Forecasting the Cross-Correlation of the CSST galaxy survey with the FAST HI Intensity Map

FUREN DENG, 1,2 YAN GONG, 1,2 YOUGANG WANG, 1,2 SHUTONG DONG, 3 YE CAO, 4 AND XUELEI CHEN^{1,2,5,6}

National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China
 School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China
 School of physics, Nankai University, Tianjin, 300071, China
 Department of Astronomy, Beijing Normal University, Beijing 100875, China
 Department of Physics, College of Sciences, Northeastern University, Shenyang 110819, China
 Center of High Energy Physics, Peking University, Beijing 100871, China

ABSTRACT

The cross-correlation of optical galaxies with the neutral hydrogen (HI) radiation intensity can enhance the signal-to-noise ratio (SNR) of the HI intensity measurement. In this paper, we investigate the cross-correlation of the galaxy samples obtained by the spectroscopic survey of the China Space Station Telescope (CSST) with the HI Intensity mapping (IM) survey of the Five-hundred-meter Aperture Spherical Telescope (FAST). Using the IllusitrisTNG simulation result at redshift $0.2 \sim 0.3$, we generate mock data of the CSST survey and a FAST L-band drift scan survey. The CSST spectroscopic survey can yield a sample of galaxies with a high comoving number density of 10^{-2} (Mpc/h)⁻³ at $z \sim 0.3$. We cross-correlate the foreground-removed radio intensity with the CSST galaxies, including both the whole sample, and red and blue galaxy sub-samples separately. We find that in all cases the HI and optical galaxies are well correlated. The total HI abundance can be measured with a high precision from this correlation. A relative error of $\sim 0.6\%$ for $\Omega_{\rm HI}$ could be achieved at $z \sim 0.3$ for an overlapping survey area of 10000 deg².

Keywords: cosmology:large-scale structure of Universe, radio lines: galaxies, telescopes

1. INTRODUCTION

Neutral hydrogen plays a key role in galaxy formation and evolution, as the fuel for the formation of denser structures such as molecular clouds and stars, and it is also a tracer of the large scale structure of the Universe. The redshifted 21 cm signal produced by neutral hydrogen (HI) contains rich cosmological information, which can be used for testing cosmological models and constraining cosmological parameters. Instead of observing individual galaxies one-by-one, HI intensity mapping is an efficient technique to perform fast surveys over large cosmic volumes in wide redshift range, by detecting integrated signal (Chang et al. 2008). A number of intensity mapping experiments are underway, such as the Tianlai (Chen 2012; Xu et al. 2015; Li et al. 2020; Wu et al. 2021), CHIME (Bandura et al. 2014; Newburgh et al. 2014; Amiri et al. 2022) and HIRAX (Newburgh et al. 2016), BINGO (Battye et al. 2012, 2016) experiments. This technique may also be applied to the new generation of general purpose radio telescopes, such as the MeerKAT (Santos et al. 2017; Wang et al. 2021) and the Square Kilometer Array (SKA) (Santos et al. 2015).

The Five-hundred-meter Aperture Spherical Telescope (FAST) (Nan et al. 2011) is a single dish telescope with currently the largest aperture (300 m during operation) and is equipped with multi-beam feed system (19 beams presently for L band) and low-noise cryogenic receivers, making it very suitable for large HI intensity mapping surveys(Bigot-Sazy et al. 2016; Hu et al. 2020a).

However, in intensity mapping observations a major obstacle is the foreground contamination. The Galactic synchrotron and free-free emission are at least three to four orders of magnitude brighter than the 21 cm signal at the L-band. In principle, the foreground can be removed using various algorithms, for example, log-polynomial fitting, PCA and ICA (for a review see e.g. Asorey et al. 2020). However, signal loss and foreground residual are inevitable, and additionally there are also calibration errors, radio frequency interference (RFI) and receiver noise that affect the observation. To-date, intensity mapping signal has not yet been positively detected in auto-correlations.

One method to reduce the contamination is to correlate the radio signal with the optical galaxy catalog. As the foreground is mostly Galactic and not correlated with the 21cm signal, such correlation will enhance the signal-to-noise ratio (SNR), making such cross-correlation easier to detect than the auto-correlations.

Indeed, positive detection of the cross-correlation has been achieved by correlating the Green Bank Telescope (GBT) data with the DEEP2 survey (Chang et al. 2010), Wiggle Z Dark Energy survey (Masui et al. 2013). More recently, the HI IM survey data from the Parkes telescope data is correlated with 2dF survey and a detection at $z \approx 0.1$ is reported (Anderson et al. 2018), though the cross-correlation is lower than expected. The GBT HI IM survey data is also correlated with the eBOSS ELG and the eBOSS LRG to constrain HI abundance and bias (Wolz et al. 2022). Some GBT quantitative results at redshift $z \approx 0.8$ are summarized in Table. 1. More recently, the Canadian Hydrogen Intensity Mapping Experiment (CHIME) detected HI cross-correlation with the luminous red galaxies (LRG) at z = 0.84, emission line galaxies (ELG) at z = 0.96, and quasars (QSO) from the eBOSS clustering catalogs at z = 1.20 by stacking the map on the angular and spectral locations of optic galaxies, though the errors in the amplitudes of correlation are still relatively large (CHIME Collaboration et al. 2022).

Besides the technical advantage, the cross-correlation of optical galaxies and HI intensity also offers useful information for science: how the neutral hydrogen and stellar mass are correlated. It is believed that on large scales both will trace the total density which is dominated by the dark matter, but on small scales the relation could be complicated, depending on a number of galaxy evolution parameters. The neutral hydrogen is linked with the current star formation rate, while the luminosity of the galaxy is determined to a large extent by the stellar mass, which has been accumulated over the entire cosmic history. By statistical analysis of the correlation between the HI and different types of optical galaxies, one may gain further insights on the formation and evolution of galaxies (Wolz et al. 2019).

The Chinese Space Station Telescope (CSST) has a planned Optical Survey (CSS-OS) (Zhan 2011; Gong et al. 2019), which includes both multi-band photometric survey and low resolution spectroscopic survey. The CSS-OS is planned to cover a total sky area of 17500 deg². It will provide a large set of galaxies over a broad redshift range, suitable for correlation with intensity mapping data. In the present paper, we investi-

gate the potential of cross-correlation between the FAST HI intensity mapping survey with the CSS-OS spectroscopic survey galaxies.

In principle, the HI auto-correlation power spectrum and the HI-optical galaxy number density crosscorrelation power spectrum are proportional to the matter density power spectrum, and can be used to probe many cosmological parameters. For example, the baryon acoustic oscillation (BAO) wiggles can be used to constrain the cosmological distance scale and Hubble expansion parameters. However, at present most studies of HI-optical galaxy cross-correlation have focused on the measurement of total neutral hydrogen density $\Omega_{\rm HI}$ at the redshift of the survey. This is because at present, foreground removal is still a very great challenge in any cosmological 21 cm observation, the residue error is usually far greater than the statistical error expected for the power spectrum. It is difficult to give an accurate estimate of the uncertainty in the HI autocorrelation power spectrum at the present time, even if we succeed in detecting it in the future survey. As for the HI-optical cross-correlation power spectrum, we are more confident of its successful measurement, which has been demonstrated by the GBT, Parkes and CHIME experiments, but by definition we would already have the optical galaxy measurement in place, which would provide a good constraint on the general cosmological parameters, it is not expected that the HI-optical crosscorrelation power could give much improvement over the optical measurement. The total neutral hydrogen density $\Omega_{\rm HI}$ can however be measured directly from the HI cross power spectrum, and it is not available from the optical observations, at least not directly. While $\Omega_{\rm HI}$ is not a fundamental parameter in the standard cosmological model, it is a physical parameter of the low redshift Universe and is of great interest to the galaxy evolution model. We shall also focus on this parameter in the present paper.

The structure of this paper is as follows. In Section 2, we discuss how to generate mock observational data from the simulation. The construction of HI catalog and optical galaxy catalog are introduced in Sec. 2.1 and Sec. 2.2, the simulation of intensity mapping is described in Sec. 2.3, and the foreground removal and transfer function calculation for the cross-correlation power spectrum are presented in Section 2.4. In Section 3, we calculate the auto-correlation and cross-correlation power spectra, and forecast the precision of cosmological parameters. We summarize and conclude in Section 4.

Survey	Reference	$10^3\Omega_{ m HI}b_{ m HI}r_{ m HI,g}$
GBT-DEEP2	Chang et al. (2010)	0.55 ± 1.5
$\operatorname{GBT-WiggleZ}$	Masui et al. (2013)	$0.43 \pm 0.07 (\mathrm{stat}) \pm 0.04 (\mathrm{sys})$
$\operatorname{GBT-WiggleZ}$	Wolz et al. (2022)	$0.58 \pm 0.09 (\mathrm{stat}) \pm 0.05 (\mathrm{sys})$
GBT-eBOSS ELG	Wolz et al. (2022)	$0.40 \pm 0.09 (\mathrm{stat}) \pm 0.04 (\mathrm{sys})$
$\operatorname{GBT-eBOSS}$ LRG	Wolz et al. (2022)	$0.35 \pm 0.08({\rm stat}) \pm 0.03({\rm sys})$

Table 1. $10^3 \Omega_{\rm HI} b_{\rm HI} r_{\rm HI, g}$ measurement at redshift $z \approx 0.8$ in literature.

We use a modern cosmological simulation, the IllustrisTNG (Pillepich et al. 2018), to study the expected optical-21cm cross-correlation. The IllustrisTNG is a set of state-of-the-art cosmological magnetohydrodynamical simulations completed with the moving mesh code AREPO (Springel 2010). It is built upon the Illustris galaxy formation model, and includes physical processes such as gas cooling, star formation, stellar evolution, metal enrichment, black hole growth, stellar winds, supernovae, and active galactic nuclei (AGNs) (Weinberger et al. 2016; Pillepich et al. 2018). It has been compared with observations and found to be in good agreement (see, e.g. Nelson et al. 2018; Diemer et al. 2018). In this work, we use the TNG300-1 data set, the corresponding box size is 205 Mpc/h, with 2500^3 re- solved elements for both dark matter and baryon. Some important simulation parameters are listed in Table 2. The IllustrisTNG adopted the Planck 2015 cosmological parameters (Ade et al. 2016), i.e., $\Omega_{\Lambda} = 0.6911$, $\Omega_m = 0.3089, \ \Omega_b = 0.0486, \ \sigma_8 = 0.8159, \ n_s = 0.9667,$ and h = 0.6774. The neutral hydrogen mass for each gas particle in the simulation is given directly in the TNG dataset, but the relative abundance of atomic and molecular hydrogen (HI and H₂ hereafter) are not given, which we shall model as described below.

In this paper we are mainly interested in the neutral hydrogen at low redshift, which can be effectively covered by the L-band multi-beam receiver of FAST (1050 to 1450 MHz, corresponding to redshift up to $z\approx 0.35$ for HI survey.). We will build our HI and optical galaxy catalog by using the simulation data from snapshot 84 (hereafter snap84) with central redshift $z\approx 0.2$, and snapshot 78 (hereafter snap78) with $z\approx 0.3$.

2.1. The HI model

We construct a catalog of galaxies from the IllustrisTNG simulation. The total mass of atomic and molecule hydrogen is directly given in the simulation for each gas particle, but the two parts are not individually given. The molecular part makes a substantial fraction (e.g. Diemer et al. (2018); Lagos et al.

(2015)), so we need to model the atomic-to-molecular transition and separate the two parts. Diemer et al. (2018) has constructed atomic and molecular hydrogen catalog for galaxies in IllusitrisTNG using five different models for the atomic-to-molecular transition, but their results are only available for snapshot with redshift z=0 and $z\geq 0.5$, which does not cover the redshift range for FAST. Following Lagos et al. (2015), we model the H₂ abundance $f_{\rm mol}$ using the ratio of H₂ surface density to total neutral hydrogen surface density, $f_{\rm mol} \equiv \Sigma_{\rm H_2}/(\Sigma_{\rm HI+H_2})$, with the H₂ model from Gnedin & Kravtsov (2011, hereafter GK11),

$$f_{\text{mol}} = \left(1 + \frac{\Sigma_c}{\Sigma_{\text{HI}+\text{H}_2}}\right)^{-2},\tag{1}$$

where Σ_c is the critical threshold surface density given by:

$$\Sigma_c = 2 \times 10^7 \frac{M_{\odot}}{\text{kpc}^2} \frac{\left[\ln(1 + gD_{\text{MW}}^{3/7} (U_{\text{MW}}/15)^{4/7}) \right]^{4/7}}{D_{\text{MW}} \sqrt{1 + U_{\text{MW}} D_{\text{MW}}^2}} (2)$$

where $D_{\rm MW}$ and $U_{\rm MW}$ are the dust-to-gas ratio and the ultraviolet (UV) field strength respectively, both in units of the local value. The factor g is given by

$$g = (1 + \alpha s + s^2)/(1+s), \tag{3}$$

where:

$$\alpha = 5 \frac{U_{\text{MW}}/2}{1 + (U_{\text{MW}}/2)^2},$$
(4)

$$s = \frac{0.04}{D_{\text{MW}} + 1.5 \times 10^{-3} \times \ln(1 + (3U_{\text{MW}})^{1.7})}.$$
 (5)

The dust-to-gas ratio is assumed to be proportional to the gas metallicity, thus

$$D_{\text{MW}} = Z_{\text{MW}} = Z/Z_{\odot},\tag{6}$$

where $Z_{\odot} = 0.0127$ is the solar metallicity. For star forming cells, the UV field strength is assumed to be proportional to the star formation rate surface density,

$$U_{\rm MW} = \Sigma_{\rm SFR} / \Sigma_{\rm SFR,local}.$$
 (7)

Name	Value
$L_{\rm box}$	205 Mpc/h
$N_{ m gas}$	2500^{3}
$N_{ m DM}$	2500^{3}
N_{TRACER}	2500^{3}
$m_{\rm baryon}$	$1.1 \times 10^7 \mathrm{~M}_{\odot}$
$m_{ m DM}$	$5.9 \times 10^7 \mathrm{~M}_{\odot}$

For quiescent cells the UV field is set to be the background value.

There are a number of way to define $U_{\rm MW}$ for the background, see e.g. Diemer et al. (2018). We define the background by comparing the flux at 1000 Å as Gnedin & Kravtsov (2011). The UV background is from the 2011 updated version of Faucher-Giguere et al. (2009), which is redshift dependant. The local UV flux value is $F=3.43\times 10^{-8}$ photons s⁻¹cm⁻²Hz⁻¹ at 1000 Å (Draine 1978).

To convert the volume density in the simulation to surface density, we approximate the size of the self-gravitating system by its Jeans length λ_J (cf. e.g. Schaye 2001), the surface density of quantity X is

$$\Sigma_X = \rho_X \lambda_J, \quad \lambda_J = \sqrt{\frac{\gamma(\gamma - 1)u}{G\rho_{\text{gas}}}}.$$
 (8)

where ρ_X is the volume density, $\rho_{\rm gas}$ is the gas density, u the internal energy per unit mass, and $\gamma=5/3$ is the adiabatic index. Although the Jeans length approximation exhibits large cell-to-cell scatter, the scatter is comparable with the scatter between different H₂ models, so this would not cause too much error, and it is computationally efficient (Diemer et al. 2018). All qualities are calculated cell-by-cell.

The HI mass function and stellar function of galaxies, and the halo mass (defined as the total mass within a sphere whose mean density is 200 times the mean density of the Universe at the time of the snapshot.) function for snap84 and snap78 are shown in Fig 1. Given the mass resolution of the TNG300 simulation(Diemer et al. 2018), in plotting the mass function we select galaxies with stellar mass larger than $5 \times 10^{10} M_{\odot}$ or gas mass larger than $5 \times 10^9 M_{\odot}$ for HI mass function, as a result below that mass our HI mass function is not accurate. However, we keep all HI ingredients, including those with smaller masses, for the intensity mapping calculation. We cut the stellar mass function at $10^8 M_{\odot}$ and the halo mass function at $3 \times 10^9 M_{\odot}$ (50 dark matter particles in TNG300). As shown by Fig. 1, there is little evolution between the two snapshots.

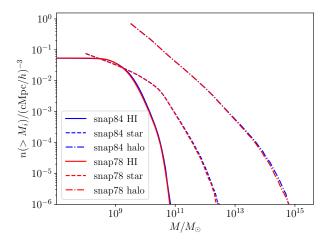


Figure 1. Cumulative galaxy HI mass function, stellar mass function, and halo mass function for the two snapshots snap84 and snap78. The left y-axis is for stellar and HI with i = HI, star and the right y-axis is for halo.

2.2. The CSS-OS Galaxy catalog

The CSST is equipped with seven photometric band filters, and slitless grating spectrographs for three bands covering 255-1000 nm. The angular resolution is $\sim 1.5'$ for photometric imaging, and $\sim 0.3'$ for spectroscopy (80% energy concentration radius). The spectroscopic resolution $R = \lambda/\Delta\lambda$ is better than 200 (see Tab. 3). The CSS-OS will include both a photometric imaging survey and a spectroscopic survey, covering a wide field of 17500 deg² in about ten years. The photometric catalog will be used to select targets for the spectroscopic survey. The photometric redshift has too large uncertainty to be of much use for cross-correlation, below we will consider the cross-correlation between the HI survey and the CSS-OS spectroscopic survey.

First we consider the construction of galaxy catalog for the CSS-OS photometric and spectroscopic surveys. The star formation process is already included in the physical model of TNG. However, to obtain the galaxies that can be detected by the CSST spectrometer and photometer, we need to calculate the spectrum for each galaxy, and convolve it with the instrument bandpass. We adopted model B in Nelson et al. (2018), which matches observation reasonably well regarding the color of galaxies. The spectrum for each star particle is calculated by the FSPS stellar population synthesis code (Conroy et al. 2009; Conroy & Gunn 2010; Foreman-Mackey et al. 2014), which models it as a single-burst simple stellar population (SSP). We adopted the Padova isochrones ((Marigo & Girardi 2007; Marigo, P. et al. 2008)), MILES stellar library (Sánchez-Blázquez et al.

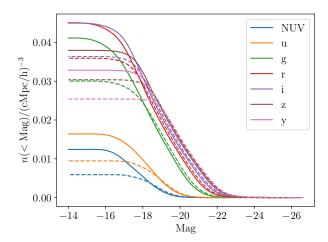


Figure 2. The expected CSST galaxy luminosity function (SNR > 10) for each photometry band. The solid line is for snap84 and the dashed line is for snap78.

(2006); Falcón-Barroso et al. (2011)), and assuming a Chabrier initial mass function (Chabrier (2003)) following Nelson et al. (2018). Cosmological parameters are also set to TNG values.

The host galaxy extinction is modeled as a power law with two dust components as described in Charlot & Fall (2000):

$$\tau_{\lambda} = \begin{cases} \tau_{1}(\lambda/\lambda_{0})^{-\alpha_{1}}, t_{\text{ssp}} \leq t_{\text{age}} \\ \tau_{2}(\lambda/\lambda_{0})^{-\alpha_{2}}, t_{\text{ssp}} > t_{\text{age}} \end{cases}, \tag{9}$$

where $\tau_1 = 1.0$, $\tau_2 = 0.3$, $\alpha_1 = \alpha_2 = 0.7$, $t_{\rm age} = 10$ Myr and $\lambda_0 = 550$ nm. The rest-frame spectrum of a star particle is attenuated by $L_{\rm rest}(\lambda) = L_i(\lambda) \exp(-\tau_{\lambda})$.

The CSST photometer is equipped seven filters: NUV, u, g, r, i, z and y, and the spectrometer is equipped with three filters: GU, GV and GI. For photometer, the AB magnitude is given by:

$$m_{\rm AB} = -2.5 \log_{10} \left(\frac{(z+1)}{4\pi r_L^2} \frac{\int LR_X \, d\ln \nu}{\int R_X \, d\ln \nu} \right) - 48.60,$$
 (10)

where $L(\nu) = L_{\rm rest}[(1+z)(1+\frac{v_{\rm los}}{c})\nu]$ is the galaxy luminosity, which is assumed to be proportional stellar mass, $v_{\rm los}$ is the line-of-sight (LOS) component of peculiar velocity, r_L is the luminosity distance, and $R_X(\nu)$ is the throughput for band X.

Precise measurement of the redshift of a galaxy depends on its emission line or absorption line features. Thus, to obtain the galaxies that can be detected by the CSST spectrometer, we choose six strong atomic lines ${\rm H}\alpha$, ${\rm H}\beta$, OIII, ${\rm H}\delta$, CaK, CaH + H ϵ , as indicators for whether a galaxy can be detected by CSST spectrometer. The first three lines are mainly detected as emission

Table 3. Wavelength range and spectral resolution for central wavelength for CSST spectrometer

Filter	Wavelength/Å	Central wavelength/Å	Resolution
GU	2550-4100	3370	287
GV	4000-6400	5250	232
GI	6200-10000	8100	207

line features, while the last three are mainly detected as absorption feature. For each feature we measure the flux within a window, and we also measure the flux of the continuum around each windows, to obtain the contrast of the emission or absorption feature. The flux contrast and continuum detected by the CSST spectrometer are given by:

$$\Delta F_{\text{line}}^{i} = \frac{1}{4\pi r_L^2} \int_{\nu_i - \frac{\Delta\nu}{2}}^{\nu_i + \frac{\Delta\nu}{2}} (L_{\text{rest}}(\nu) - L_{\text{rest}}^{\text{cont}}(\nu)) d\nu, (11)$$

$$F_{\text{cont}}(\nu) = \frac{z+1}{4\pi r_L^2} L_{\text{rest}}^{\text{cont}}((z+1)(1+\frac{v_{\text{los}}}{c})\nu)$$
 (12)

where ν_i is the rest-frame wavelength for line i, $\Delta\nu$ is the size of the window, and $L_{\rm rest}^{\rm cont}$ is the continuum luminosity. We have assumed that both the emission and absorption feature are narrow enough to be fully covered by one CSST spectrometer resolution unit.

The extinction of the Milky Way is included with the extinction law $A(\lambda)/A(V)$ given in Cardelli et al. (1989), and differential extinction E(B-V) is from the Schlegel et al. (1998) dust reddening map. For photometry, the extinction is calculated using the effective wavelength for each band, as

$$u_{\text{eff}} = \frac{\int \nu R(\nu) \ln \nu}{\int R(\nu) \ln \nu}, \qquad \lambda_{\text{eff}} = \frac{c}{\nu_{\text{eff}}}.$$

For spectroscopy, the extinction is calculated at the observed wavelength of each spectral line.

We estimate the signal-to-noise ratio (SNR) for each galaxy in the photometric surveys by making circular aperture photometry for each galaxy. For the slitless spectrometer, we model each galaxy as a point source for simplicity, though we caution that at low redshift the galaxies are generally extended sources. The SNR for each photometric aperture and spectrometer sample are calculated as the same as in Cao et al. (2018):

$$SNR = \frac{|\Delta C_s|t_{exp}\sqrt{N_{obs}}}{\sqrt{C_s t_{exp} + N_{pix}[(B_{sky} + B_{det})t_{exp} + R_n^2]}}$$
(13)

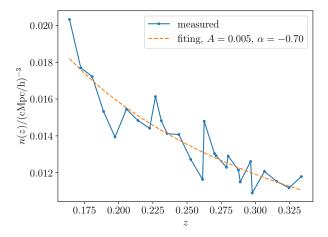


Figure 3. The number density of galaxies detected by the spectrometer, as a function of redshift.

where $N_{\rm pix} = \Delta A/l_p^2$, is the number of pixels occupied by the aperture, $l_p = 0.074~{\rm arcsec}$ is the pixel size. $B_{\rm det} = 0.02~e^-{\rm s}^{-1}{\rm pixel}^{-1}$ is the detector dark current, $t_{\rm exp}$ is exposure time and $N_{\rm obs}$ is the number observations. We set $t_{\rm exp} = 150~{\rm s}$ and $N_{\rm obs} = 4$ for spectrometer and NUV and y bands of photometer, and $t_{\rm exp} = 150~{\rm s}$ and $N_{\rm obs} = 2$ for other photometry band (Gong et al. (2019)). $B_{\rm sky}$ is the count rate from sky background per pixel:

$$B_{\rm sky} = A_{\rm eff} \int I_{\rm sky}(\nu) R_X(\nu) l_p^2 \frac{d\nu}{h\nu}, \tag{14}$$

where $A_{\rm eff}=3.14~{\rm m}^2$ is the effective area for CSST, and $I_{\rm sky}(\nu)$ is the average sky background, including earthshine and zodiacal light (Ubeda et al. 2011), C_s is the count rate from the galaxy, ΔC_s is the signal contrast. For photometer,

$$\Delta C_s = C_s = A_{\text{eff}} 10^{-0.4(m_{AB} + 48.60)} \int R_X(\nu) \frac{d\nu}{h\nu}, (15)$$

and for spectrometer and spectral line i,

$$\Delta C_s^i = A_{\text{eff}} R_X \left(\frac{\nu_i}{z+1} \right) \frac{\Delta F_{\text{line}}^i}{h\nu_i/(z+1)}, \tag{16}$$

$$C_s^i = \Delta C_s + A_{\text{eff}} \int_{\nu_o^i - \frac{\nu_R}{2}}^{\nu_o^i + \frac{\nu_R}{2}} R_X F_{\text{cont}} \frac{1}{h\nu} d\nu, \quad (17)$$

where ΔF_{line} and F_{cont} are given in (11) and (12), ν_o^i is the observed frequency for line i and ν_R is the spectral resolution of CSST spectrometer. The wavelength range and resolution for central wavelength for CSST spectrometer are given in Table 3.

The 80% energy concentration radius of the point spread function (PSF) is $R_{80}^{\rm spec} \approx 0.3$ arcsec for the

CSST spectrometer, and $R_{80}^{\rm phot} \approx 0.15$ arcsec for the CSST photometer. We model the PSF of the photometer as a gaussian function, with $\sigma_{\rm PSF}^{\rm phot} \approx 0.08$ arcsec. For photometry, the image of each galaxy within $5\sigma_{\rm PSF}^{\rm phot}$ is then convolved with the PSF before the aperture photometry, which is sufficient to model the effect of PSF. For the spectrometer, the light from one aperture is spread into a spectrum. The resolution units can be estimated as a rectangular with side length $R_{80}^{\rm spec}$ along the dispersion direction and $2R_{80}^{\rm spec}$ perpendicular to the dispersion direction.

We set the detection threshold as SNR = 10 for the photometric catalog and 5 for the spectroscopic catalog. The luminosity functions of galaxy samples for each CSST photometry band are shown in Figure 2. Because the spectroscopic target is always selected from high SNR photometric catalog, when we talk about the spectroscopic catalog, the SNR should always exceed 10 for at least one photometric band. Nevertheless, we find that almost all galaxies detected by the spectrometer with SNR larger than 5 are also detected by at least one photometric band with SNR exceeding 10.

The number density of galaxies detected by at least one spectroscopic band and at least one photometric band is shown in Figure 3. Compared with previous surveys, the galaxy number density of the CSS-OS is significantly greater. For example, the SDSS-II (main galaxy sample for DR7) is $\sim 2 \times 10^{-4} (\mathrm{Mpc/h})^{-3}$ at $z \approx 0.2$ (Ross et al. 2015), and for BOSS (DR12) it is $\sim 5 \times 10^{-4}$ to 3×10^{-4} (Mpc/h)⁻³ from z=0.2 to z=0.3 (Alam et al. 2017). The ongoing DESI survey is expected to have a galaxy number density of $\sim 9 \times 10^{-4}$ (Mpc/h)⁻³ at $z \approx 0.3$ (Levi et al. 2013), which is still significantly lower than the galaxy number density of the CSS-OS spectroscopic survey. The shot noise of power spectrum measurement would be greatly suppressed with the CSS-OS data.

On the other hand, the FAST HI galaxy survey can achieve a number density of 1×10^{-3} to $2 \times 10^{-4} \; (\mathrm{Mpc}/h)^{-3}$ from z=0.2 to z=0.3 for an integration time of 384 s per beam (Hu et al. 2020a) in the ideal case for 5σ detected galaxies, comparable to the current optical surveys, but still much lower than that of the CSS-OS.

We select galaxies that are detected in both the spectrometer and the r and g photometric bands, and show them in Figure 4. Each galaxy is shown as a point, with colors to mark their spectral characteristics: those detected only in absorption spectra are colored red, those detected only in emission spectra are colored blue, while those detected in both absorption and emission spectra are colored orange. As expected, the samples of

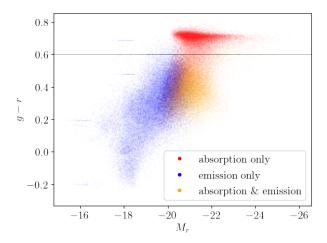


Figure 4. The color-Magnitude diagram for galaxies detected by only emission line, only absorption line or both by CSST spectrometer. The galaxies are detected by at least one spectroscopic band and both q and r bands.

detected emission line galaxies have lower absolute luminosity (larger absolute magnitude). The color of the galaxies detected only in absorption spectra are also generally redder. There are a few horizontal line structures in Figure. 4, which are artifacts from the limited time and mass resolutions of the simulation snapshots: these are some metal-poor galaxies whose stellar components are formed recently in the same time step of the simulation, but with different stellar mass. As a result, they have nearly the same colors but different absolute magnitudes. The number of such galaxies is not large in our catalog (less than 1% of the whole catalog) and will not have significant influence on our result. The optical galaxy catalog can be split into two groups by q-r=0.6as marked by the black line in the plot, corresponding to blue and red galaxies.

2.3. Intensity mapping

To simulate the intensity mapping observation process, the TNG300-1 box is divided into $2048 \times 2048 \times 128$ voxels. The angular resolution is approximately 0.6 arcmin for snap84 and 0.4 arcmin for snap78 respectively, finer than the FAST beam size at this frequency. The corresponding frequency resolution is about 0.6 MHz for snap84 and 0.5 MHz for snap78 in the frequency range of interest. The frequency resolution of the FAST raw data is usually much finer, e.g. 7.6 kHz in our own pilot survey, but this finer resolution is not needed for intensity mapping, though useful for efficient RFI flagging and also for some HI galaxy observation. It is usually re-binned to a lower frequency resolution

for intensity mapping data processing. The brightness temperature of the 21 cm signal for each voxel is given by (Villaescusa-Navarro et al. 2018) $T_b = T_0 \frac{\rho_{\rm HI}}{\rho_c^0}$, where $\rho_{\rm HI}$ is the comoving density of HI and ρ_c^0 is the critical density at z=0, and

$$T_0 = 189h \frac{H_0(1+z)^2}{H(z)} \text{ mK}$$
 (18)

The foreground is generated using GSM2016 model (Zheng et al. 2016) of the Python package PYGDSM (Price 2016). We have taken a square patch centered at RA = 12 h, $Dec = 25^{\circ}48'$ (corresponding to the zenith for the FAST location), and interpolated the foreground map to the center of each voxel.

To simulate the drift scan survey, we generate mock time-ordered data as

$$T_A(x_b, y_b, \tau) = \int T_b(x, y) \mathcal{B}(x, y; x_b, y_b, \tau) dx dy + n_T,$$

where x_b and y_b is the direction of the beam center as given in Jiang et al. (2020), $\tau = 12/\cos \delta$ s is the integration time which corresponds to the time for the sky to drift over the beam at the declination δ , and n_T is the thermal noise. Here we model the instant beam response as a Gaussian function,

$$B(x, y; x_b, y_b) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x - x_b)^2 + (y - y_b)^2}{2\sigma^2}\right),$$
(19)

where $\sigma = 0.518\lambda/(300 \text{ m})$, though the real beam may have a more complicated dependence on the frequency. The averaged beam over integration time is denoted as $\mathcal{B}(x, y; x_b, y_b, \tau)$.

We model the thermal noise n_T as gaussian white noise with standard deviation given by (Wilson et al. 2013):

$$\sigma_T = \frac{T_{\text{sys}}}{\sqrt{2\Delta\nu\tau N_{\text{scan}}}},\tag{20}$$

where $\Delta\nu$ is the frequency channel width, $N_{\rm scan}$ is the scan times for the sky area, $T_{\rm sys}$ is the system temperature. Following Hu et al. (2020b), we model the system temperature as $T_{\rm sys} = T_{\rm rec} + T_{\rm sky}$, with the receiver temperature of about 20 K, and the sky temperature off the galactic plane as

$$T_{\text{sky}} = \left(2.73 + 25.2 \left(\frac{408 \text{ MHz}}{\nu}\right)^{2.75}\right) \text{ K.} \quad (21)$$

which yields a total system temperature of about 25 K off the galactic plane, in agreement with the measurements given in Jiang et al. (2020). The minimum

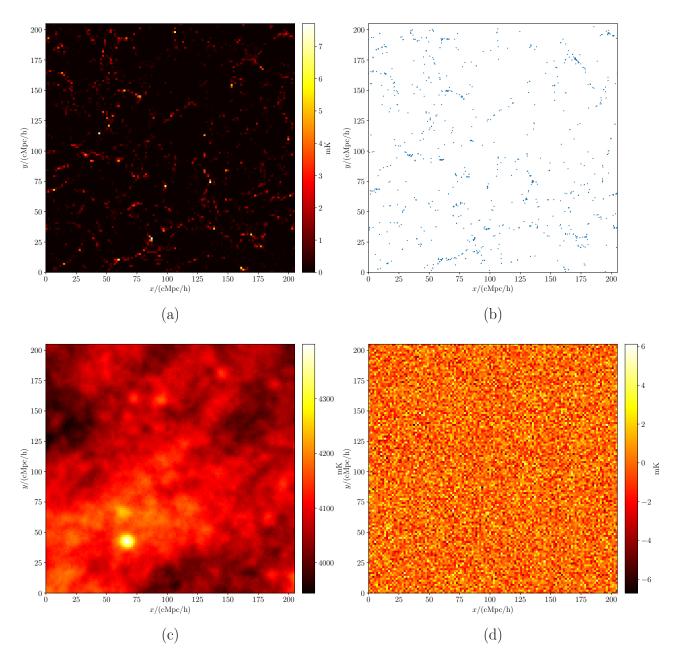


Figure 5. A slice at $\nu \approx 1185.725$ MHz for (a) the 21 cm map, (b) the galaxies that can be detected by CSST spectrometer, (c) the foreground map and (d) the thermal noise map.

variance estimator of the map is then generated by (Tegmark 1997):

$$\hat{T}_b = (A^T N^{-1} A)^{-1} A^T N^{-1} T_A, \tag{22}$$

where A is the pointing matrix and N is the covariance matrix of data noise, which is assumed diagonal. We make a coarser $128 \times 128 \times 128$ voxels map, the side length is approximately 1.6 Mpc/h, and the angular resolution is about 9.8 arcmin for snap84 and 6.6 arcmin for snap78. We adopt a coarse grid here, which is suffi-

cient for the problems studied here, as we are not concerned about the smaller scales. One slice at about 1185.725 MHz for 21 cm, foreground and thermal noise for $N_{\rm scan}=1$ are shown in Figure 5. For comparison, the galaxies at the slice that can be detected by CSST spectrometer are also shown. We note that the sky is dominated by the foreground, and the thermal noise has a magnitude larger than or comparable to the 21cm signal, making its detection challenging. We find that in our simulation box, the temperature fluctuation of 21

cm signal is about 0.36 mK for snap84 and 0.40 mK for snap78, which are much smaller than the thermal noise level of 1.47 mK for snap84 and 2.85 mK for snap78. Nevertheless, the 21 cm power spectrum increases with scale, while the thermal noise power spectrum remains flat. For $k < 0.2 \, (\mathrm{Mpc}/h)^{-1}$, the 21 cm auto power spectrum is larger than the thermal noise, but the signal-to-noise ratio is not high. It requires a few repeated scans to reduce the noise power spectrum to achieve a good signal-to-noise ratio.

2.4. Foreground removal and transfer function

In intensity mapping observation, the foreground radiation dominates over the signal, but it could be removed by exploiting the difference in the statistical properties of the foreground and the 21cm signal. Below, we simulate the foreground removal with two methods frequently employed in 21cm data analysis: the principal component analysis (PCA) and the independent component analysis (ICA). The PCA method exploits the fact that along the line of sight (LOS) the foreground is highly correlated at different frequencies while the 21 cm signal is not. Such components are associated with the principal components in the frequency-pixel covariance matrix (Masui et al. 2013; Alonso et al. 2015; Bigot-Sazy et al. 2015; Zuo et al. 2019). The ICA method exploits the fact that the foregrounds have large non-Gaussianity while the 21cm signal is nearly Gaussian (Chapman et al. 2012; Wolz et al. 2014; Alonso et al. 2015; Asorey et al. 2020).

However, foreground removal will still inevitably cause some loss of 21 cm signal, especially on large scales which interested us. This problem can be compensated at least partially if we know the expected amount of signal loss in the power spectrum. This loss transfer function can be estimated using the simulation. We simulated the mock 21 cm signal with halo and HI mass following Wolz et al. (2019), Murray (2018) and Murray (2014), and set $M_{\rm min}=10^{12.3}~M_{\odot}/h$ as did in Wolz et al. (2022).

We create 500 mock HI maps with the same box size and spatial resolution, and for each halo the HI distribution is represented by 1000 HI particles in a cored Navarro-Frank-White (cNFW) profile (Padmanabhan & Refregier 2017). Then we project the mock map to the foreground-dominant bases obtained by PCA or ICA, and subtract the projected components from the mock map to simulate the signal loss. Denoting the mock power spectrum as $\hat{P}_0^{\rm HI}(k)$, and the cross-correlation power spectrum of the mock data before and after the foreground removal procedure as $\hat{P}_X^{\rm HI}(k)$, the transfer

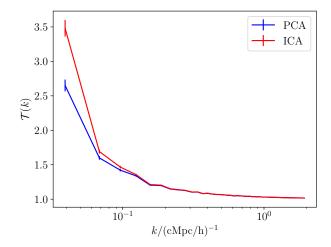


Figure 6. The transfer function $\mathcal{T}_X(k)$ for both PCA and ICA, for snap84. The errorbar is estimated directly from the standard deviation between mock data.

functions can be obtained by (Wolz et al. (2022)):

$$\mathcal{T}(k) = \frac{\langle \hat{P}_0^{\text{HI}}(k) \rangle}{\langle \hat{P}_X^{\text{HI}}(k) \rangle}.$$
 (23)

It can also be computed for both foreground removal methods, i.e. the PCA and ICA methods.

The simulated transfer functions are shown in Figure 6, with more signal loss on smaller k scales. The statistical error of the transfer function is below 3% level, which only contributes to $\sim 1\%$ of the total variance of the cross-correlation. The transfer functions are comparable for both PCA and ICA.

Note that in a foreground removal simulation, especially for the HI auto-correlations, one may obtain more optimistic results than in the analysis of real data, because the model may be too simplistic and inadequate in its representation of the complexities of the real world foreground and instrument response. Fortunately, the foreground residues are usually uncorrelated with the cosmological signal, so in the cross-correlation these are suppressed, and our forecast on the cross-correlation thus works better than on the auto-correlations.

3. POWER SPECTRUM AND COSMOLOGICAL PARAMETERS ESTIMATION

Below we infer cosmological parameters from the mock observations described above. The optical galaxy autocorrelation power spectrum is given by

$$P_g(\mathbf{k}) = V \langle \delta_g(\mathbf{k}) \delta_g^*(\mathbf{k}) \rangle - P_{\text{shot}},$$
 (24)

where $P_{\rm shot}=V/N_g=1/\langle n_g\rangle$ is the shot noise term for the galaxy survey, and $V=(205~{\rm Mpc}/h)^3$ is the box

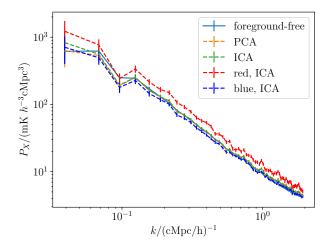


Figure 7. The HI-optical galaxy cross-correlation power spectrum. We show the foreground-free (obtained directly from the simulation box, without foreground), and after foreground removal and transfer using the two methods. The HI-red galaxy and HI-blue galaxy cross power spectra are also shown, using ICA foreground removal.

volume, while the cross-correlation power spectrum with the 21cm brightness temperature is given by

$$P_X(\mathbf{k}) = V \Re \left(\delta_T(\mathbf{k}) \delta_g^*(\mathbf{k}) \right) \tag{25}$$

The error for the auto-correlation of optical galaxies and the cross-correlation power spectrum are estimated as (Feldman et al. 1994; Wolz et al. 2017)

$$\sigma_{P_g}(k) = \frac{2\pi}{\sqrt{V_{\text{sur}}k^2\Delta k}} \left(P_g(k) + P_{\text{shot}}\right), \tag{26}$$

$$\sigma_{P_X}(k) = \frac{2\pi}{\sqrt{2V_{\text{sur}}k^2\Delta k}} \times (27)$$

$$\sqrt{P_X^2(k) + (P_T(k) + P_{\text{ns}}(k))(P_g(k) + P_{\text{shot}})},$$

where Δk is k-bin width, $P_{\rm ns}(k)$ is the power spectrum of the thermal noise.

In this paper we have used the real space position data of the simulation. In an actual observation, the peculiar motion would induce a redshift space distortion (RSD) effect. On linear scales, the RSD introduces a correction factor of the form $(1 + \beta \mu^2)^2$ in the power spectrum, where μ is the wave vector direction cosine with respect to the line of sight, $\beta = f/b$, with f the growth factor and b the bias (Kaiser 1987). With good anisotropic power spectrum measurement, in principle one can break the degeneracy between $b_{\rm HI}$ and $\Omega_{\rm HI}$ using the RSD effect(Masui et al. 2010), but we will not use this additional information here.

Although the Galactic foreground is uncorrelated with the galaxy sample, due to its much higher intensity, its

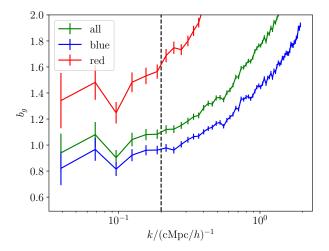


Figure 8. The galaxy bias factor b_g for all optical galaxies, red galaxies and blue galaxies in snap84. The vertical dashed line is $k = 0.2 \text{ (Mpc/h)}^{-1}$.

fluctuation still dominates the cross-power spectrum if it is not removed. The foreground removal can efficiently enhance the signal-to-noise ratio in the cross-correlation power spectrum. We find that if the foreground is subtracted to a 1% level, which is still much stronger than the 21 cm signal, the cross-correlation power spectrum can already be measured pretty well, with the signal dominating over the foreground contribution. In Fig. 7 we plot the mock observation cross power spectrum for the snap84, as well as the foreground-free spectrum. Different foreground subtraction techniques are also compared, though in this simulation there is no substantial difference. All results are quite consistent with each other. Hereafter, we only discuss the result using ICA for foreground removal.

As discussed earlier and shown in Fig. 4, here we can also divide the galaxy sample into "red" (g-r>0.6) and "blue" $(g-r\leq0.6)$ ones. The HI-red galaxy and HI-blue galaxy cross power spectra are shown in Fig. 7. The cross power spectrum for HI-red galaxy is slightly higher than the blue ones, which is not unexpected as the red galaxies have a larger bias factor, as have been found previously (see e.g. Zehavi et al. (2002); Cooray (2005); Cresswell & Percival (2008)). However, as we will show later the blue galaxies are actually more correlated with the neutral hydrogen than the red galaxies, which is consistent with previous works (e.g.Dutta et al. (2020); Wolz et al. (2022), as blue galaxies are usually more HI-rich than red galaxies.

The optical galaxy auto-correlation power spectrum is related to the matter power spectrum as (Wolz et al.

Table 4. b_g for all galaxies, blue galaxies and red galaxies for snap84 and snap78.

Catalog	snap84 ($z = 0.2$)	$\mathrm{snap}78(z=0.3)$
all	1.040 ± 0.022	1.036 ± 0.022
blue	0.926 ± 0.020	0.963 ± 0.021
-red	1.471 ± 0.032	1.508 ± 0.034

2017)

$$P_a(k) = b_a^2 P_{\text{lin}}(k), \tag{28}$$

 b_g is the bias of galaxy, $P_{\text{lin}}(k)$ is the linear matter power spectrum, while the 21cm auto-correlation power spectrum is (Wolz et al. (2017))

$$P_T(k) = T_0^2 \Omega_{\rm HI}^2 b_{\rm HI}^2 P_{\rm lin}(k),$$
 (29)

 $\Omega_{\rm HI}$ is the density parameter for HI, $b_{\rm HI}$ is the bias of HI. The cross-correlation power spectrum is (Wolz et al. 2017)

$$P_X(k) = T_0 \Omega_{\text{HI}} b_{\text{HI}} b_g r_{\text{HI},g} P_{\text{lin}}(k), \qquad (30)$$

where we introduced the galaxy-HI correlation coefficient $r_{\text{HI},g}$, to account for the fact that the optical and radio traces may not be perfectly correlated.

Given the matter power spectrum, the bias factor b_g can be measured from the optical galaxy auto-correlation power spectrum. We calculate the linear matter power spectrum by using CAMB (Lewis et al. 2000). The cosmological parameters are set to be consistent with those assumed in the IllustrisTNG simulation.

We plot the bias factor b_g in Fig. 8. As we can see, the bias factor on small scales (large k) varies with scale, due to the non-linear evolution which is more prominent on small scales. The collapse of matter into halos generates extra power at small scales compared with the linear theory and thus boosts up the small scale power spectrum, but on large scales the bias approaches a constant. We can infer the bias from the power spectrum with $k < 0.2 \; ({\rm Mpc}/h)^{-1}$ for the whole sample, and for the blue, and red galaxy sub-samples. This scale is marked by the dashed vertical lines in Fig. 8.

The average value and standard deviation of the posterior distribution of the bias factors are given in Table 4, the values are comparable with the value in literature, e.g. Cresswell & Percival (2008). For the whole catalog

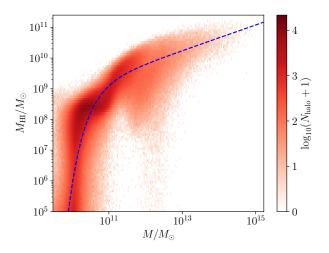


Figure 9. HI mass as a function of halo mass, the dash curve shows the result of fitting $M_{\rm HI}(M)$ using Eq. (32) for snap84

Table 5. The HI-halo mass fitting function, HI bias and HI abundance.

Snapshot	snap84 (z = 0.2)	snap78(z=0.3)
	0	
M_0	$2.2 \times 10^9 M_{\odot}$	$1.8 \times 10^9 M_{\odot}$
$M_{ m min}$	$7.2\times10^{10}M_{\odot}$	$6.3 \times 10^{10} M_{\odot}$
α	0.42	0.45
$b_{ m HI}$	0.839	0.870
$\Omega_{ m HI}$	5.20×10^{-4}	5.07×10^{-4}

with all galaxies $b_g \sim 1$, the red galaxies have a bias of $b_g \sim 1.4$, while the blue galaxies have a bias value of $b_g \sim 0.9$. The bias also decreases slightly from snap78 centered at z=0.3 to snap84 centered at z=0.2.

The $b_{\rm HI}$ can in principle be obtained from 21cm autocorrelation spectrum $P_T(k)$, if it is successfully measured in the intensity mapping survey. However, as we noted earlier, it is more difficult to measure than the crosscorrelation power spectrum. Here we do not assume that it is available. Alternatively, it can be computed if the halo HI mass function $M_{\rm HI}(M)$ can be obtained through simulations and observations of HI galaxies. Following Li & Ma (2017), we estimate the corresponding HI bias:

$$b_{\rm HI} = \frac{\int_{M_{\rm min}}^{M_{\rm max}} dM \, \frac{dn}{dM}(M, z) M_{\rm HI}(M) b(M, z)}{\int_{M_{\rm min}}^{M_{\rm max}} dM \, \frac{dn}{dM}(M, z) M_{\rm HI}(M)}, \quad (31)$$

where dn/dM is the halo mass function, for which we use the fitting function given by Tinker et al. (2008),

with $M_{\rm min}=10^8 M_{\odot}/h$ and $M_{\rm max}=10^{17} M_{\odot}/h$. The halo mass is defined as the total mass within a sphere whose mean density is 200 times the mean density of the Universe at the time of the snapshot. For b(M), we adopt the fitting formula from Tinker et al. (2010), which is in good consistency with TNG300 simulation (Springel et al. (2017)).

We plot the HI mass and the halo mass of our simulation in Fig. 9. There is a large spread in the HI-halo mass relation, but we can see there is a clear trend. There is also a clear break at the halo mass of about $7 \times 10^{10} M_{\odot}$. We fit the $M_{\rm HI}(M)$ in simulation with following formula,

$$M_{\rm HI}(M) = M_0 \left(\frac{M}{M_{\rm min}}\right)^{\alpha} \exp\left(-\frac{M_{\rm min}}{M}\right)$$
 (32)

We only fit halos with mass larger than $3 \times 10^9~M_{\odot}$, which contain more than 99% HI mass. The fitting parameters and HI bias $b_{\rm HI}$ are given in Table 5. The HI abundance is also measured directly from the simulation box, and listed in Table 5. Our result is consistent with Villaescusa-Navarro et al. (2018) for α and $M_{\rm min}$, which are the relevant parameters for bias calculation. We also measured $b_{\rm HI}$ from the auto-correlation power spectrum of HI, obtained directly from the simulation box. The result is in good agreement with the value in Table 5, and the relative error is within 3%.

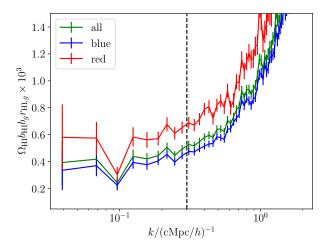


Figure 10. The $\Omega_{\text{HI}}b_{\text{HI}}b_gr_{\text{HI},g}$ for snap84. The vertical dashed line is $k=0.3~(\text{Mpc}/h)^{-1}$.

Finally we can proceed to infer the total neutral hydrogen abundance, from the cross-correlation power spectrum. In Fig. 10 we plot the quantity $\Omega_{\rm HI}b_{\rm HI}b_{\rm g}r_{\rm HI,g}$ as derived from the cross-correlation. We measure the

bias at $k \leq 0.3 (\mathrm{Mpc}/h)$ as marked by the dashed vertical line in Fig. 10. With b_g determined from the optical galaxy auto-correlation, we can then infer $\Omega_{\mathrm{HI}}b_{\mathrm{HI}}r_{\mathrm{HI},g}$. Hereafter we assume that the $M_{\mathrm{HI}}(M)$ has already be determined by other observations except a normalization factor M_0 and use the theoretical b_{HI} in Table. 5 for parameter estimation. Then the $\Omega_{\mathrm{HI}}r_{\mathrm{HI},g}$ can be constrained.

The inferred value of $\Omega_{\mathrm{HI}}r_{\mathrm{HI},g}$ for the three catalog are given in Table 6. The central value is calculated by averaging over different realizations of noise to suppress the statistic error. The error given is estimated for the volume of the simulation box. The inferred value is slightly smaller than the input value, which might be due to the foreground removal procedure. Nevertheless, we note that the difference between the input and inferred values is within the estimated total error. We find that including two more small scale (high k) modes can reduce the difference, because smaller scale modes are more robust against foreground removal and also contribute more to the final result due to lower error. However, these modes may be mildly non-linear.

The factor $r_{\mathrm{HI},g}$ quantifies how the HI are correlated with the galaxies as found in the optical survey. In our model, it should be fairly close to 1 at large scale. To check this, we calculate $r_{\mathrm{HI},g}$ by averaging the quotient $P_X(k)/\sqrt{P_g(k)P_T(k)}$ with $k<0.2~(\mathrm{Mpc}/h)^{-1}$. The result is given in Table 6, and we find that the $r_{\mathrm{HI},g}$ values are indeed quite close to 1, especially for the 'blue' catalog. Thus if we have a robust estimation of halo HI mass function $M_{\mathrm{HI}}(M)$, we should be able to infer Ω_{HI} robustly from the cross-correlation result. Although in our analysis, $r_{\mathrm{HI},g}<1$ for all cases, we notify that it can be larger than 1 if the shot noise of optic galaxies dominates over $P_g(k)$. See Wolz et al. (2017) for more details.

We estimated the contribution of thermal noise, by generating 100 different realizations of noise, but with the same input foreground and 21 cm map. We calculated the Bessel-corrected (i.e. dividing by n-1) standard deviation between the inferred value of different realizations. The error obtained this way is given as $\sigma_{\rm ns-re}$ (i.e. noise realization) in Table 6. According to Eq. (27), the error in the cross-correlation power spectrum comes from the errors in both the 21cm power spectrum and the optical galaxy power spectrum. The error in 21cm power spectrum is dominated by thermal noise. We find that $\sigma_{\rm ns-re}$ can be represented quite well by the thermal noise term in Equation (27), i.e.

$$\sigma_{\rm ns-re} \approx \frac{2\pi}{\sqrt{2V_{\rm sur}k^2\Delta k}} \frac{\sqrt{P_{\rm ns}(k)[P_g(k) + P_{\rm shot}]}}{P_{\rm lin}(k)b_qb_{\rm HI}T_0}.$$
 (33)

Catalog	snap84			$\operatorname{snap} 78$						
	Input	$Value \pm Error$	$\sigma_{ m ns-re}$	σ_c	$r_{\mathrm{HI},g}$	Input	$Value \pm Error$	$\sigma_{ m ns-re}$	σ_c	$r_{\mathrm{HI},g}$
all	0.516	0.512 ± 0.016	0.008	0.012	0.977	0.526	0.523 ± 0.018	0.014	0.013	0.990
blue	0.521	0.518 ± 0.016	0.008	0.012	0.988	0.526	0.523 ± 0.018	0.014	0.013	0.996
red	0.482	0.477 ± 0.016	0.009	0.013	0.920	0.506	0.502 ± 0.019	0.015	0.013	0.930

Table 6. The measured value and error(1σ) of $10^3\Omega_{\rm HI}r_{\rm HI,g}$ for all galaxies, as well as "blue" and "red" galaxies. We also list the thermal noise $\sigma_{\rm ns-re}$ and cosmic variance σ_c .

The impact of the thermal noise is larger for snap78, because a given comoving size corresponds to a smaller angular size at higher redshift, so the effective observation time for each pixel is smaller at higher redshift, and the thermal noise level is higher. Here we assumed the same system temperature, in fact the system temperature could also be slightly higher at higher redshifts, due to the increasing sky temperature at lower frequencies (c.f. Eq. 20). If we observe a sky area for $N_{\rm scan}$ times, the thermal noise can be reduced as $\sigma_{\rm ns-re} \propto 1/\sqrt{N_{\rm scan}}$. Thus, if $N_{\rm scan}=4$, the thermal noise contribution will be sub-dominant for both the snap84 and snap78.

The shot noise in the optical galaxy power spectrum is fairly small for the CSST, as it has a very large comoving number density. We also calculated the cosmic variance without thermal noise contribution for the simulation box size, which is given as the σ_c column of Table 6 by scaling the posterior error of $\Omega_{\rm HI}b_{\rm HI}b_gr_{\rm HI,g}$ by a factor of $b_gb_{\rm HI}$. This is quite large as our simulation box size is limited, but for the large CSST survey we are considering, the cosmic variance would be much reduced, as we shall discuss below. We do not propagate the cosmic variance of b_g into the final error, because the galaxy and HI share the same cosmological realization and therefore their cosmic variance are correlated. The error given is statistical, the modeling error for $b_{\rm HI}$ is not included.

We now extend our result by considering the realistic survey with a larger sky area. We set the sky area for the intensity mapping experiment as $10000~\rm deg^2$. The frequency range is set as 90 MHz around z=0.2 (1141 to 1231 MHz) and z=0.3 (1049 to 1139 MHz). The total survey volume is then $V_{\rm sur}\approx 2.45\times 10^8~\rm (Mpc/h)^3$ for $z\approx 0.2$ and $V_{\rm sur}\approx 5.83\times 10^8~\rm (Mpc/h)^3$. The sky area of CSS-OS is 17500 deg², though for cross-correlation the area is to be set as the smaller of the surveys, i.e. $10000~\rm deg^2$. The redshift range of CSS-OS is set the same as intensity mapping.

Table 7. The value of $10^3 \Omega_{\rm HI} r_{\rm HI,g}$ and measurement error obtained from HI-optical galaxy cross-correlation for 10000 deg² sky area.

Catalog	snap84	snap78
all	0.5118 ± 0.0029	0.5226 ± 0.0022
blue	0.5176 ± 0.0029	0.5231 ± 0.0022
red	0.4770 ± 0.0030	0.5020 ± 0.0024

The inferred value of $\Omega_{\rm HI}b_{\rm HI}r_{\rm HI,g}$ and its error is given in Table 7 by scaling errors in Table. 6 by a factor of $\sqrt{V_{\rm box}/V_{\rm sur}}$. Here the error is the total, including both thermal noise and cosmic variance, and the cosmic variance has been scaled to that of the survey area of 10,000 deg². For this much larger volume, the cosmic variance is significantly reduced, and the errors are nearly the same for the blue, red, and all galaxies samples.

Comparing these results with the existing cross-correlation measurements using GBT HI IM survey data and a number of optical surveys given in Table 1, we see that a much better precision in the measurement can be achieved with the FAST HI survey and CSST survey. The improvement is more than an order of magnitude compared with the current best result, the relative error of $\Omega_{\rm HI}$ can be suppressed to a level of $\sim 0.6\%$.

4. CONCLUSION AND DISCUSSION

The CSST is a major space astronomy project in the coming decade, and the CSS-OS is the primary sky survey planned for the CSST, with a unprecedented area and magnitude limits for optical surveys. In this work, we consider the potential of the cross-correlation of the CSS-OS spectroscopic survey with the HI survey of the FAST. The HI survey has generally been challenging, as the signal is relatively weak and prone to foreground contamination. The cross-correlation with an optical

survey will greatly enhance the signal, making the detection and measurement much more accessible. On the other hand, the HI observation will reveal the distribution of the neutral hydrogen, the basic construction material of the galaxies. The cross-correlated observation would provide a solid check on our basic understanding of the large scale structure and working of galaxy formation and evolution, and opportunities for finding deep and hidden relations between the gas, stars and dark matter.

We generate realistic samples of galaxy catalogs and HI maps from the IllustrisTNG300 simulation, and present the construction of these mocks in detail. For the optical mock data, we estimate the SNR of the CSS spectrometer for each galaxy in the simulation, to produce an SNR-limited sample. The sample is naturally divided into red and blue galaxies by their color. For the HI mock data, we have taken into account the noise, a simple beam model, and galactic foreground. We focus on redshift $0.2 \sim 0.3$, which is covered by the L-band multi-beam receiver in the FAST.

Based on the mock data, we simulated foreground removal with the PCA and ICA methods, which in the simulation produce similar results. We produce foreground-removed sky maps, to cross-correlate with the optical sample, and derive the power spectrum. Finally, we consider using this observation to measure cosmological parameters. The HI-optical cross-correlation is particularly suitable for measuring the total HI content at the low redshift Universe. We assume the bias factor b_q can be obtained from the auto-correlation of the optical galaxies, and the HI-optical cross-correlation measures $\Omega_{\rm HI}b_{\rm HI}r_{{\rm HI},g}$, where $b_{\rm HI}$ is the HI bias, and $r_{{\rm HI},g}$ is the correlation factor for the HI and optical galaxies. We can then infer the value of $\Omega_{\rm HI}$. Furthermore, by cross-correlating with galaxies of different colors, the HI content of these can be separately measured.

In the present work we have ignored the light cone effect, i.e. the evolution of galaxy statistics from the redshift corresponding to the far side of the simulation box, to that of the near side. However, the estimated error is not too much affected. We also limited our discussion mainly to the measurement of $\Omega_{\rm HI}$. The power spectrum features such as the BAO wiggles can also be used to constrain other cosmological parameters, though the optical galaxy spectrum would already provide a very good constraint on that.

With the high comoving number density that can be achieved by the CSST survey, the shot noise in the optical galaxy power spectrum is much suppressed. The cosmic variance is also anticipated to be very small for the large overlapping area of the CSST and FAST surveys. A very accurate measurement of $\Omega_{\rm HI}$ is expected, with a relative error at the 0.6% level. This illustrates the power of the upcoming CSST survey.

ACKNOWLEDGEMENT

We thank Qi Guo, Haitao Miao, Hu Zou and Xin Zhang (NAOC) for the discussions. FD and XC are supported by the Ministry of Science and Technology (MoST) inter-government cooperation program China-South Africa Cooperation Flagship project 2018YFE0120800, the China Manned Space Project with grant No. CMS-CSST-2021-B01. YG acknowledges the support of MOST-2020SKA0110402, NSFC-11822305, NSFC-11773031, NSFC-11633004, and CMS-CSST-2021-A01. YW acknowledges the the National Key R&D Program 2017YFA0402603, and the CAS Interdisciplinary Innovation Team (JCTD-2019-05). SD acknowledge Chinese Academy of Sciences College Student Innovation Training Program (No. 20214000231).

DATA AVAILABILITY STATEMENT

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Ade, P. A., Aghanim, N., Arnaud, M., et al. 2016, Astronomy & Astrophysics, 594, A13

Alam, S., Ata, M., Bailey, S., et al. 2017, Monthly Notices of the Royal Astronomical Society, 470, 2617, doi: 10.1093/mnras/stx721

Alonso, D., Bull, P., Ferreira, P. G., & Santos, M. G. 2015, MNRAS, 447, 400, doi: 10.1093/mnras/stu2474

Amiri, M., Bandura, K., Chen, T., et al. 2022, arXiv preprint arXiv:2202.01242

Anderson, C. J., Luciw, N. J., Li, Y. C., et al. 2018, MNRAS, 476, 3382, doi: 10.1093/mnras/sty346

Asorey, J., Parkinson, D., Shi, F., et al. 2020, Monthly Notices of the Royal Astronomical Society, 495, 1788

Bandura, K., Addison, G. E., Amiri, M., et al. 2014, in Proc. SPIE, Vol. 9145, Ground-based and Airborne Telescopes V, 914522, doi: 10.1117/12.2054950

Battye, R., Browne, I., Chen, T., et al. 2016, ArXiv, 1610.06826. https://arxiv.org/abs/1610.06826

- Battye, R. A., Brown, M. L., Browne, I. W. A., et al. 2012, ArXiv, 1209.1041. https://arxiv.org/abs/1209.1041
- Bigot-Sazy, M.-A., Dickinson, C., Battye, R. A., et al. 2015, MNRAS, 454, 3240, doi: 10.1093/mnras/stv2153
- Bigot-Sazy, M. A., Ma, Y. Z., Battye, R. A., et al. 2016, in Astronomical Society of the Pacific Conference Series, Vol. 502, Frontiers in Radio Astronomy and FAST Early Sciences Symposium 2015, ed. L. Qain & D. Li, 41. https://arxiv.org/abs/1511.03006
- Cao, Y., Gong, Y., Meng, X.-M., et al. 2018, Monthly Notices of the Royal Astronomical Society, 480, 2178
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, The Astrophysical Journal, 345, 245
- Chabrier, G. 2003, Publications of the Astronomical Society of the Pacific, 115, 763, doi: 10.1086/376392
- Chang, T.-C., Pen, U.-L., Bandura, K., & Peterson, J. B. 2010, Nature, 466, 463
- Chang, T.-C., Pen, U.-L., Peterson, J. B., & McDonald, P. 2008, Phys. Rev. Lett., 100, 091303, doi: 10.1103/PhysRevLett.100.091303
- Chapman, E., Abdalla, F. B., Harker, G., et al. 2012, MNRAS, 423, 2518, doi: 10.1111/j.1365-2966.2012.21065.x
- Charlot, S., & Fall, S. M. 2000, The Astrophysical Journal, 539, 718
- Chen, X. 2012, in International Journal of Modern Physics Conference Series, Vol. 12, International Journal of Modern Physics Conference Series, 256–263, doi: 10.1142/S2010194512006459
- CHIME Collaboration, Amiri, M., Bandura, K., et al. 2022, arXiv e-prints, arXiv:2202.01242. https://arxiv.org/abs/2202.01242
- Conroy, C., & Gunn, J. E. 2010, The Astrophysical Journal, 712, 833
- Conroy, C., Gunn, J. E., & White, M. 2009, The Astrophysical Journal, 699, 486
- Cooray, A. 2005, Monthly Notices of the Royal Astronomical Society, 363, 337, doi: 10.1111/j.1365-2966.2005.09457.x
- Cresswell, J. G., & Percival, W. J. 2008, Monthly Notices of the Royal Astronomical Society, 392, 682, doi: 10.1111/j.1365-2966.2008.14082.x
- Diemer, B., Stevens, A. R., Forbes, J. C., et al. 2018, The Astrophysical Journal Supplement Series, 238, 33
- Draine, B. T. 1978, The Astrophysical Journal Supplement Series, 36, 595
- Dutta, S., Khandai, N., & Dey, B. 2020, MNRAS, 494, 2664, doi: 10.1093/mnras/staa864

- Falcón-Barroso, J., Sánchez-Blázquez, P., Vazdekis, A., et al. 2011, A&A, 532, A95, doi: 10.1051/0004-6361/201116842
- Faucher-Giguere, C.-A., Lidz, A., Zaldarriaga, M., & Hernquist, L. 2009, The Astrophysical Journal, 703, 1416
- Feldman, H. A., Kaiser, N., & Peacock, J. A. 1994, ApJ, 426, 23, doi: 10.1086/174036
- Foreman-Mackey, D., Sick, J., & Johnson, B. 2014, Zenodo Gnedin, N. Y., & Kravtsov, A. V. 2011, The Astrophysical Journal, 728, 88
- Gong, Y., Liu, X., Cao, Y., et al. 2019, The Astrophysical Journal, 883, 203
- Hu, W., Wang, X., Wu, F., et al. 2020a, Mon. Not. Roy. Astron. Soc., 493, 5854, doi: 10.1093/mnras/staa650
- —. 2020b, Monthly Notices of the Royal Astronomical Society, 493, 5854, doi: 10.1093/mnras/staa650
- Jiang, P., Tang, N.-Y., Hou, L.-G., et al. 2020, Research in Astronomy and Astrophysics, 20, 064, doi: 10.1088/1674-4527/20/5/64
- Kaiser, N. 1987, Monthly Notices of the Royal Astronomical Society, 227, 1, doi: 10.1093/mnras/227.1.1
- Lagos, C. d. P., Crain, R. A., Schaye, J., et al. 2015, Monthly Notices of the Royal Astronomical Society, 452, 3815
- Levi, M., Bebek, C., Beers, T., et al. 2013, arXiv e-prints, arXiv:1308.0847. https://arxiv.org/abs/1308.0847
- Lewis, A., Challinor, A., & Lasenby, A. 2000, The Astrophysical Journal, 538, 473
- Li, J., Zuo, S., Wu, F., et al. 2020, SCIENCE CHINA Physics, Mechanics & Astronomy, 63, 1
- Li, Y.-C., & Ma, Y.-Z. 2017, PhRvD, 96, 063525, doi: 10.1103/PhysRevD.96.063525
- Marigo, P., & Girardi, L. 2007, A&A, 469, 239, doi: 10.1051/0004-6361:20066772
- Marigo, P., Girardi, L., Bressan, A., et al. 2008, A&A, 482, 883, doi: 10.1051/0004-6361:20078467
- Masui, K., Switzer, E., Banavar, N., et al. 2013, The Astrophysical Journal Letters, 763, L20
- Masui, K. W., McDonald, P., & Pen, U.-L. 2010, Phys. Rev. D, 81, 103527, doi: 10.1103/PhysRevD.81.103527
- Masui, K. W., Switzer, E. R., Banavar, N., et al. 2013, ApJL, 763, L20, doi: 10.1088/2041-8205/763/1/L20
- Murray, S. 2014, HMF: Halo Mass Function calculator, Astrophysics Source Code Library, record ascl:1412.006. http://ascl.net/1412.006
- Murray, S. G. 2018, Journal of Open Source Software, 3, 850, doi: 10.21105/joss.00850
- Nan, R., Li, D., Jin, C., et al. 2011, International Journal of Modern Physics D, 20, 989

Nelson, D., Pillepich, A., Springel, V., et al. 2018, Monthly Notices of the Royal Astronomical Society, 475, 624

- Newburgh, L. B., Addison, G. E., Amiri, M., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9145, Ground-based and Airborne Telescopes V, ed. L. M. Stepp, R. Gilmozzi, & H. J. Hall, 91454V, doi: 10.1117/12.2056962
- Newburgh, L. B., Bandura, K., Bucher, M. A., et al. 2016, in Proc. SPIE, Vol. 9906, Ground-based and Airborne Telescopes VI, 99065X, doi: 10.1117/12.2234286
- Padmanabhan, H., & Refregier, A. 2017, MNRAS, 464, 4008, doi: 10.1093/mnras/stw2706
- Pillepich, A., Springel, V., Nelson, D., et al. 2018, Monthly Notices of the Royal Astronomical Society, 473, 4077
- Price, D. C. 2016, PyGSM: Python interface to the Global Sky Model, Astrophysics Source Code Library, record ascl:1603.013. http://ascl.net/1603.013
- Ross, A. J., Samushia, L., Howlett, C., et al. 2015, Monthly Notices of the Royal Astronomical Society, 449, 835, doi: 10.1093/mnras/stv154
- Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, MNRAS, 371, 703, doi: 10.1111/j.1365-2966.2006.10699.x
- Santos, M., Bull, P., Alonso, D., et al. 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14), 19. https://arxiv.org/abs/1501.03989
- Santos, M. G., Cluver, M., Hilton, M., et al. 2017, arXiv preprint arXiv:1709.06099
- Schaye, J. 2001, The Astrophysical Journal, 559, 507
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, The Astrophysical Journal, 500, 525
- Springel, V. 2010, Monthly Notices of the Royal Astronomical Society, 401, 791
- Springel, V., Pakmor, R., Pillepich, A., et al. 2017, Monthly Notices of the Royal Astronomical Society, 475, 676, doi: 10.1093/mnras/stx3304
- Tegmark, M. 1997, The Astrophysical Journal Letters, 480, L87

- Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, The Astrophysical Journal, 688, 709
- Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, The Astrophysical Journal, 724, 878, doi: 10.1088/0004-637x/724/2/878
- Ubeda, L., et al. 2011, STScI, Baltimore
- Villaescusa-Navarro, F., Genel, S., Castorina, E., et al. 2018, ApJ, 866, 135, doi: 10.3847/1538-4357/aadba0
- Wang, J., Santos, M. G., Bull, P., et al. 2021, Monthly Notices of the Royal Astronomical Society, 505, 3698, doi: 10.1093/mnras/stab1365
- Weinberger, R., Springel, V., Hernquist, L., et al. 2016, Monthly Notices of the Royal Astronomical Society, 465, 3291
- Wilson, T. L., Rohlfs, K., & Hüttemeister, S. 2013, Tools of Radio Astronomy, doi: 10.1007/978-3-642-39950-3
- Wolz, L., Abdalla, F. B., Blake, C., et al. 2014, MNRAS, 441, 3271, doi: 10.1093/mnras/stu792
- Wolz, L., Blake, C., & Wyithe, J. S. B. 2017, Monthly Notices of the Royal Astronomical Society, 470, 3220, doi: 10.1093/mnras/stx1388
- Wolz, L., Murray, S. G., Blake, C., & Wyithe, J. S. 2019, Monthly Notices of the Royal Astronomical Society, 484, 1007, doi: 10.1093/mnras/sty3142
- Wolz, L., Pourtsidou, A., Masui, K. W., et al. 2022, Monthly Notices of the Royal Astronomical Society, 510, 3495, doi: 10.1093/mnras/stab3621
- Wu, F., Li, J., Zuo, S., et al. 2021, Monthly Notices of the Royal Astronomical Society, 506, 3455, doi: 10.1093/mnras/stab1802
- Xu, Y., Wang, X., & Chen, X. 2015, ApJ, 798, 40, doi: 10.1088/0004-637X/798/1/40
- Zehavi, I., Blanton, M. R., Frieman, J. A., et al. 2002, ApJ, 571, 172, doi: 10.1086/339893
- Zhan, H. 2011, SCIENTIA SINICA Physica, Mechanica & Astronomica, 41, 1441
- Zheng, H., Tegmark, M., Dillon, J. S., et al. 2016, Monthly Notices of the Royal Astronomical Society, 464, 3486, doi: 10.1093/mnras/stw2525
- Zuo, S., Chen, X., Ansari, R., & Lu, Y. 2019, AJ, 157, 4, doi: 10.3847/1538-3881/aaef3b