The Population of Galaxies that Contribute to The HI Mass Function

Saili Dutta\(^1\), Nishikanta Khandai\(^1\), Biprateep Dey\(^{1,2}\)

\(^1\) School of Physical Sciences, National Institute of Science Education and Research, HBNI, Jatni 752050, India
\(^2\) Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

ABSTRACT

We look at the contribution of different galaxy populations to the atomic hydrogen (HI) mass function (HIMF) and the HI density parameter, \(\Omega_{\text{HI}}\), in the local Universe. Our analysis is based on a sample of 7857 HI-selected galaxies selected from a volume common to the SDSS and ALFALFA surveys (40\% catalog – \(\alpha\).40). We define different populations of galaxies in the color\((u-r)\)-magnitude\((M_u)\) plane and compute the HIMF for each of them. Additionally, we compute the HIMF for dark galaxies; these are undetected in SDSS and represent \(\sim 2\%\) of the total sample. We find that the bright red population dominates the total HIMF for \(\log_{10}(M_{\text{HI}}h_{70}^2/M_\odot) \geq 10.4\). The full red population – bright and faint – represents about \(\sim 17\%\) of the \(\Omega_{\text{HI}}\) budget, while that of the dark population is \(\sim 3\%\). The HIMF about the knee, \(\log_{10}(M_{\text{HI}}h_{70}^2/M_\odot) \in [8, 10.4]\), is dominated by the faint and bright blue populations, the latter dominating at larger masses in this interval. Their total contribution to \(\Omega_{\text{HI}}\) is \(\sim 55 - 70\%\), the variation depending on the definition of population. The dominant populations at the low mass end, \(\log_{10}(M_{\text{HI}}h_{70}^2/M_\odot) \leq 8.0\), are the faint blue and faint bluer populations, the latter’s dominance being sensitive to its definition. The full blue (blue–bluer bright and faint) population represents \(\sim 80\%\) of \(\Omega_{\text{HI}}\). A bimodal HIMF suggested by our results is however not seen since the amplitude of the HIMF of the bright red population is small compared to that of the bright blue population.

Key words: galaxies: formation – galaxies: evolution – galaxies: luminosity function, mass function – radio lines: galaxies – surveys

1 INTRODUCTION

The relationship between gas, metals, feedback and stars in galaxies is crucial for our understanding of galaxy formation and evolution. We need a clearer picture of how dark matter halos, which harbor galaxies, are supplied with cold gas, the fuel for star formation; how local and global conditions in halos, which harbor galaxies, are responsible in processing them into stars; and finally how the gas is polluted with metals and recycled back to the intergalactic medium due to feedback processes within the galaxy.

Three basic properties that describe galaxies are: (i) the star formation rate (SFR), (ii) the stellar mass \((M_{\text{star}})\) and (iii) the cold neutral hydrogen gas mass both in molecular \((M_{\text{HI}})\) and atomic \((M_{\text{HI}})\) phases. The amount of neutral gas tells us the amount of fuel currently available for future star formation. The SFR is the current rate of forming stars from the supply of cold gas that is available, while \(M_{\text{star}}\) depends on the integrated star formation history. Within each galaxy, at sub-kpc scales, the observed correlation between surface density of molecular hydrogen, \(\Sigma_{\text{HI}}\), and the SFR surface density, \(\Sigma_{\text{SFR}}\), is stronger (Bigiel, et al. 2008; Leroy, et al. 2008) as compared to the correlation between the HI surface density, \(\Sigma_{\text{HI}}\), and \(\Sigma_{\text{SFR}}\). The HI is often distributed beyond the optical radius of the galaxy and is more diffuse whereas star formation and \(H_2\) are mostly concentrated within the optical radius and occur in clumpier regions (Leroy, et al. 2008). However when the HI gas cools and becomes denser it transitions to molecular gas, which cools further and becomes denser leading eventually to gravitational collapse to form stars. It is therefore common to correlate the total gas \((\text{HI}+H_2)\) to the SFR, the so-called Kennicutt-Schmidt law (Schmidt 1959, 1963; Kennicutt 1998, 1989) for star formation, where \(\Sigma_{\text{SFR}} \propto \Sigma_{\text{gas}}^{1.4}\).

With the aid of multiwavelength observations in the optical, ultraviolet (UV) and infrared (IR) bands followed up by spectroscopic measurements, we are able to infer scaling relations between various properties of galaxies, derived and/or observed. These scaling relations are very important since any theoretical model of galaxy formation should be able to reproduce them. In practice, these relations are
used as parameters in theoretical models of galaxy formation when studied in cosmological volumes, since the physics on small, subparsec scales is never resolved (Springel & Hernquist 2003; Di Matteo, Springel & Hernquist 2005). A multi-pronged approach is used to study the distribution of HI in the post reionization Universe; these include cosmological hydrodynamical simulations of galaxy formation (Davé, et al. 2017), semi-analytical models (SAM) of galaxy formation (Kim, et al. 2017) and halo-occupation distribution (HOD) (Paul, Choudhury & Paranjape 2018). All of these approaches invoke in some form or other some of the observed scaling relations and model predictions are tested against observations which were not used as their input. Since these scaling relations depend on the observed sample they may be biased, i.e. the relations may depend on how the sample is chosen. E.g. the $M_{\text{HI}} - M_{\text{star}}$ relation differs if a sample is chosen by HI mass (Huang, et al. 2012) or stellar mass (Catinella, et al. 2010). On the other hand quantities like the luminosity function, mass function, correlation function (to name a few) are corrected for the survey selection completeness limit, whereas the cross-correlation method constrains $\Omega_{\text{HI}b_{\text{HI}}r}$. Here $b_{\text{HI}}$ and $r$ are the HI bias parameter and galaxy-HI cross correlation coefficient. In summary these HI surveys constrain $\Omega_{\text{HI}} = 4 \pm 1.6 \times 10^{-4}$ out to $z \approx 0.4$ (Rhee, et al. 2018).

Beyond $z = 0.4$ and out to $z \approx 5$ the HI content of the Universe is derived from studying damped Lyman-α systems (DLA) seen, in absorption, in the spectra of quasars (Prochaska, Herbert-Fort & Wolfe 2005; Noterdaeme, et al. 2012; Neelam, et al. 2016; Rao, et al. 2017). Combining both these approaches at low redshift (emission) and high redshift (absorption) one sees that $\Omega_{\text{HI}}$ increases monotonically as $(1 + z)^{0.56}$ from $\Omega_{\text{HI}}(z = 0) = 4 \times 10^{-4}$ to $\Omega_{\text{HI}}(z = 5) = 1.1 \times 10^{-3}$ (for a full compilation of all observational results see Rhee, et al. 2018). The HI density at $z = 0$ is only $1.5 \times$ smaller than at $z = 2$ with the data consistent with a no-evolution picture. On the other hand the cosmic SFRD has decreased 10 fold in this interval. Clearly our picture of galaxy formation is incomplete at these redshifts since the decrease in SFRD is not commensurate with a depletion of HI. Upcoming HI surveys like the Square Kilometre Array (SKA) will help us in this direction.

However even with the existing data it is important to understand the dependence of the abundance and distribution of HI selected galaxies on different galaxy properties and environments. The clustering of HI selected galaxies in the ALFALFA survey has been measured (Martin, et al. 2012; Papastergis, et al. 2013; Guo, et al. 2017) suggesting that ALFALFA galaxies cluster weakly. The dependence of HIMF has also been explored on the environment with Jones, et al. (2018) reporting a decrease in the low mass slope with increasing density in ALFALFA whereas Said, Kraan-Korteweg & Staveley-Smith (2019) find that the low mass slope increases with increasing density in the HI Zone of Avoidance survey with the Parkes telescope. Similarly the HI velocity width function has been studied for wall and void galaxies as well as red and blue galaxies (Moorman, et al. 2014), indicating a strong dependence on both environment and galaxy color. Zwaan, et al. (2003) have looked at the dependence of the HIMF on galaxy morphology as well as on the early-late type classification. In this paper we look at the dependence of the HIMF on the different populations classified from the color-magnitude plane. This is similar in spirit to the analysis of Zwaan, et al. (2003); Moorman, et al. (2014) carried out in HI surveys and also bears resemblance to similar analysis carried out for optical surveys where the contribution of the red and blue galaxies to the galaxy stellar mass function and luminosity functions are explored (Baldry, et al. 2004; Drory, et al. 2009).

\[^1\] http://cesam.lam.fr/zCosmos/
\[^2\] http://cesam.lam.fr/vvds/
\[^3\] http://deep.ps.uci.edu/
Our paper is organized as follows. In section 2 we describe a sample common to ALFALFA and SDSS (Sloan Digital Sky Survey), in section 2 we report our measurements of the HIMF for different populations within this sample, in section 4 we discuss our results and their implication and finally we summarize and conclude in section 5. We assume a flat $\Lambda$CDM cosmology with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ and a value for the dimensionless Hubble constant $h_{70} = 0.7$.

2 DATA

At 40% data release (Haynes, et al. 2011) ALFALFA has surveyed over $\sim$2752 deg$^2$ of the sky, with comoving volume of $\sim 2.65 \times 10^6$ Mpc$^3$. Covering 40% of the targeted survey area, the $\alpha$.40 catalog contains 15555 sources in the regions $7^h30^m < R.A. < 16^h30^m$, $4^\circ < \text{dec.} < 16^\circ$, and $24^\circ < \text{dec.} < 28^\circ$ and $22^h < R.A. < 3^h$, $14^\circ < \text{dec.} < 16^\circ$, and $24^\circ < \text{dec.} < 32^\circ$. Most of these objects have optical counterparts in the Sloan Digital Sky Survey (SDSS) Data Release 7 (DR7) (Abazajian, et al. 2009).

This catalog has been listed with the following observed quantities (i) an unique entry number from the Arecibo General Catalog (AGC); (ii) right ascension and declination of the HI source and the most probable optical counterpart (OC) from SDSS DR7; (iii) heliocentric velocity ($cz_{helio}$) which is the midpoint of the HI flux density profile; (iv) velocity profile width ($W_{50}$), measured as the full width at 50% of the peak HI flux density; (v) the integrated flux density of the HI source ($S_21$), The derived quantities, contained in this catalog, are - (i) distance to the object in Mpc (D). For sources with $cz_{helio} > 6000$ km s$^{-1}$ this quantity is $cz_{helio}/H_0$, where $cz_{helio}$ is the velocity in cosmic microwave background (cmb) reference frame and $H_0$ is the Hubble constant (taken to be 70 km s$^{-1}$ Mpc$^{-1}$). And sources with $cz_{helio} < 6000$ km s$^{-1}$ this quantity has been estimated using a local flow model (Masters 2005) (ii) the HI mass ($M_{HI}$), computed as

$$M_{HI}/M_\odot = 2.356 \times 10^5 \left( \frac{D}{\text{Mpc}} \right)^2 \left( \frac{S_{21}}{\text{Jy.km.s}^{-1}} \right)$$

The $\alpha$.40 catalog also provides another property which is the Code number, with a value of 1, 2 or 9. The sources with SNR > 6.5 are referred as Code 1 objects, Code 2 objects are the detections with SNR < 6.5 and Code 9 refers to the high velocity clouds (HVC).

For our analysis we have considered only Code 1 galaxies. The number of Code 1 galaxies is 11941. We also considered a cut for $cz_{helio} < 15000$ km s$^{-1}$ to avoid radio frequency interference (RFI) generated by Federal Aviation Administration (FAA) radar at the San Juan airport (Martin, et al. 2010; Haynes, et al. 2011). This restricts the sample to redshift, $z \leq 0.05$ and reduces the sample to 10785 galaxies.

Since ALFALFA is not totally overlapping with the SDSS footprint, we have defined a common boundary for both the surveys in this work in figure 1. For this work our final area of analysis has four subregions whose vertices right ascension-declination (RA,dec) are given as (i) (123.4, 245.4), (247.16), (115.16); (ii) (113.31, 247.5, 247.5, 247.5, 28), (113.31, 28); (iii) (-27.14), (34.14), (34.14), (17.16), (-6.16); and (iv) (-2.6, 24), (13.6, 24), (13.6, 26), (-2.6, 26). The total volume of these regions is $2.02 \times 10^6$ Mpc$^3$ and it corresponds to an angular area of $\sim 2093$ deg$^2$. In this region there are 8344 galaxies in ALFALFA. This includes removal of 4 OH high-redshift impostors (Suess et al. 2016).

In fig. 2 we look at the distribution of Code 1 objects in ALFALFA in the $S_{21}$ – $W_{50}$ plane. We also show the distributions in three mass bins which correspond to the small-mass end or faint-end (thin solid line), the knee (dot-dashed line) and the high-mass end (thin-solid line). As seen from these distributions on the average velocity width increases with increasing mass. This is expected and is also seen in the $M_{HI}$ – $W_{50}$ relation (see fig. 7 of Moorman, et al. 2014). One also sees that it is very unlikely to have a low(high) mass galaxy with a large(small) velocity width. However intermediate mass objects can have the full range of velocity widths.

The broken solid line in fig.2 is the sensitivity limit and is given by a 50% completeness relation of Code 1 objects in eq. (2) (Haynes, et al. 2011). This tells us that the detection of objects not only depends on the integrated flux, but also on the observed velocity width. At fixed $S_{21}$ the detection is more likely for narrower HI profile widths.

$$\log S_{21} = \begin{cases} 0.5 \log W_{50} - 1.207 : \log W_{50} < 2.5 \\ \log W_{50} - 2.457 : \log W_{50} \geq 2.5 \end{cases}$$

After applying this completeness cut we are left with a sample of 7857 galaxies. Among these, 6076 galaxies have spectroscopic as well as photometric measurements in SDSS DR7 and 1633 galaxies have only photometric measurements. As for the remaining 148, we loosely refer to them as dark galaxies. These objects are not being identified in

![Figure 1. Common footprint of SDSS and ALFALFA used in this work. The pale grey patches is that of SDSS DR7 overlapping with the ALFALFA survey region. The black dots are the positions of ALFALFA galaxies. The common boundary used in this work is outlined by the thick black line.](image)
Figure 2. The distribution of galaxies in the $S_21 - W_{50}$ plane is shown for all Code 1 ALFALFA galaxies. The broken solid line is the 50% completeness relation (Haynes, et al. 2011). It is given in eq. 2. We have further divided the sample into three mass bins: $\log_{10} (M_{\text{HI}}/M_\odot) \in [6.0, 8.5], [8.5, 10.2], [10.2, 11.0]$. The $1 - \sigma$ contours and the peaks of the distributions for these three populations are given by thick solid line (plus-circle), the dotted dash line (dot-circle) and the thin solid line (cross-circle).

The SDSS pipeline as potential galaxies. Although follow up observations have been made on some of the dark galaxies (Cannon, et al. 2015; Janowiecki, et al. 2015; Leisman, et al. 2017), in this work we refer to all of them as dark. This translates to around $\sim 2\%$ of galaxies which are dark or have no optical counterparts in SDSS.

The ALFALFA catalog also lists the SDSS object IDS of the OC. Using these we extracted the photometric properties like the $ugriz$ values (model magnitudes), then corrected for extinction (due to our own galaxy) for these 7709 (these are non-dark galaxies) galaxies. We also kcorrect (Blanton & Roweis 2007) the magnitudes to finally obtain rest frame magnitudes. For objects which do not have spectroscopic redshifts we have supplied the HI redshifts to kcorrect them. The kcorrect code also estimates additional properties like galaxy stellar mass, integrated star formation history and metallicity for these objects.

As of writing this paper the ALFALFA team has released the 100% catalog ($\alpha$100) (Haynes, et al. 2018) which also include the RA and dec of the optical counterparts. However we find that there are many galaxies which have bright foreground stars due to which SDSS has masked the region covering the galaxy and photometric values are not provided. We therefore restrict ourselves with $\alpha$40 catalog and will revisit the $\alpha$100 sample in the future.

2.1 Subsamples and Populations of Galaxies

From our sample of 7857 galaxies we identify subsamples which define different populations of galaxies. These populations are disjoint sets and their union (including the dark galaxies) forms the full sample. The populations are based on dividing the color($u-r$)-magnitude($M_r$) plane of the HI selected galaxies into six disjoint regions. This is shown in figure 3.

In figure 3 the upper solid curve demarcates the red (above curve) from the blue (below curve) population as in Baldry, et al. (2004). This optimal divider is given by

$$C'_{ur}(M_r) = 2.06 - 0.244 \tanh \left[ \frac{M_r + 20.07}{1.09} \right]$$

The vertical solid line divides the bright (leftward of line) and faint populations. The line has been chosen so that the fraction of bright red objects over all red objects is 0.87 which we refer to as the 1.5$\sigma$ line. Similarly the lower solid curve is chosen so that the fraction of blue (above curve) galaxies over all blue galaxies is 0.87 (or the 1.5$\sigma$ cut in color). This curve has been chosen to be parallel to the curve which demarcates the red and blue populations. We refer to the objects below(above) the curve as bluer(blue) galaxies.

Similarly we define the 1$\sigma$ sample (dashed lines). This breaks the sample of HI selected galaxies (which have optical counterparts) into 6 disjoint sets in the color-magnitude plane; we call each set a single population. The number for each population is quoted. The numbers in brackets are for the faint population is a subsample of galaxies with absolute magnitudes $M_r^f$; such that $M_r^f < M_r$ ($M_r^f \geq M_r$) for some reference value $M_r$ (the vertical line in figure 3).

Similarly we have defined three populations based on color (red, blue and bluer), based on equation 3. The six populations are formed by further splitting these three into bright and faint populations. As discussed earlier, each population is dependent on the boundaries that define them which we have called a 1$\sigma$ or 1.5$\sigma$ line. We caution the reader that our definition of sample may be confusing since both these samples represent the same set of 7709 galaxies which have optical counterparts. However the number of galaxies for any given population is different for the 1$\sigma$ and 1.5$\sigma$ samples.

We have chosen to define two samples to demonstrate that the qualitative results do not depend very strongly on sample definition. However as we will show in the next section, the definition of the faint bluer population (i.e. the 1.5$\sigma$ or 1$\sigma$ sample) determines whether it contributes significantly or not to some part of the HIMF. Looking at how the observed counts change, both in relative and absolute terms, when going from the 1.5$\sigma$ to the 1$\sigma$ sample we expect that the HIMF of the bright blue population to be the affected the most. In section 4 (figure 6) we will discuss the break in the $M_{\text{HI}} - M_{\text{star}}$, at $M_{\text{star}} \approx 9$, where $M_{\text{star}}$ is the logarithm of the galaxy stellar mass in units of $M_\odot/h_0^2$. One can then use the mean $M_r - M_{\text{star}}$ relation to convert $M_{\text{star}} = 9$ to $M_r = -19$. This value is sandwiched between the vertical lines (solid and dashed) in figure 3.
The HI Mass Function

3 RESULTS

The mass function $\phi(M)$ is defined as the number density of objects in the mass range $[M, M + dM]$. The HI mass function $\phi(M_{\text{HI}})$ can be expressed as

$$\phi(M_{\text{HI}}) = \frac{dN}{VdM_{\text{HI}}}$$

where, $dN$ is the total number of galaxies having HI mass between $M_{\text{HI}}$ and $M_{\text{HI}}+dM_{\text{HI}}$ and $V$ is the survey volume of interest. The HI mass function can be parameterized as a Schechter Function

$$\phi(M_{\text{HI}}) = \phi_* \left( \frac{M_{\text{HI}}}{M_*} \right)^{-\alpha} \exp \left( -\frac{M_{\text{HI}}}{M_*} \right)$$

Here, $\alpha$ is the faint end slope, $\phi_*$ is the amplitude and $M_*$ is the characteristic HI mass.

A simple and intuitive way of calculating the HIMF is by the $1/V_{\text{max}}$ method (Schmidt 1968). The underlying assumption in this method is that the sample of galaxies detected by a survey is a representative sample of galaxies in the Universe (in the same redshift range), or in other words it assumes homogeneity. For each detected galaxy ‘i’ a maximum detectable distance $D_{\text{max}}^i$ is calculated based on inverting eq. 1 by using the mass $M_{\text{HI}}^i$ of the galaxy and the limiting flux $S_{\text{lim}}^{i}$ which is the property of the survey. In the case of ALFALFA the completeness relation (eq. 2) determines this limiting flux for the velocity width $W_{50}^i$ of the galaxy.

$D_{\text{max}}^i$ is then converted to a volume $V_{\text{max}}^i$ which is the volume in which the galaxy could be in, and still be detected by the survey. Finally the galaxies are binned in mass with relative weights $1/V_{\text{max}}^i$ to obtain the mass function. For the brighter or more massive objects $V_{\text{max}}^i$ is larger than the survey volume. In such a case the relative weight is set to unity. This method has the advantage that it is non-parametric and does not require any prior knowledge to estimate the HIMF. However since galaxies cluster, the estimate of the HIMF will be sensitive to large scale structure in the local volume. Additionally some volumes of the survey may be inaccessible due to RFI and needs further correction. An estimate of these effects can be used to recalibrate the weights $V_{\text{max}}^i$ (Martin, et al. 2010).

Maximum likelihood (Sandage, Tammann, & Yahil 1979) and Step-Wise Maximum Likelihood (SWML) (Efstathiou, Ellis, & Peterson 1988) methods on the other hand are designed to be insensitive to large-scale structure. In the former the assumption is that the galaxy sample is drawn from an underlying distribution function, e.g. the HIMF for this work $\phi(M_{\text{HI}})$, and the likelihood method determines the parameters of this function. Although in most cases a Schechter function is the chosen function one has no way of testing whether it is the optimal function to describe data (Efstathiou, Ellis, & Peterson 1988). Baldry, et al. (2012) and Drory, et al. (2009) find that a single Schechter function does not describe the galaxy stellar mass function. In the latter, i.e. SWML, the underlying distribution $\phi(M_{\text{HI}})$ has no functional form but is discretized in steps or bins of mass and a uniform distribution is assumed in each bin. Hence the value $\phi_j$ which is the value of $\phi$ in the $j^{\text{th}}$ mass bin becomes the parameter. The joint likelihood of detecting all galaxies in the sample is maximized, with respect to the parameters $\phi_j$, thus determining their values. This method works when the sample is flux-limited.

When the selection function depends on other properties of the galaxies one needs to consider an underlying bivariate or multivariate distribution for $\phi$. One has to then generalize the SWML method to higher dimensions. Loveday (2000) estimated the bivariate luminosity function $\phi(M_{\text{B}}, M_{\text{K}})$ and then marginalized over $M_{\text{B}}$ to obtain the K-Band luminosity function, $\phi(M_{\text{K}})$, starting with a $b_{\text{j}}$-selected sample in the Stromlo-APM Redshift Survey.

For a blind HI survey like ALFALFA the limiting flux (figure 2 and eq. 2) depends on the velocity width $W_{50}$. A two-dimensional SWML (2DSWML) method similar to Loveday (2000) was applied by Zwaan, et al. (2003) to estimate the HIMF for HIPASS Galaxies. The bivariate distribution in this case is $\phi(M_{\text{HI}}, W_{50})$ which can be marginalized over $W_{50}$ to obtain the HIMF (Zwaan, et al. 2003; Martin, et al. 2010; Haynes, et al. 2011; Jones, et al. 2018) or marginalized over $M_{\text{HI}}$ to obtain the HI velocity width function (Zwaan, Meyer, & Staveley-Smith 2010; Moorman, et al. 2014). The details of our implementation are given in appendix A.

In figure 4 we show our estimate of the HIMF for all the populations including the total (crosses) and the dark (filled diamonds) populations. The columns are for the 1.5σ...
Figure 4. The HIMF for the 1.5σ (left column) and the 1σ (right column) samples. Data points and error bars were estimated using the 2DSWML method (see appendix A). The curves are Schechter function fits. For comparison we have added the total HIMF (crosses) and its fit (solid line) in all the panels. For each sample the top row is for the dominant population and the bottom for the sub-dominant one. In addition to the six populations we have also added the HIMF for the dark population (filled diamonds) and its Schechter function fit (dot-dot-dot-dashed line). The symbols for the six populations are the same as in figure 3. The Schechter fits are given for the red (dotted), blue (dot-dashed) and bluer (dashed) populations. The details of the fits are given in table 1.
the faint slope in figure 4 \((a')\) differs from \(a\) in equation 5 by 1, or \(a' = a + 1\).

Since we are working with Code 1 objects in the \(\alpha = 0.40\) sample it is appropriate to compare the total HIMF to that of Martin, et al. (2010). For the HIMF we find that our results of \((M_\star \pm \sigma_{M_\star}, \Phi_\star \pm \sigma_{\Phi_\star}, \alpha \pm \sigma_\alpha) = (9.96 \pm 0.02, 5.34 \pm 0.40, -1.35 \pm 0.02)\) and \(\chi^2_{\text{reduced}} = 0.79\) (table 1). This is consistent at the 1\(\sigma\) level with Martin, et al. (2010) who found \((M_\star \pm \sigma_{M_\star}, \Phi_\star \pm \sigma_{\Phi_\star}, \alpha \pm \sigma_\alpha) = (9.96 \pm 0.02, 4.8 \pm 0.30, -1.33 \pm 0.02)\). Note however that \(\Phi_\star\) is barely within 1\(\sigma\) of each other. We attribute this difference to the choice of sample in this work which has \(\sim 25\%\) fewer galaxies than Martin, et al. (2010). We point out that when we consider the full sample our results match well (see e.g. figure A3).

The goodness of fits, \(\chi^2_{\text{reduced}}\), is given in the last column of table 1. For the dark, faint red, faint bluer populations the \(\chi^2_{\text{reduced}}\) is of order unity. In the 1\(\sigma\) sample of the faint red population the biggest contribution of \(\chi^2_{\text{reduced}}\) comes from the lowest mass bin. If we remove that point as an outlier then \(\chi^2_{\text{reduced}} = 0.57\) and \(\alpha\) flattens to \(-1.06\) from \(-1.16\), \(\Phi_\star\) increases by about 22\% from 0.31. The change in \(M_\star\) is negligible. The bright red and bright bluer populations have a low \(\chi^2_{\text{reduced}}\) irrespective of sample definition, and looking at the data points relative to their fitted curve we find that there is little variation between them. This means that the error bars are larger than the variation between data and model. The bright blue and faint blue populations on the other hand have larger \(\chi^2_{\text{reduced}}\). For the 1.5\(\sigma\) sample of the faint blue population \(\chi^2_{\text{reduced}} = 4.10\) and improves to \(\chi^2_{\text{reduced}} = 0.96\) for the 1\(\sigma\) sample. The 1\(\sigma\) sample for the faint blue population has a larger number of brighter galaxies as compared to the 1.5\(\sigma\) sample (figure 3) and because \(M_{HI}\) is correlated to \(M_\star\), (see figure 5) the high mass end of the HIMF is better represented in the 1\(\sigma\) sample, leading to a smoother change in data and a better fit. This can also be seen in the top row of figure 4. The bright blue population has \(\chi^2_{\text{reduced}} \sim 2\) irrespective of sample definition.

We start by looking at the bright population. Due to the monotonic relation between \(M_{HI}\) and \(M_\star\) (figure 5), across populations, we expect the bright population to dominate the massive end of the HIMF and be subdominant at the low mass end. This is seen in figure 4. The characteristic mass, \(M_\star\), increases systematically from the bright bluer to the bright red population. There is little change in \(M_\star\) for both the bright red and bright blue populations with respect to sample definition. We also see little change in the HIMF with respect to sample definition for \(M_{HI} \geq 10.3\) for both these populations. The bright bluer population on the other hand has \(M_\star = 9.84\) for the 1.5\(\sigma\) sample and reduces to \(M_\star = 9.52\) for the 1\(\sigma\) sample. At the low mass end the bright populations have shallower slopes \(\alpha + 1 \geq 0\). This is expected as mentioned earlier since at the low mass end we expect the faint population to dominate. At this end, the sample definition affects the bright bluer population the most, where \(\alpha = -0.92\) (1.5\(\sigma\) sample) and increases to \(\alpha = +0.86\) (1\(\sigma\) sample), the change being the smallest for the bright red population. On the other hand the amplitude is most affected for the bright blue population, it changes from \(\phi_\star = 4.85\) (1.5\(\sigma\) sample) to \(\phi_\star = 2.54\) (1\(\sigma\) sample). The change is negligible for the bright red and about \(\sim 72\%\) for the bright bluer population. We point out that changes in the HIMF with respect to sample definition can be best understood in terms of how the observed number of galaxy populations change when the boundaries in the color-magnitude plane are redrawn to define a new sample (figures 3) and the average scaling relation \(M_\star - M_{HI}\) (figure 5). For the 1.5\(\sigma\) sample the bright blue is the dominant population at the knee of the HIMF, however it contributes nearly equally with the faint blue population when we consider the 1\(\sigma\) sample. This is because the faint blue population has a net increase in observed galaxies from \(n_{gal} = 1409\) (1.5\(\sigma\) sample) to \(n_{gal} = 2422\) (1\(\sigma\) sample), the net change coming from the intersection of bright blue (1.5\(\sigma\)) and faint blue (1\(\sigma\)) (see figure 3); this is the primary reason for the increase of \(\phi_\star\) by \(\sim 4.8\times\). An interesting result is that the bright red population is the dominant population at \(M_{HI} \geq 10.4\). This result is insensitive to sample definition. We find that for \(M_{HI} \geq 10.4\) the bright red population represents \(\sim 60\%\) of total detections with \(\sim 40\%\) coming from the bright blue population which also translates to similar fractions in total HI mass at this end.

We now move to the faint population. All of them have steeper slopes as compared to their bright counterparts and do not dominate the HIMF at the high mass end. The faint red population is the most subdominant population. The observed counts of galaxies of the faint bluer population see the largest relative change with sample definition increasing from \(n_{gal} = 452\) (1.5\(\sigma\) sample) to \(n_{gal} = 1369\) (1\(\sigma\) sample), a factor \(\sim 3\times\). This results in a small change in slope from \(\alpha = -1.76\) to \(\alpha = -1.62\) but a large, \(\sim 10\times\), increase in amplitude from \(\phi_\star = 0.10\) to \(\phi_\star = 1.09\). In the 1.5\(\sigma\) sample the faint bluer population is a subdominant population, however it becomes the dominant population for masses \(M_{HI} \leq 8.1\). The faint blue population is on the other hand a dominant population below the knee of the mass function. It dominates the HIMF for \(M_{HI} \leq 8.7\) in the 1.5\(\sigma\) sample and for \(8.1 \leq M_{HI} \leq 9.7\) in the 1\(\sigma\) sample.
region & \log(M_*/M_\odot) + 2 \log h_{70} & \phi_* (10^{-3}h_{70}^3\text{M}pc^{-3}\text{dex}^{-1}) & \alpha & \chi^2_{\text{reduced}} \\
\hline
\text{total} & 9.96 \pm 0.02 & 5.34 \pm 0.40 & -1.35 \pm 0.02 & 0.79 \\
bright blue & 9.86 \pm 0.02 (9.85 \pm 0.03) & 4.85 \pm 0.42 (2.54 \pm 0.17) & -1.03 \pm 0.06 (-0.43 \pm 0.11) & 2.02 (2.17) \\
faint blue & 9.57 \pm 0.04 (9.50 \pm 0.02) & 1.22 \pm 0.22 (5.84 \pm 0.58) & -1.74 \pm 0.05 (-1.22 \pm 0.05) & 4.10 (0.96) \\
bright red & 10.04 \pm 0.04 (10.02 \pm 0.04) & 0.96 \pm 0.12 (0.95 \pm 0.12) & -1.01 \pm 0.08 (-0.86 \pm 0.10) & 0.32 (0.52) \\
bright bluer & 9.84 \pm 0.08 (9.52 \pm 0.05) & 0.46 \pm 0.13 (0.79 \pm 0.16) & -0.92 \pm 0.23 (0.86 \pm 0.38) & 0.64 (0.38) \\
faint bluer & 9.83 \pm 0.07 (9.72 \pm 0.04) & 0.10 \pm 0.03 (1.09 \pm 0.18) & -1.76 \pm 0.06 (-1.62 \pm 0.05) & 1.11 (1.40) \\
faint red & 9.89 \pm 0.07 (9.74 \pm 0.06) & [6.09 \pm 2.07] \times 10^{-2} (0.31 \pm 0.07) & -1.66 \pm 0.11 (-1.16 \pm 0.08) & 1.27 (1.36) \\
dark & 10.03 \pm 0.09 & [3.25 \pm 1.50] \times 10^{-2} & -1.92 \pm 0.09 & 1.17 \\
\hline
\end{tabular}

**Table 1.** Parameters of the Schechter function fit to the HIMF for all the populations. The estimated parameters and their uncertainties are for the 1.5σ sample and the numbers in brackets are for the 1σ sample. The goodness of fit, $\chi^2_{\text{reduced}}$, is given in the last column.

The dark population is characterized by a very steep slope $\alpha = -1.92$, large characteristic mass $M_* = 10.03$ and a very small amplitude $\phi_* = 3.25 \times 10^{-2}$ and is a subdominant population. Extrapolating to masses below $M_{\text{HI}} \lesssim 6$ our results suggest the dark population will be the dominant population. However it is unclear how far down we can extrapolate since it is unlikely that there will be too many low mass, gas rich galaxies devoid of stars which will be able to self shield themselves from the photoionizing background.

### 3.1 The contribution of different galaxy populations to $\Omega_{\text{HI}}$

We can analytically integrate the HIMF, fitted to a Schechter function to obtain the cosmic HI density parameter,

$$\Omega_{\text{HI}} = \frac{\rho_{\text{HI}}}{\rho_c} = \frac{M_* \phi_*}{\rho_c} \Gamma(\alpha + 2)$$

Alternately we can sum the binned measurements of the HIMF. Similar to Haynes, et al. (2011) we find that both methods give similar results, with the exception of the dark sample which has a very steep slope. As we argued in the previous section it is not physical to extrapolate the HIMF to very small masses. Hence we choose to quote our results by the *summed* method. We note that the results do not change if we integrate the Schechter function from $M_{\text{HI}} = 6.1$ to $\infty$. We summarize our results in table 2. Column 2 is the estimate of $\Omega_{\text{HI}}$ from each population and column 3 is the fractional contribution to $\Omega_{\text{HI}}^\text{tot}$, the values in brackets are for the 1σ sample.

For the total sample we obtain $\Omega_{\text{HI}} = (4.894 \pm 0.469) \times 10^{-4} h_{70}^{-1}$ which is consistent at 1σ with $\Omega_{\text{HI}} = (4.4 \pm 0.1) \times 10^{-4} h_{70}^{-1}$ (summed) of Martin, et al. (2010). With the addition of Code 2 objects in the $\alpha=0.40$ sample we see that it is only consistent at the 2σ level with Haynes, et al. (2011) who obtain $\Omega_{\text{HI}} = (4.2 \pm 0.1) \times 10^{-4} h_{70}^{-1}$ (summed). Since our $M_*$ is comparable with the $\alpha=0.40$ results, and $\alpha$ is only a bit steeper, the main reason for this discrepancy can be traced to $\phi_*$ (equation 6). Our value of $\phi_*$ is $\sim 10%$ higher than Martin, et al. (2010) which translates to a 10% higher estimate of $\Omega_{\text{HI}}$ at fixed $M_*$ and $\alpha$. However the relative ratios should not be sensitive to this change.

From table 2 we see that the red population (bright and faint) have a non-negligible contribution of $\sim 17%$ to $\Omega_{\text{HI}}^\text{tot}$. When combined with the dark population (3% of $\Omega_{\text{HI}}^\text{tot}$), this adds up to a non-negligible fraction of 20% (rounded). The full blue population (faint and bright blue and bluer) then contributes $\sim 80%$ (rounded) of $\Omega_{\text{HI}}^\text{tot}$. We will discuss the implications of these numbers in the next section. The dominant sample about the knee of the HIMF are the bright blue and the faint blue populations (as $M_*$ galaxies) and together they contribute most to $\Omega_{\text{HI}}$, $\sim 73%$ (1.5σ sample) and $\sim 55%$ (1σ sample).

### 4 DISCUSSION

Since we are looking at the contribution of different galaxy populations to the total HIMF we would like to see whether these populations have different scaling relations, e.g. in the $M_{\text{HI}} - M_{\text{star}}$ plane. Such relations have been explored for galaxies in the ALFALFA sample (Catinella, et al. 2010; Huang, et al. 2012; Maddox, et al. 2015) and the HI Parkes
All-Sky Survey Catalog (HICAT) (Parkash, et al. 2018). In this work the stellar masses are estimated by kcorrect which uses the population synthesis code of Bruzual & Charlot (2003). Our estimates on $M_{\text{star}}$ differ from the estimates of these authors. The $M_{\text{HI}}-M_{\text{star}}$ scaling relations are shown in figure 6.

In order to avoid crowding figure 6 we choose to compare our results for the total sample with Huang, et al. (2012) (crossed-circle, thin solid line) and the filled star (thick solid line) represents the total sample excluding the dark galaxies. The results are compared with Huang, et al. (2012) (crossed-circle and thin solid line), other data points and line styles are the same as in figure 4.

There is a clear transition in the scaling relations when going from the low mass to the high mass end. $f_{\text{HI}}$ gets suppressed for the total sample at about $M_{\text{star}} \sim 9$ consistent with Huang, et al. (2012); Maddox, et al. (2015). The transition scale also depends on the galaxy population. For the blue, bluer and red populations it occurs at $M_{\text{star}} = 10.1, 9.4, 8$ respectively. The transition scale of $M_{\text{star}} \sim 9$ corresponds to a change in the dominant morphology of galaxy populations (Maddox, et al. 2015) and also a transition between hot and cold mode accretion seen in cosmological hydrodynamical simulations (Kereš, et al. 2009).

One interesting result that we have quantified in the last section is the non-negligible HI content of red galaxies. The red galaxies dominate the HIMF at the high mass end $M_{\text{HI}} \geq 10.4$ and $\sim 17\%$ of the HI content, $\Omega_{\text{HI}}$, is locked up in them. Using the HOD framework for HI, Paul, Choudhury & Paranjape (2018) also find that the red population is the dominant population at higher masses. Since the ALFALFA sample is an HI selected sample, with the majority of the galaxies belonging to the blue cloud, one may ask: why do the rarest, gas rich galaxies, predominantly belong to the red cloud? Looking at the morphology of these gas rich red galaxies we find that these are predominantly spirals and lenticular galaxies, but there also exist some elliptical galaxies. A number of spirals have prominent bulges which would classify them as early-type spirals and there are a number of galaxies which harbor dust streaks visible on their disk plane. A significant number of galaxies are edge-on or somewhere in between edge-on and face-on. Indeed HI has been detected in early type galaxies ( Morganti, et al. 2006; Oosterloo, et al. 2007; Grossi, et al. 2009), but these do not go beyond $M_{\text{HI}} \sim 10$. Schiminovich, et al. (2010) find that the 47% of the total local SFR density is found for $M_{\text{star}} > 10$ in the GASS sample. Given that there is little correlation between $(u-r)$ color and $M_{\text{star}}$ at these masses and the fact that GASS (and ALFALFA) detections are predominantly in the blue cloud (Catinella, et al. 2010) we would expect these red gas rich galaxies to contribute a negligible fraction to the local SFR density. This does not mean that individually all bright red, gas rich galaxies have low star formation rates but, rather, their numbers are so small that their total contribution is negligible. A fraction of these red galaxies would then be the ones transitioning from a blue star-forming phase to a red passive phase with little star formation and another fraction will be dusty star forming galaxies, while the rest would be passively evolving. The amount of reddening would be enhanced if they are edge-on and would redden the color of disky galaxies which are either transitioning to the red phase or contain considerable dust on their disks (Graham & Worley 2008; Tempel, et al. 2011).

Although the bright red galaxies are the dominant population at the high mass end ($M_{\text{HI}} \geq 10.4$) the observed counts (60% of total) are only 50% more than those of the bright blue (40% of total) population. When plotted on a logarithmic scale the differences between these two mass functions are not very large (see top row of figure 4). If the inclination and reddening are important and the magnitudes are corrected for them, then a fraction of gas rich red galaxies would move to the bright blue population and would bring the mass functions of these two populations closer to each other at the high mass end. Conversely if the reddening is increased due to inclination we would expect that the transitions in the scaling relations when going from the low mass to the high mass end. $f_{\text{HI}}$ gets

![Figure 6. The $M_{\text{star}}-M_{\text{HI}}$ relation for the 1.5d sample. The bright red population as well as the faint blue and bluer population correspond to a change in the dominant morphology of galaxy populations (Maddox, et al. 2015) and also a transition between hot and cold mode accretion seen in cosmological hydrodynamical simulations (Kereš, et al. 2009).](image-url)
effects would be more dramatic in HI velocity width function. This is indeed seen for ALFALFA galaxies (Moorman, et al. 2014), where the HI velocity width function for the red and blue are well segregated at the high velocity end. We therefore believe that reddening due to dust and inclination can partially explain why the red sample is the dominant population at the high mass end of the HIMF.

The results of section 3 are essentially conditional HIMF integrated over a range in color and magnitude, which we have called the HIMF for different populations. We can repeat the exercise and compute the HIMF in finer intervals of color and magnitude to obtain a conditional HIMF (conditioned on luminosity and color). This will then tell us about the distribution of HI in the color-magnitude plane. As an application one can then make better estimates of the corrections applied to $\Omega_{\text{HI}}$ with the stacking methods (Rhee, et al. 2013, 2016, 2018) at higher redshifts. A second application would be to inform a proposed HI survey, which galaxy populations to stack on to make a tentative detection.

In the survey volume considered in this work we find that only 12% of the red population in SDSS have HI detections in ALFALFA. This number is 38% for blue galaxies. On the other hand 98% of ALFALFA galaxies have optical counterparts. The detections are due to a combination of total HI signal and observed HI velocity widths. The question we wish to ask is are the non-detections in the red cloud due to insufficient HI gas or due to large velocity widths or both? We argue that in the bright red sample the non-detections are due to insufficient HI gas as well as large widths and HI masses should decrease with either increasing stellar mass or halo mass. Although figure 6 suggests that the most massive galaxies (large stellar mass, $M_{\text{star}}$) are also the richest in terms of their gas content, this relation is biased since it is from an HI selected sample. The appropriate sample is the GASS sample which is selected on stellar mass. We look at the $M_{\text{HI}} - M_{\text{star}}$ relation in the final data release of GASS (DR5), which is summarized on table 1 of Catinella, et al. (2013). The relation has nearly a flat slope for $M_{\text{star}} \in [10.76, 11.30]$, being slightly positive if all non-detections have been assigned the limiting HI mass or halo mass. Although figure 6 suggests that the most massive galaxies (large stellar mass, $M_{\text{star}}$) are also the richest in terms of their gas content, this relation is biased since it is from an HI selected sample. The appropriate sample is the GASS sample which is selected on stellar mass. We look at the $M_{\text{HI}} - M_{\text{star}}$ relation in the final data release of GASS (DR5), which is summarized on table 1 of Catinella, et al. (2013). The relation has nearly a flat slope for $M_{\text{star}} \in [10.76, 11.30]$, being slightly positive if all non-detections have been assigned the limiting HI mass and slightly negative if all non-detections are assigned zero HI mass. The last bin however has only a handful of objects which are dominated by non-detections. However the GASS and ALFALFA catalogs are relatively shallow as compared to optical surveys like the SDSS and would miss a significant number of massive galaxies. The tail of the stellar mass function is dominated by the red galaxies and at $M_{\text{star}} = 11.3$ the number of red galaxies is $\sim 10 \times$ the blue galaxies (Baldry, et al. 2012). These galaxies are probably central red galaxies (Dror, et al. 2009) and would be in halos of $\log_{10}(M_{\text{halo}}/M_\odot) \simeq 14 - 14.5$ (Behroozi, Conroy & Wechsler 2010) with virial temperatures $T_{\text{vir}} \sim 10^7 K$ and circular velocities $V_{\text{circ}} \sim 6 - 9 \times 10^3 \text{km.s}^{-1}$. Most of these galaxies will then be the central galaxies of large groups of galaxies or clusters of galaxies. Given the large virial temperatures and large circular velocities it would be very unlikely to detect a considerable amount of neutral hydrogen in these systems. It is then very likely that the $M_{\text{HI}} - M_{\text{star}}$ relation will not asymptote to a constant as indicated in Catinella, et al. (2013) but rather decrease with increasing stellar mass. This is suggested in the results of Kim, et al. (2017).

If the average $M_{\text{HI}} - M_{\text{star}}$ becomes a non-monotonic function of stellar mass and therefore halo mass, HI abundance matching techniques, used to obtain $M_{\text{HI}} - M_{\text{halo}}$ relation (Khandai, et al. 2011; Padmanabhan & Kulkarni 2017), will break down. The HI HOD models which also assume a step-like function (with the help of the error function) (Guo, et al. 2017; Paul, Choudhury & Paranjape 2018) for the average occupation of centrals, may need to be revised. A log normal form for the mean occupation function for centrals was compared to the step-like parameterization in the context of describing quasar clustering (Shen, et al. 2013), but it was found that the HOD parameters were not well constrained. Only more direct observations will shed light on the HI content of these massive galaxies and hopefully provide better inputs for the HOD parameterization.

5 SUMMARY

In this work we have measured the HIMF of different galaxy populations picked from the color-magnitude plane. The galaxies considered were from a local volume common to ALFALFA and the SDSS surveys in the redshift range $z \in [0, 0.05]$. After putting the relevant cuts in quality of detection, volume and completeness the final sample analyzed consists of 7857 galaxies. We divided the total sample first into bright and faint populations (cut in magnitude) and these were further split into three colors: red, blue and bluer. This forms a disjoint set of six populations. A seventh population which we call a dark population is one which does not have any optical counterparts in SDSS but has a detection in ALFALFA. The union of these seven populations is the total sample of 7857 galaxies. We have further considered a second sample which redefines the six populations by shifting the boundaries defining them. We have called them the $1\sigma$ and $1.5\sigma$ samples (section 2.1). The reason for doing this is to illustrate that our sample definition does not change the qualitative results that we report. The reason for splitting the full blue cloud into four (faint/bright for blue/bluer) instead of just two (faint/bright for all blue) was because we wanted to see the systematic effect of the tail of the blue population (especially the faint end) on our results. We indeed find that based on how we define our sample the faint bluer population becomes the dominant population at the low mass end of the HIMF (figure 4 and tables 1&2).

We summarize our results below:

- At fixed color (e.g. red, blue or bluer) the HIMF of the bright population dominates over their corresponding faint counterparts at the knee and the large mass end whereas the faint populations dominate at the low mass end.
- At fixed magnitude (e.g. bright or faint) there is no systematic trend at the low mass end and the knee of the HIMF, with decreasing color, for the faint population. However for the bright population, we see that the HIMF of the high mass end increases with increasing color. The bright red population is the dominant population at this end.
- The bright red sample dominates the total HIMF at $M_{\text{HI}} \geq 10.4$. When combined with the faint red sample it contributes about $\sim 16 - 17\%$ of the $\Omega_{\text{HI}}$ budget. The dark population contributes $\sim 3\%$ to $\Omega_{\text{HI}}$. This has implications for detections done with stacking at higher redshifts which would target the blue star forming cloud for a first detection.
The total blue cloud (blue and bluer) represents about ~80% of the $\Omega_{\text{HI}}$ budget.

In the mass range about the knee, $M_{\text{HI}} \in [8, 10.4]$, the dominant populations are the faint and bright blue populations with the latter dominating at larger masses in this interval. Their total contribution to $\Omega_{\text{HI}}$ is ~55 – 70% depending on sample definitions.

The dominant populations contributing to the low mass end of the HIMF the faint blue and faint bluer populations, the latter being dominant only for the $1\sigma$ sample definition.

The bright bluer and faint red populations are subdominant populations contributing a total of ~6 – 10% to $\Omega_{\text{HI}}$.

It would be interesting to see if simulations (Davé, et al. 2017), SAMs (Kim, et al. 2017) and HOD models are able to reproduce our results which provide additional constraints on the population of HI selected galaxies. In section 4 we argued that the effect of dust and inclination are responsible for the bright red population dominating the total HIMF at large masses. This was based on the results of Catinella, et al. (2013); Moorman, et al. (2014) but needs to be confirmed with a more detailed followup. In a forthcoming paper we will report on a more detailed analysis of the HI velocity width function. We are also working on measuring the conditional (conditioned on color or magnitude or both) HIMF which will put additional constraints on the properties of gas bearing galaxies.

ACKNOWLEDGMENTS

We would like to thank R. Srianand and A. Paranjape for the many useful discussions that we had throughout this work. SD would like to thank R. Srianand for giving the hands-on training on obtaining and processing SDSS data. We would like to acknowledge discussions with J. S. Bagla and N. Kanekar. The research of NK is supported by the Ramanujan Fellowship awarded by the Department of Science and Technology, Government of India. All the analysis were done on the xanadu server funded by the Ramanujan Fellowship.

We would like to acknowledge the work of the entire ALFALFA collaboration in observing, flagging, and extracting the properties of galaxies that this paper makes use of. This work also uses data from SDSS DR7. Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Website is http://www.sdss.org/. The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions.

REFERENCES

Abazajian, K. N., et al. 2009, ApJS, 182, 543
Baldry, I. K., et al. 2004, ApJ, 600, 681
Baldry et al., 2012, MNRAS, 421, 621
Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Bigiel F., et al., 2008, AJ, 136, 2846
Blanton M. R., Roweis S., 2007, AJ, 133, 734
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Cannon, J.M., et al., 2015, Astro J 149, 72
Casey C. M., Narayanan D., Cooray A., 2014, PhR, 541, 45
Catinella B., et al., 2010, MNRAS, 403, 683
Catinella B., et al., 2013, MNRAS, 436, 34
Chang T.-C., Pen U.-L., Bandura K., Peterson J. B., 2010, Natur, 466, 463
Davé R., Rafieferantsoa M. H., Thompson R. J., Hopkins P. F., 2017, MNRAS, 467, 115
Davis M., Huchra J., 1982, ApJ, 254, 437
Di Matteo T., Springel V., Hernquist L., 2005, Nature, 433, 604
Drory N., et al., 2009, ApJ, 707, 1595
Efstathiou G., Ellis R. S., Peterson B. A., 1988, MNRAS, 232, 431
Giovanelli, R., et al., 2005, AJ, 130, 2598
Graham A. W., Worley C. C., 2008, MNRAS, 388, 1708
Grossi M., et al., 2009, A&A, 498, 407
Guo, H., et al., 2017, AJ 846, Issue 1, article id. 61, pp. 23
Haynes M. P., et al., 2011, AJ, 142, 170
Haynes M. P., et al., 2018 ApJ 861, 49
Huang S., Haynes M. P., Giovanelli R., Brinchmann J., 2012, ApJ, 756, 113
Janowiecki, S., et al., 2015, ApJ, 801, 96
Jones, M.G., Haynes, M.P., Giovanelli, R. & Moorman, C. 2018 MNRAS 477, 2
Kanekar N., Sethi S., Dwarkanath K. S., 2016, ApJL, 818, L28
Kennicutt R. C., 1989, ApJ, 344, 685
Kennicutt R. C., 1998, ApJ, 498, 541
Kereš D., Katz N., Fardal M., Davé R., Weinberg D. H., 2009, MNRAS, 395, 160
Khandai N., et al., 2011, MNRAS, 415, 2580
Kim H.-S. et al., 2017, MNRAS, 465, 111
Lah P., et al., 2007, MNRAS, 376, 1357
Lah P., et al., 2009, MNRAS, 399, 1447
Le Floc’h E., et al., 2005, ApJ, 632, 169
Leisman, L., et al., 2017, Ap J 842, 133
Leroy A. K. et al., 2008, AJ, 136, 2782
Loveday, J., 2000, MNRAS,312, 45
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
Martin, A., et al., 2010, ApJ, 723,1359
Martin A. M., Giovanelli R., Haynes M. P., Guzzo L., 2010, ApJ, 730, 38
Masters, K.L. 2005, PhD Thesis, Cornell University
Masui K. W., et al., 2013, ApJL, 763, L20
Meyer M. J., et al., 2004, MNRAS, 350, 1195
Moorman C. M., et al., 2014, MNRAS, 444, 3559
Morganti R., et al., 2006, MNRAS, 371, 157
Neelamman, et al., et al., 2016, ApJ, 818, 113
Noterdaeme P., et al., 2012, A&A, 547, L1
Oosterloo T. A., Morganti R., Sadler E. M., van der Hulst T., Serra P., 2007, A&A, 465, 787
Padmanabhan H., Kulkarni G., 2017, MNRAS, 470, 340
Papastergis E., Giovanelli R., Haynes M. P., Rodriguez-Puebla A., Jones M. G., 2013, ApJ, 776, 43
APPENDIX A: THE 2DSWML METHOD TO CALCULATE THE HI MASS FUNCTION

We now proceed in describing our implementation of the 2DSWML method (Zwaan, et al. 2003; Martin, et al. 2010). Assuming a bivariate distribution $\phi(M_{\text{HI}}, W_{50})$ the probability of detecting a galaxy ‘i’ with HI mass $M_{\text{HI}}$ and profile width $W_{50}$ at distance $D$ is

$$p_i = \int_{W_{50}=0}^{\infty} \int_{M_{\text{HI}}=M_{\text{HI},i,\text{min}}(D),W_{50}}^{\infty} \phi(M_{\text{HI}}, W_{50}) dM_{\text{HI}} dW_{50}$$

(A1)

We now need to discretize the above equation for the 2DSWML method. We will begin by considering the distribution of galaxies in bins of $M = \log_{10} \left[ M_{\text{HI}} / M_\odot \right]$ and $W = \log_{10} \left[ W_{50} / (\text{km} \cdot \text{s}^{-1}) \right]$. The number of bins are $N_M$ & $N_W$ and the bin widths are $\Delta M$ & $\Delta W$ respectively. Therefore the two dimensional distribution can be parameterized by $\phi_{jk}$, where $j = 0, 1, 2, ..., N_M - 1$ and $k = 0, 1, 2, ..., N_W - 1$. In this analysis we have taken 10 bins per dex in velocity width and 5 bins per dex in mass. Eq A1 for the 2DSWML is now

$$p_i = \frac{\sum_j \sum_k V_{ijk} \phi_{jk}}{\sum_j \sum_k H_{ijk} \phi_{jk} \Delta M \Delta W}$$

(A2)

Here $V_{ijk}$ ensures that galaxy ‘i’ is only populated in its corresponding ‘j-k’ bin.

$$V_{ijk} = \begin{cases} 1, & \text{if galaxy } i \text{ is a member of mass bin } j \\ 0, & \text{otherwise.} \end{cases} \quad \text{(A3)}$$

$H_{ijk}$ is a weight corresponding to galaxy ‘i’ in the ‘j-k’ bin and takes values from 0 to 1. It appears so that the integral in the denominator of eq A1 can be done in the entire $M$-$W$ plane after convolving with the completeness function $C^i$ in the $M$-$W$ plane. An example for one of the galaxies in our sample is shown in figure A1. The solid broken line is the completeness relation for this object in the $M$-$W$ plane.

The shaded (white) area is the area accessible (inaccessible) to this object in the $M$-$W$ plane.

$$\int_{W=W_0}^{W=W_{N_W-1}} \int_{M=M_{M=0}}^{M=M_{N_M-1}} C^i(M, W) dM dW$$

(A4)

Computing $H_{ijk}$ is a straightforward exercise and we have shown it as diagrams in table A1. The shaded area (column 2) in each case is the j-k bin of interest for which $H_{ijk}$ has to be computed. In cases 1 and 2, the values of $H_{ijk}$ are either 1 or 0. This is because the completeness line does not intersect the square and lies below or above that particular bin of interest. In cases 3-5 the square of interest intersects with the completeness curve and the shaded area is the value that $H_{ijk}$ takes (column 3) which are fractional values. The points of intersection are denoted as $W_{\text{lim}}$, $M_{\text{lim}}$ if it does not exactly intersect on the bin edges $W_k$, $M_j$. Case 6 is introduced and assumes a completeness slope greater than 1, which is not the case for ALFALFA. We also note
that the completeness relation eq. 2 has a change in slope at \( \log_{10} \left[ W_{\text{lim}} / (\text{km}\cdot\text{s}^{-1}) \right] = 2.5 \), which coincides with the edge of the bin in \( W_{\text{lim}} \) for our choice. Table A1 assumes this so that no further cases are considered.

In the 2DSWML one wishes to find \( \phi_{jk} \) that maximizes the joint probability or likelihood of finding all the galaxies in the sample simultaneously

\[
\mathcal{L} = \prod_{i=1}^{N_g} p_i
\]  

(A5)

Using eq. A1 the joint likelihood is

\[
\mathcal{L} = \prod_{i=1}^{N_g} \frac{\sum_j \Sigma_k V_{ijk} \phi_{jk}}{H_{ijk} \phi_{jk} \Delta m \Delta w}
\]

(A6)

To obtain \( \phi_{jk} \), we maximize rather the log-likelihood

\[
\ln \mathcal{L} = \sum_{i=1}^{N_g} \ln \left( \frac{\sum_j \Sigma_k V_{ijk} \phi_{jk}}{H_{ijk} \phi_{jk} \Delta m \Delta w} \right)
\]

(A7)

This gives us

\[
\phi_{jk} = \left[ \Sigma_j V_{ijk} \right] \left[ \frac{H_{ijk}}{\Sigma_m \Sigma_n H_{imn} \phi_{mn}} \right]^{-1}
\]

\[
= \frac{n_{jk}}{\Sigma_m \Sigma_n H_{imn} \phi_{mn}}
\]

(A8)

where, \( n_{jk} = \sum_j V_{ijk} \) is the number of galaxies in mass bin \( j \) and profile width bin \( k \). \( \phi_{jk} \) are iteratively determined from eq. A8. To start the iteration we set the initial value of on the RHS of eq. A8 to be \( \phi_{mn} = n_{jk} / [V_{\text{surv}} \Delta M \Delta W] \). We set a relative tolerance for convergence of \( 10^{-3} \) for all \( \phi_{jk} \). We find that \( \phi_{jk} \) converges within 20 iterations. Finally the HIMF is obtained by summing over the velocity width bins \( k \).

\[
\phi = \sum_k \phi_{jk} \Delta w
\]

(A9)

### A1 Normalization of HIMF

One drawback for likelihood methods, as opposed to the \( 1/V_{\text{max}} \) method, is that the normalization of the HIMF is not fixed. This is obvious from eq. A1. There are a number of ways to fix the normalization (Davis & Huchra 1982; Willmer 1997; Zwaan, et al. 2003; Martin, et al. 2010) which involve computing the selection function. Here we try a slightly different method. We assume that the high mass end of the HIMF is not affected by the selection function. One can test this assumption by comparing the ratio of normalized mass function from 2DSWML to that of the observed mass function which is related to the observed counts. In the top panel of figure A2 the observed HIMF is shown as filled pentagons (solid line) and the un-normalized and converged HIMF from the 2DSWML is shown as filled triangles (dashed line). This is done for the \( \alpha = 0.10 \) sample (Haynes, et al. 2018). The ratio of these is shown for the last 7 mass bins in the lower panel of figure A2. If the selection function affects the high mass end, the ratio at this end would not have a flat feature. We indeed find that in this particular example of \( \alpha = 0.10 \) the last 3 mass bins are unaffected at the level ranging from \( 0-0.003 \% \); whereas the last but third bin is relatively suppressed by around 0.4\%. For this method to work

---

**Table A1.** Values of \( H_{ijk} \) (column 3) are shown diagrammatically (column 2). The shaded region corresponds to the \( j,k \) bin of interest. Cases 1 and 2 (first two rows) take values 1 and 0, the completeness curve (line) lies below or above the square and never intersects it. Cases 3-6 give fractional values of \( H_{ijk} \) since the \( j,k \) bin of interest intersects the completeness curve. The points of intersection are denoted by \( W_{\text{lim}}^j, M_{\text{lim}}^k \) if it does not exactly intersect on the bin edges \( W_k, M_j \).
we need to test the flatness of this ratio which means that at least the last two points at the high mass end should be unaffected by the selection function. Starting from the high mass end we search sequentially at smaller bins which are unaffected by the selection function. The search is stopped when the selection function affects the particular bin at the level of 0.1% or greater. Finally we fix the normalization by matching the integrated counts over these bins to that of the observed HIMF.

We have compared this method to the one which normalizes the mass function to match the average observed counts as in Davis & Huchra (1982); Martin, et al. (2010). We find that they match at the relative level of ~ 0.4%. Finally we compare in figure A3 our result with that of α.100 HIMF (Jones, et al. 2018). As we can see our implementation with some minor modifications reproduces the HIMF of (Haynes, et al. 2018) extremely well.

A2 Error Analysis of HIMF

(i) Mass Errors: Since $M_{\text{HI}} \propto S_21D^2$ the uncertainties on both integrated flux and distances lead to uncertainties on mass. Peculiar velocities of galaxies can affect distance estimates. This effect is larger in the local volume and smaller at higher redshifts where the Hubble flow dominates over peculiar velocities. The α.40 catalog also includes radial distances (Haynes, et al. 2011), using a local volume flow model (Masters 2005) for galaxies with $cz_{\text{CMB}} < 6000\text{km.s}^{-1}$. This model has an estimated local velocity dispersion of $\sigma_v = 163\text{km.s}^{-1}$. For these galaxies we take the error in the distance to be the maximum of $\sigma_v$ and 10% of the distance. For galaxies $cz_{\text{CMB}} > 6000\text{km.s}^{-1}$ distances are estimated using the Hubble and we take the error on distances to be 10%. Using the errors on distance and fluxes and their observed values we generate 300 Gaussian realizations and recompute $M_{\text{HI}}$ for every object. We apply the 2DSWML method to all the realizations and find out the width of the distribution for every $\phi_j$ which we quote as an error.

(ii) Poisson Errors: The observed count in some of the bins is very low which means that it is important to consider errors due to Poisson counting.

(iii) Sample Variance: We estimate this error by splitting the area into 26 contiguous regions of approximately equal area each. We compute the HIMF by removing one region at a time. Finally the jackknife uncertainty for $\phi_j$ is computed as $\sigma_{\phi_j} = \frac{\sum_{i=1}^{N-1} (\bar{\phi}^j - \phi_j^i)^2}{N-1}$ where $\bar{\phi}^j$ is the jackknife mean and $\phi_j^i$ is the value for the $i^{th}$ jackknife sample.

(iv) Other Errors: There are many other sources of errors which are discussed in Jones, et al. (2018). E.g. the error associated with the 2DSWML method which one can either estimate using the information matrix (Efstathiou, Ellis, & Peterson 1988) or by making further mocks (Jones, et al. 2018). One needs to understand how these errors are correlated with a finite sample or Poisson errors. We also expect Poisson errors to be correlated to mass errors especially in the lowest and highest mass bins where the observed counts are low. Another source of uncertainty is the error in velocity width $W_{50}$. Since the peak flux, $S_{\text{peak}}$, is inversely proportional to $W_{50}$ we expect their errors to be correlated. Since we integrate over $W_{50}$ to obtain the HIMF, we do not consider errors in $W_{50}$. In order to properly account for errors one will need their covariances. We have attempted to add further sources of errors in quadrature but we find that the error bars become progressively larger and the Schechter function fits have a $\chi^2_{\text{red}} < 1$ which means that we may be overestimating the errors. For this work we stick to the errors (i) – (iv) and add them in quadrature. These errors are comparable to Martin, et al. (2010); Haynes, et al. (2011).