

# The feasibility of weak lensing and 21cm intensity mapping cross-correlation measurements

Anut Sangka<sup>1,2,3</sup>★ and David Bacon<sup>1</sup>

<sup>1</sup>*Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK*

<sup>2</sup>*National Astronomical Research Institute of Thailand, Chiangmai 50180, Thailand*

<sup>3</sup>*Department of Physics, Faculty of Science, Ubon Ratchatani University, Ubon Ratchatani 34190, Thailand*

Accepted 2024 June 7. Received 2024 February 29; in original form 2023 February 8

## ABSTRACT

One of the most promising probes to complement current standard cosmological surveys is the HI intensity map, i.e. the distribution of temperature fluctuations in neutral hydrogen. In this paper we present calculations of the two-point function between HI (at redshift  $z < 1$ ) and lensing convergence ( $\kappa$ ). We also construct HI intensity maps from  $N$ -body simulations, and measure two-point functions between HI and lensing convergence. HI intensity mapping requires stringent removal of bright foregrounds, including emission from our Galaxy. The removal of large-scale radial modes during this HI foreground removal will reduce the HI-lensing cross-power spectrum signal, as radial modes are integrated to find the convergence; here we wish to characterize this reduction in signal. We find that after a simple model of foreground removal, the cross-correlation signal is reduced by  $\sim 50$ –70 per cent; we present the angular and redshift dependence of the effect, which is a weak function of these variables. We then calculate S/N of  $\kappa$ HI detection, including cases with cut sky observations, and noise from radio and lensing measurements. We present Fisher forecasts based on the resulting two-point functions; these forecasts show that by measuring  $\kappa \Delta T_{\text{HI}}$  correlation functions in a sufficient number of redshift bins, constraints on cosmology and HI bias will be possible.

**Key words:** gravitational lensing: weak – large-scale structure of Universe – radio lines: general.

## 1 INTRODUCTION

The clustering of matter in the Universe provides an important insight into the origins and evolution of the cosmic structure. Inflation predicts that early structure formation generates a near-Gaussian random field in overdensity; evolution due to gravity causes late-time large-scale structures (LSSs) to exhibit non-Gaussian features. Two-point statistics of the density field at different redshifts capture information about the evolution of structures, and correlation functions between different pairs of cosmological probes can precisely constrain cosmological parameters (Abbott et al. 2018; Upham, Whittaker & Brown 2019; Fang et al. 2022; Pandey et al. 2022; Tröster et al. 2022). Two-dimensional surveys of the cosmic microwave background (CMB) have been effectively carried out through the last few decades (Hinshaw et al. 2013; Planck Collaboration I 2020). The complement to this is deep sky observations of the three-dimensional galaxy and dark matter fields. While conventional optical and infrared surveys have high angular resolution, long integration times are needed for these to obtain precise redshifts via spectroscopy. In contrast, photometric surveys provide faster redshift capture but less radial resolution (Fernandez-Soto et al. 2001).

To complement the low radial resolution of optical photometric surveys, alternative techniques with higher radial resolution are desirable; radio 21cm intensity mapping is a rapidly developing candidate for this purpose. Unlike most optical surveys, this

technique does not measure the brightness of individual objects, but focuses on the larger scale fluctuations in intensity of the 21cm radio signal from neutral hydrogen (HI). The temperature fluctuations can be used as a tracer for the underlying cosmic density field. This intensity mapping is a complementary technique to a photometric survey, with excellent redshift resolution but lower angular resolution (Bull et al. 2015). Hence, combining HI and optical surveys is potentially valuable, as the two techniques compensate for each other's limitations (Cunnington et al. 2019b; Square Kilometre Array Cosmology Science Working Group et al. 2020).

Recently, HI intensity mapping techniques have been actively developed (Mao et al. 2008; Harker et al. 2010; Santos et al. 2010; Wolz et al. 2016; The CHIME Collaboration 2022; Cunningham et al. 2023a). The Canadian Hydrogen Intensity Mapping Experiment (CHIME) (CHIME Collaboration 2023) has provided a detection of HI via cross-correlations with three probes of LSS, namely luminous red galaxies (LRGs), emission-line galaxies (ELG), and quasars (QSO) from the eBOSS clustering catalogues at high significant levels,  $7.1\sigma$  (LRG),  $5.7\sigma$  (ELG), and  $11.1\sigma$  (QSO). Cunningham et al. (2023a) have detected the correlated clustering between MeerKAT measurements of HI and galaxies from the WiggleZ Dark Energy Survey at  $7.7\sigma$  significance. Intensity mapping is therefore on its way to becoming an independent observational probe, providing useful information from low to high redshifts, via future surveys with radio telescopes such as MeerKAT (Pourtsidou 2017; Pourtsidou, Bacon & Crittenden 2017; Spinelli et al. 2022) and the Square Kilometre Array, SKA (Santos et al. 2015).

\* E-mail: [anut.s@ubu.ac.th](mailto:anut.s@ubu.ac.th)

The major challenge for the intensity mapping technique is that the foreground signals are much stronger than the cosmic HI brightness temperature, especially due to the galactic plane synchrotron radiation (Switzer et al. 2013; Spinelli, Bernardi & Santos 2018; Su et al. 2018). Hence several studies of two-point functions between HI and optical ( $\Delta_{\text{HI}}\delta_g$ ) have focused on the impact of foreground removal (Chapman et al. 2012; Cunnington et al. 2019b, 2020; Padmanabhan, Refregier & Amara 2020; Spinelli et al. 2022). The study by Cunnington et al. (2019b) shows that the foreground removal affects two-point function characteristics, especially when the redshift resolution is broad, as is the case in optical photometric surveys.

There are also numerous optical surveys measuring gravitational lensing shear ( $\gamma$ ) which distorts the shape of galaxy images; this is sensitive to density fluctuations of all the matter present along a line of sight, whether baryonic or dark matter. It is therefore of interest to consider the viability of the cross-correlation  $\gamma - \delta_{\text{HI}}$ , which will be able to be studied using a combination of lensing and IM surveys (Hu & Jain 2004; Abbott et al. 2018; Baxter et al. 2019; The CHIME Collaboration et al. 2022; CHIME Collaboration 2023; Cunnington et al. 2023a). The density projection along the unperturbed light ray trajectory, also known as ‘lensing convergence’  $\kappa$ , can be considered instead of  $\gamma$  as both share the same statistical properties. The two-point functions between the pairs of  $\kappa$  and HI could improve cosmological constraints and break degeneracies such as that between HI bias ( $b_{\text{HI}}$ ) and clustering amplitude.

However, removing the HI foreground potentially affects these two-point statistics, as the foreground removal effectively subtracts large-scale radial modes to which lensing is sensitive. In this paper we will calculate the cross-correlation function between convergence and 21cm intensity mapping, and will explore whether the foreground subtraction significantly hampers the cross-correlation measurement. We also explore whether the foreground removal impacts the viability of cosmological constraints from HI–HI and  $\kappa$ –HI correlations.

To achieve this, we will present theoretical and simulation approaches for calculating the  $\kappa$ –HI signal. We will then consider the effect of foreground removal on the signal, showing that the impact is significant (approximately a factor of 2 in signal reduction) but not lethal. We will then use the Fisher information matrix to make cosmological parameter forecasts for ideal and realistic surveys (including cut sky and the inclusion of telescope-specific noise), deploying the cross-correlation between convergence and intensity mapping, always including the effect of foreground removal. We discuss lensing convergence and HI simulation catalogues in Section 2, including modelling of the two-point functions. We describe the HI foreground removal and its effect on  $\kappa$ –HI two-point functions in Section 3. We present our Fisher forecasts for surveys in Section 4, effects of instrumental noise in Section 5, and present our conclusions in Section 6.

## 2 $\kappa$ -HI TWO-POINT STATISTICS: THEORY AND SIMULATIONS

In this section we discuss the relevant two-point statistics. We shall start with theoretical calculations of two-point functions of lensing convergence ( $\kappa$ ) and neutral hydrogen intensity maps (HI) in Section 2.1. We will then discuss the generation of  $\kappa$  catalogues and HI modelling from simulations of the matter overdensity  $\delta$ . The comparison between theoretical calculations and simulations is shown in Section 2.2. The simulated HI maps will be used in the next Section 3 where the foreground removal will be discussed.

### 2.1 Modelling the two-point functions

In this subsection, we describe the modelling of the two-point functions. We begin by considering how to calculate the observable quantities, namely weak lensing convergence  $\kappa$  and HI temperature fluctuations  $\Delta T_{\text{HI}}$ . We will then turn to the angular cross-power spectra. We denote  $\kappa\kappa$  as the power spectra between  $\kappa$  fields,  $\text{HI}^i\text{HI}^j$  as the cross-power spectra between HI fields, and  $\kappa\text{HI}$  as the cross-power between  $\kappa$  and HI. The dummy indices  $i$  and  $j$  refer to the  $i$ th and  $j$ th redshift bins. We will calculate the lensing convergence in an arbitrary direction on the sky  $\hat{n}$  using the Born approximation, projecting the matter overdensity  $\delta$  along an unperturbed ray direction. This can be computed by (Bartelmann & Schneider 2001)

$$\kappa(\chi_s, \hat{n}) = \frac{3\Omega_m H_0^2}{2c^2} \int_0^{\chi_s} d\chi' \frac{\chi'(\chi - \chi)}{\chi} \frac{\delta(\hat{n}, \chi')}{a(\chi')}, \quad (1)$$

where  $\chi$  is comoving distance,  $\Omega_m$  is the matter density parameter at the present epoch,  $H_0$  is the Hubble parameter today, and the subscript  $s$  refers to the source plane. For lensing of distributed sources in redshift bins  $i$ , the integrand is modified by including a source distribution, so that the integration now becomes

$$\kappa^i(\hat{n}) = \int_0^\infty d\chi' q_\kappa^i(\chi') \delta(\hat{n}, \chi'), \quad (2)$$

where the lensing weight is given by

$$q_\kappa^i(\chi) = \frac{3\Omega_m H_0^2}{2c^2} \int_0^{\chi_s} \frac{\delta(\hat{n}, \chi')}{a(\chi')} \int_\chi^\infty d\chi' \frac{\chi - \chi}{\chi} \frac{n_s^i(z(\chi')) \frac{dz}{d\chi'}}{\bar{n}_s^i}, \quad (3)$$

where  $n_s^i(z)$  is the lensing source number density, and  $\bar{n}_s^i$  is its average in the  $i$ th redshift bin.

HI will be a biased tracer of matter overdensity, so we write  $\Delta T_{\text{HI}}(\hat{n}, z) = \bar{T}_{\text{HI}}(z) b_{\text{HI}}(z) \delta(\hat{n}, z)$ , where  $b_{\text{HI}}(z)$  is the HI bias at a given redshift  $z$  and  $\bar{T}_{\text{HI}}(z)$  is the average temperature. The projected temperature fluctuation at the  $i$ th redshift bin is then

$$\Delta T_{\text{HI}}^i(\hat{n}) = \int_0^{\chi_i} d\chi' q_{\text{HI}}^i(\chi') \delta(\hat{n}, \chi'), \quad (4)$$

where

$$q_{\text{HI}}^i(\chi) = \bar{T}_{\text{HI}}(z(\chi)) b_{\text{HI}}^i(\chi) \frac{n_{\text{HI}}^i(z(\chi)) \frac{dz}{d\chi}}{\bar{n}_{\text{HI}}^i}, \quad (5)$$

where  $n_{\text{HI}}^i(z)$  is the HI source number density, and  $\bar{n}_{\text{HI}}^i$  is its average in the  $i$ th redshift bin.

Battye et al. (2013) show that for a given redshift  $z$ ,  $\bar{T}_{\text{HI}}(z)$  can be estimated by

$$\bar{T}_{\text{HI}}(z) = 44 \mu\text{K} \left( \frac{\Omega_{\text{HI}}(z) h}{2.45 \times 10^{-4}} \right) \frac{(1+z)^2}{E(z)}, \quad (6)$$

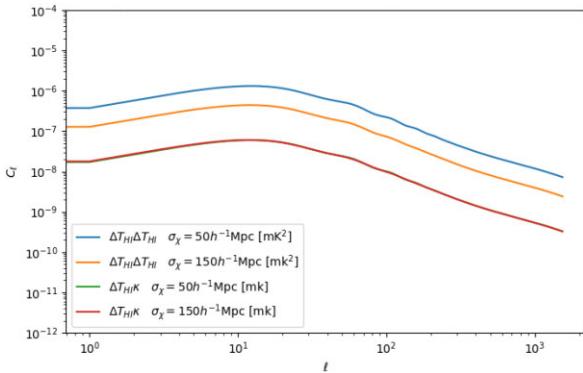
where  $E(z) = H(z)/H_0$  is the dimensionless Hubble function at redshift  $z$ . The HI density parameter could be approximated to be  $\Omega_{\text{HI}} h = 2.45 \times 10^{-4}$  (Battye et al. 2013). However, throughout this research we shall follow the fitting formula for the SKA-MID I by Square Kilometre Array Cosmology Science Working Group et al. (2020)

$$\Omega_{\text{HI}}(z) = 0.00048 + 0.00039z - 0.000065z^2. \quad (7)$$

Constraining the HI bias  $b_{\text{HI}}(z)$  will be discussed later in Section 4.

Using the Limber approximation, the angular power spectra  $C^{XY}(\ell)$  are given by

$$C^{XY}(\ell) = \int d\chi \frac{q_X(\chi) q_Y(\chi)}{\chi^2} P_\delta \left( \frac{\ell + 1/2}{\chi}, z(\chi) \right), \quad (8)$$



**Figure 1.** Power spectra  $C_\ell$  for HI and cross-power between HI and convergence, for radial HI width  $\sigma_\chi = 50$  and  $150 h^{-1} \text{Mpc}$ . The effect of the width is less important for the cross-correlation. Different  $\sigma_\chi$  corresponds to different frequency bandwidths,  $\Delta\nu$  of the radio data.

where  $P_\delta\left(\frac{\ell+1/2}{\chi}, z(\chi)\right)$  is the matter power spectrum (LoVerde & Afshordi 2008). We compute the non-linear power spectrum using the Boltzmann code CAMB (Lewis & Bridle 2002) with the Halofit extension to non-linear scales (Takahashi et al. 2012).

Since the ray-tracing simulations by Takahashi et al. (2017) which we use below adopt a comoving bin size  $\Delta\chi = 150 h^{-1} \text{Mpc}$  (see Section 2.2), we choose a radial selection function for  $n_{\text{HI}}^i(z(\chi))/\bar{n}_{\text{HI}}^i$  as a normal distribution around a central comoving position with  $3\sigma_\chi = 150 h^{-1} \text{Mpc}$ . This  $\sigma_\chi$  corresponds to the frequency bandwidth ( $\Delta\nu$ ) selected. In practice, the frequency range and bandwidth will depend on the particular radio telescope being used; for example, BINGO (Baryon Acoustic Oscillations in Neutral Gas Observations) has operational frequency from 960 to 1260 MHz (Battye et al. 2013; Wuensche & the BINGO Collaboration 2019) and MeerKAT’s frequency bandwidth lies in the ranges 900–1185 and 580–1000 MHz for L band and UHF band, respectively (Wang et al. 2021; Cunningham et al. 2023a).

We show examples for the first time of calculations of the autopower and cross-power for HI and convergence in Fig. 1 using equation (8). As expected, the autosignal depends on the radial HI width  $\sigma_\chi$ , while the cross-power is insensitive to this.

## 2.2 Lensing convergence and HI intensity maps

The full-sky gravitational lensing mock catalogues by Takahashi et al. (2017) have been used throughout this work. They are based on a multiple-lens ray-tracing approach through  $N$ -body cosmological simulations. The data sets include weak lensing maps (convergence, shear, and rotation data) up to redshift 5.3, and halo catalogues. The catalogues provide 108 realizations of  $N$ -body simulations, 35 of which are used in this research (due to storage limitations). The  $N$ -body simulations were produced with periodic boundary conditions following dark matter gravitational evolution without baryonic processes. 14 simulation boxes of side length  $L = 450, 900, 1350, \dots, 6300 h^{-1} \text{Mpc}$  are nested to represent a region of the Universe in which lensing occurs; each box contains  $2048^3$  particles. The  $\kappa$  fields are obtained by tracing the light ray path through planes with separation  $150 h^{-1} \text{Mpc}$ . By calculating the Jacobian matrix  $A$  along the light path, the lensing convergence  $\kappa$ , shear lensing  $\gamma_{1,2}$  and rotation angle  $\omega$  can be obtained, via

$$A = \begin{bmatrix} 1 - \kappa - \gamma_1 - \gamma_2 - \omega \\ -\gamma_2 + \omega \\ -\gamma_1 + \kappa + \gamma_1 \end{bmatrix}. \quad (9)$$

**Table 1.** The central redshifts for intensity and lensing convergence maps in our multiredshift bin analysis. For  $\kappa\Delta T_{\text{HI}}$ , we require that  $z_{\text{HI}} < 0.7z_\kappa$ .

$z_{\text{HI}}$	$z_\kappa$
0.02	0.44
0.08	0.78
0.13	1.77
0.18	
0.24	
0.29	
0.35	
0.41	
0.47	
0.54	
0.60	
0.68	
0.75	
0.83	
0.91	
0.99	

The convergence maps were created in the HEALPix scheme with NSIDE of 4096 (Górski et al. 2005), which contain 200 megapixels. While this resolution is appropriate to study non-linear structure and matches forthcoming galaxy surveys such as EUCLID<sup>1</sup> and DESI<sup>2</sup>, the cross-correlation between the lensing convergence and the HI intensity map is limited by the lower angular resolution of HI intensity maps expected with real radio telescopes. Therefore the resolution is reduced to NSIDE of 512; this is not only appropriate for our two-point function measurements but also decreases the storage space requirement and computational time.

We will first consider a convergence map at a specific optical lensing catalogue source redshift, which we choose as  $z \approx 0.78$ , for which the lensing will significantly occur at the redshift of an intensity map at redshift  $\approx 0.3$ . This particular choice of redshift allows us to compare our results to current and forthcoming optical and radial surveys (Santos et al. 2015; Pourtsidou et al. 2017; Baxter et al. 2019; Euclid Collaboration 2020; Square Kilometre Array Cosmology Science Working Group et al. 2020). We will then extend to multiple lensing planes (see Table 1).

We turn now to generating our IM maps. Crucially, we will emulate removal of the IM foreground by removing the radial temperature fluctuations on large scales. The foreground removal will be discussed in detail in Section 3.

First we need to make the pre-foreground-removal IM maps. Instead of calculating the individual HI masses  $M_{\text{HI}}$  from halo catalogues, we assume that HI is a biased tracer of the total matter overdensity field  $\delta(\theta, z)$  (see equations (4) and (5)),

$$\delta_{\text{HI}}(\hat{n}, z) \equiv \frac{T_{\text{HI}}(\hat{n}, z) - \bar{T}_{\text{HI}}(z)}{\bar{T}_{\text{HI}}(z)} = b_{\text{HI}}(z)\delta(\hat{n}, z), \quad (10)$$

where  $b_{\text{HI}}$  is a HI bias. For instance the parametric form for  $b_{\text{HI}}$  adopted by Cunningham et al. (2019b) is

$$b_{\text{HI}}(z) = 0.67 + 0.18z + 0.05z^2. \quad (11)$$

Since the neutral hydrogen signal is measured as the surface brightness temperature, we shall refer to the HI intensity map as

the temperature fluctuation  $\Delta T_{\text{HI}}$ :

$$\Delta T_{\text{HI}}(\hat{n}, z) = T_{\text{HI}}(\hat{n}, z) - \bar{T}_{\text{HI}}(z) = \bar{T}_{\text{HI}}(z)b_{\text{HI}}(z)\delta(\hat{n}, z). \quad (12)$$

We apply this equation to the overdensity map obtained from Takahashi et al. (2017) catalogues to create HI intensity maps. Fig. 2 shows the uncleaned and cleaned intensity maps from one realization for the zoom-in patch with area  $5 \times 5$  square degrees.

As we are interested in the two-dimensional projection of cosmological fields on the sky, together with their power spectra, it is convenient to describe these fields  $\Theta(\hat{n}, z)$  in spherical harmonics:

$$\Theta(\hat{n}, z) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{m=\ell} a_{\ell m}(z) Y_{\ell}^m(\hat{n}), \quad (13)$$

where  $Y_{\ell}^m(\hat{n})$  and  $a_{\ell m}(z)$  are spherical harmonics and their coefficients, respectively (Heavens 2003; Castro, Heavens & Kitching 2005; Pratten et al. 2016).  $\Theta(\hat{n}, z)$  represents an arbitrary cosmological field; in this work it can be either lensing convergence or HI temperature fluctuations. The angular power spectrum is then an average of  $a_{\ell m}$  over  $m$  modes:

$$C^{XY}(\ell) = \langle a_{\ell m}^X(z_1) a_{\ell m}^{Y*}(z_2) \rangle, \quad (14)$$

where  $X$  and  $Y$  stand for the cosmological fields at given redshifts  $z_1$  and  $z_2$ , respectively.

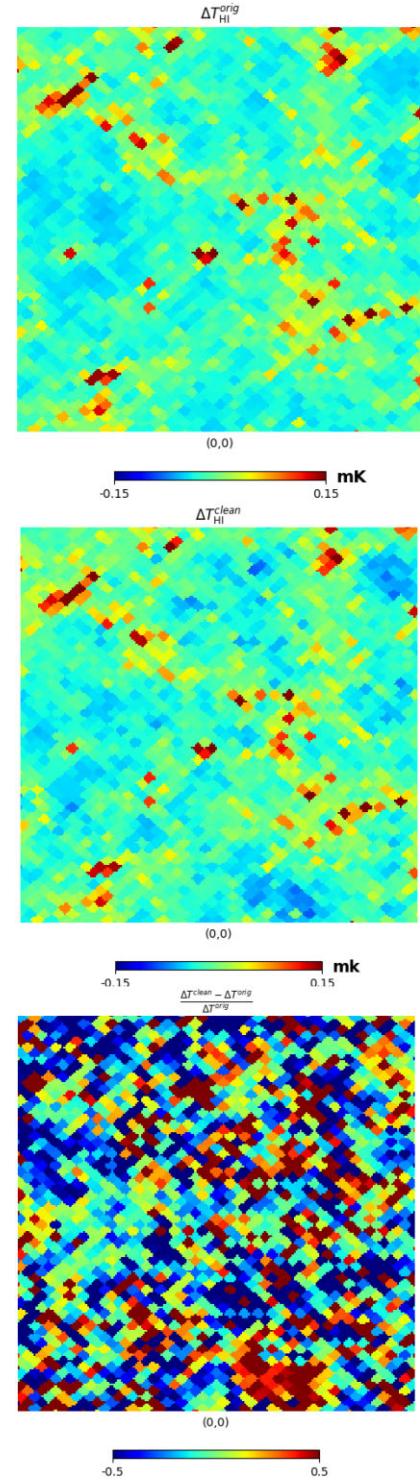
The cross-power spectrum for HI and lensing  $\kappa$  can be easily measured via HEALPix's `anafast` routine especially if the data are for the full sky [however, if the data have missing regions or a cut sky, pseudo- $C_{\ell}$  methods are required (Brown, Castro & Taylor 2005; Upham et al. 2019)].

Using this routine, we obtain cross-power measurements for the HI and  $\kappa$  fields. We measure the cross-power spectra of 35 realizations and evaluate their mean; we show the results in Fig. 3. Here the lensing convergence is measured at the central redshift 0.78 and HI is measured at the central redshift 0.3. Fig. 3 also displays a comparison between theoretical two-point statistics and the measurements from the mock catalogues. We then measure the covariance matrices  $\text{COV}(C^{XY})$  of two-point statistics from 35 realizations. The error bars are the square root of the diagonal elements of  $\text{COV}(C^{XY})$  of the estimators. The correlation matrix for  $\text{COV}(C^{XY})$  are shown in Fig. 6.

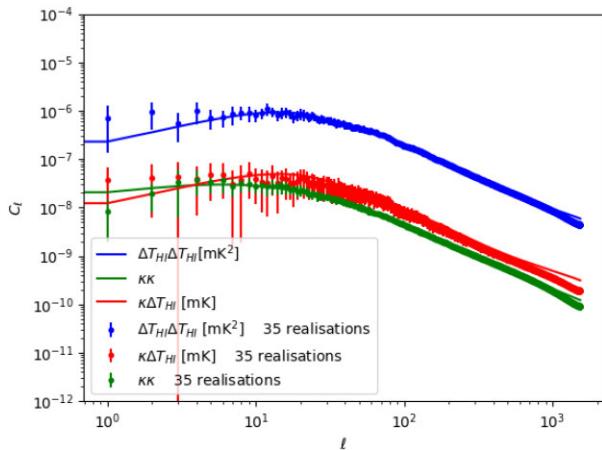
We see that the measurements from simulations agree very well with our theory curves on this plot, which indicates that our theoretical calculation and selection function  $n_{\text{HI}}^i(z(\chi))/\bar{n}_{\text{HI}}^i$  successfully match the simulations. Due to the lens shell approximation of the ray-tracing code, the measured  $C_{\ell}$  is slightly affected at very high  $\ell$  (see red line on Fig. 3).

### 3 HI FOREGROUND REMOVAL AND ITS EFFECT ON $\kappa$ HI TWO-POINT FUNCTIONS

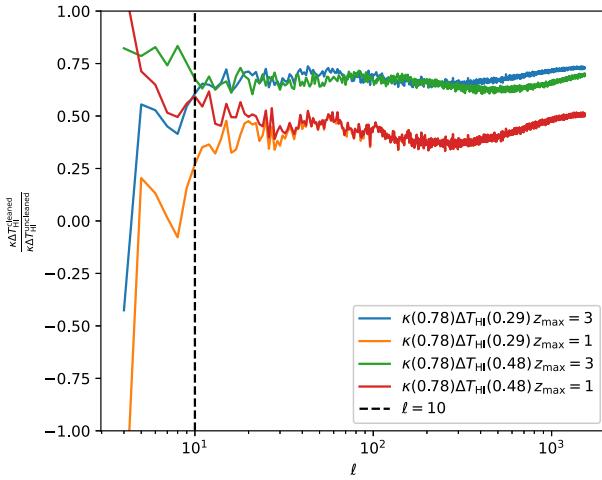
The HI signal is small compared to its foregrounds such as free-free thermal emission, extragalactic radio sources, and Galactic synchrotron. For example, the synchrotron ( $T_{\text{sync}}$ ) emission temperature, which can be modelled by  $T_{\text{sync}} \propto (1+z)^{2.7} [\text{K}]$  (Platania et al. 1998; Smoot & Debano 2017), is approximately three to four orders of magnitude larger than  $T_{\text{HI}}$  at low redshift. Thus, 21cm foreground removal is a major challenge for HI cosmology. Several studies suggest that the foreground spectrum appears to be smooth in the radial direction (Shaw et al. 2014; Cunningham et al. 2019a, b). This is equivalent to being present in the long radial wavelengths in Fourier space. We therefore remove such modes in the line-of-sight background temperature fluctuations  $\Delta T_{\text{HI}}^{\text{LoS}}(\hat{n})$ .



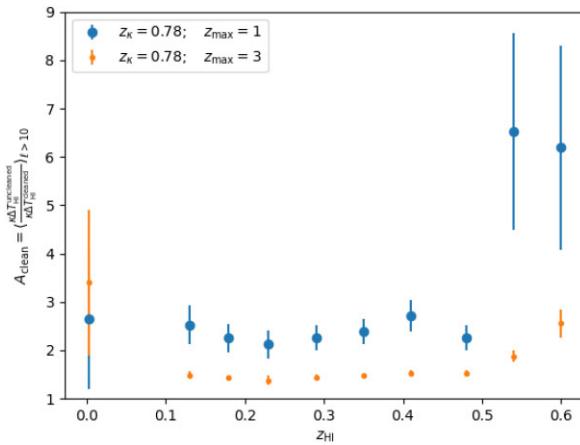
**Figure 2.** Top: the uncleaned intensity map,  $\Delta T_{\text{HI}}^{\text{orig}}$ , at  $z = 0.3$  from an example realization. The fluctuations were measured assuming  $b_{\text{HI}}(z)$  (see equations (12) and (17)). Middle: the foreground-removed intensity map,  $\Delta T_{\text{HI}}^{\text{clean}}$ , at the same redshift. The foregrounds were removed by eliminating radial long wavelength modes up to redshift  $z_{\text{max}} = 1$ . The NSIDEs of the fluctuation maps is reduced from 4096 to 512 to match the resolution of forthcoming radio surveys. Bottom: residual map of cleaned and uncleaned maps. Each of these detail maps has area  $5 \times 5$  square degrees (a small patch of the entire sky maps).



**Figure 3.** Comparison between theoretical  $C_\ell$  (see equation (8)) and measured  $C_\ell$  from our simulations. Here the lensing convergence is measured at central redshift 0.78 and HI is measured at central redshift 0.3.



**Figure 4.** The ratio between cleaned and uncleaned  $\kappa\Delta T_{HI}$  power spectra. Two maximum redshifts ( $z_{\max}$ ) for foreground removal are considered;  $z_{\max} = 1$  corresponds to current and imminent radio dishes, while  $z_{\max} = 3$  represents a future SKA survey.



**Figure 5.** The average ratio  $A_{\text{clean}}$  over  $\ell > 10$  modes, as a function of redshift of HI slice used in the cross-correlation.

Since the calculation of lensing involves integration along the light path (equation (3)), which will have a contribution from long-wavelength radial modes, the HI foreground removal is a concern for the existence of the  $\kappa$ –HI cross-correlation (i.e. we have just removed such modes from the HI signal). In this section we therefore seek to ascertain the degree to which the  $\kappa$ –HI cross-correlation survives foreground removal.

Here we follow the method for foreground removal emulation by Cunningham et al. (2019b). The cleaned intensity map  $\Delta T_{\text{HI}}^{\text{clean}}$  can be approximated as

$$\Delta T_{\text{HI}}^{\text{clean}}(\hat{n}, z) = \Delta T_{\text{HI}}^{\text{orig}}(\hat{n}, z) - \Delta T_{\text{HI}}^{\text{LoS}}(\hat{n}), \quad (15)$$

where  $\Delta T_{\text{HI}}^{\text{orig}}(\hat{n}, z)$  is the uncleaned signal in direction  $\hat{n}$  at redshift  $z$ .  $\Delta T_{\text{HI}}^{\text{LoS}}(\hat{n})$  is defined by

$$\Delta T_{\text{HI}}^{\text{LoS}}(\hat{n}) = \frac{1}{N_z} \sum_i \bar{T}_{\text{HI}}(z_i) b_{\text{HI}}(z_i) \delta(\hat{n}, z_i), \quad (16)$$

so that  $\Delta T_{\text{HI}}^{\text{LoS}}(\hat{n})$  is the mean surface brightness temperature fluctuation along the entire line of sight. This is an initial very approximate model of principal component analysis foreground removal, as most dominant components are included in the line-of-sight expectation temperature fluctuations  $\Delta T_{\text{HI}}^{\text{LoS}}(\hat{n})$ . It is worth mentioning that this blind foreground removal technique assumes the smoothness of the foreground. However, this smoothness can be hampered by non-smooth features of the beam, e.g. beamwidth of the radio dish, and some oscillating features in all bands of MeerKAT. A simple  $1/f$  dependence of the beam could generate artificial HI signals. This leads to the conclusion in Spinelli et al. (2022) that it is fundamental to develop accurate beam deconvolution algorithms and test data post-processing steps carefully before cleaning. This topic of beam deconvolution is beyond the scope of our research; here we shall assume that the  $1/f$  behaviour is sufficiently small. For more sophisticated foreground cleaning methods we encourage the reader to explore e.g. Cunningham et al. (2023b).

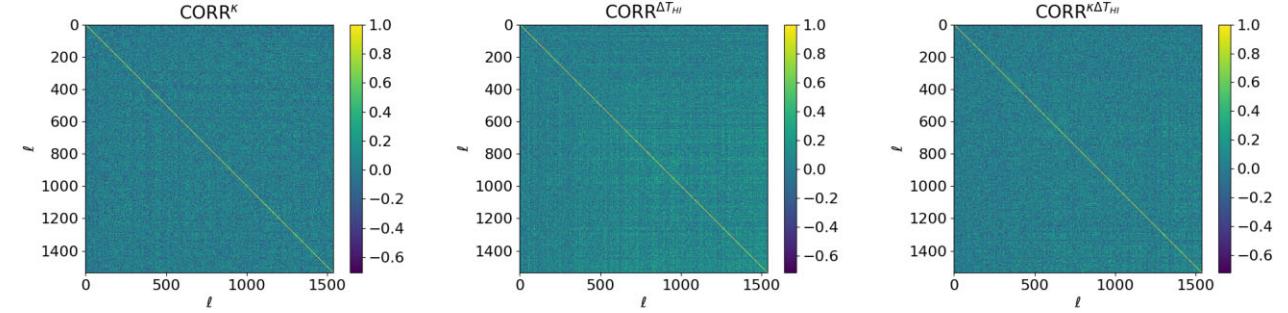
In this work we adopt the same bias model as Cunningham et al. (2019a):

$$b_{\text{HI}}(z) = \alpha(b_0 + b_1 z + b_2 z^2), \quad (17)$$

where  $\alpha, b_0, b_1$ , and  $b_2$  are set to 1, 0.67, 0.18, and 0.05, respectively. Cunningham et al. (2019a) obtained this parameter set by investigating HI as a biased tracer of the LSS via HI intensity map and optical galaxy number density cross-correlations (see eq. 39 from Cunningham et al. 2019a). We use this as a fiducial model since the HI redshift range in our work is similar to Cunningham et al. (2019a, b). Note that in this model, we solely account for the redshift evolution of HI bias and assume any transverse scale dependence of the bias is negligible. Martin et al. (2012) show that this is a good approximation for scales  $> 10 h^{-1}\text{Mpc}$ , which are our main interest.

We measure  $\Delta T_{\text{HI}}^{\text{LOS}}$  with two choices of maximum redshift,  $z_{\max} = 1$  and 3.  $z_{\max} = 3$  corresponds to futuristic HI-galaxy surveys (Square Kilometre Array Cosmology Science Working Group et al. 2020). On the other hand  $z_{\max} = 1$  is an approximate limit for HI maps with SKA1-MID and MeerKAT (Square Kilometre Array Cosmology Science Working Group et al. 2020; Cunningham et al. 2023a).

We use these intensity maps with removed foreground to calculate the autopower spectra of the intensity map ( $\Delta T_{\text{HI}}\Delta T_{\text{HI}}$ ), and the cross-power spectra between HI and  $\kappa$  ( $\kappa\Delta T_{\text{HI}}$ ). We compare the signal of removed and unremoved  $\kappa\Delta T_{\text{HI}}$ , resulting in Fig. 4 and Fig. 5. From Fig. 4, we note that foreground removal strongly affects the signal on large scales. However we find that on smaller scales,



**Figure 6.** The correlation matrices of  $C_\ell^{XY}$  measured from 35 realizations for the cleaned HI signal at central redshift  $z = 0.3$  and  $\kappa$  signal at  $z = 0.78$ . Left:  $\kappa\kappa$ , middle:  $\Delta T_{\text{HI}}\Delta T_{\text{HI}}$ , and right:  $\kappa\Delta T_{\text{HI}}$ .

at  $\ell > 10$ , the foreground removal does not erase the  $\kappa\Delta T_{\text{HI}}$  power spectrum; the signal is scaled down by a factor ( $A_{\text{clean}}$ ) which is close to constant over a range of  $\ell$  modes from 10 to 1000. Hence, in Section 4, when cosmological constraints following foreground removal are considered, the estimation of cosmological parameters is based on the signal where  $\ell > 10$ . We describe the mean signal drop  $A_{\text{clean}}$  by

$$A_{\text{clean}}(z_{\text{HI}}, z_\kappa, z_{\text{max}}) \equiv \left\langle \frac{\kappa\Delta T_{\text{HI}}^{\text{uncleaned}}}{\kappa\Delta T_{\text{HI}}^{\text{cleaned}}} \right\rangle_{10 < \ell < 1500}. \quad (18)$$

From Fig. 4, we see that the higher the maximum redshift of the survey in which we remove the LOS signal, the less is the effect on the cleaned cross-correlation signal, as more radial modes are preserved in the removal process. Figs 4 and 5 further indicate that the signal in the  $\kappa$ HI two-point correlations drops by approximately the same factor  $A_{\text{clean}}$  across a wide redshift range, if we remove the background noise up to a particular redshift  $z_{\text{max}}$ , when cross-correlating to the  $\kappa$  field at a fixed redshift. Fig. 5 also implies that  $A_{\text{clean}}(z_{\text{HI}}, z_\kappa, z_{\text{max1}}) < A_{\text{clean}}(z_{\text{HI}}, z_\kappa, z_{\text{max2}})$  if the maximum redshifts  $z_{\text{max1}} > z_{\text{max2}}$ .

## 4 FISHER FORECAST

In the previous sections, we have presented the theoretical two-point statistics for the HI-lensing cross-correlation, and have examined the impact of HI foreground removal on the cross-power spectrum. The results indicate that foreground removal reduces the two-point statistics by a modest factor.

Here we begin the exploration of  $\kappa$ -HI correlations as a tool for cosmological constraints. In particular we will make a Fisher information matrix forecast of this correlation in the case of low instrument noise (but including our foreground subtraction model); this will assess the best-case capacity of this probe to constrain cosmology, when one is dominated by LSS fluctuations in the HI and lensing fields. We will then examine more realistic cases with cut sky and the inclusion of instrumental noise.

### 4.1 The Fisher matrix

The Fisher information matrix is a useful tool to estimate the expected uncertainty in cosmological parameters for forthcoming experiments (Tegmark, Taylor & Heavens 1997; Heavens 2003). Assuming that the model parameters  $\theta_i$  are distributed by a multivariate Gaussian likelihood  $L$ , the Fisher matrix can be calculated as

$$\mathbf{F}_{ij} \equiv \left\langle \frac{\partial^2 \mathcal{L}}{\partial \theta_i \partial \theta_j} \right\rangle, \quad (19)$$

where  $\mathcal{L} = -\ln L$ . The Fisher matrix can be used to obtain the minimum uncertainty ( $\sigma_i$ ) in parameter estimation due to the Cramér–Rao inequality (Kamionkowski, Smith & Heavens 2011; Mendez et al. 2014),

$$\sigma_i \geq \sqrt{\mathbf{F}_{ii}^{-1}}, \quad (20)$$

which is equivalent to a 68 per cent confidence level. For a data set where the uncertainties are Gaussian, the Fisher matrix can be calculated by (Tegmark et al. 1997)

$$\mathbf{F}_{ij} = \frac{1}{2} \text{Tr}[\mathbf{A}_i \mathbf{A}_j + \mathbf{C}^{-1} \mathbf{M}_{ij}], \quad (21)$$

where  $\mathbf{C}$  denotes the covariance matrix of the data,  $\mathbf{A}_i \equiv \mathbf{C}^{-1} \mathbf{C}_{:,i}$ , the derivative data matrix  $\mathbf{M}_{ij} \equiv \boldsymbol{\mu}_{:,i} \boldsymbol{\mu}_{:,j}^T + \boldsymbol{\mu}_{:,j} \boldsymbol{\mu}_{:,i}^T$ , and  $\boldsymbol{\mu}$  is an expectation value of the data vector  $\mathbf{x}$ . The comma symbol means the partial derivative operator with respect to the parameter,  $\boldsymbol{\mu}_{:,i} \equiv \partial \boldsymbol{\mu} / \partial \theta_i$ . Note that all derivatives are performed at the maximum likelihood point.

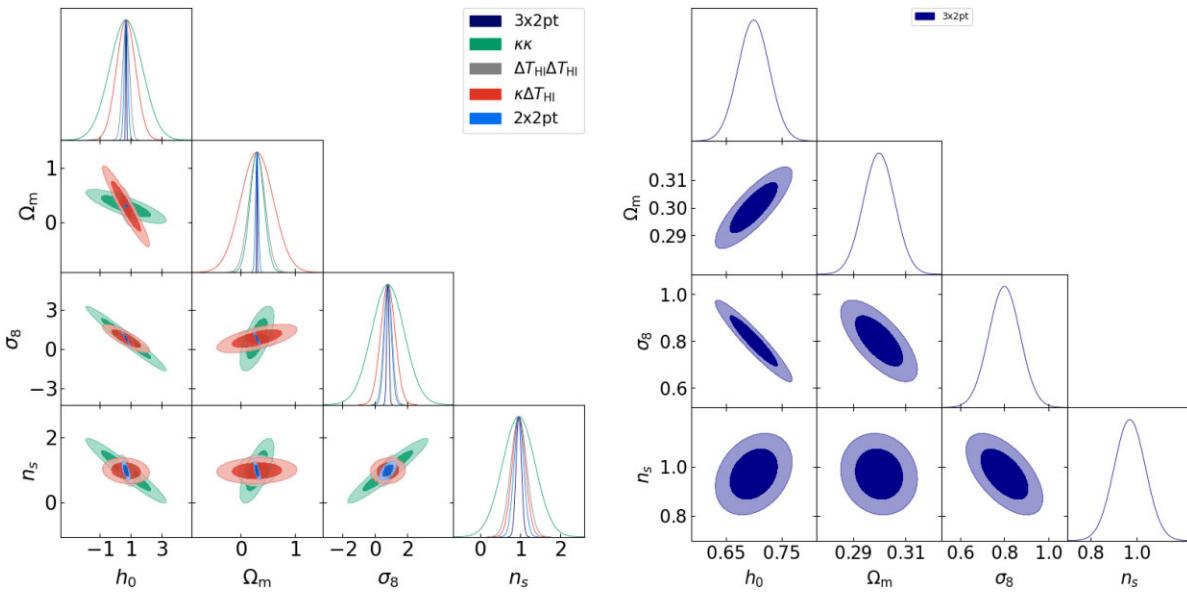
As we expect only small changes in the covariance matrix  $\text{COV}(C^{XY})$  under a modest change in cosmological parameters (see Sections 2 and 3), the first term on the right-hand side in equation (21) will be negligible. Then the Fisher matrix can be written

$$\mathbf{F}_{ij} = \sum_{XY} \frac{\partial C^{XY}}{\partial \theta_i} \text{COV}(C^{XY})^{-1} \frac{\partial C^{XY}}{\partial \theta_j}. \quad (22)$$

We calculate the cross-power spectra  $C^{XY}(\ell)$  using equation (8) with Planck 2018 cosmological parameters (Planck Collaboration I 2020). We calculate the covariance matrices of  $\kappa\kappa$ ,  $\kappa$ HI and HIHI from measured cross-power spectra of 35 realizations of the  $N$ -body simulation by Takahashi et al. (2017). All the HI temperature fluctuation maps which we use take into account foreground removal. We also calculate the correlation matrices  $\text{CORR}_{ij} = \text{COV}_{ij} / \sqrt{\text{COV}_{ii} \text{COV}_{jj}}$  and show these in Fig. 6; these do not indicate significant correlations between  $\ell$  bins.

### 4.2 Cosmological constraints for single-slice cross-correlations

In this section, the cosmological constraint viability of  $\kappa$  and HI cross-correlations is explored. We first start with the simplest observational configuration, considering only one redshift slice of HI and  $\kappa$ . We further assume that the  $b_{\text{HI}}(z)$  behaves as in equation (17). We use the Planck 2018 cosmological parameters as the fiducial cosmology (Planck Collaboration I 2020). The fiducial cosmological parameters are  $h = 0.67$ ,  $\Omega_m = 0.3$ ,  $\sigma_8 = 0.82$ ,  $\Omega_k = 0$ ,  $\Omega_\Lambda = 0.7$ ,  $\tau = 0.06$ , and  $n_s = 0.96$ . To make a covariance matrix of cross-power spectra for the Fisher matrix (equation (22)), we combine  $\ell$  modes into 15 bins; each bin contains 101  $\ell$  modes with  $11 \leq \ell \leq 1527$  and



**Figure 7.** Left: Likelihood contours for the data set described in Section 4.2; contours show 68 per cent and 95 per cent confidence levels. Right: zoom-in of likelihood contours of  $3\times 2$ -point functions and their marginalizations from the left panel.

averages over 35 realizations. We first consider the  $3\times 2$  functions for a joint analysis of  $\kappa(0.78)\kappa(0.78)$ ,  $\Delta T_{\text{HI}}(0.3)\Delta T_{\text{HI}}(0.3)$ , and  $\kappa(0.78)\Delta T_{\text{HI}}(0.3)$ , where the numbers in brackets are the central redshifts. We choose these central redshifts as examples of current HI and lensing surveys' central redshifts. The '2x2' functions refer to the same combination but exclude the weak lensing-HI cross-power spectrum.

Fig. 7 shows the joint likelihood obtained via Fisher matrices (see equation (22)). We see that single redshift slice correlations of  $\kappa - \kappa$  (green) and HI-HI (grey) provide relatively weak constraints, while 2x2pt and particularly  $3\times 2$ pt are more promising, with few-to-10-per cent constraints available on parameters in this low noise case. The zoom-in version of  $3\times 2$  pt functions is shown on the right-hand side of Fig. 7; these likelihood contours, which include the cross-correlation, show a significant improvement in cosmological constraints compared to  $\kappa\kappa$  or  $\Delta T_{\text{HI}}\Delta T_{\text{HI}}$  constraints alone. Therefore in the next section, we will examine a joint likelihood between more redshift bins, and where the HI bias ( $b_{\text{HI}}(z)$ ) is taken into account.

#### 4.3 HI bias and multiredshift bin joint likelihood analysis

It is well known that there is a degeneracy between galaxy bias,  $\Omega_m$  and  $\sigma_8$  in parameter constraints, since these parameters all affect the amplitude of the power spectrum (see equation (8)). However, they contribute differently to the evolution of the power spectrum with time; hence by measuring the power spectra in various redshifts we can break the degeneracies between them. From equation (8), we can see that while  $\Delta T_{\text{HI}}\Delta T_{\text{HI}}$  measures  $b_{\text{HI}}^2(z)$ ,  $\kappa\Delta T_{\text{HI}}$  additionally measures  $b_{\text{HI}}(z)$ . Combining the cross-bin intensity mapping power spectra with the  $\kappa\Delta T_{\text{HI}}$  cross-spectra we can therefore tighten our constraints on bias and cosmological parameters.

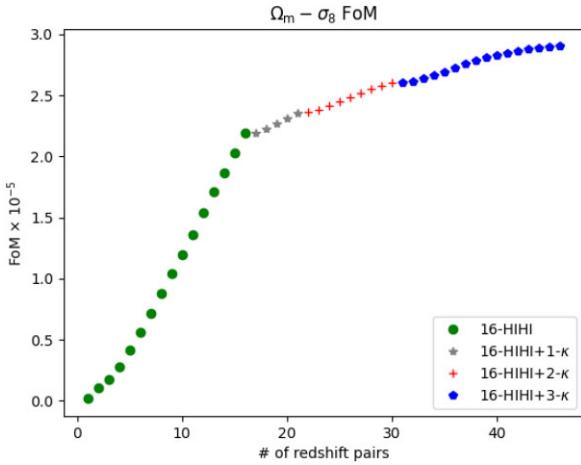
In this section we consider two different bias models, with distinct parameter sets. The first, more restricted model explores bias amplitude variation via the  $\alpha$  parameter in equation (17), setting the rest of the parameters to the best-fitting values (Cunnington et al. 2019b). In the second model,  $b_0$ ,  $b_1$ , and  $b_2$  are the bias parameters

with  $\alpha$  set to equal 1. We include these parameters when evaluating the Fisher matrices (equation (22)). Note that both  $b_{\text{HI}}$  models are scale-invariant and depend only on  $z$ . We will consider both full-sky and 300 deg<sup>2</sup> surveys to explore the viability of HIHI and  $\kappa$ HI in cosmological constraints.

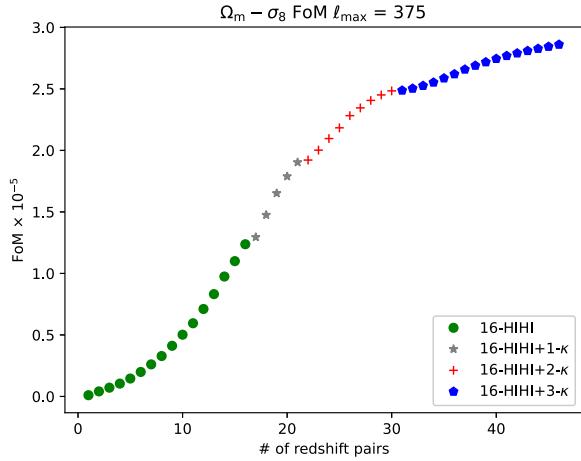
For the full-sky case, as we include more parameters for  $b_{\text{HI}}$ , we also examine more redshift bins for both HI and  $\kappa$  to obtain the best possible results. We consider the redshift range for  $\Delta T_{\text{HI}}$  which would be measured by pre-SKA and SKA-MID experiments (Santos et al. 2015; Poutsidou et al. 2017; Square Kilometre Array Cosmology Science Working Group et al. 2020). Table 1 shows the central redshifts we consider for  $\Delta T_{\text{HI}}$  and  $\kappa$  bins; the width of each bin is 150  $h^{-1}$  Mpc,  $\Delta z \approx 0.05$ . Table 1 lists both HI and  $\kappa$  central redshifts. As we have 16  $z_{\text{HI}}$ , we shall refer to '16-HIHI' which corresponds to 16 pairs of HI autocorrelation functions; we cross-correlate HI intensity maps to  $\kappa$  fields at  $z_{\kappa} = 0.44, 0.78$ , and 1.77, respectively. We refer to 16-HIHI+1- $\kappa$  as the joint analysis for 16-HIHI and  $\kappa$ HI two-point statistics at which  $z_{\kappa} = 0.44$ . We add further  $z_{\kappa}$  bins and label joint data as 16-HIHI+2- $\kappa$  and 16-HIHI+3- $\kappa$ . We calculate both the futuristic case where HI can be measured with high  $\ell_{\text{max}} \geq 1000$  and the current state of art where  $100 < \ell_{\text{max}} < 400$ .

We further calculate the figure of merit (FoM) for the  $\Omega_m - \sigma_8$  constraint. The FoM is the inverse of the area of the  $\Omega_m - \sigma_8$  contours; in this case we calculate the FoM at 95 per cent confidence level. Fig. 8 shows the FoM of the  $\Omega_m - \sigma_8$  constraints. The blue dots show the FoM from 16HIHI, where we consecutively add HI autocorrelations for the redshift bins in the order listed in Table 1. We see that all redshift bins contribute to an improved signal, with a nearly linearly increasing contribution (for this experiment, we assume that only  $z < 1$  HI slices are available). The green, red, and black dots in Fig. 8 show the FoM for  $\ell_{\text{max}} = 1530$  when we further add the cross-correlations between consecutive HI slices and the  $\kappa$  slices at  $z = 0.44, 0.78$ , and 1.78, respectively. We see that these cross-correlations significantly improve the FoM, and appear to be converging to a maximal constraint when including all slices.

We now present Fisher forecast results for the first bias model, using the redshift bins in Table 1. Fig. 10 shows the utility of



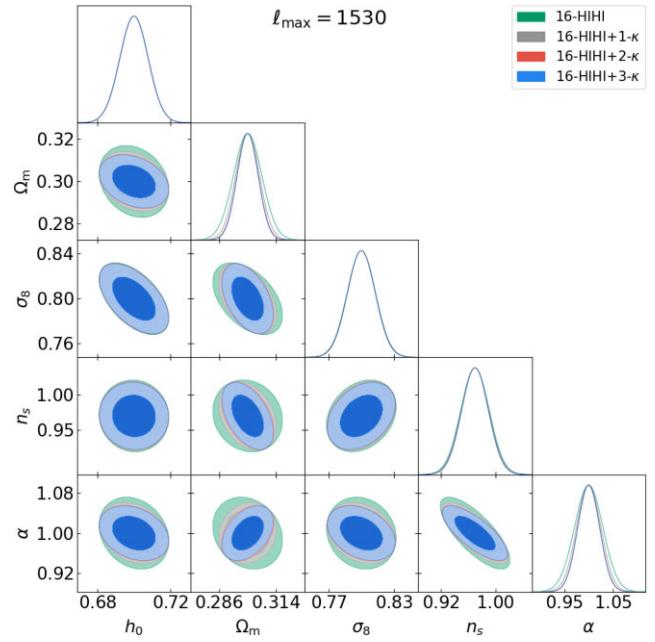
**Figure 8.** FoM for  $\sigma_8 - \Omega_m$  constraints; the horizontal axis is the number of redshift bin pairs for cosmological constraints. We show cumulative FoM when including increasing numbers of HI autocorrelation redshift bins (green); then increasing numbers of cross-correlations with convergence bins (grey, red, blue). Here  $\ell_{\max} = 1530$ .



**Figure 9.** FoM for  $\sigma_8 - \Omega_m$  constraints; the horizontal axis is the number of redshift bin pairs for cosmological constraints. We show cumulative FoM when including increasing numbers of HI autocorrelation redshift bins (green); then increasing numbers of cross-correlations with convergence bins (grey, red, blue). Here  $\ell_{\max} = 375$ .

multiredshift bin power spectra measurements with HI and  $\kappa$ . With the appropriate redshift bin size of HI ( $\Delta z_{\text{HI}}$ ) for  $\ell_{\max} = 1530$ , we can achieve tight cosmological constraints (in this low noise case) which are comparable to the optical and CMB probes of Alam et al. (2017), Abbott et al. (2018), Planck Collaboration I (2020), and Abbott et al. (2019). By including more  $\kappa$  redshift slices, the constraints are improved significantly especially for  $\Omega_m$  and  $\alpha$ . However, this also makes the contours more elliptical, as there are remaining degeneracies among parameters. The uncertainties on parameters are measured at 95 per cent confidence level and reported in Table 2.

We next consider the current state-of-the-art case, where, since the typical angular resolution is  $\sim 1^\circ$ , we set  $\ell_{\max} = 375$  (we choose this particular value as it is convenient to consider the  $\Delta\ell$  bins as 15 bins with  $\Delta\ell = 25$ ). Fig. 9 shows the cumulative FoM in this case, while Fig. 11 illustrates the likelihood contours for cosmological parameters with this maximum multipole. Comparing with Fig. 10,



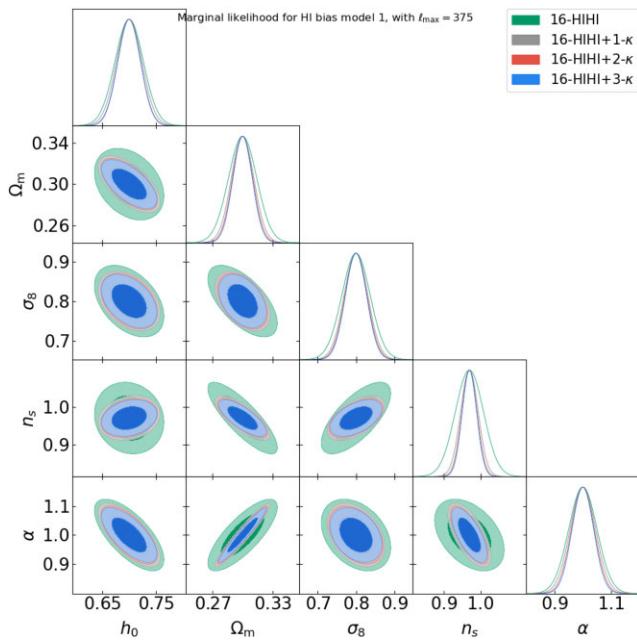
**Figure 10.** The likelihood contours for our multibin analysis with HI bias model 1 with  $\ell_{\max} = 1530$ ; the contours show 68 per cent and 95 per cent confidence levels. The cosmological parameter uncertainties at 95 per cent are measured and reported in Table 2.

**Table 2.** Cosmological forecast from Fisher analysis for 16-HIHI + 3- $\kappa$  correlations; the uncertainties on cosmological parameters and HI bias are quoted at 95 per cent confidence level. HI bias model 1 considers only the re-scaling parameter  $\alpha$ . In contrast the second model considers quadratic parameters;  $b_2$  is poorly constrained, while all other parameters are able to be measured well (see Fig. 12).

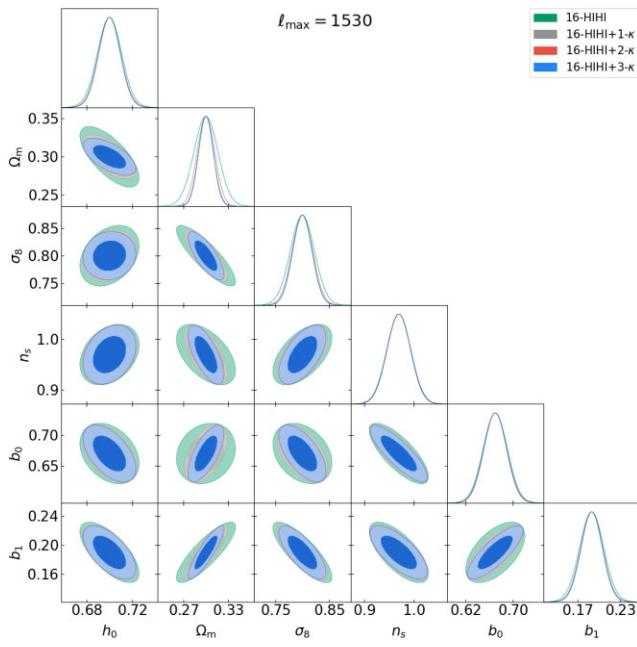
Parameters	HI bias model 1	HI bias model 2
$\Delta h_0$	$\pm 0.02$	$\pm 0.02$
$\Delta \Omega_m$	$\pm 0.01$	$\pm 0.02$
$\Delta \sigma_8$	$\pm 0.03$	$\pm 0.04$
$\Delta n_s$	$\pm 0.04$	$\pm 0.05$
$\Delta \alpha$	$\pm 0.04$	-
$\Delta b_0$	-	$\pm 0.04$
$\Delta b_1$	-	$\pm 0.03$

where  $\ell_{\max} = 1530$  we can see that there is a substantial difference in the 16-HIHI contours. However, we notice the significant improvement in parameter constraints when we add 3  $\kappa$  bins to 16-HIHI (green shade) in  $\ell_{\max} = 375$ . We are therefore seeing that by joining  $\kappa$ HI two-point statistics to HIHI autocorrelations, we can improve cosmological constraints significantly.

For bias model 2, we find that the second-order coefficient of HI bias  $b_2$  is very poorly constrained. Marginalizing over this parameter does not significantly affect the other cosmological parameter constraints ( $h_0$ ,  $\Omega_m$ ,  $\sigma_8$ , and  $n_s$ ). We set this parameter to 0.05 following Cunningham et al. (2019b). Fig. 12 shows the likelihood constraints for this model. We can see that by adding more  $\kappa$  slices, the  $h_0$  constraints do not improve much but the improvement in the  $\Omega_m - \sigma_8$  constraint can be easily noticed. The uncertainties on our HI bias models and cosmological parameters are reported in Table 2. Comparing the parameter constraints from both  $b_{\text{HI}}$  models (see Table 2), we notice that model 1 gives slightly better (but very comparable) cosmological constraints.



**Figure 11.** The likelihood contours for our multibin analysis with HI bias model 1 with  $\ell_{\max} = 375$ ; the contours show 68 per cent and 95 per cent confidence levels. Compared to Fig. 10, we can see the difference in 16-HIHI constraint. However, when combining the  $\kappa$ HI-correlation the constraints are similar to those in Fig. 10.



**Figure 12.** Constraints on cosmological parameters and  $b_{\text{HI}}(z)$  for our second bias model, for two-point functions at 68 per cent and 95 per cent levels of confidence (for our 16HIHI redshift slice case). The  $\kappa$ HI correlation functions do not significantly improve the  $h_0$ ,  $n_s$ ,  $b_0$ , and  $b_1$  constraints. For this figure, we marginalized over  $b_2$  as it is poorly constrained. However, we see significant improvement in  $\Omega_m$  and  $\sigma_8$  constraints. The parameter uncertainties at 95 per cent level are reported in Table 2.

#### 4.4 The effect of sky coverage

Now let us consider the effect of sky coverage area on two-point statistics of both HIHI auto and  $\kappa$ HI cross angular power spectra. We now consider a combined survey area  $300 \text{ deg}^2$  of lensing and HI observations, comparable to current pathfinder intensity mapping surveys. As this area is much smaller than the full-sky case, we therefore now need to use the pseudo angular power ( $\tilde{C}_\ell$ ) as an estimator of  $C_\ell$ .

Suppose the survey footprint of the observations can be expressed using the weight function  $W(\hat{n})$ . Normalized by the sky factor  $f_{\text{sky}}$  (the fraction of the sky covered by the data), the weight moments are given by

$$f_{\text{sky}} w_i = \frac{1}{4\pi} \int d\hat{n} W^i(\hat{n}), \quad (23)$$

where  $w_i$  represents the  $i$ -th moment of weighting. The power spectrum of the window function is

$$\mathcal{W}_\ell = \frac{1}{2\ell + 1} \sum_m |w_{\ell m}|^2. \quad (24)$$

For a spin-0 field  $\zeta(\hat{n})$  weighted by  $W(\hat{n})$ , a spherical harmonic coefficient  $\tilde{a}_{\ell m}$  can be expressed as (Kim & Naselsky 2010; Kim 2011)

$$\tilde{a}_{\ell m} = \int d\hat{n} \zeta(\hat{n}) W(\hat{n}) Y_{\ell m}^*(\hat{n}) \approx \Omega_p \sum_p \zeta(p) W(p) Y_{\ell m}^*(p), \quad (25)$$

where we approximate the integration over sky factor by the summation over pixel area with the surface density  $\Omega_p$ . The pseudo power spectrum estimator,  $\tilde{C}_\ell$ , is then

$$\tilde{C}_\ell = \frac{1}{2\ell + 1} \sum_m |\tilde{a}_{\ell m}|^2. \quad (26)$$

Similarly for spin-2 fields ( $\gamma_1, \gamma_2$ ), we can obtain the coefficients,  $\tilde{a}_{\pm 2, \ell m}$ , by

$$\tilde{a}_{\pm 2, \ell m} = \int [\tilde{\gamma}_1(\hat{n}) \pm i \tilde{\gamma}_2(\hat{n})] \pm 2 Y_{\ell m}^*(\hat{n}) d\hat{n}, \quad (27)$$

where

$$\tilde{\gamma}_1(\hat{n}) \pm i \tilde{\gamma}_2 = W(\hat{n}) [\gamma_1(\hat{n}) \pm i \gamma_2(\hat{n})]. \quad (28)$$

Similarly to the full-sky case,  $\tilde{a}_{\ell m}^E$  and  $\tilde{a}_{\ell m}^B$  are then

$$\tilde{a}^E = -(\tilde{a}_{2, \ell m} + \tilde{a}_{-2, \ell m})/2, \quad (29)$$

$$\tilde{a}^B = i(\tilde{a}_{2, \ell m} - \tilde{a}_{-2, \ell m})/2. \quad (30)$$

The pseudo power spectra for E and B modes are

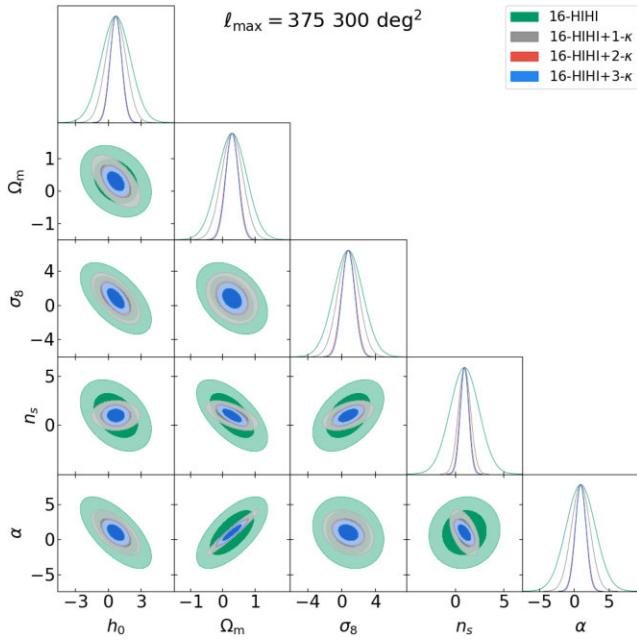
$$\tilde{C}_\ell^{E, B} = \frac{1}{2\ell + 1} \sum_m |\tilde{a}_{\ell m}^{E, B}|^2. \quad (31)$$

The pseudo power spectra  $\tilde{C}_\ell$  and true  $C_\ell$  are related by the mode-mode coupling resulting from masking ( $M_{\ell \ell'}$ ):

$$\langle \tilde{C}_\ell \rangle = \sum_{\ell} M_{\ell \ell'} \langle C_{\ell'} \rangle. \quad (32)$$

This kernel depends solely on the geometry of a cut-sky  $\mathcal{W}_\ell$  and plays a crucial role in the pseudo- $C_\ell$  method. For details concerning the mode-mode coupling, see Hivon et al. (2002) and Alonso et al. (2019).

We utilize NAMASTER (Alonso et al. 2019), which is a software package to calculate pseudo- $C_\ell$  for any spin fields, to evaluate HIHI and  $\kappa$ HI pseudo- $C_\ell$  for  $300 \text{ deg}^2$  of our simulations above, within a



**Figure 13.** Constraints on cosmological parameters and  $b_{\text{HI}}(z)$  for two-point functions at 68 per cent and 95 per cent levels of confidence, for the small area  $300 \text{ deg}^2$  case. This plot indicates that the feasibility of  $\kappa$ HI (pseudo) two-point statistics in cosmological constraints is heavily affected by the statistical incompleteness due to having a small-sky survey.

masked region  $\text{RA} = [0, 30] \text{ deg}$  and  $\text{Dec.} = +[0, 10] \text{ deg}$ , for the same redshift bins (see Table 1). We choose the same  $\ell$  bins as before up to  $\ell < 375$  with  $\theta_{\text{beam}} = 1 \text{ deg}$  convolution, and calculate the cut-sky covariance matrix using NAMASTER. Fig. 13 shows the cosmological constraints feasible of this scenario. Comparing the results to the full-sky case (Fig. 11), we can see that the statistical incompleteness due to having a cut-sky survey reduces the feasibility; with  $300 \text{ deg}^2$  sky, we cannot detect the HIHI cosmic signal. However, incorporating the cross-correlation with weak lensing improves the significance of the joint statistics (Fig. 13).

## 5 INSTRUMENT NOISE

In this section we consider the instrument noise for both lensing and HI surveys for the current state-of-the-art case. We will consider the expected thermal noise for a single-dish survey for HI measurement.

### 5.1 Single dish thermal noise

We begin by examining the noise on 3D measured power spectra for the HI autocorrelation ( $P$ ). This discussion is based on the works of Battye et al. (2013), Bigot-Sazy et al. (2015), and Santos et al. (2015). Subsequently, we derive the root mean square (rms) thermal noise expected on IM maps from the power spectra, which we will use to assess the effect of realistic noise on the ability to detect the lensing-HI cross-correlation.

The expected uncertainty ( $\sigma_p$ ) on the power spectrum  $P$  can be estimated by evaluating its expected second moment. By calculating the ratio between  $\sigma_p$  and  $P$  averaging over the radial wavenumber bin size  $\Delta k$ , we can estimate the uncertainty in the IM map measurements. Following this procedure the error on  $P$  can be estimated by the following expression (Feldman, Kaiser & Peacock

1994; Seo et al. 2010; Battye et al. 2013):

$$\frac{\sigma_p}{P} = 2\sqrt{\left(\frac{2\pi}{3}\right)V_{\text{sur}}\frac{1}{4\pi k^2 \Delta k}} \left(1 + \frac{\sigma_{\text{pix}}^2 V_{\text{pix}}}{[\bar{T}(z)]^2 W(k)^2 P}\right), \quad (33)$$

where  $V_{\text{sur}}$  is the volume of survey,  $W(k)$  is windows function,  $\bar{T}(z)$  is the average temperature at given redshift  $z$ , and  $V_{\text{pix}}$  is the pixel volume. In this work we shall ignore the contribution of shot noise and only consider the contribution from pixel (thermal) noise  $\sigma_{\text{pix}}$ . These parameters are discussed in detail in the following paragraphs.

It is essential to pick the proper size of  $\Delta k$  to optimize the viability of a single dish. As we focus on the cosmic signal the acoustic scale should be an aim. That means we require  $\Delta k/k_A < 1$ , where  $k_A$  is the wavenumber of acoustic scale. The volume of the survey  $V_{\text{sur}}$  can be computed by

$$V_{\text{sur}} = \Omega_{\text{sur}} \int_{z_{\text{min}}}^{z_{\text{max}}} dz \frac{dV}{dz d\Omega}, \quad (34)$$

where

$$\frac{dV}{dz d\Omega} = \frac{cr^2(z)}{H_0 E(z)}, \quad (35)$$

where we assume a flat universe. The window function  $W(k)$  is set by the instrument's specification. As we map the  $\delta_{\text{HI}}$  from multiple redshift bins, we can ignore the contribution of radial directions in  $W(k)$ . However for the angular direction, this is not the case. Importantly, the angular resolution of the radio beam will define the noise level. We can model  $W(k)$  by

$$W(k) = \exp\left[-\frac{1}{2}k^2 r^2(z) \frac{\theta_{\text{FWHM}}}{8 \ln(2)}\right], \quad (36)$$

where  $\theta_{\text{FWHM}}$  is the full width at half-maximum of angular resolution. Realistically,  $\theta_{\text{FWHM}}$  depends on frequency ( $\nu$ ). However, in this analysis we assume the variation of  $\theta_{\text{FWHM}}$  is small and negligible.

The pixel volume,  $V_{\text{pix}}$ , can be determined by

$$V_{\text{pix}} = \Omega_{\text{pix}} \int_{z_c - \Delta z}^{z_c + \Delta z} dz \frac{dV}{dz d\Omega}, \quad (37)$$

where  $z_c$  is the central redshift and  $\Delta z$  corresponds to frequency width ( $\Delta\nu$ ) and  $\Omega_{\text{pix}}$  is the pixel solid angle.

In radio astronomy we normally measure the signal in terms of power. The antenna temperature then generates the thermal noise; the pixel noise  $\sigma_{\text{pix}}$  can be approximated by (Seo et al. 2010; Santos et al. 2015),

$$\sigma_{\text{pix}} \approx \frac{T_{\text{sys}}}{\varepsilon \sqrt{t_p 2 \Delta \nu}}, \quad (38)$$

where  $t_p$  represents the observation time per pointing,  $\varepsilon$  (approximately 1) signifies the efficiency of the telescope, meaning that almost no signal is lost when radio radiation is transmitted to the antenna.  $T_{\text{sys}}$  represents the system temperature, which includes

$$T_{\text{sys}} = T_{\text{rx}} + T_{\text{spl}} + T_{\text{CMB}} + T_{\text{gal}}, \quad (39)$$

where we ignore the contribution from the Earth's atmosphere.  $T_{\text{spl}}$  is the spill over from ground radiation (approximately 3 K),  $T_{\text{CMB}} \approx 2.73 \text{ K}$  and galactic temperature  $T_{\text{gal}} \approx 25 \text{ K}$  ( $408 \text{ MHz}/\nu$ )<sup>2.75K</sup> (Square Kilometre Array Cosmology Science Working Group et al. 2020). The observing time per pointing  $t_p$  relates to the total observation by  $t_p = t_{\text{tot}}(\theta_B)^2 / \Omega_{\text{sur}}$ , where  $\theta_B$  is an angular pixel size.

As  $\sigma_{\text{pix}}$  is the rms thermal noise, by definition its square is the power per pixel volume ( $P_N / V_{\text{pix}}$ ). Therefore, the 3D noise power

**Table 3.** MeerKAT pilot survey specifications (Wang et al. 2021).

$\Delta\nu$	0.2 MHz
$N_{\Delta\nu}$	[200,250]
$T_{\text{rx}}$	$7.5 \times 10^3 + 10^3(\nu[\text{MHz}]/1000 - 0.75)^2$ [mK]
$t_{\text{tot}}$	10.5 h
$z$	[0.3885, 0.4623]
$N_{\text{dish}}$	64
$T_{\text{sys}}$	$16 \times 10^3$ mK
$N_{\text{pix}}$	87500
$\theta_{\text{FWHM}}$	1.48 deg
$\Omega_{\text{sur}}$	200 deg $^2$

spectrum ( $P_N$ ) is then (Battye et al. 2013; Santos et al. 2015)

$$P_N = \sigma_{\text{pix}}^2 V_{\text{pix}} = r^2 y \frac{T_{\text{sys}}^2 \Omega_{\text{sur}}}{2\varepsilon^2 t_{\text{tot}}}, \quad (40)$$

where

$$y = cH(z)^{-1} \frac{(1+z)^2}{\nu_{21}}. \quad (41)$$

If we have  $N_d$  dishes where each dish has  $N_b$  beams, we can take less time for each pointing area. The noise power spectrum is then reduced to

$$P_N(N_d, N_b) = \frac{r^2 y T_{\text{sys}}^2 \Omega_{\text{sur}}}{2\varepsilon^2 t_{\text{tot}} N_b N_d}. \quad (42)$$

To determine the optimization of a survey strategy, we can estimate the suitable  $\theta_{\text{FWHM}}$  and  $\Omega_{\text{sur}}$  that minimizes  $\delta k_A/k_A$  for acoustic scale  $k_A$ . This acoustic scale  $k_A$  can be estimated by following the work of Blake & Glazebrook (2003) and Battye et al. (2013),

$$\frac{P(k)}{P_{\text{ref}}} = 1 + Ak \exp \left[ - \left( \frac{k}{0.1h\text{Mpc}^{-1}} \right)^{1.4} \right] \sin(2\pi k/k_A), \quad (43)$$

where  $A$  is the overall amplitude which can be marginalized. The subscript ‘ref’ refers to reference cosmological parameters.

We now calculate the thermal noise of a MeerKAT-like instrument  $\sigma_{\text{pix}}^{\text{MK}} = \sigma_T$ . In this case we consider a single dish telescope consisting of 64 dishes of 13.5m diameter, operating in UHF, L, and S band. The MeerKAT pilot survey by Wang et al. (2021) focuses on L band from 856 to 1712 MHz with 4096 frequency channels. This pilot survey has 10.5 h observation time with approximately  $\sim 200$  deg $^2$  observation field (Wang et al. 2021; Cunningham et al. 2023a). The summary statistics of this MeerKAT pilot survey are listed in Table 3. Wang et al. (2021) show that for integrated frequency channels, they can achieve thermal noise  $\sigma_T \approx 2$  mK.

If we use equations ((38)) and ((39)) together with Table 3, the expected  $\sigma_{\text{pix}}$  for a single frequency channel of MeerKAT pilot survey is

$$\sigma_{\text{pix}}(\Delta\nu = 0.2; 10\text{hr}) \approx 15 \text{ mK}, \quad (44)$$

where we assume each dish has equal efficiency  $\varepsilon = 1$  and consider only a single frequency channel  $\Delta\nu = 0.2$  MHz. If we consider the whole frequency range like (Wang et al. 2021) the  $\sigma_T \approx 2$  mK.

We now consider the case where the total observation time  $t_{\text{tot}} = 1000$  h and 250 frequency channels with  $\Delta\nu = 0.2$  MHz.

## 5.2 S/N of $\kappa$ HI

In Section 5.1 we have estimated the rms thermal noise for MeerKAT-like surveys similar to the current state of the art. In this section we explore the future case, where the observation time  $t_{\text{tot}}$  can take longer than MeerKAT’s pilot survey, and we assume a full-sky survey to

estimate the best possible S/N for weak lensing-intensity mapping ( $\kappa$ HI) two-point statistics. The estimate  $\sigma_{\text{pix}}^T$  from equation (44) is 15mK for one frequency channel, which is based on the specification of the current MeerKAT survey in Table 3 and equation (38); to detect the cross-correlation we should find ways to reduce the  $\sigma_{\text{pix}}^T$  as far as possible.

We first model the S/N of  $\kappa$ HI. We can consider the zero lag noise level for  $\langle \kappa \Delta T \rangle$ , i.e. where  $\kappa$  and  $\Delta T$  are measured in the same pixel. There is no reason why the statistical noise of  $\kappa$  should be correlated with  $\Delta T$ . The noise for the cross-correlation will therefore be proportional to the product of  $\sigma_T$  and  $\sigma_e^n$ , where  $\sigma_T$  is HI thermal noise which can be estimated by equation (44). The rms noise for weak lensing  $\sigma_e^n$  can be estimated by

$$\sigma_e^n = \frac{\sigma_e}{\sqrt{n_{\text{gal}}}}, \quad (45)$$

where  $\sigma_e$  is the variance of intrinsic galaxy ellipticities and  $n_{\text{gal}}$  is galaxy number per pixel. For KiDS and DES-like surveys,  $\sigma_e \approx 0.3$ . The KiDS DR4 effective galaxy number density is  $n_{\text{eff}} = 0.325$  arcmin $^{-2}$  for the whole redshift range (Giblin et al. 2021; Heymans et al. 2021). For a pixel size 0.25 $^2$  deg $^2$ , we find  $\sigma_e^n \approx 0.03$ .

The S/N of  $\langle \kappa \text{HI} \rangle$  then can be estimated by

$$S/N = \frac{\sigma_{\text{HI}} \sigma_{\kappa}}{\sigma_T \sigma_e^n} \sqrt{N_{\text{pix}}}, \quad (46)$$

and we find that the rms lensing convergence signal  $\sigma_{\kappa}$  is similar to the rms noise  $\sigma_e^n$  on 0.25 deg scales (Gatti et al. 2021; Amon et al. 2022).

To estimate the signal of HI intensity mapping, we first recall that the HI brightness temperature fluctuations  $\delta T_{\text{HI}}$  can be expressed by

$$\delta T_{\text{HI}} = \bar{T}_{\text{HI}}(z) b_{\text{HI}}(z) \delta_m(\theta, z), \quad (47)$$

where  $\bar{T}_{\text{HI}}(z)$  is an average temperature over angular position ( $\theta$ ) for a given  $z$  and  $b_{\text{HI}}(z)$  is HI bias. As the power spectrum  $P_{\text{HI}}$  is the power of the temperature fluctuation  $\delta T_{\text{HI}}$ , the square root of  $P_{\text{HI}}$  per volume  $V_{\text{sur}}$  is effectively the root mean square of the HI true signal ( $\sigma_{\text{HI}}$ ),

$$\sigma_{\text{HI}} = \sqrt{P_{\text{HI}}/V_{\text{sur}}}. \quad (48)$$

We consider a survey similar to the MeerKAT pilot survey ( $0.39 < z < 0.46$ ) over a moderately thick  $\Delta z = 0.075$  redshift bin, with 0.25 $^2$  deg $^2$  pixel size,  $V_{\text{sur}} \sim 4000 \text{ Mpc}^3 h^{-3}$ . Assuming the foregrounds and redshift-space distortions have been appropriately dealt with, we can use

$$P_{\text{HI}} = \bar{T}_{\text{HI}}^2 b_{\text{HI}}^2 P_m \quad (49)$$

We assume an effective redshift of the survey  $z_{\text{eff}} = 0.42$ ,  $b_{\text{HI}} = 1$ , and  $\bar{T}_{\text{HI}} = 0.07$  mK (Santos et al. 2015; Wang et al. 2021; Cunningham et al. 2023a). In this case, the estimation of HI rms is then

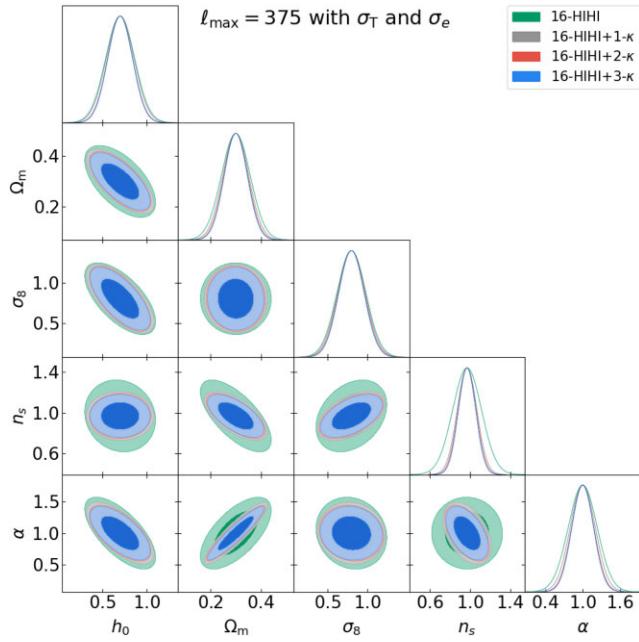
$$\sigma_{\text{HI}}(k = 0.1, z = 0.42) = 5 \mu\text{K}, \quad (50)$$

where we estimate at the  $k = 0.1 h^{-1}\text{Mpc}$  scale. This estimation generally agrees with table I in Santos et al. (2015) with slightly better signal rms, because Santos et al. (2015) use smaller channel bins than the thick redshift bin we have here.

$\sigma_T$  is approximately 15 mK for a single frequency channel given Table 3. If we stack over 200  $\Delta\nu$  channels and assume 10 h observing time, then

$$\sigma_T \approx 1.1 \text{ mK}. \quad (51)$$

Then using equation ((46)), the estimation of S/N for  $\kappa$ HI two-point statistics for KiDS-like lensing surveys and MeerKAT for  $k = 0.1$



**Figure 14.** The likelihood contours for our multibin analysis with HI bias model 1 with  $\ell_{\max} = 375$  for full-sky case including the instrument noise contributions; the contours show 68 per cent and 95 per cent confidence levels. Comparing to Fig. 11 where we ignore instrument noise contributions, we can see that  $h_0$  is now poorly constrained.

$h^{-1}$  Mpc with a pixel size  $0.25^2$  deg $^2$  covering  $\sim 200$  deg $^2$  sky, is then

$$S/N \approx 0.24. \quad (52)$$

This means that by the current state of the art, we expected to observe more instrument noise than cosmic signal from  $\kappa$ HI cross-correlations at zero lag. By including cross-correlations at different angular separations, we expect a higher total signal-to-noise ratio.

Note that this estimation is based on the pilot survey by MeerKAT which only contains 64 dishes of 13m diameter and only observes for 10 h; for the full operation of MeerKAT or SKA-Mid, we will have more dishes, longer observation time, and more frequency channels. If we assume a longer observation time such as 1000 h (as is recommended by Zhang et al. 2023) and increase the number of frequency channels to 250, the estimation of  $\sigma_{\text{MK}}$  is then

$$\sigma_{\text{MK}}^{1000} \approx 0.01 \text{ mK}. \quad (53)$$

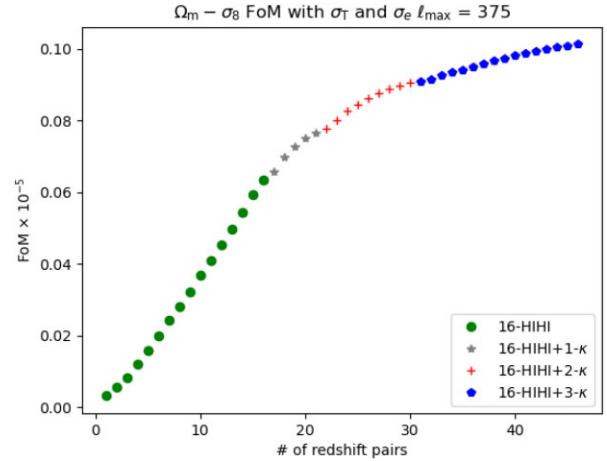
Now we have better S/N by one order of magnitude for a 3000 pixel sky. [If we use the pilot survey footprint ( $\sim 200$  deg $^2$ ), then S/N  $\approx 2.4$  for zero lag.]

We confirm this calculation by generating Gaussian random fields for the instrument noise for both  $\kappa$  and HI fields and add these noise maps using (equations 44 and 45) to the simulations described in Section 4. We utilize the full sky maps with NSIDE=128. The pixel number for this resolution is  $N_{\text{pix}} = 196\,608$ . Hence the estimation of S/N for this configuration is then

$$S/N_{\text{full}}^{1000} \approx 22. \quad (54)$$

However, when we take the LoS foreground subtraction into account, the signal would be reduced by a factor of  $\sim 3$ . This means

$$S/N_{\text{full}}^{1000} \rightarrow \approx 7. \quad (55)$$



**Figure 15.** FoM for  $\sigma_8 - \Omega_m$  constraints including instrument noise; the horizontal axis is the number of redshift bin pairs for cosmological constraints. We show cumulative FoM when including increasing numbers of HI autocorrelation redshift bins (green); then increasing numbers of cross-correlations with convergence bins (grey, red, blue). Here  $\ell_{\max} = 375$ . Comparing this figure to Fig. 9, we find that the FoM is lower by one order of magnitude.

**Table 4.** Cosmological forecast from Fisher analysis including instrument noise contributions for 16-HIHI + 3- $\kappa$  correlations; the uncertainties on cosmological parameters and HI bias are quoted at 95 per cent confident level. Here we consider only HI bias model 1.

Parameters	HI bias model 1
$\Delta h_0$	$\pm 0.28$
$\Delta \Omega_m$	$\pm 0.09$
$\Delta \sigma_8$	$\pm 0.32$
$\Delta n_s$	$\pm 0.17$
$\Delta \alpha$	$\pm 0.34$

This estimation indicates that we would expect to observe the true signal at zero lag for 1000 h exposure and large sky surveys such as SKA.

### 5.3 Fisher analysis

We now consider the feasibility of  $\kappa$ HI and HIHI two-point statistics for cosmological constraints in the presence of these current noise levels. In the S/N analysis of  $\kappa$ HI (Section 5.2), we only considered the case where  $\Delta\nu_{\text{total}} = 250 \times 0.2$  MHz. However to compare results in Section 4.3, we will adjust  $\Delta\nu$  to match  $z_{\text{HI}}$  in Table 1.

We generate Gaussian white noise fields for both  $\kappa$  and HI using the previous subsection’s calculated amplitudes. We note again that in this analysis, we consider only the full-sky case. The marginalizations of cosmological parameters are illustrated by Fig. 14. Comparing this result to the no-noise case (see Fig. 11) we can see that the cosmological feasibility of HIHI and  $\kappa$ HI are reduced significantly. We also show the resulting FoM in Fig. 15 and 2 $\sigma$ -constraints in Table 4. We note that the  $\Omega_m$  constraint shows a precise estimation, although it is degenerate with  $\alpha$ . Considering Fig. 15, we can see that the maximum of FoM when considering instrument noise is one order of magnitude less than the no-noise case (see Fig. 8). Nevertheless, all FoM plots (see Figs 8, 9, and 15) indicate that by combining  $\kappa$ HI likelihoods with HIHI, we can significantly enhance the feasibility of cosmological constraints from the HI intensity maps.

## 6 CONCLUSIONS

In this paper we have studied the two-point statistics of lensing convergence and HI intensity mapping.

We first presented the theoretical framework for calculating convergence-intensity mapping cross-correlations. Next, using realizations from an  $N$ -body simulation we have emulated HI intensity maps, and have shown that their cross-correlation with convergence maps from these simulations agree with our theoretical cross-correlation calculations.

We proceeded to study the effect of HI foreground removal on the two-point functions. We model the effect of foreground removal by removing the mean along each line of sight, which effectively represents the largest radial mode, from our HI maps, following the method of Cunningham et al. (2019b). We then measure the post-removal cross-power; we find that the foreground removal modestly reduces the  $\kappa$ HI power spectrum signal, by a factor  $A_{\text{clean}}(z_{\text{HI}}, z_{\kappa}, z_{\text{max}})$ . In the case of forthcoming HI experiments that will measure HI at  $z_{\text{max}} < 1$ ,  $A_{\text{clean}}(z_{\text{HI}} < 0.5, z_{\kappa} = 0.78, z_{\text{max}} = 1)$  is approximately 2.5 for our catalogues.

In the following section, we utilized the Fisher matrix formalism to forecast best-case cosmological constraints for the convergence-HI probe, for the maximal case of full sky and subdominant telescope noise, but while including foreground removal. We calculated the Fisher matrix for  $\kappa\kappa$ ,  $\text{HHI}$ , and  $\kappa\text{HI}$  two-point functions using the measured covariance matrices from Section 3.

We find that a single redshift slice of the HI intensity map and  $\kappa$  can constrain cosmological parameters for known bias (see Fig. 7), but when  $b_{\text{HI}}$  is a further parameter (or several), the few-slice  $3 \times 2$  point functions do not sufficiently constrain the cosmological parameters compared to current cosmological surveys such as Planck and DES (Abbott et al. 2019; Planck Collaboration I 2020).

Hence, several cross-bin correlations are required in order for this probe to be of interest. In Section 4.3, we have explored the use of several redshift bins for HI and convergence, together with the effect of  $b_{\text{HI}}$  on cosmological constraints. We consider both the current state of art where  $\ell_{\text{max}} < 400$ , and the futuristic case where  $\ell_{\text{max}} > 1000$ . Both cases show that a set of two-point functions constrains the uncertainty in cosmological parameters to a comparable level with current experiments. All FoMs show that by including the cross-correlation of a lensing survey with the 21cm signal, we can improve the HI auto constraints.

We then examined the impact of a cut-sky survey. In this analysis, we evaluated the two-point statistics based on a  $300 \text{ deg}^2$  observed patch of sky. Due to the statistical incompleteness, detecting cosmic signals becomes marginal in this context.

In Section 5.1, we explore the instrument noise affecting lensing and HI galaxy surveys. The thermal noise of a single-dish survey was calculated. In this study, we focus on instruments similar to MeerKAT for radio observations and KIDs-like surveys for optical counterparts. Our analysis demonstrates the feasibility of detecting the  $\kappa$ HI cross-correlation, provided we have sufficient sky coverage and long exposure times for the radio measurements.

Even though we have shown positive results for two-point statistics between the  $\kappa$  field and HI intensity map, there are important caveats that remain to be explored further:

(i) In this study we have created HI intensity maps based on the assumption that they are linearly biased in relation to overdensity  $\delta_m$ . A more realistic study should construct HI intensity maps by assigning HI mass  $M_{\text{HI}}$  to simulated halo catalogues. Also in future work the generated HI maps will be compared in detail to real data.

(ii) We have approximated the foreground cleaning by removing the mean fluctuation along the line of sight, which effectively represents the largest radial mode. More detailed emulation of the foreground cleaning will be studied in future.

(iii) The cosmological and  $b_{\text{HI}}$  constraint predictions have been obtained using the idealized Fisher matrix analysis; for real data, Markov Chain Monte Carlo methods are required to deal with non-Gaussian likelihoods and realistic degeneracies between parameters.

(iv) Only the  $\Lambda$ CDM model is considered in these parameter constraints. Further extensions (e.g. wCDM) should be considered in further work.

(v) Only two-point statistics have been explored in this research. However the study by Schmit, Heavens & Pritchard (2019) shows that combining the bispectrum and power spectrum can reduce the error of cosmological parameters by an order of magnitude compared to Planck.

(vi) This study has only explored the low-redshift two-point functions. The high-redshift probes at the time of the Epoch of Reionization and the CMB are not taken into account. It is an interesting matter for future work to consider whether two-point statistics between the EOR HI and CMB weak lensing can also be measured.

In conclusion,  $\kappa$ -HI cross-correlations are an intriguing additional probe for cosmology, which are not destroyed by foreground removal. This probe will be available for measurement with forthcoming HI and lensing surveys this decade.

## ACKNOWLEDGEMENTS

We acknowledge studentship support from the Thai Royal Government for AS. We thank Tzu-Ching Chang (JPL), Yu-Wei Liao (ASIAA), Steve Cunningham (University of Manchester), and Utane Sawangwit (National Astronomical Research Institute of Thailand) for extremely helpful discussions. DB was supported by STFC Consolidator Grant ST/S000550/1. Calculations were made using Chanlawan High Performance Computer at the National Astronomical Research Institute of Thailand.

## DATA AVAILABILITY

In this work we use the  $N$ -body simulations and weak lensing catalogues from Takahashi et al. (2017). The catalogues can be obtained at [http://cosmo.phys.hirosaki-u.ac.jp/takahashi/allsky\\_raying](http://cosmo.phys.hirosaki-u.ac.jp/takahashi/allsky_raying). The measured two-point functions and Python codes for calculations in Sections 3 and 4 are available in Github, at <https://github.com/AnutUoP/2020s-KappaHI>. Other data will be shared on reasonable requests to the corresponding author.

## REFERENCES

- Abbott T. M. C. et al., 2018, *Phys. Rev. D*, 98, 043526
- Abbott T. M. C. et al., 2019, *ApJ*, 872, L30
- Alam S. et al., 2017, *MNRAS*, 470, 2617
- Alonso D., Sanchez J., Slosar A., LSST Dark Energy Science Collaboration, 2019, *MNRAS*, 484, 4127
- Amon A. et al., 2022, *Phys. Rev. D*, 105, 023514
- Bartelmann M., Schneider P., 2001, *Phys. Rep.*, 340, 291
- Battye R. A. et al., 2013, *MNRAS*, 434, 1239
- Baxter E. J. et al., 2019, *Phys. Rev. D* 99, 023508
- Bigot-Sazy M.-A. et al., 2015, *MNRAS*, 454, 3240
- Blake C., Glazebrook K., 2003, *ApJ*, 594, 665
- Brown M. L., Castro P. G., Taylor A. N., 2005, *MNRAS*, 360, 1262

Bull P., Ferreira P. G., Patel P., Santos M. G., 2015, *ApJ*, 803, 21

CHIME Collaboration et al., 2023, *ApJ*, 947, 16

Castro P. G., Heavens A. F., Kitching T. D., 2005, *Phys. Rev. D*, 72, 023516

Chapman E. et al., 2012, *MNRAS*, 423, 2518

Cunnington S., Harrison I., Poursidou A., Bacon D., 2019a, *MNRAS*, 482, 3341

Cunnington S. et al., 2019b, *MNRAS*, 488, 5452

Cunnington S., Poursidou A., Soares P. S., Blake C., Bacon D., 2020, *MNRAS*, 496, 415

Cunnington S. et al., 2023a, *MNRAS*, 518, 6262

Cunnington S. et al., 2023b, *MNRAS*, 523, 2453

Euclid Collaboration, 2020, *A&A*, 642, A191

Fang X., Eifler T., Schaan E., Huang H.-J., Krause E., Ferraro S., 2022, *MNRAS*, 509, 5721

Feldman H. A., Kaiser N., Peacock J. A., 1994, *ApJ*, 426, 23

Fernandez-Soto A., Lanzetta K. M., Chen H.-W., Pascarelle S. M., Yahata N., 2001, *ApJS*, 135, 41

Gatti M. et al., 2021, *MNRAS*, 504, 4312

Giblin B. et al., 2021, *A&A*, 645, A105

Górski K. M. et al., 2005, *ApJ*, 622, 759

Harker G. et al., 2010, *MNRAS*, 405, 2492

Heavens A., 2003, *MNRAS*, 343, 1327

Heymans C. et al., 2021, *A&A*, 646, A140

Hinshaw G. et al., 2013, *ApJS*, 208, 19

Hivon E., Górski K. M., Netterfield C. B., Crill B. P., Prunet S., Hansen F., 2002, *ApJ*, 567, 2

Hu W., Jain B., 2004, *Phys. Rev.*, D70, 043009

Kamionkowski M., Smith T. L., Heavens A., 2011, *Phys. Rev. D*, 83, 023007

Kim J., 2011, *A&A*, 531, A32

Kim J., Naselsky P., 2010, *A&A*, 519, A104

Lewis A., Bridle S., 2002, *Phys. Rev. D*, 66, 103511

LoVerde M., Afshordi N., 2008, *Phys. Rev. D*, 78, 123506

Mao Y. et al., 2008, *Phys. Rev. D*, 78, 023529

Martin A. M., Giovanelli R., Haynes M. P., Guzzo L., 2012, *ApJ*, 750, 38

Mendez R. A., Silva J. F., Orostica R., Lobos R., 2014, *PASP*, 126, 798

Padmanabhan H., Refregier A., Amara A., 2020, *MNRAS*, 495, 3935

Pandey S., et al., DES Collaboration, 2022, *Phys. Rev. D*, 106, 043520

Planck Collaboration I, 2020, *A&A*, 641, A1

Platania P., Bensadoun M., Bersanelli M., De Amici G., Kogut A., Levin S., Maino D., Smoot G. F., 1998, *ApJ*, 505, 473

Poursidou A., 2017, preprint ([arXiv:1709.07316](https://arxiv.org/abs/1709.07316))

Poursidou A., Bacon D., Crittenden R., 2017, *MNRAS*, 470, 4251

Pratten G., Munshi D., Valageas P., Brax P., 2016, *Phys. Rev. D*, 93, 103524

Santos M. G., Ferramacho L., Silva M. B., Amblard A., Cooray A., 2010, *MNRAS*, 406, 2421

Santos M. et al., 2015, Proc. Sci., Cosmology from a SKA HI Intensity Mapping Survey. SISSA, Trieste, PoS#19

Schmit C. J., Heavens A. F., Pritchard J. R., 2019, *MNRAS*, 483, 4259

Seo H.-J., Dodelson S., Marriner J., McGinnis D., Stebbins A., Stoughton C., Vallinotto A., 2010, *ApJ*, 721, 164

Shaw J. R., Sigurdson K., Pen U.-L., Stebbins A., Sitwell M., 2014, *ApJ*, 781, 57

Smoot G. F., Debono I., 2017, *A&A*, 597, A136

Spinelli M., Bernardi G., Santos M. G., 2018, *MNRAS*, 479, 275

Spinelli M., Carucci I. P., Cunningham S., Harper S. E., Irfan M. O., Fonseca J., Poursidou A., Wolz L., 2022, *MNRAS*, 509, 2048

Square Kilometre Array Cosmology Science Working Group et al., 2020, *Publ. Astron. Soc. Aust.*, 37, e007

Su H. et al., 2018, *MNRAS*, 479, 4041

Switzer E. R. et al., 2013, *MNRAS*, 434, L46

Takahashi R. et al., 2012, *ApJ*, 761, 152

Takahashi R. et al., 2017, *ApJ*, 850, 24

Tegmark M., Taylor A. N., Heavens A. F., 1997, *ApJ*, 480, 22

CHIME Collaboration et al., 2022, *ApJS*, 261, 29

Tröster T. et al., 2022, *A&A*, 660, A27

Upham R. E., Whittaker L., Brown M. L., 2019, *MNRAS*, 491, 3165

Wang J. et al., 2021, *MNRAS*, 505, 3698

Wolz L., Tonini C., Blake C., Wyithe J. S. B., 2016, *MNRAS*, 458, 3399

Wuensche C. A., the BINGO Collaboration, 2019, *J. Phys. Conf. Ser.*, 1269, 012002

Zhang M., Li Y., Zhang J.-F., Zhang X., 2023, *MNRAS*, 524, 2420

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author