

# Predictions for the FAST telescope's CRAFTS extragalactic H I survey

Kai Zhang<sup>1b</sup>,<sup>1,2★</sup> Jingwen Wu,<sup>1,2★</sup> Di Li<sup>1b</sup>,<sup>1,2,3★</sup> Chao-Wei Tsai,<sup>1,2</sup> Lister Staveley-Smith<sup>1b</sup>,<sup>4,5</sup> Jing Wang,<sup>6</sup> Jian Fu,<sup>7</sup> Travis McIntyre<sup>1</sup> and Mao Yuan<sup>1,2</sup>

(FAST collaboration)

<sup>1</sup>National Astronomical Observatories, Chinese Academy of Sciences, Datun Road, Chaoyang District, Beijing 100101, China

<sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup>NAOC-UKZN Computational Astrophysics Centre, University of KwaZulu-Natal, Durban 4000, South Africa

<sup>4</sup>International Centre for Radio Astronomy Research (ICRAR), The University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia

<sup>5</sup>ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia

<sup>6</sup>Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, China

<sup>7</sup>Key Laboratory for Research in Galaxies and Cosmology, Shanghai Astronomical Observatory, CAS, 80 Nandan Road., Shanghai 200030, China

Accepted 2020 October 16. Received 2020 October 16; in original form 2020 July 20

## ABSTRACT

The Five-hundred-meter Aperture Spherical radio Telescope (FAST) has started the Commensal Radio Astronomy FasT Survey (CRAFTS). In this paper, we use the technical parameters of FAST derived from commissioning observations to simulate the completeness function for extragalactic H I survey of CRAFTS, H I galaxies from two kinds of mock catalogues are selected. One is generated by Monte Carlo simulation based on the interpolated mass–velocity width function of the ALFALFA 100 per cent (a.k.a.  $\alpha$ .100) catalogue. The other is constructed by semi-analytical  $N$ -body simulation based on the  $\Lambda$ CDM model. Our results suggest that a two-pass CRAFTS extragalactic H I survey will be able to detect nearly  $4.8 \times 10^5$  galaxies, from which the ‘faint end’ slope of the H I mass function (HIMF) can be recovered to  $10^7 M_\odot$  and the ‘knee mass’ of the HIMF can be measured to a redshift of 0.1. Considering the radio frequency interference status and sensitivity limitation, CRAFTS will be efficient in detecting H I galaxies at redshifts below 0.1, which implies a tremendous potential in exploring the galaxy interactions in different environments and the spatial distribution of H I galaxies in the local universe.

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

## 1 INTRODUCTION

The current standard  $\Lambda$  cold dark matter ( $\Lambda$ CDM) cosmological model has successfully simulated the formation history of galaxies and the cosmic structure on large scales, but there are still several controversial discrepancies between simulations and observations at small scales (Weinberg et al. 2015). One puzzle is that the observed low-mass galaxies (satellites) are much rarer than the number of subhaloes predicted by the model, which is usually referred to as the ‘missing satellites’ problem (Klypin et al. 1999; Moore et al. 1999). The ‘missing satellites’ problem can also be reflected in the discrepancy of the derived dark matter halo mass function between simulations and observations. In the  $\Lambda$ CDM model (Press & Schechter 1974), dark matter halo mass function can be well fitted by Schechter function, with a low-mass end or ‘faint end’ of  $\alpha \sim -1.9$ . However, from existing galaxy surveys, the ‘faint end’ of the optical luminosity function (LM) and the H I mass function (HIMF) are much flatter, typically between  $-1.0$  and  $-1.4$  (e.g. Blanton et al. 2005; Zwaan et al. 2005; Montero-Dorta & Prada 2009; Hill et al. 2010; Martin et al. 2010; Jones et al. 2018). The

overpredicted number of satellites might be alleviated by adding baryonic processes like photoionization and stellar feedback (e.g. Koposov et al. 2009; Macciò et al. 2010) and by considering tidal interactions with the Galactic disc (e.g. Garrison-Kimmel et al. 2017; Nadler et al. 2018; Kelley et al. 2019) in simulations. The census of dwarf galaxies has also improved due to the enhancement of survey sensitivity and sky coverage (e.g. Tollerud et al. 2008; Drlica-Wagner et al. 2020). However, this mismatch has not been fully eliminated. The H I observations provide an alternate approach to identify dwarf galaxies because of their higher neutral gas fraction (Schombert, McGaugh & Eder 2001; Geha et al. 2006; Haynes 2019).

Despite the ‘missing satellites’ problem, recent observations of the most luminous satellites around the Milky Way indicate that the dark matter subhaloes predicted from the  $\Lambda$ CDM simulation are more massive than the results inferred from stellar kinematics in the satellites, which is referred to as the ‘Too Big To Fail’ (TBTf) problem (Boylan-Kolchin, Bullock & Kaplinghat 2011, 2012). This problem is also discovered in the M31 system (Tollerud, Boylan-Kolchin & Bullock 2014) and field galaxies (Ferrero et al. 2012; Papastergis et al. 2015), which implies that this may not be caused by environmental interactions. The dark matter halo mass function can be predicted by the rotational velocity function of galaxies and the rotational velocity of a galaxy can be inferred from its velocity width of H I profile, as H I can extend farther in the radial direction.

\* E-mail: zk3kw2n@nao.cas.cn (KZ); jinwen@nao.cas.cn (JW); dili@nao.cas.cn (DL)



Larger samples are required to further investigate the origin of this phenomenon.

The large blind H I survey is the most efficient way in studying the statistical properties of H I galaxies like the HIMF, the velocity width function (WF), and the two-point correlation function (2PCF). The HIPASS (HI Parkes All Sky Survey; Barnes et al. 2001; Meyer et al. 2004) and the ALFALFA (Arecibo Legacy Fast ALFA; Giovanelli et al. 2005; Haynes et al. 2011, 2018; Giovanelli & Haynes 2015) survey are the two largest H I surveys to date, which have successfully provided the H I distribution properties in a fair cosmological volume. However, new questions have been raised due to the limited sensitivity and/or sky coverage. For example, from the HIPASS and  $\alpha$ 40 catalogues, no obvious clustering dependence of galaxies on H I mass has been identified by measuring the projected 2PCF (Basilakos et al. 2007; Meyer et al. 2007; Martin et al. 2012; Papastergis et al. 2013). However, Guo et al. (2017) found a strong positive correlation between H I mass and clustering properties through the effective volume-limited projected 2PCF measurements of the ALFALFA 70 per cent (a.k.a.  $\alpha$ .70) catalogue. The clustering measurements in the latter work are biased by the superclusters at H I masses below  $10^9 M_\odot$ , which demands a larger survey volume and sample size to achieve more robust measurements. In addition, as noted in Jones et al. (2018), the ‘faint end’ slope of the HIMF derived from the known H I-deficient Virgo cluster is flatter than that from the galaxies in its immediate surroundings. They hypothesized that the Virgo cluster might be in a gas-rich environment and the surrounded gas-rich filaments might feed the growth of the cluster, which needs deeper and wider observations to detect the low-mass galaxies in that region.

The Five-hundred-meter Aperture Spherical radio Telescope (FAST; Nan et al. 2011, Li & Pan 2016) is now the largest filled-aperture single dish (Jiang et al. 2020) and aims to carry out large-scale HI and pulsar surveys (Li et al. 2019). FAST plans to achieve high-quality H I imaging and pulsar searching simultaneously during observations by injecting a high-frequency electronic noise signal for calibration (CAL; Li et al. 2018). The survey plan is designated as the Commensal Radio Astronomy FAST Survey (CRAFTS). Benefitting from this innovative observation mode, CRAFTS will be able to conduct extragalactic H I, galactic H I imaging, pulsar search, and fast radio burst (FRB) search surveys simultaneously. The large collecting area and great sensitivity of FAST will allow a superior detectability on the H I gas in the extragalactic systems (Duffy et al. 2008; Zhang et al. 2019) and provide a high-quality data set for investigating the aforementioned issues. In the following sections unless specified, the CRAFTS refers to the two-pass CRAFTS extragalactic H I galaxy survey.

This paper is organized as follows: Section 2 describes the technical parameters of CRAFTS extragalactic H I survey and H I mass detection limit. In Sections 3 and 4, we estimate the number of H I galaxies CRAFTS may detect from Monte Carlo and semi-analytical simulation, respectively. Sections 5 and 6 discuss the recovered HIMF and source confusion rate from mock catalogues. Section 7 illustrates our improved normalization method of the HIMF, which is less susceptible to the selective effects of the survey. In Section 8, we discuss the influence of source confusion and radio frequency interference (RFI) on the survey. We will summarize our predicted results of CRAFTS in Section 9.

In this paper, we adopt a cosmology of  $H_0$  (Hubble constant) =  $70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ,  $\Omega_m$  (density parameter of matter) = 0.272, and  $\Omega_\Lambda$  (density parameter of dark energy) = 0.728 (Komatsu et al. 2011). The calculations of cosmological distances in this paper are mainly referred to Hogg (1999) and Meyer et al. (2017).

## 2 THE SIMULATED CRAFTS EXTRAGALACTIC H I SURVEY

### 2.1 Technical parameters of CRAFTS

As illustrated in Giovanelli et al. (2005), to achieve more detections, increasing the survey solid angle is more efficient than increasing the integration time. CRAFTS aims to complete a relatively shallow but wide survey within a declination (Dec.) range between  $-14^\circ$  and  $66^\circ$ , covering over  $20\,000 \text{ deg}^2$  of sky regions and reaching a redshift of 0.35. As discussed in Li et al. (2018) and Zhang et al. (2019, hereinafter Zh19), the aperture of FAST can be fully illuminated at zenith angles (ZAs) up to  $26.4^\circ$  and is only partially illuminated at ZAs up to  $40^\circ$ , so there is gain loss and an increase in beam size when ZA is above  $26.4^\circ$ . The system temperature of FAST will increase with ZA because of the radiation due to the surrounding mountain peaks entering the nearer sidelobes. To minimize the influence of gain fluctuation and RFI to the survey, CRAFTS plans to conduct a full two-pass drift-scan survey by the FAST L-band Array of 19-beam (FLAN) receiver, which will be rotated by  $23.4^\circ$  to achieve a super-Nyquist sampling while drifting (Li et al. 2018). In the now formally approved CRAFTS observing program, there is only time allowed for one pass. We aim to supplement the observation with open time proposals, which intends to be scheduled for high valued sky area (e.g. with clusters) about 6 months apart from the survey epoch.

The basic parameters of the CRAFTS extragalactic H I survey are summarized in Table 1. The technical parameters used in this paper, including gain, system temperature, and beam size, were obtained from commissioning observations and had been summarized in Zh19. Note that the effective integration time calculated in this paper is similar to equation (6) in Zh19, but we used the noise-equivalent beam area (NEBA) instead of the beam area as in Zh19. Since when combining data from multiple observations, the weighting factor should be the square of the beam response, in order to maximize the signal-to-noise ratio (SNR) (Staveley-Smith 1997). Therefore, the noise pattern will differ from the beam pattern and the NEBA will be a half of the beam area. As a result, the effective integration time in Zh19 is overpredicted by a factor of 2 from this point of view, and the sensitivity is overpredicted by a factor of  $\sqrt{2}$ , which leads the predicted average channel rms of two-pass CRAFTS to increase from 1.18 to 1.67 mJy. For comparison, after converting to the same velocity resolution of HIPASS and ALFALFA, the effective rms noise of CRAFTS is expected to be nearly 0.47 mJy (the sensitivity of HIPASS is 13 mJy with  $26 \text{ km s}^{-1}$  resolution; Barnes et al. 2001) and 0.67 mJy (the sensitivity of ALFALFA is 2.0 mJy with  $10 \text{ km s}^{-1}$  resolution; Haynes et al. 2018), respectively.

### 2.2 Flux limit for H I detection

For FAST, it is reasonable to assume that most of the observed H I galaxies are point sources (Duffy et al. 2008). If we ignore the peculiar motion of the Earth and H I sources, the frequency of an H I line we observe is  $\nu_{\text{obs}} = \nu_{\text{H I}}/(1+z)$ , where  $\nu_{\text{H I}} = 1420.4058 \text{ MHz}$  is the rest-frame H I frequency, and  $z$  is the redshift of the source. Using the non-relativistic assumption, the line-of-sight velocity can be approximated by  $cz$ , where  $c$  is the speed of light.

The velocity width of an H I line profile can be measured at 50 per cent level of each of two peaks or the full width of half-maximum (FWHM) of one peak, which is termed as  $W_{50}$ . The velocity width observed by the telescope is measured in the observed frame while the intrinsic velocity width of an H I source refers to the



**Table 1.** The technical parameters of CRAFTS extragalactic H I survey derived from commissioning observations. The two columns at the bottom reflect the change in sensitivity for the telescope at low and high ZAs. At ZAs above 26.4°, the reflecting surface is only partially illuminated. For detail descriptions, please see [Zh19](#).

Survey strategy	Two-pass drift scan	
Scan spacing	21.9 arcmin in declination	
Sky coverage	~ 20 000 deg <sup>2</sup>	
Declination range	−14° to +66°	
Number of beams	19	
Polarizations per beam	Full Stokes	
Frequency range	1050–1450 MHz	
Channel width	7.6 kHz (1.6 km s <sup>−1</sup> at 1420.4058 MHz)	
System temperature	18–26 K	
Zenith angle (ZA)	0–26.4°	26.4°–40°
Effective illuminated aperture size	300 m	250–300 m
Gain (centre beam)	16.5 K / Jy	11.0–16.5 K / Jy
Beamsize (FWHM) at 1420.4058 MHz	2.95 arcmin	2.95–3.60 arcmin

velocity width in the source’s rest frame, which is termed as  $W_{\text{obs}}$  and  $W_{\text{intri}}$ , respectively. Then the relation between  $W_{\text{obs}}$  and  $W_{\text{intri}}$  can be denoted by (Peacock 1999; Abdalla & Rawlings 2005; Meyer et al. 2017)

$$W_{\text{obs}} = (1 + z)W_{\text{intri}}, \quad (1)$$

where  $W_{\text{obs}}$  and  $W_{\text{intri}}$  are in km s<sup>−1</sup>.

The corresponding velocity width  $\Delta V_{\text{ch}}$  in observed frame within a channel width of  $\Delta f_{\text{ch}}$  can be estimated by (Meyer et al. 2017)

$$\Delta V_{\text{ch}} \simeq \frac{c(1+z)^2}{\nu_{\text{HI}}} \Delta f_{\text{ch}}. \quad (2)$$

We assume the peculiar motion speed and the velocity width of the source are much less than  $c$ . For simplicity, when calculating the velocity integrated flux of an H I line in terms of observed frame velocity using mock catalogues, we assume a top-hat line profile with a peak flux  $S_{\text{v}}^{\text{peak}}$ , so that  $S_{\text{int}}^{\text{obs}} = S_{\text{v}}^{\text{peak}} W_{\text{obs}}$ .

The SNR can be calculated by (Haynes et al. 2011)

$$\text{SNR} = S_{\text{v}}^{\text{peak}} \left( \frac{S_{\text{rms}}}{f_{\text{sno}}^{1/2}} \right)^{-1}, \quad (3)$$

where  $S_{\text{rms}}$  is the rms noise per channel,  $f_{\text{sno}}$  is the number of independent channels that the source signal can be smoothed over. As illustrated in Giovanelli et al. (2005), the SNR is best rendered when the noise is measured after smoothing a signal to a spectral resolution of about half of its linewidth, thus  $f_{\text{sno}}^1$  can be given by (Jones et al. 2015)

$$f_{\text{sno}} = \frac{1}{2 * \Delta V_{\text{ch}}} \begin{cases} W_{\text{obs}} & \text{if } W_{\text{obs}} \leq W_{\text{c}} \\ W_{\text{c}} & \text{if } W_{\text{obs}} > W_{\text{c}}, \end{cases} \quad (4)$$

where  $\log W_{\text{c}}/\text{km s}^{-1} = 2.5$  is the transition caused by the very broad spectral ‘standing waves’ resulting from the reflections in the telescope focal structure (e.g. Briggs et al. 1997; Haynes et al. 2011). Thus the velocity integrated flux limit for detecting an H I galaxy in terms of observed frame velocity can be estimated by

$$S_{\text{int,lim}}^{\text{obs}} = \text{SNR}_{\text{lim}} * S_{\text{rms}} W_{\text{obs}} / f_{\text{sno}}^{1/2}, \quad (5)$$

where  $\text{SNR}_{\text{lim}}$  is the limiting SNR for detections.

By combining equations (1), (2), (3), (4), and (5), one can obtain the velocity integrated flux limit in observed frame for detecting an

H I galaxy at a redshift of  $z$  and with an intrinsic velocity width of  $W_{\text{intri}}$ , which can be denoted by

$$\begin{aligned} \log \left( \frac{S_{\text{int,lim}}^{\text{obs}}}{\text{Jy km s}^{-1}} \right) &= \log \text{SNR}_{\text{lim}} + \log \left( \frac{S_{\text{rms}}}{\text{Jy}} \right) \\ &+ \frac{1}{2} \left[ \log \left( \frac{c}{\text{km s}^{-1}} \right) + \log \left( \frac{\Delta f_{\text{ch}}}{\text{kHz}} \right) - \log \left( \frac{\nu_{\text{HI}}}{\text{kHz}} \right) \right] \\ &+ \begin{cases} \frac{1}{2} \log \left( \frac{W_{\text{intri}}}{\text{km s}^{-1}} \right) + \frac{3}{2} \log (1 + z) & \text{if } W_{\text{obs}} \leq W_{\text{c}} \\ \log \left( \frac{W_{\text{intri}}}{\text{km s}^{-1}} \right) + 2 \log (1 + z) - 1.25 & \text{if } W_{\text{obs}} > W_{\text{c}}. \end{cases} \end{aligned} \quad (6)$$

The completeness of a survey can be defined as the fraction of cosmic sources from the underlying distributions that are detected by the survey. The specific form of the completeness function will differ from different smoothing and source extraction methods. For ALFALFA survey, the completeness function is defined by the relation between the velocity width and integrated flux density of H I galaxies (Haynes et al. 2011). As described in Rosenberg & Schneider (2002) and Haynes et al. (2011), using a 50 per cent completeness limit and a full completeness limit as the selection criterion obtains approximately similar statistical results. We estimate the selection function of CRAFTS by referring to the ALFALFA 50 per cent completeness function. By replacing the sensitivity per channel of ALFALFA 40 per cent (3.4 mJy per 24.4 kHz;<sup>2</sup> Haynes et al. 2011; Jones et al. 2015) and redshift of 0 into equation (6), an SNR of 5.75 will give a close approximation to the 50 per cent completeness function of the ALFALFA survey, which can be used to derive the HIMF of the survey samples (e.g. Martin et al. 2010; Jones et al. 2018). Thus, we adopt the  $\text{SNR}_{\text{lim}}$  of 5.75 and other technical parameters mentioned before to estimate the observed velocity integrated flux limit of CRAFTS through equation (6), from which the H I mass detection limit in terms of observed frame velocity integrated flux can be denoted by (Roberts 1975; Meyer et al. 2017)

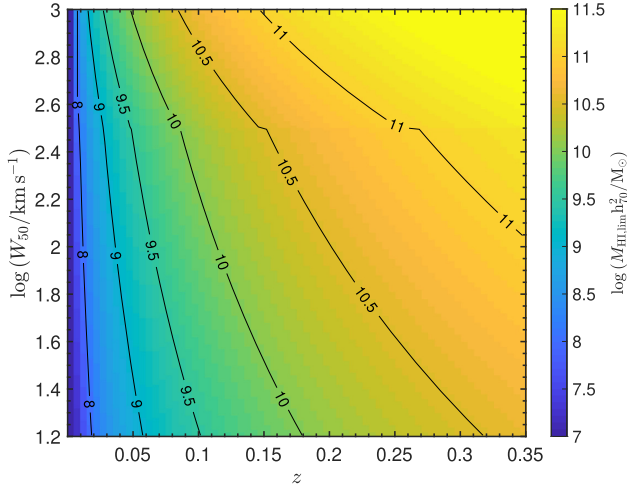
$$\left( \frac{M_{\text{HI,lim}}}{M_{\odot}} \right) = \frac{2.35 \times 10^5}{(1 + z)^2} \left[ \frac{D_{\text{L}}(z)}{\text{Mpc}} \right]^2 \left( \frac{S_{\text{int,lim}}^{\text{obs}}}{\text{Jy km s}^{-1}} \right), \quad (7)$$

where  $D_{\text{L}}(z)$  is the luminosity distance at a redshift of  $z$ .

<sup>2</sup>Noted that the average rms of ALFALFA has been revised in ALFALFA 100 per cent (Haynes et al. 2018), which is 2.8 mJy per 24.4 kHz resolution. In this work, our aim is to obtain the ALFALFA 50 per cent completeness function and compare it with Jones et al. (2015), we thus adopt the ALFALFA 40 per cent value (Haynes et al. 2011).

<sup>1</sup>After conformed by M. G. Jones privately, there might be a typo in the interpretation of  $f_{\text{sno}}$  in Jones et al. (2015).





**Figure 1.** The H I mass limit of two-pass CRAFTS survey in the redshift and intrinsic velocity width plane at a ZA of  $0^\circ$ . The solid lines represent the H I mass limit contours.

The channel width of CRAFTS is approximately 7.6 kHz and the value of  $S_{\text{rms}}$  varies with the ZA and redshift of the source (for detail discussion, see Zh19). Fig. 1 shows the H I mass detection limit in the redshift and intrinsic velocity width plane at a ZA of  $0^\circ$  for two-pass CRAFTS. The deviations of H I mass limit at  $\log W_{50}/\text{km s}^{-1} = 2.5$  is caused by the ‘standing wave’, which makes it harder to detect H I galaxies with wider velocity widths.

In this paper, we use two kinds of mock catalogues to predict for CRAFTS. One is the Monte Carlo simulation based on the ALFALFA catalogue, where no large-scale structure (LSS) of the universe is included. The low-mass end in Monte Carlo simulation extends to  $10^{6.4} M_\odot$ , which can be used to test the measurements of low-mass end of the HIMF and give a rather complete prediction on galaxy detection. Another is the semi-analytical simulation based on the  $\Lambda$ CDM model, which considers the environmental information of H I galaxies, by which we can estimate the influence of source confusion on the survey. However, the resolution of the simulation limits the low-mass end to  $10^9 M_\odot$ .

### 3 THE MONTE CARLO SIMULATION

#### 3.1 The ALFALFA mass–width function

The ALFALFA survey is the largest completed blind H I survey so far, which covers a sky region of nearly  $7000 \text{ deg}^2$  to a redshift of 0.06. Here, we use full ALFALFA or the  $\alpha.100$  catalogue (Haynes et al. 2018; Jones et al. 2018) to calculate the mass–width function (MWF; Papastergis et al. 2011) of the H I galaxies, which gives the distribution of the number density of H I galaxies as a function of H I mass and velocity width. The MWF of H I galaxies is an intermediate product of the calculation of the HIMF and the velocity width function (WF), which depicts the number density distribution of H I galaxies as a function of H I mass and velocity width. The HIMF and WF can be obtained by summing up the MWF along the velocity width dimension and H I mass dimension, respectively.

As blind surveys are usually flux limited, which means the samples in the survey are selected by its flux limit (for H I surveys, there is also velocity width dependence), it is required to compensate for selective effects while estimating the volume density. Schmidt (1968) developed the ‘ $1/V_{\text{max}}$ ’ method to calculate the number density of

a flux-limited survey by weighting each detected source with the reciprocal of  $V_{\text{max}}$ , where  $V_{\text{max}}$  is the maximal comoving volume in which the source can be detected by the survey. However, the ‘ $1/V_{\text{max}}$ ’ method is very sensitive to the LSS of the universe, which can be reflected by the spatial distribution of clusters and voids. For nearby H I surveys, significant bias will be caused by the overdensity of H I galaxies in the local volume (Martin et al. 2010).

The step-wise maximum likelihood (SWML) method (Efstathiou, Ellis & Peterson 1988) is developed to reduce the effect of LSS, which assumes the number density in each separated bin is constant and maximizes the joint likelihood of detecting all samples in the survey. The detection of H I galaxies will also depend on the velocity width of the galaxies. In this case, the two-dimensional SWML (2DSWML) method (Loveday 2000; Zwaan et al. 2003) or the ‘ $1/V_{\text{eff}}$ ’ (Zwaan et al. 2005) method is developed to calculate the MWF by splitting the MWF or  $\phi$  into bins of  $m = \log(M_{\text{H I}}/M_\odot)$  and  $w = \log(W_{50}/\text{km s}^{-1})$ .  $\phi(m, w)$  can be obtained by summing up the  $1/V_{\text{eff}}$  of each H I galaxy in individual bin, where  $V_{\text{eff}}$  is the effective volume of the galaxy.  $V_{\text{eff}}$  can be obtained by maximizing the joint likelihood of detecting all galaxies in the survey and is equivalent to  $V_{\text{max}}$  in the ‘ $1/V_{\text{max}}$ ’ method (Zwaan et al. 2005). We utilize the ‘ $1/V_{\text{eff}}$ ’ method to calculate the MWF of the  $\alpha.100$  catalogue. The detail calculation procedure of the MWF can be found in Zwaan et al. (2005), Martin et al. (2010), Papastergis et al. (2011), and references therein.

The final samples to calculate the MWF are from the catalogue to calculate the 100 per cent ALFALFA HIMF (provided by M. G. Jones via private communication; Jones et al. 2018). The catalogue we use here contains 22 798 H I galaxies in total with following cuts:  $z \leq 0.05$ ,  $6.4 \leq m \leq 11$ ,  $1.2 \leq w \leq 3.0$ , and 50 per cent completeness limit for high SNR sources. The  $m$  and  $w$  range is chosen to make the derived number density closed to the fitted HIMF. The redshift cut is to avoid the strong contamination of the RFI from the aviation radar, the redshift gap caused by which would otherwise influence the performance of the 2DSWML method in calculating the HIMF (Haynes et al. 2011).

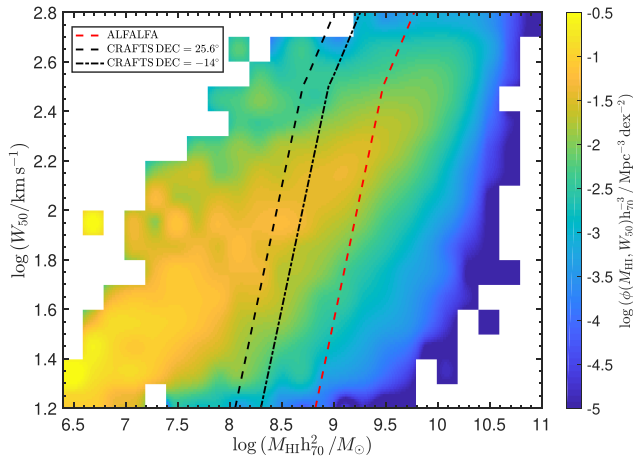
The MWF is first calculated in a bin width of 0.2 in  $m$  and  $w$  and then approximated through interpolation<sup>3</sup> into a bin width of 0.01. The interpolated MWF derived from the  $\alpha.100$  catalogue is shown in Fig. 2. The finer distribution of  $\phi(m, w)$  is used to produce the random number of  $m$  and  $w$  in Monte Carlo simulation and we keep the total number density of H I galaxies a constant before and after interpolation.

#### 3.2 Simulation and results

To estimate the number of galaxies that would be detected by CRAFTS, we generate a mock catalogue through Monte Carlo simulation, similar to the approach in Jones et al. (2015), based on the ALFALFA MWF. We assume that H I galaxies are uniformly distributed in the universe, so that the right ascension (RA), the sine of Declination (Dec.), and the comoving volume ( $V_C$ ) of H I galaxies can be generated through uniformly distributed random numbers in CRAFTS sky. For the H I mass and velocity width, we utilize the normalized MWF mentioned before as the probability density function (PDF) of  $m$  and  $w$ . We assume the number volume density of H I galaxies does not evolve with redshift and the value is

<sup>3</sup>The interpolation is performed through the function of ‘interp2’ in the software of MATLAB, the method used here is ‘makima’, which stands for the Modified Akima cubic Hermite interpolation.





**Figure 2.** The MWF derived from the  $\alpha.100$  catalogue after interpolation to a bin width of 0.01. The red dashed line represents the 50 per cent completeness H I mass limit at a redshift of 0.025 for the ALFALFA survey. Similarly the black dashed, and dash–dotted line represents the CRAFTS limit at a Dec. of  $25.6^\circ$  and  $-14.0^\circ$ , respectively. The galaxies located to the left of those lines will not be detected by the survey. The H I mass detection limit at a Dec. of  $-14.0^\circ$  is higher than that at a Dec. of  $25.6^\circ$ , mainly due to the gain loss and increase of system temperature at high ZAs.

approximately  $0.1161 h_{70}^{-3} / \text{Mpc}^{-3}$ , which is calculated by integrating the  $\alpha.100$  HIMF (Jones et al. 2018) from  $m$  of 6.4 to 11. The total number of H I galaxies can be obtained by multiplying the total number density of H I galaxies with the comoving volume of the survey. The redshift of a galaxy at a given comoving volume in the survey region can be approximated by its comoving distance. We assume the simulated H I galaxy can be detected when its H I mass is above the H I mass limit estimated by equation (7).

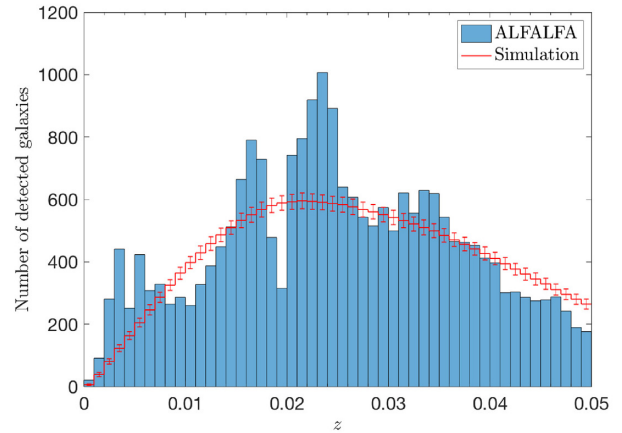
To test our model, we adopt the ALFALFA 50 per cent completeness function (Haynes et al. 2011) as the selection criterion to see whether we can recover the number of detected galaxies of the ALFALFA survey from Monte Carlo simulation. Given the ALFALFA survey mainly focuses on the local volume ( $z \leq 0.05$ ), we ignore the cosmological expansion effects. Thus, the comoving distance and luminosity distance are both approximated by  $cz/H_0$ .

The average number of detections in our 1000 simulations for ALFALFA survey is 20984, with an average fluctuation of approximately 4.7 per cent. This is approximately 8 per cent less than the observational samples, which lies above the 50 per cent completeness limit of ALFALFA and contains 22798 (Code 1) detections in total. The comparison between simulation and observation is shown in Fig. 3. The neglect of the LSS in the universe might be the main reason of the deviation in redshift distribution of detected galaxies.

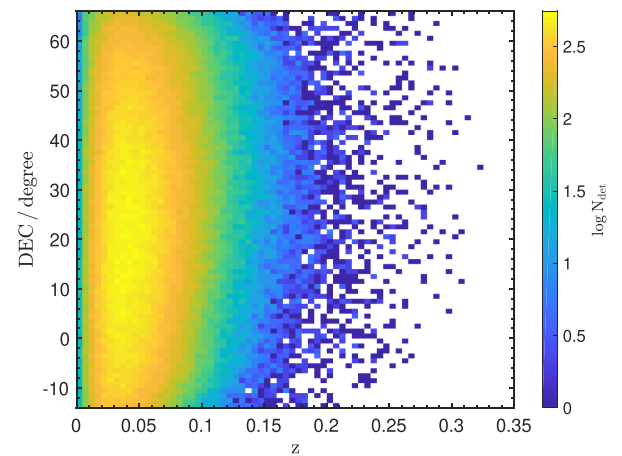
The FAST 19-beam receiver can reach 1.05 GHz (corresponding to  $z = 0.35$ ). We thus consider the influence of cosmic expansion and overlook the peculiar motions of galaxies when calculating the cosmological distances. The distribution of the number of detections for CRAFTS in the  $z$ –Dec. plane is shown in Fig. 4. CRAFTS will not be able to detect H I galaxies efficiently at redshifts above 0.1 due to the limited integration time.

#### 4 THE SEMI-ANALYTICAL SIMULATION

The mock catalogue we use is based on the semi-analytical models of galaxy formation from Fu et al. (2013) and Luo et al. (2016), which is a branch of L-Galaxies models (e.g. Kauffmann et al. 1999;



**Figure 3.** The redshift distribution of the  $\alpha.100$  catalogue and average result from 1000 simulations, which is shown as blue histogram and red solid line, respectively. The uncertainty comes from the maximal fluctuation of 1000 simulations.

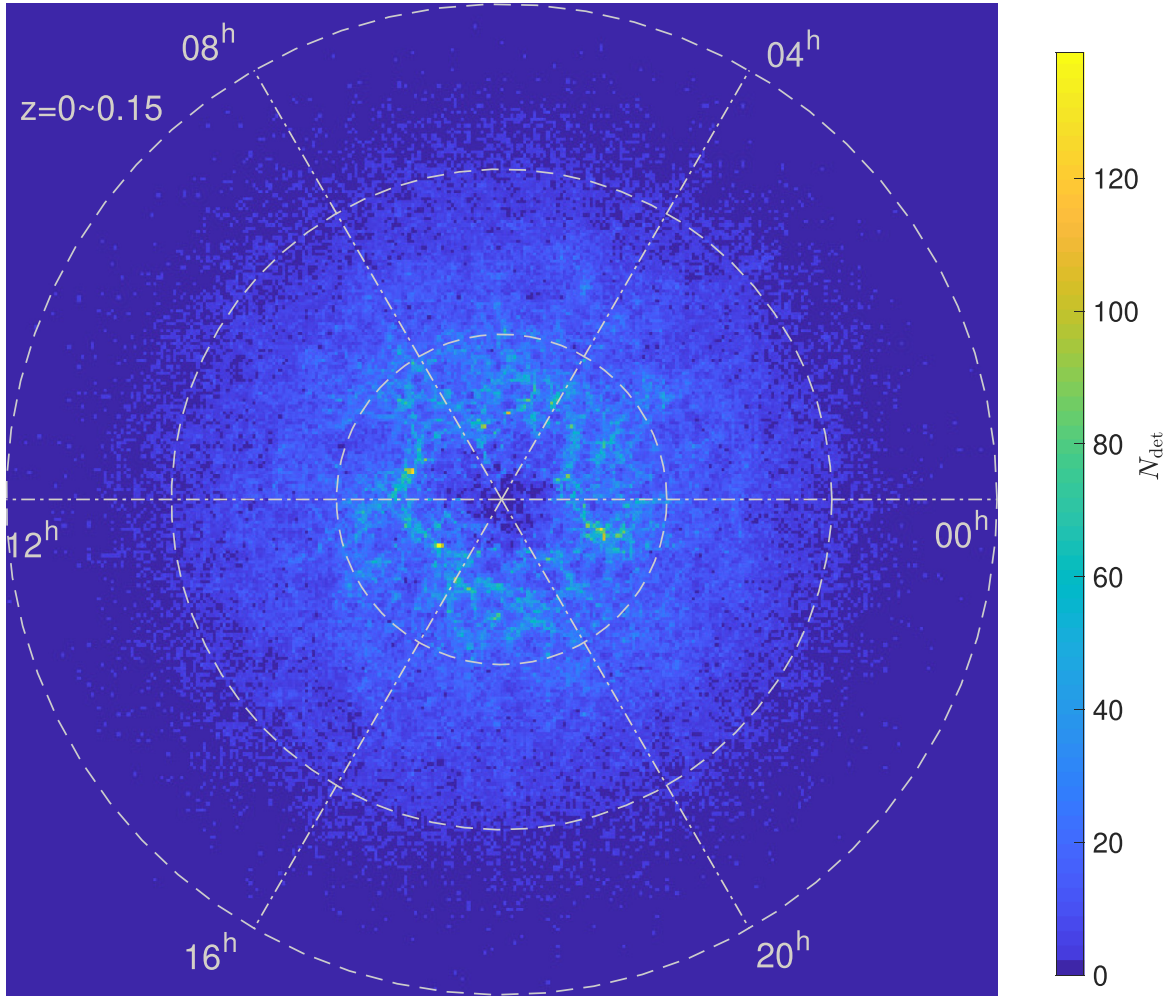


**Figure 4.** The number of detected H I galaxies of CRAFTS in the  $z$ –Dec. plane estimated from Monte Carlo simulation. The pixel size is 0.01 in redshift and  $1^\circ$  in declination. The colourbar represents the number of detected galaxies in the volume bin of a 24-h drift scan within the boundary of each pixel size. We assume the H I galaxies in the universe are uniformly distributed. The variation at different declinations implies the impact of telescope parameters on galaxy detection.

Springel et al. 2001; Croton et al. 2006; De Lucia & Blaizot 2007; Fu et al. 2010; Guo et al. 2011, 2013). This branch includes the prescriptions of atomic gas (H I) and molecular gas ( $\text{H}_2$ ) in interstellar medium (ISM; e.g. cold gas cooling, star formation law, ram pressure stripping). The mock catalogue is based on two  $\Lambda$ CDM  $N$ -body simulations: the Millennium Simulation (Springel et al. 2005) and the Millennium-II Simulation (Boylan-Kolchin et al. 2009), and we adopt the cosmology parameters from WMAP7 ( $\Omega_\Lambda = 0.728$ ,  $\Omega_m = 0.272$ ,  $\Omega_{\text{baryon}} = 0.0454$ ,  $\sigma_8 = 0.807$ , and  $h = 0.704$ ; Komatsu et al. 2011). The resolution of the simulation limits the minimum H I mass of galaxies in the mock catalogue to  $10^9 M_\odot$ . Details of the models can be found in Fu et al. (2013), Henriques et al. (2015), Luo et al. (2016), and references therein.

The WF derived from the  $\Lambda$ CDM simulation shows a large discrepancy compared to the ALFALFA observation results (Obreschkow et al. 2009; Papastergis et al. 2011), especially at the low-velocity width end, which is referred as the CDM overabundance problem





**Figure 5.** The number of detected galaxies projected into the  $z$ -RA plane derived from semi-analytical simulation for two-pass CRAFTS. The spacing between two adjacent grey dashed circles is 0.05 in redshift. The colourbar represents the number of galaxies detected in the area of each rectangle pixel, with a size of 0.001 in the plane.

(Papastergis et al. 2011). As discussed in Jones et al. (2015), the predicted number of detections estimated from the ALFALFA MWF is around 60–75 per cent of the results based on the  $\Lambda$ CDM simulation (Duffy et al. 2012b) for the future survey complemented by ASKAP (the Australian SKA Pathfinder; Johnston et al. 2008; Deboer et al. 2009). The mock catalogue we use also overpredicts the number of H I galaxies with low-velocity widths. To make our model match with existing observation results, we use the mass-conditional velocity width function (MCWF) derived from the ALFALFA survey (Jones et al. 2015) to generate the velocity widths of H I galaxies by Monte Carlo simulation. The MCWF gives the PDF of the velocity width of a galaxy at a given H I mass, which is fitted by the Gumbel distribution (Martin et al. 2010).

Fig. 5 displays the number distribution of detections projected into  $z$ -RA plane derived from semi-analytical simulation. Cosmic web features can be seen below a redshift of 0.05, which demonstrates the potential of CRAFTS in exploring the galaxy interactions in different environments and the spatial distribution of H I galaxies in the local universe. The comparison of the total number of detections and mean redshift from two simulations is presented in Table 2. The predicted results from two simulations are consistent within an H I mass range between  $10^9$  and  $10^{11} M_{\odot}$ , that is because we

utilize the MCWF from the ALFALFA survey to produce the velocity widths of galaxies in the semi-analytical simulation. Fig. 6 illustrates the number of detections of CRAFTS as a function of redshift and H I mass in two simulations. Corrected by the ALFALFA MWF, the number of detections from semi-analytical simulation is nearly 50 per cent of the predicted results without correction. The bottom panel of Fig. 6 compared the number of detections with and without correction as a function of H I mass from  $10^9$  to  $10^{11} M_{\odot}$ , which highlights the importance of better measurements of the WF and in matching the models to observations not just in terms of mass or luminosity function, but also the WF.

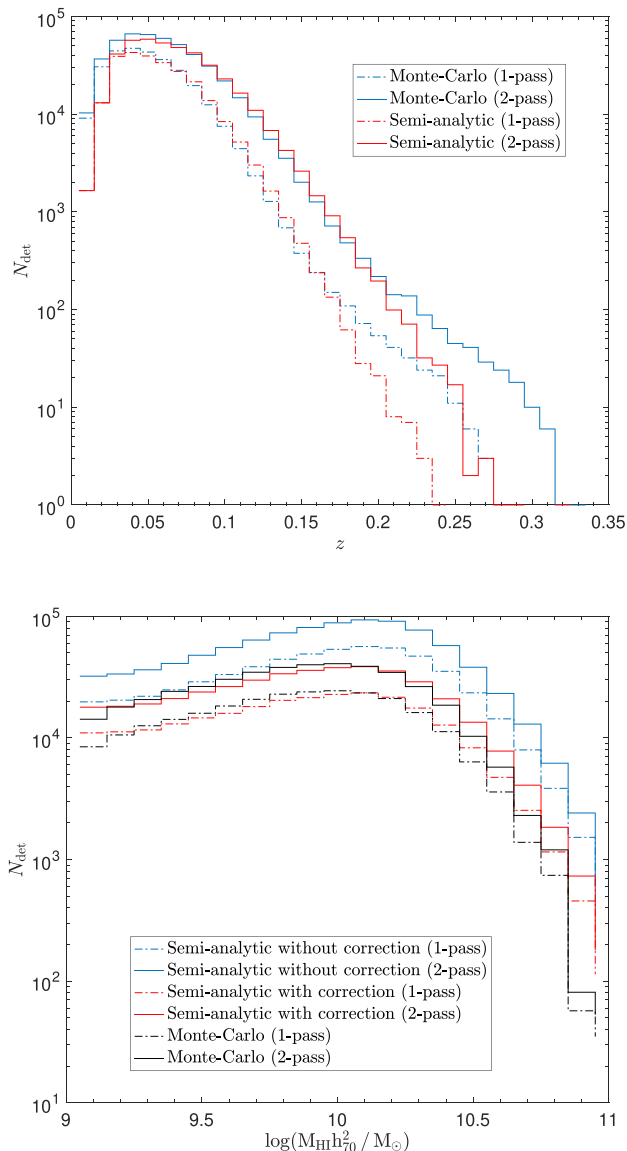
In our predictions, one-pass CRAFTS will be able to detect nearly  $2.9 \times 10^5$  H I galaxies with a mean redshift of 0.047, while for two-pass survey, around  $4.8 \times 10^5$  galaxies will be detected with a mean redshift of 0.055. We neglect the influence of RFI on source detections. The potential impact of RFI on CRAFTS is discussed in Section 8.2. The accurate number of detected galaxies from these two simulations could be different given simulations could not perfectly simulate the observed HIMF in the ALFALFA survey. However, this deviation is much smaller compared to the results without correction.

Our predicted results show that the number of H I galaxies of two-pass CRAFTS is nearly 21 times of the ALFALFA catalogue we use



**Table 2.** The predicted results of CRAFTS from Monte Carlo and semi-analytical simulations. The Monte Carlo simulation is based on the MWF of the ALFALFA survey, within an H I mass range between  $10^{6.4}$  and  $10^{11} M_{\odot}$ . The semi-analytical simulation is constructed in a frame of the  $N$ -body  $\Lambda$ CDM models with an H I mass resolution of  $10^9 M_{\odot}$  (Fu et al. 2013; Henriques et al. 2015; Luo et al. 2016). We corrected the velocity widths by the ALFALFA MCWF (Martin et al. 2010; Jones et al. 2015) considering the discrepancy in the derived velocity WF between observations and simulations. The value in the parentheses represents the number of detected galaxies with H I masses ranging from  $10^9$  to  $10^{11} M_{\odot}$  in Monte Carlo simulation.

Mock catalogue	Monte Carlo simulation		Semi-analytical simulation	
Survey strategy	One-pass	Two-pass	One-pass	Two-pass
Number of detections	290 000 (260 000)	480 000 (430 000)	250 000	420 000
Mean redshift	0.047 (0.051)	0.055 (0.060)	0.052	0.061



**Figure 6.** Top: The predicted number of galaxies detected by CRAFTS as a function of redshift. The red and blue lines represent the results from Monte Carlo and semi-analytical simulation, respectively. Bottom: The comparison of predicted results as a function of H I mass among the semi-analytical simulation, semi-analytical simulation corrected by ALFALFA MCWF (Jones et al. 2015) and Monte Carlo simulation, which are shown as blue lines, red lines, and black lines, respectively. The dash-dotted lines and solid lines are from one-pass and two-pass survey, respectively.

in this paper. That is mainly benefitted by larger sky coverage, wider bandwidth, and better sensitivity of CRAFTS. The sky coverage of CRAFTS is nearly  $23\,000 \text{ deg}^2$ , while the ALFALFA catalogue we use covers approximately  $6600 \text{ deg}^2$ . The ALFALFA catalogue reaches a redshift of 0.05, while CRAFTS will be able to detect galaxies efficiently to a redshift of 0.10. Our calculation indicates that two-pass CRAFTS will detect nearly 65 000 galaxies in the survey volume of ALFALFA catalogue, which is  $\sim 3$  times of the ALFALFA catalogue. The expected average sensitivity of CRAFTS is approximately 1.67 mJy per 7.6 kHz, which is  $\sim 3.0$  times of the ALFALFA survey (2.8 mJy per 24.4 kHz; Haynes et al. 2018) after smoothing to same channel width.

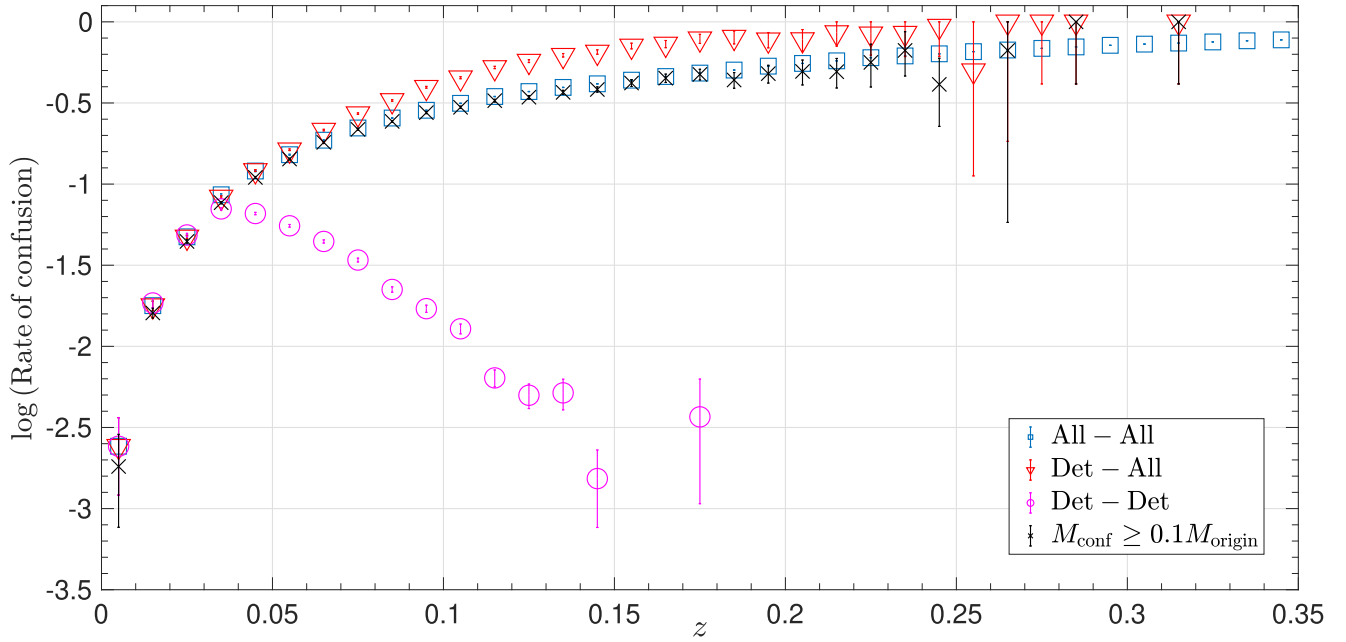
## 5 SOURCE CONFUSION

Source confusion occurs when multiple sources are located within the same beam at similar velocity channels and cannot be distinguished. For H I galaxy surveys, two galaxies are confused when they are projected within one beam in the sky and overlapped in the line-of-sight direction by their velocity widths (e.g. Duffy et al. 2012a, b; Jones et al. 2015). In a shallow survey like CRAFTS, confusion effects may manifest at higher redshifts due to a large beam size at  $L$  band and broader velocity widths of galaxies at high- $z$ . This may cause an overestimation of H I mass and underestimation of the number of detectable galaxies presented in a beam, which might bias the observed HIMF and WF.

The actual source confusion relies on the distribution of H I mass, velocity width, and clustering property. As the Monte Carlo simulation does not contain any clustering property of H I galaxies, which will underestimate the rate of confusion, we only calculate the results from semi-analytical simulation. Our semi-analytical models are based on  $N$ -body cosmological simulations, which can reflect the spatial distribution of galaxies/clusters in real universe. In our previous papers of L-Galaxies SAM (e.g. Fu et al. 2017; Henriques et al. 2020), our results of HIMF can approximately fit the results of ALFALFA, and the models can also reproduce the 2PCF when we consider the process of stripping of the cold gas by ram pressure (Luo et al. 2016). The H I velocity WF in our mock catalogue is corrected by ALFALFA observational results. Our predictions about source confusion could act as a reference for the future survey.

We followed the method in Jones et al. (2015) to define whether two galaxies are confused. To judge whether two galaxies are overlapped in the line-of-sight direction, we use the simulated channel number of the observed galaxies in the backend. We convert the redshift and velocity width of a galaxy to the channel number that the galaxy occupies in the backend of FAST, which is obtained by dividing the observed velocity width by the channel velocity width. The channel velocity width can be approximated through





**Figure 7.** The rate of confusion of CRAFTS calculated from semi-analytical simulation. The ‘All–All’ represents the rate of confusion of simulated galaxies blended with any other simulated galaxies (blue rectangles). The ‘Det–All’ represents the rate of confusion of detected galaxies blended with other simulated galaxies (red inverted triangles). The ‘Det–Det’ represents the rate of confusion between detections (magenta circles). The rate of confusion when the additional confused mass  $M_{\text{conf}}$  is beyond 10 per cent of its original mass  $M_{\text{origin}}$  is shown as black crosses. The uncertainty is from the Poisson counting error.

equation (2), we assume the channel velocity width is constant for each galaxy. Two galaxies are considered to be confused if the occupied channels have any overlaps (thus overlap in velocity space even just by the edges of flat-top profiles) and their separation in the sky is less than the beam size. The actual source confusion rate might be improved as CRAFTS will do a super-Nyquist sampling so that some confused galaxies can be actually recognized after post-survey reductions.

The rate of confusion is obtained by dividing the number of blended sources by the total number of sources in different redshift ranges with an interval of 0.01. Following Duffy et al. (2012b), we calculate the rate of confusion in three different cases. The ‘All–All’ case illustrates the confusion between any simulated galaxies, no matter they are detected or not, which is shown as blue rectangles in Fig. 7. The ‘Det–All’ case represents the detected galaxies blended with any other simulated galaxies, normalized by the total number of detected galaxies, which is shown as red inverted triangles in Fig. 7. The ‘Det–Det’ case stands for the detected galaxies blended with other detections, which is shown as magenta circles in Fig. 7. We also calculate the rate of confusion when the additional H I mass caused by confusion is beyond 10 per cent of its original mass, which is shown as black crosses in Fig. 7. The uncertainty in Fig. 7 is Poissonian.

A noticeable feature in Fig. 7 is that ‘Det–All’ is higher than ‘All–All’ at same redshifts, which indicates that the detected galaxies are more likely to be blended with other galaxies. This result is consistent with Duffy et al. (2012b), who interpreted the effect mainly as the result of that high-mass galaxies tend to be located in denser environments and the clustering property is overpredicted by the simulation. However, we have found a similar blending feature in Monte Carlo simulation, where no clustering property is included. Our mock catalogue has been corrected by the ALFALFA MCWF, as shown in Fig. 2, most of low-velocity width galaxies are located

in the ‘faint end’ of the HIMF, which are hard to detect and occupy a large fraction of number densities. So the velocity widths of detected galaxies are wider than most of galaxies in the mock catalogue, which makes the detected galaxies more likely to be blended with other galaxies in the line-of-sight direction.

In Jones et al. (2016), they estimated the expected average confused mass for FAST would be  $\sim 10^9 M_{\odot}$  by a redshift of  $\sim 0.1$  by assuming a cylinder volume with a velocity width of  $600 \text{ km s}^{-1}$ . The average confused mass in our mock catalogue around the redshift of 0.1 is nearly  $2.0 \times 10^9 M_{\odot}$ , which is consistent with their results. Our prediction implies that ‘Det–Det’ will reduce nearly 2 per cent of detected galaxies in semi-analytical simulation, which suggests that source confusion will not play a significant role in the detection of H I galaxies for CRAFTS.

## 6 THE RECOVERY OF THE HIMF

The HIMF depicts the number density of H I galaxies as a function of H I mass, which can be well fitted by the Schechter function (Schechter 1976). The HIMF could be expressed in the form of

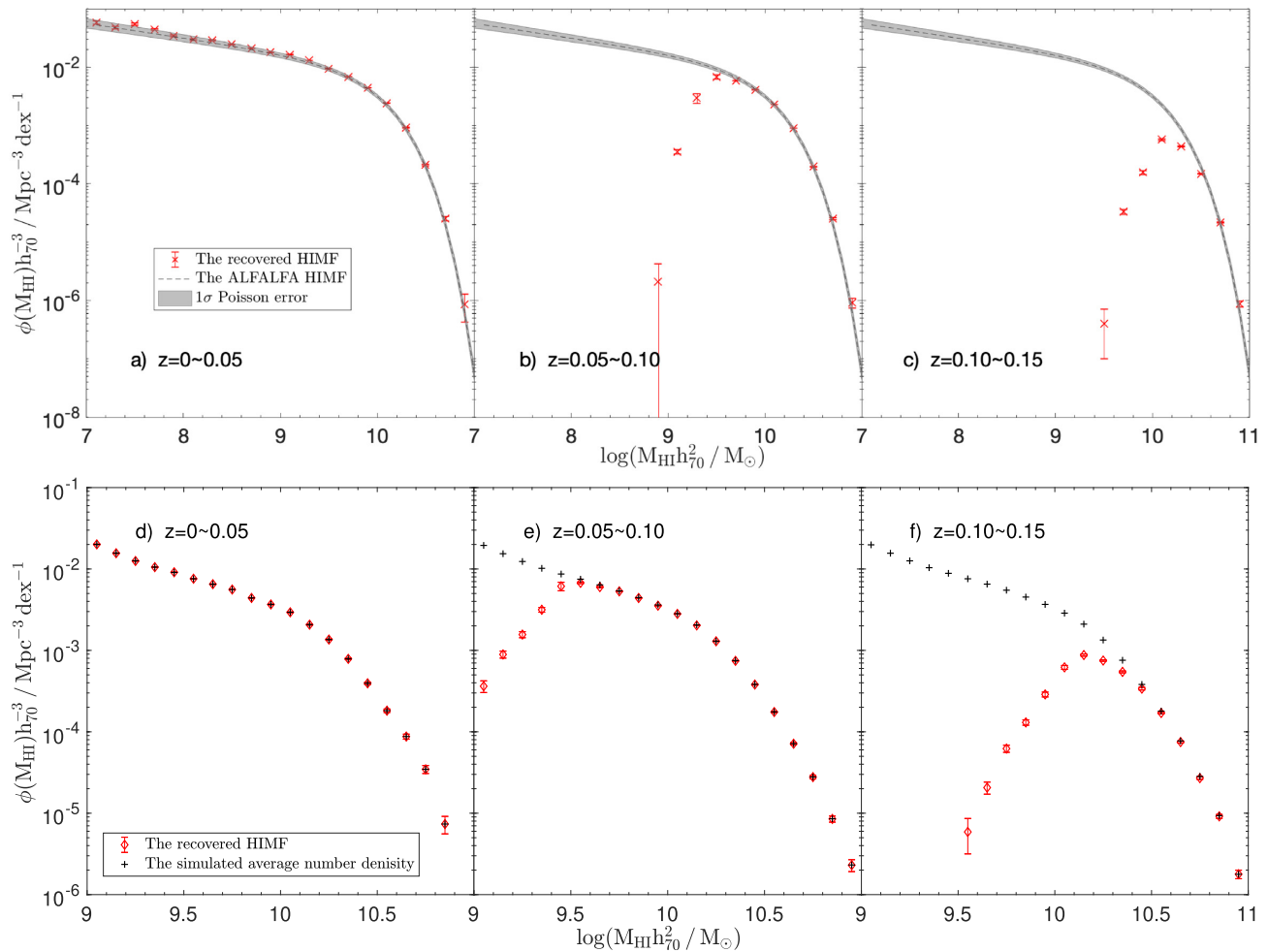
$$\phi(M_{\text{H I}}) = \frac{dN_{\text{gal}}}{dV \log M_{\text{H I}}}, \quad (8)$$

$$= \ln(10) \phi_* \left( \frac{M_{\text{H I}}}{M_*} \right)^{\alpha+1} e^{-\left( \frac{M_{\text{H I}}}{M_*} \right)}, \quad (9)$$

where  $dN_{\text{gal}}$  is the average number of galaxies in comoving volume element  $dV$ ,  $\phi_*$  is the normalization constant,  $M_*$  is the ‘knee mass’ (we will often use  $m_* = \log M_*/M_{\odot}$ ), and  $(\alpha + 1)$  is the low-mass slope, which is usually referred as the ‘faint end’ slope of the HIMF.

Measuring the HIMF is one of main science goals of CRAFTS. As discussed in Zh19, beam size of FAST is slightly small at low





**Figure 8.** The recovered HIMF of CRAFTS calculated through the ‘ $1/V_{\text{eff}}$ ’ method, the uncertainty is Poissonian. Top three panels: The recovered HIMF from the Monte Carlo simulation based on the ALFALFA MWF, the red crosses represent the observed HIMF, the ALFALFA 100 per cent HIMF (Jones et al. 2018) is represented as black dashed line, with  $1\sigma$  Poisson error shown as grey shaded area. Bottom three panels: The HIMF recovered from semi-analytical simulation. The red diamonds stand for the observed HIMF and the black pluses show the average number density of galaxies in the mock catalogue.

frequencies to adjust the wide bandwidth of the receiver, which will add complexity to the calculation of the simulated integration time and selection function. For example, at ZAs below  $26.4^\circ$ , the actual beam size of FAST at a frequency of 1050 MHz will be nearly 0.11 arcmin smaller than the theoretical beam size if we assume beam size is inversely proportional to frequency.

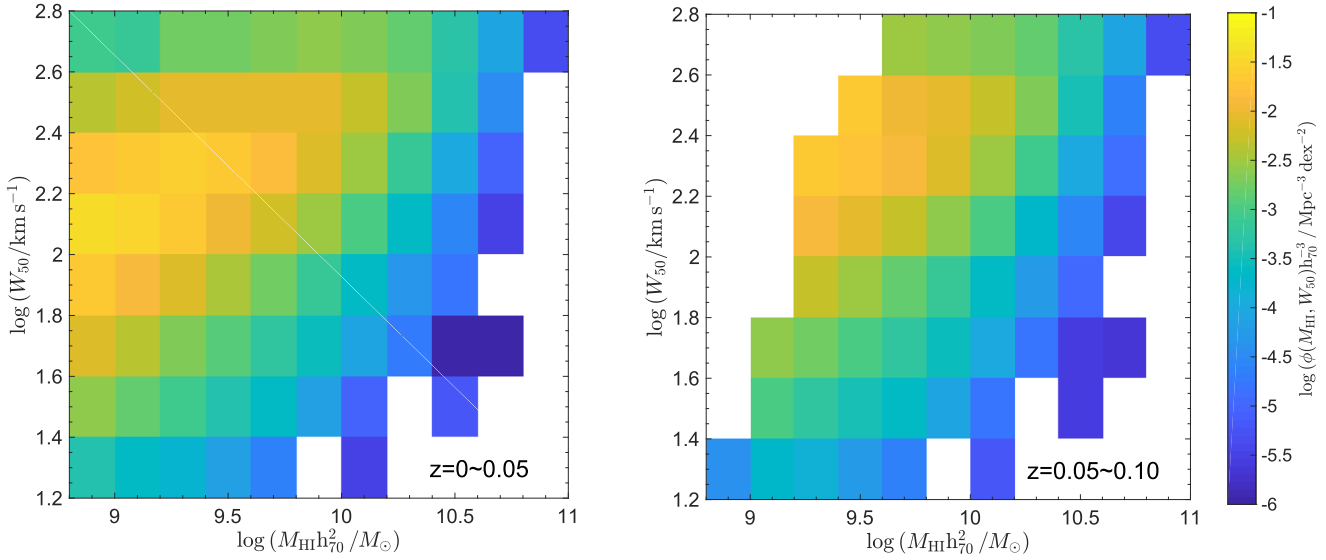
To simplify the calculation of the HIMF, we assume that the beam size of FAST is inversely proportional to frequency, we change the parameter value of equation (1) and (2) in Zh19 to  $m_1 = 3.24$  arcmin when ZA is below  $26.4^\circ$  and  $m_2 = 0$ , where  $m_1$  is the beam size at a frequency of 1250 MHz and  $m_2$  is the correction parameter. This will slightly decrease the sensitivity of the survey but give a more simplified form of the simulated selection function.

We use the ‘ $1/V_{\text{eff}}$ ’ method mentioned previously to calculate the HIMF from the selected samples of two-pass CRAFTS. The survey volume is separated into three parts from a redshift of 0 to 0.15 with an interval of 0.05. The recovered HIMF from Monte Carlo and semi-analytical simulation is shown in Fig. 8. As in Monte Carlo simulation, CRAFTS is not sensitive enough to fully recover the number density of galaxies with  $m$  below 7, we calculate the HIMF

from  $m$  of 7 to 11 with a bin size of 0.2, same as the bin size to calculate the ALFALFA HIMF. The result is shown in the top three panels of Fig. 8. The HIMF from semi-analytical simulation, limited by its mass resolution, is obtained from  $m$  of 9 to 11 with an interval of 0.1. The comparison of the recovered HIMF and average number density from semi-analytical simulation is presented in the bottom three panels of Fig. 8. From Fig. 8, we can see that CRAFTS will be able to recover the ‘faint end’ of the HIMF to  $10^7 M_\odot$  and measure the ‘knee mass’ to a redshift of 0.1, which will be the first time to achieve such a deep observation of the global ‘knee mass’.

Fig. 9 illustrates the measured MWF by CRAFTS from Monte Carlo simulation at  $z = 0-0.05$  and  $0.05-0.10$ . At higher redshifts and low H I masses, CRAFTS is not sensitive to galaxies with wide H I profiles. These galaxies are not detected, so cannot be included in HIMF calculation, which can be reflected by the blank area at  $z = 0.05-0.10$ . The low-mass end of the recovered HIMF therefore underestimates the true HIMF at  $z > 0.05$ , which can be seen from Fig. 8. For a single-dish telescope like FAST, the measurement of the HIMF might be biased by source confusion, which will be discussed in Section 8.1.





**Figure 9.** The MWF revealed by CRAFTS from Monte Carlo simulation at two redshift slices. Limited by its sensitivity, low-mass galaxies with wide velocity widths cannot be detected by CRAFTS at higher redshifts and will not be included in MWF calculation, which can be reflected by the blank area in the right-hand panel. This will cause the depression of HIMF measurements at low masses in Fig. 7.

## 7 THE NORMALIZATION OF THE HIMF

In 2DSWML, the non-normalized density parameter can be obtained from the iteration of

$$\phi_{jk} = \frac{n_{jk}}{\sum_i \frac{H_{ijk}}{\sum_m \sum_n H_{imn} \phi_{mn}}}, \quad (10)$$

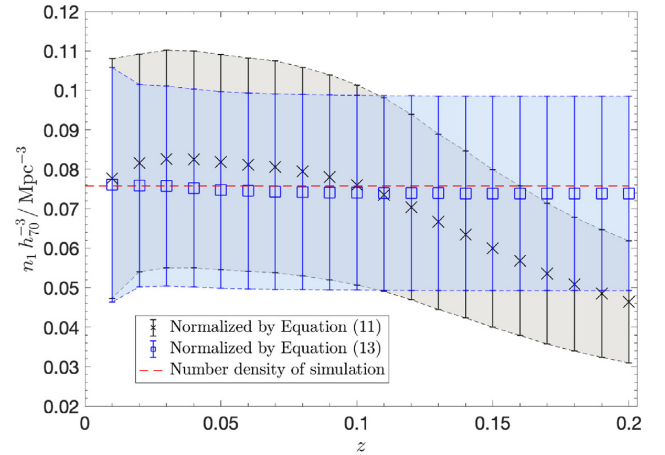
where  $\phi_{jk}$  is the non-normalized density parameter in  $m$  bin of  $j$  and  $w$  bin of  $k$ ,  $n_{jk}$  is the total number count of galaxies in bin  $jk$ , and  $H_{ijk}$  is the fraction of areas of bin  $jk$  in  $(m, w)$  plane that lies above the selection function of galaxy  $i$ . The detail form of  $H_{ijk}$  can be found in the appendix of Martin et al. (2010), Papastergis et al. (2013), and references therein.

The denominator of equation (10) represents the total number of galaxies located in bin  $jk$  in the survey volume, which is corrected by the selection function of each galaxy in the sample (Zwaan et al. 2005; Papastergis et al. 2013). Assuming the number density of galaxies is constant, the denominator of equation (10) is proportional to the survey volume and can be treated as the ‘effective volume’. Similar to the standard ‘ $1/V_{\text{max}}$ ’ method, equation (10) can be converted in the form of summing the reciprocal of the ‘effective volume’ of each galaxy in bin  $jk$  (Zwaan et al. 2005).

The normalization of the HIMF is obtained by matching the total number density to the estimator of average number density. There are three estimators described in Davis & Huchra (1982), which are most suitable in different situations. As the selection function of the ALFALFA is better understood, the estimator which is less prone to bias is adopted, which can be defined as (Davis & Huchra 1982; Martin et al. 2010)

$$n_1 = V_{\text{survey}}^{-1} \int \frac{n(D) dD}{S(D)} = V_{\text{survey}}^{-1} \sum_i \frac{1}{S(D_i)}, \quad (11)$$

where  $n(D) dD$  is the number of galaxies in a spherical shell of thickness  $dD$  and radius  $D$ ,  $V_{\text{survey}}$  is the total volume of the survey, and  $S(D)$  is the selection function at the distance of  $D$ . This method assumes galaxies are uniformly distributed and considers galaxy  $i$  located at the centre of the corresponding shell. Thus,  $1/S(D_i)$

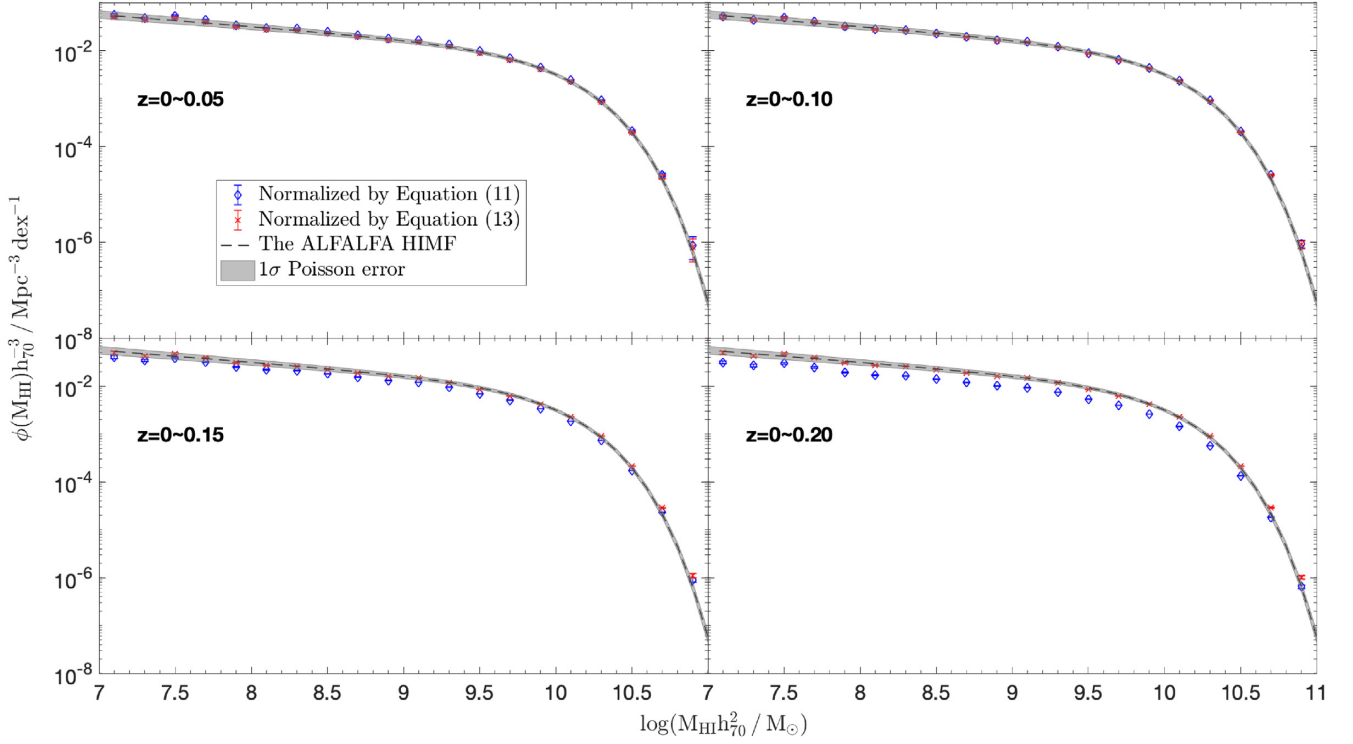


**Figure 10.** The average number density in different volumes derived from CRAFTS samples in Monte Carlo simulation. The horizontal axis represents the redshift boundary of survey volume. We make total number density of galaxies constant in the simulation, which is shown as red dashed line. The black crosses are the results calculated by equation (11) (Martin et al. 2010), we can see the fluctuation trend of  $n_1$  is similar to that of the number of detected galaxies as a function of redshift, which may imply that the variation is caused by the non-uniform distance distribution of detected galaxies. The blue rectangles represent the results calculated by equation (13), which looks more stable. The uncertainty is Poissonian.

represents the total number of galaxies in the shell centred at  $D_i$ , where  $D_i$  is the distance of galaxy  $i$ .

As the survey is flux limited, the number of detected galaxies will decrease at high redshifts, which means  $\sum_i 1/S(D_i)$  due to the selective effects is not proportional to the comoving volume. The distance distribution of distant galaxies is much sparser than those nearby, which will cause a variation of  $n_1$ . We use the galaxies detected by CRAFTS in Monte Carlo simulation to calculate  $n_1$  in different survey volumes. The result of equation (11) is represented





**Figure 11.** The comparison of HIMFs normalized by equations (11) (Martin et al. 2010) and (13) in different redshift ranges, which is shown in blue diamonds and red crosses, respectively. The HIMFs are calculated from CRAFTS detections in Monte Carlo simulation. The dashed line represents the ALFALFA 100 per cent HIMF (Jones et al. 2018), with  $1\sigma$  Poisson counting error shown as grey shade area. As can be seen from the bottom two panels, when including samples from higher redshifts, the HIMF would be underestimated by equation (11).

as black crosses in Fig. 10, from which we can see a noticeable decline of  $n_1$  as survey volume increases.

The 2DSWML method splits number density into different mass and velocity width bins. For galaxies located at same bin, the selective effects will not differ too much therefore the number of detection will not be strongly biased by distance. We can also calculate the average number density in each bin and then sum up to obtain the total number density. Thus,  $n_1$  can be interpreted by

$$n_1 = \sum_j \sum_k n_{1,jk} = \sum_j \sum_k V_{\text{survey},jk}^{-1} \int \frac{n_{jk}(D) dD}{S_{jk}(D)}, \quad (12)$$

where  $n_{1,jk}$  is the average number density in bin  $jk$ ,  $V_{\text{survey},jk}$  is the survey volume of bin  $jk$ , the boundary of which can be obtained from the most distant galaxy in bin  $jk$ . Similar to equation (11),  $n_{jk}(D) dD$  is the number of galaxies belonging to bin  $jk$  in a spherical shell of thickness  $dD$  and radius  $D$ , and  $S_{jk}(D)$  is the fraction of galaxies detectable in bin  $jk$  at the distance of  $D$ . In 2DSWML,  $n_{1,jk}$  can be interpreted by

$$n_{1,jk} = V_{\text{survey},jk}^{-1} \sum_l \frac{1}{H_{ljk}}, \quad (13)$$

for every galaxy  $l$  in bin  $jk$  and  $H_{ljk} > 0$ .

The total average number density calculated from equation (13) is shown as blue rectangles in Fig. 10. One can see the results are much more stable as the survey volume increases, within 5 per cent difference compared to the mean number density integrated from the input HIMF. For comparison, the maximum deviation of  $n_1$  calculated from equation (11) is nearly 10 per cent at redshifts below 0.1. Fig. 11 compares the HIMFs from Monte Carlo simulation within four volumes normalized by equations (11) and 13, from which we

can see equation (11) would lead to an underestimation of HIMF when including sources in larger volumes.

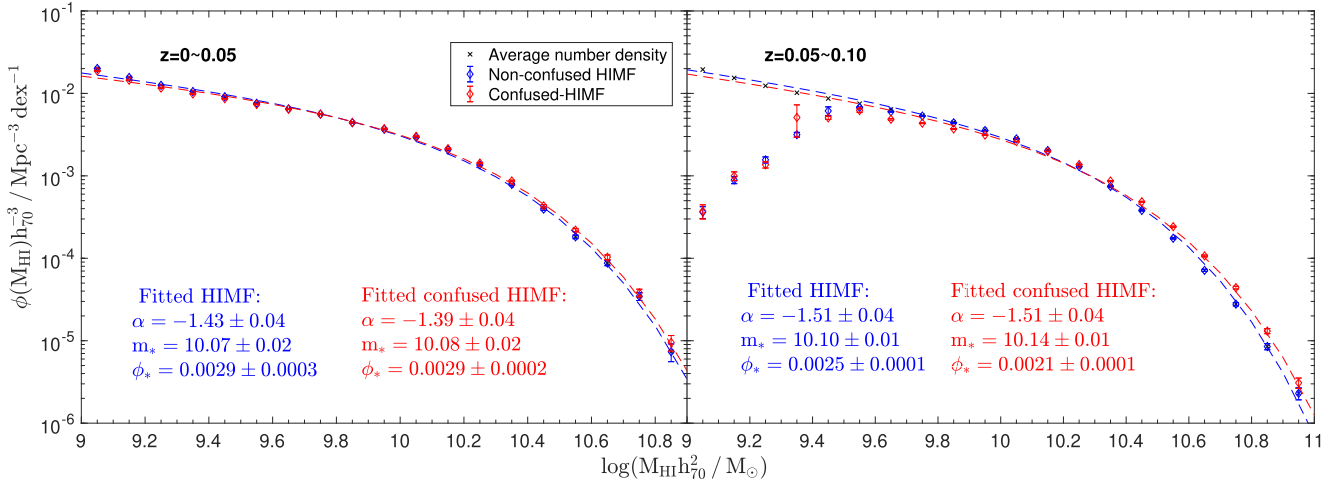
Equation (13) normalizes the HIMF by the detections in every distinct bin, thus if there are a small number of detected galaxies in a bin, it may induce large uncertainty in the calculation of the total average number density, which will influence the accuracy of the normalization. While calculating the number density in different redshift slices, the redshifts of some galaxies are very close to the lower boundary of the redshift bin, which may underestimate the survey volume of individual bin and overestimate the total number density. Equation (11) normalizes the HIMF by including all of the galaxies in the survey volume and the detected galaxies in every redshift bin tend to be uniformly distributed. Thus, equation (11) will give a more accurate estimation while calculating HIMFs in different redshift slices. Our results in Figs 8 and 12 are both calculated from equation (11). These two methods could be complementary in normalizing the HIMF.

## 8 DISCUSSION

### 8.1 The impact of confusion effect on CRAFTS HIMF measurements

The measured HIMF might be biased by confusion effect from real distribution, especially at higher redshifts. The actual source confusion relies on the distribution of H I mass, velocity width, and clustering property. In this work, we use the samples from semi-analytical simulation to roughly estimate the potential influence of source confusion on the observed HIMF of CRAFTS. Following Jones et al. (2015), we sum up the H I mass of blended sources





**Figure 12.** The comparison of confused and non-confused HIMF or the observed HIMF with and without considering source confusion effects simulated for CRAFTS. The HIMF is calculated from semi-analytical simulation by the ‘ $1/V_{\text{eff}}$ ’ method, the uncertainty is from the Poisson counting error. The black crosses represent the average number density of the mock catalogue, the blue and red diamonds stand for the non-confused and confused HIMF, respectively. The fitted Schechter functions are denoted by dashed lines. At  $z = 0.05\text{--}0.10$ , the HIMFs are fitted in an  $m$  range of 9.7–11.0 and the ‘faint end’ slope is adopted from the average number density of the catalogue.

together as the final confused mass to achieve an upper limit of the impact of confusion on HIMF measurements. For detections blended with non-detections, we assume other properties like velocity widths and positions stay constant after considering source confusion. For cases of confusion between two detections, we count blended sources as a single source, consider the boundary of blended H I profiles in the line-of-sight direction as the final confused velocity width and adopt the average value of positions. The criteria of source confusion are described in Section 5.

We calculate the confused HIMF to a redshift of 0.1, that is because, as illustrated in Fig. 8, CRAFTS will be able to measure the ‘knee mass’ of the HIMF to that redshift. The result is shown in Fig. 12, from which we can see larger uncertainties are produced by source confusion at higher redshift end. As mentioned previously, CRAFTS will not be able to recover the number density of low-mass galaxies at  $z = 0.05\text{--}10$ , we thus fit the HIMF in an  $m$  range from 9.7 to 11.0 by adopting the ‘faint end’ slope of the average number density of the mock catalogue and assuming the ‘faint end’ slope does not change for confused HIMF. The deviation of  $m_*$  caused by source confusion is within  $1\sigma$  and  $4\sigma$  Poisson counting error at redshifts of  $z = 0\text{--}0.05$  and  $0.05\text{--}0.10$ , respectively. As our assumptions have overestimated the confused H I mass to a large extent, the actual impact of source confusion could be significantly weaker.

Jones et al. (2015) found that confusion will steepen the ‘faint end’ slope of the HIMF, but our fitted results from semi-analytical simulation seems to differ from their results. That is likely because the low-mass end of the HIMF we measured only reaches  $10^9 M_\odot$ , which is not the actual ‘faint end’ of the HIMF. The ‘faint end’ slope of the HIMF is mainly constrained by low-mass galaxies, only a few of which can be detected at very low redshifts due to the limited sensitivity of the telescope. While at low redshifts, the angular size of the beam corresponds to a smaller physical size, thus making confusion less likely. As the H I mass increases, source confusion will have a larger impact on detected galaxies. From the left-hand panel of Fig. 12, we can see the decreased number density caused by source confusion when  $m$  is below  $m_*$ . The decreased number density at  $m < m_*$  will drag down the ‘faint end’ slope while fitting, which is consistent with the results in Jones et al. (2015). Considering

the extremely low number density at high-mass end, a small number of added sources will lead to notable changes of observed number density, which could influence the observed ‘knee mass’ of the HIMF.

## 8.2 The RFI status at FAST site

The impact of RFI on source detection can be vital on FAST, especially considering the receiver of FAST covers a wide range of frequencies. Currently, the severest contamination is from satellites and aircraft navigation beacons, which corresponds to redshifts from 0.1 to 0.2 for H I observations. Previous data show that these kinds of strong RFI will occupy nearly half of the observation time. Considering the RFI contamination and sensitivity limitation, CRAFTS will be difficult to detect galaxies at redshifts above 0.1. If we only consider the H I galaxies at redshifts below 0.1, in Monte Carlo simulation, CRAFTS will be able to detect approximately  $2.8 \times 10^5$  and  $4.4 \times 10^5$  galaxies for one-pass and two-pass survey, with a mean redshift of 0.04 and 0.05, respectively.

## 8.3 Compared with other planned large-scale H I surveys

Apart from FAST, several large telescopes including ASKAP (Johnston et al. 2007, 2008), MeerKAT (the South African Meer-Karoo Array Telescope; Jonas et al. 2016), and Apertif (APERture Tile In Focus) upgradation of WSRT (Westerbork Synthesis Radio Telescope; Oosterloo, Verheijen & van Cappellen 2010) also plan to complement large-scale H I surveys.

Those surveys can be divided into two types. One is to target a small area with long integration time to achieve depth. LADUMA<sup>4</sup> (Looking at the Distant Universe with MeerKAT Array; Blyth et al. 2016) intends to study the cosmic evolution of H I galaxies to a redshift of  $z \sim 1.4$  with a single pointing of 3424 h; MIGHTEE-HI<sup>5</sup> (H I component of the ‘MeerKAT International GHz Tiered Extragalactic Exploration’ survey; Jarvis et al. 2016) aims to cover a

<sup>4</sup><http://www.laduma.uct.ac.za/>

<sup>5</sup><http://idia.ac.za/mightee/>



sky coverage of  $20 \text{ deg}^2$  to  $z \sim 0.5$  with 16 h per pointing; Apertif<sup>6</sup> plans to complement a medium-deep survey covering  $\sim 450 \text{ deg}^2$  to a redshift of 0.26, observed for  $7 \times 12 \text{ h}$ ; carried by ASKAP, DINGO<sup>7</sup> (Deep Investigation of Neutral Gas Origins; Johnston et al. 2008; Duffy et al. 2012b) is arranged in two tiers, named as DINGO-Deep and DINGO-Ultradeep, which plans to cover  $150 \text{ deg}^2$  ( $5 \times 500 \text{ h}$ ) to redshifts  $0 < z < 0.26$  and  $60 \text{ deg}^2$  ( $2 \times 2500 \text{ h}$ ) to redshifts  $0.1 < z < 0.43$ , respectively. The other type is relatively shallow of large sky areas. WNSHS<sup>6,8</sup> (the Westerbork Northern Sky HI Survey; Koribalski 2012) aims to cover  $\sim 3500 \text{ deg}^2$  to a redshift of 0.26 with 12 h per pointing, detecting about 90 000 galaxies out to a redshift of  $z \sim 0.1$ ; WALLABY<sup>9</sup> (The Widefield ASKAP L-band Legacy All-sky Blind Survey; Johnston et al. 2008; Duffy et al. 2012b; Koribalski et al. 2020) plans to observe 75 per cent of the sky ( $-90^\circ < \text{Dec.} < +30^\circ$ ) to redshifts of  $z \lesssim 0.26$ , detecting  $\sim 500\,000$  H I galaxies.

As discussed previously, FAST will be easily contaminated by source confusion while observing H I galaxies at high redshifts compared to interferometers, limited by its relatively large beam size (e.g. the resolution of WALLABY is  $\sim 30 \text{ arcsec}$ ). However, a major advantage of FAST is its large collecting area, which allows the detection of very low mass galaxies nearby. CRAFTS ( $-14^\circ < \text{Dec.} < +66^\circ$ ) and WALLABY ( $-90^\circ < \text{Dec.} < +30^\circ$ ) cover approximately 95 per cent of the sky in total, the combination of their final catalogues will enlarge the census of H I galaxies unprecedentedly, which is excellent in investigating the statistical properties of H I galaxies in the local universe. The combined observational data from CRAFTS and WALLABY will benefit from the higher resolution of ASKAP and better sensitivity of FAST, which will give a full sampling of nearby faint H I galaxies and their surroundings (e.g. Koribalski et al. 2020).

## 9 CONCLUSIONS

In this paper, we have briefly introduced the technical parameters of CRAFTS extragalactic H I galaxy survey derived from commissioning observations, and predicted the potential of the survey on H I galaxy detection and HIMF measurements. We summarized our main results as the following:

- (i) CRAFTS will be able to detect approximately  $2.9 \times 10^5$  and  $4.8 \times 10^5$  galaxies ranging from  $10^{6.4}$  to  $10^{11} M_\odot$  for one-pass and two-pass survey, with a mean redshift of nearly 0.047 and 0.055, respectively. Considering the RFI situation and sensitivity limitation, CRAFTS will effectively survey H I galaxies at redshifts below 0.1.
- (ii) CRAFTS will recover the ‘faint end’ slope of nearby galaxies to  $10^7 M_\odot$  and measure the global ‘knee mass’ to a redshift of 0.1.
- (iii) From our simulation, source confusion would reduce approximately 2 per cent of detections for CRAFTS. After considering source confusion effects, the measured ‘knee mass’ of the HIMF would be overestimated by  $1\sigma$  and  $4\sigma$  Poisson counting error at redshifts of 0–0.05 and 0.05–0.10, respectively.
- (iv) The large survey area and outstanding sensitivity performance enable CRAFTS to better explore the galaxy interactions in different environments and the spatial distribution of H I galaxies in the local universe. CRAFTS will also enlarge the census of low-mass galaxies

and constrain the ‘faint end’ slope of nearby HIMF, shedding some light on the ‘missing satellite’ problem.

## ACKNOWLEDGEMENTS

The author would like to thank helpful discussion with Marko Krčo, Michael G. Jones, Ying Zu, and Sean E. Lake. This work is supported by the National Natural Science Foundation of China grant No. 11988101, No. 11725313, No. 11673029, No. 11590783, No. 11973051; the National Key Research and Development Program of China grant No. 2017YFA0402600, No. 2016YFA0400702; the Open Project Program of the Key Laboratory of FAST, NAOC, Chinese Academy of Sciences; the International Partnership Program of Chinese Academy of Sciences grant No. 114A11KYSB20160008. JW thanks the support from National Natural Science Foundation of China grant No. 12073002 and No. 11721303. JF acknowledges the support by the Youth Innovation Promotion Association, CAS and Shanghai Committee of Science and Technology grant No. 19ZR1466700.

## DATA AVAILABILITY

The data underlying this article will be shared on reasonable requests to the corresponding authors.

## REFERENCES

- Abdalla F. B., Rawlings S., 2005, *MNRAS*, 360, 27
- Barnes D. G. et al., 2001, *MNRAS*, 322, 486
- Basilakos S., Plionis M., Kovac K., Voglis N., 2007, *MNRAS*, 378, 301
- Blanton M. R., Lupton R. H., Schlegel D. J., Strauss M. A., Brinkmann J., Fukugita M., Loveday J., 2005, *ApJ*, 631, 208
- Blyth S. et al., 2016, *Proc. Sci.*, On the Pathway to the SKA. SISSA, Trieste, p. PoS#4
- Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, *MNRAS*, 398, 1150
- Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2011, *MNRAS*, 415, L40
- Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2012, *MNRAS*, 422, 1203
- Briggs F. H., Sorar E., Kraan-Korteweg R. C., van Driel W., 1997, *Publ. Astron. Soc. Aust.*, 14, 37
- Croton D. J. et al., 2006, *MNRAS*, 365, 11
- Davis M., Huchra J., 1982, *ApJ*, 254, 437
- Deboer D. R. et al., 2009, *Proc. IEEE*, 97, 1507
- De Lucia G., Blaizot J., 2007, *MNRAS*, 375, 2
- Drlica-Wagner A. et al., 2020, *ApJ*, 893, 47
- Duffy A. R., Battye R. A., Davies R. D., Moss A., Wilkinson P. N., 2008, *MNRAS*, 383, 150
- Duffy A. R., Moss A., Staveley-Smith L., 2012a, *Publ. Astron. Soc. Aust.*, 29, 202
- Duffy A. R., Meyer M. J., Staveley-Smith L., Bernyk M., Croton D. J., Koribalski B. S., Gerstmann D., Westerlund S., 2012b, *MNRAS*, 426, 3385
- Efstathiou G., Ellis R. S., Peterson B. A., 1988, *MNRAS*, 232, 431
- Ferrero I., Abadi M. G., Navarro J. F., Sales L. V., Gurovich S., 2012, *MNRAS*, 425, 2817
- Fu J., Luo Y., 2017, *Proc. IAU Symp.* 321, Formation and Evolution of Galaxy Outskirts. Kluwer, Dordrecht, p. 131
- Fu J., Guo Q., Kauffmann G., Krumholz M. R., 2010, *MNRAS*, 409, 515
- Fu J. et al., 2013, *MNRAS*, 434, 1531
- Garrison-Kimmel S. et al., 2017, *MNRAS*, 471, 1709
- Geha M., Blanton M. R., Masjedi M., West A. A., 2006, *ApJ*, 653, 240
- Giovanelli R., Haynes M. P., 2015, *A&AR*, 24, 1
- Giovanelli R. et al., 2005, *AJ*, 130, 2598
- Guo H., Li C., Zheng Z., Mo H. J., Jing Y. P., Zu Y., Lim S. H., Xu H., 2017, *AJ*, 846, 61

<sup>6</sup><http://old.astron.nl/radio-observatory/apertif-surveys>

<sup>7</sup><https://dingo-survey.org/survey-design/>

<sup>8</sup>[http://www.astron.nl/phisc2014/Documents/Surveys/jozsa\\_dwingeloo\\_wnshs.pdf](http://www.astron.nl/phisc2014/Documents/Surveys/jozsa_dwingeloo_wnshs.pdf)

<sup>9</sup><https://www.atnf.csiro.au/research/WALLABY/>



- Guo Q. et al., 2011, *MNRAS*, 413, 101
- Guo Q., White S., Angulo R. E., Henriques B., Lemson G., Boylan-Kolchin M., Thomas P., Short C., 2013, *MNRAS*, 428, 1351
- Haynes M. P., 2019, *Proc. IAU Symp. 344, Dwarf Galaxies: From the Deep Universe to the Present*. Kluwer, Dordrecht, p. 3
- Haynes M. P. et al., 2011, *AJ*, 142, 170
- Haynes M. P. et al., 2018, *ApJ*, 861, 49
- Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, *MNRAS*, 451, 2663
- Henriques B. M. B., Yates R. M., Fu J., Guo Q., Kauffmann G., Srisawat C., Thomas P. A., White S. D. M., 2020, *MNRAS*, 491, 5795
- Hill D. T., Driver S. P., Cameron E., Cross N., Liske J., Robotham A., 2010, *MNRAS*, 404, 1215
- Hogg D. W., 1999, preprint ([arXiv:9905116](https://arxiv.org/abs/9905116))
- Jarvis M. et al., 2016, *Proc. Sci., On the Pathway to the SKA*. SISSA, Trieste, p. PoS#6
- Jiang P. et al., 2020, *Res. Astron. Astrophys.*, 20, 64
- Johnston S. et al., 2007, *Publ. Astron. Soc. Aust.*, 24, 174
- Johnston S. et al., 2008, *Exp. Astron.*, 22, 151
- Jonas J., MeerKAT Team, 2016, *Proc. Sci., On the Pathway to the SKA*. SISSA, Trieste, p. PoS#1
- Jones M. G., Papastergis E., Haynes M. P., Giovanelli R., 2015, *MNRAS*, 449, 1856
- Jones M. G., Haynes M. P., Giovanelli R., Papastergis E., 2016, *MNRAS*, 455, 1574
- Jones M. G., Haynes M. P., Giovanelli R., Moorman C., 2018, *MNRAS*, 477, 2
- Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999, *MNRAS*, 303, 188
- Kelley T., Bullock J. S., Garrison-Kimmel S., Boylan-Kolchin M., Pawlowski M. S., Graus A. S., 2019, *MNRAS*, 487, 4409
- Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999, *ApJ*, 522, 82
- Komatsu E. et al., 2011, *ApJS*, 192, 18
- Koposov S. E., Yoo J., Rix H.-W., Weinberg D. H., Macciò A. V., Escudé J. M., 2009, *AJ*, 696, 2179
- Koribalski B. S., 2012, *Publ. Astron. Soc. Aust.*, 29, 359
- Koribalski B. S. et al., 2020, *Ap&SS*, 365, 118
- Li D., Pan Z., 2016, *Radio Sci.*, 51, 1060
- Li D. et al., 2018, *IEEE Microw. Mag.*, 19, 112
- Li D., Dickey J. M., Liu S., 2019, *Res. Astron. Astrophys.*, 19, 016
- Loveday J., 2000, *MNRAS*, 312, 557
- Luo Y., Kang X., Kauffmann G., Fu J., 2016, *MNRAS*, 458, 366
- Macciò A. V., Kang X., Fontanot F., Somerville R. S., Koposov S., Monaco P., 2010, *MNRAS*, 402, 1995
- Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, *ApJ*, 723, 1359
- Martin A. M., Giovanelli R., Haynes M. P., Guzzo L., 2012, *ApJ*, 750, 38
- Meyer M. J. et al., 2004, *MNRAS*, 350, 1195
- Meyer M. J., Zwaan M. A., Webster R. L., Brown M. J. I., Staveley-Smith L., 2007, *ApJ*, 654, 702
- Meyer M., Robotham A., Obreschkow D., Westmeier T., Duffy A. R., Staveley-Smith L., 2017, *Publ. Astron. Soc. Aust.*, 34, 52M
- Montero-Dorta A. D., Prada F., 2009, *MNRAS*, 399, 1106
- Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, *ApJ*, 524, L19
- Nadler E. O., Mao Y.-Y., Wechsler R. H., Garrison-Kimmel S., Wetzel A., 2018, *ApJ*, 859, 129
- Nan R. et al., 2011, *IJMPD*, 20, 989
- Obreschkow D., Croton D., De Lucia G., Khochfar S., Rawlings S., 2009, *ApJ*, 698, 1467
- Oosterloo T. A., Verheijen M., van Cappellen W., 2010, in van Leeuwen J., ed., *Proc. ISKAF Science Meeting*, Assen, the Netherlands
- Papastergis E., Martin A. M., Giovanelli R., Haynes M. P., 2011, *ApJ*, 739, 38
- Papastergis E., Giovanelli R., Haynes M. P., Rodríguez-Puebla A., Jones M. G., 2013, *ApJ*, 776, 43
- Papastergis E., Giovanelli R., Haynes M. P., Shankar F., 2015, *A&A*, 574, A113
- Peacock J. A., 1999, *Cosmological Physics*. Cambridge Univ. Press, Cambridge
- Press W. H., Schechter P., 1974, *ApJ*, 187, 425
- Roberts M. S., 1975, in Sandage M., Kristan J., eds, *Galaxies and the Universe*. University of Chicago Press, Chicago
- Rosenberg J. L., Schneider S. E., 2002, *ApJ*, 567, 1
- Schechter P., 1976, *ApJ*, 203, 297
- Schmidt M., 1968, *ApJ*, 151, 393
- Schombert J. M., McGaugh S. S., Eder J. A., 2001, *AJ*, 121, 2420
- Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, *MNRAS*, 328, 726
- Springel V. et al., 2005, *Nature*, 435, 629
- Staveley-Smith L., 1997, *Publ. Astron. Soc. Aust.*, 14, 111
- Tollerud E. J., Bullock J. S., Strigari L. E., Willman B., 2008, *ApJ*, 688, 277
- Tollerud E. J., Boylan-Kolchin M., Bullock J. S., 2014, *MNRAS*, 440, 3511
- Weinberg D. H., Bullock J. S., Governato F., Kuzio de Naray R., Peter A. H. G., 2015, *Proc. Natl. Acad. Sci.*, 112, 12249
- Zhang K. et al., 2019, *Sci. China Phys. Mech. Astron.*, 62, 959506 (Zh19)
- Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, *MNRAS*, 359, L30
- Zwaan M. A. et al., 2003, *AJ*, 125, 2842

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