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

L. Wolz,1,2 C. Blake2,3 and J. S. B. Wyithe1,2

1School of Physics, University of Melbourne, Parkville, VIC 3010, Australia
2ARC Centre of Excellence for All-Sky Astrophysics (CAASTRO), University of Melbourne, Parkville, VIC 3010, Australia
3Centre for Astrophysics & Supercomputing, Swinburne University of Technology, PO 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 that 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 21-cm 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 Square Kilometre Array Pathfinder 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 $M_{\text{HI}}>10^8 M_\odot$, 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 data set to infer complementary HI gas information from existing optical and infrared observations.

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

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 have been determined by blind HI surveys via their 21-cm 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 ultraviolet (UV) and optical counterparts that 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 SFR peaks at $z\sim2$ from the combination of UV and infrared 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 (Delhaize et al. 2013; Rhee et al. 2013, 2016) and by HI column density measurements from Damped Lyman-alpha systems at higher redshifts (Rao, Turnshek & Nestor 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 evolution that 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. Fu et al. 2010; Lagos et al. 2011; Poping, Somerville & Trager 2014) 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 21-cm 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 21-cm emission on large scales, converted into 3D diffuse maps of the HI density (Battye, Davies & Weller 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, Loeb & Geil 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 in data taken by the Green Bank Telescope (GBT) 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. 2013), CHIME (Bandura et al. 2014) and Tianlai (Chen 2012). The cross-correlation with optical galaxies provides further insights into the HI properties of the next decade. A rich variety of 21-cm experiments are proposed such as spectral stacking, HI line absorption, HI galaxy surveys and intensity mapping experiments are in preparation spanning a wide range of redshifts, such as EAGLE (Bahé et al. 2016; Crain et al. 2016). Simulated 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.

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 that 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 Section 2 and demonstrate the detectability of the signal with upcoming ASKAP observations in Section 3. In Section 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 Section 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 summarize and conclude in Section 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 multwavelength surveys in Fonseca et al. (2015).

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

\[
T_{\text{HI}}(x) = \frac{3}{2} \frac{\hbar^3 A_{10} (1+z)^2}{2\pi m_{\text{H}} k_B T(z)} \times \rho_{\text{HI}}(x) = C \times \rho_{\text{HI}}(x)
\]

in terms of Planck constant \( \hbar \), the speed of light \( c \), the Einstein coefficient \( A_{10} \), the mass of the hydrogen atom \( m_{\text{H}} \), Boltzmann constant \( k_B \), the rest-frame frequency of the transition \( \nu_{21} \) 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 \( n_i(x) \) across the observed volume \( V \). We divide the volume into \( N_i \) cells with positions \( x_i \) such that the number of galaxies in the \( i \)-th cell is \( N_i(x_i) = n_i(x_i) \times (V/N_i) \). If \( M_i(x_i) \) is the HI mass associated with each galaxy in cell \( i \), the HI brightness temperature is given according to equation (1) as

\[
T_{\text{HI}}(x_i) = C \times N_i(x_i) M_i(x_i) (N_i/V).
\]

We consider the cross-correlation between the galaxy density contrast

\[
\delta_g(x_i) = \frac{N_i(x_i) - \langle N_i \rangle}{\langle N_i \rangle}
\]

and the brightness overheating

\[
\delta_T(x_i) = T_{\text{HI}}(x_i) - \langle T_{\text{HI}} \rangle,
\]

\cite{10.1093/mnras/stx270}
where the angled brackets denote mean values and \( \langle N \rangle = N/N_\circ \). Our estimator for the cross-power spectrum \( P_s \) at wavevector \( k \) between the gridded galaxy distribution and intensity map is

\[
\hat{P}_s(k) = V \text{Re} \{ \bar{\delta}_g(k) \bar{\delta}_T(k) \}.
\]

(5)

The Fast Fourier Transforms of \( N(x_i) \) and \( T_{HI}(x_i) \) are given by

\[
\hat{N}(k) = \sum_i N(x_i) e^{ikx_i},
\]

(6)

\[
\hat{T}(k) = \sum_i T_{HI}(x_i) e^{ikx_i} = C \frac{N_c}{V} \sum_i N(x_i) M(x_i) e^{ikx_i},
\]

(7)

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

\[
\langle \hat{P}_s(k) \rangle = V \text{Re} \{ \bar{\delta}_g(k) \bar{\delta}_T(k) \}
\]

\[
= C \sum_{i,j} \langle (N_i - \langle N_i \rangle) (N_j - \langle N_j \rangle) \rangle e^{ik(x_i - x_j)}
\]

\[
= C \sum_{i,j} \langle (N_i M_j - \langle N_i \rangle \langle M_j \rangle) \rangle e^{ik(x_i - x_j)},
\]

(8)

where \( \langle N_i \rangle \equiv N(x_i) \) and \( \langle M_i \rangle \equiv M(x_i) \) 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 \( H I \) mass, we have \( \langle N_i M_j \rangle = \langle N_i \rangle \langle M_i \rangle \), which allows us to write assuming Poisson statistics:

\[
\langle N_i^2 M_j \rangle - \langle N_i \rangle \langle M_i \rangle = \langle N_i M_i \rangle.
\]

(9)

For the terms with \( j \neq i \), we define the galaxy-temperature cross-correlation function \( \xi_{x,T} \) as

\[
\langle (N_i N_j M_j - \langle N_i \rangle \langle N_j M_j \rangle) \rangle = \langle N_i \rangle \xi_{x}(x_i, x_j),
\]

(10)

where \( \xi_{x} \) can be related to the cross-power spectrum as

\[
\xi_{x}(x_i, x_j) = \frac{1}{(2\pi)^3} \int P_x(k) e^{-i k(x_i - x_j)} \, d^3 k.
\]

(11)

After some algebraic calculations we obtain

\[
\langle \hat{P}_s(k) \rangle = P_s(k) + C \sum_i N(x_i) M(x_i)
\]

(12)

such that the estimator depends on both the underlying clustering power spectrum and a shot noise term \( SN \) which depends on the \( H I \) content of the galaxy sample. This can be re-expressed in terms of the average brightness temperature \( T_{HI,\circ} \) generated by the optical galaxies:

\[
SN = C \sum_i N(x_i) M(x_i) = \frac{T_{HI,\circ}}{\bar{\rho}_{HI}}.
\]

(13)

Note that in the context of this paper, we use the SN as an abbreviation for shot noise. The above equation demonstrates that the cross-shot noise is a direct measure of the brightness temperature that is connected to the \( H I \) masses located in the galaxy sample via equation (2).

Our models for the auto- and cross-power spectra of the galaxies and \( H I \) brightness temperature are defined as

\[
P_g(k) = b_g(k)^2 P_m(k)
\]

\[
P_T(k) = T_{HI}^2 b_{HI}(k)^2 P_m(k)
\]

\[
P_s(k) = b_s(k) \bar{b}_g T_{HI} b_{HI} P_m(k)
\]

(14)

in terms of the galaxy bias \( b_g(k), H I \) bias \( b_{HI}(k), \) mean \( H I \) brightness temperature \( T_{HI} \), and matter power spectrum \( P_m(k) \). The cross-correlation power spectrum is determined by the large-scale amplitudes of the biases \( b_{HI} \) and \( b_g \), and \( b_s(k) \) is used to express the scale dependence of the bias. In terms of this model, \( \langle \hat{P}_s \rangle = P_s + SN \).

For intensity mapping observations, the mean \( H I \) brightness temperature can be expressed as a function of the cosmic \( H I \) energy density \( \Omega_{HI} \) as

\[
\bar{T}_{HI} = C \times \Omega_{HI} \rho_{HI}.
\]

(15)

We note that the combination \( T_{HI} b_{HI} \), equivalent to \( \Omega_{HI} b_{HI} \), is degenerate in the present formalism. In general, the parameters could be distinguished when considering redshift space distortions (RSDs), an effect introduced by the peculiar velocities of the \( H I \)-emitting objects that modulates the large scales of the power spectrum. The amplitude of the modulation is proportional to \( b_{HI} \) and allows to constrain the global \( H I \) density and the \( H I \) bias individually; see e.g. Wyithe (2008) and Masui, McDonald & Pen (2010). In this study, we choose to neglect RSDs as they should be irrelevant for the amplitude of the shot noise on the 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 \( H I \) 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 ASKAP composed of thirty-six 12-m dishes capable of observing a field of view of 30 deg with baselines up to 6 km. The Deep Investigation of Neutral Gas Origins (DINGO) survey will conduct deep \( H I \) observations for \( z < 0.4 \) with Early Science starting in 2017 aiming to measure \( H I \) contents of galaxies via spectral stacking techniques. The overlapping optical galaxy data set 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 data sets. 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

\[
P_N(k, \lambda) = \sigma^2_T(k, \lambda) V_{pix} d^2 u / \Omega_{ODV}
\]

(16)

where \( k_L \) is the wavevector component of the power spectrum in the plane of the sky, \( d^2 u = 2\pi \Delta v \, du \) is a 2D pixel in visibility space where \( k_L = 2\pi u/r \) and \( r \) is the comoving distance to the observed redshift, \( \Omega_{ODV} \) is the field of view of an observation in steradians and \( V_{pix} \) is the 3D volume corresponding to \( \Omega_{ODV} \) and the frequency width of each channel \( \Delta \nu \), which we evaluate as

\[
V_{pix} = r^2 \Omega_{ODV} r_s \Delta \nu / \nu.
\]

(17)

where \( r_s = c \Delta \nu / \lambda^2 \). The temperature variance \( \sigma^2_T \) in equation (16) is given by

\[
\sigma^2_T(k, \lambda) = \frac{\lambda^2 T_{sys}}{A_{dish} \sqrt{\Delta \nu n(u) du \int \bar{N}_{beam} N_{pol}}}.
\]

(18)

where \( \lambda = 0.21 \times (1 + z) \) is the observed wavelength (in metres) 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 $N_{\text{pol}}$ polarizations and $n(u)$ is the number density of baselines in visibility space normalized such that

$$\int n(u) \, du = N_{\text{dish}}(N_{\text{dish}} - 1)/2$$

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

$$P_\nu(k) = \int P_\nu(k_x) \, d\Omega_x = \int \frac{k_x n_x}{k \sqrt{k^2 - k_x^2}} \, dk_x$$

(20)

For ASKAP observations we assume a total collecting area $A_{\text{tot}} = N_{\text{dish}} A_{\text{dish}} = 4000$ m² composed of $N_{\text{dish}} = 36$ dishes with $N_{\text{beam}} = 30$ beams and $N_{\text{pol}} = 2$ polarizations covering a field-of-view $\Omega_{\text{FOV}} = 30$ deg², an instrumental system temperature $T_{\text{inst}} = 90$ K and frequency channel width $\Delta \nu = 20$ kHz. We consider two components: a ‘deep’ survey mapping total survey area $\Omega_{\text{surv}} = 150$ deg² and redshift range $0 < z < 0.26$ with total observing time $t_{\text{tot}} = 2500$ h, and an ‘ultra-deep’ survey mapping $\Omega_{\text{surv}} = 60$ deg² and $0.1 < z < 0.43$ with $t_{\text{tot}} = 5000$ h. We take the effective redshift for the noise calculation as the mean redshift in the range. We derive the integration time as

$$t_{\text{int}} = t_{\text{tot}} \frac{\Omega_{\text{FOV}}}{\Omega_{\text{surv}}}$$

(21)

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

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

(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 $P_{g}(k)$ of the intensity map and an overlapping optical spectroscopic redshift survey, in a Fourier bin $k$ of width $\Delta k$ containing $N_{\text{mode}}$ unique Fourier modes, is

$$\sigma(P_g) = \sqrt{\frac{1}{2 N_{\text{mode}}} \left( P_x^2 + (P_T + P_N) \right) \left( \frac{1}{n_g} \right)}$$

(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 equation (14) to model these power spectra. The number of unique modes is

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

(24)

in terms of the total survey volume $V_{\text{surv}}$, which we evaluate in the survey cone of solid angle $\Omega_{\text{surv}}$, using a ultra-flat ΛCDM cosmological model with matter density $\Omega_m = 0.3$. For our forecast, we assume representative values $b_g = 1$, $b_h \Omega_{HI} = 0.43 \times 10^{-3}$ (Switzer et al. 2013), $b_{ig} \Omega_{HI} = 1$ 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 equation (13) assuming that each optical galaxy contains H I mass $M_{HI} = 10^7 h^{-2}$ M⊙ and generates the matter power spectrum $P_m(k)$ using a Planck fiducial cosmological model, and assume

$$T_{HI} = 197 \, \text{mK} \times \frac{(1 + z)^2}{H(z)}$$

(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 $k < k_{\text{min}} = 2\pi u_{\text{min}}/r$, where $u_{\text{min}}$ 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 $P_{gx}(k)$ and noise $\sigma(P_{gx})(k)$ for these configurations is plotted in Fig. 1. 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 equation (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, $N_{\text{mode}}$ would be increased; however, with constant observing time, this would increase the instrumental noise. The H I 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. Equation (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 equation (23). However, one should keep in mind that we measure a statistical average of the mean H I mass of galaxies and that a smaller sample could be less representative for the selected galaxy type. We conclude that the critical parameter 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 Mpc/h)³ with particle mass resolution of 8.6 × 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 H I 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 H I 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 H I intensity maps and the optical samples. The presence of sub-structure complicates the cross-shot noise relations and we will present a full treatment of these effects in a follow-up study. We generated the mock data sets as follows.

(i) Intensity Maps

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

(ii) 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 Sloan Digital Sky Survey optical r band. The (NUV − r) colour is a measure of the SFR. Thus the derived scaling relation can be interpreted as the average H I 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 resulted in galaxy samples with N_{g} = {774889, 6214418, 694146, 263815, 187110, 662317, 965585} number of objects.

5 H I 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 \( P_{\alpha\beta}(k) \) 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, \( P_I(k) \), 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, \( P_{g,i}(k) \) on scales \( k < 1.5 \) Mpc⁻¹.

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

\[
\hat{r}(k) = \frac{\hat{P}_{\alpha\beta}(k)}{\sqrt{P_I(k)P_{g,i}(k)}} = \frac{P_{\alpha\beta}(k) + SN_i}{\sqrt{P_I(k)P_{g,i}(k)}}
\]

Using equation (14), we can rewrite the cross-correlation coefficient \( r_i(k) \) as

\[
r_i(k) = \frac{b_{\alpha\beta}(k)b_{g,i}(k)b_{HI}(k)}{\bar{b}_{HI}(k)b_{g,i}(k)}.
\]

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

If both the H I and galaxy bias were 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 \( b_{g,i} \) cancel out. Any excess power above 1 measured in \( \hat{r}(k) \) 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 \( b_{\alpha\beta}(k)/\bar{b}_{HI}(k)b_{g,i}(k) \).

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 dependence of the cross-correlation bias would follow \( b_{i}(k) \sim \bar{b}_{g,i}(k) \) and the shape of the cross-correlation coefficient could be approximated as \( r(k) \sim 1/b_{g,i}(k) \).

In more general terms, the functional form of the bias \( b_{i}(k) \) 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 H I 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 H I-rich objects in dense regions predicted by this star-formation recipe. This implies that \( b_{i}(k) \sim b_{HI}(k) \) and, therefore, the shape of the cross-correlation coefficient \( r(k) \) is proportional to \( 1/b_{HI}(k) \). For this case, we can assume that \( r_i(k) \) is independent of the galaxy sample \( i \) and converges to the same scale-dependent form \( \hat{r}(k) \sim 1/b_{HI}(k) \). In the following, we present an empirical, model independent technique to approximate \( \hat{r}(k) \) using the autopower spectra.
We estimate the functional scale dependence of $\bar{r}(k) \sim 1/b_{Hi}(k)$ using $\sqrt{P_{g,i}^{\bar{k}}(k)/P_T(k)}$, 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-normalized 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 $f(k) = \sum_{i}^{N} p_i k^i$ with $N = 3$ to the average of all bias quotients $\overline{B} = \sum_i B_i$ 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-fitting $f_0(k)$ 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 equation (26) as the solid lines. We determine the cross-shot noise by fitting the observed $\hat{r}_i(k)$ to

$$F_i(k) = f_{bi}(k) + \frac{\Delta N_i}{\sqrt{P_{g,i}^{\bar{k}}(k)/P_T(k)}},$$  

where $f_{bi}(k)$ models $\hat{r}_i(k)$. The result of the best-fitting $F_i(k)$ is plotted in Fig. 3 as the dashed lines. We can see that the best-fitting 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 H\textsuperscript{i}-poor, which results in a small brightness temperature.

We can convert the fitted cross-shot noise $\Delta N_i$ into an average H\textsuperscript{i} mass per galaxy for each sample $i$ using $SN_i = \overline{T_{Bi,gal}}/n_{Bi}$ in combination with equation (2). The results are shown in Fig. 4. The fitted average H\textsuperscript{i} 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 H\textsuperscript{i} masses above $10^8 M_{\odot}$. Within the current pipeline, it is not possible to extend H\textsuperscript{i} mass measurements beyond this threshold and for the reddest colour bin, we cannot determine an upper limit on the H\textsuperscript{i} mass. This is due to the fact that the H\textsuperscript{i} cross-shot noise lies within the variance of the polynomial fit itself.

Figure 2. We show the quotient of the galaxy bias and the H\textsuperscript{i} bias determined by taking the square root of the galaxy power spectrum divided by the intensity mapping power spectrum $\sqrt{P_{g,i}^{\bar{k}}(k)/P_T(k)}$. The different colours refer to the power spectra selected by galaxy colour ($NUV - r$), which is an indicator for the SFR of the sample, from blue, star-forming to red, quiescent objects. The dashed coloured lines indicate the best-fitting polynomial to the quotient of each individual sample and the solid black line marks the best-fitting polynomial to the averaged bias quotient.

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 H\textsuperscript{i} cross-shot noise $SN = \overline{T_i}/n_{Bi}$. The dashed lines are the the best-fit of the polynomial plus the estimated shot noise parameter $SN/\sqrt{P_{g,i}^{\bar{k}}(k)P_T(k)}$.

Figure 4. The estimated average H\textsuperscript{i} 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.
6 CONCLUSIONS

In this study, we verify the possibility of determining H I scaling relations for high redshifts using H I 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 H I mass of the galaxy population. We forecast the detectability of the H I cross-shot noise term of the cross-power spectrum for upcoming ASKAP DINGO observations. We use semi-analytic simulations to demonstrate the measurement of the H I 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 H I 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 H I 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 H I and galaxy bias terms from their respective autopower spectrum.

We conclude that the H I 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 H I 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 is facing.

ACKNOWLEDGEMENTS

We thank Paul Geil and Martin Meyer for useful discussions throughout this project. We thank the anonymous referee for their useful suggestions to improve the final form of the article. LW is supported by an Australian Research Council (ARC) Discovery Early Career Researcher Award (DE170100356). Some of this work was performed at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1066293. This research was conducted by the Australian Research Council Centre of Excellence for All-sky Astrophysics (CAASTRO) through project number CE110001020.

REFERENCES

Bahé Y. M. et al., 2016, MNRAS, 456, 1115
Bandura K. et al., 2014, in Stepp L. M., Gilmozzi R., Hall H. J., eds, Proc. SPIE Conf. Ser. Vol. 9145, Ground-based and Airborne Telescopes V. SPIE, Bellingham, p. 914522
Battye R. A., Davies R. D., Weller J., 2004, MNRAS, 355, 1339
Battye R. A. et al., 2013, MNRAS, 434, 1239
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. et al., 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., 2017, J. Cosmology Astropart. Phys., 001
Castorina E., Villaescusa-Navarro F., 2016, preprint (arXiv:1609.05157)
Catinella B. et al., 2010, MNRAS, 403, 683
Chang T.-C., Pen U.-L., Peterson J. B., McDonald P., 2008, MNRAS, 100, 091303
Chang T.-C., Pen U.-L., Bandura K., Peterson J. B., 2010, Nature, 466, 463
Chen X., 2012, MNRAS, 12, 256
Cortese L., Catinella B., Boissier S., Boselli A., Heins S., 2011, MNRAS, 415, 1797
Craun R. A. et al., 2016, preprint (arXive-prints)
Croton D. J. et al., 2016, ApJS, 222, 22
Delhaize J., Meyer M. J., Staveley-Smith L., Boyle B. J., 2013, MNRAS, 433, 1398
Driver S. P. et al., 2011, MNRAS, 413, 971
Fernández X. et al., 2016, ApJ, 824, L1
Fonseca J., Camera S., Santos M. G., Maartens R., 2015, ApJ, 812, L22
Fu J., Guo Q., Kauffmann G., Krumholz M. R., 2010, MNRAS, 409, 515
Hall A., Bonvin C., 2017, Phys. Rev. D, 95, 043530
Jaffe Y. L., Smith R., Candlish G. G., 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., Steierwalt S., 2010, ApJ, 723, 1359
Masui K. W., McDonald P., Pen U.-L., 2010, Phys. Rev. D, 81, 103527
Masui K. W. et al., 2013, MNRAS, 763, L20
McQuinn M., Zahn O., Zaldarriaga M., Hernquist L., Furlanetto S. R., 2006, ApJ, 653, 815
Meiring J. D. et al., 2011, ApJ, 732, 35
Noterdaeme P. et al., 2012, A&A, 547, L1
Padmanabhan H., Refregier A., 2017, MNRAS, 464, 4008
Popping G., Somerville R. S., Trager S. C., 2014, MNRAS, 442, 2398
Pourtsidou A., Bacon D., Crittenden R., Metcalf R. B., 2016, MNRAS, 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., Zwaan M. A., Briggs F. H., Chengalur J. N., Lah P., Oosterloo T., van der Hulst T., 2013, MNRAS, 435, 2693
Rhee J., Lah P., Chengalur J. N., Briggs F. H., Colless M., 2016, MNRAS, 460, 2675
Saintonge A. et al., 2011, MNRAS, 415, 32
Saintonge A. et al., 2016, MNRAS, 462, 1749
Santos M. et al., 2015, Proc. Sci., Cosmology from a SKA HI intensity mapping survey. SISSA, Trieste, PoS#019
Sarkar T. G., Datta K. K., Pal A. K., Choudhury T. R., Bharadwaj S., 2016, J. Astrophys. Astron., 37, 26
Sarkar D., Bharadwaj S., Ananthi Amalika S., 2016, MNRAS, 460, 4310
Switzer E. R. et al., 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

This paper has been typeset from a TeX/LaTeX file prepared by the author.