalactic sources, which are substantially easier to model and mask in the subsequent power spectrum analysis. A more comprehensive treatment of these short baselines is deferred to later iterations of the analysis.

Each EW baseline samples a largely disjoint set of m -modes, with the central mode determined by the baseline length and the width of the sensitive band set by the diameter of the cylinder. Equation (42) combines the baselines in the m -mode domain by multiplying each baseline's visibility by the complex conjugate of its beam transfer function $(B_{pfedm}^{m\text{-mode}})^*$ and weighting by the redundancy factor W_e^{EW} . The multiplication by $(B_{pfedm}^{m\text{-mode}})^*$ applies a matched filter using the known instrumental response and automatically retains only the m -modes to which that baseline is sensitive (setting insensitive modes to zero). The additional redundancy-based weights W_e^{EW} approximate optimal inverse-variance weighting across baseline types. Together, these operations yield maps with the full m -mode coverage of the array and close-to-optimal thermal noise, substantially lower than that of the individual hybrid-beamformed visibilities.

The resulting m -mode maps are transformed back to local ERA via

$$M_{pfdr} = \frac{1}{a_{pf}} \sum_m M_{pfdm}^{m\text{-mode}} e^{jm\phi_r}, \quad (43)$$

where the polarization- and frequency-dependent normalization factor

$$a_{pf} = \sum_m \sum_e W_e^{\text{EW}} |B_{pfedm}^{m\text{-mode}}|^2 \quad (44)$$

ensures that a point source located at the reference decl. d_{ref} has a peak value equal to its flux density. We adopt the decl. of Cygnus A as the reference decl., consistent with the normalization convention used throughout the pipeline.

The synthesized beam along the ϕ -direction is given by

$$B_{\text{synth}, pfd}^{\text{EW}}(\phi - \phi_r) = \frac{1}{a_{pf}} \sum_m \sum_e W_e^{\text{EW}} |B_{pfedm}^{m\text{-mode}}|^2 \times e^{jm(\phi - \phi_r)}. \quad (45)$$

The sky signal is attenuated by the primary beam during observation and is then multiplied by the known beam model in the matched-filtering step (Equation (42)). Consequently, the reconstructed sky brightness in our maps is modulated as a function of decl. by the square of the primary beam response on the meridian (i.e., at zero hour angle), relative to its value at the reference decl. In CHIME Collaboration et al. (2023), we corrected for this modulation by dividing the maps by the

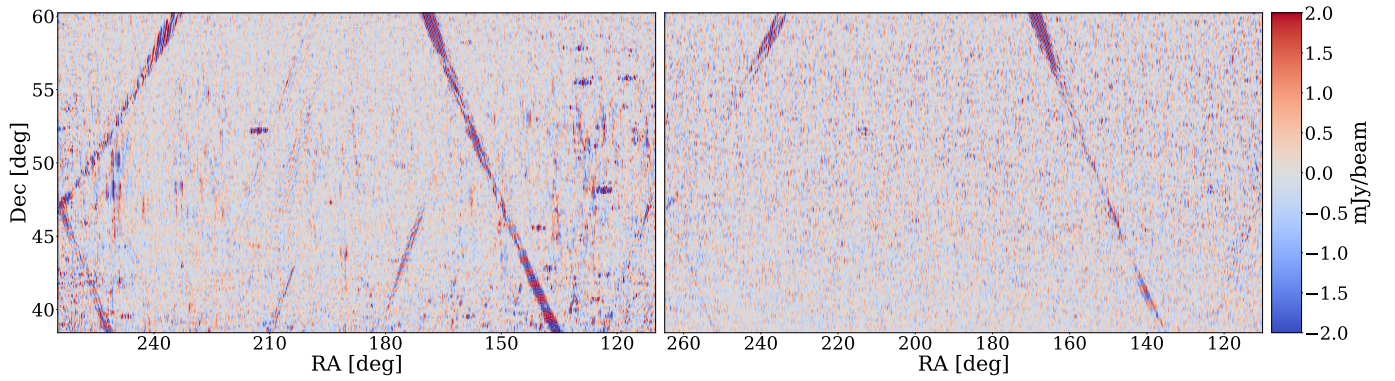


Figure 6. Foreground-filtered intensity maps near the NGC at 678.5 MHz in YY polarization. Left: map produced using the original CHIME Collaboration et al. (2023) pipeline, which formed the basis of the high signal-to-noise cross-correlation detection with eBOSS galaxies and quasars. Right: the same dataset reprocessed with the updated pipeline described in Section 3, showing a significant reduction in artifacts due to RFI and foreground leakage, which yields a cleaner input for auto power spectrum analysis. Bright continuum sources in the far sidelobes still produce high-delay signal that passes through the foreground filter, appearing as characteristic “U-shaped” tracks that are masked for this analysis (see Section 5.2).

square of this meridian beam model, thereby normalizing the response across decl. This normalization reduced beam-induced spectral structure in the data and allowed us to use a less aggressive foreground filter. In this analysis, we do not implement this correction for two reasons. First, when we later Fourier transform along the decl. axis during power spectrum estimation, the meridian beam response provides desirable downweighting of declinations where the primary beam response is low. Optimal power spectrum estimation requires weighting by the beam response and the inverse noise variance, as is done here. Second, correcting for the beam response at this stage is unnecessary because foreground filtering has already been applied and we will forward model the effect of the beam on the 21 cm signal. This second reason also allows us to omit the step of deconvolving the EW beam response along local ERA, which was done in CHIME Collaboration et al. (2023).

We employ an updated primary beam model (rev_03) relative to the model used in CHIME Collaboration et al. (2023) (rev_02) when calculating the beam transfer function. Construction of this model follows the procedure described in Appendix B of CHIME Collaboration et al. (2023), with several improvements. We applied enhanced RFI excision and a revised NS beamforming window function when forming the dataset used to fit the sky model. We then fit for a beam model over the full CHIME band, enabling better correction of flux errors in the input sky model.

For the remainder of this work, we restrict our analysis to a region of sky near the North Galactic Cap (NGC), which transits close to zenith and lies at high Galactic latitude, far from the Galactic plane. Specifically, we examine a field with $N_\phi = 1763$ local ERA bins spanning $\phi \in [110^\circ, 265^\circ]$ and $N_y = 231$ telescope y bins spanning $y \in [-0.19, 0.19]$, corresponding to declinations $\theta \in [38.4, 60.2]$. The survey area of this field is

$$A_s \simeq 0.67 \text{ sr} \simeq 2.2 \times 10^3 \text{ deg}^2. \quad (46)$$

Figure 6 shows an example of the resulting map at 678.5 MHz in YY polarization for this field. Both panels are derived from the same underlying dataset, frequency channel, and polarization; the difference lies solely in the analysis pipeline. The left panel uses the processing from CHIME Collaboration et al. (2023), which yielded a high signal-to-noise detection of 21 cm emission in cross correlation with the

eBOSS ELG, LRG, and QSO samples, while the right panel shows the same data reprocessed with the improved pipeline described in this work. The measured noise in a single frequency channel is on average $0.5 \text{ mJy beam}^{-1}$. The updated map exhibits markedly fewer artifacts from RFI and foreground leakage, providing a substantially cleaner dataset for subsequent auto power spectrum analysis.

3.7. Forming (Pseudo) Stokes I and Q

We combine the linearly polarized maps derived from the co-polar visibilities ($p \in \{XX, YY\}$) to form pseudo-Stokes maps ($q \in \{I, Q\}$). For each frequency channel f , decl. grid point d , and local ERA sample r ,

$$M_{qfdr}^{\text{Stokes}} = \sum_p T_{qp} M_{pfd r}, \quad (47)$$

with

$$T \equiv \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \quad (48)$$

We compute only I and Q (not U or V). These are *pseudo*-Stokes maps because no primary beam deconvolution is applied; they are beam-weighted combinations of the co-polar maps. Note that Stokes maps are masked at frequency f if either XX and YY are masked at that f (see Equation (39)).

3.8. Detection and Masking of Narrowband Absorption Features

We have identified numerous narrowband absorption-like features in our data that are stable across sidereal days and localized in both sky position and frequency. These features are candidate 21 cm absorbers that require further confirmation, though the brightest corresponds to a previously known 21 cm absorber system. A detailed methodology for the identification and characterization of these systems across the full CHIME band, along with their population statistics, will be presented in a forthcoming publication (CHIME Collaboration et al. 2026, in preparation). Here we briefly outline the procedure used to identify and mask these features for the power spectrum analysis.

We generate a foreground-filtered map from the full 94 sidereal day stack, and in parallel, a jackknife map by differencing sidereal day stacks formed from the even and odd

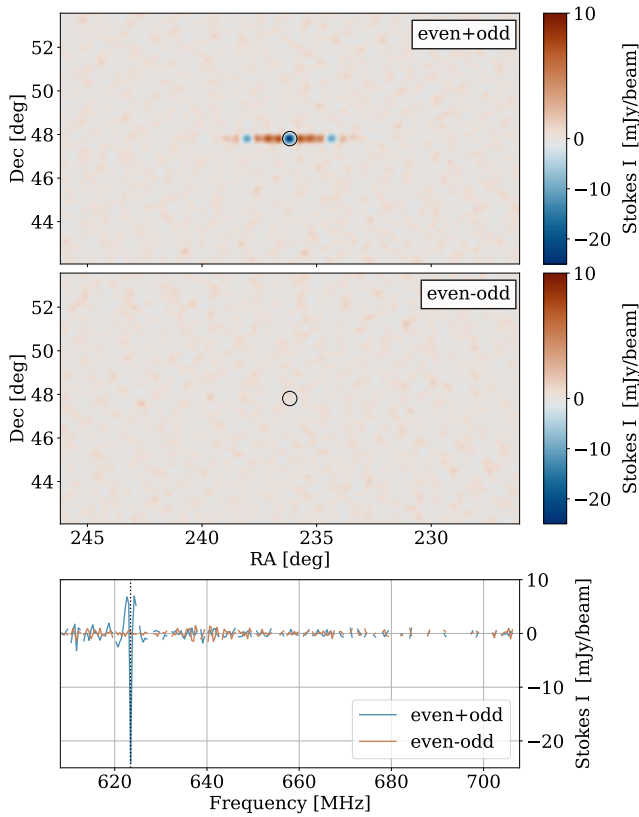


Figure 7. CHIME detection of the known 21 cm absorption system B1543+480 at $z = 1.277$ (S. J. Curran et al. 2013). Top: foreground-filtered Stokes I map at 623.4 MHz in the vicinity of B1543+480. The source is clearly seen across the full extent of the CHIME primary beam in R.A. In the R.A. direction, the point spread function response to this source shows a strong negative excursion at the central lobe; immediately outside it, symmetric positive shoulders caused by masked intracylinder baselines; and, farther out, symmetric negative responses at the grating lobes. A diverging color scale with unequal slopes for positive and negative values is used to enhance contrast. Middle: foreground-filtered Stokes I even-odd jackknife map, obtained by differencing the sidereal day stacks from the even and odd partitions. The absorption signal disappears in the difference, indicating its presence in both partitions and supporting a celestial origin. Bottom: spectrum at the map pixel nearest the source location, showing the full sidereal day stack and the even-odd jackknife. A narrow absorption line is evident in the stack but not in the jackknife. Positive shoulders bracketing the line arise from high-pass filtering along frequency to remove smooth foregrounds.

partitions. The voxels in both the combined and jackknife maps are then normalized by the per-voxel standard deviation derived from the propagated fast-cadence noise estimate (see Appendix A.1). If the maps contained only noise, the distribution of voxel values in both of these 3D maps would be Gaussian with unit standard deviation. Instead, we find many outliers that appear in the combined map but not in the jackknife, indicating that these signals are stable in sidereal coordinates and likely associated with celestial sources.

Manual inspection of these outliers shows that they are localized in the sky, appear as point sources in the spatial domain, and exhibit narrowband absorption-like features in their frequency spectra (see Figure 7 for an example). In the observed region of the NGC between 608 and 708 MHz, we identify 40 such candidates with an $S/N > 7$. Note that we see evidence of bright foreground sources leaking through far sidelobes of the primary beam and contaminating the NGC region (see Figure 6). To mitigate this, we apply a spatial mask as described in Section 5.2, which removes approximately

33% of the sky area. All 40 candidates are detected in the remaining unmasked regions. Future improvements in primary beam modeling and foreground filtering will enable searches across substantially larger sky areas without requiring such aggressive spatial masking. Crossmatching the 3D positions of our candidates with existing catalogs of known absorber systems, we find one match (shown in Figure 7) with the known 21 cm absorption system B1543+480 at $z = 1.277$, previously reported in S. J. Curran et al. (2013), which provides confidence in our detection methodology. The remaining 39 features are new candidate systems that will require further confirmation.

These narrowband features are typically confined to a single 390 kHz wide coarse channel of CHIME data. They also exhibit spectral sidelobes due to convolution with the foreground filter, similar to the stacked H 121 cm emission signal as shown in CHIME Collaboration et al. (2023). For each candidate absorber s , we generate a 3D taper $T_{\text{abs}}^s(\nu, y, \phi)$ that accounts for both spatial location and spectral position, masking 13 channels on either side of the peak absorption channel to cover spectral sidelobes. We apply this taper to the map, which zeros 1.35% of the voxels in our data cube. The details of the construction of the taper are described in Appendix B. We also update the global mask to reflect any zeroed data,

$$G_{fdr}^{\text{abs}} = G_f^{\text{stack}} \times \Theta \left(\prod_s T_{\text{abs}}^s(\nu_f, y_d, \phi_r) \right), \quad (49)$$

where the product runs over all candidate absorbers s , so that G_{fdr}^{abs} vanishes exactly where the combined absorber taper is zero.

The effect of masking these candidates on the distribution of normalized voxel values is shown in Figure 8. Note that we have already applied the spatial mask for point-source foregrounds to the data (see Section 5.2); here we study the effect of additionally masking the narrowband features. Before masking (blue histogram), the combined map shows a substantial excess of voxels at high S/N due to these features. After masking (red histogram), the distribution becomes consistent with the jackknife map (green histogram), both showing approximately Gaussian profiles centered at an $S/N = 0$. We note that both the masked combined map and the jackknife map have a distribution of voxels that is 10%–15% broader than the theoretical Gaussian with unit standard deviation, which is expected because our noise model does not incorporate sky temperature variations (see Section 5.4). The close agreement between the masked map and the jackknife confirms that our masking procedure effectively removes these narrowband features while preserving the noise properties of the data.

4. Stacking on the eBOSS QSO Catalog

While the improvements to the offline pipeline described in Section 3 successfully remove many systematics present in our earlier processing, it is essential to verify that the new pipeline does not inadvertently suppress or remove the cosmological 21 cm signal. To test this, we repeat the cross-correlation (stacking) analysis of CHIME Collaboration et al. (2023) using the newly processed dataset that will later be used for the auto power spectrum measurement. This allows us to quantify any signal loss introduced by the updated pipeline.

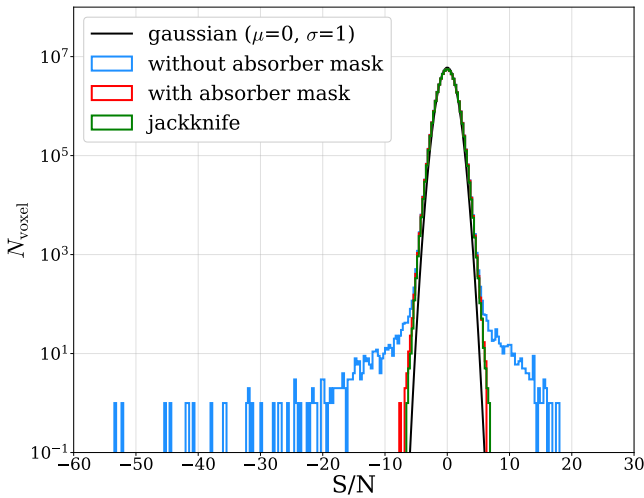


Figure 8. Histogram of voxel values in the foreground-filtered map after normalizing by the per-voxel standard deviation derived from the fast-cadence noise estimate. The blue histogram shows the distribution before masking H I absorbers, revealing pronounced negative tails (down to an $S/N \approx -60$) from absorption systems and corresponding positive tails (up to an $S/N \approx 20$) produced by the high-pass filtering applied both spectrally and spatially (i.e., through the exclusion of intracylinder baselines), which generates adjacent positive excursions around strong negative features. The red histogram shows the distribution after masking absorbers, which is consistent with the jackknife map (green), obtained by differencing the even and odd sidereal day stacks. A Gaussian with mean $\mu = 0$ and standard deviation $\sigma = 1$ is shown for comparison. Both the masked and jackknife distributions exhibit a 10%–15% excess width relative to the Gaussian, as expected because the noise estimate does not account for sky temperature variations. The agreement between the masked map and jackknife validates the absorber identification and masking procedure.

In CHIME Collaboration et al. (2023), we reported the detection of the cosmological 21 cm emission by stacking CHIME’s foreground-filtered maps on the angular and spectral positions of LRGs, ELGs, and QSOs from the eBOSS (K. S. Dawson et al. 2016) clustering catalog, achieving detection significances ranging from 7σ to 11σ depending on the tracer. In the current analysis, we first apply a spatial mask, as described in Section 5.2, for consistency with the autocorrelation analysis, and perform stacking on the QSO sample (B. W. Lyke et al. 2020; A. J. Ross et al. 2020), which provides the highest source density (33,119 objects) within our redshift range ($1.01 < z < 1.34$), compared to the LRG and ELG catalogs (see Figure 2 of CHIME Collaboration et al. 2023).

Figure 9 shows the stacked signal as a function of frequency offset. The black line shows the data, and the red line shows the best-fit signal model. The dark- and light-gray bands represent the 68% and 95% confidence intervals derived from stacking on 10,000 random mock catalogs. The bottom panel shows residuals after subtracting the best-fit model. The signal model, fitting procedure, and estimation of detection significance are described in CHIME Collaboration et al. (2023). Following the same procedure, we again detect the 21 cm signal in cross correlation with the eBOSS QSO sample, with a detection significance of 9.1σ .

Despite the substantial improvements in data quality, the resulting detection significance is slightly lower than that reported in CHIME Collaboration et al. (2023). This is expected for several reasons. First, we analyze a narrower frequency range (608–708 MHz) compared to the band (585–800 MHz) used previously, resulting in fewer objects

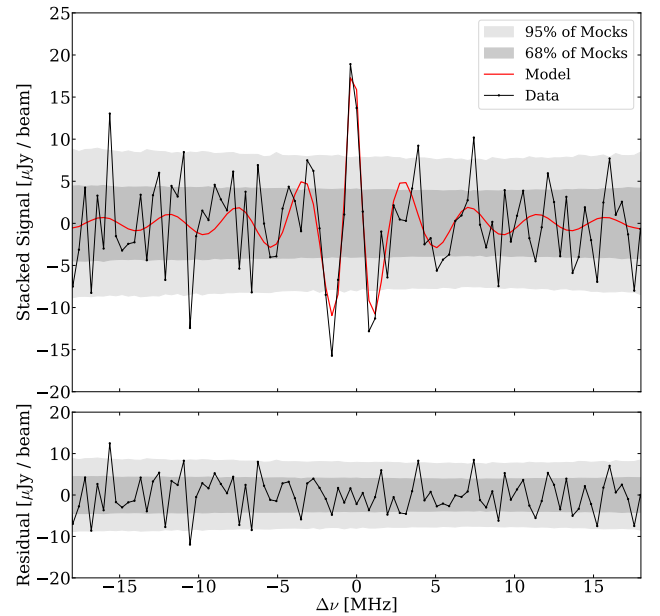


Figure 9. Top: stacked signal as a function of frequency offset for the eBOSS QSO catalog. The data are shown in black, and the best-fit model is shown in red. Bottom: residuals obtained by subtracting the best-fit model from the data. In both panels, the dark- and light-gray bands indicate the central 68% and 95% of values, respectively, obtained by applying the same stacking procedure to 10,000 random mock catalogs.

to stack on and a smaller cosmological volume. Second, we apply a spatial mask that excludes approximately 33% of the sky area (see Section 5.2), which was not applied in the previous analysis. Third, the foreground filtering differs between the two analyses: the previous analysis employed a decl.- and polarization-dependent delay cutoff ranging from ~ 120 to 180 ns in the decl. range analyzed here (see Figure 10 of CHIME Collaboration et al. 2023), while the current analysis uses a uniform 200 ns cutoff, resulting in greater signal attenuation.

To quantify the impact of these changes, we reanalyzed the dataset of CHIME Collaboration et al. (2023) using the same processing steps: applying a 200 ns delay cutoff for foreground filtering, selecting the frequency range of 608–708 MHz, applying the absorber mask and spatial masks of bright point-source foregrounds, and then performing stacking on the QSO catalog. This yields a detection significance of 8σ , compared to 9.1σ achieved with the newly processed data using identical analysis methods. This improvement demonstrates the enhanced sensitivity achieved by the current pipeline despite the more conservative foreground filtering.

The successful recovery of the 21 cm signal in cross correlation with the updated dataset demonstrates that the improved processing pipeline preserves the cosmological signal while removing systematic contamination. This provides confirmation that the pipeline does not introduce significant signal loss, giving us confidence in proceeding with the 21 cm auto power spectrum measurement presented in the following sections.

5. Power Spectrum Estimation

In this section, we describe how we estimate the power spectrum from the foreground-filtered maps.

5.1. Delay Spectrum Estimation

We first form delay spectra by Fourier transforming each map pixel along the frequency axis. Because many channels are masked and a foreground filter has been applied along frequency, a naive discrete Fourier transform that zero-fills masked samples and ignores the filter will result in additional mixing between delay modes, inflating covariance between neighboring $k_{\parallel}$ bins. Our goal is to construct delay estimates with compact, well-behaved window functions in $k_{\parallel}$ so that subsequent binning remains efficient. We therefore estimate delay spectra using a linear matched filter and inverse-covariance-weighted transform. This transform explicitly accounts for the mask and foreground filter and uses the propagated post-filter frequency–frequency covariance to weight frequency channels, thereby mitigating mode mixing and reducing the variance of the delay estimates relative to a naive transform.

Let M_{qfdr}^{Stokes} denote the foreground-filtered map, with indices q (polarization), f (frequency channel), d (decl.), and r (local ERA). In what follows, we consider each map pixel (q, d, r) independently and, to simplify notation, suppress these indices throughout this subsection. We define the frequency-domain data vector for this pixel as

$$[\mathbf{m}]_f = M_{qfdr}^{\text{Stokes}} \quad \text{for } f \in \{1, \dots, N_{\nu}\}, \quad (50)$$

where N_{ν} is the total number of frequency channels. We model this data vector as

$$\mathbf{m} = \mathbf{R} \tilde{\mathbf{m}} + \mathbf{n}, \quad (51)$$

where $\tilde{\mathbf{m}}$ denotes the underlying 21 cm signal in delay space, $\mathbf{n}$ is the (post-filter) instrumental noise in frequency space, and $\mathbf{R}$ is the forward operator that maps delay to frequency and applies the same frequency axis operations as our pipeline,

$$\mathbf{R} = \mathbf{W} \mathbf{G} \mathbf{H} \mathbf{F}. \quad (52)$$

It consists of the following components:

1. $\mathbf{F}$: delay-to-frequency Fourier operator, given by

$$[\mathbf{F}]_{fa} = e^{j2\pi\nu_f\tau_a} / \sqrt{N_{\nu}}. \quad (53)$$

2. $\mathbf{H}$: ϕ -dependent, stacked foreground filter (see Equation (37)).
3. $\mathbf{G}$: (y, ϕ)-dependent masking operator that zeros channels that are missing or masked by an earlier stage of the pipeline,

$$[\mathbf{G}]_{ff'} = G_{f'rd}^{\text{abs}} \delta_{ff'}. \quad (54)$$

4. $\mathbf{W}$: operator that selects the sub-band of interest,

$$[\mathbf{W}]_{ff'} = \begin{cases} \delta_{ff'}, & \text{if } \nu_{\text{lower}} \leq \nu_f \leq \nu_{\text{upper}}, \\ 0, & \text{otherwise.} \end{cases} \quad (55)$$

Our estimate of the delay-domain 21 cm signal for the pixel is obtained by applying a fixed linear operator to the frequency-domain data vector,

$$\hat{\mathbf{m}} = \mathbf{D} \mathbf{m}. \quad (56)$$

The operator $\mathbf{D}$ implements inverse-covariance weighting followed by the adjoint frequency-to-delay projection,

$$\mathbf{D} \equiv \mathbf{R}^{\dagger} (\mathbf{C}_{\text{reg}} + \mathbf{N})^{-1}. \quad (57)$$

Here $\mathbf{N} \equiv \langle \mathbf{nn}^{\dagger} \rangle$ is the frequency–frequency noise covariance, which can be expressed as

$$\mathbf{N} = \mathbf{W} \mathbf{G} \mathbf{\Sigma} \mathbf{G}^{\dagger} \mathbf{W}, \quad (58)$$

with $\mathbf{\Sigma}$ obtained by propagating the sidereal day averaged covariance from Equation (38) through the mapmaking procedure.

In addition to the propagated noise covariance, we include a small regularization term $\mathbf{C}_{\text{reg}}$ in the total frequency–frequency covariance entering Equation (57). This matrix is chosen to approximate the expected 21 cm covariance, calibrated from end-to-end simulations, and is small compared to the propagated noise covariance in the well-measured subspace. We have verified that the measured power spectrum is insensitive to the overall normalization of $\mathbf{C}_{\text{reg}}$ over a wide range around this calibration (see Appendix C for details).

Because fully masked channels produce zero rows and columns, $\mathbf{C}_{\text{reg}} + \mathbf{N}$ in Equation (57) is singular. We therefore perform the inversion on the unmasked frequency subspace (which is symmetric positive definite), compute the inverse via a Cholesky decomposition, and embed the result into the full matrix with zeros on masked rows and columns. This is equivalent to the Moore–Penrose pseudoinverse under masking (J. C. A. Barata & M. S. Hussein 2012).

This procedure is strictly linear in the data. In earlier work, we instead used a Gibbs sampling algorithm to infer both the delay power spectrum and the corresponding delay-transform Wiener filter from the data (CHIME Collaboration et al. 2023, Appendix A). While viable, this type of nonlinear, data-dependent filter estimation can bias power spectrum estimates if not implemented with care (Z. S. Ali et al. 2018; C. Cheng et al. 2018). Here we avoid that class of systematics by constructing the delay transform operator from the propagated noise covariance and a simple, simulation-inspired model for the 21 cm covariance.

We repeat this for every pixel (q, d, r) and assemble

$$\tilde{\mathbf{M}}_{qadr}^{\text{Stokes}} \equiv [\hat{\tilde{\mathbf{m}}}]_a, \quad (59)$$

where a indexes delay. The pixel-dependent delay-transform operators $\mathbf{D}_{qdr}$ are also saved and applied to the 21 cm and noise simulations in order to interpret the measured power spectrum.

5.2. Spatial Mask

Residual contamination in the delay-filtered maps is spatially associated with the brightest continuum sources in the sky. Because our beamforming provides good localization in the telescope y -direction and the positions of these sources are well known, we can excise the affected regions using three types of smoothly tapered spatial masks.

1. *Transit mask.* To suppress contamination from bright sources near transit, we apply an elliptical tapered mask, $T_{\text{transit}}^{(s)}(y, \phi)$, to each continuum source s in the specfind v3 catalog (Y. Stein et al. 2021) with flux density greater than 10 Jy at 600 MHz. The extent of

each mask is set by the primary beamwidth in ϕ and the synthesized beamwidth in y , both evaluated at the lowest frequency (608.2 MHz).

2. *Track mask.* We observe residual foreground contamination from the brightest sources in the sky (e.g., Cygnus A, Cassiopeia A, Taurus A, Virgo A) even when they lie in the far sidelobes of the primary beam, appearing as “U”-shaped tracks in the foreground-filtered maps (see Figure 6). This arises from sparse sampling of EW spatial modes by the array, which prevents precise localization of sources in the EW direction and allows bright off-axis sources to contribute high-delay power that survives foreground filtering. To suppress this contamination, we construct extended track masks, $T_{\text{track}}^{(s)}(y, \phi)$, for each catalog source s with flux density greater than 60 Jy at 600 MHz, following the full path of the source across the sky.
3. *Galactic plane mask.* For a specific range of local ERA in our field, we observe residual contamination in the foreground-filtered maps that is spatially associated with the track of the inner Galactic plane as it moves through the far sidelobes. This contamination is apparent at delay modes just below the threshold used in our final analysis, appearing at $\tau \lesssim 280$ ns, which is within the transition region of the delay filter (see Section 5.3). Nevertheless, to guard against possible leakage into the higher delay modes used for power spectrum estimation, we conservatively apply a spatial mask, $T_{\text{gal}}(y, \phi)$, to exclude these regions. We later verified that removing this mask has a negligible effect on the measured power spectrum.

The combined spatial mask for bright continuum foreground sources is constructed by multiplying the individual mask functions:

$$T_{\text{base}}(y, \phi) = T_{\text{gal}}(y, \phi) \times \prod_{s \in \mathcal{S}_{10}} T_{\text{transit}}^{(s)}(y, \phi) \times \prod_{s' \in \mathcal{S}_{60}} T_{\text{track}}^{(s')}(y, \phi), \quad (60)$$

where $\mathcal{S}_{10}$ and $\mathcal{S}_{60}$ denote the sets of catalog sources above the 10 and 60 Jy flux density thresholds, respectively. This mask excludes approximately 33% of the sky area. The details of the mask construction, including the mathematical formulation and parameter choices, are provided in Appendix B.

5.3. Power Spectrum Estimation

The final stage of the analysis is the estimation of the cosmological power spectrum from the delay-space map. The procedure involves transforming the data into a 3D cosmological wavenumber space, forming a cross power spectrum to remove noise bias, and then averaging the result into cylindrical and spherical bins to produce the 2D and one-dimensional (1D) power spectra.

We first define a final spatial window function by multiplying the smoothly tapered mask $T_{\text{base}}(y, \phi)$ constructed in Section 5.2 with a 2D apodization $T_{\text{edge}}(y, \phi)$ to reduce edge effects in the subsequent Fourier transform. The apodization is separable, $T_{\text{edge}}(y, \phi) = \text{Tukey}(y) \text{Tukey}(\phi)$, where Tukey is a 1D Tukey window function with cosine fraction 0.5 (i.e., a flat central region over 50% of the field with a cosine roll-off to

zero at the edges):

$$T(y, \phi) = T_{\text{base}}(y, \phi) \times T_{\text{edge}}(y, \phi). \quad (61)$$

We apply the same window function identically to all delay modes of the map:

$$\tilde{M}_{\text{qadr}}^{\text{windowed}} = T(y_d, \phi_r) \tilde{M}_{\text{qadr}}^{\text{Stokes}}. \quad (62)$$

We then transform the map from its native coordinates (ϕ, y) to spatial Fourier modes (u, v) by applying a discrete 2D Fourier transform to the map for each delay mode τ_a :

$$\tilde{V}_q(u, v, \tau_a) = \frac{1}{N_y N_\phi} \sum_d \sum_r \tilde{M}_{\text{qadr}}^{\text{windowed}} e^{j2\pi(u \Delta x_r + v \Delta y_d)}, \quad (63)$$

where

$$\Delta x_r \equiv \cos \bar{\theta} (\phi_r - \bar{\phi}), \quad \Delta y_d \equiv y_d - \bar{y}$$

are local flat-sky Cartesian coordinates, defined as offsets from the center of the field $(\bar{\phi}, \bar{y})$. The factor $\cos \bar{\theta}$ uses the average decl. of the field to convert offsets in local ERA into approximate angular distances in the EW direction for the entire field. We denote the result by $\tilde{V}$ to emphasize that the 2D Fourier transform of the delay-space map returns us to a visibility-like representation in delay space.

The discrete Fourier transform in Equation (63) invokes the flat-sky approximation, which is valid for our analysis because we exclude the intracylinder baselines (as mentioned in Section 3.6) (K. K. Datta et al. 2007; M. W. Eastwood et al. 2019). This exclusion removes sensitivity to large angular scales probed by the shortest pure NS baselines, restricting our measurements to $k_\perp$ modes beyond 0.08 h Mpc^{-1} , or $\ell > 200$ (see Figure 12), where the curvature of the sky is negligible over the relevant field of view.

The instrumental coordinates (u, v, τ) are then converted to comoving cosmological wavenumbers $\mathbf{k} = (k_x, k_y, k_\parallel)$ using the standard relations for the central redshift of our observation as (M. F. Morales & J. Hewitt 2004; A. Liu & J. R. Shaw 2020)

$$k_x = \frac{2\pi u}{D_c(z)}, \quad k_y = \frac{2\pi v}{D_c(z)}, \quad k_\parallel = \frac{2\pi \nu_{21} H_0 E(z)}{c(1+z)^2} \tau, \quad (64)$$

where $\nu_{21} = 1420.40575 \text{ MHz}$ is the rest-frame frequency of the 21 cm line, $D_c(z)$ is the comoving transverse distance, $E(z) = \sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda + \Omega_k(1+z)^2}$, z corresponds to the central redshift of the band, and H_0 is the Hubble constant (D. W. Hogg 1999).

As described in Section 3.5.4, to remove noise bias, we split the 94 sidereal days into two interleaved partitions, labeled “even” and “odd.” Each partition is processed independently through an identical analysis pipeline. The power spectrum is then computed by cross-correlating the resulting data cubes. Since the noise is uncorrelated between different sidereal days, this produces an estimate of the power spectrum without any noise bias. The 3D power spectrum is then given by (F. G. Mertens et al. 2020)

$$\hat{P}_q(k_x, k_y, k_\parallel) = \mathcal{V}_s \tilde{V}_q^{\text{even}}(k_x, k_y, k_\parallel) \tilde{V}_q^{\text{odd},*}(k_x, k_y, k_\parallel), \quad (65)$$

where $\mathcal{V}_s$ is the observed comoving volume of the data cube in units of $h^{-3} \text{Mpc}^3$, given by

$$\mathcal{V}_s = \frac{\mathcal{A}_s D_c^2(z) N_\nu \Delta_\nu \Delta D}{f_{\text{window}}}. \quad (66)$$

Here $\mathcal{A}_s$ is the survey area defined in Equation (46), $N_\nu \Delta_\nu$ is the total bandwidth of the observation, and $\Delta D = \frac{c(1+z)^2}{H(z)\nu_{21}}$ converts that bandwidth into comoving line-of-sight depth. The factor $f_{\text{window}} = \frac{(\sum_{dr} T_{dr})^2}{\sum_{dr} T_{dr}^2}$ accounts for the reduction in survey area due to the window function.

Assuming isotropic and translationally invariant signal statistics, we average the 3D power spectrum into cylindrical and spherical bins. The cylindrically averaged 2D power spectrum $\hat{P}_q(k_\perp, k_\parallel)$ is obtained by averaging $\hat{P}_q(k_x, k_y, k_\parallel)$ in annular bins of constant perpendicular wavenumber $k_\perp = \sqrt{k_x^2 + k_y^2}$.

The foreground filter strongly attenuates power at delays below the cutoff, $\tau_{\text{cut}} = 200 \text{ ns}$, where spectrally smooth foregrounds are expected to dominate. At higher delays ($\tau > \tau_{\text{cut}}$), the attenuation decreases only gradually with delay, resulting in residual suppression of the cosmological signal across a broad transition region (see Section 3.5 of A. Ewall-Wice et al. 2021 for further discussion on signal attenuation). To quantify the effective attenuation outside the nominal cutoff, we compute the rms amplitude of the filter residual as a function of delay, following the methodology of A. Ewall-Wice et al. (2021). In Figure 10, we plot the rms of the filter residual, $z^\tau = \mathbf{H}\mathbf{x}^\tau$, where $\mathbf{x}^\tau$ is a complex sinusoid with delay τ and amplitude equal to unity. Here $\mathbf{H}$ is the sidereal day averaged filter described in Equation (37). The filter depends on the RFI mask and can vary with local ERA, as different local ERA bins have different flagged frequency channels. The frequency flagging fraction varies between 15% and 21% over local ERA bins. We compute the filter rms for different local ERA bins and show them in different colors in Figure 10, with the median over all local ERA bins shown in black. The filter exhibits an overall attenuation, and the scatter across local ERA bins reflects variations induced by the changing RFI mask. Crucially, this local ERA-dependent overall attenuation does not bias our power spectrum estimates because the filter is applied as a linear operator identically to the data and 21 cm simulations. Any suppression of the signal is therefore properly forward modeled, ensuring an unbiased comparison. We find that beyond $\tau \sim 280 \text{ ns}$, the median filter attenuation is less than 10% relative to the high-delay plateau as shown in Figure 10. To avoid these attenuated modes beyond the filter cutoff (τ_{cut}), we generate a conservative 2D mask $W(k_\perp, k_\parallel)$, where we exclude all delay modes $|\tau| < 280 \text{ ns}$, which corresponds to $k_{\parallel, \text{min}} = 0.35 h \text{ Mpc}^{-1}$.

Finally, we compute the spherically averaged power spectrum by averaging the masked 2D power spectrum in logarithmically spaced spherical shells of constant wavenumber magnitude, $k = |\mathbf{k}| = \sqrt{k_\perp^2 + k_\parallel^2}$, given by

$$\hat{P}_q(k_i) = \frac{\sum_j W(k_{\perp,j}, k_{\parallel,j}) \hat{P}_{2D}(k_{\perp,j}, k_{\parallel,j})}{\sum_j W(k_{\perp,j}, k_{\parallel,j})}, \quad (67)$$

where the sum is over all $(k_\perp, k_\parallel)$ bins that fall within the i th k -shell. Since the data are already optimally weighted in the mapmaking and delay-transform estimation stages, we use a

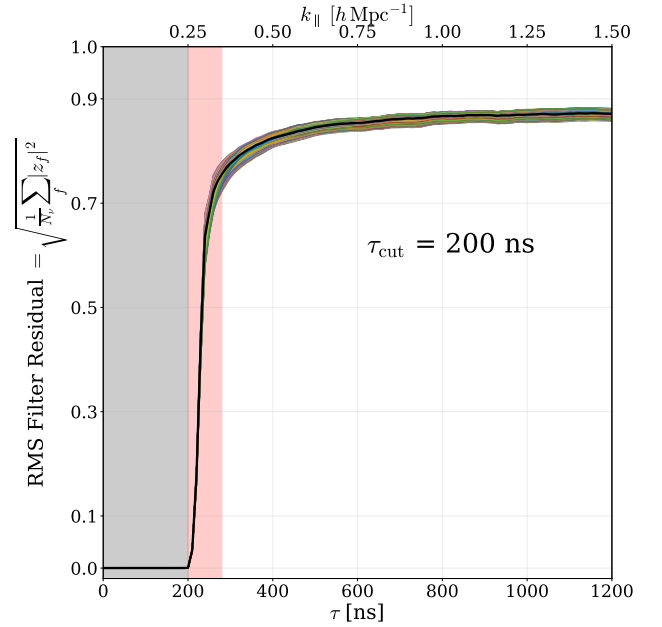


Figure 10. rms of the residual after applying the stacked foreground filter (Equation (37)) to a sinusoid of unit amplitude and delay τ . At the top, we show the $k_\parallel$ modes corresponding to the delays estimated at the central redshift of the band ($z = 1.16$). The filter depends on the RFI mask, which varies over local ERA and ranges from 15% to 21%. Different colors correspond to different local ERA bins, and the black line is the median over local ERA. There is an overall local ERA-dependent attenuation, reflecting the effect of varying the RFI mask. However, it does not bias the power spectrum estimate, as the application of the filter is a linear operation and is applied to data and 21 cm simulation identically. We see that below the filter cutoff (200 ns), the delay modes are strongly attenuated (gray shaded region) and there is a transition zone (pink shaded region) where the filter response gradually increases from near-zero and reaches a plateau at high-delay modes. Beyond 280 ns, the attenuation is less than 10% relative to the high-delay plateau. In this analysis, we exclude all delay modes below 280 ns.

uniform weight while averaging the power spectrum into spherical bins.

To convert the power spectrum estimates into physical units of $\text{K}^2 h^{-3} \text{Mpc}^3$, we employ a simulation-based approach. First, we simulate a sky map corresponding to a flat power spectrum with unit variance ($P(k) = 1 \text{ K}^2 h^{-3} \text{Mpc}^3$). From this map, we generate mock visibility data following the procedure outlined in CHIME Collaboration et al. (2023; see also J. R. Shaw et al. 2015). This simulated visibility data is processed through the same analysis pipeline used for the actual data to estimate its power spectrum. The resulting power spectrum gives the conversion factor, in both 2D ($\hat{P}_c(k_\perp, k_\parallel)$) and 1D ($\hat{P}_c(k)$). We apply this single factor to the estimated power spectra from the observed data, the cosmological 21 cm signal simulation, and the noise realization. This procedure places all quantities in consistent physical units, enabling a direct, one-to-one comparison.

5.4. Noise Covariance Estimation

Our goal in this section is to obtain an accurate estimate of the noise covariance, i.e., the covariance between the power measured in different k bins that arises from stochastic fluctuations in the data. The covariance model includes contributions from receiver noise and sky/self-noise. Residual RFI, foreground leakage, and other systematics are not included as additional covariance terms; they are mitigated

through masking and filtering, with their residual impact constrained by the validation tests presented in Section 7. The resulting noise covariance is needed to assess the detection significance of the measured power spectrum, to perform internal consistency and null tests, and to propagate uncertainties to any fitted model parameters.

At the visibility level, we have a reliable estimate of the noise variance from fast cadence, even-odd differencing (see Appendix A.1), and this variance is propagated through the subsequent linear operations in the pipeline. However, those same operations (NS beamforming, foreground filtering, sidereal rebinning and stacking, mapmaking, delay transforming, spatial masking, spatial Fourier transforming, and k -space binning) also couple samples along multiple axes, such that the noise acquires covariances in frequency, local ERA, telescope y coordinate, etc. Tracking those covariances would require storing an array with an additional axis for every data dimension, which is computationally prohibitive for our data volumes. Instead, we estimate the covariance by forward (Monte Carlo) propagation: we generate random noise realizations at the visibility level, process them using the same pipeline that was applied to the data, and take the sample covariance of the resulting power spectra. This is complicated by the fact that the daily beamforming and foreground filtering introduce covariance at the individual-day level, and it is not feasible to generate and process noise realizations for every day. Instead, we require a stack-level noise description that already incorporates the correlations produced by those daily operations.

To provide this, the pipeline propagates and outputs the frequency–frequency noise covariance of the sidereal-stacked, foreground-filtered, hybrid-beamformed visibilities, $\Sigma_{pff'}^{\text{stack}}$, where p is the polarization index, e is the EW baseline index, r is the local ERA index, and f, f' are frequency indices (see Equation (38)). This quantity already contains the correlations introduced by the daily foreground filtering, but it no longer carries an explicit NS baseline index because the beamforming step has collapsed that axis. We applied the foreground filter after beamforming so that a single frequency–frequency operator could be used for all telescope y coordinates. This both reduced the computational cost of constructing the filter and made it feasible to propagate the resulting frequency–frequency noise covariance, since it is identical for every telescope y coordinate. Consequently, Σ^{stack} tells us how the noise is correlated across frequency after beamforming, but not how that noise should be apportioned among the individual NS baselines before beamforming. This apportioning is important because the noise assigned to each NS baseline will determine the noise variance as a function of k_y .

To recover a visibility-level description, we distribute the propagated frequency–frequency noise covariance over the individual NS baselines. We assume that the noise covariance on an NS grid point n scales with the redundancy $\mathcal{R}_{pen}$ of that baseline,

$$\sigma_{pfnr}^2 \propto \frac{1}{\mathcal{R}_{pen}},$$

which captures the dominant trend seen in the data, with residual structure at the $\lesssim 10\%$ level (see Figure 1). Under this

assumption, we define a scaling factor

$$J_{pff'} \equiv \sum_n \frac{W_{fn}^{\text{NS}} W_{f'n}^{\text{NS}}}{\mathcal{R}_{pen}}, \quad (68)$$

where W_{fn}^{NS} is the beamforming window given by Equations (10) and (12). This quantity is used to construct a visibility-level covariance

$$\Sigma_{pff'}^{\text{vis}} = \frac{\Sigma_{pff'}^{\text{stack}}}{J_{pff'}}. \quad (69)$$

This is precisely the scaling factor required so that, if each baseline n is assigned noise with covariance $\Sigma_{pff'}^{\text{vis}}/\mathcal{R}_{pen}$ and we apply the same beamforming window W_{fn}^{NS} used on the data, the beamformed noise will have the measured frequency–frequency covariance $\Sigma_{pff'}^{\text{stack}}$.

The matrix $\Sigma_{per}^{\text{vis}}$ is Hermitian and positive definite, so we take its Cholesky factorization

$$\Sigma_{per}^{\text{vis}} = \mathbf{L}_{per} \mathbf{L}_{per}^T, \quad (70)$$

where $\mathbf{L}_{per}$ is lower triangular with positive diagonal entries. We generate independent, circularly symmetric complex Gaussian noise,

$$Z_{pfenr} \sim \mathcal{G}(0, 1), \quad (71)$$

and impose the desired frequency correlations and redundancy scaling via

$$V_{pfnr}^{\text{noise}} = \frac{1}{\sqrt{\mathcal{R}_{pen}}} \sum_{f'} \mathbf{L}_{pff'} Z_{pf' enr}. \quad (72)$$

Applying the NS beamforming operator (Equation (9)) to V_{pfnr}^{noise} then produces hybrid-beamformed noise visibilities whose frequency–frequency covariance matches the propagated Σ^{stack} , while retaining a baseline dependence consistent with the assumed redundancy model. For each noise realization, we construct a map, apply the absorber mask, perform the delay transform, and apply the spatial mask, using exactly the same pipeline that was applied to the data (see Section 3). This is done separately using the covariance matrix Σ^{stack} from the even and odd partition of sidereal days.

Spatially varying sky/self-noise. The construction above assumes that visibility noise is uncorrelated between baselines. While receiver noise, which originates independently in each antenna’s electronics, is indeed uncorrelated between baselines, the contribution from sky temperature is correlated, leading to covariance between different baselines. This effect, often termed “self-noise,” results in a nonuniform noise distribution in the final synthesized map, where the noise power is modulated by the sky brightness distribution (S. R. Kulkarni 1989). Consequently, when we beamform (Section 3.2.2) to a particular location in the sky, certain regions will be noisier than others depending on the local sky intensity.

To account for this spatial variation, we construct a jackknife map by differencing the data map constructed from the even and odd partitions of sidereal days. This cancels the sky signal, which is coherent across sidereal days, leaving a realization of the noise that captures the spatially dependent contribution from T_{sky} . We scale this difference by the square

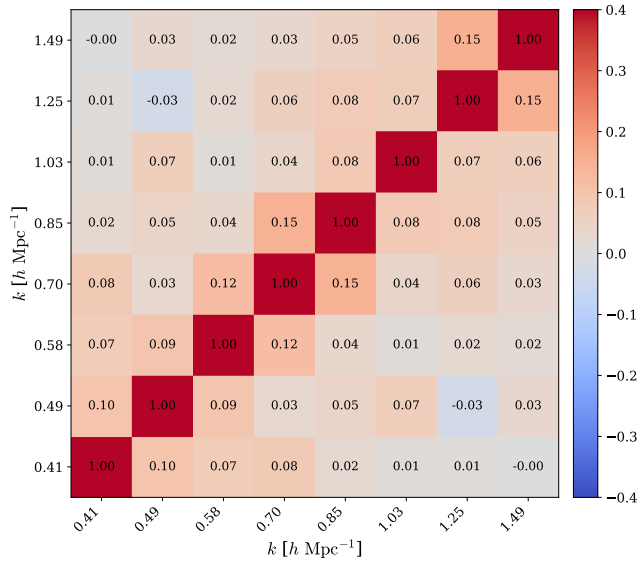


Figure 11. Correlation matrix of the power spectrum k bins, estimated from 1000 noise-only simulations. The matrix is nearly diagonal, with off-diagonal correlations largest for neighboring bins but $\lesssim 15\%$. Such correlations are expected due to the windows and masks applied by the pipeline in the map and frequency domain. These correlations are accounted for when fitting the signal model to the data (Section 6.2.2).

root of the inverse variance propagated from the fast-cadence estimate of the noise. We then compute the root-mean-square (rms) of this normalized difference map along the frequency axis to produce a 2D spatial map, $\sigma_{\text{spatial}}(\theta, \phi)$, characterizing the noise variation across the field relative to our expectation. We apply this spatial scaling factor to the delay transform of each noise realization map before computing its power spectrum. This enforces the correct large-scale variation of the noise across the field. The assumption here is that the sky noise modulation is spectrally smooth, so a single factor per sky pixel is adequate; in practice there is a residual frequency dependence in the sky noise at the $\lesssim 10\%$ level, which we ignore.

Monte Carlo estimate of the noise covariance. Next, the spatial transform of each noise realization is taken, the even and odd partitions are cross-correlated to remove the overall bias, and the resulting power spectrum is binned in an identical manner as the data. One pass through this pipeline produces one noise power spectrum (2D and 1D). We repeat this $N_{\text{samp}} = 1000$ times with independent draws of Z to build up an empirical estimate of the power spectrum noise covariance:

$$\hat{\mathbf{C}}_{\text{N}} = \frac{1}{N_{\text{samp}} - 1} \sum_{m=1}^{N_{\text{samp}}} (\mathbf{p}_m - \bar{\mathbf{p}})(\mathbf{p}_m - \bar{\mathbf{p}})^T, \quad (73)$$

where $\mathbf{p}_m$ is the vector of bandpowers for the m th realization and $\bar{\mathbf{p}}$ is their mean. Figure 11 shows the corresponding correlation matrix. It is nearly diagonal, with off-diagonal correlations $\lesssim 15\%$. These correlations arise primarily from multiplicative operations applied by the pipeline in the spectral/spatial domain (windowing, RFI masking, absorber and point-source masking), which mix Fourier modes. We use $\hat{\mathbf{C}}_{\text{N}}$ in the likelihood calculations used to fit the signal model to the measured power spectrum (Section 6.2.2) and to estimate the detection significance (Section 6.3). In subsequent figures, the error bars shown on individual bandpowers correspond to

the square root of the diagonal elements of $\hat{\mathbf{C}}_{\text{N}}$, estimated from the standard deviation of the 1000 noise power spectra.

Limitations. This procedure is not exact, for several reasons. (i) The real dependence of the noise variance on NS baseline does not follow $1/\mathcal{R}_{\text{pen}}$ perfectly, so individual $k_{\parallel}$ bins can differ from the true variance at the $\lesssim 10\%$ level even though the average trend is captured. (ii) Correlations introduced by the time-to- ϕ rebinning, where each time sample can contribute to two adjacent ϕ bins, are not explicitly modeled. (iii) Sky/self-noise is enforced via a single spatial template that is assumed constant over frequency, whereas the true sky noise has a mild frequency dependence that is not captured in our noise model. Nevertheless, the resulting noise realizations agree with the jackknife-measured noise level to better than $\lesssim 10\%$ in any domain, and by construction, capture the dominant correlations introduced by the pipeline, while remaining unbiased on average.

6. Results

Building on the data reduction pipeline described above, we now present the resulting measurement of the cosmological 21 cm power spectrum and its theoretical interpretation.

6.1. Power Spectrum Measurement

The cylindrically averaged 2D power spectrum, $P(k_{\perp}, k_{\parallel})$, is shown in Figure 12. The power spectrum is estimated from the Stokes I map over the frequency range of 608.2–707.8 MHz ($z = 1.34$ – 1.01), corresponding to a central redshift of $z \sim 1.16$. The left panel shows the real component of the power spectrum, and the right panel shows the S/N, defined as $P(k_{\perp}, k_{\parallel})/P_{\text{N}}(k_{\perp}, k_{\parallel})$. Although we do not yet have the sensitivity for detection of the 21 cm power spectrum in this $(k_{\perp}, k_{\parallel})$ domain with high S/N, which would enable us to constrain line-of-sight effects, the cylindrically binned 2D power spectrum remains an effective diagnostic tool for identifying any residual systematic contamination. Note that we employ very fine binning along the line-of-sight direction $k_{\parallel}$, without further averaging beyond the native spectral resolution ($\Delta k_{\parallel} = 0.012 \text{ h Mpc}^{-1}$). A coarser binning in $k_{\parallel}$ would reduce the noise variance and could reveal tentative evidence of the 21 cm signal in this 2D $(k_{\perp}, k_{\parallel})$ space. However, we defer such an analysis, as well as the coherent integration of multiyear observations to further improve sensitivity, to forthcoming work.

The measured power spectrum is complex-valued as we cross-correlate two data cubes with independent noise and hence can fluctuate to negative values. The power spectrum of a coherent sky signal is positive by construction and confined to the real component, leaving the imaginary component to be noise dominated. Figure 12 shows the real component of the 2D power spectrum fluctuating between positive and negative values, with the S/N varying between $\pm 3\sigma$. The 2D power spectrum shows excellent agreement with expected thermal noise and shows no evidence of any significant systematic contamination.

We further average the 2D power spectrum in spherical bins of constant wavenumber k , assuming isotropy of the cosmological signal. The spherically binned 1D power spectrum is shown in Figure 13. The top panel shows the real part of the power spectrum in red, and the black line is the best-fit model of the cosmological H I signal, as discussed in

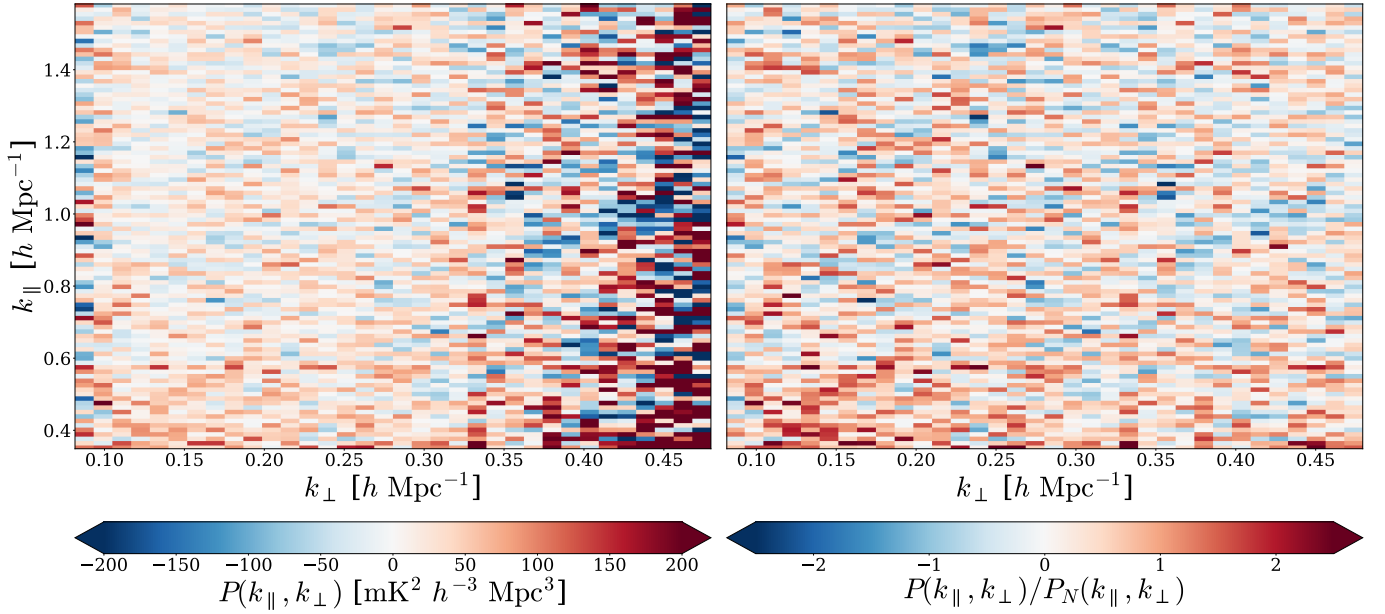


Figure 12. Left panel: the real part of the cylindrically averaged power spectrum for Stokes I data. Right panel: the power spectrum normalized by the power spectrum of expected thermal noise. The measured 2D power spectrum is consistent with the expected thermal noise and shows no evidence of significant contamination from residual foregrounds or other systematics.

Section 6.2. The bottom panel shows the result of subtracting the best-fit model from the data. We find that the residual power spectrum is consistent with noise. Since $k_{\perp}$ is limited to $<0.48 h \text{ Mpc}^{-1}$, while $k_{\parallel}$ extends to $1.58 h \text{ Mpc}^{-1}$ (see Figure 12), the spherically binned 1D power spectrum at $k > 0.4 h \text{ Mpc}^{-1}$ is primarily sensitive to the line-of-sight modes.

We also divide the full bandwidth into two independent sub-bands of approximately 50 MHz each. The upper sub-band spans 658.2–707.8 MHz ($z = 1.16$ – 1.01), centered at $z \sim 1.08$, whereas the lower sub-band spans 608.2–658.2 MHz ($z = 1.34$ – 1.16), centered at $z \sim 1.24$. Although the spatial mask remains identical to that used for the full band, the RFI masks are different. Our RFI flagging pipeline removes approximately 50% of frequency channels in the upper sub-band but only 17% in the lower sub-band. Figure 14 shows the measured power spectrum for both sub-bands. The top panel shows the real component of the power spectrum along with the best-fit model, while the bottom panel shows the residuals. Note that we generate an independent signal model for each of these sub-bands and fit them to the data. We find an excellent fit to both the sub-bands, as well as for the full frequency band. The description of the signal model and fitting procedures, including detection significance, for both the full band and the sub-bands is described below.

6.2. Theoretical Interpretation

Our companion paper (CHIME Collaboration et al. 2026a) describes our theoretical model for interpreting the measured power spectrum, along with how this model is evaluated, the constraints we have obtained on its parameters, and the information this provides about the clustering of H I. In this paper, we use the modeling approach from CHIME Collaboration et al. (2026a) to determine the best-fit power spectrum model that is plotted alongside our measurements, and to assess the significance of our detection. Therefore, in this subsection we briefly describe the relevant aspects of the

modeling, referring the reader to CHIME Collaboration et al. (2026a) for further details.

6.2.1. Power Spectrum Model

We interpret the measured power spectrum using the following model:

$$P_{21}(k, \mu; z) = \bar{T}_b(z)^2 [b_{\text{HI}}(z) + f(z)\mu^2]^2 \times P_m(k, z) D_{\text{HI}}^{\text{FoG}}(k\mu; z)^2, \quad (74)$$

where

$$P_m(k, z) = \left[\frac{D^+(z)}{D^+(z_{\text{fid}})} \right]^2 \{ P_L(k, z_{\text{fid}}) + \alpha_{\text{NL}} [P_{\text{NL}}(k, z_{\text{fid}}) - P_L(k, z_{\text{fid}})] \}. \quad (75)$$

The 21 cm power spectrum is proportional to the square of the mean brightness temperature $\bar{T}_b(z)$, which is itself proportional to the mean H I density $\Omega_{\text{HI}}(z)$. The linear H I bias $b_{\text{HI}}(z)$ relates the large-scale clustering of H I to the clustering of matter, and is added to the Kaiser term $f\mu^2$ (N. Kaiser 1987) to incorporate the leading effect of redshift-space distortions on large scales. We take $\Omega_{\text{HI}}(z_{\text{eff}})$ and $b_{\text{HI}}(z_{\text{eff}})$, with $z_{\text{eff}} = 1.16$, as free parameters in our analysis, assuming that the redshift dependence of $\Omega_{\text{HI}}(z)$ and $b_{\text{HI}}(z)$ follows the N. H. M. Crighton et al. (2015) fitting function and the IllustrisTNG $b_{\text{HI}}(z)$ fitting function from CHIME Collaboration et al. (2023), respectively.

We model the scale dependence of the power spectrum using two factors. The first is a matter power spectrum, Equation (75), that uses a single parameter α_{NL} to interpolate between the linear version and the nonlinear model from A. J. Mead et al. (2021), using the linear growth factor $D^+(z)$ to rescale a fiducial model evaluated at $z_{\text{fid}} = 1$. We use the CAMB package (A. Lewis et al. 2000) to compute these matter power spectra. The second is the square of a Lorentzian

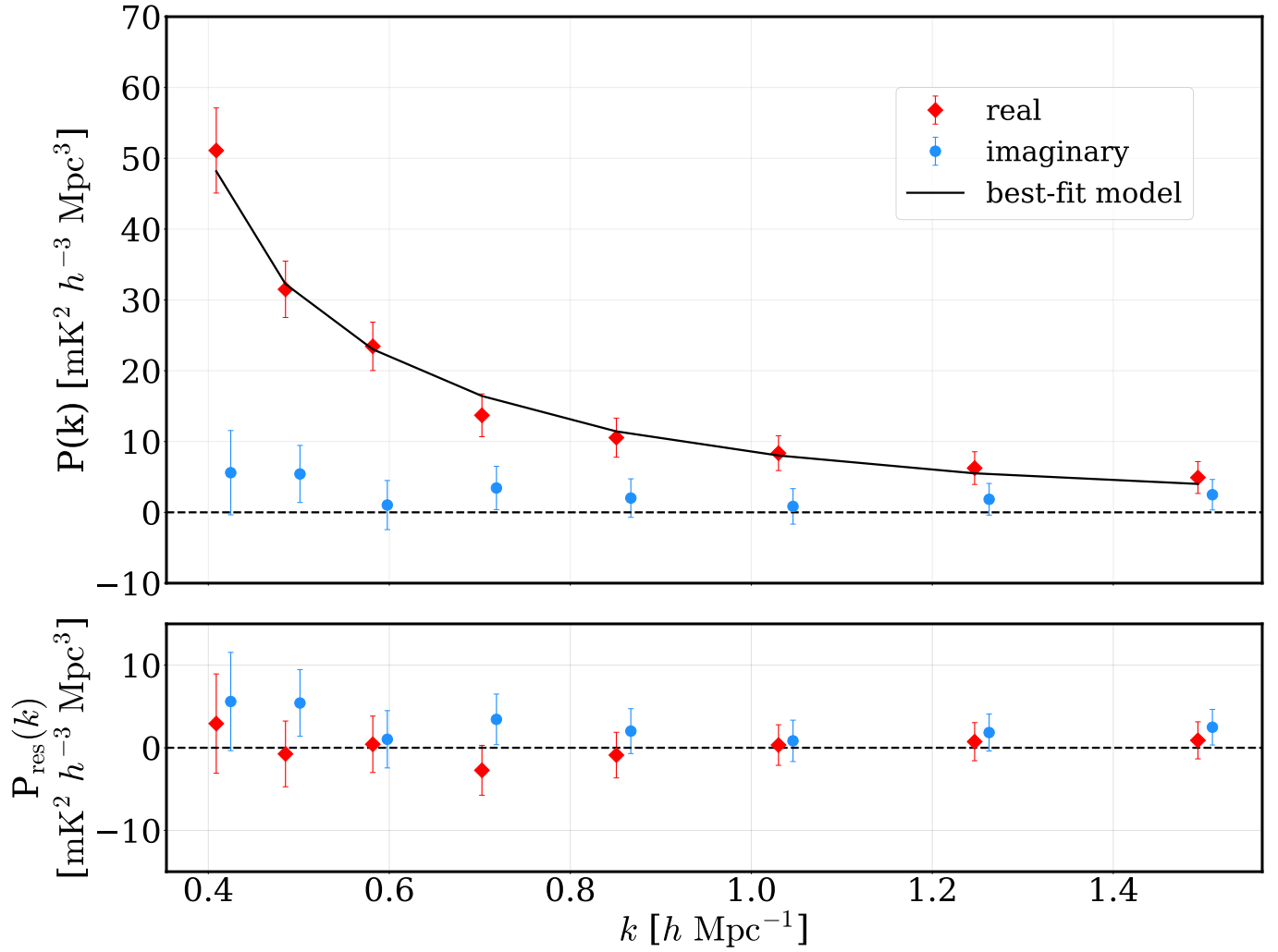


Figure 13. Top: the real component of the spherically averaged 1D Stokes I power spectrum (in red) estimated over a frequency range of 608.2–707.8 MHz ($z = 1.34\text{--}1.01$), corresponding to a central redshift of $z \sim 1.16$. The best-fit model to the data points is shown in black. The uncertainties on each measured bandpower are derived from the standard deviation of 1000 noise power spectra (see Section 5.4). The best-fit model, described in Section 6.2, yields a p -value of 0.92, indicating an excellent fit. Bottom: the residual power spectrum, after subtracting the best-fit model from the data (in red). In both panels, we show the imaginary components of the power spectrum (in blue), slightly offset in k for clarity. The power spectrum is complex valued, and a coherent signal from the sky will be confined to the real component, while the imaginary component should be noise dominated (see Section 7.1). Both residuals and the imaginary component are consistent with the estimated noise level.

damping factor $D_{\text{HI}}^{\text{FoG}}(k\mu; z)$, with damping scale $\sigma_{\text{FoG}}(z)$, that approximates “finger-of-God” redshift-space distortions (J. C. Jackson 1972) from small-scale H I velocities. Along with α_{NL} , we use $\alpha_{\text{FoG}} \equiv \sigma_{\text{FoG}}(z_{\text{eff}})/\sigma_{\text{FoG}}^{(\text{fid})}(z_{\text{eff}})$ as a free parameter, where $\sigma_{\text{FoG}}^{(\text{fid})}(z)$ is the fitting function from the CHIME Collaboration et al. (2023), based on simulation measurements from D. Sarkar & S. Bharadwaj (2019). In summary, the power spectrum model in Equations (74) and (75) has four free parameters: Ω_{HI} , b_{HI} , α_{NL} , and α_{FoG} .

This model is undoubtedly simplistic, and we emphasize that α_{NL} and α_{FoG} should be viewed as phenomenological shape parameters, intended to capture a range of plausible shapes for the nonlinear H I power spectrum without having easily interpretable physical meanings. There are several reasons why this model is appropriate for our analysis:

1. At the S/N and spatial scales of our measurement, any parameters describing the shape of the power spectrum (in our case, α_{NL} and α_{FoG}) are highly degenerate with

parameters describing the amplitude (Ω_{HI} and b_{HI}). Therefore, using a more detailed model with a larger number of parameters likely would not yield further useful information.

2. In CHIME Collaboration et al. (2026a), we describe several simulation-based tests that demonstrate that, at the S/N of our measurement, our model allows us to obtain unbiased constraints on the power spectrum amplitude parameter $\mathcal{A}_{\text{HI}}^2$ (see below) when the shape parameters α_{NL} and α_{FoG} are marginalized over.
3. Any power spectrum model must be propagated through a transfer function that is specific to the analysis we have performed. As described in CHIME Collaboration et al. (2026a), this propagation can be carried out efficiently for the specific model from Equations (74) and (75), using a simulation-based template approach based on the CHIME-eBOSS stacking analysis in CHIME Collaboration et al. (2023).

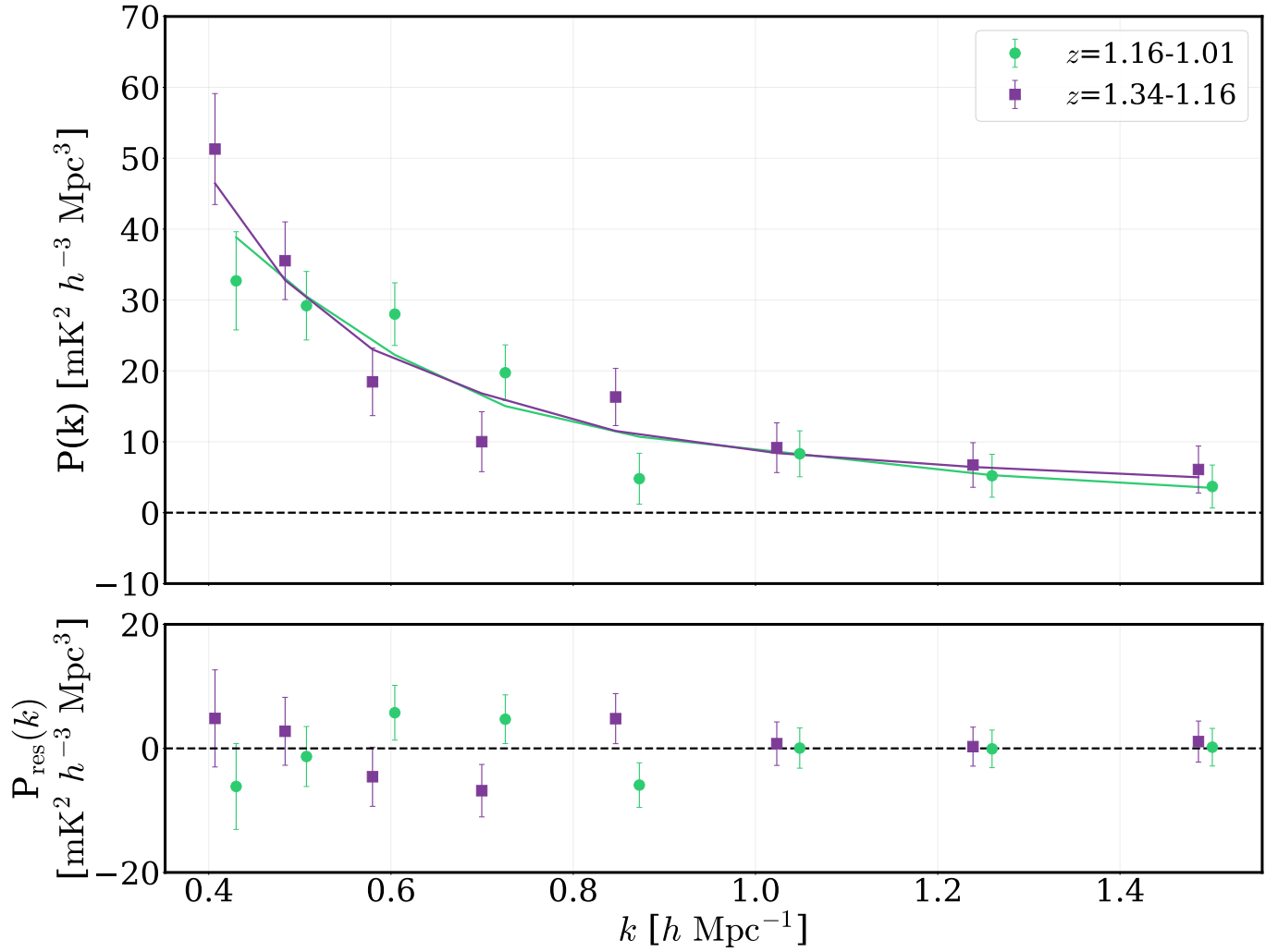


Figure 14. The power spectrum of the upper sub-band spanning 658.2–707.8 MHz ($z = 1.16$ – 1.01) and lower sub-band spanning 608.2–658.2 MHz ($z = 1.34$ – 1.16). The top panel shows the real component of the power spectrum and the best-fit model for each sub-band. The bottom panel shows the residual obtained after subtracting the best-fit model from the data. In both sub-bands, the model agreement is excellent, and the residuals are consistent with the expected noise levels. The imaginary component of the power spectrum for both sub-bands is also consistent with noise and is not shown for clarity.

6.2.2. Model Fitting

We use the `emcee` package (D. Foreman-Mackey et al. 2013) to perform Markov Chain Monte Carlo (MCMC) sampling from the posterior distribution for the four parameters in the model from Section 6.2.1. We use a Gaussian likelihood,

$$\mathcal{L}(\theta) = \frac{1}{|2\pi\mathbf{C}_N|^{1/2}} \exp\left(-\frac{\chi^2(\theta)}{2}\right), \quad (76)$$

with

$$\chi^2(\theta) \equiv [\mathbf{d} - \mathbf{s}(\theta)]^T \mathbf{C}_N^{-1} [\mathbf{d} - \mathbf{s}(\theta)], \quad (77)$$

where $\mathbf{d}$ is the data vector, $\mathbf{s}(\theta)$ is the model prediction evaluated at parameter values θ , and $\mathbf{C}_N$ is the noise covariance described in Section 5.4. When evaluating Equation (77), we multiply $\mathbf{C}_N^{-1}$ by a “Hartlap factor” $(n_{\text{mocks}} - n_{\text{data}} - 2)/(n_{\text{mocks}} - 1) \approx 0.99$ to remove bias due to the finite number of noise realizations used to estimate $\mathbf{C}_N$ (J. Hartlap et al. 2007). We run chains with 12 walkers for 5 million samples each, discarding the first 15% of samples for each walker as burn-in and verifying that the chains have

converged using the multivariate $\hat{R}$ statistic from S. P. Brooks & A. Gelman (1998), with $\hat{R} - 1 < 0.01$. We use broad, uniform priors for the parameters: $\alpha_\Omega \sim \mathcal{U}(0, 20)$, $\alpha_b \sim \mathcal{U}(0, 10)$, $\alpha_{\text{NL}} \sim \mathcal{U}(0, 5)$, and $\alpha_{\text{FoG}} \sim \mathcal{U}(0, 10)$, where for Ω_{HI} and b_{HI} , we have defined priors in terms of $\alpha_\Omega \equiv \Omega_{\text{HI}}/\Omega_{\text{HI}}^{\text{fid}}$ and $\alpha_b \equiv b_{\text{HI}}/b_{\text{HI}}^{\text{fid}}$.

We take the best-fit parameters to be those from the maximum a posteriori sample from the MCMC chains. For the analysis of the full band, this yields $(\alpha_\Omega, \alpha_b, \alpha_{\text{NL}}, \alpha_{\text{FoG}}) = (2.8, 0.24, 0, 0.29)$. For the upper sub-band, we obtain $(0.17, 8.8, 5.0, 1.2)$, while for the lower sub-band, we obtain $(0.42, 4.1, 0.084, 0.048)$. We use these parameters to generate the best-fit model curves that are plotted alongside the measurements in this paper. Note that some of these parameter values are quite extreme (compared to expected values of order unity), and/or close to the boundary of the prior. However, these best-fit values do not correspond to maximum values of the marginalized posteriors for each parameter, and there are many other points in parameter space where the values are closer to expectations and the posterior is comparably high. Therefore, the appearance of extreme best-fit values is not a cause for concern. For a complete picture of the parameter values inferred from the data, one should refer to the

discussion of the marginalized posteriors in CHIME Collaboration et al. (2026a).

The best-fit parameters in the previous paragraph yield χ^2 values of 1.6, 7.1, and 6.0 for the full band, upper sub-band, and lower sub-band, respectively. We convert these into p -values using a Monte Carlo-based procedure, whereby we refit the signal model to 1000 mock datasets comprising the best-fit signal prediction and a random noise realization, and compare the χ^2 values from the data to the ensemble of mock-based values (see Appendix D for details). This procedure yields p -values of 0.92, 0.25, and 0.31 for the three bands, indicating acceptably good fits in all cases.

6.2.3. Power Spectrum Amplitude

The power spectrum model in Equation (74) has a strong degeneracy between Ω_{HI} and b_{HI} , such that the marginalized posterior on either parameter is very broad. The form of the degeneracy can be read off from Equation (74): the amplitude of the power spectrum is determined by the product of Ω_{HI}^2 (since $\bar{T}_b \propto \Omega_{\text{HI}}$) and $(b_{\text{HI}} + f\mu^2)^2$. This implies that the angle-averaged power spectrum $P_{21}(k; z)$ is approximately proportional to the amplitude parameter²⁶

$$\mathcal{A}_{\text{HI}}^2 \equiv 10^6 \Omega_{\text{HI}}^2 (b_{\text{HI}} + \langle f\mu^2 \rangle)^2, \quad (78)$$

where the angle brackets surrounding $f\mu^2$ denote the appropriate angular average. Our notation reflects the fact that this parameter is the square of the $\mathcal{A}_{\text{HI}}$ parameter used to describe the amplitude of the CHIME-eBOSS stacking signal in CHIME Collaboration et al. (2023). We use $\langle f\mu^2 \rangle = 0.761$ to compute $\mathcal{A}_{\text{HI}}^2$, since it corresponds to the best-constrained definition of this parameter and also agrees with a direct average of $f\mu^2$ over the set of μ bins used in our measurement.

In CHIME Collaboration et al. (2026a), we perform a full Bayesian analysis to infer $\mathcal{A}_{\text{HI}}^2$ from the measured power spectrum, and find that $\mathcal{A}_{\text{HI}}^2 = 3.55^{+0.96}_{-1.32}(\text{stat.}) \pm 0.61(\text{sys.})$ for the full frequency band, $\mathcal{A}_{\text{HI}}^2 = 2.47^{+1.59}_{-1.08}(\text{stat.}) \pm 0.43(\text{sys.})$ for the upper sub-band, and $\mathcal{A}_{\text{HI}}^2 = 3.24^{+6.79}_{-2.11}(\text{stat.}) \pm 0.56(\text{sys.})$ for the lower sub-band. The statistical uncertainties are determined from the 68% highest posterior density intervals of the marginalized posteriors for $\mathcal{A}_{\text{HI}}^2$, while the systematic uncertainty incorporates several contributions related to the estimated accuracy of our calibration and modeling framework (see CHIME Collaboration et al. 2026a for details).

Adopting a literature-informed prior on b_{HI} can enable our constraints on $\mathcal{A}_{\text{HI}}^2$ to be propagated into inferences about Ω_{HI} , and vice versa; we discuss this further in CHIME Collaboration et al. (2026a). We also compare the constraining power of the auto spectrum and stacking the same data on eBOSS quasars in Section 7.10.

6.3. Detection Significance

We use a Monte Carlo-based likelihood ratio test to assess the detection significance of our power spectrum measurements in the full band and each sub-band. For each of the 1000

²⁶ To be precise, the angle-averaged power spectrum is proportional to $b_{\text{HI}}^2 + 2b_{\text{HI}}\langle f\mu^2 \rangle + \langle f^2\mu^4 \rangle$, but we show in CHIME Collaboration et al. (2026a) that the approximation $\langle f^2\mu^4 \rangle \approx \langle f\mu^2 \rangle^2$ is sufficiently accurate over the range of μ of our measurement.

Table 2
Power Spectrum Detection Significance in Each Band

ν Band (MHz)	z	S/N		
		F -based	χ^2 -based	Amplitude-based
608.2–707.8	1.34–1.01	12.4	13.0	13.6
658.2–707.8	1.16–1.01	8.6	8.8	9.6
608.2–658.2	1.34–1.16	9.1	9.4	10.0

Note. We quantify the detection significance by computing $\Delta\chi^2 \equiv \chi_{\text{null}}^2 - \chi_{\text{best-fit}}^2$ for each of 1000 simulated noise power spectra, determining the best-fit distribution of those $\Delta\chi^2$ values, and using this distribution to compute an S/N value corresponding to the $\Delta\chi^2$ value for the data. An F distribution provides a good fit to the $\Delta\chi^2$ values (after appropriate scaling; see Appendix D for details), and we use the associated S/N values for each band (shown in bold) as our main results. For comparison, we also quote S/N values based on the best-fit χ^2 distribution, and based on scaling the best-fit model prediction with a free amplitude.

noise power spectra described in Section 5.4, we compute

$$\Delta\chi^2 \equiv \chi_{\text{null}}^2 - \chi_{\text{best-fit}}^2, \quad (79)$$

where χ_{null}^2 is given by Equation (77) with $s = 0$, corresponding to a noise-only model, and $\chi_{\text{best-fit}}^2$ is

$$\chi_{\text{best-fit}}^2 \equiv \min_{\theta} \chi^2(\theta). \quad (80)$$

We then compare the $\Delta\chi^2$ value for the data, computed in the same manner, with an analytical probability distribution function fitted to the 1000 values from the noise power spectra.

Table 2 shows the results of three versions of this procedure. The most conservative version, which we use for our main results, is based on fitting an F distribution to the 1000 $\Delta\chi^2$ values (after appropriate scaling). This is motivated by the fact that the noise covariance used in Equation (79) and Equation (80) is estimated from simulations and therefore adds its own source of scatter to the χ^2 values (see Appendix D for further discussion). The second S/N column in Table 2 is based on fitting a χ^2 distribution instead of an F distribution, which is conceptually simpler but ignores scatter in the noise covariance estimate; therefore, the inferred distribution of $\Delta\chi^2$ values has a lighter tail than for the F distribution, and this slightly inflates the detection significance inferred from the $\Delta\chi^2$ value for the data.

Finally, in the third S/N column of Table 2, we quote an even simpler S/N estimate, based on scaling the best-fit power spectrum prediction (with all model parameters fixed) with an amplitude prefactor A , and computing σ_A/A . As expected, these values are higher than those from the other methods, since they do not account for the freedom to determine the four model parameters from the data.

7. Validation

In this section, we describe a comprehensive set of validation tests used to establish the robustness and reliability of our power spectrum measurement. These tests assess whether the data are consistent with the cosmological 21 cm signal plus thermal noise. In all cases, consistency is evaluated statistically using the appropriate noise covariance estimated from the Monte Carlo noise realizations described in Section 5.4. Several are null tests, in which the data are processed such that any cosmological signal cancels and the

resulting power spectrum should be consistent with zero. The remaining tests assess whether the measured power spectrum remains consistent under independent data splits and changes to the analysis pipeline, as expected for the cosmological 21 cm signal. Statistically significant deviations in these tests would indicate residual systematics or an inaccurate noise model.

For each null test, we construct the null bandpower vector $\mathbf{d}_{\text{null}}$, which should be statistically consistent with zero. We test this consistency by computing the χ^2 under the null hypothesis that data are drawn from a zero-mean distribution with the appropriate noise covariance:

$$\chi^2 = \mathbf{d}_{\text{null}}^T \mathbf{C}_N^{-1} \mathbf{d}_{\text{null}}. \quad (81)$$

The noise covariance matrix $\mathbf{C}_N$ for each null test is estimated by applying the identical analysis pipeline on 1000 noise realizations, as described in Section 5.4. We then compute the p -value from the χ^2 statistics with 8 dof, as we have 8 bandpower k -modes probed in this analysis.²⁷ We mention the p -values for all validation tests in Table 3. These p -values show that the validation tests are statistically consistent with the estimated thermal noise covariances, with no evidence that residual systematics dominate the measured auto power spectrum.

7.1. Imaginary Component of the Power Spectrum

The power spectrum is complex-valued because we cross-correlate two independent maps, each constructed from a stack of either even or odd sidereal days. If the sky is stationary and the instrument response is identical on even and odd sidereal days, then the cross power from the sky is real-only, and the imaginary part is consistent with noise. Departures from this expectation can arise from, for example, relative spectral shifts or small astrometric misalignments between the maps (due to pointing offsets introduced by phase calibration errors), which rotate real power into the imaginary component (M. Kolopanis et al. 2019; HERA Collaboration et al. 2022). The imaginary component of the Stokes I power spectrum is shown in the bottom panel of Figure 13 and yields $\chi^2/\text{dof} = 0.82$ with a corresponding p -value of 0.58. This indicates excellent consistency with thermal noise and no evidence for residual phase calibration-induced offsets. In the null tests below, we likewise evaluate the imaginary component against the noise-only expectation.

7.2. Consistency between Even and Odd Days

As described previously (Section 5.3), we split the total 94 nights of data into two even and odd subsets and cross-correlate them to estimate the power spectrum. To assess temporal stability and check for consistency between these subsets, we further subdivide each into two halves and construct differenced hybrid-beamformed visibility datasets, $\tilde{V}_{\text{diff},1}^p$ and $\tilde{V}_{\text{diff},2}^p$ (following CHIME Collaboration et al. 2023). We then process these datasets through the identical analysis pipeline (described in Section 3 and Section 5) and generate the power spectrum by cross-correlating them.

²⁷ In each case, we have verified with the KS and AD tests that the 1000 χ^2 values are well described by a χ^2 distribution. This justifies our usage of χ^2 statistics, as opposed to the F distribution we use for detection significance calculations.

Table 3

Statistical p -values for the Validation Tests Described in Section 7, Which Assess Whether Jackknife Combinations and Differences between Data Splits or Mask Choices Are Consistent with Our Noise Model

Validation Tests		
Data Selection	Selection Criteria	p -value
Imaginary component	$\text{Im}[P(k)]$	0.58
(Even–odd) nights	$\text{Re}[P(k)]$	0.90
	$\text{Im}[P(k)]$	0.55
Stokes Q	$\text{Re}[P(k)]$	0.21
	$\text{Im}[P(k)]$	0.33
R.A. bins	(RA2—RA1)	0.54
Decl. bins	(Dec2—Dec1)	0.48
Baseline selections	[(1+2+3)-cylinder]	0.75
	—[1-cylinder]	
Flux Cuts		
Mask type	Threshold (Jy)	p -value
Track mask	20	0.20
	25	0.16
	30	0.77
	40	0.74
Transit mask	2	0.85
	3	0.53
	4	0.42
	6	0.68

We expect the cosmological 21 cm signal to remain constant across sidereal days and therefore cancel in the difference between subsets. The noise is independent between different days and hence will remain in each of the differenced visibility datasets. However, this will not result in any positive bias in a cross power spectrum between the two differenced datasets $\tilde{V}_{\text{diff},1}^p$ and $\tilde{V}_{\text{diff},2}^p$, because each will have an independent noise realization. Transient RFI is expected to be independent between subsets and will persist in the difference. Systematic foreground residuals exhibit different behavior depending on their origin: residuals arising from chromatic instrumental effects or seasonal variations in the instrument response will cancel, as these remain stable on day-to-day timescales. However, foreground contamination driven by day-to-day instrumental variability will not cancel and may appear in the difference.

Therefore, in the absence of transient RFI or temporal instrumental instability, the resulting power spectrum should be consistent with zero. Figure 15 shows both real and imaginary components of the power spectrum from these differenced datasets. Under the null hypothesis that the observed signal remains the same on different subsets, we find $\chi^2/\text{dof} = 0.43$ with a p -value of 0.90 for the real component and $\chi^2/\text{dof} = 0.86$ with a p -value of 0.55 for the imaginary component. Both are fully consistent with noise, indicating the absence of transient RFI or day-to-day instrumental variations.

7.3. Power Spectrum of Stokes Q Data

A critical potential contaminant of the measurement of the 21 cm power spectrum is polarized foreground emission. The

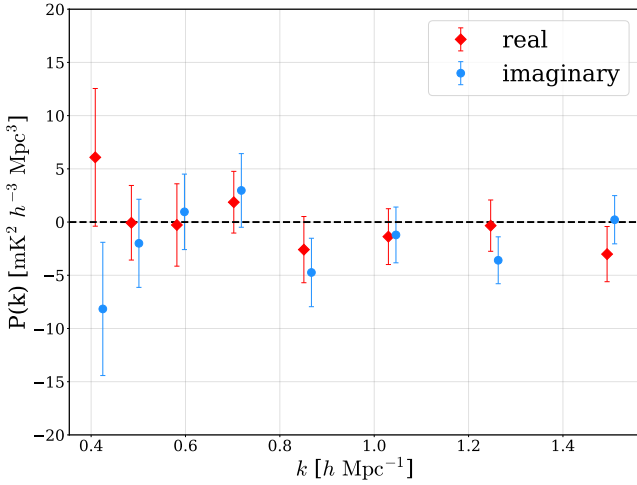


Figure 15. The real and imaginary components of the power spectrum estimated from the difference of interleaved nights. For visual clarity, we apply a small offset to the data points along k . These are consistent with the null hypothesis that the data are drawn from the noise covariance.

cosmological H 121 cm signal is intrinsically unpolarized. In contrast, polarized foreground emission can undergo Faraday rotation as it propagates through magnetized plasma in the interstellar medium, introducing a chromatic structure that varies as λ^2 and potentially mimics the frequency dependence of the cosmological signal (D. F. Moore et al. 2013; J. R. Shaw et al. 2015; K. M. B. Asad et al. 2015; D. F. Moore et al. 2017; Z. E. Martinot et al. 2018). This frequency structure can leak into Stokes I measurements through various instrumental nonidealities: nonorthogonal or rotated antenna feeds, cross-talk between feeds, and asymmetries between the XX and YY beam patterns. These effects provide mechanisms for spectrally nonsmooth, Faraday-rotated polarized emission to contaminate the unpolarized 21 cm signal.

J. R. Shaw et al. (2015) provide the mathematical formalism to study the leakage between all four Stokes parameters (I , Q , U , V) for a cylindrical drift-scan telescope and establish the calibration accuracy and beam modeling requirements necessary to detect the cosmological signal. They showed that if the two orthogonal feeds satisfy two key constraints—equal normalized power patterns everywhere on the sky ($|A_{XX}|^2 = |A_{YY}|^2$) and orthogonal polarization orientations across the entire field of view ($A_{XX} \cdot A_{YY} = 0$)—then the measured visibility data is sensitive only to the total intensity Stokes I and insensitive to the other polarization modes. In other words, if there are no significant deviations from these two constraints, the Stokes I component should measure the unpolarized cosmological 21 cm signal without polarized contamination. Of course, a real instrument like CHIME does not perfectly satisfy these two conditions. We quantify the potential contamination from polarized leakage due to instrument nonidealities by measuring the power spectrum of Stokes Q maps.

We process Stokes Q maps for even and odd day datasets through the identical pipeline used for Stokes I and measure the power spectrum. The real and imaginary components of the power spectrum are shown in Figure 16. Under the null hypothesis that the Stokes Q power spectrum is drawn from the noise covariance and consistent with zero, we find p -values of 0.21 and 0.33 for the real and imaginary components, respectively. The consistency of the Stokes Q power spectrum

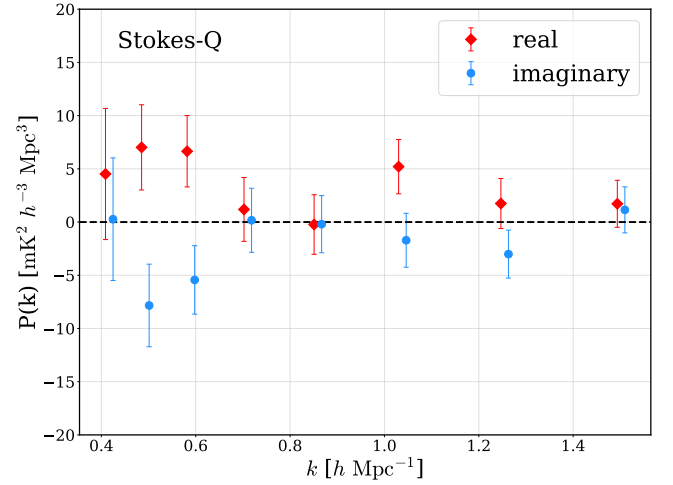


Figure 16. Here we show the power spectrum estimated from the Stokes Q maps. In the absence of significant deviation between XX and YY beam patterns, we expect that the power in Stokes Q should be consistent with noise fluctuations. We find that the p -value is greater than 0.2, under the null hypothesis that the Stokes Q power spectrum is drawn from the noise covariance and consistent with zero. This indicates that we do not detect any significant power in the Stokes Q component and rule out polarized foreground contamination in the Stokes I measurement through leakage from Stokes Q to Stokes I . While the real and imaginary components show apparent structure, it is not statistically significant and remains consistent with noise fluctuation. The measured power spectrum in Stokes Q provides an upper bound on $Q \leftrightarrow I$ leakage and indicates that contamination in the Stokes I measurement due to any polarized foreground leakage is subdominant.

with noise, combined with the highly significant detection in Stokes I , provides strong validation that the measured signal originates from intrinsically unpolarized H I emission, and any polarized foreground leakage is subdominant. If polarized foregrounds were responsible for the Stokes I detection through instrumental leakage, they would necessarily produce detectable power in Stokes Q at a comparable or higher level, which we do not observe. The measured Stokes Q power spectrum serves as an upper bound on potential $Q \leftrightarrow I$ leakage.

This test demonstrates that any polarized foreground leakage is negligible and subdominant to the detected Stokes I measurement of the 21 cm signal.

7.4. Power Spectrum of Two Independent R.A. Bins

To test the presence of any R.A.-dependent systematics, we divide the data into two independent R.A. bins and measure the power spectrum separately for each subset. The full NGC region spans 110° – 265° in R.A. (Figure 6), which we partition into RA1 (110° – 190°) and RA2 (190° – 265°). Each subset is processed through the identical analysis pipeline described in Section 5.

The frequency mask applied to the full R.A. range data is the same for both R.A. bins. However, the spatial mask, which depends on the location of the foreground sources in the sky, is different between the two R.A. bins. Consequently, the flagging fraction is also different. The spatial mask excludes approximately 25% of the sky area in RA1, compared to 42% in RA2.

Residual RFI and instrumental systematics that are localized in R.A. or local sidereal time will cause excess variance in the power spectrum for one R.A. bin but may not be present in the other. Additionally, contamination from Galactic foreground depends on R.A. and may cause systematic foreground leakage

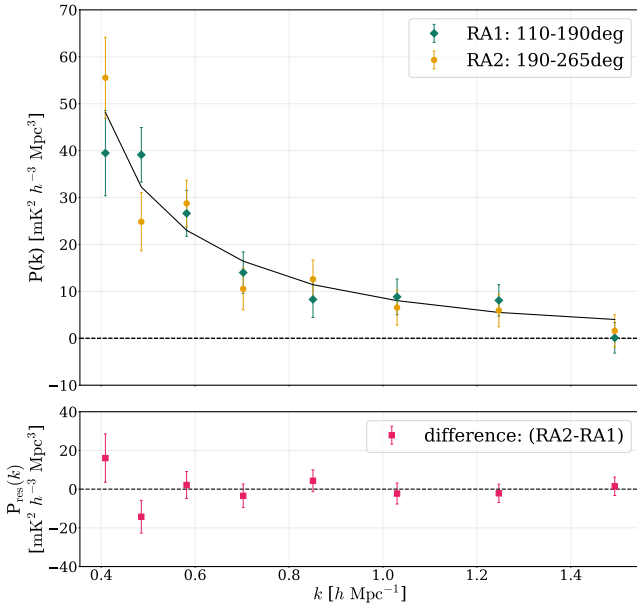


Figure 17. The top panel shows the real component of the power spectrum of two R.A. bins, RA1: 110°–190° and RA2: 190°–265°. We also show the best-fit H I model to the data of the whole R.A. range (the same best-fit model curve as shown in Figure 13). We subtract the power spectrum of RA2 from RA1, both for data and noise realizations, and construct the noise covariance from the difference noise power spectra. The residual power spectrum, along with the 1σ uncertainty, is shown in the bottom panel. We find that the residual power spectrum is consistent with arising from noise fluctuations.

into high-delay modes that could manifest differently across the two R.A. bins. Simulations of Galactic foreground emission show more than 40% higher variance in RA1 compared to RA2 due to the proximity to the Galactic plane. Even after excluding the intracylinder baselines, which are dominated by Galactic emission (see Section 3.6), the differences persist between R.A. bins. However, the cosmological 21 cm signal is isotropic and R.A. independent and therefore should be identical across both R.A. bins. Hence, for a true detection of the cosmological 21 cm signal, we expect the measured power spectrum should be similar between different R.A. bins.

Figure 17 shows the measured power spectra for both R.A. bins. The top panels show the real component of the power spectrum. In the bottom panel, we show the residual power spectrum estimated by taking the difference between RA2 and RA1. To assess whether this residual is consistent with noise, we construct a noise covariance matrix by taking the difference between 1000 noise power spectra for the two R.A. bins. The uncertainty on the residual power spectrum (shown as error bars in the bottom panel) is estimated as the 1σ scatter of these 1000 differenced noise realizations. Using this covariance matrix, we compute the χ^2 statistic for the residual power spectrum. Under the null hypothesis that the observed signal is identical in both R.A. bins and the residual is consistent with noise fluctuations alone, we obtain a p -value of 0.54. This indicates that the residual is fully consistent with noise, with no evidence for systematic differences between the two R.A. bins.

This test demonstrates that there is no evidence of R.A.-dependent systematics present in the data. Importantly, since Galactic emission is substantially brighter in RA1 compared to RA2, the observed consistency between R.A. bins provides

strong evidence that our detected signal is not dominated by residual Galactic foregrounds.

7.5. Power Spectrum of Two Independent Decl. Bins

To test for decl.-dependent systematics, we divide the data into two independent decl. bins, Dec1 (38.4–49.2) and Dec2 (49.3–60.2), and measure the power spectrum for each subset independently following the procedure described in Section 5.

The contamination from the point-source foregrounds leaking through the main lobe and far sidelobes of the telescope beam will be different between the two decl. bins. We also expect the Galactic plane emission will have a decl.-dependent structure (similar to Section 7.4). Additionally, CHIME Collaboration et al. (2023) demonstrated that the primary beam response contains large ($\sim 50\%$) ripples in frequency and decl., due to multipath interference (see their Figure 9). If foreground contamination or any other decl.-dependent systematics are present, we would expect excess variance in the measured power spectrum as a function of decl. due to all of these effects. In contrast, the cosmological 21 cm signal is isotropic and should yield identical power spectra in both bins.

Figure 18 shows the real component of the measured power spectra for both decl. bins in the top panel, while the bottom panel shows the difference between them. We compute the noise covariance matrix by taking the difference between 1000 noise power spectra of these two decl. bins. Under the null hypothesis that the observed signal is identical in both bins and the difference between them is consistent with noise, we obtain a p -value of 0.46. This shows that the residual is in agreement with the noise.

This test demonstrates that we measure the same cosmological 21 cm signal in two decl. bins, and there is no evidence of decl.-dependent systematics.

7.6. Power Spectrum with Different Baseline Selections

In this section, we test for the presence of baseline-dependent foreground contamination in the measured power spectrum. Any residual foreground leakage is expected to exhibit baseline-dependent structure. The projected geometric delay is larger for longer baselines, and because of this, spectrally smooth foregrounds will appear at larger delay or higher $k_{\parallel}$ modes for those baselines, resulting in a characteristic “foreground wedge” (A. R. Parsons et al. 2012). Although we beamform along the NS direction, which removes this dependency for NS baselines, geometric delays remain associated with EW baselines (see Section 3.2.2). This results in foreground contamination at higher delay modes for longer EW baselines, and this can potentially evade the foreground filter. Additionally, bright point sources drifting through the far sidelobes of the telescope beam can leak power at higher delays due to large projected delays associated with longer EW baseline separations (see Section 7.8). Hence, if any potential foreground leakage contaminates our measurement, we would expect systematic differences between power spectra measured using different baseline subsets.

To test for this effect, we generate two maps using different baseline configurations: one using only 1-cylinder ($b^{\text{EW}} \in 22$ m) baselines and another using the combined (1 + 2 + 3)-cylinder ($b^{\text{EW}} \in 22, 44, 66$ m) baselines employed in our fiducial analysis. We then process both maps through an

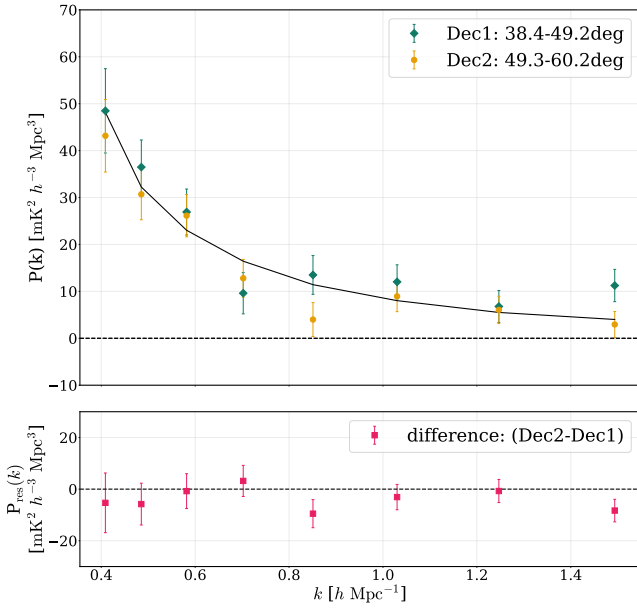


Figure 18. The top panel shows the real component of the power spectrum of two decl. bins, Dec1.: 38.4–49.2 and Dec2: 49.3–60.2, along with the best-fit H I model to the data of the whole decl. range (the same best-fit model curve as shown in Figure 13) in black. We subtract the power spectrum of Dec2 from Dec1, both for data and 1000 noise realizations, and construct the noise covariance from the difference noise power spectra. The different power spectrum, along with the 1σ uncertainty, is shown in the bottom panel. We find that the residual power spectrum is consistent with the noise fluctuations.

identical analysis pipeline and estimate the 2D and 1D power spectra. The cylindrically binned 2D power spectrum for the 1-cylinder baseline map has more limited $k_{\perp}$ coverage compared to the (1+2+3)-cylinder map due to the shorter maximum baseline length. To enable a consistent comparison, we bin both power spectra to 1D spherical bins using identical bin edges corresponding to the 1D k -mode ranges probed by the 1-cylinder map.

The real component of the power spectrum for two baseline selections is shown in the top panel of Figure 19, where in the bottom panel we show the difference between them. We also show the best-fit model after rebinning using the same 1D bin edges as used for the data. We offset the data points slightly along k for visual clarity. We construct a noise covariance matrix by taking the difference between 1000 noise power spectra of these two baseline selections, and the uncertainty on the residual power spectrum (bottom panel) is estimated by taking the standard deviation of these differenced noise power spectra. Under the null hypothesis that the observed signal is identical in the 1-cylinder and (1+2+3)-cylinder baseline configurations and the difference between them is consistent with noise fluctuations alone, we obtain a p -value of 0.75. This indicates that there is no systematic difference between power spectra measured using different subsets of baselines, and the difference between them is fully consistent with noise.

If residual foreground leakage contaminates our measurement, we would expect the (1+2+3)-cylinder baseline configuration, which includes the longest 66 m EW baselines, to show enhanced power compared to the 1-cylinder baseline configuration that includes only the shortest 22 m EW baselines. We do not see any such evidence in the measured power spectrum for these two baseline selections. This test demonstrates that we measure the cosmological 21 cm signal

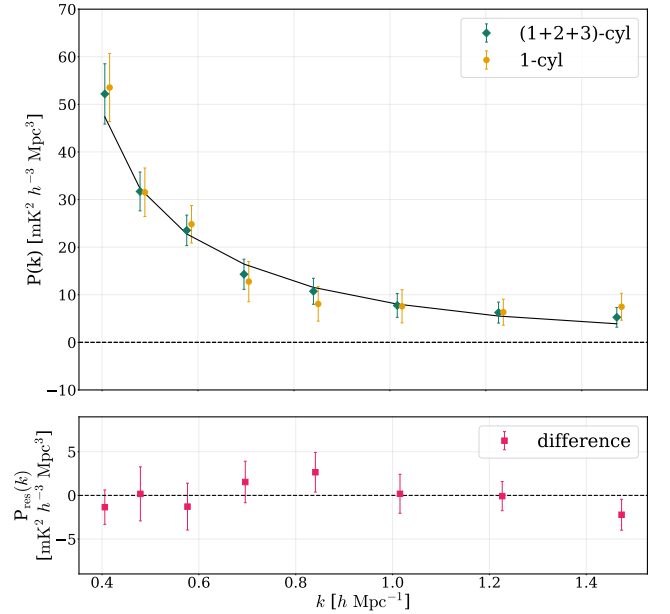


Figure 19. The top panel shows the real component of the spherically averaged power spectrum measured using 1-cylinder baselines only and the combined (1+2+3)-cylinder baselines employed in our fiducial analysis. The black curve shows the best-fit signal model rebinned to the same k bins corresponding to the 1-cylinder map. Data points are slightly offset in k for visual clarity. Both baseline configurations yield consistent measurements across all scales. In the bottom panel, we show the difference between these two power spectra. The uncertainty on each bandpower is estimated from the standard deviation of 1000 differenced noise power spectra. The residuals yield a p -value of 0.75 under the null hypothesis that the difference is consistent with noise fluctuations. This indicates there is no baseline-dependent foreground contamination present in our measurement.

consistently using different baseline selections, with no indication of residual baseline-dependent systematics.

7.7. Power Spectrum of Two Independent Sub-bands

We report, in Section 6.1, the detection of the cosmological 21 cm signal in two independent sub-bands: the upper sub-band centered at $z \sim 1.08$ and the lower sub-band centered at $z \sim 1.24$. The spherically averaged power spectrum for these sub-bands is shown in Figure 14. We find that the power spectra are broadly consistent with each other. Fitting a signal model independently to each sub-band yields best-fit power spectrum amplitudes ($\mathcal{A}_{\text{HI}}^2$) that are consistent within 1σ , as discussed in Section 6.2.3. The 1D marginalized posterior distributions for the power spectrum amplitude parameter for both sub-bands are shown in Figure 20.

These independent measurements also provide a critical test for frequency-dependent systematic contamination. If residual foreground contamination were present, we would expect a significantly large bias in the lower sub-band ($z = 1.24$), since foregrounds—dominated by synchrotron emission—follow a power-law scaling with frequency. However, the intrinsic 21 cm signal is not expected to evolve substantially over this narrow redshift range. Based on the fitting functions for $\Omega_{\text{HI}}(z)$ and $b_{\text{HI}}(z)$ that we adopt in Section 6.2.1, we compute the theoretically predicted evolution of the power spectrum amplitude between the two redshifts. The model predicts that the amplitude at $z = 1.24$ is only $\sim 13\%$ higher than at $z = 1.08$.

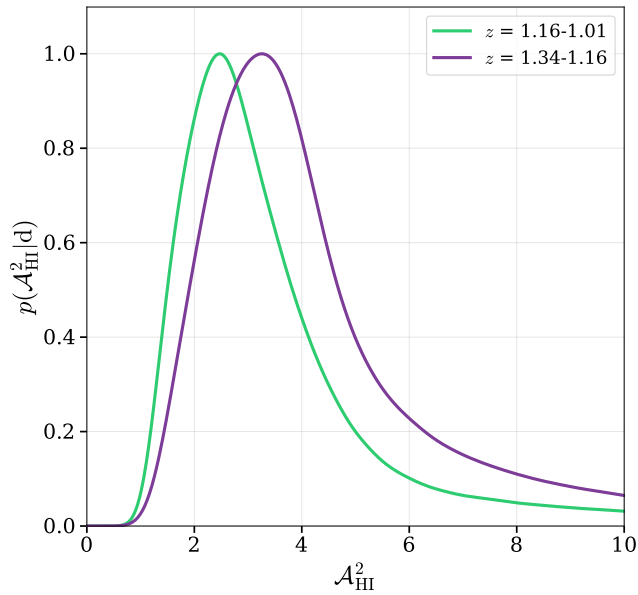


Figure 20. The marginalized 1D posterior of the power spectrum amplitude parameter (A_{HI}^2) derived from the best-fit model for two sub-bands. The upper sub-band spanning 658.2–707.8 MHz ($z = 1.16$ – 1.01), centered at $z = 1.08$ and lower sub-band spanning 608.2–658.2 MHz ($z = 1.34$ – 1.16), centered at $z = 1.24$. We find that the amplitude parameters are consistent with each other within 1σ . Although we expect mild evolution between two redshifts, the current S/N is insufficient to probe the redshift evolution conclusively.

The measured best-fit power spectrum amplitudes are consistent with each other within 1σ . The current signal-to-noise ratio is insufficient to probe redshift evolution conclusively. Future work with substantially more data can probe the evolution of the 21 cm signal over redshifts. However, the consistency between the two sub-bands provides evidence against frequency-dependent foreground contamination. Significant foreground leakage would manifest as excess power at lower frequencies, which we do not observe, and demonstrates that there is no frequency-dependent foreground leakage.

7.8. Changing the Track Mask for Different Flux Cuts

The dominant contamination in our foreground-filtered maps arises from bright radio point sources drifting through the far sidelobes of the primary beam (see Figure 6). At low elevation, the projection of the source direction onto the EW baseline yields a large geometric delay, pushing sidelobe pickup into high-delay modes. Although most far-sidelobe signals are removed by the fringe-rate filter applied during the mapmaking process (see Equation (42)), there are low-elevation intervals for which the apparent fringe rate lies within its passband; during those intervals, the large EW delay places power beyond our delay cutoff, allowing leakage to survive both filters and appear prominently at high delay. Because the far-sidelobe primary beam response is low, this effect is visible only for the brightest sources. To mitigate it, we apply spatial masks that exclude the predicted sky tracks of these sources (see Section 5.2).

A critical concern is whether fainter sources—below our baseline flux threshold—contribute residual foreground power that could be misinterpreted as cosmological 21 cm signal. To test this hypothesis, we investigate the impact of fainter sources that lie below our nominal flux threshold. We generate increasingly aggressive track masks by progressively lowering

the flux density threshold and measure the resulting power spectra. Each mask is defined by a different flux density cutoff, f_{cut} , applied to an external source catalog. This systematic variation allows us to directly assess the sensitivity of our power spectrum measurements to the exclusion of sources with progressively lower flux densities.

We construct the spatial mask for a given flux cutoff by combining our base mask (Section 5.2) with the individual track masks of all sources exceeding that cutoff:

$$T_{\text{f}_{\text{cut}}}(y, \phi) = T_{\text{base}}(y, \phi) \times \left[\prod_{s \in S_{f_{\text{cut}}}} T_{\text{track}}^s \right], \quad (82)$$

where $S_{f_{\text{cut}}}$ is the set of catalog sources with flux densities exceeding f_{cut} . For this test, we generate masks using flux thresholds of $f_{\text{cut}} = \{20, 25, 30, 40\}$ Jy, compared against our baseline threshold of 60 Jy (base mask) used for the main analysis. These progressively aggressive masks flag 58%, 52%, 43%, and 40% of spatial pixels, respectively, compared to 33% for the base mask. Note that the transit masks for sources above 10 Jy remain fixed across all tests.

For each mask configuration, we estimate the power spectrum following the procedure in Section 5. If residual foreground contamination from fainter sources dominates our measurement, we expect the measured power to decrease systematically as we increase the spatial masking fraction (i.e., as we exclude more potential foreground sources). Conversely, if we are detecting the cosmological 21 cm signal—which is statistically isotropic—the measured power should remain stable across different mask configurations, with only the statistical uncertainty increasing due to reduced sky coverage. The top panel of Figure 21 presents the real component of the power spectra for all mask configurations along with the base mask and the corresponding best-fit signal model (same as Figure 13). We generate 1000 noise realizations for each mask configuration separately and estimate the power spectrum uncertainty from these simulations. We observe no systematic trend in the measured power as a function of masking fraction; all power spectra remain consistent within their 1σ uncertainties.

Because most of the unmasked samples are identical for any two masks, the corresponding power spectra are computed from nearly the same underlying data; therefore, the error bars in the top panel of Figure 21 are strongly correlated across masks. To quantify consistency more rigorously, we form residual power spectra by subtracting the base mask spectrum from each test-mask spectrum. We apply the same paired subtraction to 1000 noise realizations, using the same realization for both masks before differencing. We then estimate the covariance of these residuals from the ensemble of differenced noise realizations and use it in Equation (81) to test consistency with zero. The bottom panel of Figure 21 shows the residual spectra constructed in this way, with correspondingly smaller error bars. All residuals yield $p > 0.1$, indicating that the small changes observed in the power spectra as we apply increasingly aggressive masks are noise driven rather than due to residual foreground contamination.

This null test provides strong evidence that fainter sources do not contribute measurable foreground leakage through sidelobes in our analysis. The stability of the measured power spectrum across varying degrees of spatial masking

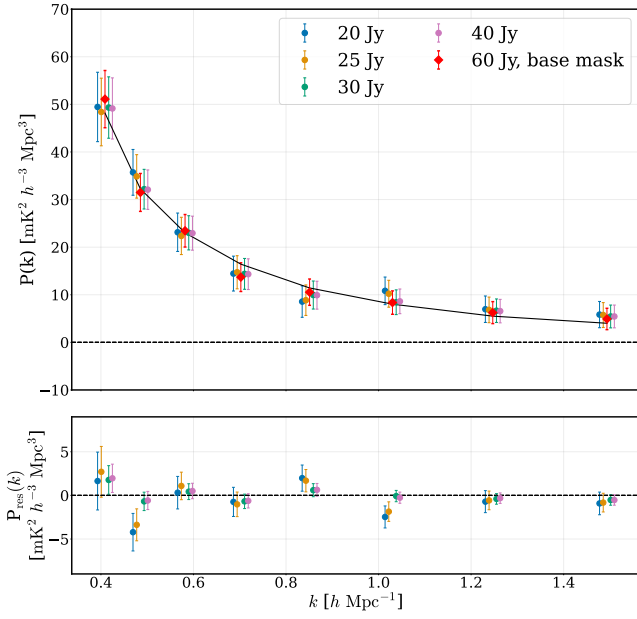


Figure 21. Power spectra made with different spatial sidelobe masks generated with different flux cuts. We show the real part of the power spectrum in the top panel, and in the bottom panel, we show the residual after subtracting the power spectrum for different cuts from the base mask. In all these different sidelobe masks, we kept the main lobe mask fixed and chose a flux cut for the main lobe mask to be 10 Jy. For visual clarity, we apply a small offset to the data points along k . The residuals across different flux cuts are statistically consistent with zero, indicating no systematic dependence on the sidelobe masking threshold.

demonstrates that we are detecting the cosmological 21 cm signal rather than residual sidelobe contamination.

7.9. Changing the Transit Mask for Different Flux Cuts

In this section, we test for multiplicative systematics that could imprint high-delay structure on foreground emission in the NGC field. Several instrumental effects could produce such systematics: (i) narrowband features in per-feed gain solutions caused by residual RFI during calibration; (ii) imperfect handling of the RFI mask during foreground filtering; (iii) chromatic primary beam response due to mutual coupling between feeds; (iv) time-variable cable reflections (e.g., from the 50 m coax with round-trip delay of 390 ns) that produce bandpass ripples at high delay. Our test is agnostic to the specific cause and instead targets their generic consequence: modulation of otherwise spectrally smooth foregrounds, resulting in high-delay leakage that scales with foreground brightness and is therefore largest for bright point sources in the main lobe.

To accomplish this, we design a jackknife test similar to the one described in Section 7.8, where we progressively mask fainter sources in the main lobe of the synthesized beam by lowering the flux density threshold when generating the transit mask (as described in Section 5.2). The combined mask is given by

$$T_{\text{cut}}(y, \phi) = T_{\text{base}}(y, \phi) \times \left[\prod_{s \in S_{f_{\text{cut}}}} T_{\text{transit}}^s \right], \quad (83)$$

where $S_{f_{\text{cut}}}$ is the set of catalog sources with flux densities exceeding f_{cut} . We test four flux thresholds, $f_{\text{cut}} = \{2, 3, 4, 6\}$ Jy, and then compare against our base mask, where a 10 Jy

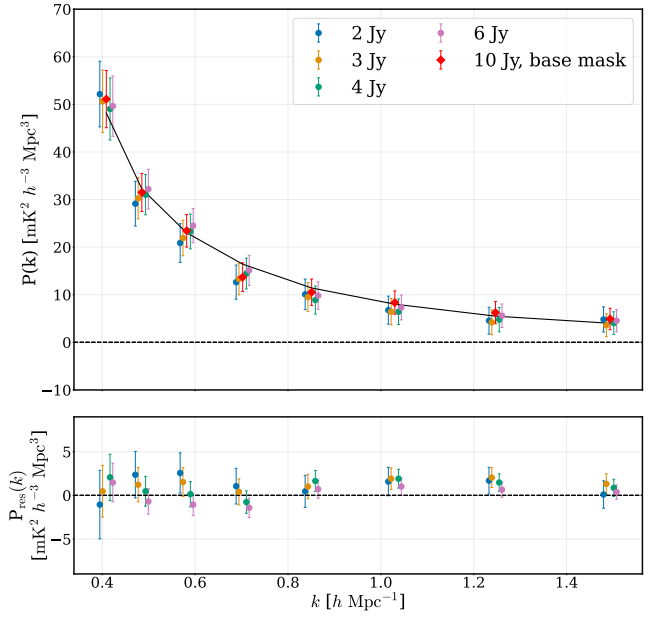


Figure 22. Power spectra made with different transit masks generated with different flux density thresholds. We show the real part of the power spectrum in the top panel, and in the bottom panel, we show the residual after subtracting the power spectrum for different cuts from the base mask. For visual clarity, we apply a small offset to the data points along k . We find that for different cuts, the residual is consistent with noise.

threshold is used for the transit mask. These transit masks exclude 53%, 45%, 42%, and 39% of sky area, respectively. We then compute the power spectrum of the data with each of these masks applied.

The resulting power spectra are shown in Figure 22. The top panel shows the real component of the power spectrum for all transit mask configurations, with the best-fit signal model to the base mask overlaid. We observe no systematic decrease in the measured power as a function of the transit mask flux threshold; all measurements remain consistent within their 1σ uncertainties. The bottom panel shows residuals relative to the base mask, computed following the same procedure as described in Section 7.8. All residuals yield p -values > 0.3 , confirming consistency with zero.

If multiplicative high-delay systematics dominated the measured power, we would expect a predictable decrease as we lower the transit mask threshold. For a multiplicative effect, the contaminant from each source scales with its flux S , so the power contributed by the source scales roughly as S^2 . With typical near-Euclidean differential counts ($dN/dS \propto S^{-\beta}$ with $\beta \simeq 2.5$, (G. de Zotti et al. 2010)), the residual leaked power after masking sources brighter than S_c scales approximately as $S_c^{3-\beta}$. Thus, reducing the threshold from 10 to 2 Jy would be expected to lower any such leakage by a factor $(2/10)^{3-\beta} \approx (0.2)^{0.5} \simeq 0.45$ (a $\sim 55\%$ reduction). We observe no such trend in our data, which argues against multiplicative high-delay leakage from main lobe sources dominating the measured power. By contrast, if we are measuring the cosmological 21 cm signal, the power spectrum should be stable to these mask changes, aside from increased statistical uncertainty from reduced sky area, as observed.

These results indicate that multiplicative high-delay systematics, if present, are subdominant at our sensitivity and do not materially affect the measured power spectrum.

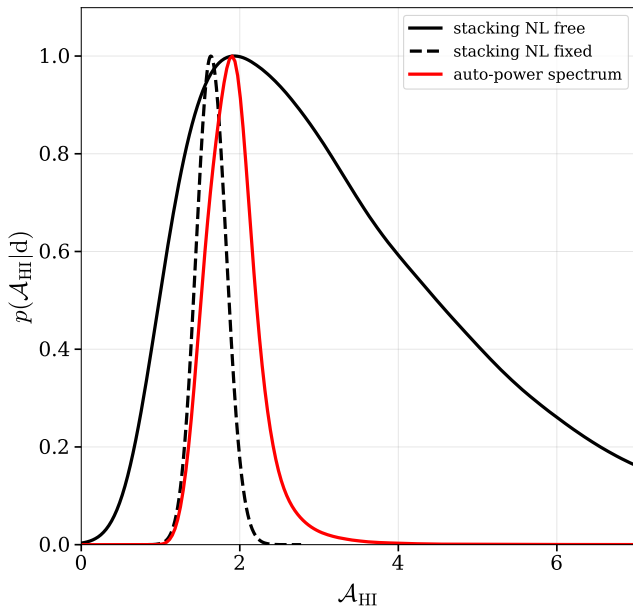


Figure 23. The posterior of the effective clustering amplitude parameter ($\mathcal{A}_{\text{HI}}$) derived from stacking on the QSO catalog from eBOSS. We show the constraint for both models, where all parameters are allowed to vary, and when the nonlinear parameters (NL) are fixed to their fiducial values (same as CHIME Collaboration et al. 2023). We also show the constraint on the power spectrum amplitude parameter derived from the autocorrelation measurement by taking the square root of the $\mathcal{A}_{\text{HI}}^2$ (see Equation (78)). We find that the derived amplitude parameters from cross-correlation and autocorrelation analyses are consistent with each other within 1σ .

7.10. Consistency of H I Clustering Amplitude between Autocorrelation and Cross Correlation

As mentioned in Section 4, we perform the stacking analysis on the same data used for the auto power spectrum measurement. We stack the masked foreground-filtered map on 33,119 QSOs from the eBOSS clustering catalog within the redshift range of 1.01–1.34 (608–708 MHz), corresponding to the sky region analyzed in our auto power spectrum measurement. We detect the 21 cm signal in this cross correlation with QSOs, with detection significance of 9.1σ .

Following the procedure of CHIME Collaboration et al. (2023), we use this cross-correlation measurement to constrain the effective clustering amplitude of neutral hydrogen, defined as $\mathcal{A}_{\text{HI}} \equiv 10^3 \Omega_{\text{HI}}(b_{\text{HI}} + \langle f\mu^2 \rangle)$. We find $\mathcal{A}_{\text{HI}} = 1.93^{+2.50}_{-0.93}$ when nonlinear parameters are free to vary, and $\mathcal{A}_{\text{HI}} = 1.63^{+0.19}_{-0.18}$, when those parameters are fixed to their fiducial values. We then compare this effective clustering amplitude parameter with the value we derived from the auto power spectrum measurement (see Equation (78)).

Figure 23 compares the marginalized posterior distributions of $\mathcal{A}_{\text{HI}}$ derived from both the stacking analysis and the auto power spectrum analysis. We take the square root of the power spectrum amplitude parameter ($\mathcal{A}_{\text{HI}}^2$) for one-to-one comparison. The two measurements are statistically consistent within 1σ . This consistency provides strong evidence against contamination from residual systematics. Systematic effects that are uncorrelated between CHIME and eBOSS, such as residual foreground leakage, RFI, or other instrumental artifacts, would bias the autocorrelation measurement but not the cross correlation. Such contamination would result in a significantly larger amplitude in the auto power spectrum relative to the stacking measurement. The observed agreement

between the two independent analyses therefore demonstrates that there is no residual systematic contamination present in our data.

Furthermore, we find that the marginalized posterior on $\mathcal{A}_{\text{HI}}$ is much narrower for the auto power spectrum than for the stacking analysis with all model parameters free to vary (nonlinear-free). For diagnostic purposes, we also consider a variant of the stacking model in which the nonlinear parameters are fixed to fiducial values (NL-fixed).²⁸ In this case, the width of the stacking posterior becomes comparable to that for the auto power spectrum, highlighting the extent to which degeneracies between $\mathcal{A}_{\text{HI}}$ and the nonlinear parameters contribute to the broadening of the NL-free stacking constraint.

Several factors likely contribute to the tighter constraints obtained from the auto power spectrum analysis relative to the NL-free stacking analysis. First, the auto power spectrum detection has a higher overall S/N than the stacking measurement (13.1σ versus 9.1σ). Second, the auto power spectrum scales quadratically with the amplitude parameter, $P_{\text{HI}} \propto \mathcal{A}_{\text{HI}}^2$, whereas the stacked signal (or cross power spectrum) scales linearly, $P_{\text{HI,g}} \propto \mathcal{A}_{\text{HI}}$, so at fixed noise covariance the auto measurement is intrinsically more sensitive to $\mathcal{A}_{\text{HI}}$. Third, the stacking analysis must marginalize over a larger set of nuisance parameters, including the linear bias of the QSO catalog b_g (which is well constrained), the shot-noise contribution parameterized by M_{10} (which is not), and two separate fingers-of-God (FoG) parameters (compared to a single FoG parameter for the auto power spectrum). This larger nuisance-parameter space leads to more severe prior volume effects on $\mathcal{A}_{\text{HI}}$. Finally, although both analyses average over modes to form a 1D statistic, in the current spectral stacking implementation most of the signal to noise is concentrated at small line-of-sight separations, so only a small number of broad, highly overlapping linear combinations of $k_{\parallel}$ are well constrained. By contrast, the power spectrum estimator retains several k bins (dominated by different ranges of $k_{\parallel}$) with more distinct responses to the model parameters, providing additional effective scale information that helps to break parameter degeneracies.

7.11. Linearity

We have designed the pipeline that converts measured visibilities into foreground-filtered maps to be linear in the data wherever possible. For a fixed linear pipeline, the 21 cm signal, foregrounds, and noise propagate independently. The effect of the pipeline on the cosmological 21 cm signal can therefore be characterized by applying the same operations to signal-only simulations. This captures the signal attenuation and mode mixing induced by the pipeline, without requiring the simulations to reproduce the particular 21 cm realization present in the data or its chance correlations with foregrounds.

This is the main distinction between our approach and data-driven foreground cleaning methods, where the modes removed from the maps are inferred from the observed sky. In such methods, the filter can depend on chance correlations between the cosmological signal and the foregrounds, so applying the final filter to signal-only simulations does not

²⁸ As in CHIME Collaboration et al. (2023), the mean of the NL-fixed posterior is determined by the fiducial values of the nonlinear parameters and should not be interpreted as an estimate of the true $\mathcal{A}_{\text{HI}}$, since the actual values of those parameters are unknown.

necessarily capture the full signal loss. The exceptions in our pipeline are several steps where masks, bandpass corrections, and noise crosstalk estimates are derived from the data. These steps are formally nonlinear, and the relevant question is whether the data-defined choices have appreciable sensitivity to the cosmological 21 cm signal. We discuss these cases below.

RFI masking. The RFI masks described in Section 3.1, Section 3.4, and Appendix A are derived from the data, and are therefore formally nonlinear. This nonlinearity would affect the signal transfer only if the mask selection statistics had appreciable sensitivity to the cosmological 21 cm signal. The methods used to mask the majority of RFI are based on thresholding statistics in which steady celestial emission is removed or strongly suppressed. Since the cosmological 21 cm signal is part of the steady sky, these same operations also suppress the 21 cm contribution before thresholding. The radiometer mask uses even-odd differences of 30 ms samples, over which the sky contribution is effectively constant and cancels to first order. The fringe-rate mask thresholds residual power above the maximum fringe rate of steady celestial sources, where neither foregrounds nor 21 cm emission contribute significantly.

The remaining RFI masks are not constructed from a subspace orthogonal to the 21 cm signal, since they are based on a noise-normalized χ^2 statistic formed from high-delay residuals. However, these masks remove only 6% of the data beyond the masks described above, and the masked samples are primarily associated with single-day outliers rather than persistent sky structure. In this statistic, the 21 cm contribution scales as its signal-to-noise squared, averaged over baselines, which is far below the thermal noise scatter. The thresholded features are instead high-significance excursions localized in time or frequency, unlike the low-amplitude stochastic fluctuations expected from the cosmological 21 cm field and consistent with residual RFI. We therefore do not expect the RFI mask to be biased appreciably by the cosmological signal. Once determined from the data, the masks are held fixed and applied to the simulated 21 cm realizations, so the effect of the missing samples is included in the forward model.

HyFoReS bandpass correction. The HyFoReS bandpass correction described in Section 3.3 is the step most closely related to data-driven foreground cleaning. A potential concern is that correlations between the template and the 21 cm field could introduce a component of the cosmological signal into the subtracted leakage model. The possible signal loss is limited by the construction of the template and by the low dimensionality of the fit. The template excludes the modes used in the power spectrum analysis, so the coefficient estimate contains no direct self-correlation of the high-delay 21 cm field. Any remaining signal dependence must therefore arise either from correlations between the high-delay 21 cm field and the foreground structure in the template, or from correlations between low- and high-delay 21 cm modes. The latter contribution is further suppressed by the small signal-to-foreground ratio in the template. Moreover, at each frequency, HyFoReS fits only the amplitude of this single angular template across the full field. The possible signal loss is therefore restricted to the component of the cosmological field that follows this specific foreground-dominated angular pattern. As an empirical check, we repeat the quasar-stacking analysis described in Section 4, with the HyFoReS correction

omitted from the pipeline. We find no appreciable change in the recovered stacked signal between the two versions of the processing.

Absorber masking. The absorber mask described in Section 3.8 is also data-dependent, since the masked regions correspond to outliers identified in the maps. These candidates are localized both angularly and spectrally, appearing as unresolved sources confined to a narrow range of frequencies. This morphology is distinct from the cosmological 21 cm field. The candidates are also individually detected at high significance ($S/N > 7$), whereas the 21 cm fluctuations implied by the measured power spectrum are far below the per-voxel noise level. Finally, the mask removes only 1.35% of the map voxels. We tested the possible signal-dependent bias from this type of selection using simulated maps containing both 21 cm signal and noise. With a 5σ threshold, more aggressive than the 7σ threshold used to identify the absorber candidates, we find no appreciable change in the recovered power spectrum; signal suppression becomes apparent only when the clipping threshold is lowered further. This test therefore finds no detectable selection-induced signal loss at the threshold used in the analysis. Once defined, the same mask is applied to the simulated 21 cm realizations, so the effect of the missing voxels is included in the forward model.

Noise crosstalk removal. The noise crosstalk removal described in Section 3.5.2 estimates and subtracts a constant additive offset using a weighted median over a reference range in local ERA. Although the median is a nonlinear estimator, any cosmological signal included in the estimate is subtracted as a constant in local ERA and is therefore confined to the $m = 0$ mode. This mode is not used in the power spectrum measurement, since the intracylinder baselines that provide sensitivity to the lowest m -modes are excluded from the mapmaking. The crosstalk subtraction can therefore remove only cosmological modes outside the region probed by this analysis.

8. Conclusions

In this paper, we have presented the first detection of the cosmological 21 cm intensity mapping signal in autocorrelation at $z \sim 1$ using data from CHIME. Using 94 nights of observation from January through November 2019, we have measured the 21 cm auto power spectrum over a frequency range of 608.2–707.8 MHz, corresponding to a mean redshift of $z \sim 1.16$ ($z = 1.34$ – 1.01), with a statistical significance of 12.4σ . We have also independently measured the 21 cm signal in two sub-bands, obtaining 9.1σ and 8.6σ detections at mean redshifts of $z \sim 1.24$ and $z \sim 1.08$, respectively. This result marks a significant milestone, moving beyond reliance on cross correlation with external survey data and demonstrating that the 21 cm auto power spectrum can be measured directly at these redshifts.

This detection was made possible by substantial advancements in our data processing pipeline, specifically designed to overcome the immense challenge of separating the faint cosmological signal from astrophysical foregrounds that are orders of magnitude brighter. Key methodological advances include (1) novel RFI detection and masking algorithms combining radiometer tests, fringe-rate filtering, and spectral filtering methods (Section 3.1); (2) implementation of achromatic beamforming techniques to maintain consistent

synthesized beams across frequency (Section 3.2.2); (3) application of foreground filtering before time averaging to minimize spectral leakage (Section 3.2.3); and (4) deployment of the HyFoReS algorithm to correct residual bandpass errors (Section 3.3). These developments form part of a broader effort to extract 21 cm auto power spectra from interferometric data and may inform analyses with other low- and high-redshift 21 cm experiments, including CHORD (K. Vanderlinde et al. 2019), Tianlai (J. Li et al. 2020; F. Wu et al. 2021), HIRAX (D. Crichton et al. 2022), MeerKat (S. Paul et al. 2026), uGMRT (A. Chakraborty et al. 2021), the Ooty Wide Field Array (C. R. Subrahmanya et al. 2017), HERA (D. R. DeBoer et al. 2017), Murchison Widefield Array (S. J. Tingay et al. 2013), LOFAR (M. P. van Haarlem et al. 2013) and the upcoming Square Kilometre Array telescopes.

The robustness of our detection has been established through a comprehensive suite of validation tests. Our validation framework includes (1) the imaginary component of the power spectrum is consistent with noise, indicating the absence of phase calibration errors; (2) temporal stability through even-odd day comparisons; (3) the power spectrum derived from Stokes Q data is also consistent with noise, ruling out any significant polarized foreground leakage; (4) the consistency of the signal across independent R.A. bins rules out R.A.-dependent systematics; (5) the consistency of the signal across independent decl. bins rules out any decl.-dependent systematics; (6) the consistency of the signal across different baseline configurations (1-cylinder versus combined (1+2+3)-cylinder baselines) rules out baseline-dependent foreground contamination; (7) the consistency of the signal across independent sub-bands rules out frequency-dependent systematic contamination; (8) the signal remains consistent with varying spatial track masks and rules out foreground leakage through sidelobes of the beam; and (9) the consistency of the signal with varying transit masks rules out the narrowband systematic contamination.

We also briefly discuss the theoretical interpretation of the measurements (Section 6.2), and the detailed modeling of the 21 cm signal is outlined in our companion paper (CHIME Collaboration et al. 2026a). Given the strong degeneracy between the cosmic H I abundance (Ω_{HI}) and the linear bias of H I (b_{HI}), we constrain their combination through the amplitude parameter: $\mathcal{A}_{\text{HI}}^2 = 10^6 \Omega_{\text{HI}}^2 (b_{\text{HI}}^2 + \langle f\mu^2 \rangle)^2$, where $\langle f\mu^2 \rangle$ is a factor capturing the angular average of the Kaiser redshift-space distortion term over the wavenumbers we have measured. We find that $\mathcal{A}_{\text{HI}}^2 = 3.55^{+0.96}_{-1.32}(\text{stat.}) \pm 0.61(\text{sys.})$ for the full frequency band ($z \sim 1.16$), $\mathcal{A}_{\text{HI}}^2 = 2.47^{+1.59}_{-1.08}(\text{stat.}) \pm 0.43(\text{sys.})$ for the upper sub-band ($z \sim 1.08$), and $\mathcal{A}_{\text{HI}}^2 = 3.24^{+6.79}_{-2.11}(\text{stat.}) \pm 0.56(\text{sys.})$ for the lower sub-band ($z \sim 1.24$). These constraints are dominated by statistical uncertainties from the limited signal to noise of our current measurement. In CHIME Collaboration et al. (2026a), we also compare our measurements with predictions from the IllustrisTNG hydrodynamical simulations.

As an additional validation test, we verify that the amplitude parameter constraint from the auto power measurement is consistent with the analogous amplitude constraint from a cross correlation of the same dataset with eBOSS quasars. This provides strong evidence against contamination from systematics that are uncorrelated between CHIME and external large-scale structure tracers. The auto power-based constraint is significantly tighter than the cross-correlation-based constraint, which can be attributed to several factors: the auto power

signal to noise is higher than that of the cross correlation; the auto power scales quadratically with $\mathcal{A}_{\text{HI}}$ while the cross correlation only scales linearly; the cross-correlation signal model involves a larger number of nuisance parameters; and the auto power retains more scale-dependent information than our stacking-based cross-correlation method.

The CHIME archive now contains over 7 yr of data. Incorporating these additional observations should improve the statistical precision of the measurement, enabling tighter constraints on the abundance and clustering of H I and their evolution with redshift. Parallel efforts are also underway to extend the analysis to the lower-frequency portion of the CHIME band (400–600 MHz), corresponding to redshifts $z \sim 1.4$ –2.5.

The present analysis is sensitive to scales $k_{\perp} \gtrsim 0.08 h \text{ Mpc}^{-1}$ and $k_{\parallel} \gtrsim 0.35 h \text{ Mpc}^{-1}$ at a central redshift of $z \sim 1.16$. The lower $k_{\perp}$ limit is set by the exclusion of intracylinder baselines, while the lower $k_{\parallel}$ limit is set by the 200 ns delay cutoff used in the foreground filter. With these analysis choices, sensitivity to the linear modes relevant for baryon acoustic oscillation (BAO) measurements is largely removed, while the retained modes enable a high-S/N measurement on nonlinear scales, $0.4 h \text{ Mpc}^{-1} \lesssim k \lesssim 1.5 h \text{ Mpc}^{-1}$. The BAO feature appears at multiples of the characteristic spacing $k_{\text{BAO}} = 2\pi/r_{\text{drag}} \approx 0.064 h \text{ Mpc}^{-1}$ (see also Section 5.1 and Figure 14 of CHIME Collaboration et al. 2023).

Recovering a broader range of modes will likely require improved calibration of the primary beam, including its far-sidelobe and polarization response, together with improved methods for modeling and removing foreground contamination. Extending the measurement to lower $k_{\perp}$ will additionally require better subtraction of noise crosstalk on the intracylinder baselines. In parallel, continued development of real-time RFI masking algorithms should reduce the fraction of future observations lost to RFI.

As CHIME continues operations and improvements to instrumental calibration and data processing are implemented, we anticipate increasingly precise constraints on the abundance and clustering of neutral hydrogen over a broader range of redshifts and spatial scales, as well as on the cosmological parameters governing large-scale structure.

Acknowledgments

We thank the Dominion Radio Astrophysical Observatory (DRAO), operated by the National Research Council Canada, for their gracious hospitality and expertise. The DRAO is situated on the traditional, ancestral, and unceded territory of the Syilx Okanagan people. We are fortunate to live and work on these lands.

CHIME is funded by grants from the Canada Foundation for Innovation (CFI) 2012 Leading Edge Fund (project 31170), the CFI 2015 Innovation Fund (project 33213), and by contributions from the provinces of British Columbia, Québec, and Ontario. Long-term data storage and computational support for analysis are provided by WestGrid,²⁹ SciNet,³⁰ and the Digital Research Alliance of Canada,³¹ and we thank their staff for flexibility and technical expertise, which has

²⁹ <https://www.westgrid.ca/>

³⁰ <https://www.scinethpc.ca/>

³¹ <https://www.alliancecan.ca/>

been essential to this work, particularly Martin Siegert, Lixin Liu, and Lance Couture.

Additional support was provided by the University of British Columbia, McGill University, and the University of Toronto. CHIME also benefits from NSERC Discovery Grants to several researchers, funding from the Canadian Institute for Advanced Research (CIFAR), and from the Dunlap Institute for Astronomy and Astrophysics at the University of Toronto, which is funded through an endowment established by the David Dunlap family. This material is partly based on work supported by the NSF through grant Nos. 2008031, 2006911, 2006548, 2510770, 2510771, 2510772, and 2510773, and by the Perimeter Institute for Theoretical Physics, which in turn is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Research and Innovation. M.D. is supported by a CRC Chair, NSERC Discovery Grant, CIFAR, and by the FRQNT Centre de Recherche en Astrophysique du Québec (CRAQ). U.P. is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) (funding reference Nos. RGPIN-2019-06770, ALLRP 586559-23, RGPIN-2025-06396), Canadian Institute for Advanced Research (CIFAR), Ontario Research Fund-Research Excellence (ORF-RE Fund, 72074697), AMD AI Quantum Astro. A.C. and A.O. are supported by the NSERC Brockhouse Prize. J.M.P. acknowledges the support of an NSERC Discovery Grant (No. RGPIN-2023-05373). We acknowledge the support of NSERC (funding reference No. 569654). We thank the Beus Center for Cosmic Foundations at Arizona State University for supporting a workshop where part of this work was completed. K.W.M. holds the Adam J. Burgasser Chair in Astrophysics.

We thank the Sloan Digital Sky Survey (SDSS) and eBOSS collaborations for publicly releasing the galaxy and quasar catalogs and supporting mock catalogs used in this work. Funding for the SDSS-IV has been provided by the Alfred P. Sloan Foundation, the US Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High Performance Computing at the University of Utah. The SDSS website is at www.sdss.org.

Software: CAMB (A. Lewis et al. 2000, 2025a), caput (J. R. Shaw et al. 2025b), ch_pipeline (J. R. Shaw et al. 2025c), cora (J. R. Shaw et al. 2025a), Cython (S. Behnel et al. 2011), draco (J. R. Shaw et al. 2025d), driftscan (J. R. Shaw et al. 2025e), emcee (D. Foreman-Mackey et al. 2013), FFTW (M. Frigo & S. G. Johnson 2005), GetDist (A. Lewis 2025; A. Lewis et al. 2025b), h5py (A. Collette et al. 2023), hankl (M. Karamanis & F. Beutler 2021), HDF5 (The HDF Group 1997), healpy (A. Zonca et al. 2019), kotekan (A. Renard et al. 2021), Matplotlib (J. D. Hunter 2007; The Matplotlib Development Team 2025), mpi4py (L. Dalcin & Y.-L. L. Fang 2021; M. Rogowski et al. 2023), NumPy (C. R. Harris et al. 2020), OpenMPI (E. Gabriel et al. 2004), SciPy (P. Virtanen et al. 2020), zarr (A. Miles et al. 2020), Skyfield (B. Rhodes 2019).

Appendix A

Algorithms for RFI Detection and Masking

A.1. Radiometer Test Method

The first step in our RFI excision procedure is to construct a test statistic that measures deviations in the measured variance

of the visibilities from the expectation for Gaussian, uncorrelated noise. The correlator begins by applying a four-tap polyphase filterbank to the raw voltage timestream from each feed, in which each $2.56 \mu\text{s}$ block of 2048 samples (at 1.25 ns spacing) is transformed into 1024 complex frequency channels. These channelized samples, produced at a $2.56 \mu\text{s}$ cadence, form the baseband data. They are spatially cross-correlated to generate visibilities, which are accumulated to a cadence of 30 ms. At this stage, two intermediate quantities are accessible internally as data move from GPU to CPU: the unnormalized visibilities $v_{fab\tau}$ and the corresponding number of accumulated baseband samples $n_{f\tau}$. Note that the number of samples need not be constant across time or frequency, as it can be reduced by packet loss in network transport or by real-time RFI excision. When averaging these 30 ms visibilities into coarse 9.9404 s integrations, the correlator also computes variance statistics from the same fine-resolution products, as described below.

Let M denote the number of fine samples per coarse integration ($M \simeq 332$), and define the fine-sample index set for coarse time t as

$$\mathcal{I}_t \equiv \{ \tau \in \mathbb{Z} : M t \leq \tau < M(t+1) \}.$$

Let $\mathcal{E}_t \subset \mathcal{I}_t$ denote the even offsets in this interval,

$$\mathcal{E}_t \equiv \{ \tau \in \mathcal{I}_t : (\tau - M t) \bmod 2 = 0 \},$$

so that for every $\tau \in \mathcal{E}_t$, the pair $(\tau, \tau + 1)$ gives one even-odd pair of fine samples. The correlator accumulates four quantities:

$$v_{fabt} \equiv \sum_{\tau \in \mathcal{I}_t} v_{fab\tau}, \quad (\text{A1})$$

$$N_{ft} \equiv \sum_{\tau \in \mathcal{I}_t} n_{f\tau}, \quad (\text{A2})$$

$$\Delta v_{fabt}^2 \equiv \sum_{\tau \in \mathcal{E}_t} |v_{fab,\tau+1} - v_{fab\tau}|^2, \quad (\text{A3})$$

$$\Delta n_{ft}^2 \equiv \sum_{\tau \in \mathcal{E}_t} (n_{f,\tau+1} - n_{f\tau})^2. \quad (\text{A4})$$

From these quantities, it then outputs for each coarse integration an estimate of the normalized visibilities,

$$V_{fabt} \equiv \frac{v_{fabt}}{N_{ft}}, \quad (\text{A5})$$

and an estimate of the variance of the visibilities,

$$\sigma_{fabt}^2 \equiv \frac{1}{N_{ft}^2} \left[\Delta v_{fabt}^2 - \frac{\Delta n_{ft}^2}{N_{ft}^2} |v_{fabt}|^2 \right], \quad (\text{A6})$$

where the second term in Equation (A6) removes the bias that arises when the even and odd accumulations are formed from unequal numbers of baseband samples. The 30 ms cadence is short enough that the sky visibility is approximately constant across each fine-sample pair, and therefore cancels in the difference $v_{fab,\tau+1} - v_{fab\tau}$, leaving a combination of thermal noise and RFI. The expected variance for Gaussian, uncorrelated noise can be obtained directly from the autocorrelations of each feed. Specifically, for visibility V_{fabt} between feeds a

and b ,

$$\text{Var}\{V_{fabt}\} = \frac{V_{faat} V_{fbbt}}{N_{ft}}, \quad (\text{A7})$$

where V_{faat} and V_{fbbt} are the autocorrelations. Any measured excess relative to this prediction indicates a departure from the Gaussian, uncorrelated noise assumption and is taken as evidence for RFI.

Because the array contains relatively short baselines, RFI tends to appear coherently across them, whereas the receiver noise contributions remain independent. Averaging the variance estimates over all baselines within a given polarization product, therefore, enhances sensitivity to RFI while suppressing uncorrelated noise fluctuations. The radiometer test statistic ϱ_{pft} is defined as the square root of the ratio of the baseline-averaged measured variance to the baseline-averaged, autocorrelation-based prediction,

$$\varrho_{pft} \equiv \left(\frac{\sum_{ab \in \mathcal{B}_p} \sigma_{fabt}^2}{1N_{ft} \sum_{ab \in \mathcal{B}_p} V_{faat} V_{fbbt}} \right)^{1/2}, \quad (\text{A8})$$

where $\mathcal{B}_p$ denotes the set of all baselines formed from input pairs (a, b) whose polarizations correspond to the polarization product $p \in \{XX, XY, YY\}$.

Figure A1 (top panel) shows the radiometer test statistic for the XY polarization over a full sidereal day. The measured variance is consistently at least 2.3% larger than the expected variance from the radiometer equation, a behavior observed across days and polarizations whose origin remains unknown. The autocorrelation-based normalization removes nearly all large-scale variations associated with the instrumental band-pass and changing system temperature, but residual variations at the $\sim 0.05\%$ level persist and track the transit of the Galactic plane. The scatter in the statistic also increases during the transit of bright sources (Sun, Cyg A, Cas A, Tau A), whose correlated contributions to the system temperature do not average down across baselines.

Under ideal assumptions of Gaussian noise that is uncorrelated in time and between baselines, the mean of ϱ would be unity and its variance could be computed analytically. In practice, however, correlations between baselines introduced by the sky noise complicate this calculation and would require a detailed sky-dependent model. Rather than attempt to model these effects, we estimate the background and scatter of ϱ directly from the data using robust median filters. For each polarization p , the algorithm then constructs a time-frequency mask by iteratively identifying outliers in ϱ relative to these empirically determined background and scatter levels.

Before the main iterative procedure, two preliminary masks are applied. First, a static frequency mask excludes channels known to be persistently contaminated in the CHIME environment. Second, for each frequency channel, the 15th percentile of ϱ over the sidereal day is computed and then smoothed across frequency with a rolling median of width 191 channels. Channels whose percentile deviates from this baseline by more than 5 times a robust local estimate of scatter (computed using an MAD) scaled by 1.4826, estimated with the same 191 channel window) are masked uniformly in time. This step suppresses channels with persistently abnormal behavior over the sidereal day.

Outlier identification proceeds through a five-iteration process. At the start of each iteration, the mask from the previous iteration is applied so that previously identified outliers are excluded when calculating statistics. A 2D weighted median filter is used to estimate a smooth background of the radiometer test statistic,

$$\mu_{\varrho, pft} \equiv \left[\text{MedFilt}_{w_t, w_f, 1}(\varrho) \right]_{pft}, \quad (\text{A9})$$

where $\text{MedFilt}_{w_t, w_f, 1}$ denotes a median filter with time width $w_t = 181$ samples, frequency width $w_f = 37$ channels, and polarizations treated independently, applied while excluding samples flagged by the current mask $\mathcal{Q}_{pft}$. The local scatter is then estimated with a weighted MAD filter,

$$\sigma_{\varrho, pft} \equiv 1.4826 \times \left[\text{MedFilt}_{w'_t, w'_f, 1}(|\varrho - \mu_{\varrho}|) \right]_{pft}, \quad (\text{A10})$$

with $w'_t = 31$ samples and $w'_f = 101$ channels, again excluding masked samples. The background-subtracted residual is

$$\delta\varrho_{pft} \equiv \varrho_{pft} - \mu_{\varrho, pft}, \quad (\text{A11})$$

and the normalized deviation is

$$z_{pft} \equiv \frac{\delta\varrho_{pft}}{\sigma_{\varrho, pft}}. \quad (\text{A12})$$

Thresholds are progressively lowered over the five iterations, so that the brightest excursions are flagged first; this prevents large outliers from biasing the subsequent estimates of $\mu_{\varrho, pft}$ and $\sigma_{\varrho, pft}$.

Figure A1 (middle panel) shows the normalized deviation z after the final iteration for the same sidereal day and baseline. The background variations have been removed, making this metric a highly sensitive indicator of RFI. The CHIME site exhibits a striking diversity of RFI morphologies, including persistent contamination from TV channels and cellular bands, frequent broadband impulses, narrowband carriers with varying duty cycles, extended emission from Meridian satellites in Molniya orbits, and brief 6 MHz wide features produced by distant TV stations reflecting off objects in the sky.

A preliminary MAD mask is first constructed by flagging samples with $|z_{pft}| > \eta^{(k)}$. The threshold begins at $\eta^{(0)} = 25.3125$ and is reduced by a factor of 1.5 at each iteration, ending with $\eta^{(4)} = 5$ on the final pass. This preliminary mask also defines the basis for flagging digital TV channels: within each 6 MHz band beginning at 398 MHz, the fraction of frequency bins with $|z_{pft}| > \eta_{tv}^{(k)}$ is evaluated. If more than half of the bins in a channel exceed this level, the entire TV channel is masked for that time sample. The threshold $\eta_{tv}^{(k)}$ is significantly lower than the $\eta^{(k)}$ threshold, and is chosen such that the false positive rate is equal to that given by the $\eta^{(k)}$ threshold for the single-sample flagging assuming a Gaussian noise model. The union of this MAD-based mask and the TV-band mask is then carried forward as the starting point for a variant of the SumThreshold procedure (A. R. Offringa et al. 2010).

The SumThreshold algorithm operates directly on the background-subtracted residuals $\delta\varrho_{pft}$ (Equation (A11)), while propagating the per-sample variance $\sigma_{\varrho, pft}^2$ (Equation (A10)) through convolutions. For each window length

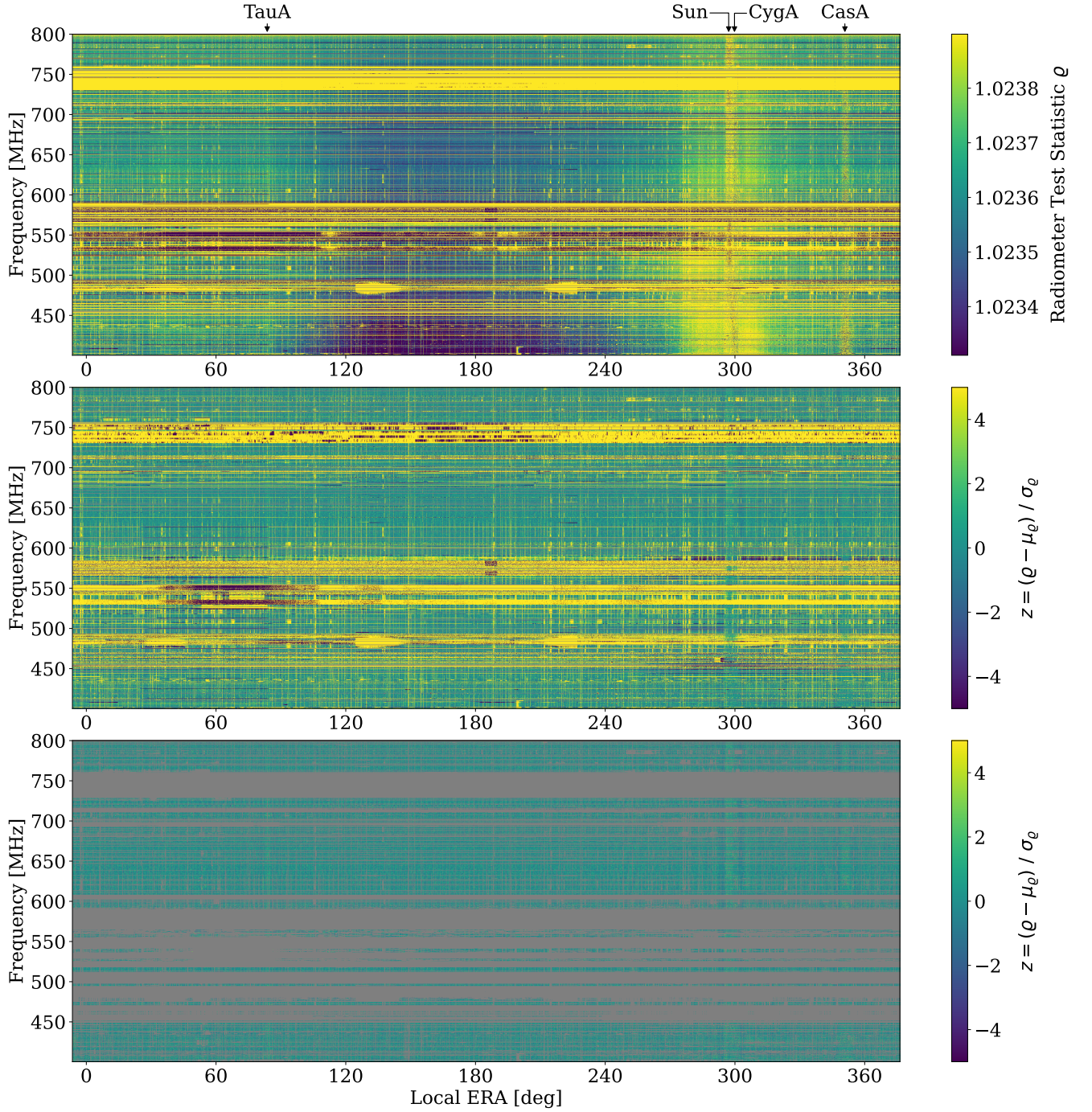


Figure A1. Top: radiometer test statistic ρ (Equation (A8)) as a function of frequency and local ERA for XY polarization baselines on 2019 January 15. Frequencies that are missing due to nonoperational GPU nodes appear in gray. The color scale is offset by 1.023, implying the measured noise is at least 2.3% higher than predicted from the autocorrelations. The mild dependence of ρ on frequency and local ERA tracks the Galactic plane’s transit through the field of view. During transits of bright sources, the scatter in ρ increases due to partial noise correlation across baselines. The displayed dynamic range is narrow ($<0.1\%$), and bright RFI features that clip the scale are often much larger than the maximum value shown. Middle: robust z-score map (Equation (A12)), after subtracting a slowly varying background μ_ρ (Equation (A9)) and normalizing by an MAD-based scatter estimate σ_ρ (Equation (A10)). Bottom: same quantity as the middle panel, with masked times and frequencies displayed in gray. The fraction of data missing or masked relative to the full time–frequency grid is 12% missing, 27% automatically masked by the static persistent RFI mask, and 25% masked by the algorithm, for a total of 64% missing or masked.

$l = 1, 2, 4, \dots, l_{\max}(=64)$, we form

$$D_{l,pft} = [K_l \otimes \delta\rho]_{pft}, \quad (\text{A13})$$

$$\Sigma_{l,pft}^2 = [K_l^2 \otimes \sigma_\rho^2]_{pft}, \quad (\text{A14})$$

where K_l denotes a top-hat kernel of width l , and masked samples are set to zero in both convolutions. A detection is

declared whenever

$$|D_{l,pft}| > \eta^{(k)} \Sigma_{l,pft}. \quad (\text{A15})$$

Windows that trigger a detection are symmetrically expanded to cover all l contributing samples. Convolutions are applied first along the time axis and then along the frequency axis, with masks from earlier passes carried forward to later ones.

This variance-propagating formulation differs from the canonical `SumThreshold` algorithm (A. R. Offringa et al. 2010), which lowers the per-sample threshold for longer windows; here the scaling with l arises naturally from variance propagation. The strength of the `SumThreshold` method lies in its sensitivity to low-level but extended contamination: by summing over contiguous samples, it identifies faint RFI whose aggregate power is inconsistent with Gaussian noise even when no individual sample exceeds threshold. The use of rectangular windows ensures that the flagged regions naturally match the morphology of broadband or time-extended RFI.

After `SumThreshold`, the algorithm blends the MAD and `SumThreshold` masks depending on the local ERA. During transits of bright sources (within one beamwidth of Cas A, Cyg A, Tau A, and B0329+54, and within three beam widths of the Sun), the MAD mask is preferred to avoid overflagging astrophysical signals broadened by the primary beam. At all other times, the `SumThreshold` mask is used. For times near source transits, the mask is expanded with a 1D scale-invariant rank operator (A. R. Offringa et al. 2012b) applied along the time axis to prevent streak-like RFI from being underflagged during transits; the default aggressiveness is $\eta = 0.2$. Unlike the general scale-invariant rank option, this restricted version is applied only in the time direction and only during transit intervals.

The background and scatter are recomputed after each iteration using the updated mask, and the process is repeated for a total of five iterations. The resulting mask Q_{pft}^{rad} for each polarization is then combined with a logical OR to produce the global time-frequency mask Q_{ft}^{rad} . Figure A1 (bottom panel) shows this mask overlaid on the normalized deviation z . This final mask incorporates the static exclusions and the 1D outlier test, as well as the MAD, digital TV band, and `SumThreshold` detections, and is used to excise contaminated data from subsequent analysis.

A.2. Fringe-rate Filtering Method

The second RFI flagging step targets transient RFI based on the assumption that this type of RFI has an impulse-like time-conjugate Fourier response—that is, power from transient RFI is expected to appear in more rapidly varying Fourier modes than the background sky, whose contribution should be limited by the fringe pattern of the instrument. Although not all signals with this Fourier behavior are necessarily caused by RFI, we do not have to differentiate for this particular analysis.

In order to reduce the volume of data that we need to filter, we compute and use only the Stokes I polarization parameter for each baseline pair (a, b) :

$$V_{fabt}^I = V_{fabt}^{XX} + V_{fabt}^{YY}. \quad (\text{A16})$$

Empirically, we see no significant differences in the resulting mask if all polarization pairs are included.

An approximate sky map is produced using a fast Fourier transform (FFT) across the flattened baseline axis, ordered based on two levels of sorting: first, by signed NS feed pair separation, and second by increasing EW cylinder separation:

$$\hat{V}_{fkt}^I = \text{FFT}\{V_{f(ab)t}^I\}. \quad (\text{A17})$$

This process is significantly more computationally efficient compared to forming a true ringmap, where each frequency is first beamformed along the NS telescope axis using a weighted discrete Fourier transform, followed by a weighted FFT in the EW direction. The omission of baseline weighting and the process of concatenating all EW separations in Equation (A17) produces an incorrect map—the formed beams do not exactly map to the declinations expected from the telescope layout, and the measured intensity of the sky is incorrect. These errors, however, are irrelevant in the context of this process; we simply wish to localize RFI that may not be otherwise visible in the visibilities prior to flagging, and the beam axis is not included in the final mask.

As with the radiometer test method, we first apply a static frequency mask to exclude channels known to be consistently bad. For each frequency f and beam k , the background sky is estimated by constructing and applying a low-pass filter with cutoff μ_c set by the fringe rate of the longest EW CHIME baseline b_{EW} ,

$$\mu_c = \frac{f}{c} \frac{b_{\text{EW}}}{\cos \Lambda}, \quad (\text{A18})$$

where Λ is the latitude of the telescope. The filter uses a flat-top window and order equal to μ_s/μ_c , where $\mu_s = T^{-1}$ for sampling rate $T \approx 10$ s. We divide the filtered data by the filtered mask in order to normalize for samples that were masked prior to this stage. The low-pass filtered background is then subtracted from the unfiltered data

$$\bar{V}_{ft}^I = \frac{M_{ft}^I \hat{V}_{ft}^I \otimes K_{\mu_c}}{M_{ft}^I \otimes K_{\mu_c}}, \quad (\text{A19})$$

$$\delta \hat{V}_{ft}^I = \hat{V}_{ft}^I - \bar{V}_{ft}^I, \quad (\text{A20})$$

where M_{ft}^I is the prior data mask.

Outlier detection is done independently for each frequency using a robust MAD metric derived from the background-subtracted data using the method described in Equations (A9)–(A12). An example of the initial approximate sky map, the background-subtracted map, and the resulting MAD metric is shown in the top, middle, and bottom panels of Figure A2, respectively. In this case, the 1D windows for background estimation (Equation (A9)) and scatter estimation (Equation (A10)) have time widths $w_t = 101$ samples and $w_t = 51$ samples, respectively. Rather than using a single threshold, however, flagging is done based on a dual-threshold criterion: a sample is flagged if it exceeds some high threshold or if it both exceeds a lower threshold and is connected (in an adjacency context) to a sample which exceeds the higher threshold. This approach proves to be more robust to overflagging than using a single threshold, while consistently flagging RFI events spanning multiple time samples in their entirety.

The final time-frequency mask $Q_{ft}^{\text{t-filt}}$ is obtained by flagging a frequency time if the fraction of flagged beams for that sample exceeds some threshold τ ,

$$Q_{ft}^{\text{t-filt}} = \Theta\left(\frac{1}{n_k} \sum_k Q_{fkt}^{\text{t-filt}} - \tau\right), \quad (\text{A21})$$

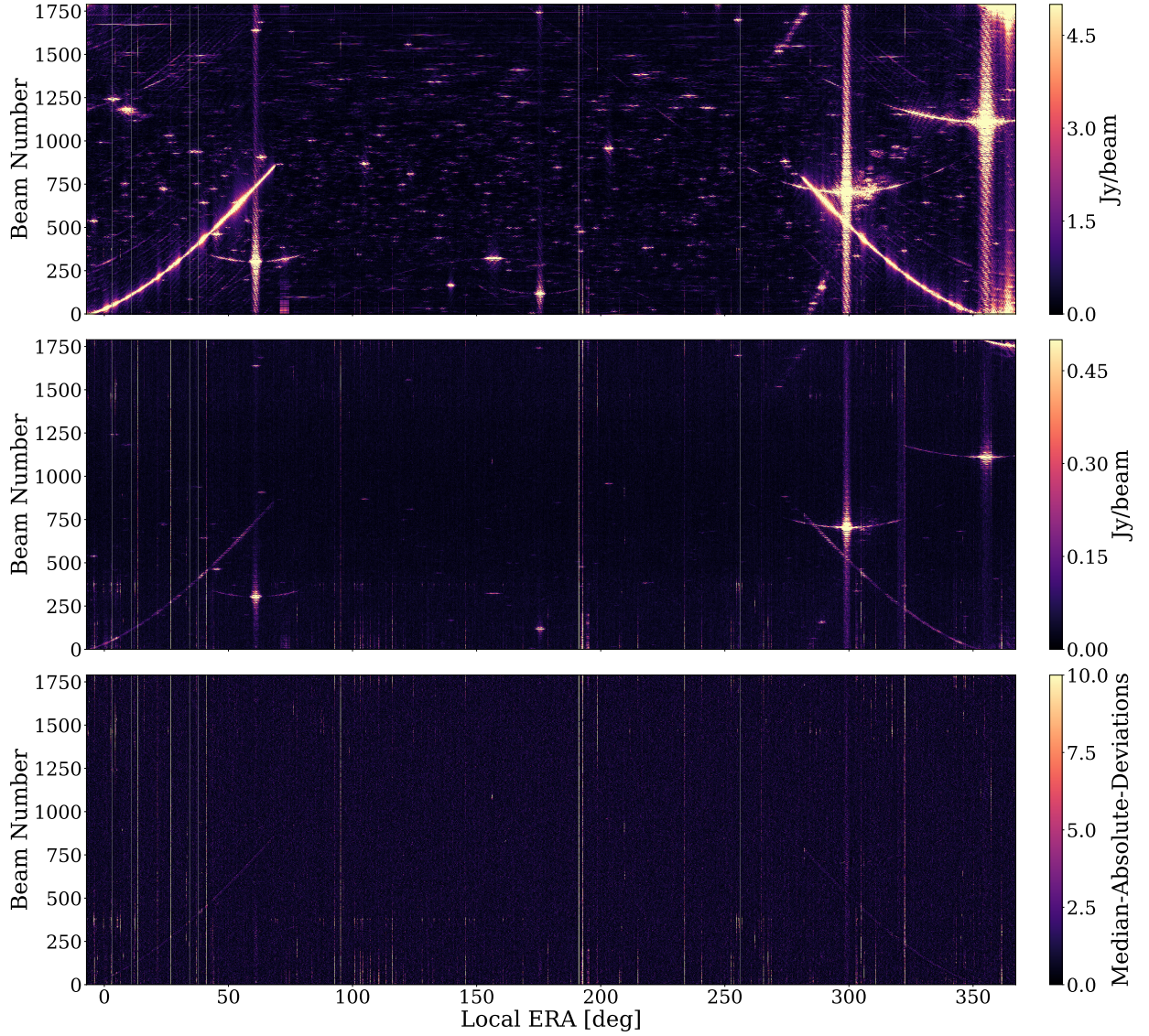


Figure A2. Top: map produced using Equation (A17) from the sum of XX- and YY-polarization intercyylinder baselines on 2021 March 25 at 706.25 MHz. Time samples that were flagged in real time appear as thin gray lines. RFI features can be seen as bright vertical lines that do not clearly correspond to a bright source and are most frequently seen near the horizons. The flagging algorithm uses a map that includes intracylinder baselines—these are omitted here in order to make RFI and point sources more visible for demonstration purposes. Middle: map produced using all baselines, after subtracting a low-pass filtered background $\bar{V}_{ft}^I$ (Equation (A19)). Transient RFI is strongly visible against the residual sky. Additional RFI is visible, originating from intracylinder baselines not included in the top panel. Bottom: robust MADs map derived from the middle panel using the method described in Equations (A9)–(A12).

where n_k is the total number of beams, $\Theta(\cdot)$ is the Heaviside function, and Q_{fkt}^{t-filt} is the beam-dependent mask. A fraction $\tau = 0.01$ was empirically determined to adequately flag RFI events while minimizing false positives. Similarly to the variance-

based mask, this mask is used to excise contaminated data from the analysis. The effectiveness of the background subtraction and final masking is illustrated in Figure (A3) for a six-hour interval containing contributions from Cyg A, Cas A, and the Sun.

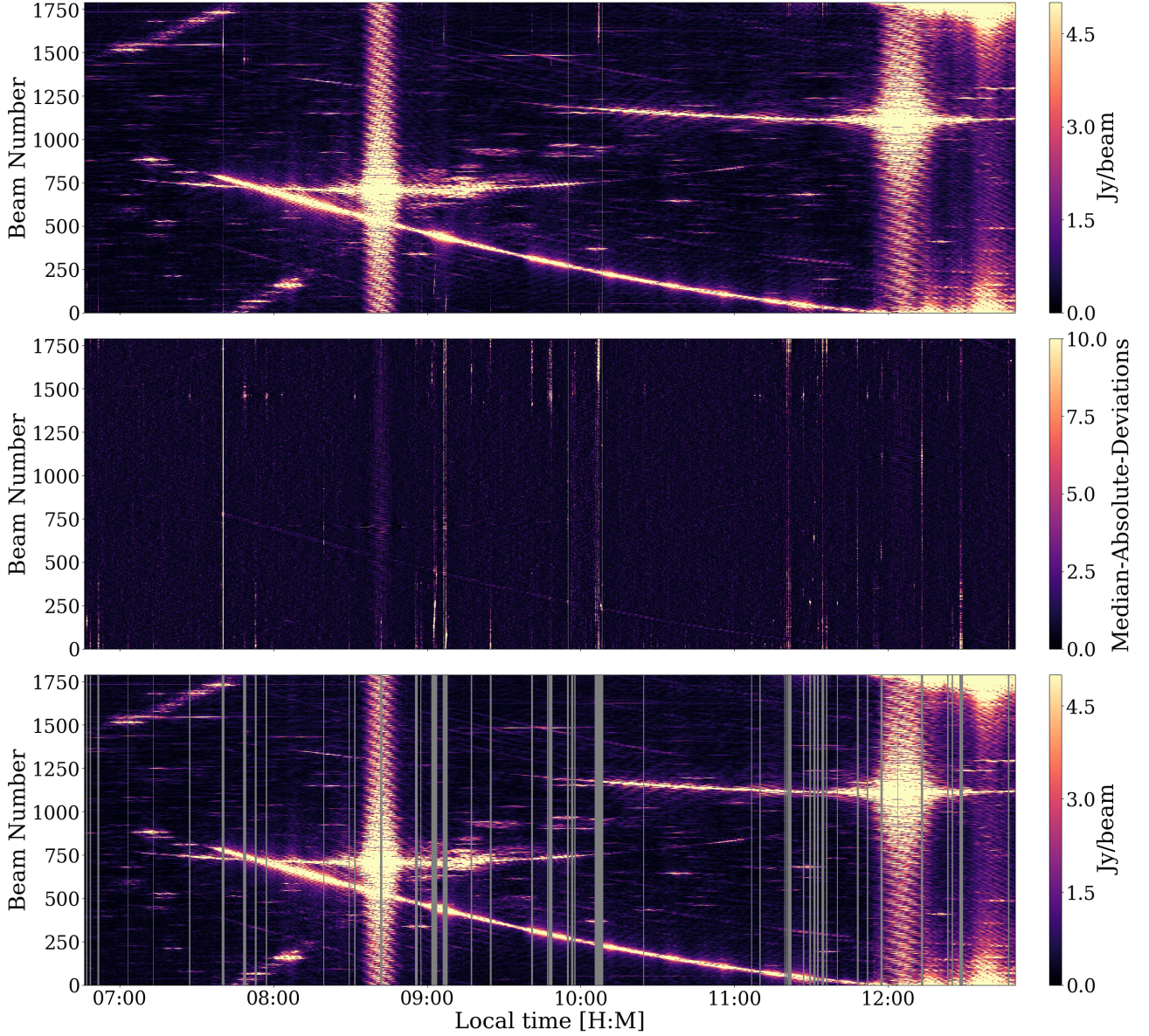


Figure A3. Top and middle: A 6 hr slice of the map and corresponding MAD metric shown in Figure A2. A time range during which Cyg A, Cas A, and the Sun are chosen to demonstrate the effectiveness of the background subtraction process. Bottom: the same map shown in the top panel, after application of the final RFI mask produced using Equation (A21), which is reduced across beams.

A.3. Spectral Filtering Method

Our final step examines whether, after applying the previously determined masks, the data are consistent with the expected level of thermal noise at high delays. The purpose of this step is to isolate residual RFI and other nonthermal signals by applying an aggressive high-pass filter in delay that removes the contribution from the smooth radio sky. Foregrounds are spectrally smooth, whereas RFI typically exhibits sharp spectral features; applying an aggressive high-pass filter along frequency removes the foreground signal while allowing RFI to pass through and remain detectable.

Let $Q_{fi} \equiv Q_{fi}^{\text{rad}} \vee Q_{fi}^{\text{t-filt}}$ denote the combined binary mask from Appendices A.1 and A.2, and define the good data flag

$G_{fi} \equiv 1 - Q_{fi}$. The diagonal masking operator is then given by

$$[G_t]_{ff'} \equiv G_{fi} \delta_{ff'}. \quad (\text{A22})$$

The per-time DAYENU high-pass operator (A. Ewall-Wice et al. 2021) is then

$$H_t^{\text{rfi}} = (G_t C G_t)^+, \quad (\text{A23})$$

where $^+$ denotes the Moore–Penrose pseudoinverse, and C is the model frequency–frequency covariance of smooth foregrounds from Equation (16), with $\tau_{\text{cut}} = 400$ ns and ridge $\epsilon = 10^{-12}$. The value of τ_{cut} is set by the geometric horizon delay for the longest baseline in the array (≈ 100 m, corresponding to $\tau_{\text{hor}} \approx 330$ ns), plus an additional buffer of

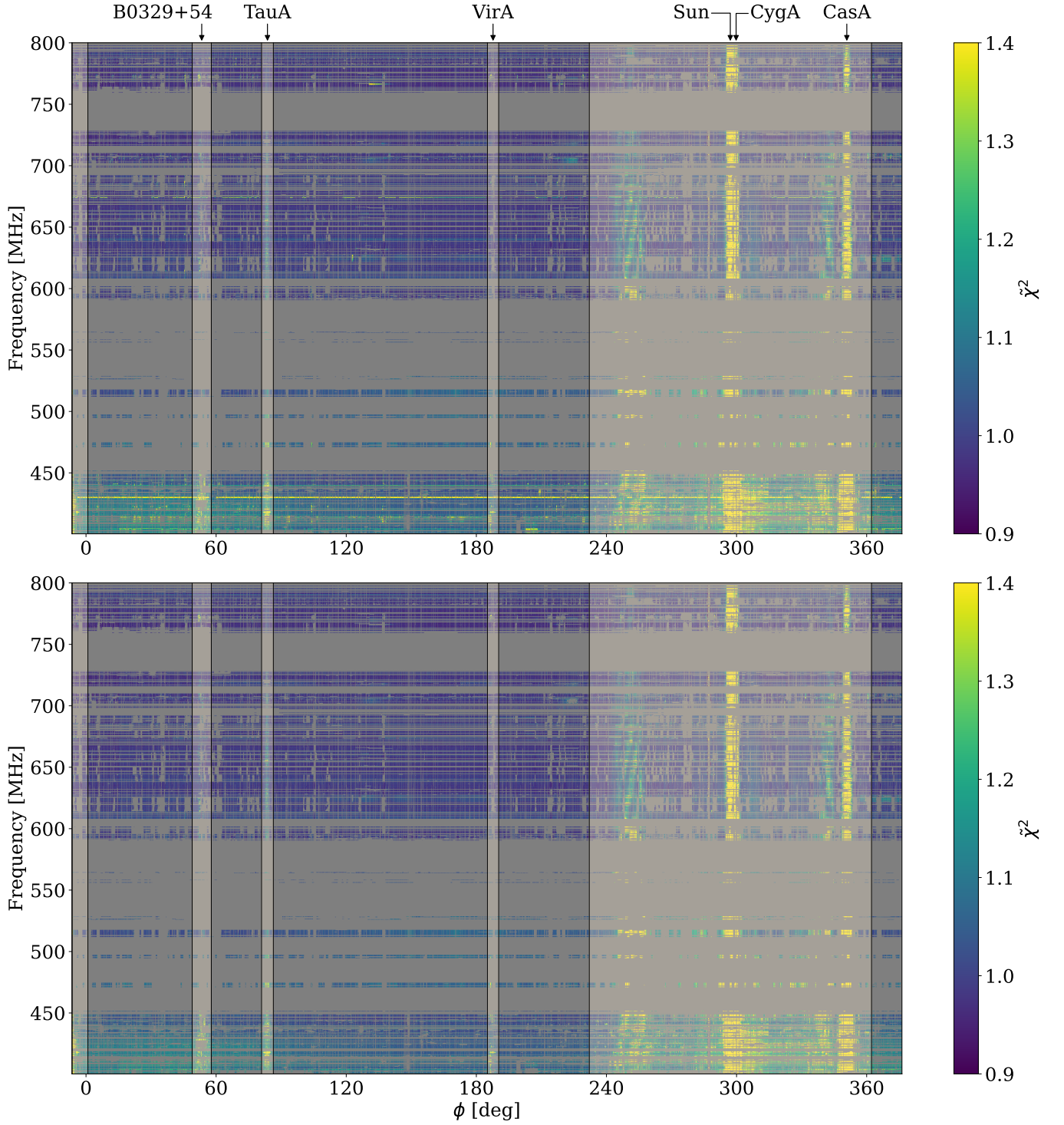


Figure A4. χ^2/dof test statistic ($\tilde{\chi}^2$) as a function of local ERA and frequency on 2019 January 15. Masked samples are shown in gray. Local ERAs when the Sun is above the horizon and when bright sources transit are shaded. Top: combined mask from the radiometer and fringe-rate filters, $Q^{\text{rad}} \vee Q^{\text{f-tilt}}$, applied prior to filtering and constructing $\tilde{\chi}^2$. Bottom: same as above, but including the additional mask $Q^{\text{f-tilt}}$, derived directly from $\tilde{\chi}^2$ and applied after filtering. No $Q^{\text{f-tilt}}$ mask is constructed within the shaded intervals.

≈ 70 ns to allow for primary beam leakage. This choice effectively removes smooth sky contributions, even from sources near the horizon, while leaving nonsmooth signals, such as RFI, present at high delays.

The filter is applied to both visibilities,

$$V_{pf'ent}^{\text{hpf,rft}} = \sum_{f'} H_{ff't}^{\text{rft}} V_{pf'ent}, \quad (\text{A24})$$

and to the noise variance estimate to obtain the inverse variance of the filtered visibilities,

$$w_{pfent}^{\text{hpf,rfi}} = \left(\sum_{f'} \left| H_{ff't}^{\text{rfi}} \right|^2 \sigma_{pf'ent}^2 \right)^{-1}. \quad (\text{A25})$$

Note that the variance estimates used here already have the radiometer mask Q^{rad} applied, so they closely match the expectation for uncorrelated Gaussian noise derived from the autocorrelations, and they have been median-filtered along the time axis using a 5 minute window to suppress sample variance, following the standard data processing pipeline. From these two quantities, we can construct a χ^2/dof test statistic by averaging over baselines:

$$\tilde{\chi}_{ft}^2 = \frac{1}{N_b} \sum_{p \text{ pen}} w_{pfent}^{\text{hpf,rfi}} \left| V_{pfent}^{\text{hpf,rfi}} \right|^2, \quad (\text{A26})$$

where the sum over p includes all polarization products $p \in \{\text{XX}, \text{XY}, \text{YY}\}$, the sum over n includes NS baselines $n \in [-239, 239]$, and the sum over e includes only intercylinder baselines $e \in [1, 2, 3]$. The quantity N_b denotes the total number of baselines included in the sum. If the data are consistent with thermal noise at high delay, then $\mathbb{E}[\tilde{\chi}^2] = 1$ and $\text{Var}[\tilde{\chi}^2] = 2/N_b$.

Figure A4 (top panel) shows the $\tilde{\chi}^2$ test statistic over a full sidereal day. Outliers in this statistic are identified through a simple two-step procedure. First, for each frequency we compute the median of $\tilde{\chi}^2$ over time, using only unmasked samples, and compare it to a smooth baseline across frequency. The baseline is obtained using the *asymmetric reweighted penalized least squares* method (S.-J. Baek et al. 2015), which iteratively estimates a smooth trend by minimizing a weighted combination of squared curvature and residuals, with the weights updated to downweight positive outliers. We use a regularization parameter of 10^5 . Channels whose medians deviate from this baseline by more than 5 MAD are masked uniformly in time, removing channels that exhibit persistent excess power at high delay throughout the sidereal day.

Next, we subtract a moving weighted median in time to remove slow background variations in the test statistic,

$$\delta\tilde{\chi}_{ft}^2 \equiv \tilde{\chi}_{ft}^2 - \left[\text{MedFilt}_{w_t,1}(\tilde{\chi}^2) \right]_{ft}, \quad (\text{A27})$$

where the window size is $w_t = 601$ samples (100 minutes). We then flag times and frequencies where $|\delta\tilde{\chi}_{ft}^2| > 5\sigma_{\tilde{\chi}^2}$, with $\sigma_{\tilde{\chi}^2} = \sqrt{2/N_b}$ representing the expected scatter.

Elevated values of $\tilde{\chi}^2$ are observed during several specific periods. Around the transits of bright sources (Cyg A, Cas A, Tau A, and Vir A), excess power at high delay is primarily due to residual bandpass errors that couple foregrounds to high-delay modes. In the main pipeline, these effects are mitigated using the HyFoRes algorithm (Section 3.3) and by masking the brightest sources (Section 5.2). Elevated values are also observed when the Sun is in the primary beam during transit, as well as when it is in the far sidelobes near sunrise and sunset. In the latter case, although the beam response is small, the Sun’s extreme brightness combined with the near-horizon geometry of the sidelobes produces significant power at high delays. Finally, the bright pulsar B0329+54 exhibits significant power at high delays, which is thought to arise from its

Table A1
Percentage of Nighttime Data Masked

Method	400–800 MHz	608.2–707.8 MHz
Static	31.6	3.5
Radiometer	13.1	11.9
Fringe rate	3.7	3.8
Radiometer $\wedge$ fringe rate ^a	14.4	17.3
Spectral	2.8	2.2
Total	65.6	38.7

Note.

^a The radiometer test and fringe-rate filtering masks are constructed independently; this row gives the percentage of data masked by both methods. All other rows give the percentage masked exclusively by that method.

strong diffractive scintillation (e.g., N. Wang et al. 2008). For these reasons, we do not construct a mask during daytime, as these data are not used in the cosmological analysis. We also do not generate a mask during the transits of bright sources or B0329+54 within a 1σ window around transit, where σ is the approximate width of a Gaussian fit to the primary beam.

The resulting mask from this procedure is denoted $Q^{\text{f-tilt}}$, where samples identified as outliers in the high-delay $\tilde{\chi}^2$ test are flagged. Unlike the radiometer-based method, no iterative background re-estimation is performed. Figure A4 (bottom panel) shows the $\tilde{\chi}^2$ test statistic after this mask has been applied. An additional 1.5% of the data outside of the regions mentioned above was masked on this particular sidereal day.

Finally, we define the overall RFI mask as the logical OR of the three masks introduced in this appendix, $Q_{ft} \equiv Q_{ft}^{\text{rad}} \vee Q_{ft}^{\text{f-tilt}} \vee Q_{ft}^{\text{f-filt}}$. This global mask is used to excise contaminated data from all subsequent analyses. Table A1 summarizes the percentage of nighttime data excised by each method and by their combination, for both the full CHIME band (400 MHz to 800 MHz) and the analysis sub-band (608.2–707.8 MHz) over the 94 sidereal days described in Section 2.4. In this table, “Static” denotes the fixed mask that removes persistently contaminated frequency channels. The radiometer and fringe-rate filtering masks are constructed independently, and there is substantial overlap in the samples that both methods identify as contaminated. Among the data excised exclusively by one method or the other, the fringe-rate filtering method is particularly effective at transient, localized RFI (e.g., scattered TV channels), whereas the radiometer test method captures many short-duration, broadband events and other miscellaneous RFI. The spectral filtering method is applied after these stages and typically excises only a small additional fraction of data.

Appendix B

Details of the Spatial and Spectral Mask Construction

In this appendix, we provide the technical details for constructing the masks used to mitigate contamination from bright continuum foreground sources (Section 5.2) and narrowband absorption features (Section 3.8).

For each source s , we construct a smoothly tapered mask defined by the source coordinates $(\nu_s, \theta_s, \alpha_s)$, where ν_s is the peak-frequency channel of the narrowband feature (for absorber candidates), and α_s and θ_s are the R.A. and decl. of the source in CIRS coordinates at the mean epoch of the observations used to construct the map. For continuum sources, we construct purely spatial masks in the (y, ϕ) plane

that are applied uniformly across frequency. For absorber candidates, we use a 3D mask that depends on both the spatial coordinates and the observed frequency.

We first define a generalized (normalized) distance d from a map voxel at coordinates (ν, y, ϕ) to the effective location of the source at coordinates $(\nu_s, y_s(\phi_s), \phi_s)$:

$$d^{(s)}(\nu, y, \phi) = \left[\left(\frac{\nu - \nu_s}{w_\nu} \right)^2 + \left(\frac{y - y_s(\phi_s)}{w_y} \right)^2 + \left(\frac{(\phi - \phi_s) \cos \theta_s}{w_\phi} \right)^2 \right]^{1/2}. \quad (\text{B1})$$

The parameters w_ν , w_y , and w_ϕ set the extent of the mask in the ν , y , and ϕ directions, respectively. For continuum sources, we enforce frequency independence by taking $w_\nu \rightarrow \infty$, so that the first term in Equation (B1) vanishes and $d^{(s)}$ depends only on (y, ϕ) . For absorber candidates, we use a finite w_ν (defined below), so that the taper is genuinely 3D.

The spatial extents w_y and w_ϕ are defined differently in the two directions. The extent in y is set by the width of the synthesized beam, whereas the extent in ϕ is determined by the width of the primary beam, since we have significant grating lobes along ϕ . To obtain a single spatial mask that is valid across all frequencies, we conservatively set w_y and w_ϕ using the beam widths at the lowest frequency in our analysis band (608.2 MHz), ensuring that the masked regions are sufficiently large at all frequencies. We adopt $w_y = 0.018$ (corresponding to approximately 1° at zenith), which is roughly 3 times the FWHM of the NS synthesized beam, and $w_\phi = 4.8$, which is roughly twice the FWHM of the primary beam.

We then apply a smooth cosine taper as a function of the generalized distance d to ensure a gradual transition from fully masked to unmasked regions:

$$T(d) = \begin{cases} 0, & d \leq d_{\text{in}}, \\ 0.5 \left[1 + \cos \left(\frac{\pi(d - d_{\text{in}})}{w_{\text{taper}}} \right) \right], & d_{\text{in}} < d \leq d_{\text{out}}, \\ 1, & d > d_{\text{out}}, \end{cases} \quad (\text{B2})$$

where d is the normalized distance defined in Equation (B1), $d_{\text{in}} = 1.0$ defines the inner radius where the taper begins, $d_{\text{out}} = d_{\text{in}} + w_{\text{taper}}$ defines the outer radius where the taper reaches unity, and $w_{\text{taper}} = 1.0$ is the taper width parameter. This yields a smooth transition from fully masked regions ($T = 0$) to unmasked regions ($T = 1$) over a distance equal to the mask window size.

As mentioned in Section 5.2, we construct several families of tapered masks, each targeting a specific source of contamination:

1. *Transit mask.* For all continuum sources in the *specfind* v3 catalog (Y. Stein et al. 2021) with flux densities greater than 10 Jy at 600 MHz, we construct a spatial mask $T_{\text{transit}}^{(s)}(y, \phi)$ for each source s by treating it as a point at

$$y_s \equiv \sin(\theta_s - \Lambda), \quad \phi_s \equiv \alpha_s. \quad (\text{B3})$$

We evaluate the distance metric $d^{(s)}$ using Equation (B1) under the continuum source convention $w_\nu \rightarrow \infty$, and apply the taper function in Equation (B2) to obtain $T_{\text{transit}}^{(s)}(y, \phi)$, which is applied uniformly across all frequency channels.

2. *Track mask.* For the brightest continuum sources, with flux densities greater than 60 Jy at 600 MHz, we construct extended track masks $T_{\text{track}}^{(s)}(y, \phi)$ that follow the full path of each source s across the sky. The track is parameterized in terms of a local ERA coordinate ϕ_s , which denotes the source location in local ERA, as

$$y_s(\phi_s) \equiv \cos \Lambda \sin \theta_s - \sin \Lambda \cos \theta_s \cos(\phi_s - \alpha_s). \quad (\text{B4})$$

We sample ϕ_s on a grid that is upsampled by a factor of 10 relative to the native map resolution in ϕ , and restrict this grid to the range of local ERA over which the source is above the horizon. For each sampled point $(y_s(\phi_s^{(k)}), \phi_s^{(k)})$, we evaluate Equation (B1) with $w_\nu \rightarrow \infty$ to obtain a distance $d_k^{(s)}(y, \phi)$ at each map pixel. The mask value at a given pixel (y, ϕ) is then set by the minimum distance to any point on the track,

$$d_{\text{min}}^{(s)}(y, \phi) = \min_k d_k^{(s)}(y, \phi),$$

and the taper function in Equation (B2) is applied to $d_{\text{min}}^{(s)}$ to define $T_{\text{track}}^{(s)}(y, \phi)$.

3. *Galactic plane mask.* To mask sidelobe pickup from the inner Galactic plane within the NGC field, we construct a Galactic plane mask $T_{\text{gal}}(y, \phi)$ using the same track mask procedure. We select a set of dimmer known continuum sources in the inner Galactic plane whose tracks approximately coincide with the residual contamination observed in the maps at delays within the transition region of the delay filter. For each such source g , we build a track mask $T_{\text{track}}^{(g)}(y, \phi)$ as above (again with $w_\nu \rightarrow \infty$), and define

$$T_{\text{gal}}(y, \phi) = \prod_g T_{\text{track}}^{(g)}(y, \phi). \quad (\text{B5})$$

4. *Absorber mask.* As described in Section 3.8, we identify 40 narrowband absorption candidates with an S/N > 7 within the NGC region. For each candidate system s with sky position (θ_s, α_s) and observed frequency ν_s (corresponding to the peak of the feature), we construct a 3D mask $T_{\text{abs}}^{(s)}(\nu, y, \phi)$ using the distance definition in Equation (B1). The spectral extent is defined as

$$w_\nu = n_\nu \Delta_\nu, \quad (\text{B6})$$

where $\Delta_\nu = 390.625$ kHz is the channel width and we adopt $n_\nu = 13$ to cover 13 channels on both sides of the peak channel, encompassing the spectral sidelobes (see Section 3.8). We use the same spatial parameters w_y and w_ϕ as for the transit masks, and apply the taper function in Equation (B2) to the full 3D distance to obtain $T_{\text{abs}}^{(s)}(\nu, y, \phi)$.

We compute the rms of the map over the frequency axis, and it is shown in Figure B1. The top panel shows the unmasked data, clearly revealing systematic contamination from the brightest point sources. These sources produce characteristic “U”-shaped tracks in the map as they drift through the far sidelobes of the primary beam. Additionally, numerous point-source-like features are visible, which are predominantly

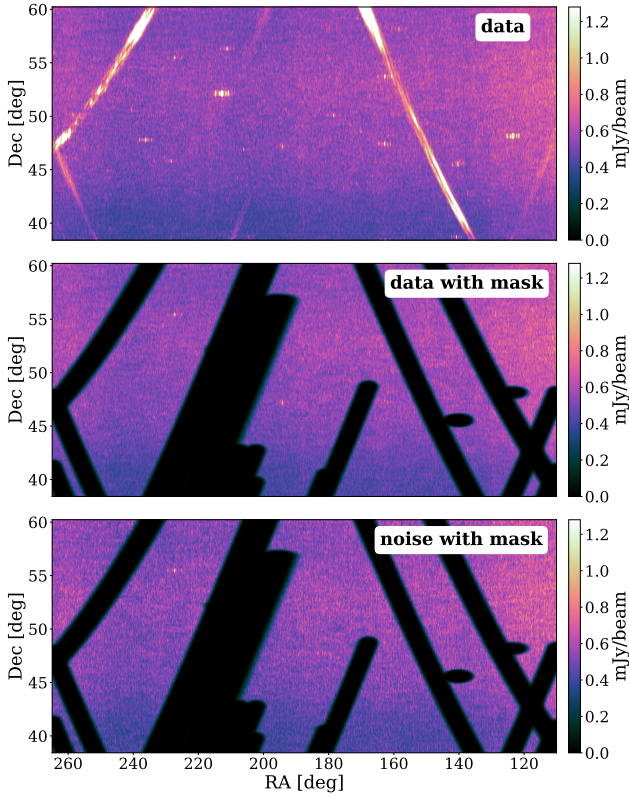


Figure B1. Top panel: rms of the map over the frequency axis, without applying any mask. The “U”-shaped tracks correspond to bright continuum foreground sources in the far sidelobes of the primary beam, and the point-source-like features are predominantly narrowband absorption features. Middle panel: same data after applying the combined spatial mask to mitigate continuum foreground contamination and the 3D absorber mask to remove narrowband absorption features, flagging $\sim 33\%$ of the spatial area and 1.35% of voxels, respectively. Bottom panel: a single noise realization, scaled to account for sky temperature variations (see Section 5.4), with the identical mask applied. After masking, the data and noise are in excellent agreement. A point-source-like feature is visible in both the masked data and noise maps (near R.A. $\sim 227^\circ$, decl. $\sim 55^\circ$). This arises from an unknown variable source that also appears in the even-odd jackknife map used to derive the spatial scaling of the noise simulations, and is therefore imprinted in the noise realization as well. To assess its impact, we generate an additional transit mask to exclude this feature and recompute the power spectrum, finding no significant change in the measured bandpowers. We conclude that this source does not bias our results.

candidate absorbers. The middle panel shows the same data after applying both the spatial mask (to mitigate continuum foreground source contamination) and the spectral mask (to remove narrowband absorption features). The bottom panel presents a single realization of the noise simulation, scaled to account for spatial variations in sky temperature (as detailed in Section 5.4), with the identical mask applied. The masked data and noise realization exhibit excellent agreement.

However, both the masked data and noise maps contain a single point-source-like feature at the same location. This feature is not associated with any known bright point source in the sky, but exhibits variability across sidereal days. Because it is variable, this source appears in the even-odd jackknife map, which is used to derive the spatial scaling factor for the noise simulations (Section 5.4). Consequently, this variability is imprinted in the noise realizations as well. The physical origin of this source requires further investigation. To assess its

impact on this analysis, we generate an additional spatial transit mask to exclude this feature and recompute the power spectrum. We find no significant deviation in the measured bandpowers, confirming that this source does not bias our result.

Appendix C

Varying the Signal Prior in the Delay Spectrum Estimation

In Section 5.1, we add a small regularization matrix $\mathbf{C}_{\text{reg}}$ to the propagated noise covariance $\mathbf{N}$ when forming inverse-covariance weights along the frequency axis, $(\mathbf{C}_{\text{reg}} + \mathbf{N})^{-1}$. The role of $\mathbf{C}_{\text{reg}}$ is to provide a smooth, simulation-informed approximation to the expected 21 cm frequency–frequency covariance while remaining subdominant to the noise, so that it stabilizes the matrix inverse without significantly changing the weighting of the data. Here we describe how $\mathbf{C}_{\text{reg}}$ is constructed.

We start from a simple model for the 21 cm signal covariance in delay space, which we take to be diagonal with constant variance:

$$[\mathbf{S}]_{aa'} = A_{\text{prior}} \delta_{aa'}. \quad (\text{C1})$$

The parameter A_{prior} sets the overall variance of the 21 cm prior in delay space. We calibrate A_{prior} using an ensemble of 21 cm simulations passed through our telescope model and analysis pipeline. From the resulting maps, we form a sample frequency–frequency covariance by averaging over local ERA and across 21 cm realizations, then Fourier transform it along the frequency axis to obtain an effective delay-space covariance. We find that the variance decreases approximately exponentially with delay. Rather than imposing this scale dependence explicitly, we adopt a flat prior in delay space with constant amplitude $A_{\text{prior}} = 5 \times 10^{-6}$, chosen to match the mean variance observed in simulations over the range of delays used for power spectrum estimation.

The same simulations also show significant structure in the frequency–frequency covariance induced by the primary beam. To ensure that $\mathbf{C}_{\text{reg}}$ reflects this behavior, we introduce a diagonal matrix $\mathbf{B}$ whose entries are the rms of the EW synthesized beam for the pixel under consideration,

$$[\mathbf{B}]_{ff'} = \delta_{ff'} \left[\frac{1}{2N_\phi} \sum_p \sum_r |B_{\text{synth}, pfd}^{\text{EW}}(\phi_r)|^2 \right]^{1/2}, \quad (\text{C2})$$

where $B_{\text{synth}, pfd}^{\text{EW}}(\phi)$ is given by Equation (45). We then construct the regularization matrix as

$$\mathbf{C}_{\text{reg}} = \mathbf{W} \mathbf{G} \mathbf{H} \mathbf{B} \mathbf{F} \mathbf{S} \mathbf{F}^\dagger \mathbf{B} \mathbf{H}^\dagger \mathbf{G} \mathbf{W}, \quad (\text{C3})$$

where the $\mathbf{W}$, $\mathbf{G}$, $\mathbf{H}$, and $\mathbf{F}$ operators implement the sub-band selection, frequency channel mask, foreground filter, and delay-to-frequency transform, respectively, and are defined in Section 5.1. Compared to the raw frequency–frequency signal covariance measured directly from the end-to-end 21 cm simulations, this construction captures the dominant frequency structure introduced by the primary beam while smoothing over sample variance fluctuations.

It is important to verify that our results do not depend sensitively on the choice of A_{prior} . To test this, we estimate power spectra for a wide range of A_{prior} values between 10^{-7} and 5×10^{-5} . The resulting 1D power spectra are shown in

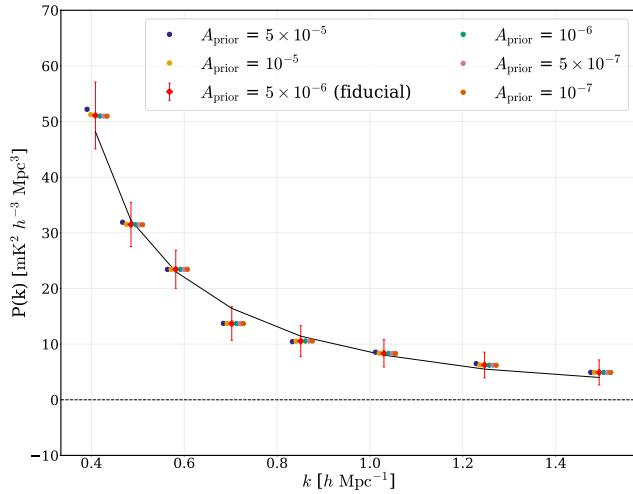


Figure C1. Effect of varying the signal prior variance A_{prior} on the 1D power spectrum. The different-colored points show the measured $P(k)$ for several choices of A_{prior} between 10^{-7} and 5×10^{-5} , while the red open diamond corresponds to our fiducial choice $A_{\text{prior}} = 5 \times 10^{-6}$. The solid black line shows the best-fit model from Section 6.2. The curves are nearly indistinguishable, with differences far smaller than the statistical uncertainties, indicating that the measurement is effectively insensitive to the choice of A_{prior} over this range.

Figure C1. As A_{prior} is decreased, the recovered $P(k)$ rapidly converges to a stable curve; our fiducial choice $A_{\text{prior}} = 5 \times 10^{-6}$ lies well within this convergence regime. Differences between the power spectra for different A_{prior} are small compared to the statistical uncertainties, demonstrating that A_{prior} acts only as a mild regularization of the inverse in Equation (57) and does not bias the final power spectrum estimates.

Appendix D

Details of Detection Significance and Goodness-of-fit Calculations

In this appendix, we provide further details of the detection significance calculations presented in Section 6.3 and goodness-of-fit results quoted in Section 6.2.3.

D.1. Detection Significance

When computing χ^2_{null} and $\chi^2_{\text{best-fit}}$ in Equations (79) and (80) for each noise realization, we recompute a “leave-one-out” covariance C_N that excludes that realization d . We determine $\chi^2_{\text{best-fit}}$ by using the BFGS algorithm. To ensure that the algorithm does not get stuck in a local minimum, we run it 1000 times with different initial guesses widely distributed in the four-dimensional (4D) model parameter space,³² and retain the parameter values that result in the minimum χ^2 value out of these 1000 runs. We compute $\Delta\chi^2$ for the measured power spectrum in the same manner.

In simple cases, Wilks’ theorem states that, under the null hypothesis, $\Delta\chi^2$ for the noise power spectra is drawn from a

χ^2 distribution with $n_{\text{dof}} = n_{\text{signal}} - n_{\text{null}}$ dof, where n_{signal} and n_{null} are the numbers of parameters in the signal and null models, respectively (4 and 0 in our case). However, parameter degeneracies generally break the conditions required for Wilks’ theorem to hold, so it is necessary to characterize the distribution of $\Delta\chi^2$ values via Monte Carlo.

A common prescription is to assume that $\Delta\chi^2$ is still χ^2 distributed, but with an *effective* number of degrees of freedom $n_{\text{dof}} < n_{\text{signal}} - n_{\text{null}}$ that captures the reduction due to parameter degeneracies. However, since we have estimated the noise covariance from a set of noise realizations, we expect the relevant distribution to be F instead of χ^2 (T. W. Anderson 2003; E. Sellentin & A. F. Heavens 2016). In this case, the analog of Wilks’ theorem with an effective number of degrees of freedom is

$$\frac{n - p + 1}{n_{\text{dof}} n} \Delta\chi^2 \sim F_{n_{\text{dof}}, n-p+1}, \quad (\text{D1})$$

where $n = n_{\text{mocks}} - 1$ and $p = n_{\text{data}}$ denotes the length of the data vector.³³ Intuitively, fluctuations in the covariance matrix used to compute each χ^2 value will lead to a slightly heavier tail in the distribution of $\Delta\chi^2$ values, slightly decreasing the detection significance inferred from $\Delta\chi^2$ computed from the data compared to what would be obtained from a χ^2 distribution.

We verify this expectation by fitting Equation (D1) to the sets of 1000 noise realizations for the full band and each sub-band, determining the best-fit value of n_{dof} in each case, and using the Kolmogorov–Smirnov (KS) and Anderson–Darling (AD) tests to assess how well the associated F distribution describes each set of $\Delta\chi^2$ values.³⁴ For both sub-bands, the KS and AD tests both pass (with p -values greater than 0.05), while for the full band, the AD test passes while the KS test fails ($p = 0.005$).

To investigate whether this failure is a statistical fluctuation in the original set of 1000 noise realizations, we generate two other independent sets of 1000 realizations, and repeat the above tests. For the other two sets, the KS and AD tests both pass (with $p > 0.2$). Using two-sample KS and AD tests, we verify that each pair of noise realization sets is consistent with being drawn from the same distribution. We also verify that all three sets are consistent with being drawn from a Gaussian distribution with the associated covariance matrix, by comparing the set of χ^2_{null} values to a χ^2 distribution with n_{data} dof. Finally, we perform a “leave-one-block-out” comparison: for each set of 1000 noise realizations, we find the best-fit F distribution for the concatenation of the two other sets, and compute the KS and AD p -values comparing this

³² For this calculation, we use a parameter set that includes Ω_{HI}^2 instead of Ω_{HI} , allowing for $\Omega_{\text{HI}}^2 < 0$ to allow the best-fit model for a given noise power spectrum to have a negative amplitude. Due to parameter degeneracies, the $\chi^2_{\text{best-fit}}$ parameter values sometimes lie at extreme parameter values. To ensure that such minima are found, we generate the 1000 initial guesses using Latin hypercube sampling of the 4D space of the (base-10) logarithm of each parameter, with $\log(\Omega_{\text{HI}}^2)$ bounded by $[-2, 6]$ and $\log(b_{\text{HI}})$, $\log(\alpha_{\text{NL}})$, and $\log(\alpha_{\text{FoG}})$ bounded by $[-2, 2]$.

³³ Equation (D1) can be understood from the fact that the $(n - p + 1)/n$ prefactor and the “denominator degrees of freedom” $n - p + 1$ of the F distribution arise from the Wishart distribution from which the estimated covariance matrix is drawn, while the n_{dof}^{-1} prefactor and “numerator degrees of freedom” n_{dof} arise from the vector being tested. Equation (D1) is a version of Hotelling’s T^2 distribution where the vectors being tested are not multivariate Gaussian (in which case n_{dof} would be replaced by p), but are instead determined by n_{dof} , as in Wilks’ theorem with a known covariance.

³⁴ Note that standard KS and AD p -values are unreliable when the reference distribution is determined from the values being tested, so we calibrate our reported p -values via Monte Carlo. For each test, we draw 10,000 sets of 1000 (appropriately scaled) $\Delta\chi^2$ values from the best-fit F distribution corresponding to our 1000 noise realizations. We refit n_{dof} to each set and compute the KS or AD test statistic. We then compute a p -value by comparing this collection of test statistic values with that computed from the original 1000 noise realizations and fitted distribution.

distribution to the first set. In all cases, these tests pass with $p > 0.05$. Therefore, we conclude that the original set of 1000 full-band noise realizations is a mild outlier, with only the self-fit to an F distribution displaying unexpected behavior. For other parts of the analysis in this paper, we use this original set of noise realizations, and have checked that the results change by a negligible amount if the full set of 3000 realizations is used instead.

For each set of noise realizations, we use the best-fit F distribution to compute the p -value corresponding to the $\Delta\chi^2$ value computed from the data, and use the inverse cumulative distribution function of a standard normal distribution to translate this into an “effective number of sigmas” as an interpretable measure of the signal to noise. We quote the mean over the three sets of noise realizations in the first S/N column of Table 2, and use these values as our primary estimate of the S/N of our measurements. The standard error on these mean values is 0.2σ , 0.1σ , and 0.2σ for the three bands.

In several cases, the $\Delta\chi^2$ values for each set of noise realizations are inconsistent with χ^2 distributions (the KS test yields $p < 0.05$). Nevertheless, for comparison with the F -derived results, we quote the S/N obtained from the best-fit χ^2 distribution in the second S/N column of Table 2. As expected, these values are slightly higher than those from the F distributions, due to the lighter tail of the χ^2 distribution. For each band, the best-fit n_{dof} values for the F and χ^2 distributions are equal to two significant figures, falling between $n_{\text{dof}} = 3.5$ and 3.7.

D.2. Goodness of Fit

To assess the goodness of fit of the best-fit signal models described in Section 6.2.2, we add the best-fit model for a given band to each of our base set of 1000 noise realizations. We then refit the four-parameter model to each “signal+noise” realization, using the L-BFGS-B algorithm to minimize the χ^2 and taking the lowest result across 1000 random initial guesses in the parameter space defined by our priors in Section 6.2.2. We compute an empirical p -value from the fraction of these χ^2 values that exceed the value from the data. For additional robustness, we repeat this exercise with the other two sets of 1000 noise realizations from Appendix D.1, and report the mean p -value from the three sets. For the full band, upper sub-band, and lower sub-band, we find $p = 0.92$, 0.25 , and 0.31 , with standard deviations of 0.01 , 0.03 , and 0.03 , respectively.

ORCID iDs

Mandana Amiri  <https://orcid.org/0000-0001-6523-9029>
 Kevin Bandura  <https://orcid.org/0000-0003-3772-2798>
 Arnab Chakraborty  <https://orcid.org/0000-0002-7758-9859>
 Jean-François Cliche  <https://orcid.org/0000-0001-6509-8430>
 Matt Dobbs  <https://orcid.org/0000-0001-7166-6422>
 Simon Foreman  <https://orcid.org/0000-0002-0190-2271>
 Liam Gray  <https://orcid.org/0000-0003-3986-954X>
 Mark Halpern  <https://orcid.org/0000-0002-1760-0868>
 Alex S Hill  <https://orcid.org/0000-0001-7301-5666>
 Gary Hinshaw  <https://orcid.org/0000-0002-4241-8320>
 Carolin Höfer  <https://orcid.org/0000-0003-4887-8114>
 Albin Joseph  <https://orcid.org/0000-0003-4179-4073>
 Nolan Kruger  <https://orcid.org/0009-0004-2241-0550>

T. L. Landecker  <https://orcid.org/0000-0003-1455-2546>
 Rik van Lieshout  <https://orcid.org/0009-0005-1033-3292>
 Joshua MacEachern  <https://orcid.org/0000-0001-8064-6116>
 Kiyoshi W. Masui  <https://orcid.org/0000-0002-4279-6946>
 Juan Mena-Parra  <https://orcid.org/0000-0002-0772-9326>
 Kyle Miller  <https://orcid.org/0009-0005-2291-510X>
 Nikola Milutinovic  <https://orcid.org/0000-0001-8292-0051>
 Arash Mirhosseini  <https://orcid.org/0000-0002-2626-5985>
 Laura Newburgh  <https://orcid.org/0000-0002-7333-5552>
 Anna Ordog  <https://orcid.org/0000-0002-2465-8937>
 Ue-Li Pen  <https://orcid.org/0000-0003-2155-9578>
 Tristan Pinsonneault-Marotte  <https://orcid.org/0000-0002-9516-3245>
 Alex Reda  <https://orcid.org/0000-0001-6967-7253>
 Andre Renard  <https://orcid.org/0000-0003-3463-7918>
 Kana Sakaguri  <https://orcid.org/0000-0001-5667-8118>
 J. Richard Shaw  <https://orcid.org/0000-0002-4543-4588>
 Shabbir Shaikh  <https://orcid.org/0000-0001-6731-0351>
 Seth R. Siegel  <https://orcid.org/0000-0003-2631-6217>
 Saurabh Singh  <https://orcid.org/0000-0001-7755-902X>
 David Spear  <https://orcid.org/0009-0003-6054-8035>
 Yukari Uchibori  <https://orcid.org/0009-0003-4114-1301>
 Keith Vanderlinde  <https://orcid.org/0000-0003-4535-9378>
 Haochen Wang  <https://orcid.org/0000-0002-1491-3738>
 Donald V. Wiebe  <https://orcid.org/0000-0002-6669-3159>
 Dallas Wulf  <https://orcid.org/0000-0001-7314-9496>

References

- Abdurashidova, Z., Aguirre, J. E., Alexander, P., et al. 2022, *ApJ*, 925, 221
 Ahmed, Z., Alonso, D., Amin, M. A., et al. 2019, arXiv:1907.13090
 Ali, Z. S., Parsons, A. R., Zheng, H., et al. 2018, *ApJ*, 863, 201
 Anderson, C. J., Luciw, N. J., Li, Y. C., et al. 2018, *MNRAS*, 476, 3382
 Anderson, T. W. 2003, *An Introduction to Multivariate Statistical Analysis*, Wiley Series in Probability and Statistics (3rd ed.; Wiley)
 Asad, K. M. B., Koopmans, L. V. E., Jelić, V., et al. 2015, *MNRAS*, 451, 3709
 Baek, S.-J., Park, A., Ahn, Y.-J., & Choo, J. 2015, *Ana*, 140, 250
 Bale, S. D., Bassett, N., Burns, J. O., et al. 2023, arXiv:2301.10345
 Barata, J. C. A., & Hussein, M. S. 2012, *BrJPh*, 42, 146
 Behnel, S., Bradshaw, R., Citro, C., et al. 2011, *CSE*, 13, 31
 Brinkerink, C. D., Arts, M. J., Bentum, M. J., et al. 2025, arXiv:2504.03418
 Brooks, S. P., & Gelman, A. 1998, *Journal of Computational and Graphical Statistics*, 7, 434
 Carucci, I. P., Bernal, J. L., Cunningham, S., et al. 2025, *A&A*, 703, A222
 Ceccotti, E., Offringa, A. R., Mertens, F. G., et al. 2025, *MNRAS*, 544, 1255
 Chakraborty, A., Datta, A., Roy, N., et al. 2021, *ApJL*, 907, L7
 Chang, T.-C., Pen, U.-L., Bandura, K., & Peterson, J. B. 2010, *Natur*, 466, 463
 Chen, X., Gao, F., Wu, F., et al. 2024, *RSPTA*, 382, 20230094
 Chen, Z., Cunningham, S., Pourtsidou, A., et al. 2025, *ApJS*, 279, 19
 Cheng, C., Parsons, A. R., Kolopanis, M., et al. 2018, *ApJ*, 868, 26
 CHIME Collaboration, Amiri, M., Bandura, K., et al. 2022a, *ApJS*, 261, 29
 CHIME Collaboration, Amiri, M., Bandura, K., et al. 2022b, *ApJ*, 932, 100
 CHIME Collaboration, Amiri, M., Bandura, K., et al. 2023, *ApJ*, 947, 16
 CHIME Collaboration, Amiri, M., Bandura, K., et al. 2026a, arXiv:2603.25680
 CHIME Collaboration, Amiri, M., Bandura, K., et al. 2024a, *ApJ*, 963, 23
 CHIME Collaboration, Amiri, M., Chakraborty, A., et al. 2024b, *ApJ*, 976, 163
 CHIME Collaboration, Amiri, M., Chakraborty, A., et al. 2026b, *ApJ*, 998, 310
 CHIME/FRB Collaboration, Amiri, M., Bandura, K., et al. 2018, *ApJ*, 863, 48
 CHIME/FRB Collaboration, Amiri, M., Bandura, K. M., et al. 2021, *ApJS*, 255, 5
 Collette, A., Kluyver, T., Caswell, T. A., et al. 2023, h5py/h5py: v3.8.0, Zenodo, doi:10.5281/zenodo.7560547
 Cosmic Visions 21 cm Collaboration, Ansari, R., Arena, E. J., et al. 2018, arXiv:1810.09572
 Crichton, D., Aich, M., Amara, A., et al. 2022, *JATIS*, 8, 011019

- Crichton, N. H. M., Murphy, M. T., Prochaska, J. X., et al. 2015, *MNRAS*, **452**, 217
- Cunnington, S., Li, Y., Santos, M. G., et al. 2023, *MNRAS*, **518**, 6262
- Curran, S. J., Whiting, M. T., Tanna, A., et al. 2013, *MNRAS*, **429**, 3402
- Dalcin, L., & Fang, Y.-L. L. 2021, *CSE*, **23**, 47
- Datta, K. K., Choudhury, T. R., & Bharadwaj, S. 2007, *MNRAS*, **378**, 119
- Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, *AJ*, **151**, 44
- de Oliveira-Costa, A., Tegmark, M., Gaensler, B. M., et al. 2008, *MNRAS*, **388**, 247
- de Zotti, G., Massardi, M., Negrello, M., & Wall, J. 2010, *A&ARv*, **18**, 1
- DeBoer, D. R., Parsons, A. R., Aguirre, J. E., et al. 2017, *PASP*, **129**, 045001
- Di Matteo, T., Perna, R., Abel, T., & Rees, M. J. 2002, *ApJ*, **564**, 576
- Eastwood, M. W., Anderson, M. M., Monroe, R. M., et al. 2019, *AJ*, **158**, 84
- Ewall-Wice, A., Kern, N., Dillon, J. S., et al. 2021, *MNRAS*, **500**, 5195
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, **125**, 306
- Frigo, M., & Johnson, S. G. 2005, *IEEEP*, **93**, 216
- Gabriel, E., Fagg, G. E., Bosilca, G., et al. 2004, in Recent Advances in Parallel Virtual Machine and Message Passing Interface, Proc. 11th European PVM/MPI Users' Group Meeting, ed. D. Kranzlmüller, P. Kacsuk, & J. Dongarra (Springer), 97
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, **585**, 357
- Hartlap, J., Simon, P., & Schneider, P. 2007, *A&A*, **464**, 399
- HERA Collaboration, Abdurashidova, Z., Aguirre, J. E., et al. 2022, *ApJ*, **925**, 221
- Hogg, D. W. 1999, arXiv:astro-ph/9905116
- Hunter, J. D. 2007, *CSE*, **9**, 90
- Jackson, J. C. 1972, *MNRAS*, **156**, 1P
- Josaitis, A. T., Ewall-Wice, A., Fagnoni, N., & de Lera Acedo, E. 2022, *MNRAS*, **514**, 1804
- Kaiser, N. 1987, *MNRAS*, **227**, 1
- Karamanis, M., & Beutler, F. 2021, arXiv:2106.06331
- Kern, N. S., Parsons, A. R., Dillon, J. S., et al. 2019, *ApJ*, **884**, 105
- Kolopanis, M., Jacobs, D. C., Cheng, C., et al. 2019, *ApJ*, **883**, 133
- Kulkarni, S. R. 1989, *AJ*, **98**, 1112
- Lewis, A. 2025, *JCAP*, **2025**, 025
- Lewis, A., Challinor, A., & Lasenby, A. 2000, *ApJ*, **538**, 473
- Lewis, A., Torrado, J., Ringeval, C., et al. 2025b, cmbant/getdist: GetDist GUI, v1.7.4, Zenodo, doi:10.5281/zenodo.17432819
- Lewis, A., Vahreschild, D. A., Mead, A., et al. 2025a, cmbant/CAMB: v1.6.5, Zenodo, doi:10.5281/zenodo.17573409
- Li, J., Zuo, S., Wu, F., et al. 2020, *SCPMA*, **63**, 129862
- Li, L.-C., Staveley-Smith, L., & Rhee, J. 2021, *RAA*, **21**, 030
- Liu, A., Parsons, A. R., & Trott, C. M. 2014, *PhRvD*, **90**, 023018
- Liu, A., & Shaw, J. R. 2020, *PASP*, **132**, 062001
- Lyke, B. W., Higley, A. N., McLane, J. N., et al. 2020, *ApJS*, **250**, 8
- Marinacci, F., Vogelsberger, M., Pakmor, R., et al. 2018, *MNRAS*, **480**, 5113
- Martinot, Z. E., Aguirre, J. E., Kohn, S. A., & Washington, I. Q. 2018, *ApJ*, **869**, 79
- Masui, K. W., Shaw, J. R., Ng, C., et al. 2019, *ApJ*, **879**, 16
- Masui, K. W., Switzer, E. R., Banavar, N., et al. 2013, *ApJL*, **763**, L20
- The Matplotlib Development Team 2025, Matplotlib: Visualization with Python, v3.10.7, Zenodo, doi:10.5281/zenodo.17298696
- Mead, A. J., Brieden, S., Tröster, T., & Heymans, C. 2021, *MNRAS*, **502**, 1401
- Mertens, F. G., Mevius, M., Koopmans, L. V. E., et al. 2020, *MNRAS*, **493**, 1662
- Mertens, F. G., Mevius, M., Koopmans, L. V. E., et al. 2025, *A&A*, **698**, A186
- Miles, A., Kirkham, J., Durant, M., et al. 2020, zarr-developers/zarr-python: v2.4.0, Zenodo, doi:10.5281/zenodo.3773450
- Mohammed, N., Ordog, A., Booth, R. A., et al. 2024, *ApJ*, **971**, 100
- Moore, D. F., Aguirre, J. E., Kohn, S. A., et al. 2017, *ApJ*, **836**, 154
- Moore, D. F., Aguirre, J. E., Parsons, A. R., Jacobs, D. C., & Pober, J. C. 2013, *ApJ*, **769**, 154
- Morales, M. F., Hazelton, B., Sullivan, I., & Beardsley, A. 2012, *ApJ*, **752**, 137
- Morales, M. F., & Hewitt, J. 2004, *ApJ*, **615**, 7
- Naiman, J. P., Pillepich, A., Springel, V., et al. 2018, *MNRAS*, **477**, 1206
- Nelson, D., Pillepich, A., Springel, V., et al. 2018, *MNRAS*, **475**, 624
- Offringa, A. R., de Bruyn, A. G., Biehl, M., et al. 2010, *MNRAS*, **405**, 155
- Offringa, A. R., de Bruyn, A. G., & Zaroubi, S. 2012a, *MNRAS*, **422**, 563
- Offringa, A. R., van de Gronde, J. J., & Roerdink, J. B. T. M. 2012b, *A&A*, **539**, A95
- Oh, S. P., & Mack, K. J. 2003, *MNRAS*, **346**, 871
- Parsons, A. R., & Backer, D. C. 2009, *AJ*, **138**, 219
- Parsons, A. R., Pober, J. C., Aguirre, J. E., et al. 2012, *ApJ*, **756**, 165
- Paul, S., Santos, M. G., Chen, Z., & Wolz, L. 2026, *ApJL*, **1005**, L56
- Pen, U.-L., Staveley-Smith, L., Peterson, J. B., & Chang, T.-C. 2009, *MNRAS*, **394**, L6
- Pillepich, A., Nelson, D., Hernquist, L., et al. 2018, *MNRAS*, **475**, 648
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, **641**, A6
- Polidan, R. S., Burns, J. O., Ignatiev, A., et al. 2024, *AdSpR*, **74**, 528
- Renard, A., Shaw, R., Ng, C., et al. 2021, Kotekan: A framework for high-performance radiometric data pipelines, 2021.11, Zenodo, doi:10.5281/zenodo.5842660
- Rhodes, B. 2019, Skyfield: High precision research-grade positions for planets and Earth satellites generator, Astrophysics Source Code Library, ascl:1907.024
- Rogowski, M., Aseeri, S., Keyes, D., & Dalcin, L. 2023, *ITPDS*, **34**, 611
- Ross, A. J., Bautista, J., Tojeiro, R., et al. 2020, *MNRAS*, **498**, 2354
- Sarkar, D., & Bharadwaj, S. 2019, *MNRAS*, **487**, 5666
- Sellentin, E., & Heavens, A. F. 2016, *MNRAS*, **456**, L132
- Shaw, J. R., Gray, L., Foreman, S., et al. 2025a, cora, v25.10.0, Zenodo, doi:10.5281/zenodo.17438047
- Shaw, J. R., Sigurdson, K., Pen, U.-L., Stebbins, A., & Sitwell, M. 2014, *ApJ*, **781**, 57
- Shaw, J. R., Sigurdson, K., Sitwell, M., Stebbins, A., & Pen, U.-L. 2015, *PhRvD*, **91**, 083514
- Shaw, J. R., Masui, K., Gray, L., et al. 2025b, caput, v25.3.1, Zenodo, doi:10.5281/zenodo.15150776
- Shaw, J. R., Masui, K., Hincks, A. D., et al. 2025c, ch_pipeline, v25.10.0, Zenodo, doi:10.5281/zenodo.17438050
- Shaw, J. R., Gray, L., Siegel, S. R., et al. 2025d, draco, v25.10.0, Zenodo, doi:10.5281/zenodo.17438067
- Shaw, J. R., Foreman, S., Nitsche, R., et al. 2025e, driftscan, v25.3.1, Zenodo, doi:10.5281/zenodo.15150698
- Springel, V., Pakmor, R., Pillepich, A., et al. 2018, *MNRAS*, **475**, 676
- Stein, Y., Vollmer, B., Boch, T., et al. 2021, *A&A*, **655**, A17
- Subrahmanya, C. R., Manoharan, P. K., & Chengalur, J. N. 2017, *JApA*, **38**, 10
- The HDF Group 1997, Hierarchical Data Format, version 5, <https://www.hdfgroup.org/solutions/hdf5/>
- Thompson, A. R., Moran, J. M., & Swenson, G. W., Jr. 2001, Interferometry and Synthesis in Radio Astronomy (2nd ed.; Wiley)
- Tingay, S. J., Goeke, R., Bowman, J. D., et al. 2013, *PASA*, **30**, e007
- Tramonte, D., & Ma, Y.-Z. 2020, *MNRAS*, **498**, 5916
- Trott, C. M., Nunhokee, C. D., Null, D., et al. 2025, arXiv:2508.04164
- van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, *A&A*, **556**, A2
- Vanderlinde, K., Liu, A., Gaensler, B., et al. 2019, *Canadian Long Range Plan for Astronomy and Astrophysics White Papers*, LRP 2020, 28
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
- Wang, H., Masui, K., Bandura, K., et al. 2025, *PhRvD*, **111**, 103531
- Wang, H., Mena-Parra, J., Chen, T., & Masui, K. 2022, *PhRvD*, **106**, 043534
- Wang, H., Phoompuang, P., Masui, K. W., Chakraborty, A., & Foreman, S. 2026, *PhRvD*, **113**, 043550
- Wang, N., Yan, Z., Manchester, R. N., & Wang, H. X. 2008, *MNRAS*, **385**, 1393
- Wolz, L., Pourtsidou, A., Masui, K. W., et al. 2022, *MNRAS*, **510**, 3495
- Wu, F., Li, J., Zuo, S., et al. 2021, *MNRAS*, **506**, 3455
- Zonca, A., Singer, L., Lenz, D., et al. 2019, *JOSS*, **4**, 1298