












# MIGHTEE-H I: the first MeerKAT H I mass function from an untargeted interferometric survey

Anastasia A. Ponomareva <sup>1</sup>★, Matt J. Jarvis <sup>1,2</sup>, Hengxing Pan <sup>2</sup>, Natasha Maddox <sup>3,4</sup>,  
 Michael G. Jones <sup>5</sup>, Bradley S. Frank, <sup>6,7,8</sup> Sambatriniaina H. A. Rajohnson <sup>6</sup>, Wanga Mulaudzi <sup>6,9</sup>,  
 Martin Meyer, <sup>10</sup> Elizabeth A. K. Adams, <sup>11,12</sup> Maarten Baes <sup>13</sup>, Kelley M. Hess, <sup>11,14</sup> Sushma Kurapati, <sup>6</sup>  
 Isabella Prandoni <sup>15</sup>, Francesco Sinigaglia, <sup>16,17</sup> Kristine Spekkens, <sup>18</sup> Madalina Tudorache <sup>1</sup>,  
 Ian Heywood, <sup>1,7,19</sup> Jordan D. Collier <sup>8,20,21</sup> and Srikrishna Sekhar <sup>8,22</sup>

*Affiliations are listed at the end of the paper*

Accepted 2023 April 25. Received 2023 April 21; in original form 2023 March 3

## ABSTRACT

We present the first measurement of the H I mass function (HIMF) using data from MeerKAT, based on 276 direct detections from the MeerKAT International GigaHertz Tiered Extragalactic Exploration (MIGHTEE) Survey Early Science data covering a period of approximately a billion years ( $0 \leq z \leq 0.084$ ). This is the first HIMF measured using interferometric data over non-group or cluster field, i.e. a deep blank field. We constrain the parameters of the Schechter function that describes the HIMF with two different methods:  $1/V_{\max}$  and modified maximum likelihood (MML). We find a low-mass slope  $\alpha = -1.29^{+0.37}_{-0.26}$ , ‘knee’ mass  $\log_{10}(M_*/M_{\odot}) = 10.07^{+0.24}_{-0.24}$  and normalization  $\log_{10}(\phi_*/\text{Mpc}^{-3}) = -2.34^{+0.32}_{-0.36}$  ( $H_0 = 67.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ) for  $1/V_{\max}$ , and  $\alpha = -1.44^{+0.13}_{-0.10}$ , ‘knee’ mass  $\log_{10}(M_*/M_{\odot}) = 10.22^{+0.10}_{-0.13}$  and normalization  $\log_{10}(\phi_*/\text{Mpc}^{-3}) = -2.52^{+0.19}_{-0.14}$  for MML. When using  $1/V_{\max}$  we find both the low-mass slope and ‘knee’ mass to be consistent within  $1\sigma$  with previous studies based on single-dish surveys. The cosmological mass density of H I is found to be slightly larger than previously reported:  $\Omega_{\text{HI}} = 5.46^{+0.94}_{-0.99} \times 10^{-4} h_{67.4}^{-1}$  from  $1/V_{\max}$  and  $\Omega_{\text{HI}} = 6.31^{+0.31}_{-0.31} \times 10^{-4} h_{67.4}^{-1}$  from MML but consistent within the uncertainties. We find no evidence for evolution of the HIMF over the last billion years.

**Key words:** surveys – galaxies: evolution – galaxies: ISM – galaxies: luminosity function, mass function – radio lines: galaxies.

## 1 INTRODUCTION

Hydrogen is the most abundant element in the Universe, constituting around 75 per cent of the total baryonic matter. It is not just the primary building block of all the structure we see in the Universe, but it also plays a crucial role in the formation and evolution of galaxies. Stars in galaxies are born in dense giant molecular clouds, which themselves form due to the cooling of neutral hydrogen. It is unclear, however, where galaxies acquire the fuel to keep forming stars and how star-forming gas gets recycled. Moreover, observational studies show that the rate at which new stars are born in galaxies has been continuously decreasing for the last several billion years (e.g. Madau & Dickinson 2014; Neeleman et al. 2016; Walter et al. 2020). This decrease must be connected to the amount of the cold gas available to form stars.

The evolution in the cosmic H I density ( $\Omega_{\text{HI}}$ ) is one of the major factors in understanding of the cosmic star formation rate (SFR) density and the mass assembly of galaxies, since H I serves as a raw material for the buildup of stellar mass (Maddox et al. 2015; Pan et al. 2022).  $\Omega_{\text{HI}}$  also provides insights into the processes governing the distribution and evolution of cool gas in the Universe. At higher redshifts ( $z > 0.2$ )  $\Omega_{\text{HI}}$  has been measured indirectly using either H I

spectral stacking (Delhaize et al. 2013; Bera et al. 2018; Rhee et al. 2018; Chowdhury et al. 2020), damped Lyman  $\alpha$  (DLA) absorption line systems (Péroux et al. 2003; Noterdaeme et al. 2012; Grasha et al. 2020), or the [C II]-to-H I conversion factor (Heintz et al. 2021, 2022). Overall, these studies agree that the H I mass density of the Universe shows minor evolution with time, in contrast to the molecular hydrogen, which exhibits strong evolution and mirrors that of the global SFR density (Péroux & Howk 2020).

At low redshift  $\Omega_{\text{HI}}$  is measured directly by summing up the amount of gas in galaxies, usually via determining the H I mass function (HIMF), which is the neutral hydrogen equivalent of the stellar mass function (Baldry et al. 2012). The HIMF defines the number of galaxies per cubic Mpc as a function of H I mass, and its shape determines how the neutral gas in the Universe is distributed over galaxies of different H I masses. At  $z = 0$  the shape of the HIMF has been extensively studied (Zwaan et al. 2003; Martin et al. 2010; Jones et al. 2018; Said, Kraan-Korteweg & Staveley-Smith 2019). These studies have shown that the HIMF follows a Schechter function with a power-law low-mass slope ( $\alpha$ ) and an exponential fall-off at the high-mass end, beyond a ‘knee’ mass ( $M_*$ ; Zwaan et al. 1997). Even though an agreement between various studies on the faint-end slope of the HIMF has never been reached, they agree that the HIMF depends on the morphological type of galaxies and on the environment where galaxies reside. For example, the overall HIMF tends to have a steeper low-mass slope than when just the Local

\* E-mail: [anastasia.ponomareva@physics.ox.ac.uk](mailto:anastasia.ponomareva@physics.ox.ac.uk)

Group or individual groups of galaxies are considered (Zwaan et al. 2005; Jones et al. 2016, 2020; Busekool et al. 2021). Therefore, a way to probe an evolution of galaxies over cosmic times is to study variations of the HIMF as a function of morphology, environment, and redshift.

The HIMF is complementary to the stellar mass function in a way that it provides additional insights on galaxies' assembly processes, since the correlation between halo mass and neutral gas mass is very different from that between halo mass and stellar mass (Guo et al. 2020; Yasin et al. 2022). Well-constrained stellar and HIMFs put major constraints on the theoretical models of galaxy formation and evolution, as any successful theory of galaxy formation and evolution should be able to reproduce both mass functions simultaneously at any redshift (Crain et al. 2017; Diemer et al. 2018; Davé et al. 2020).

To date the most accurate HIMF in the Local Universe ( $z \leq 0.06$ ) has been measured by the Arecibo Legacy Fast ALFA (ALFALFA) survey (ALFALFA; Martin et al. 2010; Jones et al. 2018). The resulting HIMF from the ALFALFA 100 per cent survey (hereafter ALFALFA 100; Jones et al. 2018) indicated that most H I gas in the Local Universe resides in the high stellar mass galaxies. Additionally, this study demonstrates the effect of the environment on the HIMF, indicating that the low-mass slope is particularly sensitive. Jones et al. (2018) also report a change in the 'knee' mass when only part of the sample is used, which is also attributed to the environmental dependence. Other observational studies have also shown a flattening of the low-mass slope in high-density environments such as groups and clusters (Pisano et al. 2011; Westmeier et al. 2017; Busekool et al. 2021), suggesting that the shape of the HIMF depends on the local and global environment (Jones et al. 2020).

At redshifts beyond the Local Universe ( $z > 0.05$ ), statistical measurements from direct detections in emission become increasingly challenging, due to the intrinsic faintness of the H I line. However, using associated 21-cm absorption can provide complementary information and push H I studies to higher redshifts, although these studies are also limited due to the need of strong continuum background sources (Gupta et al. 2006; Maccagni et al. 2017; Aditya et al. 2021). To date, only two surveys have been able to provide measurements of the HIMF parameters beyond the Local Universe. One is the Arecibo Ultra-Deep Survey (AUDS; Xi et al. 2021), spanning a redshift range  $0 < z < 0.16$ . Another is the Blind Ultra-Deep H I Environmental Survey (BUDHIES; Gogate 2022), which has constructed the HIMF and measured  $\Omega_{\text{H I}}$  at  $z \sim 0.2$  using direct H I detections for the first time, but is centred on two galaxy cluster fields. In general, the results of these surveys are in agreement that there is little to no evolution of the H I content in the Universe up to  $z = 0.2$ . However, they find a somewhat different low-mass slopes and 'knee' masses. For instance, the results from the AUDS survey are in good agreement with those from ALFALFA 100, especially at the low-mass end, while BUDHIES finds a somewhat steeper  $\alpha$  and lower  $M_*$ . However, their results are subject to various significant uncertainties, such as completeness corrections and cosmic variance, as well as the environment in which the galaxies reside. Bera et al. (2022) have studied the HIMF of star-forming galaxies at  $z \sim 0.35$  using stacking technique, and found a significant evolution of the HIMF over the last 4 Gyr, especially at the high-mass end. In particular, in agreement with BUDHIES, they find a lower 'knee' mass and steeper low-mass slope in comparison to the results at  $z = 0$  from ALFALFA 100.

The advent of the next-generation deep, blind H I surveys using new telescopes such as Australian Square Kilometre Array Pathfinder (ASKAP; Johnston et al. 2008), APERTure Tile In Focus (APERTIF; Adams et al. 2022), MeerKAT (Jonas 2009), and eventually the Square Kilometre Array (SKA) will improve our understanding

of the redshift evolution of the HIMF and  $\Omega_{\text{H I}}$ . Surveys such as the Looking At the Distant Universe with the MeerKAT Array (LADUMA; Blyth et al. 2018), the Deep Investigation of Neutral Gas Origins (DINGO; Meyer 2010; Rhee et al. 2023), and the COSMOS H I Large Extragalactic Survey (CHILES; Hess et al. 2019; Dodson et al. 2022) are specifically designed to systematically study H I in galaxies over a large range of redshifts. Another such survey is the MeerKAT International GigaHertz Tiered Extragalactic Exploration (MIGHTEE; Jarvis et al. 2018), which when complete will detect more than 1000 galaxies in H I up to  $z \sim 0.6$ , therefore allowing the systematic study of the evolution of the neutral gas content of galaxies over the past five billion years.

In this paper, we present the first measurement of the HIMF using data from the MeerKAT telescope. We use the MIGHTEE Early Science data in order to construct the HIMF over the last billion years ( $0 \leq z \leq 0.084$ ) and calculate the cosmic H I mass density over this redshift range. This is also the first HIMF measured using interferometric data over non-group or cluster field, i.e. a deep blank field. Our work demonstrates the capabilities of MeerKAT and provides a benchmark for the future H I evolutionary studies with the SKA pathfinders.

This paper is organized as follows. In Section 2, we describe the MIGHTEE survey and the Early Science data. In Section 3, we present the H I mass measurements, and in Section 4, we describe how we measure the HIMF. In Section 5, we present the results and best-fitting parametrization of the Schechter function. A summary and conclusions are presented in Section 6.

Throughout this paper, we assume  $\Lambda$  cold dark matter ( $\Lambda$ CDM) cosmology parameters of  $H_0 = 67.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ,  $\Omega_{\text{m}} = 0.315$ , and  $\Omega_{\Lambda} = 0.685$  (Planck Collaboration VI 2020).

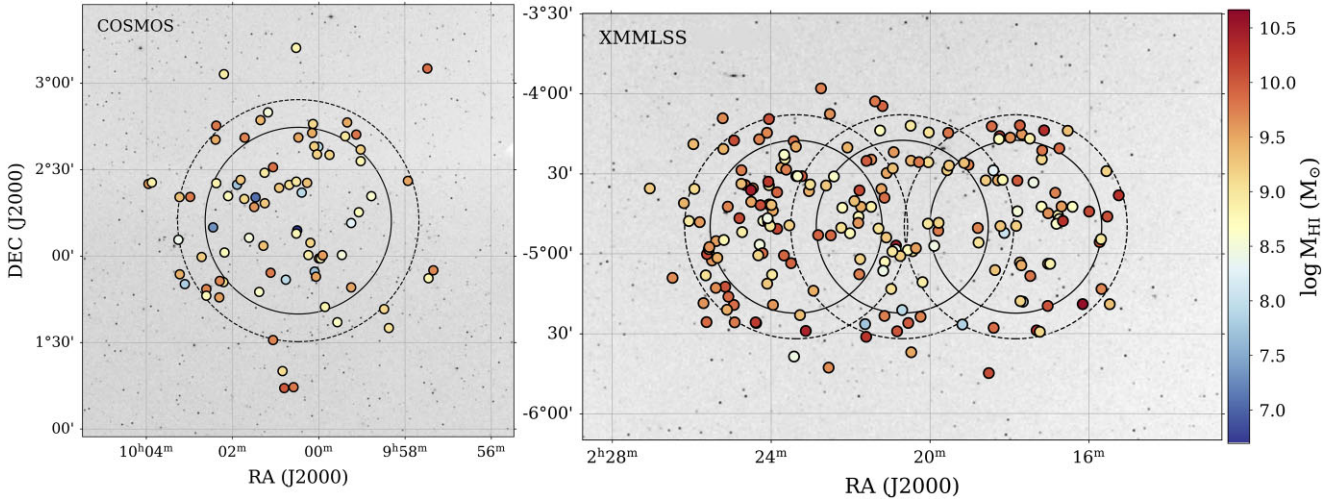
## 2 THE MIGHTEE SURVEY

The MeerKAT International GigaHertz Tiered Extragalactic Exploration (MIGHTEE) is a MeerKAT survey of four deep, extragalactic fields [Cosmic Evolution Survey (COSMOS), *XMM*-Large Scale Structure (*XMM*-LSS), *Extended Chandra Deep Field* South (ECDFS), European Large Area ISO Survey – South 1 (ELAIS-S1); Jarvis et al. 2018]. MeerKAT is a radio interferometer that consists of 64 offset Gregorian dishes and equipped with three receivers covering the frequency range from 580 to 3500 MHz (Jonas 2009). MIGHTEE is simultaneously a spectral line, continuum, and polarization survey.

For this study we use the Early Science data, which were collected as part of the H I emission project within the MIGHTEE survey. A detailed description of MIGHTEE-H I is presented in Maddox et al. (2021).

The Early Science MIGHTEE observations were conducted between mid-2018 and mid-2019 in L band ( $900 < \nu < 1670 \text{ MHz}$ ) with a limited spectral resolution (208 kHz, which corresponds to  $44 \text{ km s}^{-1}$  at  $z = 0$ ). A full description of the H I line data reduction strategy and data quality assessment will be presented in Frank et al. (in preparation).

The summary of the Early Science data used in this paper can be found in table 1 of Rajohnson et al. (2022). Briefly, they consist of observations of the COSMOS and *XMM*-LSS fields, and cover the redshift range  $0 \leq z \leq 0.084$ . The 30 per cent area of the main lobe of the primary beam of the COSMOS field is  $1.5 \text{ deg}^2$  that corresponds to one MeerKAT pointing, although we note that the full width at half-maximum (FWHM) of the primary beam is equal to  $\sim 0.9 \text{ deg}^2$  at 1420.405 MHz. The *XMM*-LSS field was covered by three overlapping pointings resulting in the observed area of  $3.3 \text{ deg}^2$  (Fig. 1). The total integration time for the COSMOS field was  $\sim 17 \text{ h}$ ,



**Figure 1.** The MIGHTEE-H I detections in COSMOS (left) and *XMM*-LSS (right) fields colour coded by their H I mass. The concentric circles represent the main lobe of the MeerKAT primary beam. The inner solid circle indicates the FWHM of the primary beam  $\sim 0.9 \text{ deg}^2$  and the outer 30 per cent level  $\sim 1.5 \text{ deg}^2$ .

and 13 h were spent on each pointing of the *XMM*-LSS field. The average rms noise per 208 KHz channel across the COSMOS field is  $\sigma_{\text{sv}} = 49 \mu\text{Jy}$ , and for the *XMM*-LSS field that average rms noise is  $\sigma_{\text{sv}} = 81 \mu\text{Jy}$ , corresponding to  $3\sigma$  H I column density sensitivity of  $4.05 \times 10^{19} \text{ atoms cm}^{-2}$  for COSMOS and  $9.83 \times 10^{19} \text{ atoms cm}^{-2}$  for *XMM*-LSS, respectively.

The source finding was performed visually by inspecting the H I data cubes with the Cube Analysis and Rendering Tool for Astronomy (CARTA; Comrie et al. 2020), and was not guided by the available deep optical information. The total Early Science sample consists of 276 objects, each of which has an identified optical counterpart in the very deep multiwavelength data over these fields. However, we note that a counterpart was not required for a source to be considered genuine.

Although even the full MIGHTEE survey will not be able to compete with the sky coverage of the ALFALFA survey ( $\sim 6900 \text{ deg}^2$  versus  $\sim 32 \text{ deg}^2$ ), the MIGHTEE Early Science flux limit is a factor of 10 deeper than the approximate flux limit of the large-area ALFALFA survey, and extends further in redshift (see fig. 8 in Maddox et al. 2021). For example, a galaxy with an H I mass equal to  $10^9 M_{\odot}$  and H I line width equal to  $100 \text{ km s}^{-1}$  would be detectable by ALFALFA out to a distance of  $\sim 80 \text{ Mpc}$  (Haynes et al. 2011; Jones et al. 2018), while a galaxy with the same parameters will be detectable by MIGHTEE to  $\sim 428 \text{ Mpc}$  (Maddox et al. 2021). This allows us to study both the low-mass slope and the ‘knee’ of the HIMF out to larger redshifts.

The total cosmological volume of the Early Science MIGHTEE survey is  $\sim 7000 \text{ Mpc}^3$ . The detections in both fields span approximately the same redshift range from  $z_{\text{min}} = 0.004$  to  $z_{\text{max}} = 0.082$  in COSMOS and  $z_{\text{max}} = 0.084$  in *XMM*-LSS, and lie almost exclusively within the area of  $1.5 \text{ deg}^2$  of each pointing (Fig. 1).

### 3 H I MASS MEASUREMENTS

The total H I mass of each galaxy was calculated following the prescription from Meyer et al. (2017):

$$\left(\frac{M_{\text{H I}}}{M_{\odot}}\right) = \frac{2.356 \times 10^5}{1+z} \left(\frac{D_{\text{L}}}{\text{Mpc}}\right)^2 \left(\frac{S}{\text{Jy km s}^{-1}}\right), \quad (1)$$

where  $D_{\text{L}}$  is the cosmological luminosity distance to the source,  $z$  is redshift, and  $S$  is the integrated H I flux density.

The integrated H I flux density has been calculated using the moment-0 maps constructed individually for each source, taking diffuse low column density emission into account. The detailed description of how the moment-0 maps were constructed is presented in Ponomareva et al. (2021) and Rajohnson et al. (2022). The error on the integrated flux  $S$  was calculated by projecting the source mask, used to construct the moment-0 map, to four emission-free regions around the detection. Then, the uncertainty in the integrated flux of a galaxy was defined as the mean rms scatter of the four flux measurements in these regions (Ramatsoku et al. 2016). As a result, the typical uncertainty on the H I mass varies from  $\sim 5$  per cent for the high-mass galaxies to  $\sim 20$  per cent for the lowest mass objects ( $M_{\text{H I}} \leq 10^8 M_{\odot}$ ). We note that due to MeerKAT’s excellent combination of sensitivity and  $uv$ -plane coverage any missed H I flux is negligible as compared to the single-dish telescopes. For example, we find an excellent agreement between the total fluxes of the overlapping sources from MIGHTEE and ALFALFA (Frank et al., in preparation).

The cosmological distance ( $D_{\text{L}}$ ) to each source has been calculated using the adopted cosmology, following the prescription of Meyer et al. (2017, equation 10). According to Tully et al. (2014) peculiar velocities are a negligible fraction of observed velocities at  $z > 0.03$ . In our sample we have 50 objects below this redshift with a mean  $\log_{10}(M_{\text{H I}}/M_{\odot}) = 8.3$ . The galaxy with the lowest systemic velocity of our sample has  $V_{\text{sys}} = 1238 \text{ km s}^{-1}$ , while a typical peculiar motion is  $\sim 300 \text{ km s}^{-1}$  (Darling & Truebenbach 2018). Adopting this value as the uncertainty on the systemic velocity results in an uncertainty on  $\log_{10}(M_{\text{H I}}) \sim 0.06 \text{ dex}$  for the galaxies at  $z < 0.03 \text{ dex}$  and  $z \sim 0.02 \text{ dex}$  for the whole sample. The overall resulting uncertainties on the H I mass due to peculiar velocities are therefore much smaller than the bin size used for our mass function calculation (0.3 dex), and also subdominant compared to the Poisson statistics combined with the sample variance (see Section 4.3) for our sample. We therefore do not attempt to correct for such peculiar motion.

## 4 CONSTRUCTING THE H I MASS FUNCTION

### 4.1 Completeness

Prior to performing a statistical analysis of any sample, determining the completeness is important. The completeness is usually calcu-



lated per H I mass bin and is an estimate of how many galaxies in that bin have been detected from the population as a whole given the limitations of the data (Gogate 2022).

Many factors can affect whether or not a galaxy is detected in the data. Effects such as primary beam attenuation, radio frequency interference (RFI), non-uniform noise distribution, and limited spatial (and/or spectral) resolution all play a role in our ability to detect sources, particularly near the sensitivity limit of the survey. Furthermore, in H I surveys in particular, our ability to detect a galaxy depends not only on its total intrinsic flux, but also on its orientation, i.e. galaxies with lower inclinations have higher flux per channel than more inclined galaxies. Low intrinsic flux and high inclinations greatly weaken the possibility of a galaxy to be detected (although the specifics depend on the method adopted for source finding), and therefore H I galaxy samples tend to be biased towards the most gas-rich and low-inclination sources in the survey volume.

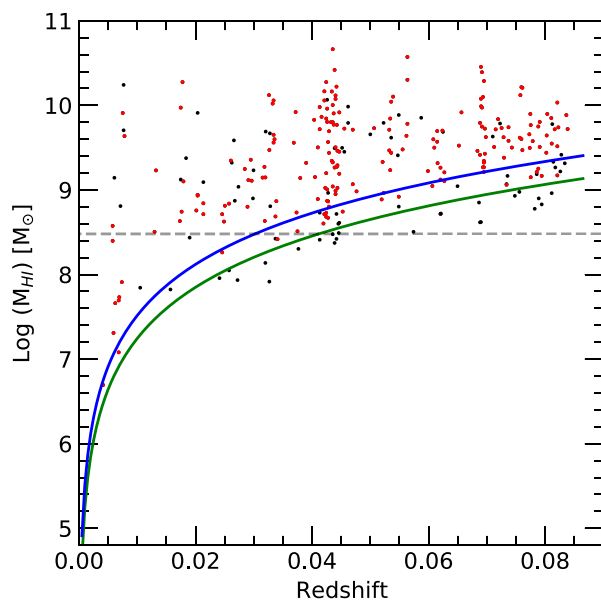
The most reliable way to determine the completeness of a survey is to inject artificial but realistic sources into the image cubes and recover them with exactly the same method that was used to find real sources. The recovery rate of the artificial sources as a function of their H I mass can then be used to correct the underlying HIMF of a survey for completeness.

This method works well when an automated source finder is available and can be trusted to find sources with high reliability, whilst robustly differentiating real sources from false positives. While there are ongoing efforts to build such source finders for the large H I interferometric surveys (e.g. Westmeier et al. 2021), there is still no definitive solution, especially for the early MeerKAT data with low-velocity resolution (Healy et al. 2021). For the MIGHTEE Early Science data, we elected to use visual source finding instead of automated methods. A group of people within the MIGHTEE team have examined the data visually, creating source catalogues, which were then merged with duplicates removed. Each source then has been cross-matched with the optical catalogue of each observed field. Therefore, we do not perform the completeness correction using injected sources in our data, since it would require the same group of people to repeat the visual source finding on a much larger sample, and it would become prohibitive. However, despite MeerKAT's superb combination of sensitivity and spatial resolution, we will undoubtedly be missing sources close to our adopted flux limit, however conversely, the sources we have identified are much more unlikely to be contaminated by artefacts, particular as they have all been cross-identified to counterparts in the exceptionally deep optical (Aihara et al. 2019), and near-infrared (McCracken et al. 2012; Jarvis et al. 2013) data over these fields (see Adams et al. 2022, for a full description of the combined data set).

Instead, rather than calculate the incompleteness, we adopt a conservative approach and limit our sample to a flux limit above which we are confident that we are very close to 100 per cent complete. We calculate a limiting line flux density ( $S_{\text{lim}}$ ) for each detection as

$$S_{\text{lim}} = \sqrt{\frac{W_{50}}{dv}} \sigma_{S_v} dv, \quad (2)$$

where  $W_{50}$  is the full width at half-maximum (FWHM) of the H I line, measured as described in Ponomareva et al. (2021),  $dv$  is the velocity resolution at the redshift of the source, and  $\sigma_{S_v}$  is the mean measured rms noise. Since the sensitivity of the telescope decreases with radius from the pointing centre, low-mass galaxies are preferentially detected within the FWHM of the primary beam (as seen in Fig. 1). Therefore, we measure rms noise in two areas: within the full width half-power (inner region) and within the 30 per cent



**Figure 2.** H I masses as a function of redshift for our sample. Black points indicate the original sample, while red points indicate the sample used to construct the HIMF. The curved green line indicates the  $5\sigma$  flux limit used for the simulations of the expected number counts in MIGHTEE-H I (Maddox, Jarvis & Oosterloo 2016; Maddox et al. 2021). The curved blue line represents the median flux limit  $5S_{\text{lim}}$  (see Section 4.1). The grey dashed horizontal line indicates an additional mass cut-off below which the sample galaxies were discarded (see Section 4.2).

of the total primary beam area (outer region). For COSMOS we find that within the half-power radius  $\sigma_{S_v} = 45 \mu\text{Jy}$  and at the 30 per cent power radius  $\sigma_{S_v} = 59 \mu\text{Jy}$ . For XMM-LSS we find  $\sigma_{S_v} = 75 \mu\text{Jy}$  in the inner region and  $\sigma_{S_v} = 87 \mu\text{Jy}$  in the outer region. To ensure that the sources used to construct the HIMF are detected at least with  $5S_{\text{lim}}$  independent of their position in the pointing, we remove from the sample all sources with line flux that falls below this limit (based on the relevant field and  $\sigma_{S_v}$  for each source). After the line flux cut and exclusion of the sources detected outside the  $1.5 \text{ deg}^2$  area (Fig. 1) the sample decreases to 203 sources out of 276, which we use to construct the HIMF. By adopting this approach we are excluding regions of the survey where we are marginally incomplete to the lower H I masses, but also not probing as deeply as we potentially could close to the pointing centre. This helps us to mitigate not only redshift-dependent completeness uncertainties, but also completeness uncertainties associated with the position of a galaxy in the field. Fig. 2 shows the distribution of the H I mass of our sample before and after the flux cut, as well as median flux limit ( $5S_{\text{lim}}$ ). To check the robustness of our results with this method to mitigate the incompleteness in our sample, we also adopt a flux limit of  $8S_{\text{lim}}$ , resulting in a sample of 174 sources. We find that the results are consistent within the uncertainties (Table 2).

#### 4.2 $1/V_{\text{max}}$ method

The number density of galaxies as a function of their H I mass can be described as

$$\phi(M_{\text{HI}}) = \frac{dN_{\text{gal}}}{dV d \log_{10}(M_{\text{HI}})}, \quad (3)$$

where  $dN_{\text{gal}}$  is the average number of galaxies in the volume  $dV$ , with H I masses that fall within each logarithmic bin in  $M_{\text{HI}}$ .

To date, two different methods have been used in order to convert observed number count of galaxies as a function of their H I mass into the intrinsic number. One method, the two-dimensional stepwise maximum likelihood estimator (2DSWML; Zwaan et al. 2005; Martin et al. 2010), has been widely used for large surveys with high number counts of galaxies per H I mass bin, as it can incorporate independent methods to account for the effects of the large-scale structure (LSS) and sample variance.

However, if a survey does not have a large number of sources, maximum likelihood methods tend to produce large errors when only a few galaxies per bin are present (Zwaan et al. 1997; Busekool et al. 2021). Therefore, for our sample we elect to use the so-called  $1/V_{\max}$  method (Schmidt 1968). The principle of this method is to estimate the maximum comoving volume ( $V_{\max}$ ) that corresponds to the maximum redshift ( $z_{\max}$ ) at which a galaxy of a certain mass could be detected within a given survey.

To determine  $z_{\max}$  we use the limiting line flux of each galaxy ( $S_{\text{lim}}$ ) calculated in Section 4.1. Then, we iteratively evaluate the line flux ( $S$ ) of each detection over the redshift range of the entire sample (see Fig. 2) until it reaches the value below the detection threshold equal to  $5S_{\text{lim}}$ . The redshift at which this condition is met is assigned as  $z_{\max}$ . We then calculate  $V_{\max}$  for each detection using  $z_{\max}$  from the previous step. If  $V_{\max}$  exceeds the volume corresponding to the upper redshift boundary of the survey volume ( $z = 0.084$ ), then  $V_{\max}$  is set to  $V(z = 0.084)$ . The HIMF is later constructed by summing  $1/V_{\max}$  in logarithmic bins of H I mass. To further ensure 100 per cent completeness in our sample, we include an additional low-mass cut-off at  $M_{\text{HI}} = 3 \times 10^8 M_{\odot}$ . We adopt this cut-off as the most likely source of incompleteness in our sample are low-velocity/low-mass systems where the H I line width may only extend over 1–2 channels in our  $44 \text{ km s}^{-1}$  spectral resolution data. Although there are relatively few galaxies below this mass (Fig. 2), we err on the side of caution and only fit the HIMF to galaxies with H I mass above this limit. Furthermore, the exclusion of these low-mass galaxies does not adversely effect our ability to constrain the low-mass slope of the HIMF, as will be seen in Section 5.

### 4.3 Sample variance

Sample variance (sometimes referred to as cosmic variance in this context) is often used to describe the inhomogeneity of the Universe. In other words, the matter in the Universe is not distributed homogeneously and it contains regions of high and low density, which can introduce a systematic bias in observational estimates of the volume density of galaxies. For astronomical surveys that cover large enough volumes and sample all possible environments, sample variance averages out. However, it is a significant source of uncertainty for deep galaxy surveys, which tend to cover relatively small areas. In this case, the mass function tends to be biased by the specific volume and would not be representative of the universal mass function at a particular redshift (Somerville et al. 2004; Moster et al. 2011).

Calculating the sample variance is not a trivial task. For example, ALFALFA being a large area survey assumed the sample variance uncertainty as the difference between the HIMF of the Spring and Fall skies (Jones et al. 2018).

For MIGHTEE we evaluate the uncertainty due to the sample variance ( $\sigma_v$ ) following the prescriptions from the ‘Cosmic variance cookbook’ by Moster et al. (2011). This prescription uses  $\Lambda$ CDM predictions to estimate the clustering strength for a given number density at a known average redshift. According to this prescription the sample variance can be estimated by multiplying the dark matter

**Table 1.** The sample variance of the combined MIGHTEE Early Science fields, and for the individual COSMOS and XMM-LSS fields, for the full range of H I mass bins.

$\log(M_{\text{HI}})$	Fractional sample variance ( $\sigma_v$ )		
	MIGHTEE	COSMOS	XMM-LSS
6.5–8.5	0.20	0.36	0.25
8.5–9.5	0.22	0.38	0.27
9.5–10.0	0.24	0.41	0.29
10.0–10.5	0.25	0.45	0.33
10.5–11.0	0.27	0.47	0.34

cosmic variance ( $\sigma_{\text{dm}}$ ) at a given redshift by the linear galaxy bias ( $b$ ) at that redshift:

$$\sigma_v = b(M_{\star}, z)\sigma_{\text{dm}}(z, \Delta z = 0.1)\sqrt{0.1/\Delta z}, \quad (4)$$

where  $M_{\star}$  is the stellar mass range of the sample,  $z$  is the mean redshift of the survey, and  $\Delta z$  is the size of the redshift bin. The last term enables a sample variance calculation of different redshift bin sizes. Therefore, the sample variance depends on the stellar mass range within a given sample. Since we have measured stellar masses of our sample galaxies (Maddox et al. 2021; Pan et al. 2022), to evaluate  $\sigma_v$  we use the stellar mass bins of the MIGHTEE sample. Then we convert stellar mass bins into the equivalent H I mass bins following the  $M_{\text{HI}}-M_{\star}$  relation from Maddox et al. (2015, 2021), see also Pan et al. (2022).

The resulting values for the sample variance of the full MIGHTEE Early Science sample, as well as for the individual COSMOS and XMM-LSS fields, are shown in Table 1. As expected, there is a clear trend and the sample variance decreases with increasing survey area, being the largest for the COSMOS field, and the smallest for the full MIGHTEE sample. As a result, the sample variance introduces the averaged uncertainties of the volume densities of  $\sim 24$  per cent for combined fields,  $\sim 41$  per cent for the COSMOS field, and  $\sim 30$  per cent for the XMM-LSS field. These uncertainties are in agreement with Driver & Robotham (2010), who have shown that the survey should be at least 10 ‘ultra-deep’ fields for the effect of sample variance to be below 20 per cent for stellar mass selected samples. We note that H I-rich galaxies cluster differently than described in equation (4) due to the lack of H I in galaxies in very dense environments (Papastergis et al. 2013). Therefore, our sample variance constraints are conservative upper limit estimates. We add these uncertainties in quadrature to the Poisson errors for each H I mass bin prior to fitting the HIMF.

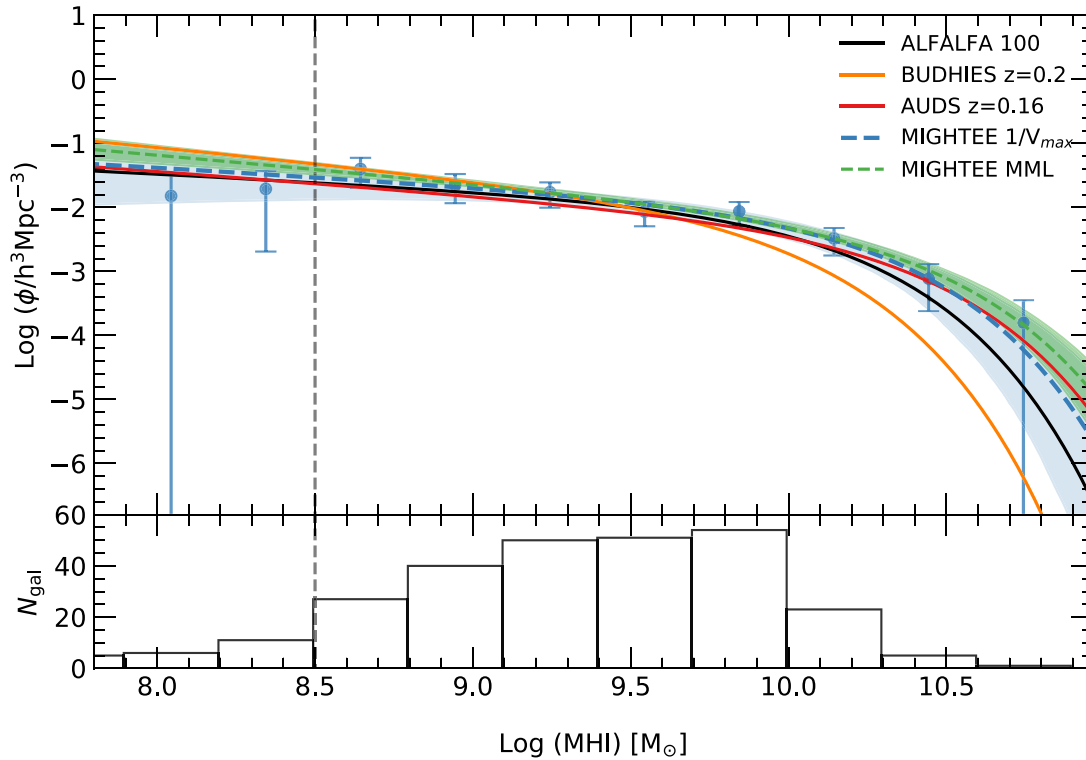
### 4.4 Fitting $1/V_{\max}$ data

It is widely accepted that a Schechter function can very well describe the shape of the HIMF (Zwaan et al. 1997):

$$\phi(M_{\text{HI}}) = \ln(10)\phi_{\star} \left(\frac{M_{\text{HI}}}{M_{\star}}\right)^{\alpha+1} e^{-\left(\frac{M_{\text{HI}}}{M_{\star}}\right)}, \quad (5)$$

where  $\phi_{\star}$  is the normalization constant,  $M_{\star}$  is the ‘knee’ mass, and  $\alpha$  is the low-mass slope.

In order to determine the best fit we perform a simple  $\chi^2$  minimization, which also enables us to determine the goodness of fit. However, to fully explore the posterior probability distribution of each of the parameters in the Schechter function, along with their degeneracies, we use MULTINEST (Feroz & Hobson 2008; Feroz, Hobson & Bridges 2009), based on the nested sampling technique (Skilling 2004). MULTINEST produces the posterior samples from distributions with an associated error estimate. We use default initial



**Figure 3.** The HIMF of the MIGHTEE Early Science data is shown with the blue points. The best-fitting relation based on the  $1/V_{\max}$  method is shown with the blue dashed line. The best-fitting relation based on the MML method is shown with the green dashed line. The ALFALFA 100 HIMF from Jones et al. (2018) is shown with the black line. The HIMF measured from BUDHIES (Gogate 2022) at  $z = 0.2$  is shown with the orange line, and the HIMF from AUFS (Xi et al. 2021) at  $z = 0.16$  is shown with the red line. The histogram in the bottom panel shows the distribution of H I mass in the MIGHTEE data. The vertical dashed line indicates the mass limit below which the data points were discarded prior to the fit (see Section 4.1). The  $1\sigma$  uncertainty of the  $1/V_{\max}$  fit, sampled from the MULTINEST posteriors (Fig. 4) is shown with the blue shaded area. The  $1\sigma$  uncertainty of the MML fit sampled from  $10^3$  bootstrap iterations is shown with the green shaded area.

parameters, such as tolerance = 0.5 and live points = 1000 (Buchner et al. 2014). The prior distributions for  $\phi_*$ ,  $M_*$ , and  $\alpha$  are set in the following ranges:  $\log_{10}(\phi_*)$ : uniform  $\in [-3, 0]$ ;  $\log_{10}(M_*)$ : uniform  $\in [9.5, 10.5]$ ; and  $\alpha$ : uniform  $\in [-2.25, -0.25]$ . We find that the maximum likelihood of the best-fitting HIMF agrees with the best fit determined by minimizing the  $\chi^2$  as expected, but we use the posteriors from MULTINEST to highlight the degeneracy between parameters.

#### 4.5 Modified maximum likelihood method

As already mentioned above, there is no definitive method to correct small samples of galaxies for the effects of the LSS when measuring the HIMF. While 2DSWML can mitigate these effects for the large surveys, it introduces large errors when there are only a few galaxies per bin of H I mass present (Section 4.2). However, to highlight degeneracies associated with fitting an HIMF we use a modified maximum likelihood (MML) method (Obreschkow et al. 2018) in addition to  $1/V_{\max}$ . MML method was specifically developed to infer generative distribution functions from uncertain and biased data. This method can accurately recover the mass function of galaxies, while simultaneously dealing with observational uncertainties and to some extent, unknown cosmic LSS. The main difference of this method from  $1/V_{\max}$  is that it is free of binning and it recovers the shape of the mass function by accounting for the individual  $1/V_{\max}$  for each galaxy, thus removing the need for binning the data. Moreover, it attempts to account for the effects of LSS (or sample variance)

by using the distance distribution of the data to model the mean density of the survey volume at comoving distance  $r$  relative to the mean density of the Universe (Baldry et al. 2012; Wright et al. 2017).

For our study we use the R-implementation of the MML (`dftools`) described in detail in Obreschkow et al. (2018). For the fit we provide our  $1/V_{\max}$  values with their associated uncertainties, as well as the distances to our galaxies. To determine the asymmetric uncertainties of the fit we use 1000 bootstrap iterations with a fixed seed for the random number generator (Obreschkow et al. 2018). The results of this method in comparison to  $1/V_{\max}$  are presented in Section 5, together with comparisons to the literature.

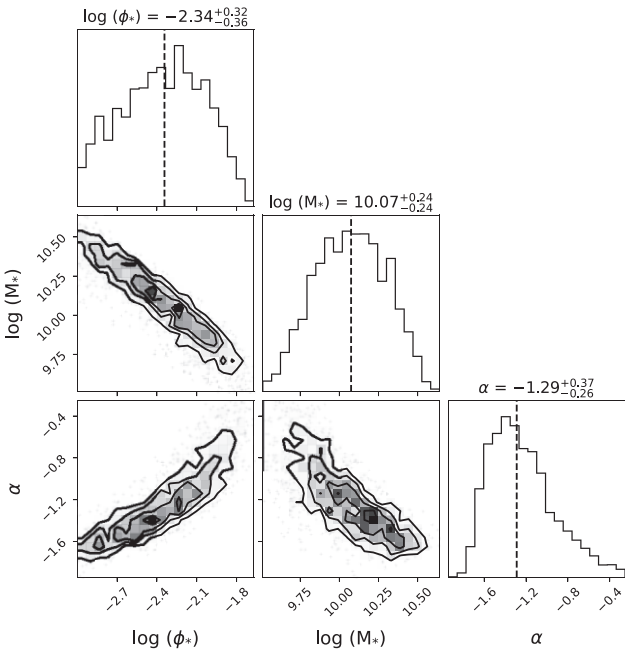
## 5 RESULTS

### 5.1 The MIGHTEE HIMF over $0 < z \leq 0.084$

Fig. 3 shows the HIMF measured using the MIGHTEE Early Science data together with the best-fitting Schechter function obtained using  $1/V_{\max}$  method (blue line) and MML method (green line), along with the H I mass distribution of the sample. The best-fitting parameters ( $\phi_*$ ,  $M_*$ , and  $\alpha$ ) for both measurements of the Schechter function parametrization for MIGHTEE and other surveys used for comparison are presented in Table 2. The posterior distributions for the Schechter function parameters obtained for  $1/V_{\max}$  method are shown in Fig. 4.

**Table 2.** The best-fitting parameters of a Schechter function parametrization of the HIMF and resulting  $\Omega_{\text{H1}}$  for MIGHTEE (this work), ALFALFA 100 (Jones et al. 2018), BUDHIES (Gogate 2022), and AUDES (Xi et al. 2021). The parameters of the three literature surveys have been scaled to  $H_0 = 67.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$  for the ease of comparison.

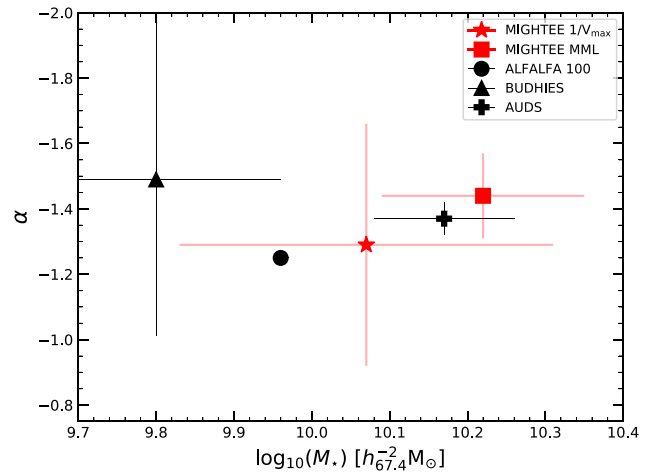
Survey (sample size)	$\log_{10}(\phi_*/h_{67.4}^3 \text{ Mpc}^{-3})$	$\log_{10}(M_*/h_{67.4}^{-2} M_\odot)$	$\alpha$	$\gamma$	$\chi^2_\nu$	$\Omega_{\text{H1}} \times 10^{-4} h_{67.4}^{-1}$
MIGHTEE $1/V_{\text{max}}^{5S_{\text{lim}}}$ (203)	$-2.34^{+0.32}_{-0.36}$	$10.07^{+0.24}_{-0.24}$	$-1.29^{+0.37}_{-0.26}$	–	0.98	$5.46^{+0.94}_{-0.99}$
MIGHTEE $1/V_{\text{max}}^{8S_{\text{lim}}}$ (174)	$-2.36^{+0.35}_{-0.38}$	$10.10^{+0.22}_{-0.24}$	$-1.40^{+0.42}_{-0.24}$	–	1.1	–
MIGHTEE MML (203)	$-2.52^{+0.19}_{-0.14}$	$10.22^{+0.10}_{-0.13}$	$-1.44^{+0.13}_{-0.10}$	–	1.2	$6.31^{+0.31}_{-0.31}$
MIGHTEE (evolutionary fit)	$-2.19^{+0.18}_{-0.26}$	$9.97^{+0.2}_{-0.18}$	$-0.96^{+0.29}_{-0.25}$	$1.15^{+1.75}_{-1.44}$	0.92	–
MIGHTEE COSMOS (53)	$-2.93^{+0.29}_{-0.27}$	$10.16^{+0.46}_{-0.33}$	$-1.58^{+0.49}_{-0.41}$	–	2.1	$2.84^{+1.33}_{-1.05}$
MIGHTEE XMM-LSS (150)	$-2.34^{+0.25}_{-0.31}$	$10.08^{+0.23}_{-0.23}$	$-1.13^{+0.40}_{-0.27}$	–	1.2	$4.77^{+0.87}_{-0.92}$
ALFALFA 100 (23621)	$-2.33(\pm 0.02 \pm 0.07)$	$9.96(\pm 0.01 \pm 0.005)$	$-1.25(\pm 0.02 \pm 0.1)$	–	–	$4.05 \pm 0.1$
BUDHIES (42)	$-2.30 \pm 0.03$	$9.80 \pm 0.16$	$-1.49 \pm 0.48$	–	–	$4.26 \pm 4.6$
AUDES (247)	$-2.60 \pm 0.01$	$10.17 \pm 0.09$	$-1.37 \pm 0.05$	–	–	$3.69 \pm 0.3$



**Figure 4.** The posterior distributions of the MIGHTEE HIMF parameters ( $\phi_*$ ,  $M_*$ , and  $\alpha$ ) obtained with MULTINEST. Black contours on the 2D histograms indicate  $1\sigma$ ,  $2\sigma$ , and  $3\sigma$  confidence levels. Dashed lines on histograms indicate the best-fitting values.

Overall, the HIMF is very well fit by the Schechter function in both cases, and the results of the two different methods are consistent within the uncertainties. The MML method presents much smaller uncertainties on the parameters since it accounts for the effects of the LSS with an implicit calculation based on the mean galaxy number density. In contrast, for the  $1/V_{\text{max}}$  method we account for cosmic variance in the error budget of the binned points. For both methods reduced  $\chi^2 \approx 1$ . The results from both methods are also consistent, within the uncertainties, with the results from ALFALFA 100, AUDES, and BUDHIES (Fig. 3).

When comparing the HIMFs from different surveys, samples, and redshifts, the most important comparisons arise from the characterization of the low-mass slope ( $\alpha$ ) and the ‘knee’ mass ( $M_*$ ) since they describe the overall shape of the HIMF. For example, using ALFALFA 100 Jones et al. (2018) found the low-mass slope to be



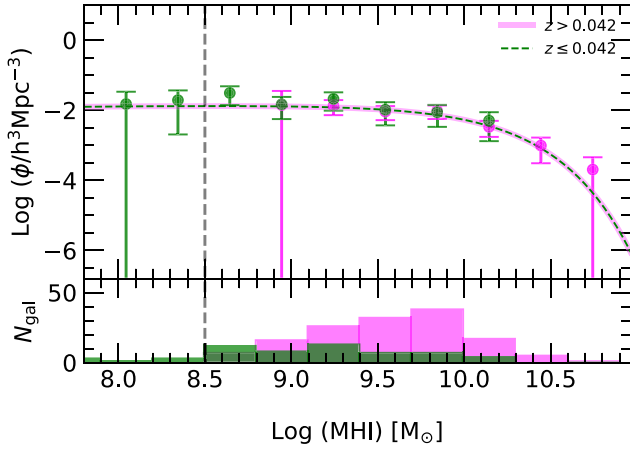
**Figure 5.**  $M_*$ - $\alpha$  measurements with associated uncertainties for different surveys. MIGHTEE measurements are shown with red symbols: star indicates the measurements obtained from  $1/V_{\text{max}}$  method and square is from using the MML method (see Section 4). The parameters of the three literature surveys are shown with the black symbols and have been scaled to  $H_0 = 67.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$  for the ease of comparison.

significantly flatter in the Fall sky than in the Spring sky due to the Virgo Cluster, suggesting that  $\alpha$  is sensitive to the environment. Using AUDES, Xi et al. (2021) found a very similar low-mass slope to ALFALFA 100, even though their sample covers a much larger redshift range ( $0 < z \leq 0.16$ , as opposed to  $0 < z \leq 0.06$  for ALFALFA 100) and a much smaller area.

Using the  $1/V_{\text{max}}$  method we find the best-fitting low-mass slope to be similar to both of these studies, and particularly almost identical to the slope measured by ALFALFA 100 ( $\alpha_{\text{A100}} = -1.25(\pm 0.02 \pm 0.1)$ ) and  $\alpha_{\text{MIGHTEE}} = -1.29^{+0.37}_{-0.26}$ . The MML yields a somewhat steeper slope that is consistent with the low-mass slope from BUDHIES, but mainly due to the fact that the low-mass slope from the BUDHIES data is relatively poorly constrained, due to having a mass limit of  $M_{\text{H1}} = 10^9 M_\odot$ . Fig. 5 shows the  $M_*$ - $\alpha$  comparison between our measurements and those from the literature.

We find the best fit for the ‘knee’ mass of the HIMF to be consistent, within the uncertainties, with previous studies and is in excellent agreement with AUDES, which probes twice the redshift range of MIGHTEE (Fig. 3). The ‘knee’ mass is responsible for the counts of





**Figure 6.** The HIMF constructed for the evolutionary fit (equation 6) is shown with the green ( $z = 0.021$ ) and magenta lines ( $z = 0.063$ ). The green points represent galaxies from the low-redshift bin ( $z \leq 0.04$ ), while magenta points are galaxies from the high-redshift bin ( $z > 0.04$ ). The lower panel shows the observed counts for ‘high’- $z$  sample in magenta and low- $z$  sample in green. The vertical dashed line indicates the completeness cut-off.

the high-mass galaxies, which require a large volume to be detected in sizeable numbers. Although, with the MIGHTEE Early Science sample we can constrain the ‘knee’ mass (Fig. 4), the uncertainties are still relatively large (around an order of magnitude larger than ALFALFA 100) due to the limited volume of the Early Science data.

Both our methods ( $1/V_{\max}$  and MML) yield consistent results (Table 2). Therefore, we proceed further by comparing only  $1/V_{\max}$  to the literature studies because it is more sensitive to the uncertainties, and we can investigate the goodness of the fit with the posterior distributions of the Schechter function parameters that highlight possible degeneracies (Fig. 4).

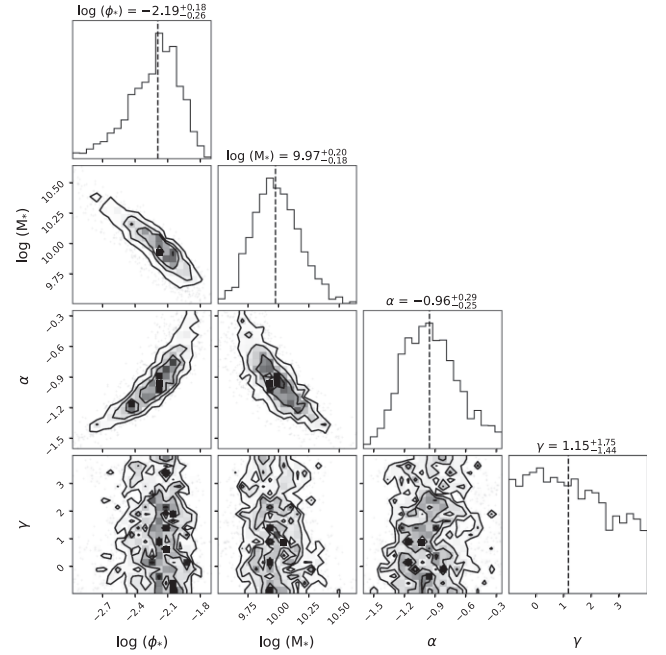
## 5.2 Evolution of the HIMF with redshift

As a proof of concept we investigate whether or not there is any evidence for evolution in the HIMF as a function of redshift. To do so we divide our main sample into two: a low-redshift sample ( $z \leq 0.04$ ) and ‘high’-redshift sample ( $z > 0.04$ ). For both samples we recalculate  $V_{\max}$  using only the volume corresponding to the new redshift range. We also adjust the Poisson counting errors to the new samples.

To quantitatively assess the possible evolution of the HIMF, instead of fitting the two samples separately, we rather fit them simultaneously with the modified Schechter function that includes a  $(1+z)^\gamma$  density evolution term (Pan et al. 2020):

$$\phi(M_{\text{HI}}) = \ln(10) \phi_* \left( \frac{M_{\text{HI}}}{M_*} \right)^{\alpha+1} e^{-\left( \frac{M_{\text{HI}}}{M_*} \right)} (1+z)^\gamma, \quad (6)$$

where all the parameters are the same as in equation (5),  $z$  is the mean redshift of each sample ( $z = 0.021$  for the low- $z$  sample and  $z = 0.063$  for the ‘high’- $z$  sample), and  $\gamma$  is an evolutionary parameter describing how the HIMF evolves with redshift. We note that the evolutionary term  $(1+z)^\gamma$  can be appended to  $M_*$  (resulting in a horizontal shift in the HIMF) instead of to the overall density (vertical shift). Given our sample size and where the bulk of our galaxies reside, we have a higher chance of detecting the overall density evolution due to the fact that we have better statistics on the low-mass slope than the high-mass turnover. However, we test this by



**Figure 7.** The posterior distributions of the fitted evolutionary modified Schechter function parameters ( $\phi_*$ ,  $M_*$ ,  $\alpha$ , and  $\gamma$ ) obtained with MULTINEST. The contours and vertical dashed lines are the same as in Fig. 4.

adopting a characteristic mass evolution and find similar results, consistent with zero evolution.

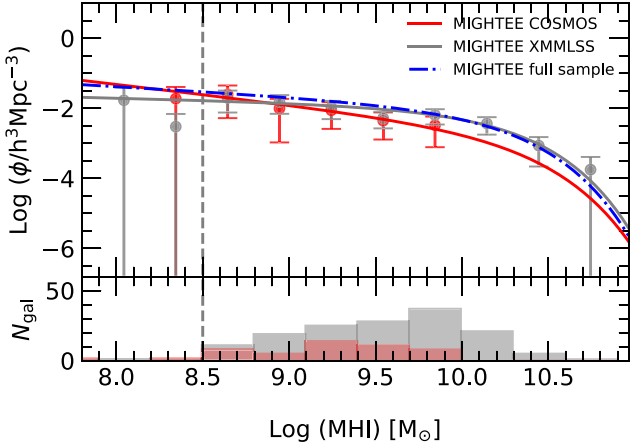
Fig. 6 shows the resulting HIMF for both the low- and high- $z$  samples. From Fig. 6 it is already clear that, we do not detect any significant evolution of the HIMF over the redshift range of our sample. The posterior distributions of fitted parameters are shown in Fig. 7. While all other parameters of the HIMF are well constrained and consistent with the fit of the entire sample (Table 2),  $\gamma$  is not constrained and has very large associated uncertainties (Fig. 7). We calculate  $\chi^2_v = 0.92$  using the parameters of the model including  $\gamma$  compared to  $\chi^2_v = 0.98$  when the evolutionary term is not considered. Therefore, invoking Occam’s Razor, within our redshift range we do not find any evidence for evolution of the HIMF. This result is in agreement with various models of galaxy formation and evolution that predict that the major evolution of H I content of the Universe has occurred between  $z = 2$  and  $z = 0$  (Yates, Péroux & Nelson 2021), and 1 billion yr in lookback time is not enough to be able to detect any evolution of the HIMF.

## 5.3 HIMF over different fields

Next, we investigate the variations for the HIMF over two distinct fields, which combined make up our main sample, COSMOS ( $1.5 \text{ deg}^2$ ) and XMM-LSS ( $3.3 \text{ deg}^2$ ). We construct the HIMF for each field separately and adjust the uncertainties due to cosmic variance and Poisson source counts of each sample. The HIMFs for two different fields, together with the HIMF for the main sample, are shown in Fig. 8.

We find that while the HIMF of both fields is well constrained at the low-mass range, due to the small volume probed by the COSMOS data, the ‘knee’ mass of the COSMOS HIMF is poorly constrained and suppressed, in comparison to the ‘knee’ mass of the XMM-LSS sample. This results in a steeper low-mass slope than when the whole sample is considered (Table 2). Moreover, the COSMOS HIMF also





**Figure 8.** The HIMF of the COSMOS field is shown with red points and its best-fitting relation is shown with the red line. The *XMM*-LSS HIMF is shown with grey points and grey line. The total MIGHTEE parametrized HIMF is shown with the blue line. The lower panel shows the observed counts for COSMOS in red and for *XMM*-LSS in grey. The vertical dashed line indicates the completeness cut-off.

has much larger uncertainties due to the significantly smaller volume than for the *XMM*-LSS field, which results in lower number counts in the same H I mass bin coupled with a higher sample variance (Table 1).

The ‘knee’ mass of the full sample is completely dominated by the *XMM*-LSS field due to the much larger volume coverage, and a relatively few high-H I-mass galaxies in our whole sample have been detected only in the *XMM*-LSS field. Therefore, we do not find any difference between the *XMM*-LSS ‘knee’ mass and the one for the whole sample (Fig. 8).

Because of our limited volume we cannot conclude if the low-mass slope of the HIMF ( $\alpha$ ) is sensitive to the environment, as was found by Jones et al. (2020). Given a relatively small sample, we find the measured values of  $\alpha$  to have large uncertainties, and therefore the low-mass slopes from the two different fields are consistent with the low-mass slope of the entire sample within errors. Interestingly,  $\alpha$  of the main sample sits in between the steeper low-mass slope of the COSMOS field, dominated by low-mass galaxies, and the shallower low-mass slope of the *XMM*-LSS field, dominated by the higher H I mass galaxies. From Figs 4 and 7 it is clear that the Schechter function parametrization  $\alpha$  and  $M_*$  are highly degenerate, and it is only by using the larger volume but slightly shallower *XMM*-LSS data in conjunction with the deeper and narrower COSMOS data, that we can overcome this degeneracy and reduce the uncertainty on each individual parameter.

#### 5.4 Cosmic H I density ( $\Omega_{\text{H I}}$ )

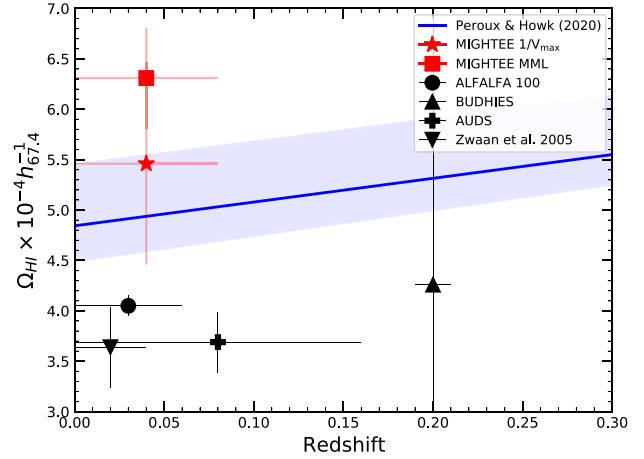
The HIMF can be used to calculate the cosmic H I density ( $\Omega_{\text{H I}}$ ) by integrating over the best-fitting Schechter function.

First, the comoving H I density ( $\rho_{\text{H I}}$ ) is defined as

$$\rho_{\text{H I}} = \Gamma(\alpha + 2) \phi_* M_* \quad (7)$$

where  $\Gamma$  is Euler Gamma function and  $\phi_*$ ,  $M_*$ , and  $\alpha$  are the HIMF parameters of the best-fitting Schechter function. Subsequently  $\Omega_{\text{H I}}$  is calculated as

$$\Omega_{\text{H I}} = \frac{8\pi G}{3H_0^2} \rho_{\text{H I}}, \quad (8)$$



**Figure 9.** Neutral gas density ( $\Omega_{\text{H I}}$ ) as a function of redshift for different H I surveys (emission only). Horizontal error bars indicate the redshift coverage of each survey. The best-fitting cosmic H I density from a compilation of H I emission (direct and stacking) and Ly $\alpha$  absorption from  $z = 0$  to  $z = 5$  is shown with the blue line. The blue shaded region indicates 95 per cent confidence interval. Adopted from Péroux & Howk (2020).

where  $G$  is the gravitational constant and  $H_0$  is the Hubble constant. Using the parameters of the HIMF for our main MIGHTEE sample we obtain  $\Omega_{\text{H I}} = 5.46^{+0.94}_{-0.99} \times 10^{-4}$  for the  $1/V_{\text{max}}$  method and  $\Omega_{\text{H I}} = 6.31^{+0.31}_{-0.31} \times 10^{-4}$  using MML method. We remind the readers that in our study we use  $H_0 = 67.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$  and therefore the direct comparison with other surveys that use  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$  would not be accurate. If we scale the values of  $\Omega_{\text{H I}}$  found by the other surveys to  $H_0 = 67.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , we find that our measurements are consistent within  $2\sigma$  with  $\Omega_{\text{H I}} = 4.05 \pm 0.1 \times 10^{-4}$  found by ALFALFA 100 (Jones et al. 2018). Fig. 9 shows  $\Omega_{\text{H I}}$  values measured using direct detections from various H I surveys scaled to  $H_0 = 67.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$  for the ease of comparison. We also show the best-fitting relation for  $\Omega_{\text{H I}}$  as a function of  $z$ , obtained by Péroux & Howk (2020) by fitting the compilation of various  $\Omega_{\text{H I}}$  measurements from  $z = 0$  to  $z = 5$ , zoomed in to the relevant redshift range to demonstrate where local measurements lie with respect to the global trend. Although we find a slightly higher value of  $\Omega_{\text{H I}}$  using  $1/V_{\text{max}}$ , it is consistent with previous measurements within the large uncertainties, which are dominated by the sample variance, and also with predictions from hydrodynamical cosmological simulations (Villaescusa-Navarro et al. 2018; Diemer et al. 2019) and semi-analytic models (Popping et al. 2019). For the MML method we find a steeper  $\alpha$ , higher ‘knee’ mass, and similar normalization, therefore the  $\Omega_{\text{H I}}$  from MML is larger than when  $1/V_{\text{max}}$  is used, even though consistent within uncertainties (Fig. 9).

To investigate how different samples affect the measured  $\Omega_{\text{H I}}$ , we calculate it separately for the two different fields. In Section 5.3, we found that while COSMOS has a steeper low-mass slope, the *XMM*-LSS field defines the ‘knee’ mass of the entire sample. We find  $\Omega_{\text{H I}}(\text{COSMOS}) = 2.84^{+1.33}_{-1.05} \times 10^{-4}$  and  $\Omega_{\text{H I}}(\text{XMM-LSS}) = 4.77^{+0.87}_{-0.92} \times 10^{-4}$ . These are both lower than what we measure for the whole sample due to highly suppressed  $M_*$  in COSMOS and flatter  $\alpha$  in *XMM*-LSS.

This highlights the importance of fully sampling all parts of the mass function, high-mass galaxies (and therefore large enough volume) are needed to constrain the ‘knee’ of the mass function, whereas the lower mass galaxies are needed to decouple the low-mass slope from variations in  $\phi_*$ .

## 6 SUMMARY AND CONCLUSIONS

In this paper, we present the first HIMF determined from data using the MeerKAT telescope and the first using interferometry data over a non-targeted overdensity. We use the Early Science data from the MIGHTEE survey and construct the HIMF over the last billion years ( $0 \leq z \leq 0.084$ ). We investigate the properties of the HIMF in different fields, as well the possible evolution with redshift. Our main results can be summarized as follows.

(i) Visual source finding is still widely used to assess the reliability of the detected sources. However, moving forward it is not sustainable for when the modern surveys will reach their full capacity.

(ii) We use two different methods to measure the parameters of the HIMF ( $1/V_{\max}$  and MML) and we find that the first MeerKAT HIMF is in excellent agreement with previous single-dish and interferometric studies, even though all these studies probe different area, volume, environment, and redshift range. Using  $1/V_{\max}$  we find an identical low-mass slope  $\alpha = -1.29^{+0.37}_{-0.26}$  in comparison to findings by ALFALFA 100 ( $\alpha = -1.25 \pm 0.02 \pm 0.01$ ) and ‘knee’ mass ( $\log_{10}(M_*) = 10.07 \pm 0.24$ ) that is consistent within the uncertainties when comparing to the ALFALFA 100 survey ( $\log_{10}(M_*) = 9.96 \pm 0.01 \pm 0.005$ ), as well as when comparing to the higher redshift ( $z = 0.16$ ) AUDS ( $\log(M_*) = 10.17 \pm 0.09$ ). We note that we scale the parameters of the literature values to  $H_0 = 67.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$  for the ease of comparison.

(iii) As a proof of concept we investigate whether or not there is any evidence for evolution in the HIMF as a function of redshift. As expected from cosmological models of galaxy formation and evolution, we find no evidence for evolution of the HIMF over the last billion years. This result is also consistent with the studies of the HIMF at higher redshift range – neither AUDS ( $z = 0.16$ ) nor BUDHIES ( $z = 0.2$ ) has found any evidence for the evolution of the HIMF. However, the evolution in the shape of the HIMF (suppressed ‘knee’ mass and steeper low-mass slope) is expected at  $z \sim 0.35$  according to findings by Bera et al. (2022). This will be tested with the full capacity of the MIGHTEE survey.

(iv) We investigate the properties of the HIMF in two distinct fields (COSMOS and XMM-LSS). We find  $\alpha$  to be steeper in COSMOS due to the suppressed ‘knee’ mass. We find that  $M_*$  is highly sensitive to the volume observed by a survey, as it requires a sampling of galaxies beyond the ‘knee’, and it is poorly constrained for COSMOS that covers a smaller volume than XMM-LSS.

(v) We find the cosmic H I density  $\Omega_{\text{HI}} = 5.46^{+0.94}_{-0.99} \times 10^{-4}$  to be slightly higher than reported by previous studies, though consistent within uncertainties. We find  $\Omega_{\text{HI}}(\text{COSMOS}) = 2.84^{+1.33}_{-1.05} \times 10^{-4}$  and  $\Omega_{\text{HI}}(\text{XMM-LSS}) = 4.77^{+0.87}_{-0.92} \times 10^{-4}$ , which highlight the importance of fully sampling all parts of the mass function, large enough volume is needed to constrain the ‘knee’ of the mass function, whereas the depth is needed to constrain the low-mass slope.

New observational facilities such as MeerKAT have the potential to transform our knowledge of H I in the Universe way before the SKA era. Even with the Early Science MIGHTEE data we were able to measure the HIMF and estimate  $\Omega_{\text{HI}}$  over the period of the last billion years. At the same time, the MIGHTEE Large Survey Program is well underway, and will give us an opportunity to extend the current study up to  $z = 0.5$ . With the wealth of excellent ancillary data, including large 4-m Multi-Object Spectrograph Telescope (4MOST) spectroscopic survey (Driver et al. 2019), the MIGHTEE data will be crucial for our understanding of the evolution of the HIMF and  $\Omega_{\text{HI}}$ , especially using the combination of direct detections and new Bayesian stacking techniques (Pan et al. 2020).

## ACKNOWLEDGEMENTS

AAP, MJJ, and IH acknowledge support of the STFC consolidated grants [ST/S000488/1] and [ST/W000903/1] and MJJ and IH support from a UKRI Frontiers Research Grant [EP/X026639/1]. SHAR is supported by the South African Research Chairs Initiative of The Department of Science and Technology and National Research Foundation. MJJ, HP, and IH acknowledge support from the South African Radio Astronomy Observatory (SARAO) that is a facility of the National Research Foundation (NRF), an agency of the Department of Science and Innovation, South Africa. AAP and MJJ acknowledge support from the Oxford Hintze Centre for Astrophysical Surveys that is funded through generous support from the Hintze Family Charitable Foundation. NM acknowledges support of the LMU Faculty of Physics. IP acknowledges support from INAF under the Large Grant 2022 funding scheme (project ‘MeerKAT and LOFAR Team up: a Unique Radio Window on Galaxy/AGN co-Evolution’). SK is supported by the South African Research Chairs Initiative of The Department of Science and Technology and National Research Foundation. KMH acknowledges financial support from the grant CEX2021-001131-S funded by MCIN/AEI/10.13039/501100011033, from the coordination of the participation in SKA-SPAIN, funded by the Ministry of Science and Innovation (MCIN) and from grant PID2021-123930OB-C21 funded by MCIN/AEI/10.13039/501100011033, by ‘ERDF A way of making Europe’ and by the ‘European Union’. The MeerKAT telescope is operated by the South African Radio Astronomy Observatory, which is a facility of the National Research Foundation, an agency of the Department of Science and Innovation. We acknowledge use of the Inter-University Institute for Data Intensive Astronomy (IDIA) data intensive research cloud for data processing. IDIA is a South African university partnership involving the University of Cape Town, the University of Pretoria, and the University of the Western Cape. The authors acknowledge the Centre for High Performance Computing (CHPC), South Africa, for providing computational resources to this research project. This work is based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under ESO programme ID 179.A-2005 (Ultra-VISTA) and ID 179.A-2006 (VIDEO) and on data products produced by CALET and the Cambridge Astronomy Survey Unit on behalf of the Ultra-VISTA and VIDEO consortia. Based on observations collected at the European Southern Observatory under ESO programmes 179.A-2005 (UltraVISTA), and 179.A-2006 (VIDEO). Based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/IRFU, at the Canada–France–Hawaii Telescope (CFHT) that is operated by the National Research Council (NRC) of Canada, the Institut National des Science de l’Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This work is based in part on data products produced at Terapix available at the Canadian Astronomy Data Centre as part of the Canada–France–Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS. The Hyper SuprimeCam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS),

Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University.

This research has made use of NASA's Astrophysics Data System Bibliographic Services. This research made use of ASTROPY<sup>1</sup> a community-developed core PYTHON package for astronomy.

## DATA AVAILABILITY

The MIGHTEE Early Science spectral cubes are available by request to the corresponding author. The data catalogue that was used in this paper can be found as supplementary material (<https://shorturl.at/bglpj>). It includes ID, coordinates, observed frequency, total HI flux density, and its associated error (see Section 3 for details) for every detected source.

## REFERENCES

- Adams E. A. K. et al., 2022, *A&A*, 667, A38
- Aditya J. N. H. S., Jorgenson R., Joshi V., Singh V., An T., Chandola Y., 2021, *MNRAS*, 500, 998
- Aihara H. et al., 2019, *PASJ*, 71, 114
- Baldry I. K. et al., 2012, *MNRAS*, 421, 621
- Bera A., Kanekar N., Weiner B. J., Sethi S., Dwarakanath K. S., 2018, *ApJ*, 865, 39
- Bera A., Kanekar N., Chengalur J. N., Bagla J. S., 2022, *ApJ*, 940, L10
- Blyth S. et al., 2018, Proc. Sci., LADUMA: Looking at the Distant Universe with the MeerKAT Array. SISSA, Trieste, PoS(MeerKAT2016)004
- Buchner J. et al., 2014, *A&A*, 564, A125
- Busekool E., Verheijen M. A. W., van der Hulst J. M., Tully R. B., Trentham N., Zwaan M. A., 2021, *MNRAS*, 501, 2608
- Chowdhury A., Kanekar N., Chengalur J. N., Sethi S., Dwarakanath K. S., 2020, *Nature*, 586, 369
- Comrie A. et al., 2020, CARTA: The Cube Analysis and Rendering Tool for Astronomy (1.4.0). Zenodo. Available at: <https://doi.org/10.5281/zenodo.4034416>
- Crain R. A. et al., 2017, *MNRAS*, 464, 4204
- Darling J., Truebenbach A. E., 2018, *ApJ*, 864, 37
- Davé R., Crain R. A., Stevens A. R. H., Narayanan D., Saintonge A., Catinella B., Cortese L., 2020, *MNRAS*, 497, 146
- Delhaize J., Meyer M. J., Staveley-Smith L., Boyle B. J., 2013, *MNRAS*, 433, 1398
- Diemer B. et al., 2018, *ApJS*, 238, 33
- Diemer B. et al., 2019, *MNRAS*, 487, 1529
- Dodson R. et al., 2022, *AJ*, 163, 59
- Driver S. P., Robotham A. S. G., 2010, *MNRAS*, 407, 2131
- Driver S. P. et al., 2019, *The Messenger*, 175, 46
- Feroz F., Hobson M. P., 2008, *MNRAS*, 384, 449
- Feroz F., Hobson M. P., Bridges M., 2009, *MNRAS*, 398, 1601
- Gogate A., 2022, PhD thesis, Univ. Groningen, Groningen, the Netherlands
- Grasha K., Darling J., Leroy A. K., Bolatto A. D., 2020, *MNRAS*, 498, 883
- Guo H., Jones M. G., Haynes M. P., Fu J., 2020, *ApJ*, 894, 92
- Gupta N., Salter C. J., Saikia D. J., Ghosh T., Jeyakumar S., 2006, *MNRAS*, 373, 972
- Haynes M. P. et al., 2011, *AJ*, 142, 170
- Healy J., Deb T., Verheijen M. A. W., Blyth S. L., Serra P., Ramatsoku M., Vulcani B., 2021, *A&A*, 654, A173
- Heintz K. E., Watson D., Oesch P. A., Narayanan D., Madden S. C., 2021, *ApJ*, 922, 147
- Heintz K. E. et al., 2022, *ApJ*, 934, L27
- Hess K. M. et al., 2019, *MNRAS*, 484, 2234
- Jarvis M. J. et al., 2013, *MNRAS*, 428, 1281
- Jarvis M. et al., 2018, Proc. Sci., The MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) Survey. SISSA, Trieste, PoS(MeerKAT2016)006
- Johnston S. et al., 2008, *Exp. Astron.*, 22, 151
- Jonas J. L., 2009, *IEEE Proc.*, 97, 1522
- 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
- Jones M. G., Hess K. M., Adams E. A. K., Verdes-Montenegro L., 2020, *MNRAS*, 494, 2090
- Maccagni F. M., Morganti R., Oosterloo T. A., Geréb K., Maddox N., 2017, *A&A*, 604, A43
- McCracken H. J. et al., 2012, *A&A*, 544, A156
- Madau P., Dickinson M., 2014, *ARA&A*, 52, 415
- Maddox N., Hess K. M., Obreschkow D., Jarvis M. J., Blyth S. L., 2015, *MNRAS*, 447, 1610
- Maddox N., Jarvis M. J., Oosterloo T. A., 2016, *MNRAS*, 460, 3419
- Maddox N. et al., 2021, *A&A*, 646, A35
- Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, *ApJ*, 723, 1359
- Meyer M., 2010, Proc. Sci., PoS(PRA2009)015 Exploring the HI Universe with ASKAP. SISSA, Trieste, PoS(PRA2009)015
- Meyer M., Robotham A., Obreschkow D., Westmeier T., Duffy A. R., Staveley-Smith L., 2017, *Publ. Astron. Soc. Aust.*, 34, e052
- Moster B. P., Somerville R. S., Newman J. A., Rix H.-W., 2011, *ApJ*, 731, 113
- Neeleman M., Prochaska J. X., Ribaldo J., Lehner N., Howk J. C., Rafelski M., Kanekar N., 2016, *ApJ*, 818, 113
- Noterdaeme P. et al., 2012, *A&A*, 547, L1
- Obreschkow D., Murray S. G., Robotham A. S. G., Westmeier T., 2018, *MNRAS*, 474, 5500
- Pan H., Jarvis M. J., Allison J. R., Heywood I., Santos M. G., Maddox N., Frank B. S., Kang X., 2020, *MNRAS*, 491, 1227
- Pan H. et al., 2022, preprint ([arXiv:2210.04651](https://arxiv.org/abs/2210.04651))
- Papastergis E., Giovanelli R., Haynes M. P., Rodríguez-Puebla A., Jones M. G., 2013, *ApJ*, 776, 43
- Péroux C., Howk J. C., 2020, *ARA&A*, 58, 363
- Péroux C., McMahon R. G., Storrie-Lombardi L. J., Irwin M. J., 2003, *MNRAS*, 346, 1103
- Pisano D. J., Barnes D. G., Staveley-Smith L., Gibson B. K., Kilborn V. A., Freeman K. C., 2011, *ApJS*, 197, 28
- Planck Collaboration VI, 2020, *A&A*, 641, A6
- Ponomareva A. A. et al., 2021, *MNRAS*, 508, 1195
- Popping G. et al., 2019, *ApJ*, 882, 137
- Rajohnson S. H. A. et al., 2022, *MNRAS*, 512, 2697
- Ramatsoku M. et al., 2016, *MNRAS*, 460, 923
- 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
- Rhee J. et al., 2023, *MNRAS*, 518, 4646
- Said K., Kraan-Korteweg R. C., Staveley-Smith L., 2019, *MNRAS*, 486, 1796
- Schmidt M., 1968, *ApJ*, 151, 393
- Skilling J., 2004, in Fischer R., Preuss R., von Toussaint U., eds, AIP Conf. Proc. Vol. 735, Bayesian Inference and Maximum Entropy Methods in Science and Engineering. Am. Inst. Phys., Melville, NY, p. 395
- Somerville R. S., Lee K., Ferguson H. C., Gardner J. P., Moustakas L. A., Gialavisco M., 2004, *ApJ*, 600, L171
- Tully R. B., Courtois H., Hoffman Y., Pomarède D., 2014, *Nature*, 513, 71
- Villaescusa-Navarro F. et al., 2018, *ApJ*, 866, 135
- Walter F. et al., 2020, *ApJ*, 902, 111
- Westmeier T. et al., 2017, *MNRAS*, 472, 4832
- Westmeier T. et al., 2021, *MNRAS*, 506, 3962
- Wright A. H. et al., 2017, *MNRAS*, 470, 283
- Xi H., Staveley-Smith L., For B.-Q., Freudling W., Zwaan M., Hoppmann L., Liang F.-H., Peng B., 2021, *MNRAS*, 501, 4550
- Yasin T., Desmond H., Devriendt J., Slyz A., 2022, preprint ([arXiv:2206.15443](https://arxiv.org/abs/2206.15443))

<sup>1</sup><http://www.astropy.org>

Yates R. M., Péroux C., Nelson D., 2021, *MNRAS*, 508, 3535  
 Zwaan M. A., Briggs F. H., Sprayberry D., Sorar E., 1997, *ApJ*, 490, 173  
 Zwaan M. A. et al., 2003, *AJ*, 125, 2842  
 Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, *MNRAS*, 359, L30

## SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

### suppl\_data

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

<sup>1</sup>*Oxford Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK*

<sup>2</sup>*Department of Physics and Astronomy, University of the Western Cape, Robert Sobukwe Road, Bellville 7535, South Africa*

<sup>3</sup>*School of Physics, H.H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, UK*

<sup>4</sup>*Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, D-81679 Munich, Germany*

<sup>5</sup>*Steward Observatory, University of Arizona, 933 North Cherry Avenue, Rm. N204, Tucson, AZ 85721-0065, USA*

<sup>6</sup>*Department of Astronomy, University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa*

<sup>7</sup>*South African Radio Astronomy Observatory, 2 Fir Street, Observatory 7925, South Africa*

<sup>8</sup>*The Inter-University Institute for Data Intensive Astronomy (IDIA), and University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa*

<sup>9</sup>*Anton Pannekoek Institute for Astronomy, University of Amsterdam, NL-1098 XH Amsterdam, the Netherlands*

<sup>10</sup>*International Centre for Radio Astronomy Research (ICRAR), The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia*

<sup>11</sup>*ASTRON, the Netherlands Institute for Radio Astronomy, Oude Hoogeveesedijk 4, NL-7991 PD Dwingeloo, the Netherlands*

<sup>12</sup>*Kapteyn Astronomical Institute, PO Box 800, NL-9700 AV Groningen, the Netherlands*

<sup>13</sup>*Sterrenkundig Observatorium, Universiteit Gent, Krijgslaan 281 S9, B-9000 Gent, Belgium*

<sup>14</sup>*Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía s/n, E-18008 Granada, Spain*

<sup>15</sup>*INAF-IRA, Via P. Gobetti 101, I-40129 Bologna, Italy*

<sup>16</sup>*Department of Physics and Astronomy, Università degli Studi di Padova, Vicolo dell'Osservatorio 3, I-35122 Padova, Italy*

<sup>17</sup>*INAF – Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, I-35122 Padova, Italy*

<sup>18</sup>*Department of Physics and Space Science, Royal Military College of Canada, PO Box 17000, Station Forces, Kingston K7K 7B4, Canada*

<sup>19</sup>*Department of Physics and Electronics, Rhodes University, PO Box 94, Makhanda 6140, South Africa*

<sup>20</sup>*School of Science, Western Sydney University, Locked Bag 1797, Penrith, NSW 2751, Australia*

<sup>21</sup>*CSIRO Astronomy and Space Science, PO Box 1130, Bentley, WA 6102, Australia*

<sup>22</sup>*National Radio Astronomy Observatory, 1003 Lopezville Road, Socorro, NM 87801, USA*

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