Determining the HI content of galaxies via intensity mapping cross-correlations

Document Version
Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):
Wolz, L., Blake, C., & Wyithe, J. S. B. (2017). Determining the HI content of galaxies via intensity mapping cross-correlations. *Monthly Notices of the Royal Astronomical Society.* https://doi.org/10.1093/mnras/stx1388

Published in:
Monthly Notices of the Royal Astronomical Society

Determining the HI content of galaxies via intensity mapping cross-correlations

L. Wolz,[1,2]★ C. Blake,[2,3] and J.S.B. Wyithe[1,2]

[1] *School of Physics, University of Melbourne, Parkville, VIC 3010, Australia*
[2] *ARC Centre of Excellence for All-Sky Astrophysics (CAASTRO)*
[3] *Centre for Astrophysics & Supercomputing, Swinburne University of Technology, P.O. Box 218, Hawthorn, VIC 3122, Australia*

ABSTRACT

We propose an innovative method for measuring the neutral hydrogen (HI) content of an optically-selected spectroscopic sample of galaxies through cross-correlation with HI intensity mapping measurements. We show that the HI-galaxy cross-power spectrum contains an additive shot noise term which scales with the average HI brightness temperature of the optically-selected galaxies, allowing constraints to be placed on the average HI mass per galaxy. This approach can estimate the HI content of populations too faint to directly observe through their 21cm emission over a wide range of redshifts. This cross-correlation, as a function of optical luminosity or colour, can be used to derive HI-scaling relations. We demonstrate that this signal will be detectable by cross-correlating upcoming Australian SKA Pathfinder (ASKAP) observations with existing optically-selected samples. We also use semi-analytic simulations to verify that the HI mass can be successfully recovered by our technique in the range $M_{HI} > 10^8 M_\odot$, in a manner independent of the underlying power spectrum shape. We conclude that this method is a powerful tool to study galaxy evolution, which only requires a single intensity mapping dataset to infer complementary HI gas information from existing optical and infra-red observations.

Key words: galaxies: evolution – radio lines: galaxies – large-scale structure of Universe

1 INTRODUCTION

Galaxy evolution is driven by a range of processes including star formation, feedback from supernovae and active galactic nuclei, and environmental effects such as gas stripping and galaxy mergers. Galactic star formation activity is fuelled by cold gas reservoirs and the star formation efficiency is driven by the cooling and cycling mechanism of the gas within the galaxy. Before collapsing into stars, atomic hydrogen (HI) is converted into its molecular phase. Understanding the mechanisms of atomic to molecular gas conversion and their relation to the star formation mechanism is the goal of theorists and observers alike.

In the local Universe ($z < 0.2$), the Cosmic HI density and the HI mass function has been determined by blind HI surveys via their 21cm-line emission down to masses $10^7 M_\odot$ (HIPASS Zwaan et al. 2005, ALFALFA Martin et al. 2010). Targeted surveys have combined HI detections of galaxies with their UV and optical counterparts which estimate the star formation rate (SFR) of the objects. The derived HI scaling relations establish strong relations between the atomic gas fractions and their stellar masses and star forming activities, see e.g. Catinella et al. (2010). Spectral stacking

studies in Brown et al. (2015) suggest that the SFR is the primary driver of the correlations. It has also been shown that there exists a strong environmental dependence of the HI scalings, e.g. in Cortese et al. (2011) and Brown et al. (2017) which is driven by HI stripping processes caused by the dynamics in dense environments (Jaffé et al. 2015). Molecular hydrogen gas is traced via the CO(1-0) line emission and Saintonge et al. (2011, 2016) demonstrated that the molecular gas fraction is strongly correlated with SFR rather than stellar mass of their hosts.

The evolution of neutral gas with cosmic time is still poorly observationally constrained, given that the highest confirmed HI detection is at $z = 0.376$ (Fernández et al. 2016). Therefore, although it is well-established that the cosmic star formation rate peaks at $z \sim 2$ from the combination of UV and IR observations (Madau & Dickinson 2014), it is less clear how the evolution of the SFR is linked to the cold gas in either atomic or molecular form. The HI abundance history is also constrained for redshifts $z < 0.5$ by HI stacking techniques (Rhee et al. 2013; Delhaize et al. 2013; Rhee et al. 2016) and by HI column density measurements from Damped Lyman-alpha systems at higher redshifts (Rao et al. 2006; Prochaska & Wolfe 2009; Meiring et al. 2011; Noterdaeme et al. 2012). Although the high-redshift behaviour of the HI density remains poorly constrained, the observations indicate a relatively low redshift evo-

★ E-mail: lwolz@unimelb.edu.au

lution which is therefore uncorrelated with the SFR history. This suggests that the molecular gas phase, and its correlations, are the key to understanding galaxy evolution across Cosmic time.

The latest numerical simulations include refined models of the cold gas evolution and recipes for the hydrogen partitioning process in both semi-analytic (e.g. Lagos et al. 2011; Popping et al. 2014; Fu et al. 2010) and hydrodynamical approaches such as EAGLE (Bahé et al. 2016; Crain et al. 2016). Simulated global HI properties are in good agreement with $z = 0$ observations, within their volume and resolution limitations. Better higher-redshift observations can test the predicted redshift evolution of atomic and molecular gas in different scenarios, and distinguish between star formation and feedback recipes.

The Square Kilometre Array[1] (SKA), and its pathfinder projects such as the Australian SKA Pathfinder (ASKAP)[2], will provide key insights into HI evolution beyond current redshift limitations within the next decade. A rich variety of 21cm experiments are proposed such as spectral stacking, HI line absorption, HI galaxy surveys and intensity mapping experiments (Santos et al. 2015). Intensity mapping describes observations of the unresolved flux of redshifted 21cm emission on large scales, converted into 3D diffuse maps of the HI density (Battye et al. 2004; Chang et al. 2010). The primary goal of intensity mapping experiments is cosmological analysis to measure the baryon acoustic oscillation scales and the large scale structure fluctuations to constrain current cosmological models (Chang et al. 2008; Wyithe et al. 2008). The intensity maps depend on both the cosmic HI density as a function of redshift, and the bias of the HI density with respect to the underlying structure. The HI intensity mapping signal has been detected in cross-correlation with the WiggleZ Dark Energy Survey at $z \approx 0.8$ in Masui et al. (2013) which also provided the first measurement of the HI density at this redshift (Switzer et al. 2013). Many new intensity mapping experiments are in preparation spanning a wide range of redshifts, such as BINGO (Battye et al. 2012), CHIME (Bandura et al. 2014), and Tianlai (Chen 2012). The cross-correlation with optical galaxies has the advantage of being independent of the systematic effects of the intensity maps caused by astrophysical signal contaminations, instrument uncertainties, and data processing. Therefore, the measurement of the cross-correlation is highest priority in the upcoming intensity mapping projects.

Intensity Mapping datasets provide fundamental information on the global HI distribution such as the clustering properties and the cosmic energy density, however, they are entangled with the cosmological signal such that one can not fully understand one without the other. The cosmological community has identified this unique aspect of intensity mapping experiments by, for example, examining the potential of HI bias measurements in Castorina & Villaescusa-Navarro (2016) and Sarkar et al. (2016). Accurately modelling the HI distribution with respect to the underlying dark matter halos is addressed in e.g. Padmanabhan & Refregier (2017). Cross-correlations of intensity mapping data with galaxy surveys can provide further inside into the HI properties and increase detectability as shown, for instance in Carucci et al. (2016); Hall & Bonvin (2016); Pourtsidou et al. (2016); Sarkar et al. (2016); Wolz et al. (2016).

In this paper, we propose a new interpretation of the cross-correlation signal of intensity maps with galaxy surveys, which uses this signal to measure the average HI content of the optically-selected galaxies. This measurement is not based on the large-scale structure cross-correlation signal but originates from the shot noise term in the power spectrum measurement. The shot noise term is caused by Poisson sampling noise whose amplitude scales as the inverse of the sampled galaxy density. In the case of intensity mapping cross-correlations, the cross-shot noise term is additionally weighted by the average brightness temperature of the galaxy population. This means by measuring the shot noise amplitude on the cross-power spectrum, we can determine the average HI content of the galaxy sample. From one intensity mapping data set, the cross-correlation with multiple galaxy survey catalogues binned into selection criteria, such as colour, size or stellar mass, can supply multiple HI scaling relations. These HI measurements can be obtained for redshifts inaccessible in direct detection by the current generation of telescopes and provide new insights into galaxy evolution mechanisms.

We present a derivation of the HI cross-shot noise term in Sec. 2, and demonstrate the detectability of the signal with upcoming ASKAP observations in Sec. 3. In Sec. 4 we present semi-analytical simulations for creating mock HI intensity maps and colour-selected galaxy catalogues at $z = 0.9$. We use these mock catalogues in Sec. 5 to demonstrate that we can successfully recover the HI shot noise contribution to the cross-power spectra as a function of colour, using a model-independent fitting pipeline. We summarise and conclude in Sec. 6.

2 POWER SPECTRUM FORMALISM

In this section we derive the contribution to the cross-power spectrum of an HI intensity mapping survey and an overlapping optical galaxy redshift survey, that arises from the imprint of the HI content of the optically-selected galaxies in the intensity map. We note that similar expressions have been derived in the context of combining multi-wavelength surveys in Fonseca et al. (2015).

The HI brightness temperature $T_{\mathrm{HI}}(\boldsymbol{x})$ at position $\boldsymbol{x}$, resulting from HI mass density $\rho_{\mathrm{HI}}(\boldsymbol{x})$, is given by

$$T_{\mathrm{HI}}(\boldsymbol{x}) = \frac{3hc^3 A_{10}}{32\pi m_H k_B \nu_{21}^2} \frac{(1+z)^2}{H(z)} \times \rho_{\mathrm{HI}}(\boldsymbol{x}) = C \times \rho_{\mathrm{HI}}(\boldsymbol{x}) \quad (1)$$

in terms of Planck's constant h, the speed of light c, the Einstein coefficient A_{10}, the mass of the hydrogen atom m_H, Boltzmann's constant k_B, the rest-frame frequency of the transition ν_{21}, and the Hubble parameter $H(z)$ at redshift z (e.g. Bull et al. (2015)).

We assume that the HI is contained within N galaxies with number density distribution $n_g(\boldsymbol{x})$ across the observed volume V. We divide the volume into N_c cells with positions $\boldsymbol{x}_i$, such that the number of galaxies in the ith cell is $N(\boldsymbol{x}_i) = n_g(\boldsymbol{x}_i) \times (V/N_c)$. If $M(\boldsymbol{x}_i)$ is the HI mass associated with each galaxy in cell i, the HI brightness temperature is given according to Equ. 1 as

$$T_{\mathrm{HI}}(\boldsymbol{x}_i) = C\, N(\boldsymbol{x}_i)\, M(\boldsymbol{x}_i)\, (N_c/V) \quad (2)$$

We consider the cross-correlation between the galaxy density contrast

$$\delta_g(\boldsymbol{x}_i) = \frac{N(\boldsymbol{x}_i) - \langle N \rangle}{\langle N \rangle} \quad (3)$$

and the brightness over-temperature

$$\delta_T(\boldsymbol{x}_i) = T_{\mathrm{HI}}(\boldsymbol{x}_i) - \langle T_{\mathrm{HI}} \rangle \quad (4)$$

where the angled brackets denote mean values, and $\langle N \rangle = N/N_c$.

Our estimator for the cross-power spectrum $P_\times$ at wavevector $\boldsymbol{k}$ between the gridded galaxy distribution and intensity map is

$$\hat{P}_\times(\boldsymbol{k}) = V \operatorname{Re}\{\tilde{\delta}_g(\boldsymbol{k})\,\tilde{\delta}_T^*(\boldsymbol{k})\} \tag{5}$$

The Fast Fourier Transforms of $N(\boldsymbol{x}_i)$ and $T_{\mathrm{HI}}(\boldsymbol{x}_i)$ are given by

$$\tilde{N}(\boldsymbol{k}) = \sum_i N(\boldsymbol{x}_i)\, e^{i\boldsymbol{k}.\boldsymbol{x}_i} \tag{6}$$

$$\tilde{T}(\boldsymbol{k}) = \sum_i T_{\mathrm{HI}}(\boldsymbol{x}_i)\, e^{i\boldsymbol{k}.\boldsymbol{x}_i} = \frac{C\,N_c}{V} \sum_i N(\boldsymbol{x}_i)\, M(\boldsymbol{x}_i)\, e^{i\boldsymbol{k}.\boldsymbol{x}_i} \tag{7}$$

The expectation value of our cross-power spectrum estimator is then

$$
\begin{aligned}
\langle \hat{P}_\times(\boldsymbol{k}) \rangle &= V \operatorname{Re}\{\langle \tilde{\delta}_g(\boldsymbol{k})\,\tilde{\delta}_T^*(\boldsymbol{k}) \rangle\} \\
&= \frac{C}{N} \sum_{i,j} \langle (N_i - \langle N_i \rangle)\left(N_j M_j - \langle N_j M_j \rangle\right)\rangle\, e^{i\boldsymbol{k}.(\boldsymbol{x}_i - \boldsymbol{x}_j)} \\
&= \frac{C}{N} \sum_{i,j} \left(\langle N_i N_j M_j \rangle - \langle N_i \rangle \langle N_j M_j \rangle\right) e^{i\boldsymbol{k}.(\boldsymbol{x}_i - \boldsymbol{x}_j)} \tag{8}
\end{aligned}
$$

where for brevity we have written $N_i \equiv N(\boldsymbol{x}_i)$ and $M_i \equiv M(\boldsymbol{x}_i)$, and omitted the real part.

For the terms with $j = i$, if we assume that the number of galaxies in each cell is uncorrelated with their HI mass, we have $\langle N_i M_i \rangle = \langle N_i \rangle \langle M_i \rangle$, which allows us to write assuming Poisson statistics:

$$\langle N_i^2 M_i \rangle - \langle N_i \rangle \langle N_i M_i \rangle = \langle N_i M_i \rangle \tag{9}$$

For the terms with $j \neq i$, we define the galaxy-temperature cross-correlation function $\xi_\times$ as

$$\langle N_i N_j M_j \rangle - \langle N_i \rangle \langle N_j M_j \rangle = \langle N_i \rangle \xi_\times(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{10}$$

where $\xi_\times$ can be related to the cross-power spectrum as

$$\xi_\times(\boldsymbol{x}_i, \boldsymbol{x}_j) = \frac{1}{(2\pi)^3} \int P_\times(\boldsymbol{k})\, e^{-i\boldsymbol{k}.(\boldsymbol{x}_i - \boldsymbol{x}_j)}\, d^3\boldsymbol{k} \tag{11}$$

After some algebra we obtain

$$\langle \hat{P}_\times(\boldsymbol{k}) \rangle = P_\times(\boldsymbol{k}) + \frac{C}{N} \sum_i N(\boldsymbol{x}_i)\, M(\boldsymbol{x}_i) \tag{12}$$

such that the estimator depends on both the underlying clustering power spectrum and a shot noise term SN which depends on the HI content of the galaxy sample. This can be re-expressed in terms of the average brightness temperature $\overline{T}_{\mathrm{HI},g}$ generated by the optical galaxies:

$$\mathrm{SN} = \frac{C}{N} \sum_i N(\boldsymbol{x}_i)\, M(\boldsymbol{x}_i) = \frac{\overline{T}_{\mathrm{HI},g}}{n_g} \tag{13}$$

The above equation demonstrates that the cross-shot noise is a direct measure of the brightness temperature which is connected to the HI masses located in the galaxy sample via Equ. 2.

Our models for the auto- and cross-power spectra of the galaxies and HI brightness temperature are defined as

$$
\begin{aligned}
P_g(k) &= b_g(k)^2\, P_m(k) \\
P_T(k) &= \overline{T}_{\mathrm{HI}}^2\, b_{\mathrm{HI}}(k)^2\, P_m(k) \\
P_\times(k) &= b_\times(k)\, \bar{b}_g\, \overline{T}_{\mathrm{HI}}\, \bar{b}_{\mathrm{HI}}\, P_m(k) \tag{14}
\end{aligned}
$$

in terms of the galaxy bias $b_g(k)$, HI bias $b_{\mathrm{HI}}(k)$, mean HI brightness temperature $\overline{T}_{\mathrm{HI}}$, and matter power spectrum $P_m(k)$. The cross-correlation power spectrum is determined by the large-scales

amplitudes of the biases $\bar{b}_{\mathrm{HI}}$ and $\bar{b}_g$, and $b_\times(k)$ is used to express the scale-dependence of the bias. In terms of this model, $\langle \hat{P}_\times \rangle = P_\times + \mathrm{SN}$. For intensity mapping observations, the mean HI brightness temperature can be expressed as a function of the cosmic HI energy density Ω_{HI} as

$$\overline{T}_{\mathrm{HI}} = C \times \Omega_{\mathrm{HI}}\, \rho_{\mathrm{crit}}. \tag{15}$$

We note that the combination $\overline{T}_{\mathrm{HI}}\, b_{\mathrm{HI}}$, equivalent to $\Omega_{\mathrm{HI}} b_{\mathrm{HI}}$, is degenerate in the present formalism. In general, the parameters could be distinguished when considering redshift space distortions (RSD), an effect introduced by the peculiar velocities of the HI-emitting objects which modulates the large scales of the power spectrum. The amplitude of the modulation is proportional to b_{HI} and allows to constrain the global HI density and the HI bias individually, see e.g Wyithe (2008); Masui et al. (2010). In this study, we choose to neglect RSDs as they should be irrelevant for the amplitude shot noise on cross-power spectrum. We note, however, that measuring b_{HI} through RSD of the intensity mapping power spectrum could improve fitting of the cross-correlation coefficient. RSD will be considered in future, more detailed studies of the HI cross-shot-noise fitting.

3 DETECTABILITY FORECAST

We assessed the potential of upcoming surveys to detect the shot noise contribution to the cross-power spectrum by computing a forecast on the detectability if the shot-noise signal. We consider planned observations by the Australian SKA pathfinder (ASKAP) composed of 36 12m-dishes capable of observing a field-of-view of 30 deg with baselines up to 6km. The Deep Investigation of Neutral Gas Origins (DINGO) survey will conduct deep HI observations for $z < 0.4$ with Early Science starting in 2017 aiming to measure HI contents of galaxies via spectral stacking techniques. The overlapping optical galaxy dataset will be provided by the Galaxy and Mass Assembly (GAMA) survey (Driver et al. 2011) in the same regions which provides a variety of galaxy datasets. This survey would provide an ideal test ground to establish our new method.

The noise power spectrum P_N for intensity mapping by an interferometer is derived by McQuinn et al. (2006) and Bull et al. (2015). We write the relation in the form

$$P_N(k_\perp) = \sigma_T^2(k_\perp)\, V_{\mathrm{pix}}\, d^2u/\Omega_{\mathrm{FOV}} \tag{16}$$

where $k_\perp$ is the wavevector component of the power spectrum in the plane of the sky, $d^2u = 2\pi u\, du$ is a 2D pixel in visibility space where $k_\perp = 2\pi u/r$ and r is the co-moving distance to the observed redshift, Ω_{FOV} is the field-of-view of an observation in steradians, and V_{pix} is the 3D volume corresponding to Ω_{FOV} and the frequency width of each channel $\Delta\nu$, which we evaluate as

$$V_{\mathrm{pix}} = r^2\, \Omega_{\mathrm{FOV}}\, r_\nu\, \frac{\Delta\nu}{\nu} \tag{17}$$

where $r_\nu = \frac{c\,(1+z)}{H(z)}$. The temperature variance σ_T^2 in Equ. 16 is given by

$$\sigma_T^2(k_\perp) = \frac{\lambda^2\, T_{\mathrm{sys}}}{A_{\mathrm{dish}} \sqrt{\Delta\nu\, n(u)\, d^2u\, t_{\mathrm{int}}\, N_{\mathrm{beam}}\, N_{\mathrm{pol}}}} \tag{18}$$

where $\lambda = 0.21 \times (1 + z)$ m is the observed wavelength at redshift z, T_{sys} is the system temperature, A_{dish} is the collecting area of each dish, t_{int} is the integration time, the telescope has N_{beam} beams and

N_{pol} polarizations, and $n(u)$ is the number density of baselines in visibility space normalized such that

$$\int n(u)\, d^2u = N_{dish}(N_{dish}-1)/2 \qquad (19)$$

in terms of the total number of dishes N_{dish}. We average the noise power spectrum in spherical k-bins by computing

$$P_N(k) = \int P_N(k_\perp)\, d\Omega_k = \int_0^k \frac{k_\perp\, P_N(k_\perp)}{k\,\sqrt{k^2 - k_\perp^2}}\, dk_\perp \qquad (20)$$

For ASKAP observations we assume a total collecting area $A_{tot} = N_{dish} A_{dish} = 4000$ m^2 composed of $N_{dish} = 36$ dishes with $N_{beam} = 30$ beams and $N_{pol} = 2$ polarizations covering a field-of-view $\Omega_{FOV} = 30$ deg^2, an instrumental system temperature $T_{inst} = 90$ K and frequency channel width $\Delta\nu = 20$ kHz. We consider two components: a "deep" survey mapping total survey area $\Omega_{surv} = 150$ deg^2 and redshift range $0 < z < 0.26$ with total observing time $t_{tot} = 2500$ hrs, and an "ultra-deep" survey mapping $\Omega_{surv} = 60$ deg^2 and $0.1 < z < 0.43$ with $t_{tot} = 5000$ hrs. We take the effective redshift for the noise calculation as the mean redshift in the range. We derive the integration time as

$$t_{int} = t_{tot} \times \frac{\Omega_{FOV}}{\Omega_{surv}} \qquad (21)$$

and obtain the total system temperature after adding in the sky brightness as

$$T_{sys} = T_{inst} + 60\text{K}\left(\frac{300\,\text{MHz}}{\nu}\right)^{2.55} \qquad (22)$$

where $\nu = (1400\,\text{MHz})/(1+z)$ is the observing frequency. We note that the sky temperature term is negligible compared to the receiver temperature of the ASKAP dishes.

The noise in the measurement of the cross-power spectrum $P_\times(k)$ of the intensity map and an overlapping optical spectroscopic redshift survey, in a Fourier bin k of width Δk containing N_{mode} unique Fourier modes, is

$$\sigma(P_\times) = \frac{1}{\sqrt{2\,N_{mode}}}\sqrt{P_\times^2 + (P_T + P_N)\left(P_g + \frac{1}{n_g}\right)} \qquad (23)$$

in terms of the temperature power spectrum $P_T(k)$ and galaxy power spectrum $P_g(k)$, where $1/n_g$ is the galaxy shot noise power in terms of the galaxy number density n_g. We use Equ. 14 to model these power spectra. The number of unique modes is

$$N_{mode} = \frac{V_{surv}}{(2\pi)^3}\, 2\pi k^2\, \Delta k \qquad (24)$$

in terms of the total survey volume V_{surv}, which we evaluate in the survey cone of solid angle Ω_{surv} using a fiducial flat ΛCDM cosmological model with matter density $\Omega_m = 0.3$. For our forecast we assume representative values $b_g = 1$, $b_{HI}\Omega_{HI} = 0.43 \times 10^{-3}$ (Switzer et al. 2013), $b_\times(k) = 1$ and $n_g = 10^{-3}\, h^3$ Mpc^{-3}, which is given as a representative number from the GAMA galaxy number density at z 0.3. We compute the shot noise contribution SN using Equ. 13 assuming that each optical galaxy contains HI mass $M_{HI} = 10^9\, h^{-2}\, M_\odot$ and generate the matter power spectrum $P_m(k)$ using a *Planck* fiducial cosmological model, and assume:

$$\bar{T}_{HI} = 197\,\text{mK} \times \frac{(1+z)^2}{H(z)} \qquad (25)$$

(e.g. Bull et al. (2015)).

We evaluate the forecast in 30 logarithmically spaced k-bins

between $k = 0.1$ and $100\, h$ Mpc^{-1}. We note that the interferometer does not provide any measurements for scales $k < k_{\perp,min} = 2\pi u_{min}/r$, where u_{min} is the minimum visibility. We compute an average ASKAP visibility distribution $n(u)$ given the 2D baseline distribution of the array.

The cross-power spectrum signal $P_\times(k)$ and noise $\sigma(P_\times)(k)$ for these configurations is plotted in Fig. 2. We see that the ASKAP-DINGO intensity-mapping observations, cross-correlated with an optical survey such as GAMA, provide accurate measurements of the cross-power spectrum and the ability to detect the cross-shot noise signal with high significance for small scale modes $k > 10\, h$ Mpc^{-1} where the shot noise is dominating the observed signal, such that we do not require to model the underlying clustering power spectra.

The errors in Equ. 23 on the cross-power spectrum primarily depend on the observed number of modes, i.e. the survey volume, as well as noise on the intensity maps. By increasing the survey volume, N_{mode} would be increased, however, with constant observing time, this would increase the instrumental noise. The HI cross-shot noise is best measured on the small scales of the power spectrum where the cosmological signal falls off and a sufficient number of modes can be sampled in a small volume. Therefore, a deep survey with lower receiver noise on the intensity maps is preferable over a big survey volume. Equ. 23 also incorporates a galaxy shot noise factor, which implies that the errors on the cross-power spectrum increase for smaller galaxy densities but, in turn, the amplitude of cross-shot noise is likewise increased by the same factor since it scales at the inverse of the galaxy density. Therefore, a smaller galaxy sample could theoretically boost the measurability of the cross-shot noise as long as the shot-noise term is larger than Equ. 23. However, one should keep in mind, that we measure a statistical average of the mean HI mass of galaxies and that a smaller sample could be less representative for the selected galaxy type. We conclude that the critical parameters when aiming to measure the cross-shot noise is the ratio of the cross-shot noise to the cosmological signal and the sensitivity of the radio interferometer to these scales.

4   SIMULATIONS

For illustration of our detection technique, we employ the simulation presented in Wolz et al. (2016) which is based on the semi-analytic models used for SAGE, see Croton et al. (2016). The simulation box is of size $(500\,\text{Mpc}/h)^3$ with particle mass resolution of $8.6 \times 10^8 h^{-1} M_\odot$. The star-formation model follows the Blitz-Rosolowsky recipe, see Blitz & Rosolowsky (2006) for details. In summary, the star formation activity depends on the gas surface density, the molecular hydrogen fraction and the molecular gas depletion time. In this model, the partitioning of the hydrogen in atomic and molecular states is determined by the hydrostatic pressure of the galaxy derived by stellar surface density, gas surface density and the gas velocity dispersion. For a detailed description of the simulations and a study of their galaxy properties and intensity mapping constraints, we refer the reader to Wolz et al. (2016).

We showcase the HI measurement concept using a simulation snapshot at $z = 0.905$, which is the target epoch of intensity mapping experiments by the GBT, Parkes Telescope and potentially ASKAP. Direct HI measurements at these redshifts are not obtainable with current technologies. For the purposes of this test, we study the clustering of only central galaxies, removing all satellite galaxies from the simulation for both probes, the HI intensity maps and the optical samples. The presence of sub-structure complicates the
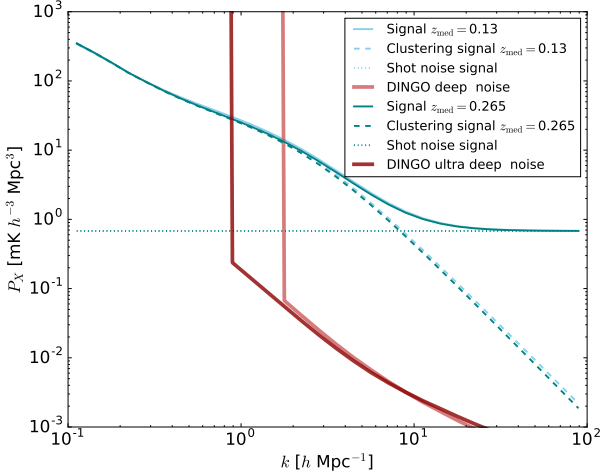
Figure 1. The forecast for a intensity mapping cross-power spectrum measurement between the ASKAP-DINGO survey and an overlapping optical survey, for the "deep" and "ultra-deep" DINGO configurations. The HI cross-shot noise contribution is marked as the dotted line assuming an average HI mass $10^9\,h^{-2}\,M_\odot$ per optical galaxy. The noise in the cross-power spectrum is marked as the red solid lines, which are up to three orders of magnitudes smaller than the expected shot noise signal for both DINGO surveys.

cross-shot noise relations, and we will present a full treatment of these effects in a follow-up study.

We generated the mock datasets as follows:

- **Intensity Maps**

We bin all HI-emitting objects in the simulation box in a grid with resolution $N_{pix} = 512^3$ and convert the HI mass in each grid cell into a brightness temperature using Equ. 1. The resulting intensity maps are in units of K. In order to perform an accurate test of the theory, we neglect telescope effects.

- **Galaxy Samples**

We simulate the galaxy magnitudes in optical and UV filters by applying the photometric mock galaxy code described by Tonini et al. (2012) to the semi-analytic simulation output. We cross-correlated mock galaxy samples with the intensity map in bins of $(NUV - r)$ colour, using the GALEX NUV band and the SDSS optical r-band. The $(NUV - r)$ colour is a measure of the SFR. Thus the derived scaling relation can be interpreted as the average HI mass as a function of SFR. The galaxy sample can alternatively be binned using many other properties such as luminosity, stellar mass, halo mass, environment, molecular content, metallicity or morphology, depending on the available data. We split the galaxy population by $(NUV - r)$ colour in bins with divisions at $\{-1, 0, 1, 2, 3, 4, 5, 6\}$, defining samples from blue star-forming to red quiescent galaxies. Our selection process results in galaxy samples with $N_g = \{774889, 6214418, 694146, 263815, 187110, 662317, 965585\}$ number of objects.

5 HI CROSS SHOT NOISE MEASUREMENT

In this section, we present the application of the theoretical concept to the semi-analytic simulation and demonstrate an empirical approach to perform the shot noise measurement from the cross-power

spectrum in a manner independent of the underlying cosmological power spectrum by employing the cross-correlation coefficient.

The cross-correlation shot noise is a scale-independent additive term to the cross-power spectrum. It can be straightforwardly determined from $\hat{P}_\times(k)$ if the amplitude of the shot noise is high enough to dominate the signal and/or the cross-power is measured on small enough scales where the cosmological signal becomes negligible compared to the shot noise. For any other cases, the estimation of the amplitude of the cross-shot noise directly from the cross-power spectrum requires knowledge of the underlying cosmology and bias.

In this study, we demonstrate the measurement of the shot noise from the observed cross-correlation coefficient $\hat{r}(k)$ to eliminate the dependence on the underlying clustering signal and, hence, the model dependence. We employ the measured power spectrum of the intensity map, $P_T(k)$, the galaxy power spectra for the N colour cuts labelled by i, $P_{g,i}(k)$, and the cross-power spectra of the galaxy samples and intensity map, $\hat{P}_{\times,i}(k)$ on scales $k < 1.5\,h\,\mathrm{Mpc}^{-1}$.

The observed cross-correlation coefficient $\hat{r}(k)$ includes a shot noise term and can be written as

$$\hat{r}_i(k) = \frac{\hat{P}_{\times,i}(k)}{\sqrt{P_T(k)\,P_{g,i}(k)}} = \frac{P_{\times,i}(k) + \mathrm{SN}_i}{\sqrt{P_T(k)\,P_{g,i}(k)}} = r_i(k) + \frac{\mathrm{SN}_i}{\sqrt{P_T(k)\,P_{g,i}(k)}} \tag{26}$$

Using Equ. 14, we can rewrite the cross-correlation coefficient $r_i(k)$ as

$$r_i(k) = \frac{b_{\times,i}(k)\bar{b}_{g,i}\bar{b}_{HI}}{b_{HI}(k)b_{g,i}(k)} \tag{27}$$

In order to measure the shot noise contribution from $\hat{r}_i(k)$, we require a model for the scale-dependence of $r_i(k)$.

If the HI and galaxy bias were both scale-independent, the cross-correlation coefficient could be expected to be only mildly scale-dependent, for instance, caused by environmental effects and to converge to 1 on large scales since the amplitudes of the biases $\bar{b}_{HI}$ and $\bar{b}_{gal}$ cancel out. Any excess power above 1 measured in $\hat{r}_i(k)$ would be due to the shot noise term.

In the case of scale-dependent bias, we need to determine the scale-dependence of the quotient $b_\times(k)/(b_{HI}(k)b_g(k))$.

For illustrative purposes, let us consider the cross-correlation where the intensity maps have scale-independent bias but the galaxy sample is scale-dependently biased on large scales. This would, however, result in no upward bias in the cross-correlation of both probes since the increased power of the large scale Fourier modes is not present in the intensity maps and therefore is not correlated. In this case, the scale dependency of the cross-correlation bias would follow $b_\times(k) \sim b_{HI}(k)$ and the shape of the cross-correlation coefficient could be approximated as $r(k) \sim 1/b_g(k)$.

In more general terms, the functional form of the bias $b_\times(k)$ of the cross-correlation power spectrum follows the scale-dependence of the less dominantly biased power of both probes. The shape of the resulting cross-correlation coefficient $r(k)$ is driven by the inverse of the bias which is not reflected in the cross-correlation bias, i.e. the inverse of the more dominant scale-dependent power of the bias.

In our simulations, after removal of the satellites, the galaxy bias is fairly scale-independent whereas the HI intensity maps exhibit a strong scale-dependent bias on smaller scales $k > 0.5\,h$ Mpc^{-1} which is caused by the increased clustering of HI-rich objects in dense regions predicted by this star formation recipe. This implies that $b_\times(k) \sim b_g(k)$ and, therefore, the shape of the cross-correlation coefficient $r_i(k)$ is proportional to $1/b_{HI}(k)$. For this case, we can assume that $r_i(k)$ is independent of the galaxy sample
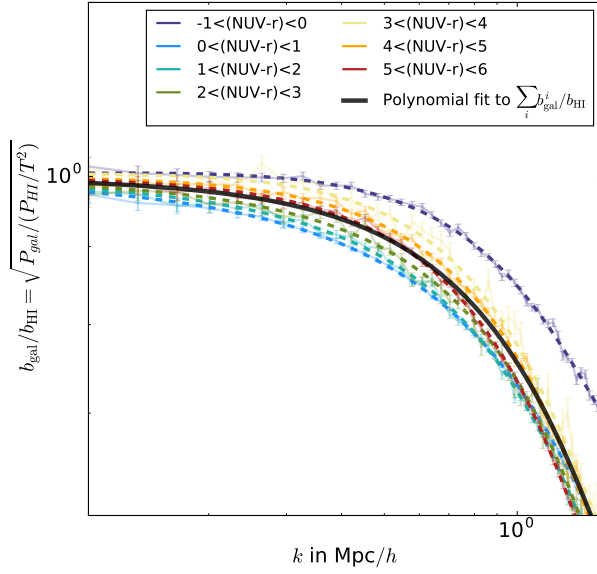
Figure 2. We show the quotient of the galaxy bias and the HI bias determined by taking the square-root of the galaxy power spectrum divided by the intensity mapping power spectrum $\sqrt{\hat{P}_{g,i}(k)/\hat{P}_T(k)}$. The different colours refer to the power spectra selected by galaxy colour $(NUV-r)$ which is an indicator for the star-formation rate of the sample, from blue, star-forming to red, quiescent objects. The dashed coloured lines indicate the best-fit polynomial to the quotient of each individual sample and the solid black line marks the best-fit polynomial to the averaged bias quotient.

i and converges to the same scale-dependent form $\bar{r}(k) \sim 1/b_{HI}(k)$. In the following, we present an empirical, model independent technique to approximate $\bar{r}(k)$ using the auto-power spectra.

We estimate the functional scale-dependence of $\bar{r}(k) \sim 1/b_{HI}(k)$ using $\sqrt{P_{g,i}(k)/P_T(k)}$ which is equal to $B_i(k) = b_{g,i}(k)/b_{HI}(k)$ for each population i. In Fig. 2, we show $B_i(k)$ for each colour bin i where each curve has been re-normalised to 1 on large scales. It can be seen that the functional form of the quotient is preserved over the different colour cuts i within a small variance. This confirms that the galaxy bias can be assumed as scale-independent though the amplitude varies considerably with colour cut. We perform a polynomial fit $f(k) = \sum_j^N p_j k^j$ with $N = 3$ to the average of all bias quotients $\bar{B} = \sum_i B_i$ and employ the resulting coefficients p_1, p_2, p_3 to determine the scale-dependence of the model for $\bar{r}(k)$. The resulting best fit $f_{bf}(k)$ is shown as the black curve in Fig. 2. Note that we set $p_0 = 1$ since an overall normalization factor has already been removed.

In Fig. 3, we show the cross-correlation coefficient $\hat{r}_i(k)$ for all colour cuts i measured as in Equ. 26 as the solid lines. We determine the cross-shot noise by fitting the observed $\hat{r}_i(k)$ to

$$F_i(k) = f_{bf}(k) + \frac{\hat{SN}_i}{\sqrt{P_{g,i}(k)P_T(k)}}, \qquad (28)$$

where $f_{bf}(k)$ models $\bar{r}(k)$. The result of the best fit $F_i(k)$ is plotted in Fig. 3 as the dashed lines. We can see that the best-fit model and measurement are in good agreement. For red galaxy samples with $NUV - r > 4$, the cross-shot noise amplitude is extremely small since galaxies in these bins are HI-poor which results in a small brightness temperature.

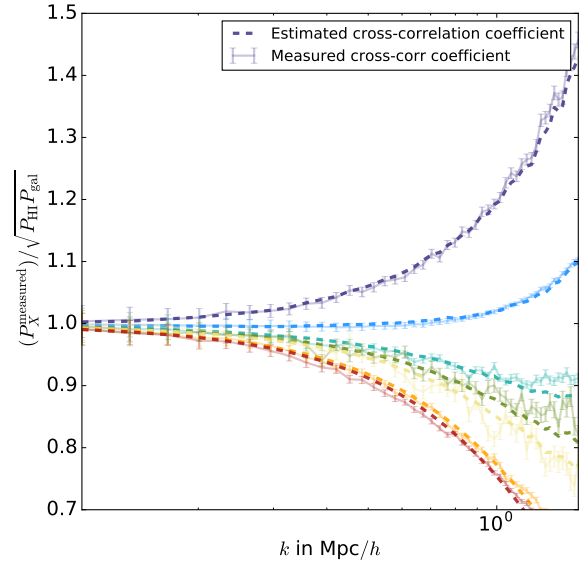We can convert the fitted cross-shot noise $\hat{SN}_i$ into an average

Figure 3. The cross-correlation coefficient $\hat{r}(k)$ of the intensity maps and the different galaxy samples. The colours from blue to red indicate the colour bins in $\Delta(NUV - r) = 1$ from -1 to 6. The splitting of the measurements seen in this plot is caused by the HI cross-shot noise $SN = \bar{T}_i/n_g$. The dashed lines are the the best-fit of the polynomial plus the estimated shot noise parameter $\hat{SN}/\sqrt{P_g(k)P_T(k)}$.
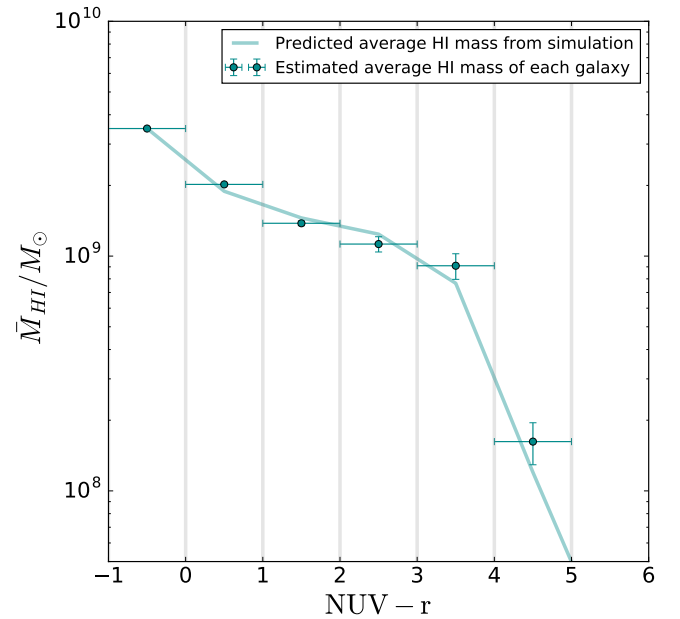

Figure 4. The estimated average HI mass per galaxy as a function of the colour $(NUV - r)$ of the population marked as the circles compared to the simulated model marked with the dashed lines. The errors are given by the model uncertainties of the polynomial fitted to the bias quotient and the variance given by each individual shot noise fit.

HI mass per galaxy for each sample i using $SN_i = \overline{T}_{HI,g,i}/n_{g,i}$ in combination with Equ. 2. The results are shown in Fig. 4. The fitted average HI mass $\overline{M}_{HI,i}$ are marked in symbols and the true values measured directly from the simulation box are given as the solid line. The error bars on this measurement are composed of the standard error given by the weighted least-square fit of the measurement to the model, in addition to the error on the model itself which can be estimated from the covariance of the polynomial fit. The method performs very well over two orders of magnitude for HI masses above $10^8 M_\odot$. Within the current pipeline, it is not possible to extend HI mass measurements beyond this threshold and for the reddest colour bin, we can not determine an upper limit on the HI mass. This is due to the fact that the HI cross-shot noise lies within the variance of the polynomial fit itself.

6   CONCLUSIONS

In this study, we verify the possibility of determining HI scaling relations for high redshifts using HI intensity mapping cross-correlations with galaxy populations. We show that the shot noise term of the cross-correlation power spectrum is proportional to the average HI mass of the galaxy population. We forecast the detectability of the HI cross-shot noise term of the cross-power spectrum for up-coming ASKAP DINGO observations. We use semi-analytic simulations to demonstrate the measurement of the HI cross-shot noise for a range of different galaxy samples. Our model-independent fitting pipeline can detect the shot noise on scales $k < 1.5\,h\,\mathrm{Mpc}^{-1}$ for HI masses above $10^8 M_\odot$. Shot noise measurements for $k > 1.5\,h\,\mathrm{Mpc}^{-1}$ are less sensitive to the fitting method since the noise amplitude dominates the cosmological signal on smaller scales.

In future work, we plan to extend our studies to satellite galaxies, investigating the effect of the central-satellite term on the HI cross-shot noise, within a full halo occupation formalism. We will also develop a more sophisticated method to estimate the scale-dependence of the cross-correlation coefficient by accurate modelling of the HI and galaxy bias terms from their respective auto-power spectrum.

We conclude that the HI measurements via intensity mapping cross-correlations have a strong potential to deliver a variety of new constraints for galaxy evolution models at high redshifts. The HI shot noise measurements provide a complimentary use of the cross-correlation power spectrum to cosmological tests, independent of the systematic challenges the first generation of intensity mapping experiments are facing.

This paper has been typeset from a TEX/LATEX file prepared by the author.

REFERENCES

Bahé Y. M., Crain R. A., Kauffmann G., Bower R. G., Schaye J., Furlong M., Lagos C., Schaller M., Trayford J. W., Dalla Vecchia C., Theuns T., 2016, MNRAS, 456, 1115

Bandura K., Addison G. E., Amiri M., Bond J. R., Campbell-Wilson D., Connor L., Cliche J.-F., Davis G., Deng M., Denman N., Dobbs M., Fandino M., Gibbs K., Gilbert A., Halpern M., et al. 2014, 9145, 22

Battye R. A., Brown M. L., Browne I. W. A., Davis R. J., Dewdney P., Dickinson C., Heron G., Maffei B., Pourtsidou A., Wilkinson P. N., 2012, MNRAS, astro-ph.CO

Battye R. A., Davies R. D., Weller J., 2004, MNRAS, 355, 1339

Blitz L., Rosolowsky E., 2006, ApJ, 650, 933

Brown T., Catinella B., Cortese L., Kilborn V., Haynes M. P., Giovanelli R., 2015, MNRAS, 452, 2479

Brown T., Catinella B., Cortese L., Lagos C. d. P., Davé R., Kilborn V., Haynes M. P., Giovanelli R., Rafieferantsoa M., 2017, MNRAS, 466, 1275

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

Carucci I. P., Villaescusa-Navarro F., Viel M., 2016, arXiv.org

Castorina E., Villaescusa-Navarro F., 2016, arXiv.org, p. arXiv:1609.05157

Catinella B., Schiminovich D., Kauffmann G., Fabello S., Wang J., Hummels C., Lemonias J., Moran S. M., Wu R., Giovanelli R., Haynes M. P., Heckman T. M., Basu-Zych A. R., et al., 2010, MNRAS, 403, 683

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, MNRAS, 100, 091303

Chen X., 2012, MNRAS, 12, 256

Cortese L., Catinella B., Boissier S., Boselli A., Heinis S., 2011, MNRAS, 415, 1797

Crain R. A., Bahe Y. M., Lagos C. d. P., Rahmati A., Schaye J., McCarthy I. G., Marasco A., Bower R. G., Schaller M., Theuns T., van der Hulst T., 2016, ArXiv e-prints

Croton D. J., Stevens A. R. H., Tonini C., Garel T., Bernyk M., Bibiano A., Hodkinson L., Mutch S. J., Poole G. B., Shattow G. M., 2016, ApJS, 222, 22

Delhaize J., Meyer M. J., Staveley-Smith L., Boyle B. J., 2013, MNRAS, 433, 1398

Driver S. P., Hill D. T., Kelvin L. S., Robotham A. S. G., Liske J., Norberg P., Baldry I. K., Bamford S. P., Hopkins A. M., Loveday J., Peacock J. A., Andrae E., Bland-Hawthorn J., et al., 2011, MNRAS, 413, 971

Fernández X., Gim H. B., van Gorkom J. H., Yun M. S., Momjian E., Popping A., Chomiuk L., Hess K. M., Hunt L., Kreckel K., Lucero D., Maddox N., Oosterloo T., et al., 2016, ApJ, 824, L1

Fonseca J., Camera S., Santos M. G., Maartens R., 2015, arXiv.org

Fu J., Guo Q., Kauffmann G., Krumholz M. R., 2010, MNRAS, 409, 515

Hall A., Bonvin C., 2016, arXiv.org

Jaffé Y. L., Smith R., Candlish G. N., Poggianti B. M., Sheen Y.-K., Verheijen M. A. W., 2015, MNRAS, 448, 1715

Lagos C. D. P., Baugh C. M., Lacey C. G., Benson A. J., Kim H.-S., Power C., 2011, MNRAS, 418, 1649

Madau P., Dickinson M., 2014, ARA&A, 52, 415

Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, ApJ, 723, 1359

Masui K. W., McDonald P., Pen U.-L., 2010, Physical Review D, 81, 103527

Masui K. W., Switzer E. R., Banavar N., K B., Blake C., Calin L. M., Chang T. C., Chen X., Li Y. C., Liao Y. W., Natarajan A., Pen U. L., Peterson J. B., Shaw J. R., Voytek T. C., 2013, MNRAS, 763, L20

McQuinn M., Zahn O., Zaldarriaga M., Hernquist L., Furlanetto S. R., 2006, ApJ, 653, 815

Meiring J. D., Tripp T. M., Prochaska J. X., Tumlinson J., Werk J., Jenkins E. B., Thom C., O'Meara J. M., Sembach K. R., 2011, ApJ, 732, 35

Noterdaeme P., Petitjean P., Carithers W. C., Pâris I., Font-Ribera A., Bailey S., Aubourg E., Bizyaev D., Ebelke G., Finley H., Ge J., Malanushenko E., Malanushenko V., et al., 2012, A&A, 547, L1

Padmanabhan H., Refregier A., 2017, Monthly Notices of the Royal Astronomical Society, 464, 4008

Popping G., Somerville R. S., Trager S. C., 2014, MNRAS, 442, 2398

Pourtsidou A., Bacon D., Crittenden R., Metcalf R. B., 2016, Monthly Notices of the Royal Astronomical Society, 459, 863

Prochaska J. X., Wolfe A. M., 2009, ApJ, 696, 1543

Rao S. M., Turnshek D. A., Nestor D. B., 2006, ApJ, 636, 610

Rhee J., Lah P., Chengalur J. N., Briggs F. H., Colless M., 2016, MNRAS, 460, 2675

Rhee J., Zwaan M. A., Briggs F. H., Chengalur J. N., Lah P., Oosterloo T., van der Hulst T., 2013, MNRAS, 435, 2693

Saintonge A., Catinella B., Cortese L., Genzel R., Giovanelli R., Haynes M. P., Janowiecki S., Kramer C., Lutz K. A., Schiminovich D., Tacconi L. J., Wuyts S., Accurso G., 2016, MNRAS, 462, 1749

Saintonge A., Kauffmann G., Kramer C., Tacconi L. J., Buchbender C., Catinella B., Fabello S., Graciá-Carpio J., Wang J., Cortese L., Fu J., Genzel R., Giovanelli R., Guo Q., Haynes M. P., et al., 2011, MNRAS, 415, 32

Santos M., Bull P., Alonso D., Camera S., Ferreira P., Bernardi G., Maartens R., Viel M., Villaescusa-Navarro F., Abdalla F. B., Jarvis M., Metcalf R. B., Pourtsidou A., Wolz L., 2015, Advancing Astrophysics with the Square Kilometre Array (AASKA14), p. 19

Sarkar D., Bharadwaj S., S A., 2016, arXiv.org

Sarkar T. G., Datta K. K., Pal A. K., Choudhury T. R., Bharadwaj S., 2016, arXiv.org

Switzer E. R., Masui K. W., Bandura K., Calin L. M., Chang T. C., Chen X. L., Li Y. C., Liao Y. W., Natarajan A., Pen U. L., Peterson J. B., Shaw J. R., Voytek T. C., 2013, MNRAS, 434, L46

Tonini C., Bernyk M., Croton D., Maraston C., Thomas D., 2012, ApJ, 759, 43

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

Wyithe J. S. B., 2008, MNRAS, 388, 1889

Wyithe S., Loeb A., Geil P., 2008, MNRAS, 383, 1195

Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, MNRAS, 359, L30