The Astrophysical Journal Supplement Series, 278:32 (11pp), 2025 June https://doi.org/10.3847/1538-4365/adc8a4

Investigating Cross Correlations between Cosmic Microwave Background Lensing and 21 cm Intensity Mapping

Alessandro Marins[1,2], Chang Feng[1,2], and Filipe B. Abdalla[1,2]
[1] Department of Astronomy, University of Science and Technology of China, Hefei 230026, People's Republic of China; changfeng@ustc.edu.cn
[2] School of Astronomy and Space Science, University of Science and Technology of China, Hefei 230026, People's Republic of China
Received 2024 December 10; revised 2025 March 31; accepted 2025 April 1; published 2025 May 14

Abstract

The neutral hydrogen (H I) signal is a crucial probe for astrophysics and cosmology, but it is quite challenging to measure from raw data because of bright foreground contaminants at radio wavelengths. Cross-correlating the radio observations with large-scale structure (LSS) tracers could detect faint cosmological signals since they are not correlated with the foreground, but exquisite component separation procedures must be performed to reduce the variance induced by the foreground. In this work, we adopt the lensing of the cosmic microwave background (CMB) as the LSS tracer and investigate the cross correlation of CMB lensing and H I observations at the postreionization epoch. We use simulations to study lensing and H I cross correlations in the context of next-generation CMB and intensity-mapping experiments. We investigate the impact of the component separation based on linear combinations of the H I observations at different frequencies and estimate the signal-to-noise ratios for the cross correlation measurements in different scenarios.

Unified Astronomy Thesaurus concepts: H I line emission (690); Gravitational lensing (670); Astronomy data analysis (1858); Dark matter distribution (356); Dark energy (351); Radio astronomy (1338)

1. Introduction

The 21 cm emission of neutral hydrogen (H I) can trace the evolution of our Universe from the dark ages to the postreionization epochs and is a crucial probe for astrophysics and cosmology. However, the radio foreground is a few orders of magnitude brighter than the H I signal at radio wavelengths. By cross-correlating observations at radio frequencies with large-scale structure (LSS) tracers, such as galaxies, evidence for a 21 cm signal has been found (T.-C. Chang et al. 2010; K. W. Masui et al. 2013; C. J. Anderson et al. 2018; L. Wolz et al. 2022). A recent H I detection was also performed through cross correlation with a galaxy survey where the H I observation was taken from the MeerKAT telescope (M. G. Santos et al. 2005) for several hours of observation and over an effective survey area of approximately 200 deg^2, as reported in S. Cunnington et al. (2023). The cross correlation between H I intensity mapping (IM) and LSS tracers has the benefit of easily mitigating different systematics and will be complementary to recent autopower spectrum measurements (S. Paul et al. 2023).

The cosmic microwave background (CMB) lensing is a crucial LSS tracer of the underlying dark matter distribution. It can map the accumulated deflections of CMB photons from the last scattering surface to the observers in different sky directions. Moreover, CMB lensing is a unique probe in that it is less affected by astrophysics and nonlinearities. A cross correlation between H I IM and CMB lensing could both confirm the detection of the 21 cm signal and provide complementary cosmological information. However, previous studies revealed particular challenges regarding cross correlations (N. G. Karaçaylı & N. Padmanabhan 2019; C. Modi et al. 2019; K. Moodley et al. 2023; R. Kothari & R. Maartens 2024). Foreground contaminants are expected to dominate the lowest radial modes in Fourier space, and several methods have been proposed to recover these modes without considering H I signal suppression after blind foreground removal algorithms (N. G. Karaçaylı & N. Padmanabhan 2019; C. Modi et al. 2021; C. Guandalin et al. 2022; K. Moodley et al. 2023).

Cosmic-shear measurements can be obtained by measuring shapes of background galaxies that are distorted by intervening dark matter (M. Bartelmann & P. Schneider 2001). Similar to CMB lensing, the cosmic shear is also correlated with the H I. As demonstrated in A. Sangka & D. Bacon (2024), the cross correlation of the H I IM from the midfrequency of the Square Kilometer Array (SKA-MID) and MeerKAT, and cosmic shear from optical surveys, could also be detected after precise foreground removal of the H I IM observations.

In this work, we use end-to-end simulations to study complex cross correlations between the H I and CMB lensing, incorporating blind foreground removal procedures, such as principal component analysis (S. Zuo et al. 2023), independent component analysis (ICA; D. Maino et al. 2002), general morphological component analysis (J. Bobin et al. 2007), and generalized needlet internal linear combination methods (Planck Collaboration et al. 2016a). In this work, we adopt the FastICA method as a benchmark test.

The organization of this work is as follows: we describe the theoretical models for H I and CMB lensing in Section 2 and the simulations and validations in Section 3. We describe the radio sky model and study the foreground removal procedures in Section 4. We present cross-correlation results in Section 5 and discuss simulations with experimental specifications for H I IM and CMB lensing in Section 6. We conclude in Section 7.

2. Theoretical Models for the Cross-correlation Signals

We define the three-dimensional dark matter distribution as $\delta(\boldsymbol{x}) = \delta(\chi \boldsymbol{n})$, where χ is the comoving distance, $\boldsymbol{n}$ is a direction

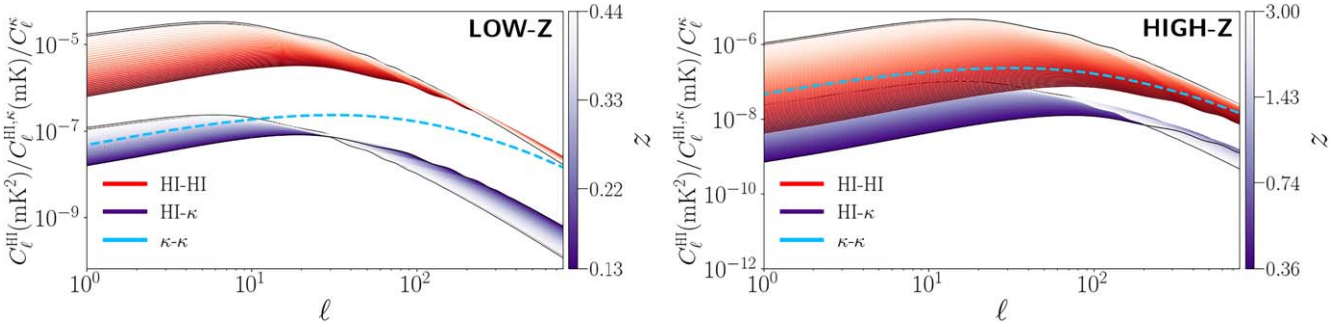

Figure 1. The red gradient colors describe the theoretical H I power spectra in units of mK^2, the blue dashed line is the theoretical CMB lensing power spectrum κ–κ, and the purple gradient colors correspond to the theoretical H I–κ cross correlations in units of mK. The gradient evolution changes from darker colors for the lower redshifts to whiter colors for the higher redshifts. In this work, we consider two representative scenarios for the H I IM: LOW-Z corresponds to H I observations from 980 to 1260 MHz ($0.13 < z < 0.45$), and HIGH-Z from 350–1050 MHz ($0.39 < z < 3.05$).

in the sky, and $\boldsymbol{x}$ denotes the real-space coordinates. The CMB lensing convergence field κ is a projected field, which is a sum of the weighted density field as

$$\kappa(\boldsymbol{n}) = \int d\chi W_\kappa(\chi)\delta(\chi\boldsymbol{n}), \tag{1}$$

where the weighting function is

$$W_\kappa(z) = \frac{3\Omega_m H_0^2}{2cH(z)}\frac{\chi(z)}{a(z)}\frac{\chi(z_*) - \chi(z)}{\chi(z_*)}. \tag{2}$$

Here, Ω_m is the fraction of matter-energy density, $H(z)$ is the Hubble constant at redshift z, H_0 is the Hubble constant today, c is the speed of light, a is the scale factor, which is $1/(1+z)$, and z_* refers to the redshift at the last scattering surface.

The H I brightness temperature field is a biased tracer of the underlying matter distribution, so it can be expressed as

$$\delta T_{\rm H\,I}(\boldsymbol{n}, \nu_j) = \int dz b_{\rm H\,I}(z)\bar{T}_{\rm H\,I}(z)W_{\Delta_j}(z)\delta(\chi\boldsymbol{n}), \tag{3}$$

with a bias $b_{\rm H\,I}(z)$. $W_{\Delta_j}(z)$ is the window function derived from the radio bandwidth Δ_j, which is assumed to be a top-hat window. The brightness temperature described in Equation (3) is essentially a projected observable with a redshift weighting by

$$W_{{\rm H\,I},j}(z) = b_{\rm H\,I}(z)\bar{T}_{\rm H\,I}(z)W_{\Delta_j}(z), \tag{4}$$

where the mean H I brightness temperature (S. R. Furlanetto et al. 2006) is

$$\bar{T}_{\rm H\,I}(z) = 188.8\frac{(1+z)^2}{E(z)}\Omega_{\rm H\,I}(z)h\ [\rm mK]. \tag{5}$$

Here, $\Omega_{\rm H\,I}$ is the energy density fraction of neutral hydrogen, h is the dimensionless Hubble constant and $E(z)$ is $H(z)/H_0$.

The general expression of the angular power spectrum for two generic LSS tracers A and B can be described as

$$C_\ell^{A,B}(\nu_i, \nu_j) = \int \frac{dzH(z)}{c}\frac{W_{A,i}(z)W_{B,j}(z)}{\chi^2(z)}P(k, z), \tag{6}$$

which has been simplified by the Limber approximation $k = (\ell + 1/2)/\chi$ where k refers to spatial and temporal coordinates and ℓ is the multipole. In this work, $A = \kappa$ and $B =$ H I. We show a representative plot for the auto- and cross-power spectra in Figure 1. These analyses assume full-sky coverage and a range of frequencies compatible with BINGO (E. Abdalla et al. 2022), hereafter LOW-Z, and SKA1-MID (SKA Cosmology Science Working Group et al. 2020), hereafter HIGH-Z. We note that these two names refer to two different redshift ranges with an overlapping region around $z \sim 0.35$ in Figure 2. We assume that $\Omega_{\rm H\,I}(z) = 4.86 \times 10^{-4}$ (P. Bull et al. 2015) and $b_{\rm H\,I}(z) = 1$. The cosmological parameter set is (H_0, $\Omega_b h^2$, $\Omega_c h^2$, n_s) = (67.5, 0.022, 0.122, 0.965). The matter power spectrum is calculated via the public code CAMB.[3] As found in L. Wolz et al. (2019), the nonlinear effects of the cross-power spectrum only start to dominate at very small scales up to $k \sim 10\,h\,{\rm Mpc}^{-1}$. Also, as we will show in the following text, the signal-to-noise ratios (SNRs) are mainly dominated by the linear scales below $\ell < 200$ for the specific configurations considered in this work; thereby, the nonlinear scales can be safely ignored. Because of these reasons, we only consider a linear matter power spectrum for this study. In addition, we neglected the shot-noise contributions of the power spectra as they are found to be small at all redshifts from the state-of-the-art hydrodynamic simulations (F. Villaescusa-Navarro et al. 2018; L. Wolz et al. 2019). In Figure 2, the normalized redshift distributions of CMB lensing (Equation (2)) and the H I (Equation (4)) are shown as a solid line and step functions, respectively. We note that the galaxy bias $b_{\rm H\,I}(z, k)$ could be both redshift and scale dependent and the H I fraction $\Omega_{\rm H\,I}(z)$ could be redshift dependent as well. In particular, there are also nonnegligible redshift-space distortions. However, we do not include these complexities to simplify the validation of

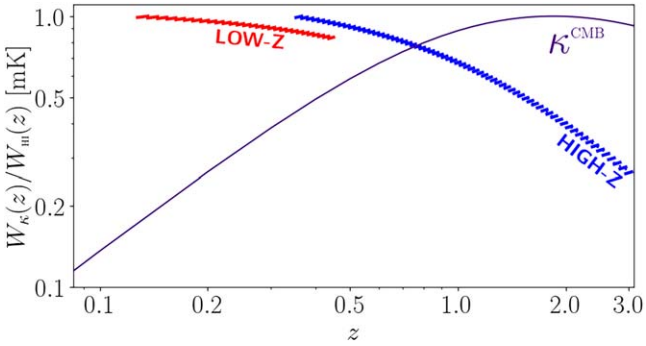

Figure 2. A window function of CMB lensing κ described in Equation (2) is shown as a purple line, and the H I window functions described in Equation (4) are shown as red and blue steps for the two scenarios considered in this work.

[3] https://camb.readthedocs.io/en/latest/

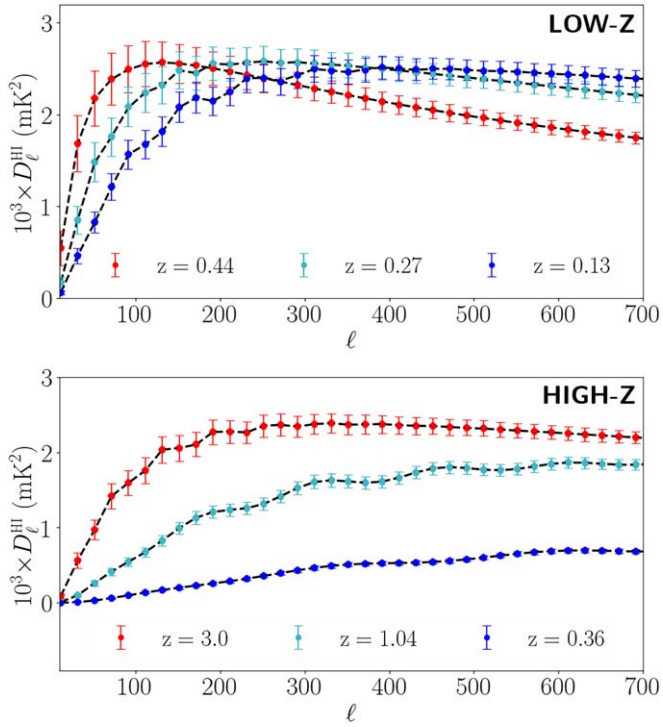

Figure 3. Validations of the correlated simulations as described in Equations (7) and (8). The simulated band powers are consistent with the theoretical calculations, and the band-power errors are estimated from 100 realizations. Here, we define $D_\ell = \ell(\ell+1)C_\ell/2\pi$.

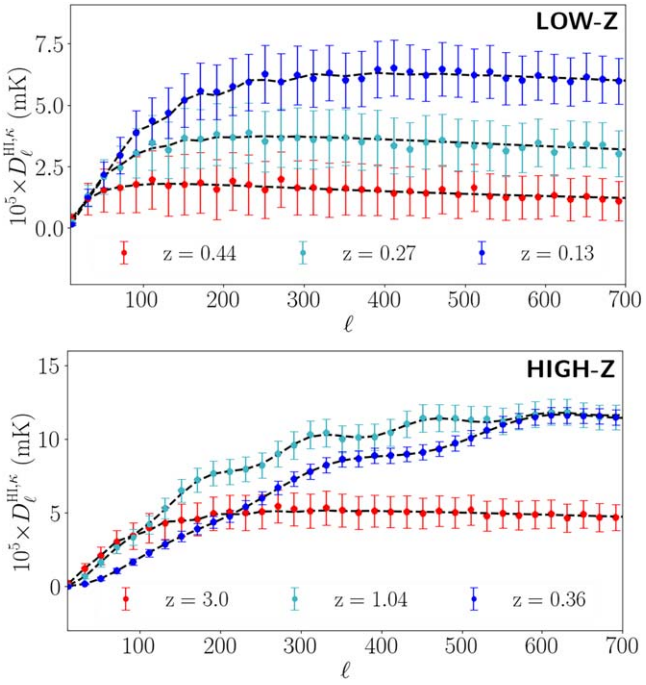

Figure 4. Validations of the cross correlations between CMB lensing and the H I. The correlated simulations are defined in Equations (7) and (8), and the error bars are estimated from 100 realizations. The error bars are multiplied by a factor of 0.2 for visualization purposes.

the cross-correlation algorithms. We defer the detailed investigation of how the theoretical models can affect the reconstruction of the cross correlation to future work.

3. Simulations of the Cross Correlations

To simulate correlated H I and κ fields, we first derive all the theoretical power spectra and generate Gaussian fields with random seeds $\zeta^{j(1)}$ and $\zeta^{(2)}$ (M. Kamionkowski et al. 1997). Here, the index j refers to a specific frequency. Specifically, the correlated fields are generated from

$$a_{\ell m}^{\rm H\,I}(\nu_j) = \sqrt{C_{\ell,\rm fid}^{\rm H\,I}(\nu_j)}\ \zeta_{\ell m}^{j(1)}, \tag{7}$$

$$a_{\ell m}^{\kappa} = \sum_{j=1}^{n_M} \frac{C_\ell^{\rm H\,I,\kappa}(\nu_j)}{\sqrt{C_\ell^{\rm H\,I}}(\nu_j)}\,\zeta_{\ell m}^{j(1)} + \sqrt{C_\ell^{\kappa} - \sum_{j=1}^{n_M} \frac{(C_\ell^{\rm H\,I,\kappa}(\nu_j))^2}{C_{\ell,\rm fid}^{\rm H\,I}(\nu_j)}}\ \zeta_{\ell m}^{(2)} \tag{8}$$

for n_M H I channels correlated with κ. The random variables satisfy the relations

$$\begin{aligned} \langle \zeta_{\ell m}^{j(1)} (\zeta_{\ell' m'}^{k(1)})^* \rangle &= \delta_{jk}\delta_{\ell\ell'}\delta_{mm'}, \\ \langle \zeta_{\ell m}^{(2)} (\zeta_{\ell' m'}^{(2)})^* \rangle &= \delta_{\ell\ell'}\delta_{mm'}, \\ \langle \zeta_{\ell m}^{j(1)} (\zeta_{\ell' m'}^{(2)})^* \rangle &= 0. \end{aligned} \tag{9}$$

The lensing field from this procedure can be split into

$$a_{\ell m}^{\kappa} = \sum_{j=1}^{n_M} a_{\ell m}^{\rm corr}(\nu_j) + a_{\ell m}^{\rm uncorr}. \tag{10}$$

We show the H I autopower spectra and H I-lensing cross-power spectra that are averaged from 100 realizations in Figures 3 and 4, respectively. These simulated power spectra agree with the input theory curves shown by the dashed black lines.

4. Foreground Removal

In this work, we consider two main sources of radio foregrounds, i.e., synchrotron and free–free emissions, which are simulated by `PySM3`[4] (B. Thorne et al. 2017). For the free–free emission, we use a map template generated from `COMMANDER`[5] (Planck Collaboration et al. 2016b), which can extract a degree-scale map of free–free emission at 30 GHz from Planck 2015 data. The synchrotron simulation is an extrapolation from a 408 MHz synchrotron template (M. Remazeilles et al. 2015). In addition to the Galactic foreground contamination, the extragalactic point sources contaminating the H I observations might also be correlated with the LSS. However, as found in C. Feng et al. (2017), the bright radio sources can be easily identified using the existing source catalogs and removed from the maps without losing too much sky area, thus, its contamination can be mostly suppressed. Also, the extragalactic free–free contamination has a very smooth spectral energy distribution and its brightness is a few orders of magnitude lower than the dominant Galactic synchrotron emission (M. G. Santos et al. 2005), therefore the blind foreground removal procedure used in this work can fully eliminate its impact on the 21 cm maps. Because of these reasons, we do not additionally consider the extragalactic foreground contamination.

Meanwhile, variability of the foreground, such as transients, could also affect the component separation. However, sophisticated techniques, such as matched filters, can be applied to the time streams and the time-domain features induced by the variable sources can be effectively removed. Alternatively, the

[4] https://pysm3.readthedocs.io
[5] https://github.com/Cosmoglobe/Commander

time-stream chunk with variable sources can be just flagged and removed as a conservative approach. In addition, maps from IM experiments are generated from time streams collected at a finite amount of integration time on small sky pixels that are much larger than optical surveys. The effects of the variable sources will be further averaged out.

The radio sky model at the jth channel is thus

$$x(\nu_j, \boldsymbol{n}) = \sum_{k=1}^{N_{\rm fg}} a_{jk}(\boldsymbol{n}) s_k(\boldsymbol{n}) + x_{\rm H\,I}(\nu_j, \boldsymbol{n}), \tag{11}$$

where $a_k(\boldsymbol{n})$ describes the spectral energy distribution of the kth foreground component, $N_{\rm fg}$ is the total number of foreground components, and $s_k(\boldsymbol{n})$ is its spatial response in a given sky direction $\boldsymbol{n}$. This model can be described with matrices

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} + \boldsymbol{X}_{\rm H\,I}, \tag{12}$$

where $\boldsymbol{X} = [\boldsymbol{x}(\nu_0), \boldsymbol{x}(\nu_1), \cdots, \boldsymbol{x}(\nu_{n_M-1})]^{\rm T}$ is constructed from n_M frequency channels. By minimizing the noise variance in the reconstructed map, an optimal linear combination of measurements at different frequencies can be achieved by

$$\hat{\boldsymbol{X}}_{\rm FG} = \boldsymbol{W}_{\rm FG}\boldsymbol{X}, \tag{13}$$

with the optimal weights $\boldsymbol{W}_{\rm FG} = \hat{\boldsymbol{A}}\boldsymbol{W}$ where $\boldsymbol{W} = (\hat{\boldsymbol{A}}^{\rm T}\hat{\boldsymbol{A}})^{-1}\hat{\boldsymbol{A}}^{\rm T}$. The radio sky model contains two major components $\boldsymbol{X} = \boldsymbol{X}_{\rm H\,I} + \boldsymbol{X}_{\rm FG}$, as described by Equation (12). Therefore, the H I component can be obtained by subtracting the reconstructed foreground from the raw data via

$$\hat{\boldsymbol{X}}_{\rm H\,I} = \boldsymbol{X} - \hat{\boldsymbol{X}}_{\rm FG} = (\boldsymbol{1} - \boldsymbol{W}_{\rm FG})\boldsymbol{X}. \tag{14}$$

With the two-component model, we can express the two reconstructed maps with the optimal weights as

$$\hat{\boldsymbol{X}}_{\rm FG} = \boldsymbol{W}_{\rm FG}\boldsymbol{X}_{\rm FG} + \widetilde{\boldsymbol{L}}_{\rm H\,I}, \tag{15}$$

and

$$\hat{\boldsymbol{X}}_{\rm H\,I} = \widetilde{\Delta\boldsymbol{L}}_{\rm FG} + \boldsymbol{X}_{\rm H\,I} - \widetilde{\boldsymbol{L}}_{\rm H\,I}, \tag{16}$$

where we define the two leakage terms $\widetilde{\boldsymbol{L}}_{\rm H\,I} = \boldsymbol{W}_{\rm FG}\boldsymbol{X}_{\rm H\,I}$ and $\widetilde{\Delta\boldsymbol{L}}_{\rm FG} = (\boldsymbol{1} - \boldsymbol{W}_{\rm FG})\boldsymbol{X}_{\rm FG}$.

Equation (16) indicates that the reconstructed H I signal is not only contaminated by foreground residuals but also a weighted sum of H I signals at all frequencies. The leakage $\widetilde{\boldsymbol{L}}_{\rm H\,I}$ is present due to the foreground removal procedures and is a generic contribution regardless of different blind methods. FastICA is a computationally efficient blind foreground removal algorithm, so it is adopted for this analysis. FastICA is a fast ICA fixed-point algorithm that uses negentropy to measure non-Gaussianity and forms an optimal linear combination $\hat{\boldsymbol{S}} = \boldsymbol{W}\boldsymbol{X}$, where all n_s components are mutually independent (A. Hyvärinen et al. 1999). In this work, we use the nonquadratic function $g(\cdot) = \log\cosh(\cdot)$, with 20 iterations and set 1% tolerance with FastICA provided by the scikit-learn package.[6]

From Equation (16), the estimated cross-power spectrum at each channel j is

$$\begin{aligned} C_\ell^{\hat{\rm H\,I},\kappa}(\nu_j) &= C_\ell^{\rm H\,I,\kappa}(\nu_j) + (C_\ell^{\widetilde{\rm FG},\kappa}(\nu_j) - C_\ell^{\widetilde{\rm H\,I},\kappa}(\nu_j)) \\ &\approx C_\ell^{\rm H\,I,\kappa}(\nu_j) - C_\ell^{\widetilde{\rm H\,I},\kappa}(\nu_j). \end{aligned}$$

Here, the cross-power spectra $C_\ell^{\widetilde{\rm FG},\kappa}$ and $C_\ell^{\widetilde{\rm H\,I},\kappa}$ correspond to cross correlations of κ and the foreground leakage and that of κ and H I leakage, respectively. The term $C_\ell^{\widetilde{\rm FG},\kappa}(\nu_j)$ vanishes since the κ field and foreground are uncorrelated. However, the weighted sum of the H I signals due to foreground removal gives rise to an additional κ–H I correlation with a minus sign. This term for the channel ν in harmonic space is

$$a_{\ell m}^{\widetilde{\rm H\,I}}(\nu) = \sum_{\nu'=1}^{n_{\rm ch}} \omega_{\nu\nu'} a_{\ell m}^{\rm H\,I}(\nu'), \tag{17}$$

where $\omega_{\nu\nu'}$ represents the components of $\boldsymbol{W}_{\rm FG}$, i.e., $\boldsymbol{W}_{\rm FG} = \{\omega_{\nu\nu'}\}$. This term implies that the leakage is due to the mixing of H I observations at different frequencies with each frequency weighted by the νth row of $\boldsymbol{W}_{\rm FG}$.

As described in Equation (8), κ can be correlated with the H I fields at different frequencies, thus the cross correlation of κ and the H I leakage is not vanishing, i.e.,

$$C_\ell^{\hat{\rm H\,I},\kappa}(\nu) \approx C_\ell^{\rm H\,I,\kappa}(\nu) - \sum_{\nu'=1}^{n_{\rm M}} \omega_{\nu\nu'} \langle a_{\ell m}^{\rm H\,I}(\nu'), a_{\ell m}^{\kappa} \rangle, \tag{18}$$

which is just

$$C_\ell^{\hat{\rm H\,I},\kappa}(\nu) = C_\ell^{\rm H\,I,\kappa}(\nu) - \sum_{\nu'} \omega_{\nu\nu'} C_\ell^{\rm H\,I,\kappa}(\nu'), \tag{19}$$

$$= \sum_{\nu'} (\delta_{\nu\nu'} - \omega_{\nu\nu'}) C_\ell^{\rm H\,I,\kappa}(\nu'). \tag{20}$$

This equation can be expressed in a matrix formalism as

$$\boldsymbol{C}_\ell^{\rm H\,I\text{-}\kappa} = (\mathbf{I} - \boldsymbol{W}_{\rm FG})\boldsymbol{C}_\ell^{\rm H\,I\text{-}\kappa,0}, \tag{21}$$

where $\boldsymbol{C}_\ell^{\rm H\,I\text{-}\kappa,0}$ is the input cross-power spectrum. The whole matrix $\boldsymbol{W}_{\rm FG}$ is sparse and may not be invertible.

As presented in S. Cunnington et al. (2019), an approximation of a constant $\boldsymbol{W}_{\rm FG}$ mixing matrix for a foreground removal procedure was adopted to avoid spurious effects from blind algorithms, which cannot distinguish the radial modes of the foreground from that of the H I. The authors showed that removing the mean H I fluctuations along the line of sight cannot effectively mitigate the edge effect caused by blind foreground removals but can conserve the spectrum shape and partially recover its power spectrum.

Removing the mean H I fluctuations along the line of sight is a good approximation of the procedure described in Equation (16), especially when we are interested in cross-correlation analysis, as shown in Figure 5. This approach assumes that H I leakage is an average among all frequency channels. Therefore, the reconstructed H I map in Equation (16) becomes

$$\hat{\boldsymbol{X}}_{\rm H\,I} = \boldsymbol{X}_{\rm H\,I} - \widetilde{\boldsymbol{L}}_{\rm H\,I} \approx \boldsymbol{X}_{\rm H\,I} - \bar{\boldsymbol{X}}_{\rm H\,I}. \tag{22}$$

The second term refers to the average over all frequencies. In particular, the H I brightness temperature in S. Cunnington et al. (2019) is

$$\begin{aligned} \delta\hat{T}_{\rm H\,I}(\boldsymbol{n}, \nu) &= \delta T_{\rm H\,I}(\boldsymbol{n}, \nu) - \delta\bar{T}_{\rm H\,I}(\boldsymbol{n}) \\ &= \delta T_{\rm H\,I}(\boldsymbol{n}, \nu) - \frac{1}{n_{\rm ch}} \sum_{\nu'} \delta T_{\rm H\,I}(\boldsymbol{n}, \nu'), \end{aligned} \tag{23}$$

which is similar to our procedure in Equation (16), except that they chose a uniform weighting $\omega_{\nu\nu'} = 1/n_{\rm ch}$ for all ν',

[6] https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.FastICA.html

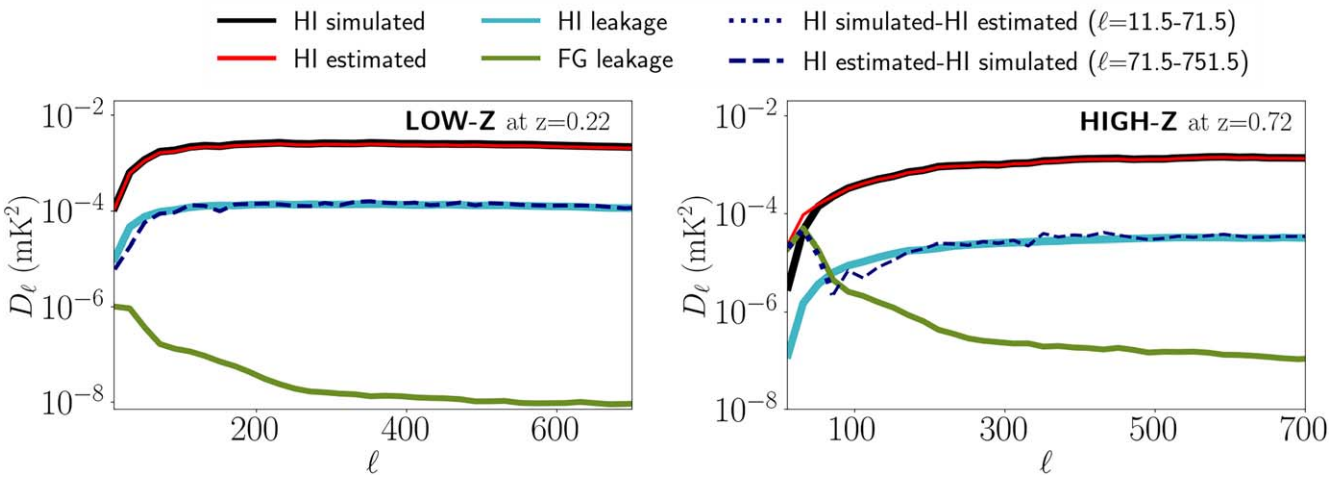

Figure 5. Different components of the H I autopower spectra for LOW-Z (left) at $z = 0.22$ and HIGH-Z (right) at $z = 0.72$.

whereas we derive this foreground removal efficiency matrix precisely from simulations.

We calculate the mixing matrix from a noiseless radio sky model with foreground components that consist of free–free and synchrotron emission. The results are shown in Figure 6, where the $\boldsymbol{W}_{\rm FG}$ matrices are computed for $n_{\rm s} = 2$–5 independent components for the LOW-Z and HIGH-Z scenarios. We find that the first and last channels in the mixing matrix have spuriously higher values than the remaining channels because of limited information at the boundaries, which is exactly the edge effect.

5. Impact of Foreground Removal on the Cross Correlations

Equation (20) indicates that each channel contributes to the cross-power spectrum. The effects of the leakage term are shown in Figure 7 for LOW-Z and HIGH-Z. The dashed curves from blue to yellow are the contributions from different numbers of mixed channels, i.e., $\sum_{\nu'=1}^{N_{\rm eff}} \omega_{\nu\nu'} C_\ell^{\rm H\,I,\kappa}(\nu')$ where $N_{\rm eff}$ is gradually increased from $N_{\rm eff} = 2$ to $N_{\rm eff} = n_M$. When $N_{\rm eff} = n_M$, the leakage term is almost the same as the solid blue line which is the input theory curve. This is because the blind approaches only minimize the variances of the foreground components, resulting in suboptimal reconstructions of the H I signal. There might be new schemes to simultaneously optimize both the foreground and H I reconstructions, achieving minimal cross-correlation leakages. However, this discussion is beyond the scope of this work.

In light of the relationship between leakage and the number of frequency channels, one straightforward way to reduce leakage is to reduce the effective number of frequencies, for example by correlating only the lensing map with a subset of H I frequency channels. In Figures 8 and 9, we show the diagonal elements and a particular row of the mixing matrix (Figure 6) in the left and right plots, respectively. The mixing matrices are calculated by varying the independent ICA modes from $n_s = 2$ to $n_s = 5$. In comparison, the right plot indicates that the weighting of each frequency after component separation is different from a uniform weighting, which would cause biases in the cross-power-spectrum estimation.

For LOW-Z, neighboring frequencies around channel 20 can reach the same level of mixing with different numbers of independent components, as shown in Figure 8. Therefore, we choose a subset of the frequency channels between 18 and 24 labeled by the gray region. For HIGH-Z, the edge effect is clearly observed from the first few channels corresponding to higher redshifts in Figure 9. Similar to the discussions for the LOW-Z, the region between channels 46 and 51 can reach roughly the same level of mixing, so we choose the median frequency as the 48th channel. The exact mixing obviously deviates from a uniform mixing that is inversely proportional to the number of frequencies, i.e., $\omega_{\nu\nu'} \sim n_M^{-1} \approx 0.01$ when $n_M = 70$.

Therefore, if we assume that the lensing field is only correlated with a subset of the H I channels, the resulting cross-power spectra will become nonvanishing since the leakage terms for such a scenario are greatly suppressed, as shown in Figure 10. This scenario can be easily achieved if the CMB lensing field is replaced by a cosmic-shear field or a narrow tomographic bin of the LSS tracer.

However, if the leakage term $\sum_{\nu'=1}^{N_{\rm eff}} \omega_{\nu\nu'} C_\ell^{\rm H\,I,\kappa}(\nu')$ can be precisely estimated, the cross-power spectrum between the CMB lensing and H I can still be measured after debiasing the leakage contributions even if all H I frequency channels are correlated with the CMB lensing.

We use both numerical simulations and theoretical calculations to estimate the leakage of the cross correlation. Since the mixing matrix only depends on the foregrounds, we can assume that it is not dependent on H I and κ realizations, i.e., $\langle \omega_{\nu\nu'} \rangle \approx \omega_{\nu\nu'}$. We calculated the mixing matrices from both a single realization and 100 realizations and found that the difference is negligible. The leakage term for a specific frequency channel j is estimated by an average over realizations:

$$\langle \hat{L}_\ell^{\rm H\,I(\nu_j),\kappa} \rangle_{\rm sim} = \sum_{\nu'} \omega_{\nu_j\nu'} \langle C_\ell^{\rm H\,I,\kappa}(\nu') \rangle. \tag{24}$$

In the limit of a sufficiently large number of simulations, the expression above converges to the theoretical expectation:

$$\langle \hat{L}_\ell^{\rm H\,I(\nu_j),\kappa} \rangle_{\rm th} = \sum_{\nu'} \omega_{\nu_j\nu'} C_{\ell,\rm fid}^{\rm H\,I,\kappa}(\nu'). \tag{25}$$

Here, fid refers to the fiducial model.

To mimic an LSS tracer from a galaxy survey, we show the leakage estimations for the 26th channel ($z \sim 0.17$) of LOW-Z in Figure 11 and the 66th channel ($z \sim 0.41$) of HIGH-Z in

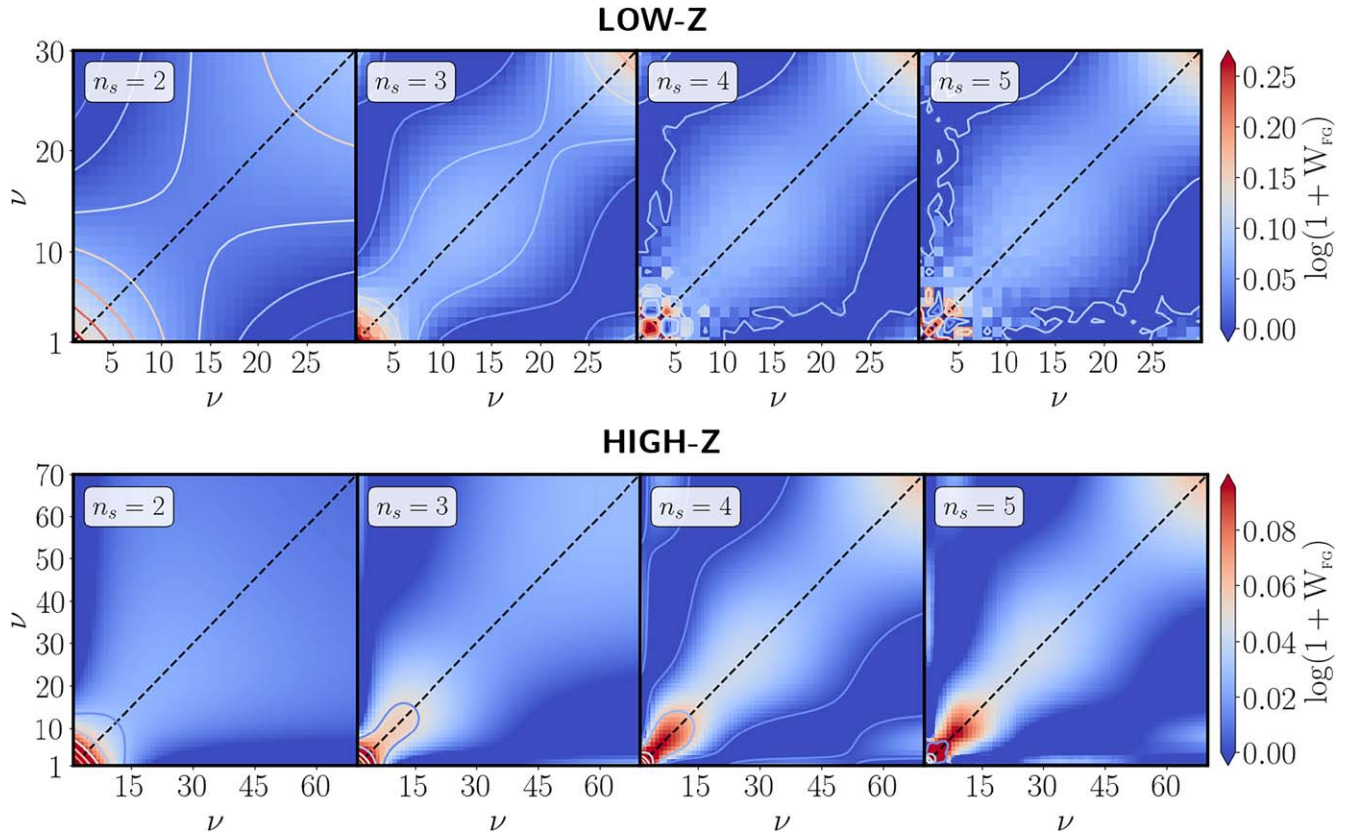

Figure 6. The mixing matrix $\boldsymbol{W}_{\rm FG}$ due to the foreground removal is obtained via the FastICA method. The subpanels are the matrices with different numbers of independent components (n_s). The x- and y-axes correspond to the matrix dimension related to the H I channel indices, and the color bar reflects the relative values of the mixing matrices.

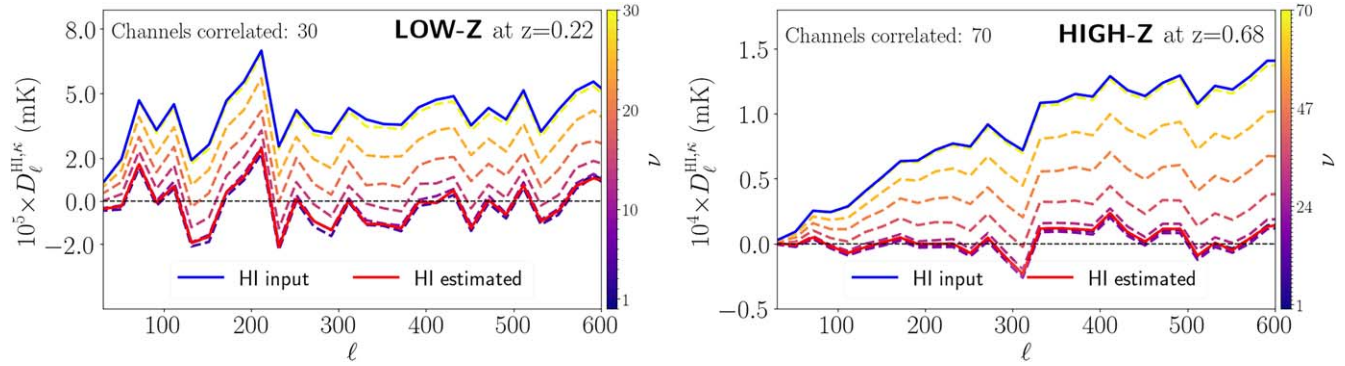

Figure 7. Cross correlations between H I and κ with the H I simulated directly from input theory (blue solid) and the reconstructed H I from FastICA (red solid). We assume that κ is correlated with all H I channels. The left plot shows the cross-power spectrum of a specific frequency channel, i.e., the 20th channel, for the LOW-Z scenario. The right plot shows the cross-power spectrum of the 48th channel for the HIGH-Z scenario. The colored dashed lines are calculated from Equation (19) with variations in the effective number of frequency channels. The results are obtained with $n_s = 4$.

Figure 12. The solid blue curves which are the κ–H I cross-power spectra from the input simulations are consistent with the theoretical predictions. The left plot shows the tests with all H I channels correlated with κ, and the right plot is the case when only a few H I channels are correlated with κ. Obviously, the cross-power spectrum for the latter case becomes nonvanishing.

6. Experimental Configurations

In the previous section, we investigated the cross-correlation leakages for cases with all channels and with a subset of H I observations correlated with κ. To assess the detectability of the cross-power spectra, we assume a LOW-Z experiment similar to the BINGO telescope (F. B. Abdalla et al. 2022) and a HIGH-Z experiment similar to the SKA-MID band 1 (SKA Cosmology Science Working Group et al. 2020). We assume a lensing template from next-generation CMB experiments such as CMB-S4 (K. Abazajian et al. 2019) and a cosmic-shear template from the Rubin Observatory's Legacy Survey of Space and Time (LSST; LSST Science Collaboration et al. 2009).

The H I IM noise power spectrum is assumed to be $N_\ell = \sigma_{\rm pix}^2 \Omega_{\rm pix}$, which is further deconvolved by a Gaussian beam $b_\ell = \exp[-\ell(\ell+1)\theta_{\rm FWHM}^2/(16\log 2)]$ with a resolution

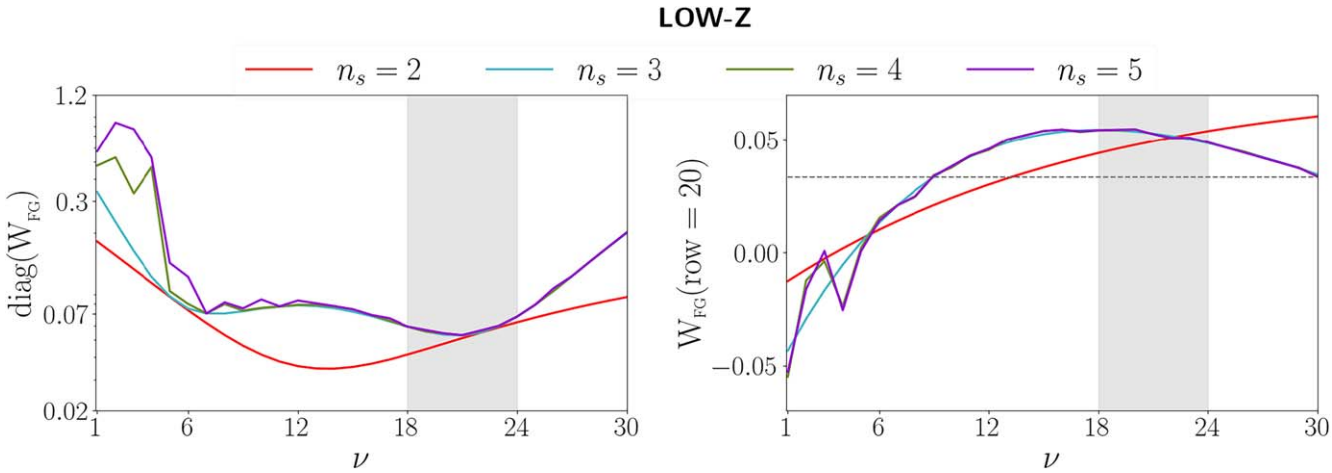

Figure 8. Mixing-matrix elements of $\boldsymbol{W}_{\rm FG}$ for the LOW-Z scenario. The left plot represents the diagonal matrix elements for two to five independent components. The right plot shows the matrix elements from the 20th row of the mixing matrix described in the second term of Equation (19). The gray-shaded regions represent a subset of frequencies tested for the leakage suppression. The dashed black line in the right panel represents the uniform weighting with an amplitude of $1/30$.

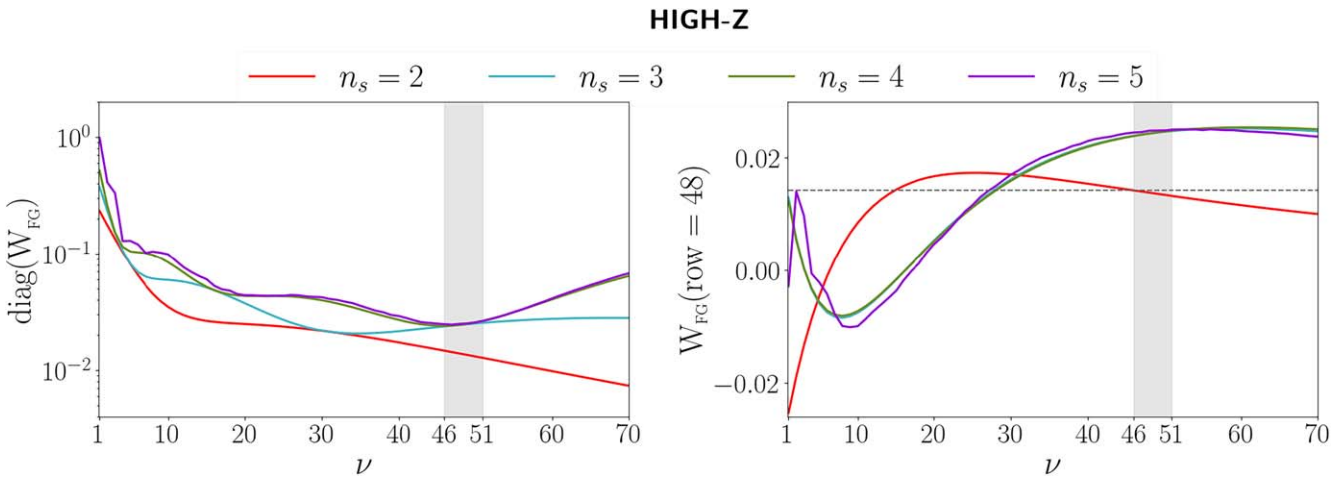

Figure 9. Mixing-matrix elements of $\boldsymbol{W}_{\rm FG}$ for the HIGH-Z scenario. The left plot represents the diagonal matrix elements for two to five independent components. The right plot shows the matrix elements from the 48th row of the mixing matrix described in the second term of Equation (19). The gray-shaded regions represent a subset of frequencies tested for the leakage suppression. The dashed black line in the right panel represents the uniform weighting with an amplitude of $1/70$.

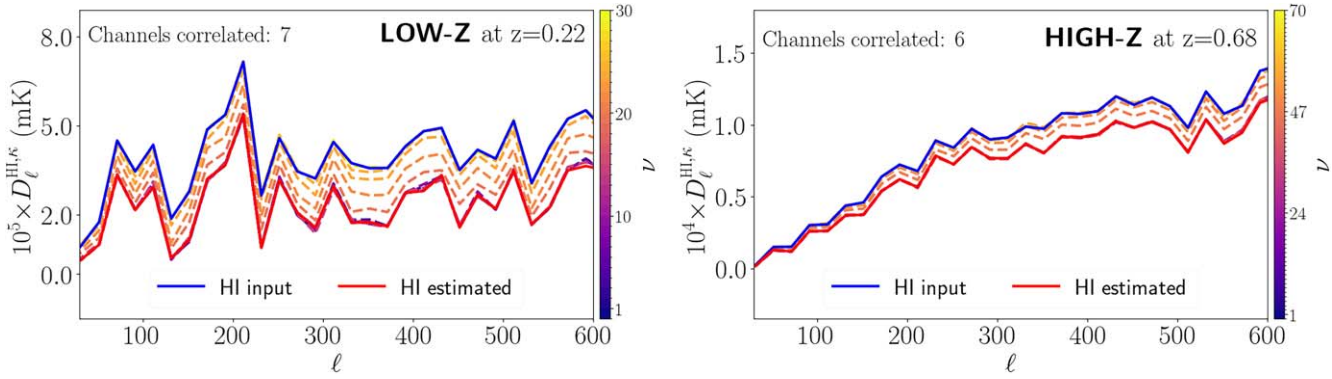

Figure 10. Cross correlations between H I and κ with the H I simulated directly from input theory (blue solid) and the component-separated H I (red solid). We assume that the κ field is only correlated with H I channels at a subset of frequencies. The left plot shows the cross-power spectrum of the 20th channel for the LOW-Z scenario, and the right plot shows the cross-power spectrum of the 48th channel for the HIGH-Z scenario. The colored dashed lines are different leakage terms calculated from Equation (19) with variations in the effective number of frequency channels. The results are obtained with $n_s = 4$.

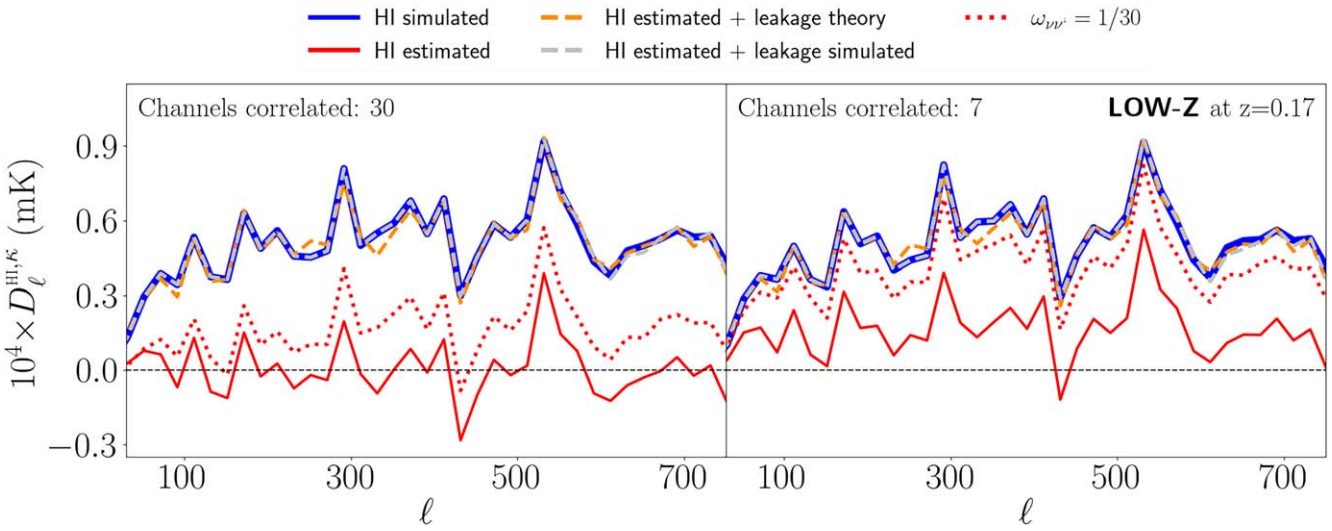

Figure 11. Angular cross-power spectra for the LOW-*Z* configuration with an effective redshift of 0.17 (channel 26) when κ is correlated with both all the H I channels (left) and only seven channels (right), i.e., channels 18–24. In addition to the cross-power spectra with H I before (blue) and after (red) the foreground removal procedure, we also show the debiased cross-power spectra with Equation (24) (gray) and Equation (25) (orange), as well as a uniform weighting scheme (dashed red). The results are obtained with $n_s = 4$.

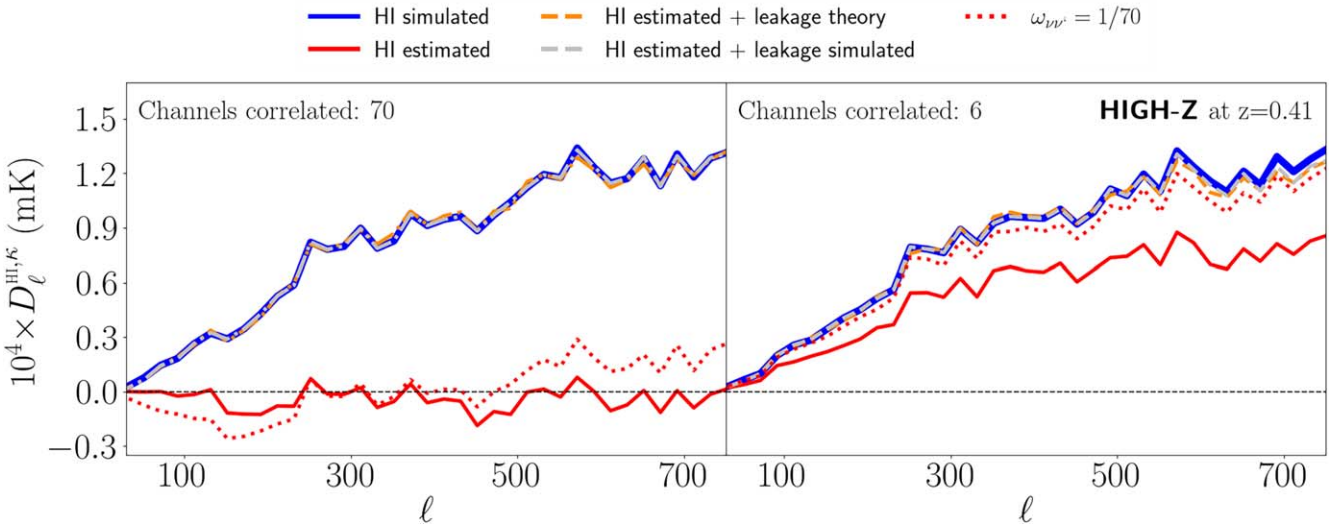

Figure 12. Angular cross-power spectra for the HIGH-*Z* configuration at an effective redshift of 0.41 (channel 66) when κ is correlated with both all the H I channels (left) and only six channels (right), i.e., channels 46–51. In addition to the cross-power spectra with H I before (blue) and after (red) the foreground removal procedure, we also show the debiased cross-power spectra with Equation (24) (gray) and Equation (25) (orange), as well as a uniform weighting scheme (dashed red). The results are obtained with $n_s = 4$.

θ_{FWHM}. The sensitivity of a pixel is determined by

$$\sigma_{\mathrm{pix}} = K \frac{T_{\mathrm{sys}}}{\sqrt{\Delta\nu t_{\mathrm{pix}}}}, \tag{26}$$

where K is noise performance, T_{sys} is a system temperature, and $\Delta\nu$ is a bandwidth. The pixel integration time is

$$t_{\mathrm{pix}} = \epsilon t_{\mathrm{sur}} N_{\mathrm{beams}} \frac{\Omega_{\mathrm{pix}}}{\Omega_{\mathrm{sur}}}. \tag{27}$$

Here, Ω_{sur} is the survey coverage area, ϵ is the coverage efficiency, N_{beams} is the number of beams, and the pixel area is assumed to be $\Omega_{\mathrm{pix}} \approx \pi\theta^2_{\mathrm{FWHM}}/4$. White-noise realizations are drawn from the noise power spectra calculated with the parameters listed in Table 1.

As found in F. B. Abdalla et al. (2022) and S. D. Matshawule et al. (2021), sophisticated optical designs can achieve approximately Gaussian beams with low sidelobes for IM experiments such as BINGO and SKA-MID/MeerKAT. Therefore, in this work, we assume a frequency-independent Gaussian beam profile with a constant resolution. Frequency-dependent beam uncertainties can undermine smooth frequency correlations of bright foreground contaminants, thereby degrading the performance of foreground removal procedures. However, for IM experiments targeting the postreionization Universe, the frequency ranges are usually narrow and the ratios of two different frequencies are close to unity, thus the beam resolution differences among different frequency channels can be approximated as beam calibration errors. Details of real optical designs and constructions of the telescopes are required to accurately model beam shapes, and this

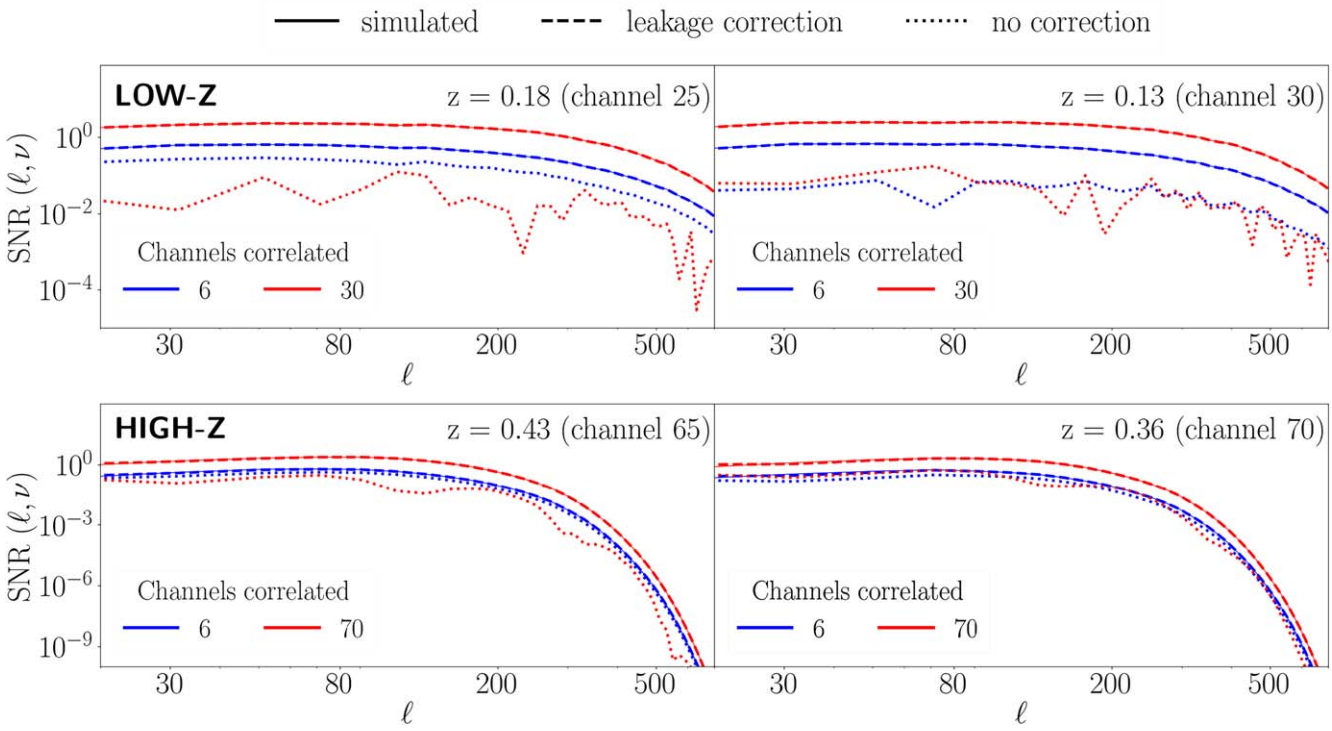

Figure 13. The averaged signal-to-noise ratios (SNRs) per multipole (Equation (30)) from 100 realizations. In this figure, we show the SNRs at two different redshifts for both the LOW-Z and HIGH-Z scenarios. In particular, we compare the SNRs for cross correlations with the CMB lensing in red and with tomographic tracers in blue.

Table 1
Experimental Specifications for the H I IM Surveys Investigated in This Work

	LOW-Z	HIGH-Z
Angular resolution ($\theta_{\rm FWHM}$)	40'	2°
Frequency range	980–1260 MHz	350–1050 MHz
Number of channels (n_ν)	30	70
Bandwidth ($\Delta\nu$)	9.33 MHz	10 MHz
Sky coverage ($\Omega_{\rm sur}$)	41,253 deg^2	41,253 deg^2
Number of beams ($N_{\rm beams}$)	1084	270
System temperature ($T_{\rm sys}$)	70 K	24–62 K
Total coverage time ($t_{\rm sur}$)	1 yr	1 yr
Coverage efficiency (ϵ)	1	1
Noise performance (K)	$\sqrt{2}$	1
Noise per pixel ($\sigma_{\rm pix}$)	60 μK	80–204 μK

Note. Two typical scenarios are considered.

investigation is beyond the aims of this work. Therefore, we defer the detailed investigations of systematic issues, including the $1/f$ noise and beam chromaticity, to the future work.

The CMB lensing can be reconstructed with quadratic estimators (W. Hu et al. 2007). In this work, we consider only the *EB* estimator since its Gaussian bias is the smallest for high-sensitivity polarization experiments (W. Hu & T. Okamoto 2002). Different Gaussian biases can be calculated via `quicklens` code.[7] We assume a CMB-S4-like noise level, i.e., $\sqrt{2}\,\mu$K-arcmin in polarization, and a 1′ Gaussian beam (K. Abazajian et al. 2019) for full-sky coverage.

For a galaxy-lensing-like survey, the shot-noise contribution should be considered due to the finite sample of resolved sources and can be expressed as

$$\sigma_N = \sigma_\epsilon \sqrt{\frac{A_{\rm pix} N_{\rm bins}}{n_{\rm gal}}}, \tag{28}$$

where σ_ϵ is the standard deviation of the observed ellipticities in the survey, $n_{\rm gal}$ is the surface number density of galaxies (per angular resolution), $A_{\rm pix}$ is the angular resolution, and $N_{\rm bins}$ is the number of bins. We use the LSST parameter set (LSST Science Collaboration et al. 2009) but consider a full-sky survey with a set of experimental parameters $(\sigma_\epsilon^2,\, n_{\rm gal}/[\rm gal/arcmin^2],\, A_{\rm pix}/[\rm arcmin^2],\, N_{\rm bins}) = (0.26,\, 27,\, 5.49,\, 6)$.

The covariance matrix for the cross-power spectrum can be expressed as (L. Knox 1995)

$$\Delta^2(\hat{C}_\ell^{A,B}) = \frac{1}{(2\ell+1) f_{\rm sky}\,\Delta\ell}\left\{ \hat{C}_\ell^{A,B}\hat{C}_\ell^{A,B} + \left[\hat{C}_\ell^{A} + \frac{\mathcal{N}_\ell^A}{(b_\ell^A)^2}\right]\left[\hat{C}_\ell^{B} + \frac{\mathcal{N}_\ell^B}{(b_\ell^B)^2}\right]\right\}, \tag{29}$$

where $b_\ell^{\mathcal{X}}$ is the beam profile for an observable X, $\mathcal{N}_\ell^X$ is its noise power spectrum, $f_{\rm sky} = 1$, and $\Delta\ell = 20$. Thus, the SNR per multipole is

$$\mathrm{SNR}_{\rm A,B}(\nu,\,\ell) = \sqrt{\left\langle \frac{\hat{C}_\ell^{A,B}(\nu)}{\Delta\hat{C}_\ell^{A,B}(\nu)} \right\rangle^2}, \tag{30}$$

with $\Delta\hat{C}_\ell^{A,B}$ given by Equation (29), and the brackets indicate an average over the realizations. In this work, $A = \kappa$ and $B =$ H I. We have assumed FastICA with $n_s = 4$ for all the

[7] https://github.com/dhanson/quicklens/

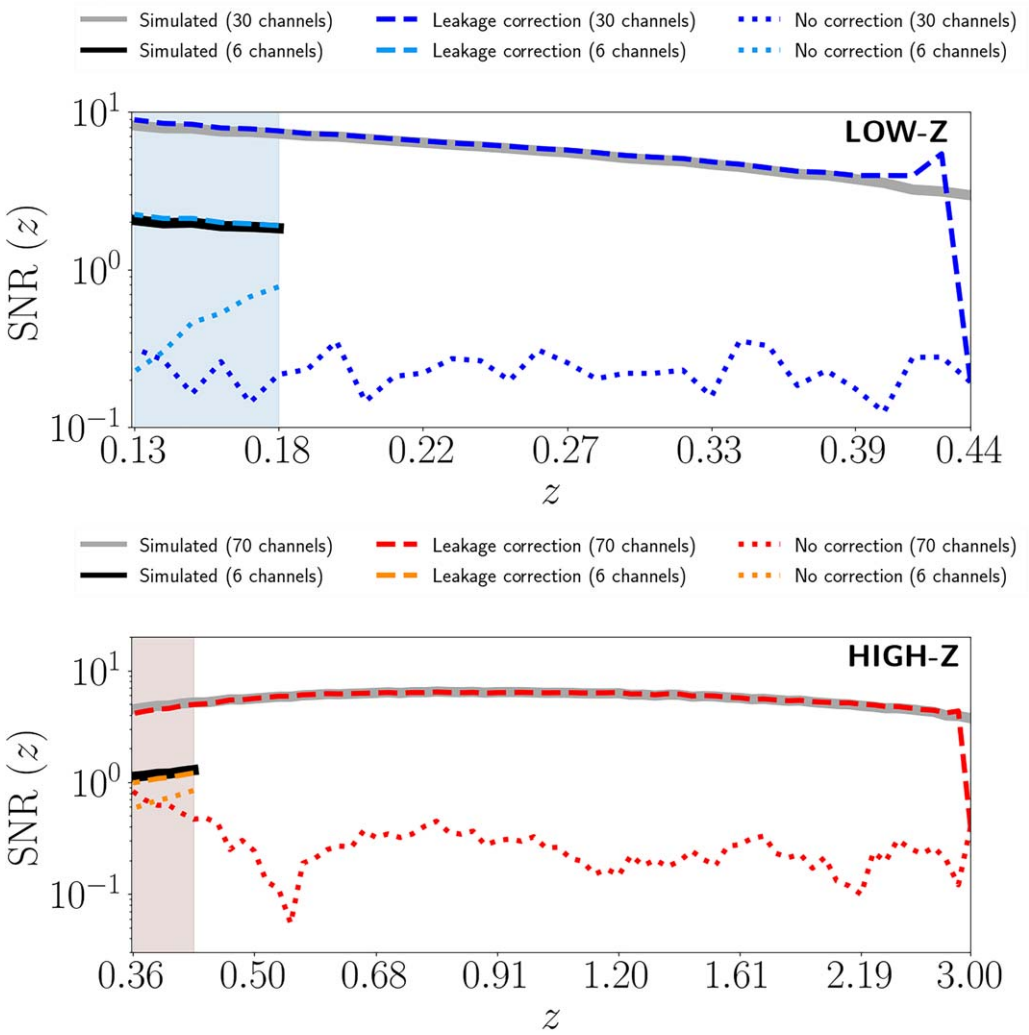

Figure 14. The SNRs per effective redshift or frequency channel. The effective redshift is taken as the mean redshift of a given channel. The shaded regions represent the galaxy-survey-like scenario where only a few H I frequency channels are correlated with the CMB lensing.

cases since they provided equivalent results from $n_s = 3$ to 5 for LOW-Z and the optimal result for HIGH-Z.

The SNRs for different scenarios are shown in Figure 13. The main difference between LOW-Z and HIGH-Z is caused by the resolutions of the H I surveys, and the SNRs of LOW-Z decrease more slowly than those of HIGH-Z. The SNRs can be further improved once the leakage terms are estimated and the corrections are applied.

In addition to the SNRs per multipole, we can also obtain the cumulative SNRs for each frequency band (or a tomographic band), i.e., $\mathrm{SNR}_{\mathrm{A,B}}(\nu) = \sqrt{\sum_\ell \mathrm{SNR}^2_{\mathrm{A,B}}(\nu, \ell)}$ as shown in Figure 14, and even an overall SNR as $\mathrm{SNR}_{\mathrm{A,B}} = \sqrt{\sum_{\nu,\ell} \mathrm{SNR}^2_{\mathrm{A,B}}(\nu, \ell)}$ listed in Table 2. These results indicate that the cross correlation between H I and CMB lensing could be a promising probe only if it can be precisely debiased.

CMB lensing can trace all the dark matter distribution in the Universe while the galaxy surveys can achieve tomographic mapping of the dark matter distribution. Different cross- and autopower spectra formed by a set of three observables $\{\mathrm{H\,I}, \kappa, g\}$ can further improve the detectability of the faint H I signals and would be important probes to fundamental physics in late Universe.

Table 2
The Global SNR for Each Configuration

H I Regime	Corr. Channels[a]	n_s[b]	SNR		
			Not corrected	Corrected	Simulated
LOW-Z	6	3	1.3	5.0	4.7
		4	1.3	5.0	4.7
		5	1.3	5.0	4.7
	30	3	1.3	33.0	31.7
		4	1.3	33.3	31.7
		5	1.3	34.2	31.7
HIGH-Z	6	3	0.7	0.8	3.0
		4	1.8	2.7	3.0
		5	1.0	1.6	3.0
	70	3	5.2	21.2	47.6
		4	2.6	47.5	47.6
		5	1.8	21.2	47.6

Notes. For LOW-Z the SNRs with leakage corrections are slightly higher than the theoretical expectations. This is because we neglected the variance associated with the leakage correction terms. An exact computation of the leakage correction variance is beyond the scope of this work.
[a] Number of H I channels correlated with the convergence field.
[b] Number of FastICA independent components.

7. Conclusions

There have been several detections of H I and galaxy cross correlations, however, the cross correlations between H I and the CMB lensing would not only be complementary but robust cosmological probes since they are less contaminated by systematic issues. Moreover, CMB lensing can map all the LSS structures along the line of sight, directly tracing the dark matter distribution.

The H I IM analyses normally adopt blind foreground removal methods with minimal assumptions. Therefore, the foreground removal procedure inevitably removes the radial H I modes because there is degeneracy in long radial wavelengths between H I and foregrounds. The reduction of these long-wavelength modes has negligible impact on the H I auto-correlations at two different frequencies. However, it can significantly suppress the cross correlations between H I and CMB lensing.

We generated correlated simulations for the H I and κ fields and produced mock observations at different frequencies, taking into account the foreground contamination and instrumental noises. We verified that the cross-power spectra of the input simulations are consistent with the theoretical predictions. The blind foreground removal methods were applied to these mock observations, and reconstructed H I maps were cross-correlated with the simulated CMB lensing maps. To investigate different components of the cross correlations, we performed various tests for two scenarios including a case when all H I channels are correlated with κ and a case when a subset of H I channels are correlated with κ. The tests indicate that the cross-power spectra can be moderately suppressed if only a limited number of H I channels are correlated with the CMB lensing.

We made forecasts for cross correlations between the H I IM experiments including LOW-Z and HIGH-Z scenarios, and the LSS tracers including CMB lensing and cosmic shear. The SNRs per multipole were estimated for the cases without and with suppression corrections.

In addition, we theoretically modeled the leakage components which were validated with simulations. The debiased cross-power spectra are consistent with theoretical expectations. Moreover, we found that better SNRs can be achieved at linear regions such as $\ell \sim 300$ after the leakage corrections.

Future IM experiments will obtain multifrequency H I data sets with low instrumental noises, and next-generation CMB experiments will produce a high-sensitivity lensing map via the quadratic estimator technique. Therefore, it is conceptually straightforward to cross-correlate the H I observations with the LSS tracers such as the CMB lensing to infer the underlying astrophysical and cosmological information. This work explores the detectability of the cross-power spectra and performs theoretical and numerical calculations to analyze details of the complex signals, paving the way for using H I-lensing cross correlations as complementary cosmological probes.

Acknowledgments

We are grateful for the helpful discussions with Xin Zhang, Yichao Li, David Bacon, and Jiajun Zhang. We also thank Dongdong Zhang for helping to estimate the weak-lensing noise. This work is supported by the USTC's starting grant. We acknowledge the use of `HEALPix` (K. M. Górski et al. 2005), MATPLOTLIB (J. D. Hunter 2007), NUMPY (S. van der Walt et al. 2011), SCIKIT-LEARN (F. Pedregosa et al. 2011), CAMB (A. Lewis & A. Challinor 2011) and NAMASTER (D. Alonso et al. 2019).

ORCID iDs

Alessandro Marins https://orcid.org/0000-0002-6519-6038
Chang Feng https://orcid.org/0000-0001-7438-5896
Filipe B. Abdalla https://orcid.org/0000-0003-2063-4345

References

Abazajian, K., Addison, G., Adshead, P., et al. 2019, arXiv:1907.04473
Abdalla, E., Ferreira, E. G. M., Landim, R. G., et al. 2022, A&A, 664, A14
Abdalla, F. B., Marins, A., Motta, P., et al. 2022, A&A, 664, A16
Alonso, D., Sanchez, J., Slosar, A. & LSST Dark Energy Science Collaboration 2019, MNRAS, 484, 4127
Anderson, C. J., Luciw, N. J., Li, Y. C., et al. 2018, MNRAS, 476, 3382
Bartelmann, M., & Schneider, P. 2001, PhR, 340, 291
Bobin, J., Starck, J.-L., Fadili, J., & Moudden, Y. 2007, ITIP, 16, 2662
Bull, P., Ferreira, P. G., Patel, P., & Santos, M. G. 2015, ApJ, 803, 21
Chang, T.-C., Pen, U.-L., Bandura, K., & Peterson, J. B. 2010, Natur, 466, 463
Cunnington, S., Li, Y., Santos, M. G., et al. 2023, MNRAS, 518, 6262
Cunnington, S., Wolz, L., Pourtsidou, A., & Bacon, D. 2019, MNRAS, 488, 5452
Feng, C., Cooray, A., & Keating, B. 2017, ApJ, 846, 21
Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, PhR, 433, 181
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Guandalin, C., Carucci, I. P., Alonso, D., & Moodley, K. 2022, MNRAS, 516, 3029
Hu, W., DeDeo, S., & Vale, C. 2007, NJPh, 9, 441
Hu, W., & Okamoto, T. 2002, ApJ, 574, 566
Hunter, J. D. 2007, CSE, 9, 90
Hyvärinen, A., Karhunen, J., & Oja, E. 1999, Independent Component Analysis (New York: Wiley)
Kamionkowski, M., Kosowsky, A., & Stebbins, A. 1997, PhRvD, 55, 7368
Karaçaylı, N. G., & Padmanabhan, N. 2019, MNRAS, 486, 3864
Knox, L. 1995, PhRvD, 52, 4307
Kothari, R., & Maartens, R. 2024, JCAP, 2024, 089
Lewis, A., & Challinor, A., 2011 CAMB: Code for Anisotropies in the Microwave Background, Astrophysics Source Code Library, ascl:1102.026
LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv:0912.0201
Maino, D., Farusi, A., Baccigalupi, C., et al. 2002, MNRAS, 334, 53
Masui, K. W., Switzer, E. R., Banavar, N., et al. 2013, ApJL, 763, L20
Matshawule, S. D., Spinelli, M., Santos, M. G., & Ngobese, S. 2021, MNRAS, 506, 5075
Modi, C., White, M., Castorina, E., & Slosar, A. 2021, JCAP, 2021, 056
Modi, C., White, M., Slosar, A., & Castorina, E. 2019, JCAP, 2019, 023
Moodley, K., Naidoo, W., Prince, H., & Penin, A. 2023, arXiv:2311.05904
Paul, S., Santos, M. G., Chen, Z., & Wolz, L. 2023, arXiv:2301.11943
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, JMLR, 12, 2825
Planck Collaboration, Adam, R., Ade, P. A. R., et al. 2016b, A&A, 594, A10
Planck Collaboration, Aghanim, N., Ashdown, M., et al. 2016a, A&A, 596, A109
Remazeilles, M., Dickinson, C., Banday, A. J., Bigot-Sazy, M. A., & Ghosh, T. 2015, MNRAS, 451, 4311
Sangka, A., & Bacon, D. 2024, MNRAS, 532, 996
Santos, M. G., Cooray, A., & Knox, L. 2005, ApJ, 625, 575
SKA Cosmology Science Working Group, Bacon, D. J., Battye, R. A., et al. 2020, PASA, 37, e007
Thorne, B., Dunkley, J., Alonso, D., & Næss, S. 2017, MNRAS, 469, 2821
van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22
Villaescusa-Navarro, F., Genel, S., Castorina, E., et al. 2018, ApJ, 866, 135
Wolz, L., Murray, S. G., Blake, C., & Wyithe, J. S. 2019, MNRAS, 484, 1007
Wolz, L., Pourtsidou, A., Masui, K. W., et al. 2022, MNRAS, 510, 3495
Zuo, S., Chen, X., & Mao, Y. 2023, ApJ, 945, 38