The Arecibo Ultra-Deep Survey

Hongwei Xi\textsuperscript{1,2,3,4}, Lister Staveley-Smith\textsuperscript{2,4*}, Bi-Qing For\textsuperscript{2,4}, Wolfram Freudling\textsuperscript{5}, Martin Zwaan\textsuperscript{5}, Laura Hoppmann\textsuperscript{2}, Fu-Heng Liang\textsuperscript{2,6}, and Bo Peng\textsuperscript{1}

\textsuperscript{1} CAS Key Laboratory of FAST, National Astronomical Observatories, Chinese Academy of Sciences, Datun Rd., Chaoyang District, Beijing 100101, China
\textsuperscript{2} International Centre for Radio Astronomy Research (ICRAR), University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia
\textsuperscript{3} University of Chinese Academy of Sciences, Beijing 100049, China
\textsuperscript{4} ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia
\textsuperscript{5} European Southern Observatory, Karl-Schwarzschildstrasse 2, D-85748 Garching bei München, Germany
\textsuperscript{6} Department of Astronomy, Tsinghua University, Beijing 100084, China

Accepted XXX. Received YYY; in original form ZZZ.

ABSTRACT

The Arecibo Ultra Deep Survey (AUDS) is a blind HI survey aimed at detecting galaxies beyond the local Universe in the 21-cm emission line of neutral hydrogen (HI). The Arecibo L-band Feed Array (ALFA) was used to image an area of 1.35 deg\textsuperscript{2} to a redshift depth of 0.16, using a total on-source integration time of over 700 hours. The long integration time and small observation area makes it one of the most sensitive HI surveys, with a noise level of \~{} 75 \mu Jy per 21.4 kHz (equivalent to 4.5 km s\textsuperscript{-1} at redshift z = 0). We detect 247 galaxies in the survey, more than doubling the number already detected in AUDS60. The mass range of detected galaxies is \log (M_{\text{HI}} \left[ M_{\odot}\right]) = 6.32 - 10.76. A modified maximum likelihood method is employed to construct an HI mass function (HIMF). The best fitting Schechter parameters are: low-mass slope $\alpha = -1.37 \pm 0.05$, characteristic mass \log (M^* \left[ M_{\odot}\right]) = 10.15 \pm 0.09, and density $\Phi_* = (2.41 \pm 0.57) \times 10^{-3} h_{70}^3 \text{ Mpc}^{-3} \text{ dex}^{-1}$. The sample was divided into low and high redshift bins to investigate the evolution of the HIMF. No change in low-mass slope $\alpha$ was measured, but an increased characteristic mass $M^*$, was noted in the higher-redshift sample. Using Sloan Digital Sky Survey (SDSS) data to define relative galaxy number density, the dependence of the HIMF with environment was also investigated in the two AUDS regions. We find no significant variation in $\alpha$ or $M^*$. In the surveyed region, we measured a cosmic HI density $\Omega_{\text{HI}} = (3.55 \pm 0.30) \times 10^{-4} h_{70}^{-1}$. There appears to be no evolutionary trend in $\Omega_{\text{HI}}$ above $2\sigma$ significance between redshifts of 0 and 0.16.

Key words: galaxies: evolution -- galaxies: ISM -- galaxies: mass function -- radio lines: galaxies.

1 INTRODUCTION

Neutral hydrogen (HI) is the most abundant component in the interstellar medium of late-type galaxies. It is an important reservoir for feeding gas into molecular clouds and star-forming regions (Sancisi et al. 2008). The distribution function of the HI masses of galaxies (the HI mass function, HIMF) and the cosmic HI density ($\Omega_{\text{HI}}$) are crucial measurements in characterising galaxies at different redshifts, and allow quantitative comparisons with theoretical models and simulations. Observations also show that HI has an extended distribution in galaxies relative to the stellar content and other tracers, therefore better filling the dark matter halo, better tracing the outflow and inflow of material, and being more sensitive to the interaction of galaxies with the intergalactic medium and between galaxies.

Two methods are commonly used to detect and measure the HI content of galaxies. Observations of the 21-cm emission line from atomic neutral hydrogen can be used for nearby galaxies. However, the 21-cm emission line is weak at high redshift where HI is more readily detected by using ultraviolet absorption against bright background sources. The so-called Damped Lyman-\alpha Absorption (DLA) systems result from absorption by high HI column density ($>2\times10^{20}$ atoms cm\textsuperscript{-2}) clouds (Wolfe et al. 2005).

Targeted 21-cm observations have been carried out on many optically-selected samples (Catinella et al. 2008; Solanes et al. 1996; Springob et al. 2005). This method results in efficient measurements of gas in galaxies, but suffers from sample selection limitations, such as the tendency of optical surveys to miss gas-rich low sur-
face brightness galaxies which can have substantial baryon content (McGaugh 1996; Dalcanton et al. 1997; Sprayberry et al. 1997). On the other hand, blind HI surveys do not have such selection effects. The drawback is that they require a much larger amount of observing time than optically targeted surveys. Fortunately, fast widefield surveys are now made possible by modern multibeam receivers, phased-array feeds and/or interferometers. For example, Barnes et al. (2001) and Meyer et al. (2004) used the Parkes 13-beam receiver to conduct a blind survey of the entire southern sky (HIPASS). This was used by Zwaan et al. (2003) and others to derive an accurate low-redshift HIMF. Giovanelli et al. (2005) employed ALFA (Arecibo L-band Feed Array) with 7 beams to perform a large HI survey in the SDSS region (ALFALFA). Martin et al. (2010) and others used the data to derive updated HIMF parameters.

More recent work has investigated the variation of HIMF in different galaxy density environments. Zwaan et al. (2005) found a steeper faint end slope in higher density regions based on full HIPASS sample (Meyer et al. 2004). They did not find any clear dependency of the knee mass on environment. Springob et al. (2005) used HI spectra of ~3000 AGC galaxies to derive a flatter faint end slope and lower characteristic mass in higher density regions. Moorman et al. (2014) and Jones et al. (2016) found a positive correlation between the knee mass and density based on the 40% and 70% ALFALFA catalogues, respectively. However, neither found convincing evidence for any trend in the low-mass slope. Jones et al. (2018) found a similar trend in $M^*$ as Moorman et al. (2014) and Jones et al. (2016), but a steeper $\alpha$ in dense non-cluster regions. Said et al. (2019) found a flatter $\alpha$ in dense regions in the Zone of Avoidance (HIZOA, Donley et al. 2005; Staveley-Smith et al. 2016) survey, but no clear trend in $M^*$. These findings provide information, but as yet no clear overall picture, on the role of interactions, mergers and environment in galaxy formation. Moreover, all the above surveys are limited to the local Universe – e.g. HIPASS ($z < 0.04$) (Meyer et al. 2004; Wong et al. 2006) and ALFALFA ($z < 0.06$) (Giovanelli et al. 2005), so provide no information about the evolution of galaxy properties.

In this paper, we use the blind HI Arecibo Ultra Deep Survey (AUDS) to focus on galaxies of lower mass and at higher redshift than HIPASS or ALFALFA. The sensitivity of AUDS is 50-200 times better than the above surveys, but is restricted to a much smaller area and cosmic volume. AUDS also uses the Arecibo L-band Feed Array (ALFA), but our deeper observation allows us to probe the HIMF up to $z \approx 0.16$. Initial AUDS precursor observations were used to verify the ‘drift-and-chase’ observational technique, detecting 18 galaxies in redshift range $0.07 < z < 0.16$ (Freudling et al. 2011). Hoppmann et al. (2015) analysed 60 per cent of the final survey (hereafter AUDS60) to construct a catalogue of 102 galaxies. No significant evolution of the HIMF in the last billion years was found.

We present the final, complete AUDS survey (AUDS100) and organise the paper as follow. In Section 2, we describe the observations. The data reduction techniques are described in Section 3. The source-finding method and galaxy catalogue is described in Section 4, where we also investigate completeness and reliability. Two methods for deriving the HIMF are described in Section 5, together with a method to derive corrections for cosmic variance. The HIMF and cosmic HI density $\Omega_M$ derived from the AUDS data is presented in Section 6. Finally, we discuss and present our conclusions in Sections 7 and 8.

In this paper, we use $H_0 = 70$ km s$^{-1}$ Mpc$^{-1}$, $\Omega_M = 0.3$, and $\Omega_L = 0.7$.

### Table 1. The coordinates of two survey fields in AUDS.

| Field      | RA (J2000) range | Dec (J2000) range | Area |
|------------|------------------|------------------|------|
| FIELD17H   | 17:02:39.0 to 16:57:21.0 | 19:19:26 to 20:09:26 | 0.68 |
| GAL2577    | 08:22:41.5 to 08:17:18.5 | 21:45:25 to 22:35:25 | 0.68 |

### 2 OBSERVATIONS

In order to avoid solar interference and minimise radio frequency interference (RFI), only night observations were requested. To allow observations during most of the year, we therefore selected two areas located far apart, but still lying within the SDSS footprint (York et al. 2000) in order to provide complementary optical information. The field separation also helped to reduce cosmic variance. Regions with strong continuum sources were avoided. The coordinates and effective survey area of the chosen fields are listed in Table 1.

Arecibo radio telescope were employed to carry out the observation. The ALFA improved the speed of the survey by 7 times. The L-band beam size of Arecibo telescope is $\sim 3.5$ arcmin at redshift $z=0$. We used a ‘drift-and-chase’ observation mode as verified by Freudling et al. (2011). The Arecibo telescope was pointed at the western end of the field, a calibration cycle was performed, and the telescope was put in drift mode which allowed the sky to drift through the telescope beams for $\sim 230$ s, which corresponds to $\sim 1^\circ$ in Right Ascension (RA). Taking into account the ALFA footprint, the coverage in RA is $\sim 1.2^\circ$. The procedure was repeated with a shift of 20 arcsec in Declination (Dec). Due to the limited zenith angle range of the Arecibo telescope, a single observation of a field usually took several nights, after which the whole process was repeated.

Uniform (Nyquist) sampling was achieved by rotating the ALFA receiver by 19.1° with respect to the drift direction (see Figure 1 in Hoppmann et al. (2015)). In order to maintain bandpass stability, the receiver was not rotated during the observation. Rotation of ALFA was set at the start of the observation according to the observing time and the telescope azimuth and elevation coordinates.

The project started in November 2008, and finished in December 2012. Approximately one hour per night was assigned when the target area was overhead. In total, over 1000 hours observation time was assigned for the project, with an on-source time over 700 hours. The remainder was used for calibration or lost due to overheads.

The spectrometer recorded the spectral data in two sub-bands, namely S0 and S1. Each sub-band has a bandwidth of 175 MHz, which is divided into 8192 spectral channels. Each channel corresponds to a frequency resolution of 21.4 kHz and a velocity resolution of 4.5 km s$^{-1}$ at redshift $z = 0$. The two sub-bands were centered at 1.30 GHz and 1.45 GHz, covering a total bandwidth of 300 MHz from 1.225 GHz to 1.525 GHz. These sub-bands overlap a range of 25 MHz. Further observational details can be found in Hoppmann et al. (2015).

### 3 DATA REDUCTION

We used livedata to perform bandpass and flux calibration (Barnes et al. 2001). It was designed to process multibeam data from HIPASS, but was modified to include the ability to process data from ALFA. The main changes relate to the details of the data format, the Arecibo gain-elevation curve, and the different calibration strategy. HIPASS used a continuous noise-diode calibration, MNRAS 000, 1–16 (2015)
After re-calcultating the rms, pixels with values exceeding 3 times the global rms in the scan, the pixels will be flagged unless the total flagged percentage for a given spectrum or channel will be less than 80 per cent. Otherwise the spectrum is removed. (v) The ALFA correlator produced a central channel artifact corresponding to 1.3 GHz in band S1. We therefore flagged the central channel in each scan. (vi) Finally, we checked the flagged fraction in the scan file. If the flagged fraction exceeds 90 per cent, the whole scan is discarded.

We also remove an artifact caused by LIVEDATA processing of bright galaxy spectra. The bandpass calibration procedure, although robust, will slightly depress the scan-average in the frequency range of a bright galaxy, resulting in RA sidelobes. In order to remove this effect, we remove a median value for each channel in each scan file, after masking out any such galaxy profiles.

In Figure 1, we show the flagged percentage, or RFI occupancy for the two sub-bands. Aside from RFI, the method unfortunately also flags HI emission from the Milky Way, located at 1.42 GHz. However, all the raw data is archived and reusable for future research on HI in Milky Way.

The task GRIDZILLA is employed for gridding into image cubes. Since the AUDS100 data volume is too large (8192 channels $\times$ 230 integrations $\times$ 2 polarisations $\times$ 2 sub-bands $\times$ 7 beams $\times$ 11,000 scans) for GRIDZILLA to process the whole data set in memory, we divide each spectrum into 64 sub-spectra with an overlap of 22 channels between neighbouring sub-spectra to allow for Doppler shifts.

Although the LIVEDATA calibration procedure is robust, and largely immune to RFI and outliers, it only flags strong RFI signals. Therefore Hoppmann et al. (2015) developed a flagging script for removing fainter RFI in the AUDS60 data. It performed reasonably well, but an improved flagging technique was required for the higher sensitivity AUDS100 data. Here we detail the updated flagging process:

(i) We check the rms noise level in each scan. If the ratio of noise level in the two polarisations is larger than two, we completely remove the data as it is likely to be irrecoverably corrupt due to system problems or very strong RFI.
(ii) We flag data with values deviating by greater than $\pm 0.1$ Jy. After re-calcultating the rms, pixels with values exceeding $7\sigma$ are flagged. This procedure removes most RFI generated by the airport radars.
(iii) We remove $2\sigma$ outliers in the Fourier transform of the temporal values for each frequency channel in each scan. Then we perform an Inverse Fourier Transform. This procedure allows us to remove periodic RFI.
(iv) We calculate the local rms of scans in the time-frequency plane using three box sizes: $5 \times 5, 1 \times 5$, and $5 \times 1$ pixels, respectively. If the local rms exceeds three times the global rms in the scan, the pixels will be flagged unless the total flagged percentage for a given spectrum or channel will be less than 80 per cent. Otherwise the spectrum is removed.
(v) The ALFA correlator produced a central channel artifact corresponding to 1.3 GHz in band S1. We therefore flagged the central channel in each scan.

We use GRIDZILLA to process these 64 sub-spectra in parallel on a computer cluster. The final cube is generated by linear interpolation of the 64 sub-cubes, resulting in four final cubes, one for each band and each field.

As each observational period was typically only an hour, relative calibration was based on the receiver noise diode and a template gain-elevation curve for the telescope. The overhead in daily calibrator observations was otherwise too high. Final flux density calibration was therefore adjusted by comparing sources with S/N ratio above 5 in the AUDS area with 46 NVSS continuum sources (Condon et al. 1998). The absolute uncertainty in the final flux density scale using this calibration method was estimated to be 4 per cent. The ratio of integrated fluxes for the only galaxy in common with ALFALFA (Courtois & Tully 2015) is 1.02 $\pm$ 0.03, in agreement with the above. Note that the flux densities in this paper supersede those in Hoppmann et al. (2015) which were on average lower by 29 per cent compared with this paper. Therefore, using the same $h$, AUDS100 on average has a 29 percent higher HI mass than AUDS60.

4 SOURCE CATALOGUE

4.1 Source-finding methodology

Automated source-finding software, such as SoFiA (Serra et al. 2015a) is useful in finding both point sources (Westmeier et al. 2017) and extended sources (see e.g., Serra et al. 2015b; Reynolds et al. 2019). However, application to AUDS is not straightforward due to the strong variability of the noise level and baseline with position and frequency. We therefore pay close attention to the measurement of the noise level and the flattening of the baseline before passing to our own custom source-finding algorithm.

Due to the dependence on position and frequency, we require a noise estimate for each pixel. If $f_c$ represents the flagged fraction caused by RFI in a given frequency channel in the image cube, then we expect that the noise level for that channel will scale as $\sigma_c \propto 1\sqrt{1-f_c}$. Also, if $f_S$ represents the flagged fraction of a given spectrum in the image cube, then the noise level of a spectrum will be expected to scale as $\sigma_S \propto 1\sqrt{1-f_S}f_S$, where $f_S$ is proportional to the observing time at that location. Since a given pixel lies at the intersection of a spectrum and a channel, we might expect the overall noise level at that pixel to be approximated by $\sigma_p \propto 1\sqrt{1-f_S}f_S (1-f_c)$, or $\sigma_p \propto \sigma_S \sigma_c$. This gives a method
for deriving an estimate of the noise on the scale of individual pixels. Robust estimates for $\sigma_\alpha$ and $\sigma_c$ are derived using the consistent estimator $\sigma = 1.4826$ MAD, where MAD is the Median Absolute Deviation.

As an example, we plot $\sigma_p$ as a function of frequency and position for FIELD17H Band S1 in Fig. 2. It can be seen that $\sigma_p$ rises at highly-flagged frequencies between 1.22 and 1.28 GHz. It also rises at the edge of the field due to limited observational coverage, and rises due to bad baselines associated with continuum sources.

For baseline fitting, we firstly smooth each spectrum in the cube in successive steps up to 20 channels (0.43 MHz). At each step, we re-calculate the noise level for the smoothed cube, record points deviating by more than $3\sigma$ as candidate sources, and remove those points from the subsequent baseline fitting procedure. Spectra with bad baselines may end up being 100 per cent flagged in this procedure. RFI will only cause a limited number of channels to be flagged. Candidate sources extending less than 3 arcmin spatially and 128 kHz ($27 \text{ km s}^{-1}$) spectrally are removed from the list.

All sources were checked by eye and several obviously false detections were removed, resulting in 153 candidates. We then inspected the data cube by eye for additional candidates. The criteria were either: 1) the candidate has the width of a beam and is detected in one or multiple channels when viewed in the position-position plane; or 2) the candidate is detected in a 3x3 pixel spatially-averaged spectrum. This process picks up a further 94 candidates with S/N ratios slightly less than $3\sigma$. These candidates are of lower reliability (see Section 4.5) hence it is likely that some are false detections. Due to the large beamwidth (~ 3.4 arcmin) of Arecibo at this frequency, some sources are confused. We inspect the spectral and spatial morphology of these sources, and attempt to separate them into individual sources.

### 4.2 Catalogue

We detect 247 galaxies in total, of which 9 are detected in both S0 and S1 bands. We use AUDS100 to identify all AUDS galaxies, and AUDS100bpec to identify 153 AUDS galaxies only detected by our own custom source-finding algorithm. The 94 AUDS galaxies found by eye are indicated by AUDSbpec. For each source, we use a cubic polynomial to fit the baseline. A two-dimensional Gaussian is used to fit the moment map in order to derive the source central position. We assume that the galaxies we detect are not resolved. The mean and median FWHM of fitted two-dimensional Gaussian function is 3.42 and 3.46 arcmin, respectively, which is consistent with this assumption. Spectra within the FWHM ellipse are used to determine the spatially integrated spectrum for each galaxy:

$$S_i(v) = \frac{\sum_{x,y} s_i(x,y,v)}{\sum_{x,y} f(x-x_i, y-y_i)},$$

where $s_i(x,y,v)$ is the flux density of the voxel at $(x, y, v)$ of galaxy $i$, and $f(x,y)$ is the 2 dimensional Gaussian function fitted to galaxy $i$ with amplitude of unity, and centred on the fitted galaxy position $(x_i, y_i)$. A fifth-order polynomial is also fitted and removed from the spatially integrated spectrum $S_i(v)$. For the 9 galaxies detected in both bands, the spectrum with lowest noise level is used in our final result. Example spectra over a range of redshifts are shown in Figure 3.

The flux-density-weighted frequency is used to derive the redshift of each galaxy. In order to remove the influence of noise on properties of galaxy, we use the Busy Function (Westmeier et al. 2014) to fit the spatially integrated spectra, then obtain the frequency-integrated flux $S_{int}$, the peak flux density $S_{peak}$, and linewidth $W_{50}$. The HI mass is defined using:

$$M_{HI} [h_{70}^{-1} M_\odot] \approx 49.7 \times D_L [h_{70}^{-1} \text{Mpc}]^2 S_{int} [\text{Jy Hz}]$$

(Meyer et al. 2017), where $D_L$ is the luminosity distance. We list the properties of the galaxies in Table 2.

The common detected sources in both bands allow us to estimate a lower bound on the uncertainties in coordinates and recession velocity. The mean spatial and recession velocity differences are $0.22 \pm 0.20$ arcmin and $11 \pm 17 \text{ km s}^{-1}$, respectively. Considering the spatial resolution of $\sim 3.4$ arcmin and velocity resolution $\sim 4.5 \text{ km s}^{-1}$, these uncertainties are acceptable.

We compare the AUDS100 catalogue with the AUDS60 catalogue (Hoppmann et al. 2015). Using a maximum spatial and velocity separation of 3 arcmin and $250 \text{ km s}^{-1}$, respectively, 86 out of 102 AUDS60 sources have a counterpart in AUDS100. A further two sources (J081933+220624, J082011+223025) were found to have separations of more than 3 arcmin. Two confused pairs of AUDS60 sources (J165831+200304 and J165832+195702, J170122+195245 and J170123+195238) appear to correspond to only two AUDS100 sources. One source (J165569+201232) is outside the nominal survey boundary. Two weak sources (J165952+193840 and J170117+200235) have very narrow line widths, and are probably RFI artifacts. Another AUDS60 source (J170133+200336) also lies in a region with strong RFI. The remaining 8 AUDS60 sources (J081749+221114, J081957+221628, J082034+220848, J082116+221604, J082130+221945, J082136+221833, J082147+221831, J082157+221741)
Figure 3. Example spectra of nine AUDS100 galaxies at different redshifts. The red dotted lines show the best-fit Busby functions; the blue dotted lines indicate zero flux density; the double-sided arrows indicate the $W_{50}$ linewidth. The completeness parameter $C$, defined in Section 4.4, and redshift is listed for all detections. The spectra for all galaxies are available online.

J165938+193030, J165938+193512, J170133+193253) are not detected in AUDS100 and likely not real.

### 4.3 Optical counterparts

The two fields of AUDS are covered by the SDSS. By cross-matching with the SDSS DR14 (Abolfathi et al. 2018), we find that there are 191 optical galaxies with spectroscopic redshifts in FIELD17H, and 598 in GAL2577. Hoppmann (2014) did a follow-up spectroscopic observation of an additional 787 galaxies using the Anglo-Australian Telescope (AAT). These galaxies were selected from the SDSS DR7 catalogue based on $r < 20.3$ mag in the AUDS survey area. Of these, there are 229 spectra in the AUDS redshift range of sufficient quality.

Optical counterparts are matched to AUDS100 detections when the angular distance is within 1.8 arcmin, and the velocity difference is within 200 km s$^{-1}$. If an optical galaxy matches more than one HI source, the nearest one is chosen as the counterpart. In total, 97 out of the 247 galaxies have optical counterparts. For the rest of the 150 galaxies, we search SDSS for galaxies in the AUDS area with photometric redshifts less than 0.3 and angular separation within 5 arcmin. With these criteria, we find optical counterparts for all the remaining 150 galaxies. The distribution of angular distance is shown in Figure 4. All photometric counterparts lie within 2.1 arcmin, indicating that most of them are likely to be genuine associations. We fit a two-component function consisting of a Gaussian distribution and an uniform distribution, both in angular coordinates. The Gaussian distribution represents the genuine associations, while the uniform distribution represents confusing sources at different redshift. The result indicates that there may only be ~ three sources (AUDS ID: 112, 186, 206) whose optical counterparts or HI detections are in doubt. Other matches all lie within 1.6 arcmin of angular distance. Further optical observations are required to confirm these associations.
### Table 2: Properties of galaxies in AUDS100. The full version can be found online.

| AUDS ID | RA, Dec (J2000) | $cz$ km s$^{-1}$ | $S_{\text{int}}$ mJy MHz | $S_{\text{peak}}$ mJy MHz | $W_{50}$ MHz | $\log \left( \frac{M_{\text{HI}}}{10^{5} M_{\odot}} \right)$ | C | Sample | Pro | G | AUDS60 RA, Dec (J2000) (OC) |
|---------|-----------------|------------------|--------------------------|--------------------------|--------------|--------------------------------|---|--------|----|---|-----------------------------|
| 001     | 16:58:22.3+19:55:13 | 1838.6           | 0.094                    | 0.42                     | 0.23         | 6.51                          | 0.95 | ...    | N  | - |                          |
| 002     | 16:59:11.8+20:06:15 | 4346.9           | 0.187                    | 1.50                     | 0.12         | 7.56                          | 0.96 | ...    |    |   |                          |
| 003     | 16:58:02.4+19:54:07 | 4389.1           | 0.220                    | 0.85                     | 0.26         | 7.64                          | 1.00 | Sub    | Y  | - | J165806+195322 16:58:02:4+19:5362 |
| 004     | 16:58:39:2+19:56:11 | 4682.9           | 0.291                    | 0.30                     | 1.04         | 7.82                          | 0.88 | ...    |    |   |                          |
| 005     | 16:58:41.3+19:34:12 | 5631.2           | 0.082                    | 0.34                     | 0.22         | 7.43                          | 0.96 | Sub    | Y  | - |                          |
| 006     | 16:59:34:2+19:45:21 | 6012.7           | 0.608                    | 1.84                     | 0.32         | 8.36                          | 1.00 | Sub    | Y  | - | J165934+194525 16:59:34+19:4536 |
| 007     | 16:58:29.8+20:02:46 | 6022.5           | 12.985                   | 27.79                    | 0.47         | 9.69                          | 1.00 | Sub    | Y  | - | J165831+200304, J165831+200304 16:58:29+20:02:29 |
| 008     | 17:02:13.7+19:29:56 | 4917.6           | 1.915                    | 5.73                     | 0.33         | 9.26                          | 1.00 | Sub    | Y  | - | J170212+193010            |
| 009     | 16:59:51.3+19:44:28 | 9774.7           | 0.322                    | 0.77                     | 0.44         | 8.52                          | 1.00 | Sub    | Y  | - | J165950+194442 16:59:50+19:4424 |
| 010     | 17:00:18.4+19:31:39 | 9805.4           | 0.176                    | 0.39                     | 0.43         | 8.26                          | 0.97 | Sub    | Y  | - | J170015+193342 17:00:19+19:3148 |

Cols (1)-(2): AUDS100 identification, RA and Dec (J2000) coordinates from the two dimensional Gaussian fit, respectively. Col (3): Recession velocity, $cz$. Cols (4)-(6): $S_{\text{int}}$, $S_{\text{peak}}$, and $W_{50}$ - integrated flux, peak flux density and linewidth, respectively. Col (7): HI mass. Col (8): completeness estimation of each galaxy (see Section 4.4). Col (9): sample designation. Col (10): whether source is detected by the custom source finding algorithm. Col (11): Flag for confusion. Col (12): identification of AUDS60. Col (13): coordinates of the optical counterpart (see Hoppmann 2014).

b: Galaxies detected in both band S0 and S1.
1: SDSS (Hoppmann 2014).
2: AUDSOC (Hoppmann 2014).
synthetic galaxy was obtained from the data cube once the position and frequency were known. A Gaussian function of FWHM 3.4 arcmin was convolved with the model.

We use the automated procedure of Section 4.1 to search synthetic galaxies. This can give us the completeness for AUDS100$_{pro}$ sample. The number of synthetic galaxies in each frequency bin, and the detection rate of our automated source-finder. Bottom panel: the detection rate in each cell, where black is 100 per cent and white is 0 per cent.

Figure 6. An example spectral profile of an AUDS100 galaxy is shown by the black solid line. The red profiles have the same integrated flux, but are scaled to a different linewidth. The blue profiles have the same linewidth, but are scaled to different fluxes. Profiles such as these are used as templates for creating synthetic galaxies in order to measure sample completeness. In the legend, $S_{\text{int}}$ and $W_{50}$ are in the units of mJy MHz and MHz, respectively.

Figure 7. Top panel: the number of synthetic galaxy profiles inserted into the data cube at each frequency bin, and the detection rate of our automated source-finder. Bottom panel: the detection rate in each cell, where black is 100 per cent and white is 0 per cent.

Figure 8. The two dimensional sigmoid fit of Equation 3, using parameters from Table 3. The model and residual is shown in Figure 8. There are some minor differences ($\pm 0.3$) near the steepest part of the completeness function. The same procedure is performed for the higher-reliability Sub$_{pro}$ sample (defined in Section 4.5).

As we mentioned, the result can only give us the completeness for the AUDS100$_{pro}$ sample found by our source-finding program, and the high-reliability Sub$_{pro}$ sample. However, in order to estimate completeness for the full AUDS100 sample, which includes by-eye detections, we made an assumption that the completeness is also described by Equation 3, but with a shift described by a modified $C_2$ parameter.

We then re-bin the data at different values of $-C_0 \log(S_{\text{int}}/\sigma) + C_1 \log(W_{50}) + C_2$ such that the galaxies in the same bin have same completeness for one sample. The total number of galaxies that should be in each bin, corrected for completeness, is then $N_{\text{Bin}}/C_{\text{Bin}}$, where $N_{\text{Bin}}$ is the galaxy number in that bin, and $C_{\text{Bin}}$ is

| Sample     | $C_0$      | $C_1$      | $C_2$      |
|------------|------------|------------|------------|
| AUDS100$_{pro}$ | 9.98(0.73) | 7.72(0.59) | 3.76(0.22) |
| AUDS100     | 9.98(0.73) | 7.72(0.59) | 1.1(0.12)  |
| Sub$_{pro}$ | 9.95(0.83) | 7.66(0.67) | 3.74(0.25) |
| Sub         | 9.95(0.83) | 7.66(0.67) | 3.72(0.12) |
the completeness. The total number in each bin should be the same, no matter in which sample:

\[
\frac{N_{\text{Bin}}[\text{AUDS100}]}{N_{\text{Bin}}[\text{AUDS100}]} = \frac{N_{\text{Bin}}[\text{AUDS100}]}{N_{\text{Bin}}[\text{AUDS100}]} \tag{4}
\]

which allows us to estimate the completeness in each bin in the AUDS100 sample. The number of galaxies in each bin of AUDS100Pro and AUDS100 is shown in Figure 9 (upper panel), with the error bars indicating the Poisson noise. The completeness for the AUDS100 sample is also shown in Figure 9 (lower panel).

In order to avoid small number statistics, we use a wide bin width of 2.0 dex, and remove the bins where either \(N_{\text{Bin}}[\text{AUDS100}]\) or \(N_{\text{Bin}}[\text{AUDS100}]\) is smaller than 5. The completeness for both AUDS100Pro and AUDS100 is shown. We use Equation 3 to fit, and re-derive \(C_2\) for the latter, as given in Table 3. We use the same method for the Sub sample using galaxy numbers in SubPro and Sub instead.

As indicated above, the completeness should be zero for galaxies with low values of \((S_{\text{int}}/\sigma)\) and high values of \(W_{50}\) in Figure 8. To avoid the non-zero tail of the completeness function artificially affecting the effective volume, we truncate the completeness function (i.e. set to 0) at values of 0.03 and 0.35 for the AUDS100 and Sub samples, respectively. No galaxies are detected in this region.

### 4.4.1 Modified \(V/V_{\text{max}}\) test

The modified \(V/V_{\text{max}}\) test was introduced by Rauzy (2001) to examine the completeness of optical catalogues. It can also be used to assess completeness of HI catalogues (Zwaan et al. 2004; Staveley-Smith et al. 2016). The test is based on the assumption of invariant shape of the HIMF, but is insensitive to large-scale structure (changes in density only). It is based on the data alone, and is therefore an ideal method to verify the completeness derived from our simulations.

The noise level in our cubes varies at different spatial positions and frequencies. Hence, the noise level for each galaxy is different from each other. Following Hoppmann et al. (2015), and the simulations conducted above, we use the noise-scaled integrated flux \((S_{\text{int}}/\sigma)\), and linewidth \((W_{50})\) to characterise the completeness. We construct \(Z_0 = \log(S_{\text{int}}/\sigma) + \log(M_{\text{HI}})\) in order to apply the \(T_C\) method (Rauzy 2001).

In Figure 10, we show the distribution of noise-scaled integrated flux \((S_{\text{int}}/\sigma)\), and noise-scaled mean flux density \((S_{\text{int}}/\sigma W_{50})\) as a function of \(W_{50}\). It clearly shows that the galaxies with narrow linewidths are easier to detect for a given integrated flux, and that galaxies with broad linewidth are easier to detect for a given mean flux density. The radiometer equation suggests a flat distribution should result by applying the scaling \(W_{50}\), with \(\alpha = 0.5\) for integrated flux and \(W_{50}^{-1}\) for mean flux density. However, Staveley-Smith et al. (2016) suggest that baseline ripple makes galaxies with broad linewidth harder to detect, causing \(\alpha\) to deviate from 0.5. The lower panel in Figure 10 suggests \(\alpha = 0.58\) is appropriate for the AUDS100 sample. Similarly, \(\alpha = 0.69\) is derived for Sub sample. We therefore construct \(Z_1 = \log(S_{\text{int}}/\sigma W_{50}) + \log(M_{\text{HI}})\), and \(Z_2 = \log(S_{\text{int}}/\sigma W_{50}^{0.58}) + \log(M_{\text{HI}})\).

We show the results for the AUDS100 sample derived from the \(T_C\) statistic in Figure 11. The sharpest drops to \(T_C = -3\) occur at \((S_{\text{int}}/\sigma) = 0.74\) and \((S_{\text{int}}/\sigma W_{50}^{0.58}) = 1.09\) in the first and third
panels, respectively. This gives us 99.4 per cent confidence that the 100 per cent completeness limits are above these two values. $T_C$ drops more gradually for $(S_{\text{int}}/\sigma W_{50})$, indicating that a single limit does not well characterise completeness for this parameter.

We compare completeness from the $T_C$ test with our simulations in Figure 5, combining the two limits above in order to give the harshest constraints. The $T_C$ test gives a similar result to the simulation at a completeness level of 0.3 for galaxies with broad linewidths, and a similar dependence on linewidth. However, the $T_C$ test gives us a more strict constraint on galaxies with narrow linewidths, albeit this part of the plane is not as well populated. For the $Sub$ sample, we derive $T_C = -3$ limits of $(S_{\text{int}}/\sigma) = 0.97$ and $(S_{\text{int}}/\sigma W_{50}) = 1.61$, and a closer correspondence between the $T_C$ and simulation limits.

4.5 Reliability

AUDS100 sources were selected independently of SDSS, so cross-matching between the catalogues is a good measure of reliability. Combining SDSS and follow-up AAT spectroscopy, Section 4.3 shows that there are 97 (39 per cent) of AUDS100 galaxies which have optical counterparts. Unfortunately, the remaining AUDS100 sources do not have any spectroscopic observations. However, the photometric redshift analysis of Section 4.3 also shows that all of the remaining 150 (61 per cent) sources have plausible counterparts, though it is probable that – three matches are unrelated. This suggests that reliability may be as high as 98 per cent, though this is likely an upper limit.

Hoppmann et al. (2015) measured reliability by placing synthetic galaxies in the AUDS60 data cube and compared the results of three by-eye source searchers. In this paper, we use negative detections in the final cubes to internally estimate reliability. The final cubes were inspected by eye for negative detections in the same manner as for positive detections (as described in Section 4.1). 23 negative detections were found, mostly at the survey edge or in the strong RFI frequency range, and all had low SNR. The completeness of these negative detections are small relative to AUDS100 sample. This indicates that AUDS100 galaxies which are: 1) located at the edge of survey area; 2) located in the strong RFI frequency range; or 3) have low SNR, are prone to being false detections. These 23 negative galaxies give an alternative estimate of AUDS100 reliability of ~ 91 per cent.

We have also formed a high-reliability, high-completeness sub-sample of 148 AUDS100 galaxies (‘$Sub$’) using the following criteria:

- Frequency greater than 1.25 GHz (to avoid RFI),
- Peak flux density greater than 2.5$\sigma$, where $\sigma$ is the local rms in mJy, and
- Integrated flux greater than $10.5\sqrt{n}\sigma \Delta v$, where $n$ is number of channels over which the galaxy is detected and $\Delta v$ is channel width in MHz.

None of the negative detections satisfy the above criteria, so the reliability for $Sub$ should be very close to 100 per cent. In this high-reliability sub-sample, there are 83 (56 per cent) HI sources with spectroscopic redshifts from either SDSS or AUDSOC. This rate is close to that of HIPASS objects with optical spectroscopic redshifts (58 per cent) (Doyle et al. 2005), but higher than for the full AUDS sample (39 per cent).

5 HI MASS FUNCTION

The HI mass function (HIMF) is used to describe the average density of galaxies of different neutral hydrogen mass in the Universe. Most studies (Zwaan et al. 2005; Martin et al. 2010; Hoppmann et al. 2015) parameterize the HIMF with a Schechter function (Schechter 1976):

$$\Phi(M_{HI}) = \ln(10)\Phi_\Sigma M_{HI}^{\alpha+1} e^{-M_{HI}/M_*},$$

where $\alpha$ is the slope at low-mass end, $M_*$ is the characteristic mass, and $\Phi_\Sigma$ is the normalization constant. We adopt two methods to derive the HIMF, $\Sigma_1/V_{\text{max}}$ and 2D Stepwise Maximum Likelihood (2DSWML).

5.1 $\Sigma_1/V_{\text{max}}$ method

In the $\Sigma_1/V_{\text{max}}$ method (Schmidt 1968), the maximum volume ($V_{\text{max}}$) in which a galaxy can be detected is computed from its maximum detectable distance. The HIMF can then be derived, for example by summing $1/V_{\text{max}}$ over different mass bins. It is a widely used method due to its speed and simplicity of implementation. However, it is sensitive to cosmic variance, which can bias shape and amplitude of the HIMF. Example, an incorrect slope can be caused by having a different local galaxy density (mainly low mass) as compared to with distant galaxies (mainly high mass) (Martin et al. 2010). Here, we follow the procedure described in Hoppmann et al. (2015), in which effective volume ($V_{\text{eff}}$) is used instead of $V_{\text{max}}$.

The method requires a noise and a volume cube for each field. For the noise cube, we use the method described in Section 4.1 to calculate the noise level at each pixel. The lower noise level is adopted in the overlap region between S0 and S1 bands. For the volume cube, the value represents the comoving volume for that pixel. The effective volume for each galaxy is computed from

$$V_{\text{eff}} = V_{\text{co}} \times f \times C_i,$$

where $V_{\text{co}}$ is the comoving volume in each pixel, $f$ is the relative overdensity for correcting cosmic variance (see Section 5.3), and $C_i$ is the completeness for each galaxy (calculated from Equation 3). The $V_{\text{eff}}$ is the sum of all pixels in the noise and volume cubes.

The HIMF is then derived by summing the inverse of the effective volume within each mass bin, with the error given by $\sqrt{\Sigma_1/V_{\text{eff}}^2}$. A non-linear least squares (NLLS) method is then used to obtain the best-fit Schechter parameters. This method requires the binning of the data, which introduces sensitivity to the bin position and bin width. The effect is significant if the sample is small, such as for AUDS. The modified maximum likelihood (MML) method of Obreschkow et al. (2018), which can fit without binning, is therefore also adopted.

5.2 2DSWML

The 2DSWML method uses the likelihood of finding a galaxy with HI mass $M_{HI}$ and velocity linewidth $W_{50,I}$ at distance $D_i$. As with $\Sigma_1/V_{\text{max}}$, the method is model-independent, but is also robust against cosmic variance if the shape of the HIMF is invariant. It therefore provides an independent method for confirming the $\Sigma_1/V_{\text{max}}$ results (Zwaan et al. 2005; Martin et al. 2010).

The HI mass-linewidth plane is split into logarithmic mass and
linewidth bins, and galaxies in each bin are counted. The local rms and distance of each galaxy is used to compute the completeness in the mass-linewidth plane for the galaxy, and the mean value of completeness in each bin is computed. The maximum likelihood solution was derived by iterating Equation B6 in Martin et al. (2010) until the solution was stable. The errors are derived from inverse of its information matrix, as detailed in Efstathiou et al. (1988) and Loveday (2000). We apply the same constraint as used in Martin et al. (2010),

\[ g = \sum_j \sum_k \left( \frac{M_{\text{HI},j}}{M_{\text{HI,ref}}} \right) \frac{W_{50,j}}{W_{50,\text{ref}}} \phi_{jk} \Delta m \Delta w - 1 = 0, \]  

(7)

where \( m = \log(M_{\text{HI}} [h^{-2} \text{M}_\odot]) \), \( w = \log(W_{50} \text{[MHz]}) \), \( M_{\text{HI},j} \) and \( W_{50,j} \) are the central HI mass and linewidth in bin \( j,k \), \( \phi_{jk} \) is the stable solution in the \( \log(M_{\text{HI}}) - \log(W_{50}) \) plane, and \( M_{\text{HI,ref}} \) and \( W_{50,\text{ref}} \) are the mean mass and linewidth for the AUDS100 sample.

To convert the probability density function into an HIMF, we need to evaluate the amplitude of the HIMF by matching the integral of the distribution to the inferred average density of galaxies in the survey volume. The selection function, \( S(D) \), is used to normalize the survey volume and is given by following equation which gives the galaxy number weighted selection function:

\[ S(D) = \frac{\int \int C_D(m,w,\sigma) \phi dm dw}{\int \int \phi dm dw} \]  

(8)

where \( C_D(m,w,\sigma) \) is derived from Equ 2 and 3. The \( \sigma \) is median value of noise at distance \( D \). Considered the different noise at same distance in two survey field, we calculate selection function for each field. We adopt volume as weight to derive our final selection function. There are three estimators for the survey volume (Davis & Huchra 1982). We adopt the most stable estimator, \( n_3 \), to normalize \( \phi_{jk} \). The HIMF is derived by integrating \( \phi_{jk} \) over linewidth. The best fit (NLLS) parameters were then derived for the Schechter function.

### 5.3 Cosmic variance

Cosmic variance is a statistical uncertainty in estimating the number density of galaxies caused by the underlying large-scale density fluctuations. The uncertainty is in particularly significant for high-redshift, deep pencil-beam surveys. Methods for correcting for the influence of cosmic variance on the HIMF include: (1) using an optical catalogue to measure the relative density distribution (Hoppmann et al. 2015; Jones et al. 2016) and (2) using a density distribution model (Martin et al. 2010). We adopt the first method to correct for the cosmic variance, calculating the overdensity for a given redshift from:

\[ f(z) = \frac{\rho(z)}{\rho_{\text{ref}}(z)} \]  

(9)

where \( \rho(z) \) and \( \rho_{\text{ref}}(z) \) are number density of galaxies in each of the AUDS field and a reference field, respectively.

To calculate the number density of galaxies in these fields, we select late-type galaxies from the SDSS DR14 catalogue (Abolfathi et al. 2018). We preferentially select late-type galaxies using the criteria: \( u' - r'< 2.22 \), where \( u' \) and \( r' \) are model magnitudes as described in Strateva et al. (2001), and also select galaxies with redshift uncertainty \( \sigma_z < 0.01 \). Late-type galaxies better represent the galaxy population in an HI survey (Jones et al. 2018). Figure 12 shows the selected galaxies for FIELD17H (253.5° < \( \alpha < 255.8^\circ \), +19.2° < \( \delta < +21.1^\circ \)), GAL2577 (123.6° < \( \alpha < 126.3^\circ \), +21.0° < \( \delta < +23.3^\circ \)) and the much larger SDSS reference field (130.0° < \( \alpha < 236.0^\circ \), 0.0° < \( \delta < +58.0^\circ \)). The selected reference field is the same as in the cosmic variance study of Driver & Robotham (2010), for which the volume is 6.14 × 10⁷ Mpc³ out to \( z = 0.17 \).

For redshift bin of width 0.02, we calculate the ratio of the density of galaxies in AUDS and the reference field up to a redshift of 0.17. (see Figure 13). The error in the measured overdensity is calculated from Poisson statistics. Two peaks with overdensities over 2 are seen around redshift 0.01 and 0.09 in the GAL2577 field. These peaks are also consistent with HI overdensities.

### 6 RESULTS

The HIMFs derived from the \( \Sigma/V_{\text{max}} \) and 2DSWML methods are presented in Figure 14, together with the HI mass distribution for the AUDS100 sample. The binned and unbinned data from the \( \Sigma/V_{\text{max}} \) method are fitted with the NLLS and MML algorithms, respectively. Since MML cannot be applied to the binned output of 2DSWML, only the NLLS algorithm is used. Mass bins below 10⁷.6 \( h^{-2} M_\odot \) and above 10¹⁰.5 \( h^{-2} M_\odot \) generally have larger errors due to small number statistics. We present the best fit parameters for each fitted Schechter function in Table 4. The derived parameters from both fitting methods are consistent within the error. For the

---

**Figure 12.** Projected number distribution of late-type SDSS galaxies (as described in Section 5.3) in the two AUDS fields (FIELD17H and GAL2577), and the reference volume.

**Figure 13.** Relative overdensities are derived from blue SDSS galaxies. The figure shows it as a function of redshift in the two AUDS fields, FIELD17H and GAL2577, in redshift bins of width 0.02. The shaded grey represents the Poisson error.
We find that the bin width affects the parameters derived using the NLLS fitting. The effect is prominent when the sample size is small in some bins. On the other hand, MML does not use bins, and takes into account the error in the masses and is generally a more robust algorithm. Comparing the HIMFs derived from the MML fitted $\Sigma_1/V_{\text{max}}$ and the NLLS fitted 2DSWML, we find that the derived parameters are consistent within errors. We adopt the result from the MML fitted $\Sigma_1/V_{\text{max}}$ method for the rest of the paper.

We also compare with the result from AUDS60 (Hoppmann et al. 2015) and find a similar value for $\alpha$, but a smaller $M^*$ and a larger $\Phi_*$, though all within the combined errors. The difference is that we have included galaxies with $z > 0.13$, which were discarded in AUDS60 due to strong RFI. This gives us a more distant volume, that we have included galaxies with $z > 0.13$, which were discarded in AUDS60 due to strong RFI. This gives us a more distant volume,
which in turn influences the calculation of effective volume for high mass galaxies.

Figure 14 also compares with the results from HIPASS (Zwaan et al. 2005), ALFALFA40 (Martin et al. 2010), and ALFALFA100 (Jones et al. 2018). Our result shows a good consistency with HIPASS and ALFALFA40 at the low-mass end. However, ALFALFA100 gives a flatter slope, \( \alpha = -1.25 \). The knee mass \( M^* \) in AUDS100 is larger than in HIPASS and ALFALFA, but still consistent at the 2σ level with ALFALFA40 and ALFALFA100. For the normalization parameters \( \Phi_\ast \), Zwaan et al. (2005) measured \( \Phi_\ast = 4.8 \times 10^{-3} h_\odot^3 \text{Mpc}^{-3} \) for HIPASS, and Martin et al. (2010) and Jones et al. (2018) measured \( \Phi_\ast = 4.8 \times 10^{-3} h_\odot^3 \text{Mpc}^{-3} \) and \( 4.5 \times 10^{-3} h_\odot^3 \text{Mpc}^{-3} \) for ALFALFA40 and ALFALFA100, respectively. The AUDS100 value of \( \Phi_\ast = 2.41 \times 10^{-3} h_\odot^3 \text{Mpc}^{-3} \) is significantly smaller than the results from both HIPASS and ALFALFA.

To investigate if the lower normalization reflects residual cosmic variance, even after the corrections discussed in Section 5.3, we separately fit the two fields, FIELD17H and GAL2577. These fields are located in opposite parts of the sky (Figure 12), and differ in their SDSS spectroscopic redshift density (80 and 185 late-type galaxies, respectively) and HI source density (95 and 152, respectively). Table 4 and Figure 15 indicate that the two fields in fact have similar values for \( \alpha \) and \( \Phi_\ast \), but that the lower density FIELD17H has a slightly larger \( M^* \) with 1σ significance. Considering the possible influence from the slightly different mass ranges in the two sub-samples, we narrow the mass range to \( 10^{6.6} - 10^{10.3} h_\odot^{-2} \text{M}_{\odot} \), giving 89 and 149 galaxies in FIELD17H and GAL2577, respectively. The MML method gives Schechter parameters \( \log(M^* [h_\odot^{-2} \text{M}_{\odot}]) = 9.85 \pm 0.13 \) and \( \alpha = -1.18 \pm 0.13 \) in FIELD17H, and \( \log(M^* [h_\odot^{-2} \text{M}_{\odot}]) = 9.97 \pm 0.10 \) and \( \alpha = -1.33 \pm 0.07 \) in GAL2577. Both \( M^* \) and \( \alpha \) are consistent with each other within 1σ. There are numerous studies of the variation of the HIMF with environment (see Section 1). However, the overall effects of environment on the HIMF are currently not well understood and the results are not consistent between studies. Wide field surveys such as WALLABY (Koribalski et al. 2020) will be required to further examine the role of environment on the HIMF.

To assess the effects of completeness and reliability, we also calculate the HIMF using the high reliability and high completeness Sub sample discussed in Section 4.5. This gives 148 galaxies as compared to 247 galaxies in AUDS100. Figure 16 compares the two HIMFs as as fitted with the MML algorithm (\( \Sigma I/V_{\text{max}} \) method) to derive the Schechter function parameters (see Table 4). There appears to be a deficit of low and high-mass galaxies, reflected in a steep faint-end slope and smaller value for \( M^* \), respectively, but a larger value of \( \Phi_\ast \). The difference is probably due to the lack of galaxies at both mass ends of the mass distribution in the Sub sample (see bottom panel of Figure 16). If we exclude the 10 AUDS100 galaxies outside of the range \( 7.0 < \log(M_{\text{HI}} [h_\odot^{-2} \text{M}_{\odot}]) < 10.5 \) (indicated by Ref) and re-derive the HIMF, the results become consistent (see Figure 16).

To further assess the effect of completeness on the HIMF, we use sub-samples with different completeness cutoffs, then re-compute and re-fit the HIMF. The results are shown in Figure 17. No substantial variation of HIMF parameters is seen, although the slope \( \alpha \) appears to gradually steepen with increasingly complete
sub-samples. At all completeness levels, the effective volume is corrected for completeness using Equation 6, so the lack of substantial change in the HIMF parameters is an indication that our completeness corrections are reasonably accurate.

7 DISCUSSION

7.1 Cosmic HI Density $\Omega_{\text{HI}}$

HI mass density ($\rho_{\text{HI}}$) is derived by integrating the mass-weighted HIMF. It can be approximated with the analytic expression:

$$\rho_{\text{HI}} = \Gamma(\alpha + 2)\Phi_* M^\alpha$$  \hspace{1cm} (10)

in which $\Gamma$ is Euler Gamma function, and $\Phi_*$, $M^\alpha$ and $\alpha$ are the best fitted parameters of the Schechter function. The cosmic HI density is then given by:

$$\Omega_{\text{HI}} = \frac{\rho_{\text{HI}}}{\rho_c(z=0)} = \frac{8\pi G}{3H_0^2}\rho_{\text{HI}}$$  \hspace{1cm} (11)

where $G$ is gravitational constant, $4.30 \times 10^{-9}$ Mpc M$_{\odot}^{-1}$ (km s$^{-1}$)$^2$, and $H_0$ is Hubble constant, $70$ h$_{70}$ km s$^{-1}$ Mpc$^{-1}$ at $z = 0$. We also compute $\Omega_{\text{HI}}$ in a direct method by summing up the mass weighted HIMF data points. Taking into account the error of $\log(M [h_{70}^{-2} M_\odot])$, $\sqrt{\log(M_{\text{HI}} [h_{70}^{-2} M_\odot])^2/12}$, $\Omega_{\text{HI}}$(sum) has larger uncertainty than $\Omega_{\text{HI}}$(int). We list the computed $\Omega_{\text{HI}}$ for different samples in Table 4. We find $\Omega_{\text{HI}}$(int) $= 3.55(0.30) \times 10^{-2} h_{70}^{-1}$ and $\Omega_{\text{HI}}$(sum) $= 3.93(0.68) \times 10^{-2} h_{70}^{-1}$ for AUDS100, which is reasonably consistent with results from HIPASS (3.75(0.43) $\times 10^{-2} h_{70}^{-1}$) and ALFALFA (3.9(0.1) $\times 10^{-2} h_{70}^{-1}$) within a 1-$\sigma$ uncertainty.

7.2 Evolution

The sensitivity of AUDS makes it suitable for beginning to explore the evolution of $\Omega_{\text{HI}}$ within the range covered by the ALFA receiver. We divide AUDS into ten redshift bins containing equal numbers of galaxies. Due to sensitivity and cosmic variance, the span for each redshift bin is not the same. We also adopt the characteristic mass and low-mass slope of the full AUDS100 sample, and only fit for the normalization $\Phi_*$ in the different redshift bins. The results are shown in Figure 18. There is no clear trend with redshift. This is consistent with the the prediction from cosmological simulations in Lagos et al. (2014); Kim et al. (2015); Davé et al. (2017), in which the $\Omega_{\text{HI}}$ shows little evolution for $z < 2$. The upturn in the lowest-redshift bin may be an indication of existence of an ‘ankle’ at the low-mass end of HIMF, rather than a change in overall density. Most of the galaxies in this bin have log$(M_{\text{HI}} [h_{70}^{-2} M_\odot]) < 8$.

Several techniques are available for measuring $\Omega_{\text{HI}}$ at different redshifts, which gives an insight into galaxy evolution. Detecting HI in individual galaxies in the 21-cm emission line is the main technique at low redshift (e.g. Hoppmann et al. 2015; Jones et al. 2018). Due to tremendous observing time required beyond the local Universe, HI spectral stacking (Rhee et al. 2018; Hu et al. 2019) and HI intensity mapping (Chang et al. 2010; Masui et al. 2013) techniques are preferred at intermediate redshift. For high-redshift galaxies, damped Lyman alpha (DLA) absorption lines has been the main method to derive $\Omega_{\text{HI}}$ (Rao et al. 2017; Grasha et al. 2020) (21-cm absorption line measurements are also possible, but have a temperature dependence).

We show selected measurements of $\Omega_{\text{HI}}$ using the above techniques in Figure 19, together with predictions from a semi-analytic model (Kim et al. 2015) and a hydrodynamic simulation (MUFASA) (Davé et al. 2017). All the measurements of $\Omega_{\text{HI}}$ from previous studies are converted to flat cosmology with $H_0 = 70$ h$_{70}$ km s$^{-1}$ Mpc$^{-1}$, $\Omega_M = 0.3$, $\Omega_{\Lambda} = 0.7$. For the studies using the DLA method, we calculated $\Omega_{\text{HI}} = \Omega_{\text{HI,DLA}}^\text{FLA} / \mu$, where $\Omega_{\text{HI,DLA}}^\text{FLA}$ accounts for the mass contribution of helium. A gradual increase for $\Omega_{\text{HI}}$ with redshift is shown, though with a large scatter around $z \sim 1$. This trend is modelled well by MUFASA. The Kim et al. (2015) model drops at high redshifts, where there may be significant gas reservoirs in the circumgalactic medium. Our result (red star) agrees with the studies in local Universe, and is consistent with both models.

In order to investigate whether the HIMF can evolve despite the lack of any apparent change in $\Omega_{\text{HI}}$, we divide the AUDS100 sample into two equal sub-samples at $z = 0.09$. There are 123 galaxies in low-redshift sample and 124 in the high-redshift sample. The mass distribution in each sample is shown in Figure 20. We compute the HIMF using the $\Sigma_1/V_{\text{max}}$ (MML) method, and show the results in Figure 20 and Table 4. Due to the large deviation in the HIMF for the high-redshift sample in the lowest two bins, we removed the three lowest-mass galaxies solely for the purpose of fitting the Schechter function. Both the low and high-redshift samples have slopes consistent with the AUDS100 sample within the 1-$\sigma$ uncertainty, although the high-redshift sample misses the low-mass galaxies required to better constrain this. The high-redshift sample appears to have a larger characteristic mass. However, there are only two galaxies with log$(M_{\text{HI}} [h_{70}^{-2} M_\odot]) > 10.0$ in the low-redshift sample. The formal significance of the difference is only 1-$\sigma$. To mitigate the effect of the different mass ranges in the two sub-samples shown in Figure 20, we choose a common range $(10^{8.94} - 10^{10.40}) h_{70}^{-2} M_\odot$. In this range, there are 62 galaxies in the low-$z$ sample and 119 galaxies in the high-$z$ sample. The small numbers give large uncertainties for both $M^\alpha$ and $\alpha$ in the low-$z$ sample, but the significance of the difference between the high- and low-$z$ sample reduces further. However, as already noted, there remains a 2-$\sigma$ difference when comparing the mean AUDS100 results with the larger HIPASS and ALFALFA samples (see Table 5).
8 CONCLUSIONS

We employed the Arecibo Telescope to perform the Arecibo Ultra-Deep Survey (AUDS), the deepest blind HI survey so far conducted with this telescope. The survey area is 1.35 deg$^2$ with redshift coverage up to 0.16 (the limit of the ALFA receiver). The total on-source integration time exceeds 700 hours, and the final noise level is $\sim$75 $\mu$Jy at a frequency resolution of 21.4 kHz. Our final AUDS100 catalogue contains 247 galaxies with an HI mass range from $10^6$ to $10^{10.76} h^2_{70}$ M$_\odot$. For statistical purposes, a high-completeness, high-reliability sub-sample, consisting of 148 galaxies was also created. The AUDS100 catalogue consists galaxies detected by a source-finder and by manual inspection of the data cubes. Simulations were performed to measure completeness, and checked using a modified V/V$_{max}$ test.

The HI mass function (HIMF) was computed using the $\Sigma$V/V$_{max}$ and 2D stepwise maximum-likelihood (2DSWML) methods. Due to the small field-of-view, cosmic variance in AUDS is substantial, but was able to be corrected using the measured density of SDSS spectroscopic data for blue galaxies which have a similar density bias to HI-rich galaxies. Fits to the HIMF were made using different methods, including a modified maximum likelihood method, which makes better allowance for errors and limits, and is robust against binning artefacts. The best-fit parameters for a Schechter function fit are: $\Phi_\ast = 2.41(0.57) \times 10^{-3} h^3_{70}$ Mpc$^{-3}$, $\log(M^\ast/h^2_{70} M_\odot) = 10.15(0.09)$, and $\alpha = -1.37(0.05)$. Using these parameters, we compute the AUDS cosmic HI density, $\Omega_{HI}$ = 3.55(0.30) $\times 10^{-4}$ $h^4_{70}$ Mpc$^{-3}$. Our result is consistent with previous measurements from HIPASS and ALFALFA.

Evolutionary trends were inspected by dividing the AUDS100 catalogue into low and high-redshift sub-samples. A slight increase in the characteristic mass $M^\ast$ is apparent in the higher redshift bin, which becomes more significant when compared with HIPASS and ALFALFA. No evolutionary trend was detected for $\Omega_{HI}$, in line with other recent studies.

Finally, we investigated environmental trends by separately computing HIMFs in the two AUDS fields which have galaxy num-
ber densities differ by a factor of two. However, no clear changes were measured between the two fields.

Further exploration of the effect of redshift on the HIMF and $\Omega_{\text{HI}}$ will best be performed in future with deep blind HI surveys using ASKAP, FAST and MeerKAT and, eventually, the Square Kilometre Array (SKA). Stacking and intensity mapping experiments have additional dependencies on stronger evolutionary trends at optical/IR wavebands. Further exploration of the effect of environment will be best explored with wide-area blind HI surveys using ASKAP (WALLABY; Koribalski et al. (2020)), FAST (CRAFTS, Li & Duan (2019)) and WSRT/APERTIF (WEAVE-Apertif Survey; Hess et al. (2020)).

ACKNOWLEDGEMENTS

We would like to thank the China Scholarship Council (CSC) from the Ministry of Education of P.R. China and the International Centre for Radio Astronomy Research (ICRAR) in the University of Western Australia (UWA) for financial support during Hongwei Xi studying at ICRAR. The Arecibo Observatory is operated by SRI International under a cooperative agreement with the National Science Foundation (AST-1100968), and in alliance with Ana G. Méndez-Universidad Metropolitana, and the Universities Space Research Association. In light of the recent loss of the telescope, we also wish to thank the staff of the Arecibo Observatory for their tremendous support during the planning and execution of these observations, and for the careful curation and shipping of the data. Without their support and the far-reaching capability of this telescope, this research would not have been possible. Parts of this research were supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia.

DATA AVAILABILITY

Readers interested in the raw data can contact Arecibo Observatory. The derived data supporting the findings of this study are available from the corresponding author on reasonable request.

REFERENCES

Abolfathi B., et al., 2018, ApJS, 235, 42
Baker A. J., Blyth S., Holwerda B. W., LADUMA Team 2018, in American Astronomical Society Meeting Abstracts #231. p. 231.07
Barnes D. G., et al., 2001, MNRAS, 322, 486
Bird S., Garnett R., Ho S., 2017, MNRAS, 466, 2111
Blyth S., et al., 2016, in MeetKAT Science: On the Pathway to the SKA. p. 4
Braun R., 2012, ApJ, 749, 87
Catinella B., Haynes M. P., Giovanelli R., Gardner J. P., Connolly A. J., 2008, ApJ, 685, L13
Chang T.-C., Pen U.-L., Bandura K., Peterson J. B., 2010, Nature, 466, 463
Condon J. J., Cotton W. D., Greisen E. W., Yin Q. F., Perley R. A., Taylor G. B., Broderick J. J., 1998, AJ, 115, 1693
Courtois H. M., Tully R. B., 2015, MNRAS, 447, 1531
Crighton N. H. M., et al., 2015, MNRAS, 452, 217
Dalcanton J. J., Spargel D. N., Summers F. J., 1997, ApJ, 482, 659

Davé R., Rafieferantsoa M. H., Thompson R. J., Hopkins P. F., 2017, MNRAS, 467, 115
Davis M., Huchra J., 1982, ApJ, 254, 437
Delhaize J., Meyer M. J., Staveley-Smith L., Boyle B. J., 2013, MNRAS, 433, 1398
Donley J. L., et al., 2005, AJ, 129, 220
Doyle M. T., et al., 2005, MNRAS, 361, 34
Driver S. P., Robotham A. S. G., 2010, MNRAS, 407, 2131
Efstathiou G., Ellis R. S., Peterson B. A., 1988, MNRAS, 232, 431
Freudling W., et al., 2011, ApJ, 727, 40
Giovanelli R., et al., 2005, AJ, 130, 2598
Grasha K., Darling J., Leroy A. K., Bolatto A. D., 2020, MNRAS, 498, 883
Hess K. M., Falcon-Barroso J., Ascasibar Y., Perez-Martin L., Serra P., Weinmann A., Arecibo-APERTIF Team 2020, in American Astronomical Society Meeting Abstracts #235. p. 459.03
Hoppmann L., 2014, PhD thesis, University of Western Australia
Hoppmann L., Staveley-Smith L., Freudling W., Zwaan M. A., Minchin R. F., Calabretta M. R., 2015, MNRAS, 452, 3726
Hu W., et al., 2019, MNRAS, 489, 1619
Jones M. G., Papastergis E., Haynes M. P., Giovanelli R., 2016, MNRAS, 457, 4393
Jones M. G., Haynes M. P., Giovanelli R., Moorman C., 2018, MNRAS, 477, 2
Kanekar N., Sethi S., Dwarkanath K. S., 2016, ApJ, 818, L28
Kim H.-S., Wyithe J. S. B., Power C., Park J., Lagos C. d. P., Baugh C. M., 2015, MNRAS, 453, 2315
Koribalski B. S., et al., 2020, A&SS, 365, 118
Lagos C. D. P., Baugh C. M., Zwaan M. A., Lacey C. G., Gonzalez-Perez V., Power C., Swinbank A. M., van Kampen E., 2014, MNRAS, 440, 920
Lah P., et al., 2007, MNRAS, 376, 1357
Li D., Duan R., 2019, in 2019 IEEE International Conference on Microwaves, Antennas, Communications and Electronic Systems (COMCAS). pp 1–4
Loveday J., 2000, MNRAS, 312, 357
Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, ApJ, 723, 1359
Mawatari K. W., et al., 2013, ApJ, 763, L20
McGaugh S. S., 1996, MNRAS, 280, 337
Meyer M. J., et al., 2004, MNRAS, 350, 1195
Meyer M., Robotham A., Obreschkow D., Westmeier T., Duffy A. R., Staveley-Smith L., 2017, Publ. Astron. Soc. Australia, 34, 52
Moorman C. M., Vogelezang M. S., Hoyle F., Pan D. C., Haynes M. P., Giovanelli R., 2014, MNRAS, 444, 3559
Neelamman M., Prochaska J. X., Ribaudo J., Lehner N., Howk J. C., Rafelski M., Kanekar N., 2016, ApJ, 818, 113
Noteddaeml P., Petitjean P., Ledoux C., Srianand R., 2009, A&A, 505, 1087
Noteddaeml P., et al., 2012, A&A, 547, L1
Obreschkow D., Murray S. G., Robotham A. S. G., Westmeier T., 2018, MNRAS, 474, 5500
Péroux C., McMahon R. G., Storrie-Lombardi L. J., Irwin M. J., 2003, MNRAS, 346, 1103
Prochaska J. X., Wolfe A. M., 2009, ApJ, 696, 1543
Prochaska J. X., Herbert-Fort S., Wolfe A. M., 2005, ApJ, 635, 123
Rao S. M., Turnshek D. A., 2000, ApJS, 130, 1
Rao S. M., Turnshek D. A., Nestor D. B., 2006, ApJ, 636, 610
Rao S. M., Turnshek D. A., Sardane G. M., Monier E. M., Nestor D. B., 2006, ApJ, 636, 610
Rao S. M., Turnshek D. A., Sardane G. M., Monier E. M., 2017, MNRAS, 471, 3428
Rauzy S., 2001, MNRAS, 324, 51
Reynolds T. N., et al., 2019, MNRAS, 482, 3591
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
Rhee J., Lah P., Briggs F. H., Chengalur J. N., Colless M., Willner S. P., Ashby M. L. N., Le Fèvre O., 2018, MNRAS, 473, 1879
Said K., Kraan-Korteweg R. C., Staveley-Smith L., 2019, MNRAS, 486, 1796

MNRAS 000, 1–16 (2015)
Sancisi R., Fraternali F., Oosterloo T., van der Hulst T., 2008, A&ARv, 15, 189
Schechter P., 1976, ApJ, 203, 297
Schmidt M., 1968, ApJ, 151, 393
Serra P., et al., 2015a, MNRAS, 448, 1922
Serra P., et al., 2015b, MNRAS, 452, 2680
Solanes J. M., Giovanelli R., Haynes M. P., 1996, ApJ, 461, 609
Songaila A., Cowie L. L., 2010, ApJ, 721, 1448
Sprayberry D., Impey C. D., Irwin M. J., Bothun G. D., 1997, ApJ, 482, 104
Springob C. M., Haynes M. P., Giovanelli R., 2005, ApJ, 621, 215
Staveley-Smith L., Kraan-Korteweg R. C., Schröder A. C., Henning P. A., Koribalski B. S., Stewart I. M., Heald G., 2016, AJ, 151, 52
Strateva I., et al., 2001, AJ, 122, 1861
Westmeier T., Jurek R., Obreschkow D., Koribalski B. S., Staveley-Smith L., 2014, MNRAS, 438, 1176
Westmeier T., et al., 2017, MNRAS, 472, 4832
Wolfe A. M., Gawiser E., Prochaska J. X., 2005, ARA&A, 43, 861
Wong O. I., et al., 2006, MNRAS, 371, 1855
York D. G., et al., 2000, AJ, 120, 1579
Zafar T., Péroux C., Popping A., Milliard B., Deharveng J. M., Frank S., 2013, A&A, 556, A141
Zwaan M. A., et al., 2003, AJ, 125, 2842
Zwaan M. A., et al., 2004, MNRAS, 350, 1210
Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, MNRAS, 359, L30

SUPPORTING INFORMATION

Additional Supporting Information can be found in the online version of this article. The directory “Spectra” consists of all the spectra of AUDS100 HI galaxies in .png format. The directory “Table2” gives the description file, “ReadMe”, and data file, “table2.dat”.

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