The Astrophysical Journal Supplement Series, 276:33 (18pp), 2025 February https://doi.org/10.3847/1538-4365/ad9604

Expanded Generalized Needlet Internal Linear Combination (eGNILC) Framework for the 21 cm Foreground Removal

Wei-Ming Dai[1] and Yin-Zhe Ma[2]
[1] School of Physical Science and Technology, Ningbo University, Ningbo 315211, People's Republic of China
[2] Department of Physics, Stellenbosch University, Matieland 7602, South Africa; mayinzhe@Sun.ac.za
Received 2024 May 07; revised 2024 November 01; accepted 2024 November 15; published 2025 January 13

Abstract

The Generalized Needlet Internal Linear Combination (GNILC) method is a nonparametric component separation algorithm to remove the foreground contamination of the 21 cm intensity mapping data. In this work, we perform the discrete cosine transform along the frequency axis in the expanded GNILC framework (denoted eGNILC), which helps reduce the power loss in low multipoles, and further demonstrates its performance. We also calculate the eGNILC bias to modify the criterion for determining the degrees of freedom (dof) of the foreground, and embed the robust principal component analysis in mixing matrix computation to obtain a blind component separation method. We find that the eGNILC bias is related to the averaged domain size and the dof of the foreground but not the underlying 21 cm signal. In the case of no beam effect, the eGNILC bias is negligible for simple power-law foregrounds outside the Galactic plane. We also examine the eGNILC performance in the SKA Phase-I in mid-frequency (SKA-MID) and Baryon Acoustic Oscillations from Integrated Neutral Gas Observations (BINGO) simulations. We show that if the adjacent frequency channels are not highly correlated, eGNILC can recover the underlying 21 cm signal with good accuracy. With the varying Airy-disk beam applied to both SKA-MID and BINGO, the power spectra of 21 cm can be effectively recovered at the multipoles $\ell \in [20, 250]$ and $[20, 300]$, respectively. With no instrumental noise, the SKA-MID exhibits $\lesssim$20% power loss, and BINGO exhibits ~10% power loss. The varying Airy-disk beam only causes significant errors at large multipoles.

Unified Astronomy Thesaurus concepts: Radio astronomy (1338); H I line emission (690)

1. Introduction

Intensity mapping (IM) of the neutral hydrogen 21 cm line has been studied as a promising technique to probe the large-scale structure (LSS) of the Universe (P. Madau et al. 1997; R. A. Battye et al. 2004; K. W. Masui et al. 2013; M. G. Santos et al. 2005, 2015; P. Bull et al. 2015; M. A. Bigot-Sazy et al. 2015, 2016; Y.-C. Li & Y.-Z. Ma 2017; C. J. Anderson et al. 2018; S. Harper et al. 2018; X. Xu et al. 2018; D. J. Bacon et al. 2020). Instead of measuring the emission from individual galaxies, IM measures integrated emission from a voxel of many galaxies in which the total signal can be detected by radio telescopes. Therefore, the integrated H I flux density in each pixel of sky maps contains multiple galaxies' contributions to the signal.

Many experiments have been conducted or proposed to detect the H I power spectrum from auto- or cross-correlations. The techniques are mainly of two kinds, the interferometer array and the single-dish (SD) measurement. The interferometer array measures the Fourier transform of the sky brightness in uv-plane (e.g., visibilities), which directly corresponds to the power spectra of the 21 cm signal. These include the Low-Frequency Array (B. K. Gehlot et al. 2019; F. G. Mertens et al. 2020), the Precision Array for Probing the Epoch of Reionization (A. R. Parsons et al. 2010; T. G. Gogo et al. 2022), the Murchison Widefield Array (C. M. Trott et al. 2020), and the Hydrogen Epoch Reionization Array (HERA; D. R. DeBoer et al. 2017; Z. Abdurashidova et al. 2022a, 2022b; The HERA Collaboration et al. 2023). The SD experiment, on the other hand, measures the specific flux of the 21 cm signal by direct imaging, which has achieved several detections in recent years via cross-correlation, including the HIPASS cross-correlation with 6dFGS galaxies (U.-L. Pen et al. 2009), Green Bank Telescope cross-correlation with DEEP2 galaxy survey (T.-C. Chang et al. 2010) and with WiggleZ survey (K. W. Masui et al. 2013), Parkes telescope cross-correlation with 2dFGS galaxies (C. J. Anderson et al. 2018), and stacking of Parkes 21 cm fields with galaxy halos (D. Tramonte & Y.-Z. Ma 2020) and filaments (D. Tramonte et al. 2019). More recently, S. Cunnington et al. (2022) utilized 10.5 hr observations of MeerKAT (64 dishes) drifted scan over ~200 $\deg^2$ and cross-correlated it with WiggleZ galaxies, and achieved 7.7σ detection of the cross-correlation signal. In the near future, the Five-hundred-meter Aperture Spherical radio Telescope (R. Nan et al. 2011; M. A. Bigot-Sazy et al. 2016), Baryon Acoustic Oscillations from Integrated Neutral Gas Observations (BINGO; R. A. Battye et al. 2012, 2013; C. Dickinson 2014; R. Battye et al. 2016; C. Wuensche 2019; E. Yohana et al. 2019; E. Abdalla et al. 2022; C. A. Wuensche et al. 2022), and the SKA Phase-I in mid-frequency (SKA-MID; D. J. Bacon et al. 2020) will also utilize the SD IM technique to measure the 21 cm brightness of the sky and probe the LSS.

In practice, the detection of H I IM signal is a great challenge because the foreground emission is at least 4 orders of magnitude higher than the underlying signal. Fortunately, the spectral shape of the H I signal is highly fluctuated, which is in principle, separable from the otherwise smooth continuum foreground emission. Unlike the interferometric experiments, the SD experiment has to remove foreground contaminants to achieve the auto- or cross-spectrum 21 cm fluctuations. Currently, there are two main techniques for foreground removal: parametric and nonparametric methods. Parametric

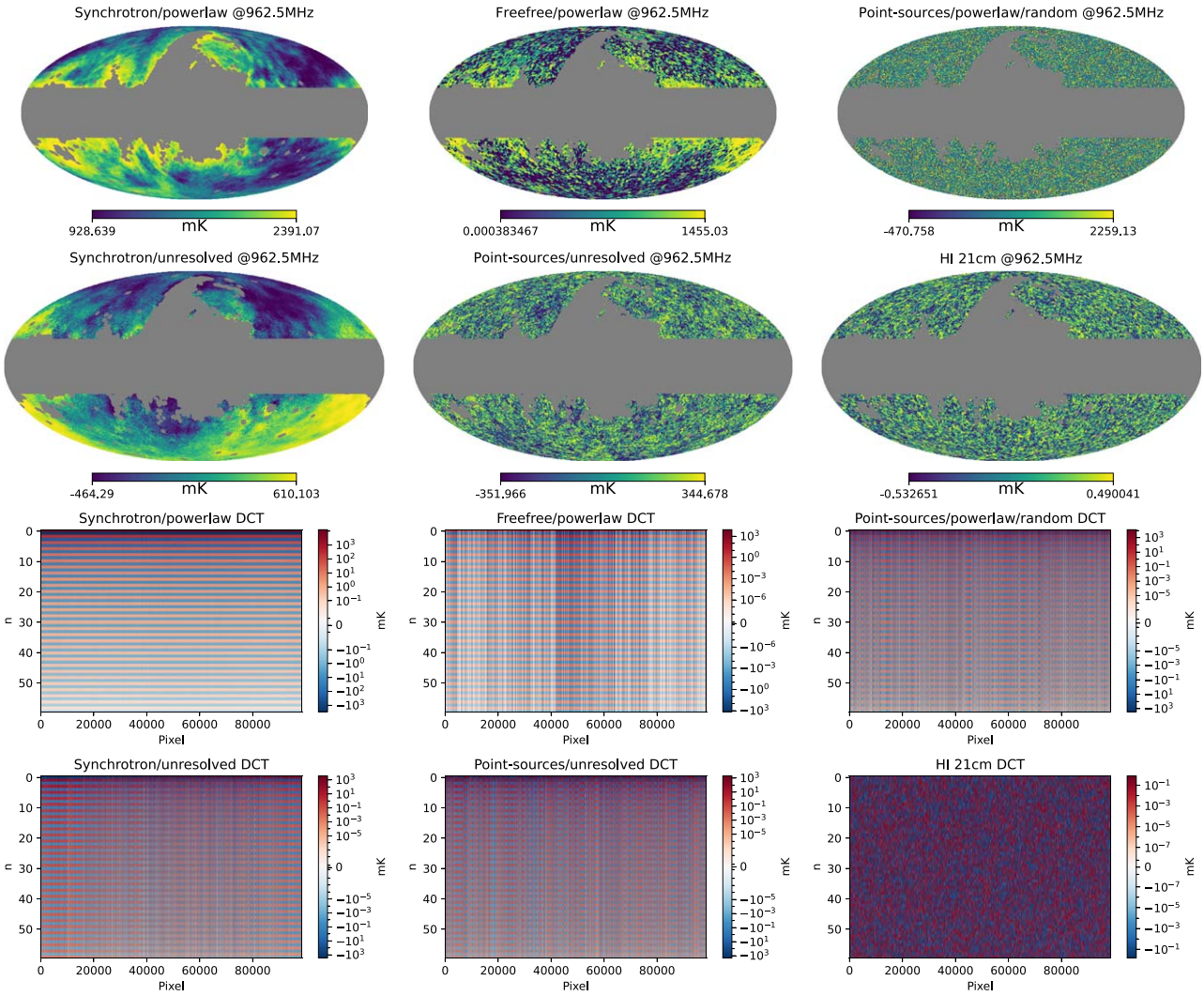

Figure 1. Sky maps of the synchrotron emission, free–free radiation, point sources, and 21 cm signal of H I at the frequency 962.5 MHz. All maps are normalized with histogram equalized color mapping. The discrete cosine transforms (DCTs) of 60 channels (962.5–1257.5 MHz) for individual catalogs are plotted in the square panels, where only unmasked pixels are used. The horizontal ticks label the pixel index, the vertical ticks label the DCT modes, and the color bar represents the DCT coefficients of maps.

methods use a (prior) model to describe the physical properties of the foregrounds. In contrast, the nonparametric methods try to minimize the use of prior knowledge of the foreground and rather use the observed data for signal extraction. Some nonparametric methods have been developed to extract 21 cm foregrounds, such as principal component analysis (PCA; K. W. Masui et al. 2013, D. Alonso et al. 2015, E. Yohana et al. 2021, S. Cunnington et al. 2022), independent component analysis (D. Alonso et al. 2015), and fast independent component analysis (D. Maino et al. 2002, L. Wolz et al. 2014).

In this paper, we study and expand the Generalized Needlet Internal Linear Combination (GNILC) method, which was first proposed in M. Remazeilles et al. (2011) and applied to the 21 cm IM in L. C. Olivari et al. (2016), to extend the generalization and robustness of the GNILC method. We denote the expanded GNILC as "eGNILC" in this paper. We will incorporate the robust PCA (RPCA) into the framework of GNILC, which becomes independent of the detailed knowledge of the underlying 21 cm signal. We name it as "RPCA-embedded" in the subsequent analysis of this paper. We will also evaluate the method in real-data scenarios and forecast its performance for BINGO and SKA-MID experiments.

This paper is organized as follows. In Section 2, we describe the models for the H I signal, foreground components, and instrumental noise. In Section 3, we present the mathematical framework of the eGNILC and make an extensive discussion of its function on foreground cleaning. Then in Section 4, we demonstrate the performance of the foreground removal methods, and apply it to the simulated data of the BINGO and SKA-MID configurations in Section 5. The conclusion is presented in Section 6.

2. Sky Maps

We simulate both H I signals and foregrounds in the interested frequency range. Figure 1 shows the simulated synchrotron emission, free–free radiation, point sources, and the H I 21 cm signal in the upper (elliptical) panels. The model of each component will be introduced in the following

subsections. Throughout this paper, the resolution of all HEALPix (K. M. Górski et al. 2005; A. Zonca et al. 2019) maps is set to $n_{\rm side} = 128$. In Figure 1 (lower, square, panels), we show the discrete cosine transform (DCT) of the 60 frequency channels spanning from 962.5–1257.5 MHz (center frequencies). To see the smoothness of each component along the frequency, we use the Python SCIPY package to conduct the DCT, which decomposes the data series at each pixel with cosine functions. We use the type-II DCT in the analysis, which is

$$y_k = 2f\sum_{n=1}^{N-1} x_n \cos\left(\frac{\pi k(2n+1)}{2N}\right), \tag{1}$$

where the scaling factor is

$$f = \begin{cases} \sqrt{1/4N} & \text{if } k = 0, \\ \sqrt{1/2N} & \text{otherwise}, \end{cases} \tag{2}$$

where N is the number of frequency channels, and the index n rum over frequencies.

2.1. H I 21 cm Intensity Map

The postreionization H I 21 cm signal traces the distribution of hydrogen. Ignoring the perturbations, the background brightness temperature of 21 cm is proportional to the number density of neutral hydrogen $\bar{n}_{\rm HI}$ (A. Hall et al. 2013),

$$\bar{T}_{\rm b}(z) = \frac{3(h_{\rm p}c)^3\bar{n}_{\rm HI}A_{10}}{32\pi k_{\rm B}E_{21}^2(1+z){\rm H}(z)}, \tag{3}$$

where $h_{\rm p}$ is Planck's constant, c is the speed of light in a vacuum, $k_{\rm B}$ is Boltzmann's constant, $A_{10} = 2.869\times10^{-15}\,{\rm s}^{-1}$ is the spontaneous emission coefficient, and $E_{21} = 5.87\,\mu$eV is the rest-frame energy of the 21 cm photon.

The perturbed 21 cm temperature is straightforward if we neglect the peculiar velocities and the Sachs–Wolfe effect (R. A. Battye et al. 2013; A. Hall et al. 2013),

$$\delta T_{\rm b}(z, \boldsymbol{r}(z)) = \bar{T}_{\rm b}(z)\delta_{\rm HI}(z, \boldsymbol{r}(z)), \tag{4}$$

where $\delta_{\rm HI}(z, \boldsymbol{r}(z))$ is the fluctuation of local H I overdensity, and $\boldsymbol{r}(z)$ is the comoving position of 21 cm sources. Making a Fourier transform and expressing $\delta_{\rm HI}(z, \boldsymbol{r}(z))$ in the spherical Bessel function basis, we have

$$\begin{aligned}\delta T_{\rm b}(z, \boldsymbol{r}(z)) &= \bar{T}_{\rm b}(z)\int\frac{d^3\boldsymbol{k}}{(2\pi)^3}\tilde{\delta}_{\rm HI}(\boldsymbol{k}, z)e^{i\boldsymbol{r}(z)\cdot\boldsymbol{k}} \\ &= 4\pi\bar{T}_{\rm b}(z)\textstyle\sum_{\ell m} i^\ell\displaystyle\int\frac{d^3\boldsymbol{k}}{(2\pi)^3}\tilde{\delta}_{\rm HI}(\boldsymbol{k}, z) \\ &\quad\times j_\ell(kr(z))Y^\star_{\ell m}(\hat{k})Y_{\ell m}(\hat{r}),\end{aligned} \tag{5}$$

where we used the decomposition of a plane wave $e^{i\boldsymbol{k}\cdot\boldsymbol{r}} = 4\pi\sum_{\ell m} i^\ell\, j_\ell(kr(z))\, Y^\star_{\ell m}(\hat{k})\, Y_{\ell m}(\hat{r})$. From Equation (5), we obtain the corresponding harmonic coefficients of redshift-slice z,

$$\begin{aligned}a_{\ell m}(z) &= 4\pi i^\ell\,\bar{T}_{\rm b}(z)\int\frac{d^3\boldsymbol{k}}{(2\pi)^3}\tilde{\delta}_{\rm HI}(\boldsymbol{k}, z) \\ &\quad\times j_\ell(kr(z))Y^\star_{\ell m}(\hat{k}).\end{aligned} \tag{6}$$

We now calculate the angular power spectra of 21 cm fluctuations. Because the density contrast is translationally and rotationally invariant, we can define the H I power spectra as

$$\begin{aligned}&\langle\tilde{\delta}_{\rm HI}(\boldsymbol{k}, z)\tilde{\delta}^*_{\rm HI}(\boldsymbol{k}', z')\rangle \\ &\quad=(2\pi)^3\delta_{\rm D}^{(3)}(\boldsymbol{k}-\boldsymbol{k}')P_{\rm HI}(k, z) \\ &\quad=(2\pi)^3\delta_{\rm D}^{(3)}(\boldsymbol{k}-\boldsymbol{k}')b^2P_{\rm m}(k)D(z)D(z'),\end{aligned} \tag{7}$$

where in the last line, we utilized the fact that the H I signal traces the underlying matter density field on large scales after reionization. $\delta_{\rm D}^{(3)}$ is the 3D Dirac delta function, $P_{\rm HI}$ and $P_{\rm m}$ are the H I and matter power spectra, $D(z)$ is the growth factor, and b is the bias between the spatial distribution of the H I and the matter field. Then, by ensemble-averaging the quadratic $a_{\ell m}$ (Equation (6)), we obtain the angular power spectrum of 21 cm fluctuations in redshift bins (z, z'; see also R. A. Battye et al. 2013),

$$\begin{aligned}C_\ell(z, z') &= \tfrac{2b^2}{\pi}\bar{T}_{\rm b}(z)\bar{T}_{\rm b}(z')D(z)D(z') \\ &\quad\times\int dk\; k^2P_{\rm m}(k)j_\ell(kr(z))j_\ell(kr(z')).\end{aligned} \tag{8}$$

We use the publicly available code Cosmology in the Radio Band (CORA)[3] to simulate the H I 21 cm intensity map at low redshifts (J. R. Shaw et al. 2014). The upper (elliptical) panels in Figure 1 show the maps at 962.5 MHz as an example. The amplitude of the 21 cm fluctuation is around 0.1 mK, which is roughly 4 orders of magnitude smaller than the foreground. The DCT of 21 cm maps (bottom-right panel) show correlations among frequencies are weak because the magnitude of each DCT mode is roughly at the same level. Therefore, a highly efficient foreground removal method and fast computational package are essential to extract the heavily contaminated signal.

Due to the limited sky coverage and Galactic contamination, observations are usually conducted in the partial sky, where the full-sky power is modulated with sky masks and mode-mixing effect. The "pseudo-C_ℓ" scheme shows that the pseudo–power spectrum $\tilde{C}_\ell$ on a cut sky is related to the full-sky power spectrum C_ℓ via

$$\langle\tilde{C}_\ell\rangle = \sum_{\ell'}M_{\ell\ell'}B^2_{\ell'}\langle C_{\ell'}\rangle, \tag{9}$$

where $M_{\ell\ell'}$ is mode-mixing matrix, and B_ℓ takes care of the telescope beam and finite pixelization effect (E. Hivon et al. 2002; G. Efstathiou 2004; K. Osato et al. 2020). The mode-coupling matrix is

$$M_{\ell_1\ell_2} = \frac{2\ell_2+1}{4\pi}\sum_{\ell_3}(2\ell_3+1)W_{\ell_3}\begin{pmatrix}\ell_1 & \ell_2 & \ell_3\\ 0 & 0 & 0\end{pmatrix}^2, \tag{10}$$

where W_ℓ is the power spectrum of the mask

$$\begin{aligned}W_\ell &= \tfrac{1}{2\ell+1}\sum_m w_{\ell m}w^*_{\ell m} \\ w_{\ell m} &= \int d^2\hat{\theta}\; w(\hat{\theta})Y^*_{\ell m}(\hat{\theta}),\end{aligned} \tag{11}$$

and $w(\boldsymbol{\theta})$ is the mask function. The last term in Equation (10) is the Wigner-3j symbol (Clebsch–Gordan coefficient; M. Abramowitz & I. A. Stegun 1972).

[3] https://github.com/radiocosmology/cora

2.2. Foregrounds

The Galactic synchrotron emission dominates the foregrounds at the frequency bands of 21 cm experiments. In addition, the free–free radio emission from ionized gas and extragalactic point sources are also much brighter than the underlying H I 21 cm signal. Because of the spectral smoothness of the foregrounds, many works have studied the scaling laws of the foregrounds in frequency direction to extrapolate the foreground at a given frequency from the template maps (C. G. T. Haslam et al. 1982; P. Platania et al. 1998; A. de Oliveira-Costa et al. 2008; M. A. Miville-Deschenes et al. 2008; A. Kogut 2012; M. Remazeilles et al. 2015). However, these templates may not be adequate for including all possible foreground components. In this work, we also add unresolved point sources to fully examine the robustness of our foreground removal method. We present each of the foreground components below. The synchrotron emission and point sources are simulated with the CORA code (J. R. Shaw et al. 2015) and the free–free radiation is simulated with PySM[4] (B. Thorne et al. 2017).

2.2.1. Extrapolation of Sky Maps

1. *Synchrotron emission*. The simulation of Galactic synchrotron emission is based on the all-sky 408 MHz Haslam map (C. G. T. Haslam et al. 1982; M. Remazeilles et al. 2015). New maps can be generated by scaling the brightness temperature map $T_{\rm Has}$ at desired frequencies with a power-law relation,

$$T_{\rm syn}(\hat{n},\,\nu) = T_{\rm Has}(\hat{n},\,408\ {\rm MHz}) \times \left(\frac{\nu}{408\ {\rm MHz}}\right)^{\alpha(\hat{n})}, \tag{12}$$

where $\alpha(\hat{n})$ is the index per pixel determined by the slope of the logarithmic intensities at 408 and 1420 MHz,

$$\alpha(\hat{n}) = \frac{\log T_{1420}(\hat{n}) - \log T_{408}(\hat{n})}{\log 1420 - \log 408}. \tag{13}$$

The 408 and 1420 MHz maps are generated by fitting the global sky model from surveys with different sky coverage and ranging from 10 MHz to 94 GHz (A. de Oliveira-Costa et al. 2008; see the first row, first column map in Figure 1 for a simulated synchrotron map).

2. *Free–free radiation*. We use a nominal model to simulate the free–free emission with the PySM code (first row, second column in Figure 1.) We first generate an intensity map at $\nu_0 = 30$ GHz, and then scale brightness temperature with a spatially constant power-law index -2.14 (B. T. Draine 2011), which is supposed to be valid for the frequency down to 10 MHz (R. Adam et al. 2016; B. Thorne et al. 2017):

$$T_{\rm b}^{\rm ff}(\hat{n}) = T_{\rm b0}^{\rm ff}(\hat{n})\left(\frac{\nu}{\nu_0}\right)^{-2.14}. \tag{14}$$

3. *Point sources*. Because bright point sources are well measured and cataloged, point sources with flux $S > 10$ Jy are assumed to be subtracted from the maps easily when we obtain an IM volume. Therefore, we generate medium-bright point sources (100 mJy $< S <$ 10 Jy) from the point-source distribution model of T. Di Matteo et al. (2002). We show an example of this generation in the first row, the third column of Figure 1. Because the discrete point-source map is not band-limited, spherical transforms of the map may be inaccurate, such as the negative pixel values shown in Figure 1. We will also introduce the unresolved foregrounds below to take care of this effect and other dim point sources.

2.2.2. Unresolved Random Sources

In addition to the power-law foreground maps, we also incorporate the extra unresolved random sources imported in the CORA code to fully study the effect of foreground removal of our computational method. The angular power spectrum of unresolved point sources can be parameterized as (J. R. Shaw et al. 2014; see also M. G. Santos et al. 2005)

$$C_\ell(\nu,\,\nu') = A\left(\frac{\ell}{\ell_0}\right)^{-\alpha}\left(\frac{\nu\nu'}{\nu_0^2}\right)^{-\beta} e^{-[\ln^2(\nu/\nu')]/2\zeta^2}, \tag{15}$$

where $\ell_0 = 100$ and $\nu_0 = 408$ MHz. The values of the parameters are assumed to be

$$\begin{aligned}[A\ ({\rm mK}^2),\ \alpha,\ \beta,\ \zeta]_{\rm syn} &= [6.6\times 10^3,\ 2.80,\ 2.8,\ 4.0\ ],\\ [A\ ({\rm mK}^2),\ \alpha,\ \beta,\ \zeta]_{\rm PS} &= [35.5,\ 2.10,\ 1.1,\ 1.0\ ],\end{aligned} \tag{16}$$

for synchrotron and point-source emissions, respectively (J. R. Shaw et al. 2014). Using Equation (15) and assuming Gaussian distribution, we simulate the unresolved synchrotron emission and point sources. They are shown in the second row, first and second columns in Figure 1, respectively.

Equation (15) shows that the unresolved random foreground depends on its frequency and spatial distribution. The DCT reflects important spectral and spatial features of the H I signal and foregrounds. For the power-law foreground contamination, the magnitudes of DCT coefficients decrease rapidly as the index n increases, which suggests that these foregrounds can mostly be determined by a few principal components. In other words, this feature makes the foreground easier to remove.

The spatial patterns are different in the DCT of power-law foregrounds because the synchrotron map is spatially smooth, while temperature fluctuations are more dramatic in the free–free and point-source maps. The unresolved maps of both synchrotron and point sources are stochastic "noises." Because the angular power spectrum of the unresolved synchrotron decreases faster (α is larger) than that of the unresolved point-source map, the map of the unresolved synchrotron foreground looks smoother than the unresolved point-source map. The horizontal stripes of the DCT images also suggest such spatial smoothness, while descending mode coefficients reflect the correlation of the ν and ν' angular power spectrum. An outstanding feature of the H I signal is that the mode coefficients do not decay visibly.

2.2.3. Instrumental Noise

The most common instrumental noise is the Gaussian distributed thermal noise. The rms of pixel noise is related to the system temperature, integration time, and frequency

[4] PySM is a publicly available Python code generating full-sky simulations of Galactic foregrounds in both intensity and polarization, relevant to CMB experiments.

Table 1
Experimental Specifications of Simulations for the Demonstration, the SKA-MID (D. J. Bacon et al. 2020; J. Wang et al. 2021) and the BINGO (R. A. Battye et al. 2012, 2013; C. Dickinson 2014; R. Battye et al. 2016; C. Wuensche 2019; E. Yohana et al. 2019; E. Abdalla et al. 2022; C. A. Wuensche et al. 2022) Configurations

Parameters	Values			Comments
	Demonstration	SKA-MID	BINGO	
$[z_{\min}, z_{\max}]$	[0.13, 0.48]	[0.33, 0.46]	[0.13, 0.48]	Redshift range
$[\nu_{\min}, \nu_{\max}]$ (MHz)	[960, 1260]	[970, 1070]	[960, 1260]	Band width
$\delta\nu$ (MHz)	15/5	0.2	5	Channel width
$\Omega_{\rm sur}$ (deg^2)	Full sky	20,000	5000	Sky coverage
$\theta_{\rm FWHM}$ (arcmin)	N/A	80@970 MHz	40@960MHz	Beam width

Note. Without loss of generality, we vary the channel width for convenience.

resolution (M. A. Bigot-Sazy et al. 2015; L. C. Olivari et al. 2016),

$$\sigma_{\rm pix} = \frac{T_{\rm sys}}{\sqrt{t_{\rm pix}\,\delta\nu}}, \tag{17}$$

where $\delta\nu$ is the width of each frequency channel, $T_{\rm sys}$ is the system temperature, and $t_{\rm pix}$ is the total integration time per pixel. We can simulate thermal noise according to different instrumental specifications in Table 1.

We notice that the $1/f$ noise also plays an important role in the radio antenna system. S. Harper et al. (2018) studied the impacts of $1/f$ noise on H I IM experiments by simulations with the following power-law model of the power spectral density (PSD),

$$\mathrm{PSD}(f,\,\omega) = \frac{T_{\rm sys}^2}{\delta\nu}\left[1 + C(\beta,\,N_\nu)\frac{(f_{\rm k}/f)^{\alpha}}{(\omega\Delta\nu)^{\frac{1-\beta}{\beta}}}\right], \tag{18}$$

where $f_{\rm k}$ is the knee frequency, α is the spectral index, ω is the inverse spectroscopic frequency wavenumber, $\Delta\nu$ is the total receiver bandwidth, β is used to parameterize the spectral index of the PSD[5], and $C(\beta, N_\nu)$ is a normalization factor. Because it is spectrally uncorrelated, pure thermal noise is irremovable but reducible with more integration time, while the $1/f$ noise is removable for data with high correlations between frequency channels. L. C. Olivari et al. (2016) explored the instrumental effects by adding thermal noise in the simulation. They presented the residual maps for the cases with no thermal noise and with thermal noise amplitudes equal to 0.05 mK and 0.08 mK. It is concluded that the larger the thermal noise amplitude is, the less accurate the reconstruction of the H I power spectrum on small angular scales is. They also suggest a solution in the case where there is a good estimate of the thermal noise power spectrum. The H I+ noise covariance matrix is used as the prior instead of the H I covariance matrix, so that we recover the H I signal plus thermal noise. We may correct for the thermal noise by using the estimate of the thermal noise power spectrum. For effects of the $1/f$ noise, readers may refer to S. Harper et al. (2018) and E. Yohana et al. (2021). Therefore, we primarily focus on removing the foreground contamination, and will not discuss the impacts of thermal noise or $1/f$ instrumental noise in the subsequent discussion.

[5] We remind the reader not to confuse the meaning of α and β here with the parameters defined in unresolved foreground sky (Section 2.2.2).

3. Foreground Removal Method

The GNILC method was first applied to 21 cm IM data by L. C. Olivari et al. (2016). We will utilize this framework in general, but replace the theoretical cross-correlation matrix of the 21 cm signal with the one estimated by the RPCA method in one case. We will also elaborate on the refinement of the GNILC method and highlight the differences and improvements.

3.1. GNILC

We denote the observed sky at the pixel p and frequency i as $x_i(p)$, which sums the fluctuation of H I 21 cm emission $s_i(p)$ and the foreground components $f_i(p)$. We remind the reader that p does not necessarily mark the pixels in the real space, and it could be the index of the needlet coefficient in the GNILC. Similarly, the index i may label any variant of frequency, such as the mode index of DCT shown in Figure 1. We may write the dependence of the frequency index in the vector form as

$$\boldsymbol{x}(p) = \boldsymbol{s}(p) + \boldsymbol{f}(p), \tag{19}$$

where $\boldsymbol{x}(p)$, $\boldsymbol{s}(p)$, and $\boldsymbol{f}(p)$ are $n_{\rm ch}$-dimensional column vectors, and $n_{\rm ch}$ is the number of frequency channels. The signal $\boldsymbol{s}$ is, of course, independent of the foreground $\boldsymbol{f}$. The bold characters denote the vectors (lowercase) or matrices (uppercase) defined in the frequency basis. From Equation (19), the covariance matrix at pixel p is

$$\boldsymbol{R}_{\boldsymbol{x}}(p) = \boldsymbol{R}_{\boldsymbol{s}}(p) + \boldsymbol{R}_{\boldsymbol{f}}(p), \tag{20}$$

where the three $\boldsymbol{R}$ s are the covariance matrices of $\boldsymbol{x}$, $\boldsymbol{s}$, and $\boldsymbol{f}$, respectively. Below we suppress the variable p for brevity.

Because of the smoothness of foreground on frequency direction, the number of independent modes of foregrounds m is much less than the frequency channel number $n_{\rm ch}$, i.e., $m \ll n_{\rm ch}$. As a consequence, the recovered H I signal $\hat{s}(p)$ may only have $n_{\rm ch} - m$ degrees of freedom (dof), regardless of its real dof. For the case where thermal noise is presented, the noise still exists in the H I 21 cm signal after the foreground removal procedure because only the frequency-correlated contaminants are removed. Because of dof $= n_{\rm ch} - m$, the estimated 21 cm signal $\hat{\boldsymbol{s}}$ is a linear combination of independent templates $\boldsymbol{t}$ as

$$\hat{\boldsymbol{s}} = \boldsymbol{S}\boldsymbol{t}, \tag{21}$$

where $\boldsymbol{S}$ is an $n_{\rm ch} \times (n_{\rm ch} - m)$ mixing matrix. Then the covariance matrix of $\hat{\boldsymbol{s}}$ is the transformation of the full-rank matrix $\boldsymbol{R}_{\rm t}$:

$$\boldsymbol{R}_{\hat{\boldsymbol{s}}} = \boldsymbol{S}\boldsymbol{R}_{\boldsymbol{t}}\boldsymbol{S}^T, \tag{22}$$

where $\boldsymbol{R}_{\rm t}$ is the covariance matrix of $\boldsymbol{t}$. One should realize that the mixing matrix ($\boldsymbol{S}$) and template set ($\boldsymbol{t}$) are not unique, because, for any given set of templates, a new set of templates ($\boldsymbol{t}'$) can be generated with an invertible linear transformation,

$$\boldsymbol{t}' = \boldsymbol{V}^{-1}\boldsymbol{t}, \tag{23}$$

$$\boldsymbol{R}_{\hat{\boldsymbol{s}}} = \boldsymbol{S}\boldsymbol{V}\boldsymbol{R}_{\boldsymbol{t}'}\boldsymbol{V}^{T}\boldsymbol{S}^{T}. \tag{24}$$

For a given mixing matrix, we want to find an "ILC weighing matrix" $\boldsymbol{W}$

$$\hat{\boldsymbol{s}} = \boldsymbol{W}\boldsymbol{x} \tag{25}$$

to extract the H I signal $\hat{\boldsymbol{s}}$ (which is generally believed to deviate from $\boldsymbol{s}$) such that the total variance is minimized, and the 21 cm signal is unbiased,

$$\min[\mathrm{Tr}(\boldsymbol{W}\boldsymbol{R}_{\boldsymbol{x}}\boldsymbol{W}^{T})], \quad \text{s.t.} \quad \boldsymbol{W}\boldsymbol{S} = \boldsymbol{S}. \tag{26}$$

The solution of this conditional minimization is analogous to the famous "map-making" equation (M. Tegmark & G. Efstathiou 1996; S. Dodelson & F. Schmidt 2020),

$$\boldsymbol{W} = \boldsymbol{S}(\boldsymbol{S}^{T}\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^{T}\boldsymbol{R}_{\boldsymbol{x}}^{-1}. \tag{27}$$

Therefore, with the weighing matrix (Equation (27)), we can implement the needlet-space ILC with the following steps:

1. *Analysis.* First, the real map at a particular frequency is needlet-transformed while the needlets have a finite spectral support adjustable at will[6] and have a good spatial localization.

$$\{x(p)\}_{p\in\mathcal{P}} \overset{\mathrm{SHT}}{\rightarrow} \{a_{\ell m}\} \overset{\times}{\rightarrow} \{b_{\ell}^{(j)}a_{\ell m}\} \overset{\mathrm{SHT}^{-1}}{\rightarrow} \{x^{(j)}(p')\}_{p'\in\mathcal{P}^{(j)}},$$

   where $x(p)$ is the data at each pixel p on the map $\mathcal{P}$, SHT is the spherical harmonics transform, SHT^{-1} is the inverse transform, $b_{\ell}^{(j)}$ is the window family that defines the needlet bases, and $\{x^{(j)}(p')\}_{p'\in\mathcal{P}^{(j)}}$ is the projected map corresponding to each needlet base. The 21 cm signal will be extracted from individual needlet map and then synthesized with the following steps.
2. *Signal extraction.* As illustrated above, we obtained the expected signal extracted from the decomposed data in the needlet space with Equation (25). As the weighing matrix is yet to be solved, Section 3.1.1 will introduce the covariance ($\boldsymbol{R}_{\boldsymbol{x}}$) and mixing matrics ($\boldsymbol{S}$), which are crucial for Equations (27) and (25) to extract the underlying signal.
3. *Synthesis.* We finally synthesize the extracted signal from each needlet base, which is the reversal of analysis.

$$\{\hat{s}^{(j)}(p')\}_{p'\in\mathcal{P}^{(j)}} \overset{\mathrm{SHT}}{\rightarrow} \{b_{\ell}^{(j)}a_{\ell m}\} \overset{\times}{\rightarrow} \{\tilde{b}_{\ell}^{(j)}b_{\ell}^{(j)}a_{\ell m}\} \overset{\mathrm{SHT}^{-1}}{\rightarrow} \{\Psi^{(j)}\hat{s}(p)\}_{p\in\mathcal{P}},$$

   where $\{\tilde{b}_{\ell}^{(j)}\}$ is the synthesis window family that ensures the reconstruction unbias condition $\sum_{j\in\mathcal{J}}\tilde{b}_{\ell}^{(j)}b_{\ell}^{(j)} = 1$, and $\Psi^{(j)}$ is the resultant smooth operator at needlet mode $j \in \mathcal{J}$. The symbol $\hat{s}$ refers to the estimate of s. At each frequency, we get the expected 21 cm map by summing up all needlet modes $\hat{s}(p) = \sum_{j\in\mathcal{J}}\Psi^{(j)}\hat{s}^{(j)}$.

[6] In other words, each needlet map is band-limited according to the choice of window function.

3.1.1. Mixing Matrix

Before determining the mixing matrix $\boldsymbol{S}$, we need to calculate covariance matrix of the signal. In general, the sky maps observed at $n_{\rm ch}$ frequency channels ($\boldsymbol{\Theta}(p)$) can be decomposed with spherical harmonics,

$$\boldsymbol{\Theta}(p) = \sum_{\ell m}\boldsymbol{a}_{\ell m}Y_{\ell m}(p) = \sum_{j\in\mathcal{J}}\boldsymbol{\Theta}^{(j)}(p), \tag{28}$$

where $Y_{\ell m}(p)$ is the spherical harmonic function, and the vector $\boldsymbol{\Theta}(p)$ represents the pth pixel values on maps at all frequency channels. $\boldsymbol{\Theta}^{(j)}(p) = \sum_{\ell m}b_{\ell}^{(j)}\boldsymbol{a}_{\ell m}Y_{\ell m}(p)$ is the jth needlet component. If the statistical isotropy holds, the covariance matrix can be calculated as

$$\begin{aligned}
\boldsymbol{R}^{(j)} &= \langle\boldsymbol{\Theta}^{(j)}\boldsymbol{\Theta}^{(j)\,T}\rangle \\
&= \langle\boldsymbol{\Theta}^{(j)}\boldsymbol{\Theta}^{(j)\,\dagger}\rangle \\
&= \sum_{\ell,m,m'}\langle(b_{\ell}^{(j)}\boldsymbol{a}_{\ell m})(b_{\ell}^{(j)}\boldsymbol{a}_{\ell m'}^{\dagger})\rangle Y_{\ell m}Y_{\ell m'}^{*} \\
&= \sum_{\ell,m,m'}\delta_{mm'}B_{\ell}^{(j)}\boldsymbol{C}_{\ell}Y_{\ell m}Y_{\ell m'}^{*} \\
&= \sum_{\ell}B_{\ell}^{(j)}\boldsymbol{C}_{\ell}\textstyle\sum_{m}Y_{\ell m}Y_{\ell m}^{*} \\
&= \sum_{\ell}\frac{2\ell+1}{4\pi}B_{\ell}^{(j)}\boldsymbol{C}_{\ell},
\end{aligned} \tag{29}$$

where the covariance matrix is pixel-independent, $B_{\ell}^{(j)} = b_{\ell}^{(j)}b_{\ell}^{(j)}$, and the matrix $\boldsymbol{C}_{\ell}$ consists of $n_{\rm ch}\times n_{\rm ch}$ cross-power spectra at ℓ.

However, foregrounds are not isotropic. Instead of Equation (29), we calculate the covariance matrix of each pixel from the sample average, because maps are smooth at a specific scale for a given choice of needlet bases. For each needlet mode j, the covariance matrix at pixel p is an average of the domain $\mathcal{D}$ around p. The effective independent pixel number is often smaller than the true pixel number in the domain because nearby pixels are correlated, especially for the needlet coefficient maps that are filtered with spectral window functions. The domain $\mathcal{D}$ can be defined by convolving maps with a symmetric Gaussian window in the real space so that the average is easily done in the harmonic domain. The size of $\mathcal{D}$ is related to the smoothing scale of the jth needlet, but the GNILC is insensitive to the detailed form of $\mathcal{D}$. We will use this average as the estimate of covariance in our analysis. However, there is a caveat to this approach: the limited number of independent samples may cause (artificial) anticorrelation between the signal and contaminants, which leads to a power loss in the recovered signal (J. Delabrouille et al. 2009; L. C. Olivari et al. 2016). We will derive this bias in Section 3.1.2.

The spectral window function determines how the needlet map is localized in both the spatial and spherical harmonics domains (refer to Appendix A). Thus, the choice of the parameter B should be in line with the spatial and spectral features of the 21 cm signal and foregrounds. In addition to the dependence on the spectral window function of the needlet space, the determination of $\mathcal{D}$ is also affected by the foregrounds, masks, and telescope beams. We will show these

effects in Sections 4 and 5. In our analysis, we use the effective $\theta_{\rm FWHM}$ to quantify the needlet, beam, pixel size, and sky coverage in order to determine the domain $\mathcal{D}$ and N_p—the effective number of independent pixels in $\mathcal{D}$. First of all, bear in mind that we want a large N_p to obtain an accurate estimate of the covariance matrix and reduce the bias (see Section 3.1.2). The independent size is limited by the map pixel size, the telescope beam, and the needlet size, which should be smaller than those three sizes. On the other hand, the maximum $\mathcal{D}$ is limited by the sky coverage (or mask), because if the $\mathcal{D}$ is significantly larger than the sky coverage, the data is averaged with the null field. In practice, the data and the null field are indeed blended together in the needlet domain, to a degree.

We can now calculate the mixing matrix $\boldsymbol{S}$. In the following, we will frequently use the square root of the real symmetric, positive-definite matrix $\boldsymbol{R}_s$, which we define via its eigendecomposition, i.e.,

$$\boldsymbol{R}_s = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T \Rightarrow \boldsymbol{R}_s^{1/2} = \boldsymbol{Q}\boldsymbol{\Lambda}^{1/2}\boldsymbol{Q}^T. \tag{30}$$

By using Equation (20), we have

$$\boldsymbol{R}_s^{-1/2}\boldsymbol{R}_x\boldsymbol{R}_s^{-1/2} = \boldsymbol{R}_s^{-1/2}\boldsymbol{R}_f\boldsymbol{R}_s^{-1/2} + \boldsymbol{I}, \tag{31}$$

where the superscript (j) is omitted for simplicity. Note that any covariance matrix is a positive semidefinite matrix, and the signal is nonzero. Thus, we can diagonalize the left-hand side of Equation (31) as

$$\begin{aligned}&\boldsymbol{R}_s^{-1/2}\boldsymbol{R}_x\boldsymbol{R}_s^{-1/2}\\ &=[\boldsymbol{U}_s\ \boldsymbol{U}_f]\begin{bmatrix}\boldsymbol{I}_{\{n_{\rm ch}-m\}} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{D}_f\end{bmatrix}\begin{bmatrix}\boldsymbol{U}_s^T\\ \boldsymbol{U}_f^T\end{bmatrix}\\ &=\boldsymbol{U}_f\boldsymbol{D}_f\boldsymbol{U}_f^T + \boldsymbol{U}_s\boldsymbol{U}_s^T,\end{aligned} \tag{32}$$

where $\boldsymbol{U}_s$ is an $n_{\rm ch} \times (n_{\rm ch} - m)$ submatrix, $\boldsymbol{I}_{\{n_{\rm ch}-m\}}$ is an $(n_{\rm ch} - m) \times (n_{\rm ch} - m)$ unit matrix, and $\boldsymbol{D}_f = {\rm diag}(\lambda_1 + 1, \cdots, \lambda_m + 1)$ $(\lambda_i \gg 1)$ are foreground modes. $\boldsymbol{U}_{\rm f}$ is an $n_{\rm ch} \times m$ submatrix, and $\boldsymbol{U}_f\boldsymbol{U}_f^T + \boldsymbol{U}_s\boldsymbol{U}_s^T = \boldsymbol{I}$. Then, the following expression gives an approximated foreground covariance matrix,

$$\begin{aligned}\boldsymbol{R}_f &= \boldsymbol{R}_x - \boldsymbol{R}_{\rm s}\\ &= \boldsymbol{R}_s^{1/2}(\boldsymbol{R}_s^{-1/2}\boldsymbol{R}_x\boldsymbol{R}_s^{-1/2} - \boldsymbol{I})\boldsymbol{R}_s^{1/2}\\ &= \boldsymbol{R}_s^{1/2}(\boldsymbol{U}_f(\boldsymbol{D}_f - \boldsymbol{I}_{\{\rm m\}})\boldsymbol{U}_f^T)\boldsymbol{R}_s^{1/2}\\ &\simeq \boldsymbol{R}_s^{1/2}(\boldsymbol{U}_f\boldsymbol{D}_f\boldsymbol{U}_f^T)\boldsymbol{R}_s^{1/2},\end{aligned} \tag{33}$$

where the last approximate equality is because of $\lambda_i \gg 1$. With this approximation, we can obtain the estimated H I 21 cm emission covariance matrix,

$$\begin{aligned}\boldsymbol{R}_{\hat{s}} &= \boldsymbol{R}_x - \boldsymbol{R}_f\\ &= \boldsymbol{R}_s^{1/2}(\boldsymbol{U}_s\boldsymbol{U}_s^T)\boldsymbol{R}_s^{1/2}\\ &= (\boldsymbol{R}_s^{1/2}\boldsymbol{U}_s)(\boldsymbol{R}_s^{1/2}\boldsymbol{U}_s)^T,\end{aligned} \tag{34}$$

where we substituted Equations (32) and (33) in the second equality.

Using Equation (24) and choosing $\boldsymbol{V} = \boldsymbol{R}_{t'}^{-1/2}$ (i.e., $\boldsymbol{V}\boldsymbol{R}_{t'}\boldsymbol{V}^{\rm T} = \boldsymbol{I}$), we have $\boldsymbol{R}_{\hat{s}} = \boldsymbol{S}\boldsymbol{S}^T$, then we obtain a proper estimate of the mixing matrix,

$$\boldsymbol{S} = \boldsymbol{R}_s^{1/2}\boldsymbol{U}_s. \tag{35}$$

So once we know the covariance matrix of the 21 cm signal, we can use Equation (35) to obtain the mixing matrix. An essential precondition is to figure out the degrees of freedom of the foregrounds, i.e., m-values, which we will elaborate on in greater detail in Section 3.1.2.

In summary, the key step of GNILC (eGNILC) is to find the eigenvectors and eigenvalues in Equation (32) to determine the mixing matrix. In Equation (35), we have only used the spatial correlations between maps to calculate the covariance matrix. But if the accuracy of component separation can be improved, the estimated 21 cm signal will be more precise. The DCT method provides such a possibility because it helps select a better basis for the GNILC by using the frequency-smoothness of the foregrounds. We will examine the improvement in Section 4.

3.1.2. Degrees of Freedom of the Foregrounds

Though we have assumed those eigenvalues of $\boldsymbol{R}_s^{-1/2}\boldsymbol{R}_x\boldsymbol{R}_s^{-1/2}$ that correspond to the 21 cm signal are unity, practically the critical eigenvalue remains unknown to us due to imperfect estimation of the covariance matrices. Given the data $\boldsymbol{x}(p)$, we can calculate the likelihood of the mode illustrated in Section 3.1.1. To select the best value of m and prevent overfitting, L. C. Olivari et al. (2016) suggested the Akaike information criterion (AIC),

$${\rm AIC}(m) = 2m - 2\log(\mathcal{L}_{\rm max}(m)), \tag{36}$$

where $\mathcal{L}_{\rm max}(m)$ is the maximum likelihood with regard to the model parameter m for the given data,

$$-2\log(\mathcal{L}_{\rm max}(m)) = \sum_{i=1}^{n_{\rm ch}-m}[\mu_i - \ln\mu_i - 1], \tag{37}$$

where we neglect the constant in the logarithmic likelihood, μ_i is the eigenvalues of $\boldsymbol{R}_s^{-1/2}\boldsymbol{R}_x\boldsymbol{R}_s^{-1/2}$, and $2m$ is the penalty term that prevents overfitting. For details of the derivation, please refer to Appendix B of L. C. Olivari et al. (2016).

The above AIC only explicitly uses the likelihood of $\boldsymbol{x}(p)$, but in fact, the 21 cm prior maps can also be used when calculating the likelihood because they are the "prior knowledge" of reconstructed signal $\hat{\boldsymbol{s}}$. Here we do not distinguish the difference between the real signal and the prior because we assume their covariance matrices are very close. We will see from the following derivation that this additional likelihood (except removable constants) is independent of the prior maps under certain assumptions.

Like the ILC bias estimation in J. Delabrouille et al. (2009), we will take a similar procedure to derive the GNILC error for the 21 cm reconstruction. The reconstructed 21 cm signal $\hat{\boldsymbol{s}}$ at each pixel is

$$\hat{\boldsymbol{s}} = \boldsymbol{s} + \boldsymbol{\delta}, \tag{38}$$

where $\boldsymbol{\delta}$ is the reconstruction error, and $\boldsymbol{s}$ is the underlying true signal. Thus, the covariance of $\hat{\boldsymbol{s}}$ is

$$\hat{\boldsymbol{R}}_s = \boldsymbol{R}_s + 2\boldsymbol{C}_{s\delta} + \boldsymbol{R}_\delta, \tag{39}$$

where $\hat{\boldsymbol{R}}_s$, $\boldsymbol{R}_s$, and $\boldsymbol{R}_\delta$ are the covariance matrices of each term in Equation (38). $\boldsymbol{C}_{s\delta}$ is the covariance between the real 21 cm signal and the reconstruction error. We note that $\hat{\boldsymbol{R}}_s$ is different from that in Equation (34), because the latter is an intermediate estimation to calculate mixing matrix.

We assume to reach a small GNILC error $\boldsymbol{\delta}$ such that the $\boldsymbol{R}_{\boldsymbol{\delta}}$ term can be neglected in Equation (39). This assumption is reasonable because we will achieve a low reconstruction bias after the foreground removal. The $\boldsymbol{C}_{s\delta}$, which couples the error to the signal and leads to signal loss, is of interest to us. For the sake of neatness, the signals are assumed/transformed to a normal distribution with mean zero. Because the covariance remains unchanged, we will not distinguish the differences in mean values, so we have $\boldsymbol{C}_{\boldsymbol{ab}} = E(\boldsymbol{ab}^{\mathrm{T}})$, where E denotes the ensemble average. From Equation (38), we can calculate the error as

$$\begin{aligned}\boldsymbol{\delta} &= \hat{\boldsymbol{s}} - \boldsymbol{s} = \boldsymbol{W}(\boldsymbol{s}+\boldsymbol{f}) - \boldsymbol{s} = \boldsymbol{W}\boldsymbol{f} \\ &= \boldsymbol{S}(\boldsymbol{S}^{\mathrm{T}}\hat{\boldsymbol{R}}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^{\mathrm{T}}\hat{\boldsymbol{R}}_{\boldsymbol{x}}^{-1}\boldsymbol{f}, \end{aligned} \tag{40}$$

where we have utilized Equations (19) and (25) in the second equality, and Equation (27) in the fourth equality. The third equality is due to the unbiasedness of the operator $\boldsymbol{W}$ (i.e., $\boldsymbol{W}\boldsymbol{s} = \boldsymbol{s}$). $\hat{\boldsymbol{R}}_{\boldsymbol{x}}$ is the estimation of $\boldsymbol{R}_{\boldsymbol{x}}$ by calculating it in a domain $\mathcal{D}$ around a pixel p, which is related to the true covariance matrix $\boldsymbol{R}_{\boldsymbol{x}}$ via $\hat{\boldsymbol{R}}_{\boldsymbol{x}} = \boldsymbol{R}_{\boldsymbol{x}} + \boldsymbol{\Delta}_{\boldsymbol{x}}$. Thus, $\boldsymbol{\Delta}_{\boldsymbol{x}}$ accounts for the error of the $\boldsymbol{R}_{\boldsymbol{x}}$ matrix. $\hat{\boldsymbol{R}}_{\boldsymbol{x}}^{-1}$ can be further expanded to first order as

$$\hat{\boldsymbol{R}}_{\boldsymbol{x}}^{-1} \simeq \boldsymbol{R}_{\boldsymbol{x}}^{-1} - \boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{\Delta}_{\boldsymbol{x}}\boldsymbol{R}_{\boldsymbol{x}}^{-1}. \tag{41}$$

Substituting Equation (41) into Equation (40), we have

$$\begin{aligned}\boldsymbol{\delta} &= \boldsymbol{S}[(\boldsymbol{S}^{T}\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1} + (\boldsymbol{S}^{T}\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1} \\ &\quad \times (\boldsymbol{S}^{T}\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{\Delta}_{\boldsymbol{x}}\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})(\boldsymbol{S}^{T}\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}] \\ &\quad \times \boldsymbol{S}^{T}(\boldsymbol{R}_{\boldsymbol{x}}^{-1} - \boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{\Delta}_{\boldsymbol{x}}\boldsymbol{R}_{\boldsymbol{x}}^{-1})\boldsymbol{f} \\ &= \boldsymbol{S}(\boldsymbol{S}^{T}\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^{T}\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{f} \qquad (42) \\ &\quad -\boldsymbol{S}(\boldsymbol{S}^{T}\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^{T}\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{\Delta}_{\boldsymbol{x}}\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{f} \qquad (43) \\ &\quad +\boldsymbol{S}(\boldsymbol{S}^{T}\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}(\boldsymbol{S}^{T}\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{\Delta}_{\boldsymbol{x}}\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S}) \\ &\quad \times(\boldsymbol{S}^{T}\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^{T}\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{f} \qquad (44) \\ &\quad +O(\Delta_{\mathrm{x}}^{2})\boldsymbol{f}, \qquad (45)\end{aligned}$$

where Equation (45) is negligible because it is the second order of $\boldsymbol{\Delta}_{\boldsymbol{x}}$. The correlation $\boldsymbol{C}_{s\delta} = \mathrm{E}(\boldsymbol{\delta}\boldsymbol{s}^{T})$ comprises correlations from Equations (42) to (45), and Equation (42) will vanish because $\boldsymbol{s}$ and $\boldsymbol{f}$ are uncorrelated.

The covariance matrix $\hat{\hat{\boldsymbol{R}}}$ of $\boldsymbol{s}$ and $\boldsymbol{f}$ is calculated in a domain $\mathcal{D}$ around a pixel p in the following equation. We first calculate the error $\boldsymbol{\Delta}_{\boldsymbol{x}}$ at the pixel p

$$\begin{aligned}\boldsymbol{\Delta}_{\boldsymbol{x}}(p) &= \hat{\boldsymbol{R}}_{\boldsymbol{x}} - \boldsymbol{R}_{\boldsymbol{x}} \\ &= \tfrac{1}{N_p}\textstyle\sum_q(\boldsymbol{s}_q+\boldsymbol{f}_q)(\boldsymbol{s}_q+\boldsymbol{f}_q)^{T} - (\boldsymbol{R}_{\boldsymbol{s}}+\boldsymbol{R}_{\boldsymbol{f}}) \\ &= \tilde{\boldsymbol{R}}_{\boldsymbol{s}} - \boldsymbol{R}_{\boldsymbol{s}} + \tilde{\boldsymbol{R}}_{\boldsymbol{f}} - \boldsymbol{R}_{\boldsymbol{f}} \\ &\quad + \tfrac{1}{N_p}\textstyle\sum_q(\boldsymbol{s}_q\boldsymbol{f}_q^{T} + \boldsymbol{f}_q\boldsymbol{s}_q^{T}), \end{aligned} \tag{46}$$

where N_p is the effective number of independent pixels in domain $\mathcal{D}$, and q sums over all pixels. $\tilde{\boldsymbol{R}}_{\boldsymbol{s}}$ is defined as $\tilde{\boldsymbol{R}}_{\boldsymbol{s}} = (1/N_p)\sum \boldsymbol{s}_q\boldsymbol{s}_q^{T}$ and the same for $\tilde{\boldsymbol{R}}_{\boldsymbol{f}}$.

Through a lengthy calculation of the actual signal covariance matrix (Appendix B), we obtain the covariance of $\hat{\boldsymbol{s}}$ as

$$\hat{\boldsymbol{R}}_{\boldsymbol{s}} = \left(1 - 2\frac{m}{N_p}\right)\boldsymbol{R}_{\boldsymbol{s}}, \tag{47}$$

which is smaller than the signal covariance $\boldsymbol{R}_{\boldsymbol{s}}$ by a factor of $2m/N_p$.

We can finally reach the AIC value of the likelihood with m-modes of foreground and N_{p} pixels as

$$\begin{aligned}\mathrm{AIC}(m, N_{\mathrm{p}}) &= 2m + n_{\mathrm{ch}}\left[\frac{1}{1-2m/N_{\mathrm{p}}} + \ln\left(1 - 2\frac{m}{N_{\mathrm{p}}}\right)\right] \\ &\quad + \sum_{i=1}^{n_{\mathrm{ch}}-m}[\mu_i - \ln\mu_i - 1], \end{aligned} \tag{48}$$

where the additional term is scaled by the channel number n_{ch}. It is not an explicit function of the 21 cm prior but the number of channels and the number of independent pixels are involved. The effect of this additional term is not noticeable if the term m/N_{p} is small (most cases), but the modified AIC might be distinguished for large n_{ch}.

3.2. RPCA Blind Foreground Removing

In Section 3.1.1, we assume the covariance matrix of 21 cm signal is computed through known maps, which means that the GNILC depends on the prior power spectrum of H I. It is verified that the GNILC method is not critically sensitive to the absence of detailed features in the prior H I power spectrum but depends on the overall amplitude of the H I power spectrum (L. C. Olivari et al. 2016). Under particular circumstances where the RPCA is efficient, the mixing matrix S can be derived from the cross-correlation matrix of data directly without the input of simulated H I 21 cm maps. S. Zuo et al. (2018) implemented RPCA to extract the H I 21 cm signal, based on the low-rank property of the foreground cross-correlation matrix and the sparsity of the H I 21 cm cross-correlation matrix. Here, we can apply the RPCA method within the GNILC framework and establish a blind method that retains the merits of the GNILC method. Let matrix $\boldsymbol{R}_{\mathrm{L}}$ and $\boldsymbol{R}_{\mathrm{S}}$ be the low-rank and sparse components of the data covariance matrix $\boldsymbol{R}_{\boldsymbol{x}}$ found by the RPCA algorithm; we obtain the estimation of 21 cm cross-correlation matrix $\boldsymbol{R}_{\hat{\boldsymbol{s}}} = \boldsymbol{R}_{\mathrm{S}}$. With this replacement, we take the standard steps of GNILC to extract the H I 21 cm signal. We test its performance with simulated data, and show the demonstration in Section 4.

The noise effect is not tested in the RPCA-embedded GNILC approach. When extracting the H I signals from the 21 cm IM data, the larger the thermal noise amplitude is, the less accurate the reconstruction of the H I power spectrum on small angular scales is (L. C. Olivari et al. 2016). A solution is also suggested in the case that we have a good estimate of the thermal noise power spectrum. The H I+ noise covariance matrix is used as the prior instead of the H I covariance matrix, so that we recover the H I signal plus thermal noise. We may correct for the thermal noise by using the estimate of the instrumental noise power spectrum. This correction can also be applied to the RPCA-embedded GNILC approach where the prior covariance is replaced with the data covariance splitting. The sparse matrix comprises the H I covariance, and the thermal covariance (diagonal for simple models), which is roughly equivalent to the H I plus noise prior. So we expect that the performance of the RPCA-embedded GNILC approach is close to the GNILC with H I signal plus thermal noise prior.

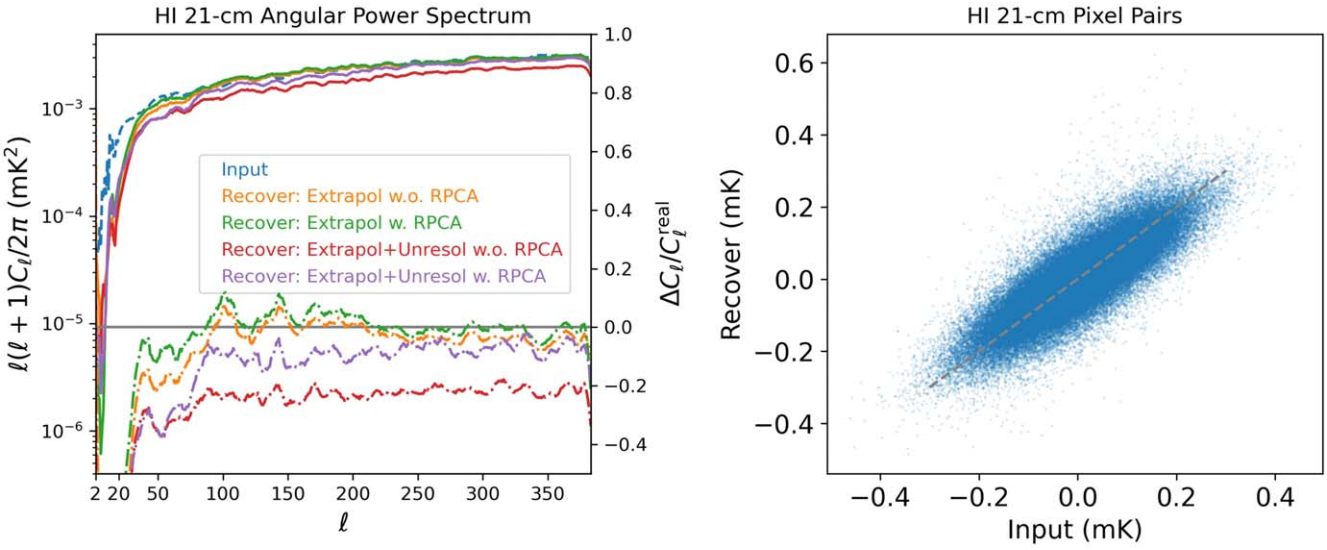

Figure 2. Demonstrations of the RPCA and foreground effects on the GNILC foreground removal method with 20 channels. The left panel shows the 21 cm angular power spectrum at 1252.5 MHz, where the blue dashed line is the input power spectrum, and the orange (green) solid line is the power spectrum recovered from "Extrapolation" power-law foregrounds without (with) RPCA. The red (purple) solid line is recovered from "Extrapolation" and "Unresolved" foregrounds combined without (with) RPCA. The dotted–dashed lines in the lower region with the same colors are the corresponding fractional differences of power spectra ($\Delta C_\ell/C_\ell^{\rm real}$) with respect to the input. The right panel shows a pixel-by-pixel comparison (~100,000 pixels) between the input map and the recovered map with the "Extrapolation" foreground model but without the RPCA method (its power spectrum is shown in the orange line on the left panel).

3.3. Summary of the eGNILC

This section summarizes the eGNILC method we developed in this paper. We will customize the GNILC to the 21 cm IM and present the general procedure of the 21 cm foreground removal. Unlike applying the internal linear combination to the CMB data set, we take additional steps to determine the mixing matrix because the 21 cm signal varies with frequency.

1. The 3D 21 cm tomography data to be cleaned is projected onto a 2D sphere, and the data array at each pixel is arranged as a vector indexed by frequency. A universal spherical mask is applied to cut off extreme foreground contamination. The data vector is therefore denoted by $\boldsymbol{x}(p)$, where p is the spatial index.
2. We are free to decompose the vector with the "Fourier" transform, e.g., the data is represented by the coefficients of DCT modes in Figure 1. Then the real map at each "frequency" is transformed to a needlet space where localization in both the spatial domain and the spherical harmonics domain is reserved. As a consequence, the indices i and p of elementary data $x_i(p)$ do not necessarily represent the frequency and pixel in real space.
3. The ILC algorithm is implemented in the general needlet domain.

   It is crucial that we figure out the imaginary mixing matrix of 21 cm signals. For 21 cm experiments, we do not know the dof of the 21 cm signal nor the foregrounds, but expect that the foregrounds are smooth along frequency. With the help of simulated 21 cm signals, this goal is achievable, as is done in Section 3.1.1, where simulated 21 cm signals are preprocessed in the same way as $x_i(p)$. However, finding the parameter m (dof of foregrounds) is an essential condition to estimate the mixing matrix. The m is determined by minimizing the AIC. We derive the GNILC bias and take the size of the averaging domain into account. Thus, the AIC is modified as in Equation (48). With no prior knowledge, we can also estimate the mixing matrix from data $\boldsymbol{x}(p)$ with the RPCA to get a blind method, as elaborated on in Section 3.2.
4. After extracting signals with ILC on each needlet base, we reverse the decomposition and sum over all modes to recover the 21 cm signal.

4. Demonstration

The basic assumption of the GNILC method is that the foregrounds are highly correlated over frequencies so that the majority of the dof of $\boldsymbol{R}_{\boldsymbol{x}}$ encodes the H I signal. Figure 2 shows the recovered H I power spectrum and the effects of increasing the complexity of foreground emissions. We also demonstrated the performance of the RPCA-embedded GNILC method. In this section, we ignored all of the instrumental effects (e.g., thermal noise and beams).

We simulate foregrounds and H I signals in the frequency range 960–1260 MHz with 20 channel bins evenly spaced. We switch the foreground between "Extrapolation" and "Extrapolation + Unresolved," the method between "RPCA" and "Simulation" to examine the foreground cleaning effect. Figure 2 shows the recovered 21 cm auto-power spectra centered at 1252.5 MHz in four cases. "Extrapolation" denotes the well-modeled sources in Section 2.2.1, and "Unresolved" denotes the unresolved random sources shown in Section 2.2.2. The GNILC method and the RPCA-embedded GNILC method are utilized in the foreground removal, which are denoted by "w.o. RPCA" and "w. RPCA" separately, according to the manner in which the covariance matrix is computed. Because almost all 21 cm IM experiments observe the partial sky, in the following comparison, we just compare the pseudo–power spectrum (L.H.S. of Equation (9)) directly, without inverting the M-matrix (Equation (10)) in Equation (9) to recover the full-sky power spectrum. In a word, we will suppress the tilde in $\tilde{C}_\ell$ in the following presentation.

The left panel of Figure 2 shows the 21 cm angular power spectrum at 1252.5 MHz, where the blue dashed line is the input power spectrum, and the orange and green solid lines are the

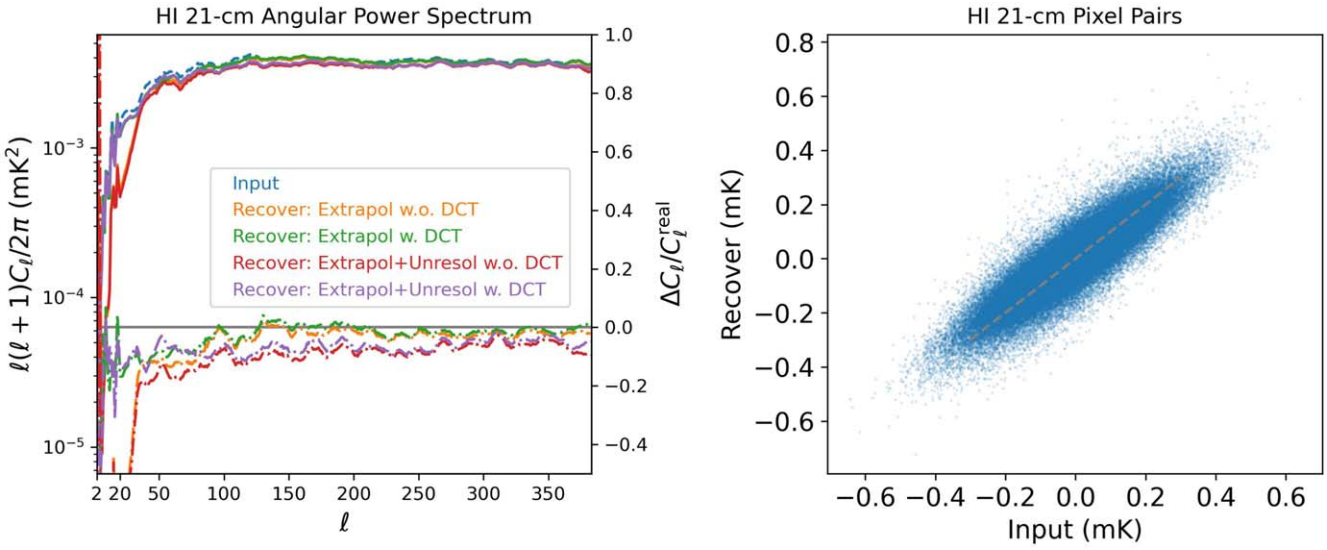

Figure 3. Same as Figure 2 but for the DCT method and 60 channels in the same frequency range.

power spectrum recovered from "Extrapolation" power-law foregrounds without and with RPCA, respectively. The red solid line is the recovered power spectrum from "Extrapolation" and "Unresolved" foregrounds without combining with RPCA, and the purple solid line is that combined with RPCA. In the lower region of the left panel, the dotted–dashed lines in the same colors are the corresponding fractional differences of power spectra with respect to the input power spectrum. For clarity of presentation, all power spectra curves are smoothed properly to remove the small wiggles in ℓ-space. The right panel makes a pixel-by-pixel comparison between the recovered H I map and the input map, in which the recovered map is the case of "Extrapolation" power-law foregrounds without RPCA. The gray dashed line is the $y = x$ diagonal line, which indicates the unbiased estimate. One can see that the recovered and input values are highly aligned alongside of the $y = x$ line, which demonstrates the level of unbiasedness in the real space (the left panel is the Fourier space).

In the lower region of the left panel of Figure 2, one can see that the level of H I power spectrum underestimation is almost zero if the contaminant is just the "Extrapolation" foreground. But this situation deteriorates if the "Unresolved" foregrounds are included, in which case, better frequency resolution would be necessary to extract the signal if the complexity of foregrounds increases. The RPCA-embedded blind GNILC method has a fairly good performance as long as the frequency resolution is good enough for foreground removal. We notice that the underestimation of the "RPCA-embedded" recovered 21 cm auto-power spectrum is smaller. This does not mean that the RPCA-estimated mixing matrix is better than that with the simulated H I signal, because the unresolved foreground power contributes to the sparse matrix in the RPCA that will be used as the input covariance of the 21 cm signals. Additionally, the RPCA takes a long time to split the low-rank and sparse components, especially when the matrix size gets larger. Moreover, the domain $\mathcal{D}$ may not be the best in this case. From the left panel of Figure 2, one can see that in the range of $\ell \leqslant 20$, the GNILC method is not suitable to clean the foreground because of the inadequate sampling on large scales.

The H I signals and foregrounds can be decomposed via the DCT as shown in Figure 1. Because the frequency transform is independent of the spatial needlet decomposition, we can do the

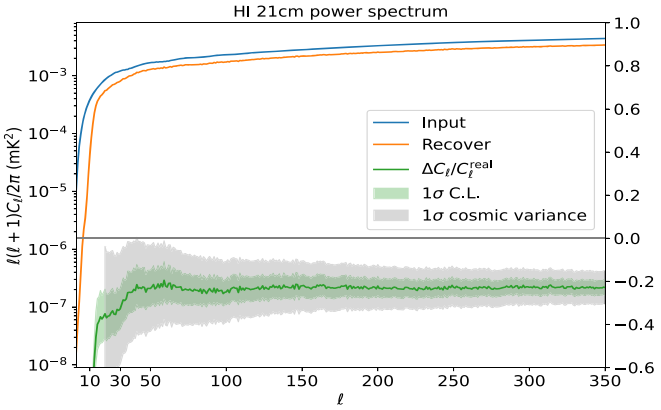

Figure 4. Fluctuations of the recovered power spectrum by the plain GNILC in 1000 realizations of H I 21 cm intensity maps, with the same foreground model and mask as in Figure 1. The simulation runs with 20 frequency channels and contains power-law and unresolved foregrounds. The cosmic variance (gray band) is calculated as $\Delta C_\ell/C_\ell = \sqrt{2/[f_{\rm sky}(2\ell+1)]}$, where $f_{\rm sky} = 0.5$ is the sky coverage factor.

DCT of frequency before the GNILC to generalize the cleaning algorithm further. Figure 3 shows the effect of whether the DCT is applied beforehand. The power spectra are presented in the same way as in Figure 2, but we change the number of frequency channels from 20 to 60. Foregrounds are allowed to vary between "Extrapolation" and "Extrapolation + Unresolved"; the DCT may be switched between "ON" and "OFF." The DCT helps to improve the 21 cm signal recovery at low ℓs where the GNILC itself lacks accuracy. This is clear for both the "Extrapolation" and the "Extrapolation + Unresolved" foregrounds, i.e., the comparison of the blue to orange and the purple to red curves. From Figures 2 to 3, we see that with 60 channels, the recovered power spectrum and pixel-by-pixel comparisons are improved.

We also studied the stability of the GNILC method. Because the blind method is comparable to the plain GNILC method regarding accuracy, we only estimated the fluctuations of the plain GNILC method to save computing time. The sky map is generated with fixed foreground maps, and the 1000 H I map samples are generated from the same power spectrum. Then the H I signal is extracted with the GNILC method from each sample. In Figure 4,

the blue curve is the theoretical angular power spectrum of H I (input), the yellow curve is the mean value of the recovered angular power spectrum of H I, and the green curve and band are the difference and standard deviation. We calculated the cosmic variance $\Delta C_\ell/C_\ell = \sqrt{2/[f_{\rm sky}(2\ell+1)]}$ (gray shadow, $f_{\rm sky} = 0.5$ is the sky coverage factor) and put the center value on the green curve for a vision comparison. The uncertainty caused by the GNILC method is insignificant because the fluctuation of the power spectrum does not override the cosmic variance.

5. Application to Single-dish Experiments

We implement the eGNILC foreground removal framework in the simulated data for two SD experiments, the SKA-MID and BINGO, both of which probe the low-z 21 cm power spectrum. The DCT transform and the modified AIC are used, and the RPCA is switched off to save computation time. We limit the sky coverage in accordance with the experimental designs and take into account the beam effect, with experimental specifications shown in Table 1. The Airy beam is truncated at 6 times the radius of the first zero, and the radius is set by best-fitting the parameter $\theta_{\rm FWHM}$. With this truncation, the last sidelobe is 40 dB below the maximum, but one should notice that the real primary beam could exhibit much more complex patterns, such as those investigated in K. M. B. Asad et al. (2021) for the MeerKAT (SKA-MID precursor) L-band beams. The Airy power pattern is (J. J. Condon & S. M. Ransom 2016)

$$P(u) = \left[\frac{2J_1(\pi u D/\lambda)}{\pi u D/\lambda}\right]^2, \tag{49}$$

where J_1 is the first-order Bessel function of the first kind, $u = \sin\theta$ is the sine of the zenith angle, λ is the wavelength, and D is the aperture width. Therefore, the beam size is scaled by the frequency ratio ν_0/ν where ν_0 is a reference frequency, which may cause significant errors on typical scales in the recovery. We will see abrupt peaks in the recovered angular power spectra of the SKA-MID and BINGO in Figures 6 and 7. The overall system temperature consists of the receiver noise temperature, the temperature of the CMB, the galaxy temperature, etc. As we stated in Section 2.2.3, the noise effect is not the major issue of this paper, and we will only extract the 21 cm signal from the foreground contaminants. The potential RFI contamination is also ignored.

To recover the signal with good accuracy, we should deconvolve the native beams and reconvolve a common beam to all maps (J. Delabrouille et al. 2009). However, in this section, we do not deconvolve the beam from the simulated data before doing the foreground removal. Perfect beam deconvolution is not easy due to nonnegligible instrumental noise, scanning strategy and map making, and beam distortion (C. Burigana & D. Saez 2003; P. A. R. Ade et al. 2016). In the real data, it is possible that the beam effect persists on small scales after beam deconvolution. We will present how the eGNILC works in a bad situation where frequency-dependent Airy beams are convolved to maps and beam deconvolution is absent, although the instrument noise is not added.

5.1. SKA-MID

SKA-MID is a dish array consisting of a set of subarrays, which comprises the South African SKA precursor MeerKAT (launched in 2018) with 64 13.5 m-diameter dishes and the supplement of 133 15 m-diameter SKA1 dishes (D. J. Bacon et al. 2020). The telescopes are configured with a compact core and three logarithmically spaced spiral arms with a maximum baseline of 150 km, which corresponds to an angular resolution $0.''3$ at the frequency of 1.4 GHz (D. J. Bacon et al. 2020). However, SKA-MID in interferometric mode does not provide enough short baselines to map the scales of interest with sufficient signal-to-noise (P. Bull et al. 2015). We use the SD mode instead, which generates the autocorrelation information from each dish.[7] The large number of dishes available with the telescopes will guarantee a large survey speed for probing the H I signal (M. G. Santos et al. 2015).

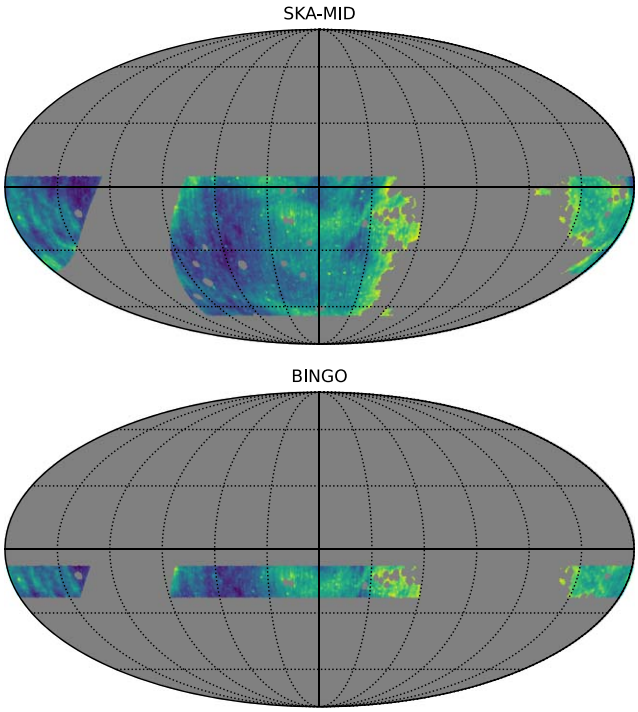

Figure 5. The sky coverage of the simulated SKA-MID and BINGO, projected in the Equatorial coordinate. The gray area is masked, which is derived from Haslam map. The valid area is reduced from the survey area of 20,000 to 10,000 deg^2 and from the survey area of 5000 to 3000 deg^2 (Table 1), respectively, because of the Galactic plane. See also Section 5.1.

J. Wang et al. (2021) calibrated 10.5 hr MeerKAT L-band (856–1712 MHz, 4096 channels) autocorrelation data, in which only the 971–1075 MHz and 1305–1504 MHz bands are free from strong RFI contamination. In this paper, we simulate the data spanning the frequency band 970–1070 MHz with channel width $\delta\nu = 0.2$ MHz, which corresponds to the redshift range of $z = [0.33, 0.46]$. These configurations are not necessarily the same as the real telescopes; because we are using the simulated data, the actual frequency channels are not critical in studying the effects of the mask and beam in the GNILC foreground removal. Figure 5 illustrates the sky coverage of the SKA-MID and BINGO with highly contaminated areas (mostly the galactic disk) removed. The sky coverage of the SKA-MID data is around 10,000 deg^2. More details are listed in Table 1.

We apply the eGNILC algorithm to the simulated SKA-MID data. The foreground consists of the "Extrapolation" with or

[7] The SD IM is an autocorrelation of each dish's voltage, which is different from the interferometric mode where the visibility is obtained by cross-multiplying the voltage.

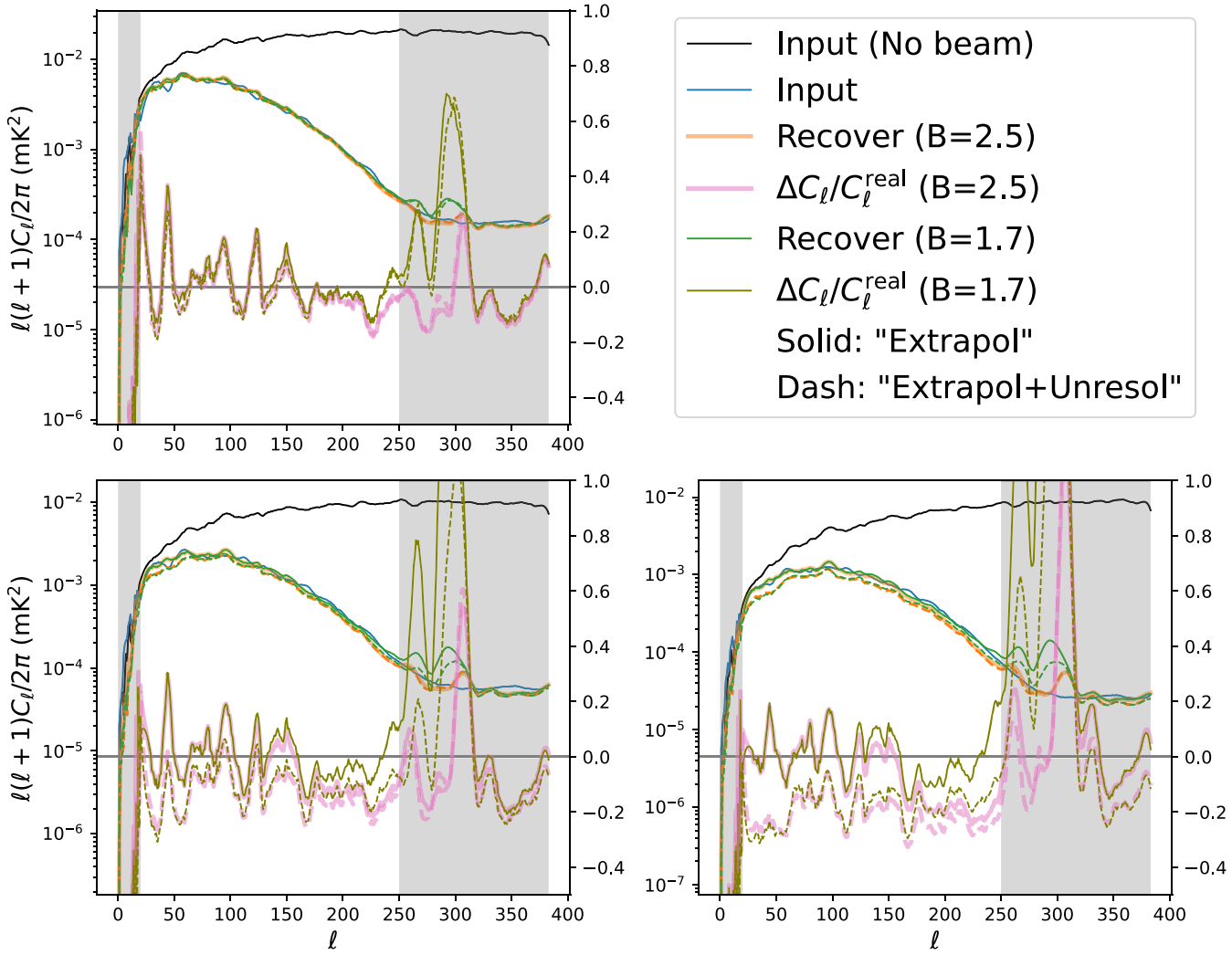

Figure 6. The recovered H I 21 cm angular power spectra of the simulated SKA-MID map with the Airy beam. The black thin line is the power spectrum of the original map (without beam), and the blue one is to convolve the black line with the beam (i.e., input spectrum). The orange thick curves and the green thin curves show the recovered power spectra with the spectral window function parameter $B = 2.5$ and $B = 1.7$, respectively. The pink thick curves and the olive thin curves show their relative differences between the input and recovered power spectrum, the values of which can be read from the right boundary line. The solid and dashed curves represent the "Extrapolation" and "Extrapolation + Unresolved" foregrounds, respectively. The upper-left panel and lower-right panel present the auto-power spectrum at frequencies 1069.9 MHz and 1069.7 MHz, respectively, and the lower-left panel shows the cross-power spectrum at these two frequencies. The left and right shadows ($\ell \lesssim 20$ and $\ell \gtrsim 250$) indicate the cutoff multipole ranges due to poor accuracy.

without the "Unresolved" components. Because the frequency resolution is $\delta\nu = 0.2$ MHz (Table 1), the neighboring 21 cm maps are highly correlated.[8] This may potentially cause practical difficulties while inverting the signal covariance because of the small eigenvalues of $\boldsymbol{R}_s$. If so, then the computational error is out of control, and consequently, the GNILC foreground removal fails in its algorithm. The solution is to separate the maps into n subsets. For the full set indexed with $\{1_1, 1_2, \ldots, 1_n; 2_1, 2_2, \ldots, 2_n; \cdots ; i_1, i_2, \ldots, i_n; \cdots \}$, n should be carefully assigned to guarantee that all maps in the subset $\{1_k, 2_k, \cdots , i_k, \cdots \}$ are uncorrelated. We set n to 10 for the 500 frequency channels, which is adequate to break the correlation among 21 cm maps within the subsets and is large enough to obtain high recovery accuracy. The eGNILC is applied to each subset to extract the H I signal.

Figure 6 presents the recovered H I 21 cm angular power spectra at 1069.9 and 1069.7 MHz, which belong to two different subsets. The original 21 cm maps, together with the foreground, are then convolved with the Airy-disk primary beam to mimic the measurement, and the recovered H I 21 cm maps are compared with the input maps, which are the masked convolved 21 cm maps. For all maps, we multiplied the same reprocessed mask from the Haslam 408 MHz map (M. Remazeilles et al. 2015). In all three panels of Figure 6, the black thin curves are their original power spectra without the beam, and the blue ones are those convolved with the beam (input) so they are damped at large ℓs. The orange thick curves are the recovered power spectra with the spectral window function parameter $B = 2.5$, while the green thin curves are recovered with $B = 1.7$. The pink thick curves and the olive thin curves show their relative differences between the input and recovered power spectrum, whose values can be read out from the right boundary line. In addition, the 21 cm power spectra recovered from the "Extrapolation" and "Extrapolation + Unresolved" foregrounds are plotted by the solid and dashed curves, respectively.

[8] One may refer to J. R. Shaw & A. Lewis (2008) and Y.-C. Li & Y.-Z. Ma (2017) for detailed calculations of the neighboring frequency band cross-correlation.

Table 2
Average Normalized Absolute Difference between the Input Power Spectrum and the Recovered Power Spectrum of the SKA-MID and BINGO, in the Multipole Ranges $20 < \ell < 250$ and $30 < \ell < 300$, Respectively

	SKA-MID				BINGO			
	Extrapol		Extrapol + Unresol		Extrapol		Extrapol + Unresol	
	$B = 2.5$	$B = 1.7$	$B = 2.5$	$B = 1.7$	$B = 2.5$	$B = 1.7$	$B = 2.5$	$B = 1.7$
1069.9 MHz × 1069.9 MHz	7.8%	6.9%	7.9%	7.0%	⋯	⋯	⋯	⋯
1069.9 MHz × 1069.7 MHz	9.5%	7.1%	13.7%	11.8%	⋯	⋯	⋯	⋯
1069.7 MHz × 1069.7 MHz	10.3%	6.8%	21.3%	17.6%	⋯	⋯	⋯	⋯
1257.5 MHz × 1257.5 MHz	⋯	⋯	⋯	⋯	8.9%	8.7%	10.5%	6.2%
1257.5 MHz × 1252.5 MHz	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
1252.5 MHz × 1252.5 MHz	⋯	⋯	⋯	⋯	8.9%	9.0%	10.0%	6.6%

From the lower-left panel of Figure 6, one can see that there is a strong correlation between the 21 cm maps from two adjacent frequency channels. The difference between "Input" and "Recovery" is 10%~20% (pink thick lines and olive thin lines; also see Table 2), which shows the good agreement between the two signals. Therefore, the eGNILC can recover the 21 cm signal very well within $\ell = [20, 250]$. The truncated Airy beam decreases the power spectrum exponentially with a flat tail. For $\ell \gtrsim 250$, the recovered 21 cm power spectra deviate from the original power spectra because the beam varies with frequency, where larger errors are observed in the recovered power spectrum. Such a region is marked with a gray shadow and is not used in further analysis. Similarly, the leftmost region $\ell \lesssim 20$ is also marked with a gray shadow for its large bias due to insufficient independent samples for the eGNILC. We change the parameter B and foreground contaminants to see how the recovered 21 cm is affected. Figure 6 shows the power spectrum bias, and Table 2 gives the average absolute difference in $\ell = [20, 250]$ between the input power spectrum and the recovered power spectrum. By changing the foregrounds from the "Extrapol" to "Extrapol + Unresol," we see the bias increases with the foreground complexity, though the difference is tiny in the auto-power spectra at 1069.9 MHz. Changing B from 2.5 to 1.7 also improves the performance of the eGNILC.

5.2. BINGO

The BINGO telescope is a dedicated instrument designed specifically for observing the BAO signal with 21 cm IM and to provide new insight into the Universe at $z \lesssim 0.5$ with a dedicated instrument (R. A. Battye et al. 2012, 2013; C. Dickinson 2014; R. Battye et al. 2016; C. Wuensche 2019; E. Yohana et al. 2019; E. Abdalla et al. 2022; C. A. Wuensche et al. 2022). The telescope consists of two ~ 40 m compact mirrors with no moving part and will be built in a very low RFI site in South America. The unique design can give an excellent polarization performance and very low sidelobe levels required for IM. BINGO will map a 15° strip on the southern hemisphere as the telescope drift scanning with 60 fixed horns.

BINGO has a better spatial resolution than SKA-MID because of its larger dish. In this paper, we simulate its performance in the frequency range of 960–1260 MHz with channel width $\delta\nu = 5$ MHz, which corresponds to the redshift range of $z = 0.13$–0.48. The sky coverage of the simulated BINGO data is about 3000 deg^2, and Figure 5 shows its valid area. We take the values of these telescope parameters as per guidelines in the previous literature (R. A. Battye et al. 2012; C. Dickinson 2014; R. Battye et al. 2016; C. Wuensche 2019; E. Yohana et al. 2019; D. J. Bacon et al. 2020; J. Wang et al. 2021; E. Abdalla et al. 2022; C. A. Wuensche et al. 2022; see also Table 1), but it just serves as a reference to present the characteristics of the two telescopes. The final telescope can have different experimental specifications than the values quoted here. Figure 7 presents the results of foreground removal and 21 cm signal recovery for BINGO in the same way as Figure 6. Because BINGO's frequency resolution is $\delta\nu = 5$ MHz, the neighboring 21 cm maps are almost uncorrelated, which shows, in the lower-left panel, that the cross-correlation is an order of magnitude lower than the autocorrelation (upper and right panels of Figure 7). The cross-power spectrum is therefore plotted linearly, and the relative difference is not presented because of the smallness of the cross powers. Because the resolution of BINGO is higher, the cutoff range due to discordant beams at high ℓ shrinks to $\ell \gtrsim 300$. After B is changed from 2.5 to 1.7, the average absolute difference and the abrupt peaks decrease, although the peaks are still distinct (see Table 2 and Figure 7). Unexpectedly, the 21 cm signal recovered from "Extrapol + Unresol" has the smallest bias. This indicates that the performance of the eGNILC rests on many key factors, e.g., mask, beam, foreground, spectral window function, and covariance calculation. Given an observation, we may vary the tuning parameters of the eGNILC to obtain the best performance.

6. Conclusions

In this paper, we developed the expanded GNILC method (eGNILC) by refining the algorithm of taking a DCT beforehand, modifying the criterion for determining the dof of foregrounds, and embedding the RPCA in mixing matrix computation. The eGNILC implements the ILC component separation method in a needlet space and is customized to 21 cm IM experiments. The map is localized in both the spatial domain and the spherical harmonics domain. Furthermore, the data vector can be decomposed along the frequency axis with the "Fourier" transform, e.g., the data is represented by the coefficients of the DCT modes in this paper. The robustness of such a customization is demonstrated in Section 4. Unlike the sole mode of CMB, the recovered 21 cm signals at different frequency channels are mixed virtually by an imaginary mixing matrix. It is crucial to determine this mixing matrix to extract the true signal successfully. This can be accomplished with the help of simulated 21 cm signals, and knowing the parameter m

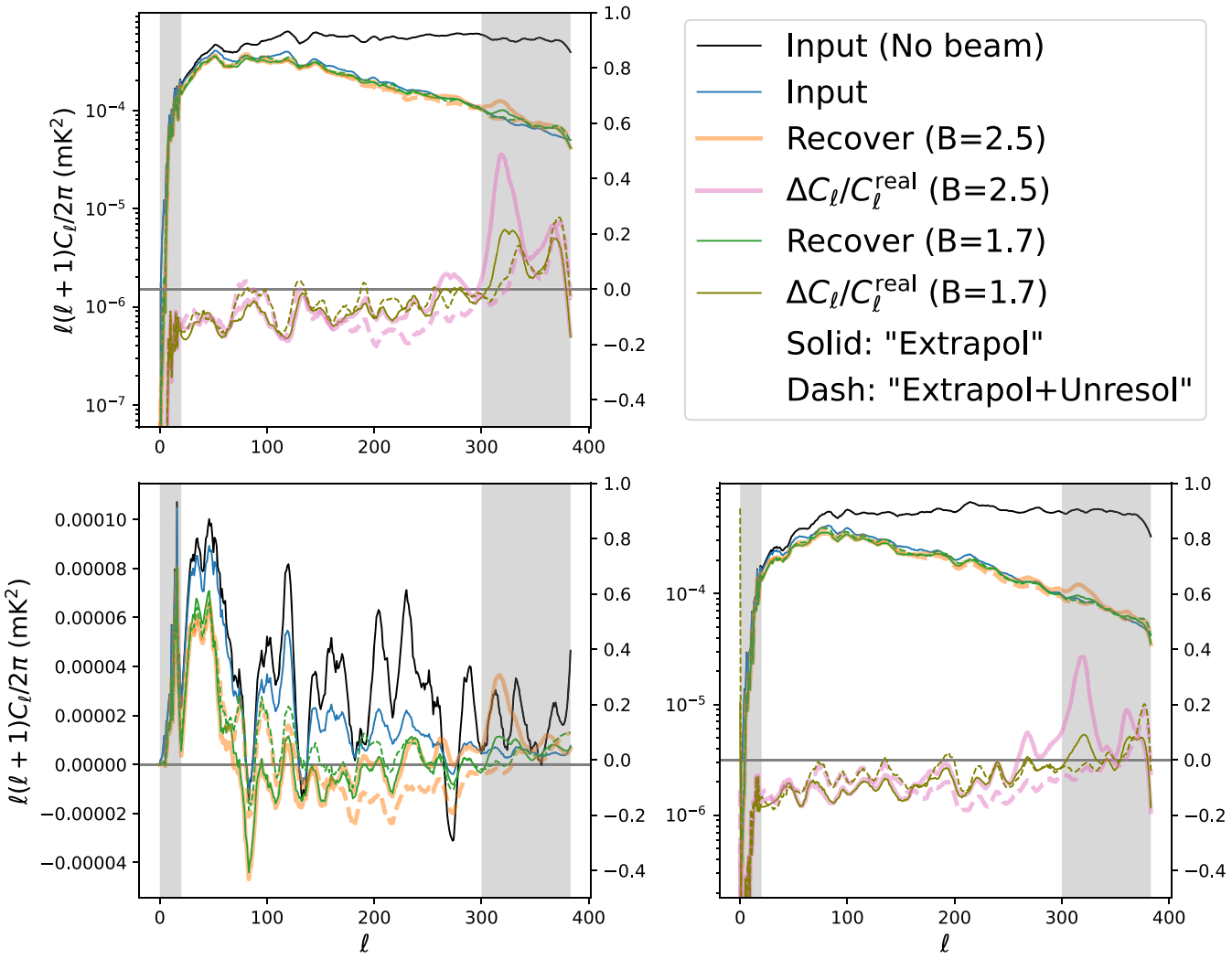

Figure 7. Same as Figure 6 but for BINGO experiment. The upper-left panel and lower-right panel present the auto-power spectrum at frequencies 1257.5 and 1252.5 MHz, respectively. The lower-left panel is the cross-power spectrum (fractional differences omitted).

(dof of foregrounds) is an essential prerequisite. The m is determined by minimizing the AIC. In the meantime, we derive the eGNILC bias, which is unique to the 21 cm data set. We find that it is not particular to a given map, but related to the choice of averaging domain that estimates the covariance matrix of data. Such bias is taken into the consideration of the total likelihood of the AIC to account for the choice of averaging domain, though it is not significant in most cases.

The GNILC method uses the relative power of the H I signals to the observations to estimate the H I subspace (see also L. C. Olivari et al. 2016). Thus, it is not sensitive to details of the prior H I maps. Nevertheless, a prior for the H I power spectra is necessary. We discussed a variant of GNILC with the prior knowledge of H I 21 cm signals substituted with the RPCA to compute the covariance of the H I signal. Thus, we obtained a blind nonparametric method in the GNILC framework. We tested the methods' response to the smooth foreground assumption and found that the RPCA-embedded GNILC performed fairly well in our demonstration. However, we note that the RPCA takes a long time to split the low-rank and sparse components, especially if the matrix size becomes larger. One should remember that the ILC bias is derived from the artificial anticorrelation between the signal and contaminants due to limited samples in the domain $\mathcal{D}$, which has no relation with the error caused by the split of low-rank and sparse components in the RPCA. Thus, we find the caveat that the modified AIC should not be used together with the RPCA in the eGNILC framework.

We demonstrated that the fluctuation of the recovered power spectrum is smaller than the cosmic variance. The eGNILC method was tested in more complex mock data that were simulated according to configurations for the SKA-MID and BINGO SD experiments. The coverage areas are plotted in Figure 5, and the effects of the telescope beam are also tested. In tests for these two experiments, the Airy-disk beam was applied to the data and then deconvolved after foreground removal. There are 500 channels in the range 970–1070 MHz for the SKA-MID ($\delta\nu = 0.2$ MHz) while there are only 60 channels in the range 960–1260 MHz for the BINGO ($\delta\nu = 5$ MHz). The procedures of foreground removal for the SKA-MID and BINGO are different because the 21 cm maps are highly correlated in the former, while they are nearly uncorrelated in the latter. Although the eGNILC takes advantage of the increasing dof of foregrounds to recover the 21 cm signals better, it could fail if the adjacent frequency

channels are highly correlated because of the computational error of inverting the highly correlated signal covariance. We avoided this problem by grouping frequency channels in Section 5.1.

With the Galactic plane masked and no telescope beam, the eGNILC bias is negligible for simple power-law foregrounds. We find that the varying Airy beam leads to significant bias in the power spectrum at typical scales. After cutting off the low-ℓ and high-ℓ multipoles limited by the number of observable samples and the telescope resolution, respectively, the test shows the customized GNILC method is applicable to the SKA-MID and BINGO SD experiments. Ignoring the system's noise, the H I signals of the SKA-MID and BINGO are effectively recovered at the multipoles in $[20, 250]$ and $[20, 300]$, respectively. With $B = 2.5$ or $B = 1.7$, the eGNILC is applied to the experiments to remove the "Extrapol" and "Extrapol + Unresol" foregrounds. The SKA-MID exhibits $\lesssim$20% power loss, and BINGO exhibits ~10% power loss. The results also suggest that beam effects should be dealt with carefully, especially at high-ℓs.

Acknowledgments

We thank the useful comment from Prof. C. Dickinson. Y.Z.M. is supported by the National Research Foundation of South Africa under grant Nos. 150580, 159044, ERC23040389081, and CHN22111069370. This work was part of the research program "New Insights into Astrophysics and Cosmology with Theoretical Models Confronting Observational Data" of the National Institute for Theoretical and Computational Sciences of South Africa.

Appendix A
Needlet Analysis

Like the Fourier transforms on the Euclid plane, the spectral analysis of data on the sphere can be conducted through spherical harmonics transform. In many situations, we need to localize the analysis in both the spatial and spherical harmonics domains (F. Guilloux et al. 2007). This requirement can be achieved through needlet analysis, which is an analogy to wavelet transform on the Euclid plane.

The usual complex spherical harmonics ($Y_{\ell m}$) ($\ell \geqslant 0$ and $-\ell \leqslant m \leqslant \ell$) on the unit sphere (denoted $\mathbb{S}$) form an orthonormal basis of the space (denoted $\mathbb{H}$) of complex-valued square-integrable functions on $\mathbb{S}$ under the Lebesgue measure $d\boldsymbol{\xi} = \sin\theta d\theta d\phi$. The convolution of a bounded axisymmetric function $H(\boldsymbol{\xi}) = h(\cos\theta)$ with an arbitrary spherical function X is well defined through

$$H^*X(\boldsymbol{\xi}) = \int_{\mathbb{S}} h(\boldsymbol{\xi}\cdot\boldsymbol{\xi}')X(\xi')d\boldsymbol{\xi}'. \tag{A1}$$

Meanwhile, the convolution theorem holds:

$$\begin{aligned} H^*X &= \textstyle\sum_{\ell m} h_\ell a_{\ell m} Y_{\ell m} = \sum_{\ell m} h_\ell \langle Y_{\ell m}, X\rangle Y_{\ell m} \\ &= \sum_{j\in\mathcal{J}} \textstyle\sum_\ell h_\ell^{(j)} \sum_{m=-\ell}^{\ell} \langle Y_{\ell m}, X\rangle Y_{\ell m} \\ &= \sum_{j\in\mathcal{J}} \textstyle\sum_\ell h_\ell^{(j)} \Pi_\ell X \\ &= \sum_{j\in\mathcal{J}} \Psi^{(j)} X, \end{aligned} \tag{A2}$$

where $a_{\ell m}$ are the multipole moments of X, h_ℓ are the Legendre series coefficients of h, $\mathcal{J}$ is a countable index set, and Π_ℓ and $\Psi^{(j)}$ are operators defined as follows:

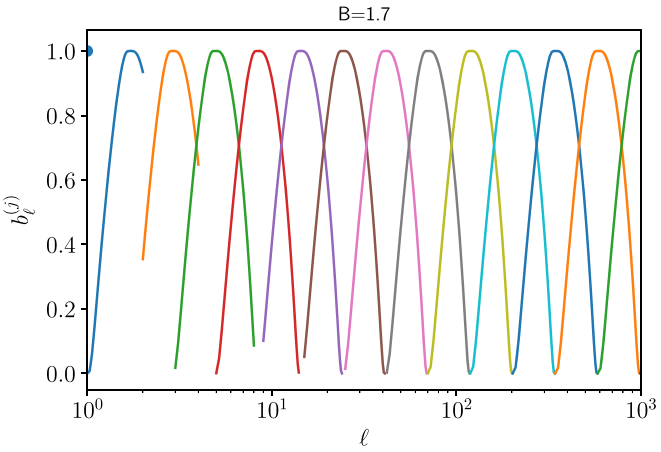

Figure 8. The window function $b_\ell^{(j)}$ as an example of $B = 1.7$.

$$\begin{aligned} \Pi_\ell X(\boldsymbol{\xi}) &= \sum_{m=-\ell}^{\ell} \langle Y_{\ell m}, X\rangle Y_{\ell m}(\boldsymbol{\xi}) \\ &= \frac{2\ell+1}{4\pi}\int_{\mathbb{S}} X(\boldsymbol{\xi}')P_\ell(\boldsymbol{\xi}\cdot\boldsymbol{\xi}')d\boldsymbol{\xi}', \end{aligned} \tag{A3}$$

$$\Psi^{(j)} = \sum_{\ell=0}^{\infty} h_\ell^{(j)}\Pi_\ell, \tag{A4}$$

where P_ℓ functions are the Legendre polynomials. With Equations (A1) and (A2), isotropic wavelet analysis can be implemented either in the spatial domain or in the harmonic domain. Here, we choose the latter by multiplying the harmonic coefficients of the field of interest X with a spectral window h_ℓ. We call this the exact reconstruction condition that ensures $\sum_{j\in\mathcal{J}}\Psi^{(j)} = \boldsymbol{Id}$, or equivalently

$$\sum_{j\in\mathcal{J}} h_\ell^{(j)} = 1. \tag{A5}$$

The filter function used in this paper is defined as follows (D. Marinucci et al. 2008; D. Pietrobon et al. 2010):

$$\rho(t) = \begin{cases} e^{-1/(1-t^2)}, & -1 < t < 1, \\ 0, & \text{otherwise}, \end{cases} \tag{A6}$$

$$s(u) = \left[\int_{-1}^{u}\rho(t)\,dt\right] \Big/ \left[\int_{-1}^{1}\rho(t)dt\right] \tag{A7}$$

$$\varphi(x) = \begin{cases} 1, & x \leqslant 1/B, \\ s[1-2B(x-1/B)/(B-1)], & 1/B < x < 1, \\ 0, & x \geqslant 1 \end{cases} \tag{A8}$$

$$\phi(x) = \varphi\left(\frac{x}{B}\right) - \varphi(x). \tag{A9}$$

Each function in the window family is $b_\ell^{(j)} = \sqrt{\phi(\ell/B^j)}$, $j \in \mathcal{J}$, complemented by $b_\ell^{(-1)} = \delta_{0\ell}$. For the band-limited HEALPix map with $\ell \leqslant \ell_{\max}$, the domain $\mathcal{J}$ is $[0, \log_B(\ell_{\max}) + 1]$. The window $b_\ell^{(j)}$ only overlaps with adjacent windows $b_\ell^{(j-1)}$ and $b_\ell^{(j+1)}$. In this case, the exact reconstruction condition for $h_\ell^{(j)} = [b_\ell^{(j)}]^2$, $\sum_{j\in\mathcal{J}} h_\ell^{(j)} = 1$ is satisfied. Figures 8 and 9 give examples of a window family

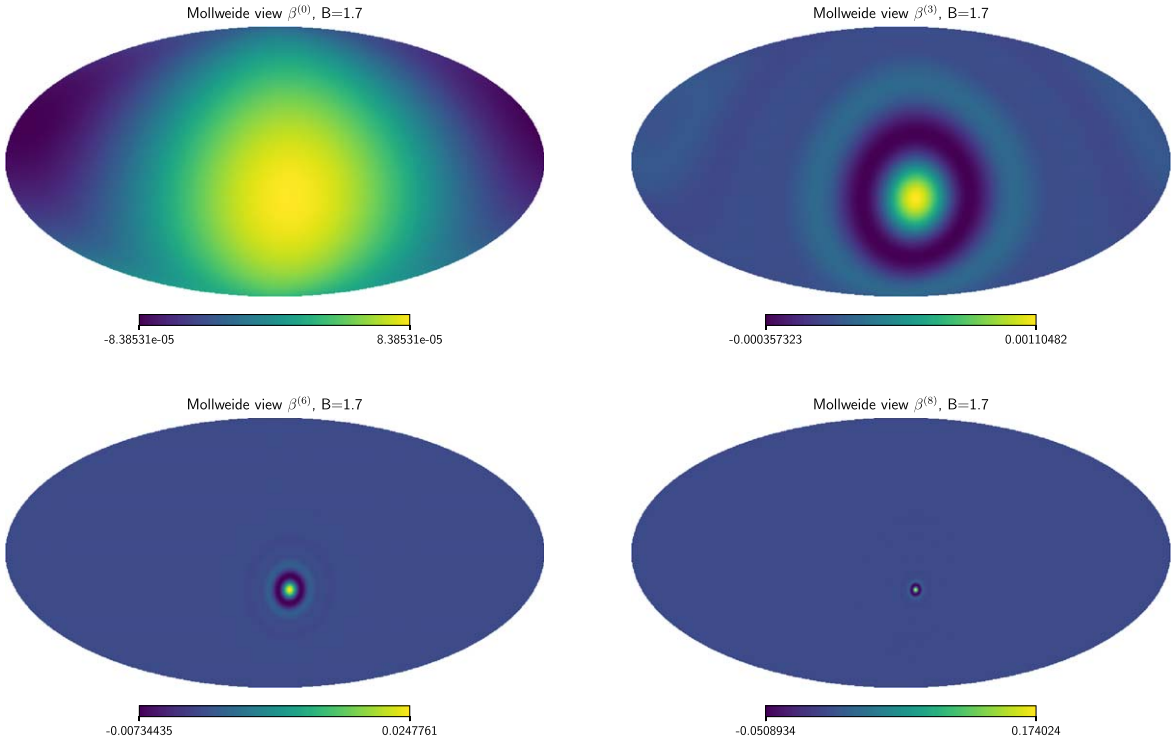

Figure 9. Mollweide-projected view of the needlet decomposition of a point source.

with $B = 1.7$ in the harmonic domain and the localization features in the spatial domain, respectively.

Appendix B
Calculation of the eGNILC Bias and AIC with m-modes of Foreground and N_p Pixel Data

We need to calculate the covariance matrix of the reconstructed 21 cm signal ($\hat{\boldsymbol{s}}$) before obtaining the AIC value of the likelihood. From Equation (39), we need to calculate $\boldsymbol{C}_{\boldsymbol{s\delta}}$, which is equal to $\mathrm{E}(\boldsymbol{\delta s}^T)$. Therefore, we need to substitute Equations (46) into Equations (42)–(45) and utilize them to calculate the expectation value. Because in Equation (42), the $\boldsymbol{f}$ is to first order, its ensemble average with $\boldsymbol{\delta}$ is zero, but Equations (44)–(45) will not give zero expectation because $\boldsymbol{\Delta}_{\boldsymbol{x}}$ also contains first-order $\boldsymbol{f}$ and $\boldsymbol{s}$ (Equation (46)). Therefore, we have

$$\begin{aligned}
\boldsymbol{C}_{\boldsymbol{s\delta}} &= \mathrm{E}(\boldsymbol{\delta s}^T)\\
&= \tfrac{1}{N_p}\textstyle\sum_q \mathrm{E}[\boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}(\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}(\boldsymbol{s}_q\boldsymbol{f}_q^T + \boldsymbol{f}_q\boldsymbol{s}_q^T)\\
&\quad\times \boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})(\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{f}_p\boldsymbol{s}_p^T\\
&\quad- \boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}(\boldsymbol{s}_q\boldsymbol{f}_q^T + \boldsymbol{f}_q\boldsymbol{s}_q^T)\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{f}_p\boldsymbol{s}_p^T]\\
&= \tfrac{1}{N_p}[\boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\mathrm{Tr}\\
&\quad\times(\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{R}_{\boldsymbol{f}})\boldsymbol{R}_{\boldsymbol{s}}\\
&\quad+ \boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{f}_p\boldsymbol{s}_p^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S}\\
&\quad\times(\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{f}_p\boldsymbol{s}_p^T\\
&\quad- \boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\mathrm{Tr}(\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{R}_{\boldsymbol{f}})\boldsymbol{R}_{\boldsymbol{s}} - \boldsymbol{S}\\
&\quad\times(\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{f}_p\boldsymbol{s}_p^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{f}_p\boldsymbol{s}_p^T]\\
&\equiv \boldsymbol{C}_{\boldsymbol{s\delta}}^{(1)} + \boldsymbol{C}_{\boldsymbol{s\delta}}^{(2)} + \boldsymbol{C}_{\boldsymbol{s\delta}}^{(3)},
\end{aligned} \tag{B1}$$

where, in the third equality, we have used the property of the expectation value $\mathrm{E}(\boldsymbol{s}_p^T\boldsymbol{s}_q) = \boldsymbol{R}_{\boldsymbol{s}}\delta_{pq}$. We have also used the matrix property (see also D. Hanson & A. Lewis 2009),

$$\langle \boldsymbol{x}^T A\boldsymbol{x}\rangle = \mathrm{Tr}[A\langle \boldsymbol{x}\boldsymbol{x}^T\rangle] = \mathrm{Tr}[AC], \tag{B2}$$

where C is the covariance matrix of the Gaussian random variable vector $\boldsymbol{x}$, and A is any matrix that is multiplicable to the data vector $\boldsymbol{x}$. The sum over q will take only the $p = q$ term out of the summation.

The identities $\boldsymbol{s} = \boldsymbol{St}$ and $\boldsymbol{R}_{\boldsymbol{s}} = \boldsymbol{SR}_{\boldsymbol{t}}\boldsymbol{S}^T$ can help us simplify Equation (B1) further. We can substitute $\boldsymbol{R}_{\boldsymbol{f}} = \boldsymbol{R}_{\boldsymbol{x}} - \boldsymbol{R}_{\boldsymbol{s}}$ and obtain the first part of the covariance matrix ($\boldsymbol{C}_{\boldsymbol{s\delta}}^{(1)}$) as[9]

$$\begin{aligned}
\boldsymbol{C}_{\boldsymbol{s\delta}}^{(1)} &= \boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\mathrm{Tr}\\
&\quad\times(\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{R}_{\boldsymbol{f}})\boldsymbol{R}_{\boldsymbol{s}}\\
&= \boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\mathrm{Tr}\\
&\quad\times(\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}(\boldsymbol{R}_{\boldsymbol{x}} - \boldsymbol{R}_{\boldsymbol{s}}))\boldsymbol{R}_{\boldsymbol{s}}\\
&= \boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\\
&\quad\times[\mathrm{Tr}(\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T)\\
&\quad- \mathrm{Tr}(\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{R}_{\boldsymbol{s}})]\boldsymbol{R}_{\boldsymbol{s}}\\
&= \boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\\
&\quad\times[\mathrm{Tr}(\boldsymbol{I}_{n_{\mathrm{ch}}-m}) - \mathrm{Tr}(\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\\
&\quad\times\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{SR}_{\boldsymbol{t}}\boldsymbol{S}^T)]\boldsymbol{R}_{\boldsymbol{s}}\\
&= \boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_{\boldsymbol{x}}^{-1}\\
&\quad\times[(n_{\mathrm{ch}} - m) - \mathrm{Tr}(\boldsymbol{R}_{\boldsymbol{x}}^{-1}\boldsymbol{R}_{\boldsymbol{s}})]\boldsymbol{R}_{\boldsymbol{s}},
\end{aligned} \tag{B3}$$

[9] Here we neglect the front $1/N_{\mathrm{p}}$ factor for now and will recover it later.

where, in the fourth equality, we have used the property that $\boldsymbol{S}$ and $\boldsymbol{R}_x$ matrices are of the dimensions $n_{\rm ch} \times (n_{\rm ch} - m)$ and $n_{\rm ch} \times n_{\rm ch}$, respectively (Equations (32) and (35)).

We now calculate the second term of the covariance matrix. By using $\boldsymbol{s} = \boldsymbol{St}$ and $\boldsymbol{R}_s = \boldsymbol{SR}_t\boldsymbol{S}^T$, we have

$$\begin{aligned}\boldsymbol{C}_{s\delta}^{(2)} &= \boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_x^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_x^{-1}\boldsymbol{f}_p\boldsymbol{s}_p^T\boldsymbol{R}_x^{-1}\boldsymbol{S} \\ &\quad \times (\boldsymbol{S}^T\boldsymbol{R}_x^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_x^{-1}\boldsymbol{f}_p\boldsymbol{s}_p^T \\ &= \boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_x^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_x^{-1}\boldsymbol{f}_p\boldsymbol{t}_p^T\boldsymbol{S}^T\boldsymbol{R}_x^{-1}\boldsymbol{S} \\ &\quad \times (\boldsymbol{S}^T\boldsymbol{R}_x^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_x^{-1}\boldsymbol{f}_p\boldsymbol{s}_p^T \\ &= \boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_x^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_x^{-1}\boldsymbol{f}_p\boldsymbol{t}_p^T\boldsymbol{S}^T\boldsymbol{R}_x^{-1}\boldsymbol{f}_p\boldsymbol{s}_p^T \\ &= \boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_x^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_x^{-1}\boldsymbol{f}_p\boldsymbol{s}_p^T\boldsymbol{R}_x^{-1}\boldsymbol{f}_p\boldsymbol{s}_p^T, \end{aligned} \tag{B4}$$

which cancels the second term of $\boldsymbol{C}_{s\delta}^{(3)}$ in Equation (B1). Substituting $\boldsymbol{R}_f = \boldsymbol{R}_x - \boldsymbol{R}_s$ into the first term of $\boldsymbol{C}_{s\delta}^{(3)}$, it becomes

$$\begin{aligned}\boldsymbol{C}_{s\delta}^{(3)} &= -\boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_x^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_x^{-1}\,{\rm Tr} \\ &\quad \times [\boldsymbol{R}_x^{-1}(\boldsymbol{R}_x - \boldsymbol{R}_s)]\boldsymbol{R}_s \\ &= -\boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{R}_x^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{R}_x^{-1} \\ &\quad \times [n_{\rm ch} - {\rm Tr}(\boldsymbol{R}_x^{-1}\boldsymbol{R}_s)]\boldsymbol{R}_s. \end{aligned} \tag{B5}$$

Combining Equations (B3) and (B4), we obtain the total covariance as

$$\boldsymbol{C}_{s\delta} = -\frac{m}{N_p}\boldsymbol{R}_s. \tag{B6}$$

Therefore, from Equation (39) the covariance of $\hat{\boldsymbol{s}}$ becomes

$$\hat{\boldsymbol{R}}_s = \left(1 - 2\frac{m}{N_p}\right)\boldsymbol{R}_s. \tag{B7}$$

This equation shows that the eGNILC bias is a function of the foreground modes number m and the effective domain size N_p. Finally, we obtain the modified AIC by taking the logarithmic likelihood of both the data and the 21 cm prior,

$$\begin{aligned}{\rm AIC}(m, N_{\rm p}) = 2m + n_{\rm ch}\Bigg[&\frac{1}{1 - 2m/N_{\rm p}} \\ &+ \ln\left(1 - 2\frac{m}{N_{\rm p}}\right)\Bigg] + \sum_{i=1}^{n_{\rm ch}-m}[\mu_i - \ln\mu_i - 1], \end{aligned} \tag{B8}$$

which is Equation (48).

ORCID iDs

Wei-Ming Dai https://orcid.org/0009-0003-7287-4380
Yin-Zhe Ma https://orcid.org/0000-0001-8108-0986

References

Abdalla, E., Ferreira, E. G. M., Landim, R. G., et al. 2022, A&A, 664, A14
Abdurashidova, Z., Aguirre, J. E., Alexander, P., et al. 2022a, ApJ, 925, 221
Abdurashidova, Z., Aguirre, J. E., Alexander, P., et al. 2022b, ApJ, 924, 51
Abramowitz, M., & Stegun, I. A. 1972, Handbook of Mathematical Functions (New York: Dover),
Adam, R., Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A10
Ade, P. A. R., Aghanim, N., Ashdown, M., et al. 2016, A&A, 594, A4
Alonso, D., Bull, P., Ferreira, P. G., & Santos, M. G. 2015, MNRAS, 447, 400
Anderson, C. J., Luciw, N. J., Li, Y. C., et al. 2018, MNRAS, 476, 3382
Asad, K. M. B., Girard, J. N., de Villiers, M., et al. 2021, MNRAS, 502, 2970
Bacon, D. J., Battye, R. A., Bull, P., et al. 2020, PASA, 37, e007
Battye, R., Browne, I., Chen, T., et al. 2016, arXiv:1610.06826
Battye, R. A., Browne, I. W. A., Dickinson, C., et al. 2013, MNRAS, 434, 1239
Battye, R. A., Brown, M. L., Browne, I. W. A., et al. 2012, arXiv:1209.1041
Battye, R. A., Davies, R. D., & Weller, J. 2004, MNRAS, 355, 1339
Bigot-Sazy, M. A., Dickinson, C., Battye, R. A., et al. 2015, MNRAS, 454, 3240
Bigot-Sazy, M. A., Ma, Y. Z., Battye, R. A., et al. 2016, in ASP Conf. Ser. 502, Frontiers in Radio Astronomy and FAST Early Sciences Symp. 2015, ed. L. Qain & D. Li (San Francisco, CA: ASP), 41
Bull, P., Ferreira, P. G., Patel, P., & Santos, M. G. 2015, ApJ, 803, 21
Burigana, C., & Saez, D. 2003, A&A, 409, 423
Chang, T.-C., Pen, U.-L., Bandura, K., & Peterson, J. B. 2010, Natur, 466, 463
Condon, J. J., & Ransom, S. M. 2016, Essential Radio Astronomy (Princeton, NJ: Princeton Univ. Press)
Cunnington, S., Li, Y., Santos, M. G., et al. 2022, MNRAS, 518, 6262
DeBoer, D. R., Parsons, A. R., Aguirre, J. E., et al. 2017, PASP, 129, 045001
Delabrouille, J., Cardoso, J. F., Jeune, M. L., et al. 2009, A&A, 493, 835
de Oliveira-Costa, A., Tegmark, M., Gaensler, B. M., et al. 2008, MNRAS, 388, 247
Dickinson, C. 2014, arXiv:1405.7936
Di Matteo, T., Perna, R., Abel, T., & Rees, M. J. 2002, ApJ, 564, 576
Dodelson, J. J., & Schmidt, F. 2020, Modern Cosmology (Amsterdam: Elsevier)
Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium (Princeton, NJ: Princeton Univ. Press)
Efstathiou, G. 2004, MNRAS, 348, 885
Gehlot, B. K., Mertens, F. G., Koopmans, L. V. E., et al. 2019, MNRAS, 488, 4271
Gogo, T. G., Ma, Y.-Z., Kittiwisit, P., et al. 2022, MNRAS, 510, 1680
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Guilloux, F., Fay, G., & Cardoso, J.-F. 2007, arXiv:0706.2598
Hall, A., Bonvin, C., & Challinor, A. 2013, PhRvD, 87, 064026
Hanson, D., & Lewis, A. 2009, PhRvD, 80, 063004
Harper, S., Dickinson, C., Battye, R., et al. 2018, MNRAS, 478, 2416
Haslam, C. G. T., Salter, C. J., Stoffel, H., & Wilson, W. E. 1982, A&AS, 47, 1
Hivon, E., Gorski, K., Netterfield, C., et al. 2002, ApJ, 567, 2
Kogut, A. 2012, ApJ, 753, 110
Li, Y.-C., & Ma, Y.-Z. 2017, PhRvD, 96, 063525
Madau, P., Meiksin, A., & Rees, M. J. 1997, ApJ, 475, 429
Maino, D., Farusi, A., Baccigalupi, C., et al. 2002, MNRAS, 334, 53
Marinucci, D., Pietrobon, D., Balbi, A., et al. 2008, MNRAS, 383, 539
Masui, K. W., Switzer, E. R., Banavar, N., et al. 2013, ApJL, 763, L20
Mertens, F. G., Mevius, M., Koopmans, L. V. E., et al. 2020, MNRAS, 493, 1662
Miville-Deschenes, M. A., Ysard, N., Lavabre, A., et al. 2008, A&A, 490, 1093
Nan, R., Li, D., Jin, C., et al. 2011, IJMPD, 20, 989
Olivari, L. C., Remazeilles, M., & Dickinson, C. 2016, MNRAS, 456, 2749
Osato, K., Shirasaki, M., Miyatake, H., et al. 2020, MNRAS, 492, 4780
Parsons, A. R., Backer, D. C., Foster, G. S., et al. 2010, AJ, 139, 1468
Pen, U.-L., Staveley-Smith, L., Peterson, J., & Chang, T.-C. 2009, MNRAS, 394, 6
Pietrobon, D., Balbi, A., Cabella, P., & Gorski, K. M. 2010, ApJ, 723, 1
Platania, P., Bensadoun, M., Bersanelli, M., et al. 1998, ApJ, 505, 473
Remazeilles, M., Delabrouille, J., & Cardoso, J.-F. 2011, MNRAS, 418, 467
Remazeilles, M., Dickinson, C., Banday, A. J., Bigot-Sazy, M. A., & Ghosh, T. 2015, MNRAS, 451, 4311
Santos, M. G., Cooray, A., & Knox, L. 2005, ApJ, 625, 575
Santos, M. G., Bull, P., Alonso, D., et al. 2015, in Proc. Advancing Astrophysics with the Square Kilometre Array (AASKA14) (Trieste: SISSA), 19
Shaw, J. R., & Lewis, A. 2008, PhRvD, 78, 103512
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
Tegmark, M., & Efstathiou, G. 1996, MNRAS, 281, 1297
The HERA Collaboration, Abdurashidova, Z., Adams, T., et al. 2023, ApJ, 945, 124
Thorne, B., Dunkley, J., Alonso, D., & Naess, S. 2017, MNRAS, 469, 2821
Tramonte, D., & Ma, Y.-Z. 2020, MNRAS, 498, 5916
Tramonte, D., Ma, Y.-Z., Li, Y.-C., & Staveley-Smith, L. 2019, MNRAS, 489, 385
Trott, C. M., Jordan, C. H., Midgley, S., et al. 2020, MNRAS, 493, 4711

Wang, J., Santos, M. G., Bull, P., et al. 2021, MNRAS, 505, 3698
Wolz, L., Abdalla, F. B., Blake, C., et al. 2014, MNRAS, 441, 3271
Wuensche, C. 2019, JPhCS, 1269, 012002
Wuensche, C. A., Villela, T., Abdalla, E., et al. 2022, A&A, 664, A15
Xu, X., Ma, Y.-Z., & Weltman, A. 2018, PhRvD, 97, 083504
Yohana, E., Li, Y.-C., & Ma, Y.-Z. 2019, RAA, 19, 186
Yohana, E., Ma, Y.-Z., Li, D., Chen, X., & Dai, W.-M. 2021, MNRAS, 504, 5231
Zonca, A., Singer, L., Lenz, D., et al. 2019, JOSS, 4, 1298
Zuo, S., Chen, X., Ansari, R., & Lu, Y. 2018, AJ, 157, 4