

The Distribution of Neutral Hydrogen in the Local Universe

By

Saili Dutta

PHYS11201404009

National Institute of Science Education and Research, Bhubaneswar

A thesis submitted to the

Board of Studies in Physical Sciences

In partial fulfillment of requirements

for the Degree of

DOCTOR OF PHILOSOPHY

of

HOMI BHABHA NATIONAL INSTITUTE



February, 2021

Homi Bhaba National Institute

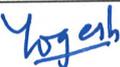
Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by Saili Dutta entitled "The Distribution of Neutral Hydrogen in the Local Universe" and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor of Philosophy.

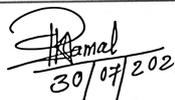
Chairman - Prof. Bedangadas Mohanty  Date 30/07/2021

Guide / Convener - Dr. Nishikanta Khandai  Date 30/07/2021

Examiner - Prof. Somnath Bharadwaj  Date 30/7/2021

Member 1 - Dr. Yogesh Srivastava  Date 30/07/2021

Member 2 - Dr. Chethan Gowdigere  Date 30/7/21

Member 3 - Dr. Pramod Kumar Samal  Date 30/07/2021

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to HBNI.

I/We hereby certify that I/we have read this thesis prepared under my/our direction and recommend that it may be accepted as fulfilling the thesis requirement.

Date : 30/07/2021

Place : Jatni, Bhubaneswar

Signature

Co-guide (if any)



Signature

Guide

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Saili Dutta

Saili Dutta

DECLARATION

I hereby declare that I am the sole author of this thesis in partial fulfillment of the requirements for a postgraduate degree from National Institute of Science Education and Research (NISER). I authorize NISER to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Saili Dutta

Saili Dutta

List of Publications arising from the thesis

Journal

1. "The Population of Galaxies that Contribute to the HI Mass Function"
Saili Dutta, Nishikanta Khandai and Biprateep Dey
MNRAS, **2020**, Vol. 494, Issue 2, May 2020, 2664-2678.
2. "The Distribution of Neutral Hydrogen in the Color-Magnitude Plane of Galaxies"
Saili Dutta, Nishikanta Khandai
MNRAS Letters, **2020**, Vol. 500, Issue 1, November 2020, 37-41
3. "The Dark Matter Halos of HI Selected Galaxies"
Saili Dutta, Nishikanta Khandai, Sandeep Rana
(Submitted to MNRAS)

Others

1. "The Effect of HI-Halo Mass Scatter on the Clustering of HI-selected Galaxies"
Sandeep Rana, Saili Dutta, Nishikanta Khandai
(Manuscript under preparation)

Saili Dutta

Saili Dutta

To
Maa & Baba

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to my thesis advisor, Dr. Nishikanta Khandai, for his continuous support, guidance, motivation, and patience throughout my Ph.D. work. It was a pleasure working with him and I have learned a plenty of things about research, the way of approaching a scientific problem, and how to solve them.

I sincerely thank Prof. Bedangadas Mohanty, who has shown faith in me and supported me in many ways during my tenure at NISER. I acknowledge my Doctoral committee members Dr. Yogesh Kumar Srivastava, Dr. Chethan N. Gowdigere, and Dr. Pramod Kumar Samal for their evaluation and feedback on my work.

I would especially like to thank Prof. Raghunathan Srianand for being involved in my research work; his experience, all the discussions, and suggestions in my work surely encouraged me to make it better. I express sincere gratitude to Prof. Jasjeet Bagla, Prof. Nissim Kanekar, and Prof. Aseem Paranjape for several useful discussions throughout my Ph.D. work. I would also like to thank Dr. Rajeshwari Dutta, Dr. Sandeep Rana, and Biprateep Dey for their help in various aspects during this Ph.D. course.

I thank NISER and DAE for providing me this opportunity, infrastructure support, and funding during the entire span of the research. The computational work was done using the high-performance computing facility at the School of Physical Sciences, NISER.

I would like to acknowledge the work of the entire ALFALFA collaboration in observing, flagging, and extracting the properties of galaxies that this thesis 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. The CosmoSim database used in this dissertation is a service by the Leibniz-Institute for Astrophysics Potsdam (AIP). The MultiDark database was developed in cooperation with the Spanish MultiDark Consolider Project CSD2009-00064. I gratefully acknowledge the

Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) and the Partnership for Advanced Supercomputing in Europe (PRACE, www.prace-ri.eu) for funding the MultiDark simulation project by providing computing time on the GCS Supercomputer SuperMUC at Leibniz Supercomputing Centre (LRZ, www.lrz.de).

Finally, I express my respect and gratefulness to my parents for their continuous support and all the sacrifices they have done for me. And special thanks to Mustakim, Manoar, and Soheli for all those memorable times we spent together.

ABSTRACT

The 21 cm line of neutral atomic hydrogen (HI) can be used as a probe to study the formation and distribution of large scale structures in the Universe. In the Post-reionization era although the hydrogen gas is almost ionized, there are still traces of neutral hydrogen locked up in galaxies and provides fuel for star formation. In the first part of the thesis, we look at the contribution of different galaxy populations to the HI mass function (HIMF) and the HI density parameter, Ω_{HI} , in the local Universe. We use a sample of 7857 HI-selected galaxies from the ALFALFA (40% catalog) survey which has a common volume with the SDSS DR7. We define the populations based on the colors ($u - r$) and magnitudes (M_r) of galaxies. We find that the low mass end of the HIMF is dominated by the faint blue galaxies, while the bright blue population contributes at moderate to high mass values, i.e., for $\log_{10}(M_{\text{HI}} h_{70}^2 / M_{\odot}) > 8.0$. 70% of the $\Omega_{\text{HI}}^{\text{tot}}$ comes from the bright and faint blue populations. The bright red population of galaxies dominates the high mass end of the HIMF, while the full red sample represents $\sim 17\%$ of the $\Omega_{\text{HI}}^{\text{tot}}$. The sample has 148 galaxies which have no optical counterparts in SDSS DR7, dark galaxies, and they also have a non-negligible ($\sim 3\%$) contribution to the $\Omega_{\text{HI}}^{\text{tot}}$. In the second part, we present the conditional HIMF conditioned on observed optical properties for the same sample excluding dark galaxies. We use the conditional HIMF to estimate the underlying distribution of Ω_{HI} , $p(\Omega_{\text{HI}})$, in the color-magnitude ($[u - r] - M_r$) plane, and we find the distribution to be skewed with a long tail towards the faint blue and bright red galaxies. We discuss how the method to obtain $p(\Omega_{\text{HI}})$ can be used to find the underlying mean scaling properties of galaxies in an unbiased way. In the third part of the thesis, we estimate the HI velocity function (HIVF) for the total, red and blue samples. The blue galaxies dominate the low velocity end of the HIVF, whereas the red population contributes to the high velocity end. The distributions of these two populations are well separated at the high velocity end suggesting that inclination plays an important role in the reddening of HI selected galaxies. Finally, we describe a model to populate HI in dark matter halos motivated and calibrated from recent observations. We use a non-monotonic mean $\langle M_{\text{HI}} - M_{\text{h}} \rangle$ relation to obtain the HI selected halo mass function, $\phi^{\text{HI}}(M_{\text{h}})$. We abundance match $\phi(M_{\text{HI}})$, $\phi^{\text{HI}}(M_{\text{h}})$, $\phi(w_{50})$, $\phi(V_{\text{rot}})$ and obtain the scaling relations for the total, red and blue populations.

Contents

Summary	xi
List of Figures	xii
List of Tables	xxi
Chapter 1 Introduction	1
1.1 Linear Perturbation Theory	8
1.2 Two Point Correlation Function	9
1.3 Mass Function : The Press-Schechter Formalism	11
1.4 Neutral Atomic Hydrogen - HI	13
1.4.1 21 cm radiation from neutral Hydrogen (HI line)	15
1.4.2 Einstein Coefficients	18
1.4.3 Absorption and Emission Coefficients in Terms of Einstein Coefficients	19
1.4.4 Einstein Coefficients, Flux and HI mass	20
1.5 Structure of the Thesis	21
Chapter 2 The Population of Galaxies that Contribute to The HI Mass Function	25
2.1 Introduction	25
2.2 Data	27
2.2.1 ALFALFA Survey	27
2.2.2 The Sample	30
2.3 Populations of Galaxies	36
2.4 The HI Mass Function (HIMF)	38
2.4.1 HIMF Estimators	38
2.5 Results	41
2.5.1 The contribution of different galaxy populations to Ω_{HI}	51

2.6	Discussion	52
2.7	Summary	60
Chapter 3	The distribution of neutral hydrogen in the color–magnitude plane of galaxies	63
3.1	Introduction	63
3.2	Data	64
3.3	Conditional HIMF	65
3.4	Results	67
3.4.1	The Distribution of Ω_{HI} in the $C_{\text{ur}} - M_{\text{r}}$ plane	71
3.5	Discussion	75
Chapter 4	The Dark Matter Halos of HI Selected Galaxies	79
4.1	Introduction	79
4.2	Data	81
4.3	Estimating the HI Velocity Width Function	82
4.3.1	Uncertainties on HIWF	83
4.4	HI Velocity Width Function	85
4.5	Scaling Relations for HI Selected Galaxies	95
4.6	Discussion and Summary	110
Chapter 5	Summary and Discussion	121
5.1	Summary	121
5.2	Future Direction	124
Chapter A	Appendix	128
A1	Estimation of HIMF using 2DSWML Method	128
A1.1	2DSWML Method : Definition	130
A1.2	Normalization of HIMF and HIWF	136
A1.3	Error Analysis	140
A2	List of Constants	143
	References	144

List of Figures

2.1	Cone-plot of $\alpha.40$ galaxies in Virgo direction region.	28
2.2	Number distribution of $\alpha.40$ galaxies as a function of a) heliocentric velocity (cz_{helio}) b) velocity width (w_{50}) c) integrated flux (S_{21}) d) signal-to-noise ratio (SNR) e) distance (dist) f) HI mass (M_{HI}).	29
2.3	Common footprint of ALFALFA and SDSS used in this thesis. The red patches is that of SDSS DR7 overlapping with the ALFALFA survey region. The blue dots are the positions of ALFALFA galaxies. The common boundary used in this work is outlined by the thick black line.	31
2.4	The distribution of galaxies in the $S_{21} - w_{50}$ plane is shown for all Code 1 ALFALFA galaxies. The broken orange solid line is the 50% completeness relation [Haynes et al., 2011]. It is given in eq. 2.1. We have further divided the sample into three mass bins: $\log_{10}[M_{HI}/M_{\odot}] \in [6.0, 8.5[, [8.5, 10.2[, [10.2, 11.0[$. The 1σ contours and the peaks of the distributions for these three populations are given by thick blue solid line (plus-circle), the dot-dashed black line (dot-circle) and the thin red solid line (cross-circle).	33
2.5	The observed distribution of galaxies in the color-magnitude plane for SDSS (left panel) and ALFALFA (right panel) in the common volume considered in this work. The 1σ and 0.25σ contours are given by the thin and thick lines. A bimodal distribution of galaxies is seen in SDSS. The solid black curve (eq. 2.2) is used to classify the galaxies into red (above curve) and blue (below curve) populations [Baldry et al., 2004]. The numbers in each panel indicate the observed counts of galaxies in this color-magnitude range for each of the populations.	35

2.6	The distribution of HI selected galaxies in this work in the color-magnitude plane. The upper solid curve demarcates the red (above curve) from the blue (below curve) population as in Baldry et al. [2004]. The vertical solid (dashed) line divides the luminous (left) and faint (right) populations, referred as 1.5σ (1σ) cut in magnitude. The lower solid (dashed) line divides the blue population into bluer (below the curve) and blue (above the curve) populations, we call it 1.5σ (1σ) cut in color. This breaks the sample of HI selected galaxies into 6 disjoint sub-samples in the color-magnitude plane. The number for each population is quoted, where the numbers in brackets are for the 1σ sample.	37
2.7	HIMFs estimated by $1/V_{\max}$ method by [Martin et al., 2010](black circles) and our result of HIMF (red squares), where we do not use the flow model [Martin et al., 2010].	39
2.8	The HIMF for the luminous blue, luminous red, and faint blue populations (1.5σ definition). Data points and error bars were estimated using the 2DSWML method (see appendix A). The red dotted and blue dot-dashed curves are the Schechter function fits for red and blue populations, respectively. For comparison we have added the total HIMF (crosses) and its fit (solid line). The symbols for the populations are the same as in figure 2.6. The details of the fits are given in table 2.1.	42
2.9	The HIMF for the luminous blue, luminous red, faint blue and faint bluer populations (1σ definition). Data points and error bars were estimated using the 2DSWML method (see appendix A). The red dotted, blue dot-dashed and cyan dashed curves are the Schechter function fits for red, blue, and bluer populations, respectively. For comparison we have added the total HIMF (crosses) and its fit (solid line). The symbols for the populations are the same as in figure 2.6. The details of the fits are given in table 2.1.	43

2.10	The HIMF for the luminous bluer, faint bluer, and faint red populations (1.5σ definition). Data points and error bars were estimated using the 2DSWML method (see appendix A). The red dotted and cyan dashed curves are the Schechter function fits for red and bluer populations, respectively. For comparison we have added the total HIMF (crosses) and its fit (solid line). 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 populations are the same as in figure 2.6. The details of the fits are given in table 2.1.	44
2.11	The HIMF for the luminous bluer and faint red populations (1σ definition). Data points and error bars were estimated using the 2DSWML method (see appendix A). The red dotted and cyan dashed curves are the Schechter function fits for red and bluer populations, respectively. For comparison we have added the total HIMF (crosses) and its fit (solid line). 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 populations are the same as in figure 2.6. The details of the fits are given in table 2.1.	45
2.12	The $M_r - M_{\text{HI}}$ relation for the 1.5σ sample. The filled star represents the total sample excluding the dark galaxies. The other data points and line styles are the same as in figure 2.8, 2.9, 2.10, and 2.11.	50
2.13	The $M_{\text{star}} - M_{\text{HI}}$ relation for the 1.5σ sample. 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 2.8, 2.9, 2.10, and 2.11.	54
3.1	Conditional HIMF as a function of increasing color thresholds (top to bottom). The thick solid line is the HIMF for the full sample. The shaded gray region does not contain data, the conditional HIMF have however been extrapolated into this regime as well.	66

3.2	The Schechter function parameters of the conditional HIMF and their uncertainties as a function of color thresholds. The solid lines are fits to the data points with a quadratic function. The top, middle and bottom panels show the dependence of M_* , ϕ_* and α respectively, on the color threshold C_{ur}^t .	68
3.3	Conditional HIMF as a function of decreasing rest frame magnitude thresholds (top to bottom). The thick solid line is the HIMF for the full sample. The shaded gray region does not contain data, the conditional HIMF have however been extrapolated into this regime as well.	70
3.4	The Schechter function parameters of the conditional HIMF and their uncertainties as a function of magnitude thresholds M_r^t . The solid lines are fits to the data points. For α (bottom) we fit with a quadratic function. For M_* (top) and ϕ_* (middle) we fit with a function of the form: $y(x) = \left[a + b \exp\left(-\frac{(x+c)^2}{2d}\right) \right] \frac{f}{(x+e)}$	72
3.5	The bottom left panel shows the distribution function $p(\Omega_{HI})$ (see eq. 3.6) in the CM plane color coded by $(\Omega_{HI}^{ij})^{\text{norm}}$ (eq. 3.8). The thick (thin) line represent the 1σ (2σ) widths of $p(\Omega_{HI})$. The dash-dot line separates the optical red (above) and blue (below) populations [Baldry et al., 2004]. The top left (bottom right) panel is the marginalized distribution of Ω_{HI} as a function of M_r (C_{ur}). The crossed circle represents the peak of the two-dimensional distribution function, $p(\Omega_{HI})$.	74

3.6	<p>HI density, Ω_{HI}, as a function of redshift, z. The black up-triangles are our estimates of Ω_{HI}. The green plus sign and red star are the estimates from Zwaan et al. [2005] for HIPASS galaxies and Jones et al. [2018] for ALFALFA galaxies, respectively, at $z \sim 0$, using 21 cm emission line. The cyan filled square is measured by Lah et al. [2007] by stacking analysis using the Giant Metrewave Radio Telescope (GMRT). The black arrow line is the upper limit estimated by Kanekar et al. [2016] using the GMRT HI stacking. The red open diamonds, violet filled circles, and orange open squares are the measurements from damped Lyman-α from the Hubble Space Telescope (HST) and the SDSS by Noterdaeme et al. [2012], Bird et al. [2017], and Rao et al. [2017], respectively. The blue dot circles are the estimates by Rhee et al. [2013] using the Westerbork Synthesis Radio Telescope (WSRT) and HI stacking. The magenta cross is the estimate of Ω_{HI} by Chowdhury et al. [2020] using upgraded GMRT(uGMRT) and DEEP2 galaxy redshift survey.</p>	76
4.1	<p>Our estimates of the HIMFs are shown for the total (black cross), blue (blue filled square) and red (red filled circle) samples. The Schechter function fits are plotted for the total (black thick solid line), red (red thin dashed line) and blue (blue thin dot-dashed line) samples for the HIMFs. The black thick dashed line is the estimate of the HIMF from Martin et al. [2010] using the $\alpha.40$ sample. To better display the data for the red and blue samples has been horizontally shifted with respect to the total sample.</p>	84

4.2	Our estimates of the HIWFs for the total (black cross), blue (blue filled square) and red (red filled circle) samples. The modified Schechter function fits are plotted for the total (black thick solid line), red (red thin dashed line) and blue (blue thin dot-dashed line) samples for the HIWFs. To better display the data for the red and blue samples has been horizontally shifted with respect to the total sample. The black thick dashed line is the estimate of the HIWF from Moorman et al. [2014] using the $\alpha.40$ sample. The open red circles (blue squares) are the unnormalized estimates of the HIWF for the red (blue) sample from Moorman et al. [2014].	86
4.3	This flowchart summarizes the steps to obtain the HIVE.	88
4.4	A comparison of our estimates with previously published results of the HIWF are shown in this figure. Our modified Schechter function fits and their 1σ uncertainty (shaded region for the total sample only) are plotted for the total (black thick solid line), red (red thin dashed line) and blue (blue thin dot-dashed line) samples. The black thick dotted line represents the estimates of HIWF from Papastergis et al. [2011]. The black thin dot-dashed line is the estimates of Zwaan et al. [2010] from the HIPASS data for the full sample. The black thin dashed line is the estimate of the HIWF of Moorman et al. [2014].	90
4.5	A comparison of our estimates of HIVE with previously published results are shown in this figure. Our modified Schechter function fits and their 1σ uncertainty (shaded region for the total sample only) are plotted for the total (black thick solid line), red (red thin dashed line) and blue (blue thin dot-dashed line) samples. The black thick dotted line represents the estimates of Papastergis et al. [2011] for HIVE. The black thin dot-dashed line is the estimate of Zwaan et al. [2010] for late-type galaxies.	92

4.6 Average HI mass - halo mass relation from Guo et al. [2020] is shown as red open circles. The black dashed line is the $M_{\text{HI}} - M_{\text{h}}$ scaling relation obtained by abundance matching our HIMF to the HMF. The blue dotted line is the $M_{\text{HI}} - M_{\text{h}}$ relation obtained by abundance matching our HIMF to the HI-selected HMF (see equation 4.5 and discussion). The blue solid line is the mean $\langle M_{\text{HI}} \rangle - M_{\text{h}}$ relation that is obtained by averaging the blue dotted line over all the halos at any given mass $[M_{\text{h}}, M_{\text{h}} + dM_{\text{h}}]$. The HI-selected HMF has been defined so that the solid line matches the observed points of Guo et al. [2020]. 100

4.7 The HMF (HI-selected HMF) is shown as the solid thick (thin) line. The arrow at $M_{\text{h}} \simeq 11$ is the resolution of the simulation catalog. For masses below that we have linearly extrapolated the HMF. The red thick dashed (blue thick dot-dashed) line represents the HMF hosting early (late) type galaxies. (See figure 4.8 and corresponding text in section 4.5 for the definition of early (late) type galaxies. The corresponding HI-selected HMF are shown in thin lines. 101

4.8 This figure shows the observed age (mean stellar age in Gyr) distribution of the total (black crosses, thick solid line), blue (blue open circles, thick dashed line), and red (red open squares, thin dashed line) sample of galaxies in ALFALFA. The intersection of the age distribution of blue and red galaxies at $t_{\text{age}} = 8.38\text{Gyr}$ is used to classify the blue and red populations as late-type and early-type galaxies respectively. 104

4.9 This figure shows the SMF. The grey open-diamonds with thin solid line represents observational estimates of the SMF from SDSS-GALEX [Moustakas et al., 2013]. The black thick solid line is the SMF from the SAGE catalog [Knebe et al., 2018] which is based on the MDPL2 simulation [Klypin et al., 2016]. The red dashed (blue dot-dashed) line is the SMF for early-type (late-type) galaxies defined as $t_{\text{age}} > 8.38$ ($t_{\text{age}} < 8.38$) Gyr in the context of HI-selected galaxies (see figure 4.8). 106

4.10	The figure shows the scaling relations obtained by abundance matching the three HI distributions, the HIMF, HIWF, and HIVE, amongst themselves. The left, middle and right panels show the scaling relation $M_{\text{HI}} - w_{50}$, $M_{\text{HI}} - V_{\text{rot}}^{\text{HI}}$, and $w_{50} - V_{\text{rot}}^{\text{HI}}$ respectively for the total (black solid), red (red dashed), and blue (blue dot-dashed) samples.	108
4.11	This figure shows the scaling relations by abundance matching the three HI distributions, namely the HIMF, HIWF, and HIVE with the HI-selected HMF (equation 4.5). The left, middle, and right panels show the scaling relation $M_{\text{h}} - M_{\text{HI}}$, $M_{\text{h}} - w_{50}$ and $M_{\text{h}} - V_{\text{rot}}^{\text{HI}}$ respectively for the total (black solid), red (red dashed) and blue (blue dot-dashed) samples.	109
4.12	The $M_{\text{HI}} - M_{\text{h}}$ scatter plot (grey open circles) for a volume-limited subsample generated from the ALFALFA survey. The halo masses have been estimated using the $M_{\text{star}} - M_{\text{h}}$ relation of Behroozi et al. [2010]. The black solid (dashed) line is the $M_{\text{HI}} - M_{\text{h}}$ relation obtained by abundance matching the HIMF with the HI-selected HMF (HMF).	111
4.13	A comparison of HI properties, $M_{\text{HI}} - V_{\text{rot}}$, for a volume limited sample of blue galaxies in ALFALFA (grey open circle) and SPARC (black open square).	113
4.14	A comparison of HI properties, $M_{\text{h}} - M_{\text{star}}$, for a volume limited sample of blue galaxies in ALFALFA (grey open circle) and SPARC (black open square). The solid line is the $M_{\text{star}} - M_{\text{h}}$ relation of Behroozi et al. [2010]	114
4.15	A comparison of $M_{\text{HI}} - M_{\text{h}}$, for a volume limited sample of blue galaxies in ALFALFA (grey open circle) and SPARC (black open square).	116
4.16	A comparison of $V_{\text{rot}} - M_{\text{h}}$, for a volume limited sample of blue galaxies in ALFALFA (grey open circle) and SPARC (black open square).	117
4.17	The black thick line is the HMF. The blue thin line is our estimate of the HI-selected HMF for late-type (or blue) galaxies based on ALFALFA data. The red dashed line is the HI-selected HMF for late-type galaxies by Li et al. [2019b] based on SPARC and HIPASS data.	119

A1	The completeness relation (<i>broken solid line</i>) is shown in the M-W plane. This particular galaxy 'i' has the following properties: $\log_{10} [M_{\text{HI}}^i/M_{\odot}] = 8.73$, $[w_{50}^i/\text{km.s}^{-1}] = 93$, $[D/\text{Mpc} = 27]$. The shaded(white) region is the area accessible(inaccessible) to this galaxy in the M-W plane.	131
A2	Comparison of the results from 2DSWML method with considering H_{ijk} s as 0 or 1 (blue solid line) and with fractional values of H_{ijk} s (red dot-dashed line).	133
A3	HIMF at different iterations. The black solid line is the observed number density. The blue filled circles connected by blue dashed line represents the true HIMF obtained for our sample. The red, pink, cyan and green filled circles are the ϕ_j s after 1, 2, 3 and 6 iterations, respectively.	135
A4	Results from 2DSWML method. Bottom left panel: the true two-dimensional distribution of ALFALFA galaxies in $M_{\text{HI}} - w_{50}$ plane. Orange dots are the observed galaxies from our sample of analysis. Top left panel: HIMF for our total sample, obtained by integrating the 2D distribution at the bottom panel over w_{50} bins. Bottom right panel: HIWF for our total sample, obtained by integrating the 2D distribution over M_{HI} bins. The errorbars and the normalization of the distribution functions are discussed in the next sections.	137
A5	Upper panel: Red dashed line is the mass function estimated using 2DSWML method, which is not normalized. The black solid line is the number density per mass bin calculated for the same sample of HI galaxies. Lower panel: The blue solid line shows the ratio of the un-normalized mass function and the number density, multiplied by 100.	139
A6	Comparison of the HIMF in the $\alpha.100$ sample. The open circles is the HIMF by Jones et al. [2018]. The open squares is the HIMF estimated by our implementation of the 2DSWML method.	141

List of Tables

2.1	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, $\chi^2_{reduced}$ is given in the last column.	47
2.2	The contribution of different populations to Ω_{HI} . Column 2 is the estimate of Ω_{HI} from a single population and column 3 is the fractional contribution to $\Omega_{\text{HI}}^{\text{tot}}$. The estimated values for the 1σ sample is in brackets.	52
4.1	Best fit values of modified Schechter parameters for $\phi(w_{50})$ and $\phi(V_{\text{rot}}^{\text{HI}})$ for total, blue and red samples.	94
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	129

Chapter 5

Summary and Discussion

This dissertation focuses on understanding the properties of HI-selected galaxies in terms of their statistical distributions. This approach provides better constraints on the population of HI-selected galaxies as these are corrected for the survey or sample selection. In this chapter we give a brief summary of the work presented in this thesis, including the main results from chapters 2, 3, 4. We conclude this chapter by discussing future research directions.

5.1 Summary

The Population of Galaxies that Contribute to the HIMF

We use a sub-sample of 7857 galaxies from the ALFALFA 40% data release. 7709 (98%) of these galaxies have detections in the optical survey – SDSS (DR7). We divide these galaxies into *six* populations based on their color ($u - r$) and magnitude (M_r) values and we label them as — luminous red, faint red, luminous blue, faint blue, luminous bluer, and faint bluer galaxies. The remaining 148 (2%) galaxies do not have optical counterparts in SDSS DR7, which we refer as *dark* galaxies. We study the contribution of these *seven* populations of galaxies to the total HIMF, $\phi(M_{\text{HI}})$ and to the total HI density parameter, Ω_{HI} . We find that the faint blue, luminous blue, and luminous red are the dominant populations. The faint blue population dominates the low mass end, and the luminous blue population dominates

the knee of the total HIMF. The high mass end of the HIMF is dominated by the luminous red galaxies. The full blue cloud (faint and luminous blue+bluer) contributes 80% to the total Ω_{HI} . The red cloud (faint and luminous) contributes 17% to the total HI budget. The remaining 3% of HI is coming from the dark galaxies. This is the first time we are able to quantitatively look at how different galaxy populations contribute to the total HIMF, and hence Ω_{HI} . The unexpected result, that we obtain, is that there is a significant amount of HI locked up in red galaxies, which are usually associated as passive, early type galaxies.

The Distribution of Neutral Hydrogen in the Color-Magnitude Plane of Galaxies

In the next chapter (chapter 3), we use the same sample of 7709 HI-selected galaxies from ALFALFA. We create sub-samples out of these galaxies by considering thresholds on color ($u - r$) or magnitude (M_r), and compute the HIMFs for each of them — conditional HIMF conditioned on color/magnitude. We extend this to two-dimensions and obtain the conditional HIMF conditioned on color and magnitude, from which we finally estimate the distribution of Ω_{HI} as a function of ($u - r$) and M_r . We find that the width of this distribution is quite broad in both color and magnitude. It is skewed with long tails towards the faint blue and the luminous red populations. We obtain the peak of this distribution at $M_r = -19.25$, $(u - r) = 1.44$. The 1σ and 2σ widths in $(u - r)$ are 0.8 and 1.1, respectively, and in M_r the widths are 3.0 and 4.8. 10% of the total Ω_{HI} is coming from the fainter ($M_r > -16$) galaxies, whereas the red galaxies contribute $\sim 17\%$ to the total HI budget. This is the first time we quantify the distribution of Ω_{HI} as a function of color and magnitude. This result can be used to correct the estimates of HI obtained through stacking which generally consider the bright blue population of galaxies. This also gives a better information to the observers on how to optimally target HI from optical surveys. Alternately this distribution can be used in modelling HI in cosmological hydrodynamical simulations.

The Dark Matter Halos of HI Selected Galaxies

In chapter 4 we show the estimates of the HIWF and the HIVE for the sample of 7709 galaxies. We divide this sample into red and blue populations, and obtain the HIWFs and HIVEs for them. We find that the lower velocity width (velocity) of the HIWF (HIVE) is dominated by the blue population of galaxies and the higher velocity width (velocity) is dominated by the red galaxies. Unlike the HIMF, the HIWFs (HIVEs) of the red and blue galaxies are well separated beyond the knee. This emphasizes the fact that the galaxies appear to be redder due to their inclinations. We then describe a model to populate HI in dark matter halos calibrated from observations of HI in massive galaxies. We argue that HI mass and halo mass are generally not correlated, but for an HI-selected sample we see a correlation between the two. Using a non-monotonic mean HI-halo mass relation calibrated from observations, we define the HI-selected HMF. We consider the HMF and the SAGE galaxy catalog from the MDPL2 simulation for this part of the work. Using abundance matching techniques, we obtain the scaling relations between $M_{\text{HI}} - M_h$, $w_{50} - M_h$, and $V_{\text{rot}} - M_h$. We define red and blue populations in the SAGE catalog based on their mean stellar ages and also obtain the scaling relations for the red and blue populations. We find that the scaling relations for gas-rich halos are similar for the red and blue populations, but for gas-poor halos the scaling relations are different. We also do a consistency check of our scaling relation for the total sample, and show that the scaling relation of $M_{\text{HI}} - M_h$ that we obtain is consistent with the data. We effectively provide an alternative halo model which is quite simple compared to traditional halo models.

Large volume cosmological hydrodynamical simulations have computed the abundances of atomic and molecular hydrogen by post-processing the outputs of the simulations. [Villaescusa-Navarro et al. \[2018\]](#) study the abundance of HI in the IllustrisTNG simulations [[Marinacci et al., 2018](#), [Naiman et al., 2018](#), [Pillepich et al., 2018](#), [Springel et al., 2018](#), [Nelson et al.,](#)

2018] using the self-shielding prescription by Rahmati et al. [2013]. Their model reproduces Ω_{HI} at higher redshift values well on average, although it underpredicts the HI content at moderate redshifts ($2 < z < 3.5$) and overpredicts it in the local universe. However, Diemer et al. [2019] show that the HIMF is overestimated by a factor of 2-3 at lower masses in the local Universe for the same simulation. The HIMF predicted [Davé et al., 2017] by the MUFASA simulation [Davé et al., 2016] using GIZMO [Hopkins, 2015], which also follows the same prescriptions as Villaescusa-Navarro et al. [2018], agrees well with the ALFALFA HIMF. But this was for a smaller simulation box, so the finite volume effects may change the results in larger boxes. However, the HIMF is a projection of higher-dimensional multivariate distribution in one dimension. Looking at the upcoming surveys, targeting different bands of the electromagnetic spectrum, the theories of galaxy formation need to go beyond one dimensional marginalized abundances. It will be useful if these cosmological simulations can reproduce the joint abundances of any two (or more) properties, e.g., HI-HI or HI-optical. The results from this thesis have outlined a method to obtain and constrain multivariate galaxy distributions. Our results can be used either as constraints to be satisfied for the theories of galaxy formation or can be used to model them in N-body simulations. Chapter 4 describes one such approach.

5.2 Future Direction

The ALFALFA survey has recently released the 100% data catalog with the optical properties from SDSS [Durbala et al., 2020]. Therefore we do not have to restrict ourselves to the $\alpha.40$ sample any more, we can carry out further analysis with the bigger $\alpha.100$ sample. The work done in this thesis can be extended in many directions. We outline some of the ideas below.

Conditional HIWF/HIMF

The abundance of galaxies corresponding to different components of galaxies — i) neutral hydrogen gas - HI [Martin et al., 2010, Zwaan et al., 2003], ii) stellar mass - M_{star} [Baldry et al., 2004], iii) molecular hydrogen gas - HII [Catinella et al., 2010, Saintonge et al., 2011], and iv) ionized gas — have been looked into. In order to better understand and constrain galaxy properties, it is important to go beyond single variable abundances which represent marginalized distributions of multi-variable distributions. In this thesis we have taken a first step in this direction and have presented the conditional HIMFs, conditioned on optical properties (color/magnitude). We would like to extend this work and estimate the conditional HIMFs conditioned on other optical properties (e.g. other SDSS colors/magnitudes). The same can be done for HIWFs and HIVFs. We can also estimate the conditional HIMFs/HIWFs/HIVFs conditioned on derived properties like - M_{star} , SFR, stellar ages, metallicity etc. These conditional distribution functions will put additional observational constraints for theories of galaxy formation.

$M_{\text{HI}} - M_{\text{star}}$ Scaling Relation

In chapter 3 we have outlined a method to estimate the distribution of Ω_{HI} in the color-magnitude plane of galaxies using conditional HIMFs. In a similar approach, using an optically-selected sample, the distribution of $\Omega_{\text{star}} = \rho_{\text{star}}/\rho_c$ can be obtained in the same color-magnitude plane by computing conditional SMFs. This method also gives the estimates of the average number density of galaxies in each color-magnitude pixel for both the HI-selected sample and optically-selected sample. Hence, if we consider both the distributions, the distribution of Ω_{HI} and the distribution of Ω_{star} , in a common color-magnitude plane, we can obtain the underlying scaling relation between M_{HI} and M_{star} . As the computation of HIMFs and SMFs take care of the corresponding selection function of the survey,

this scaling relation will be unbiased, and will not depend on sample selection and the selection function of the two surveys. This kind of scaling relation will be very useful in the context of halo modelling, and will complement the approach we have discussed in chapter 3. Using this scaling relation we can directly relate HI masses with halos via their stellar masses [Behroozi et al., 2010].

Reproducing Small Scale Clustering Using Our Halo Model

The halo model, outlined in chapter 4, does not account for the scatter around the mean HI-halo mass relation (see figure 4.12). In Rana et al. [2021, in preparation], we include the scatter as a parameter in the halo model and obtain the $M_{\text{HI}} - M_{\text{h}}$ scaling relation using the ALFALFA 70% data in the same approach as discussed in chapter 4. Using this new scaling relation we create a mock HI catalog and estimate the projected 2PCF, $w_p(r_p)$, for different M_{HI} thresholds. We then compare our clustering results with the observed results of Guo et al. [2017]. We find that a scatter of ~ 0.2 around the mean HI-halo mass relation is able to reproduce the observed large scale clustering as presented in Guo et al. [2017]. But the small scale clustering is not being reproduced well as we have not distinguished between centrals and satellites in the halo catalog. We therefore will try to improve our halo model and reproduce the small scale clustering.

Redshift-Space Distortion for HI-selected Galaxies

Since we have the HI-selected galaxies and optically-selected galaxies from ALFALFA and SDSS, respectively, from a common volume of the sky, we can measure their cross and auto correlations. This analysis has been done for the $\alpha.40$ data sample [Papastergis et al., 2013]. However the results did not discuss the implications of small scale and large scale clustering. We want to repeat this analysis for the ALFALFA 100% data and mea-

sure redshift-space distortions of HI-selected galaxies through auto correlation and cross correlation with SDSS. At small scales, the clustering results will tell us about the local environment of HI, and at large scales, it will be a useful complementary probe of cosmology. Also it will be interesting to see whether by putting our observational constraints to the simulations and using our halo model, the observed redshift space distortions can be reproduced in simulations.

Summary

In this dissertation, we look at the distribution of neutral atomic hydrogen (HI) in galaxies and dark matter halos, in the local Universe ($z \leq 0.05$). We use the Arecibo Legacy Fast ALFA (ALFALFA) 40% data ($\alpha.40$), Sloan Digital Sky Survey (SDSS) Data Release 7 (DR7) and the Semi-Analytic Galaxy Evolution (SAGE) catalog from the Multi Dark Planck 2 (MDPL2) simulation for this work.

In the first part (chapter 2), we define different populations of galaxies in terms of their colors and rest frame magnitudes and investigate their contribution to the total HI Mass Function (HIMF) and to the total HI density parameter ($\Omega_{\text{HI}}^{\text{tot}}$). We find that the faint blue population dominates the low mass end of the total HIMF, the luminous blue population dominates the knee and the luminous red population contributes to the high mass end of the HIMF. The full blue cloud contributes $\sim 80\%$ of the $\Omega_{\text{HI}}^{\text{tot}}$, whereas $\sim 17\%$ of $\Omega_{\text{HI}}^{\text{tot}}$ comes from the red cloud. The dark galaxy population, for which there are no optical counterparts in SDSS, represents $\sim 3\%$ of $\Omega_{\text{HI}}^{\text{tot}}$.

In the second part (chapter 3), we evaluate the conditional HIMFs, conditioned on color and/or magnitude and using these we estimate the underlying distribution of Ω_{HI} in the color-magnitude plane. This is the first time we quantify the distribution of Ω_{HI} in the color-magnitude plane. This distribution peaks in the blue cloud at $M_{\text{r}}^{\text{max}} = -19.25$, $C_{\text{ur}}^{\text{max}} = 1.44$, and it is skewed towards the faint blue galaxies and luminous red galaxies.

In the third part (chapter 4), we estimate the HI width functions (HIWF) and HI velocity functions (HIVF) for the full ALFALFA sample and also for the red and blue populations of galaxies. We, then, describe a model to populate HI in dark matter halos motivated and calibrated from recent observations. We use a non-monotonic mean $\langle M_{\text{HI}} - M_{\text{h}} \rangle$ relation, determined by stacking HI spectra from the ALFALFA survey for massive galaxies, to define and obtain an HI selected halo mass function (HMF). Using this HI-selected HMF we match the abundances of different properties of HI selected galaxies, to obtain the scaling relations between $M_{\text{HI}} - M_{\text{h}}$, $V_{\text{rot}} - M_{\text{h}}$, and $w_{50} - M_{\text{h}}$ for the total, red and blue populations. We also find that our method is consistent with a volume limited sample.

Chapter 1

Introduction

The most successful model explaining various observations about the Universe is the Big Bang Model. [Hubble \[1929\]](#) presented the observational evidence of an expanding Universe. He showed that all the galaxies are receding from us with a velocity (v) proportional to its distance (d) from us; $v = H_0 d$. The expansion rate H_0 is known as the Hubble's constant, whose current estimate is $H_0 = (67.4 \pm 0.5) \text{ km s}^{-1} \text{ Mpc}^{-1}$ [Planck 2018, [Aghanim et al., 2020](#)].

The energy density in our present day Universe is dominated by non-relativistic matter (baryonic matter + dark matter) and dark energy (described by the cosmological constant " Λ ") — which we will refer as the Λ -cold dark matter (Λ CDM) model. Baryons consist of standard model particles. By studying the dynamics of galaxies in the Coma cluster, [Zwicky \[1933\]](#) concluded that the observed velocity dispersion of these galaxies could be understood by the presence of matter which is invisible but interacts gravitationally — dark matter. Later [Rubin and Ford \[1970\]](#) presented the evidence of dark matter around the Andromeda galaxy by looking at its rotation curve. In its simplest form Dark energy is described by a fluid with a negative equation of state ($P = \omega \rho$). Dark energy is described by a cosmological constant (Λ) when $\omega = -1$.

The two main pillars, which the modern cosmology is based on, are - i) the *Cosmological Principle* [[Milne, 1933](#)]: the Universe is homogeneous and isotropic on large scales,

ii) *Einstein's General Theory of Relativity*: the distribution of energy in the Universe determines the structure of space-time and vice-versa. The most general metric satisfying homogeneity and isotropy is the Friedmann-Lemaitre-Robertson-Walker (FLRW) metric [Friedmann, 1922, Lemaître, 1931, Robertson, 1935, Walker, 1937],

$$ds^2 = c^2 dt^2 - a^2(t) \left[\frac{dx^2}{1 - kx^2} + x^2 (d\theta^2 + \sin^2 \theta d\phi^2) \right] \quad (1.1)$$

where, t is the cosmic time, (x, θ, ϕ) are the spatial coordinates, $a(t)$ is the scale factor, k is the spatial curvature and can take three values: i) $k = 1$, (closed Universe) ii) $k = 0$, (flat Universe) and iii) $k = -1$, (open Universe).

The physical coordinate or proper coordinate, $\vec{r}(t)$, is related to the comoving coordinate, \vec{x} , as

$$\vec{r}(t) = a(t) \vec{x} \quad (1.2)$$

The proper velocity can be written as

$$\begin{aligned} \frac{d}{dt} \vec{r}(t) &= \frac{d}{dt} [a(t) \vec{x}] \\ \dot{\vec{r}} &= \dot{a}(t) \vec{x} + a(t) \dot{\vec{x}} \end{aligned} \quad (1.3)$$

$$\dot{\vec{r}} = \vec{v}_H + \vec{v}_P \quad (1.4)$$

$\vec{v}_H = \dot{a}(t) \vec{x}$ and $\vec{v}_P = a(t) \dot{\vec{x}}$ are the Hubble flow and peculiar velocity of the object, respectively. The Hubble flow, \vec{v}_h , can be written as

$$\begin{aligned} \vec{v}_H &= \dot{a}(t) \vec{x} \\ \vec{v}_H &= \frac{\dot{a}(t)}{a(t)} a(t) \vec{x} \\ \vec{v}_H &= H(t) \vec{r}(t) \end{aligned} \quad (1.5)$$

This gives the Hubble's law where $H(t)$ is the Hubble's parameter. In the absence of perturbations (or peculiar velocity), the observed velocity is just due to the expansion of the Universe. Today i.e. at $t = t_0$, $H(t_0) = \frac{\dot{a}(t_0)}{a(t_0)} = H_0$.

The geometry of space-time is governed by the matter-energy content of the Universe, given by Einstein's field equation

$$G_{\mu\nu} + g_{\mu\nu}\Lambda = 8\pi G T_{\mu\nu} \quad (1.6)$$

Here $G_{\mu\nu}$ is the Einstein tensor, $g_{\mu\nu}$ is the metric tensor, G is the universal gravitational constant, $T_{\mu\nu}$ is the energy-momentum tensor and Λ is the cosmological constant. Using the $\{00\}^{th}$ (time) component of Einstein's equation with the FLRW metric, one obtains

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} - \frac{\Lambda}{3} \quad (1.7)$$

and using the $\{ii\}^{th}$ (spatial) components we obtain

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} = -8\pi GP + \Lambda \quad (1.8)$$

Equations 1.7 and 1.8 known as the Friedmann equations. $\rho(t)$ and $P(t)$ are the energy density and pressure, respectively. Here we have taken $c = 1$. The Friedmann equations and the equation of state determine the time dependence of $a(t)$, $\rho(t)$, $P(t)$. Using eqn. 1.7 and 1.8 we obtain

$$\frac{d}{da}(\rho a^3) = -3a^2 P \quad (1.9)$$

Given an equation of state, $P = \omega\rho$, eqn. 1.9 gives,

$$\begin{aligned} \rho(a) &\propto a^{-3(1+\omega)} \\ \rho(a) &= \rho_0 a^{-3(1+\omega)} \end{aligned} \quad (1.10)$$

For relativistic matter $\omega = \frac{1}{3}$, non relativistic matter $\omega \approx 0$, and for Λ , $\omega = -1$.

Eqn. 1.7 can be expressed as

$$\begin{aligned} \left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G}{3}\rho(t) - \frac{k}{a^2} \\ H^2(t) &= \frac{8\pi G}{3}\rho(t) - \frac{k}{a^2} \end{aligned} \quad (1.11)$$

One can solve for $a(t)$ if $\rho(t)$ is known. For the Λ CDM model $a(t)$ is monotonic with t , hence a can be used as a time variable.

We can define the critical density, ρ_c , as the density to keep the universe flat ($k = 0$), using eqn 1.11 :

$$\rho_c(t) = \rho(t)|_{k=0} = \frac{3H^2(t)}{8\pi G} \quad (1.12)$$

The present day value of the critical density is

$$\rho_{c,0} = \frac{3H_0^2}{8\pi G} \quad (1.13)$$

A dimensionless density parameter can be defined for the i^{th} component of energy density

-

$$\begin{aligned} \Omega_i(t) &= \frac{\rho_i(t)}{\rho_c} \\ \Rightarrow \Omega_i(a) &= \frac{\rho_i(a)}{\rho_c} \end{aligned} \quad (1.14)$$

Therefore eqn 1.10 becomes

$$\Omega(a) = \Omega_0 a^{-3(1+\omega)} \quad (1.15)$$

Now the total energy density, $\rho(a)$, is the sum of all components

$$\begin{aligned} \rho(a) &= \rho_{rel}(a) + \rho_m(a) + \rho_\Lambda(a) + \rho_k(a) \\ \text{i.e. } \Omega(a) &= \Omega_{rel}(a) + \Omega_m(a) + \Omega_\Lambda(a) + \Omega_k(a) \end{aligned} \quad (1.16)$$

where $\rho_{rel}(a)$ is the energy density of the relativistic particles, $\rho_m(a)$ is the matter density which includes ordinary baryonic matter ($\rho_b(a)$) and dark matter ($\rho_{dm}(a)$), $\rho_\Lambda(a)$ is the dark energy density and $\rho_k(a)$ is the density associated with curvature. Considering the equations of states for each of the component, equation 1.16 becomes

$$\Omega(a) = \Omega_{rel,0} a^{-4} + \Omega_{m,0} a^{-3} + \Omega_{\Lambda,0} + \Omega_{k,0} a^{-2} \quad (1.17)$$

where we have used the normalization, $a(t = t_0) \equiv a_0 = 1$. Using the Friedman equations (1.7 and 1.8) and equation 1.17, one gets

$$H^2(a) = \frac{\dot{a}^2}{a^2} = H_0^2 [\Omega_{rel,0} a^{-4} + \Omega_{m,0} a^{-3} + \Omega_{k,0} a^{-2} + \Omega_{\Lambda,0}] \quad (1.18)$$

The current constraints on the density parameters $[\Omega_m, \Omega_b, \Omega_\Lambda]$ are $[0.315 \pm 0.007, 0.049 \pm 0.0002, 0.685 \pm 0.007]$ [Planck 2018, Aghanim et al., 2020], which is consistent with a flat universe.

The distance between two objects is defined in terms of some observable quantities like luminosity, redshift, etc.

- *Redshift* (z)

Due to the relative motion between the observer and the emitter the wavelength of the emitted light (λ_e) will be different from the wavelength observed (λ_{obs}). This shift is quantified as redshift, z :

$$z = \frac{\lambda_{obs} - \lambda_e}{\lambda_e}$$

$$1 + z = \frac{\lambda_{obs}}{\lambda_e} \quad (1.19)$$

The observed wavelength get expanded by a factor of $\frac{a(t_{obs})}{a(t_e)}$ from time t_e to t_{obs}

$$\lambda_{obs} = \frac{a(t_{obs})}{a(t_e)} \lambda_e \quad (1.20)$$

For an observer from present time $a(t_{obs}) = 1$, and hence

$$1 + z = \frac{1}{a(t_e)} \quad (1.21)$$

- *Comoving distance* (χ)

The distance which does not change with time due to the expansion of the Universe is known as the comoving distance (χ). It is defined as the ratio of proper distance

and the scale factor:

$$\chi = \int_{t_e}^{t_{obs}} \frac{c dt}{a(t)} \quad (1.22)$$

where, t_e and t_{obs} are the emitted and observed time, respectively. The above equation can be written as

$$\begin{aligned} \chi &= \int_{t_e}^{t_{obs}} \frac{c}{a(t)} \frac{dt}{da} da = \int_{a_e}^{a_{obs}} \frac{c}{a \dot{a}} da \\ &= \int_{a_e}^{a_{obs}} \frac{c}{a^2 H(a)} da \end{aligned}$$

In terms of redshift,

$$\chi = \int_{z_{obs}}^{z_e} \frac{c dz}{H(z)} \quad (1.23)$$

Here, a_e and a_{obs} are the scale factors and z_e and z_{obs} are the redshifts corresponding to the emitter and observer, respectively. For present day observation, $a_{obs} = 1$ and $z_{obs} = 0$, and the comoving distance is

$$\chi = \int_{a_e}^1 \frac{c da}{a^2 H(a)} = \int_0^{z_e} \frac{c dz}{H(z)} \quad (1.24)$$

- *Angular diameter distance (D_A)*

Let us consider an object of constant proper length l perpendicular to the line of sight of the observer and the angular distance between the end points of that object is $d\theta$. Then using the small angle approximation ($d\theta \ll 1$), the angular diameter distance, D_A , of that object is defined as

$$D_A = \frac{l}{d\theta} \quad (1.25)$$

For an expanding universe, the length of the object is

$$l = a(t_e) \chi d\theta \quad (1.26)$$

where χ is the the comoving distance of that object from the observer and t_e is the emitted time. Then the angular diameter distance can be written as

$$D_A = a(t_e) \chi = \frac{\chi}{1+z} \quad (1.27)$$

where z is the redshift of the object.

- *Luminosity distance (D_L)*

In a static universe, the observed flux, S , in a time interval dt emitted from a source of intrinsic luminosity, L , can be written as

$$S = \frac{L}{4\pi D^2} \quad (1.28)$$

where D is the distance between the source and the observer.

In an expanding universe, the observed flux from an object from redshift z decreases by a factor of $(1+z)^2$ — i) due to the expansion of the Universe, the wavelength of the emitted photon increases making the energy to be decreased by $(1+z)$, and ii) the expansion of the Universe makes the detection time of the photons bigger by $(1+z)$. Therefore the above equation in an expanding universe can be written as

$$\begin{aligned} S &= \frac{L}{4\pi D^2 (1+z)^2} \\ &= \frac{L}{4\pi D_L^2} \end{aligned} \quad (1.29)$$

$D_L = D(1+z)$ is known as the luminosity distance, where $D = \chi$, the comoving distance.

The luminosity distance, angular diameter distance and comoving distance of an object relate as

$$D_L = \chi(1+z) = D_A(1+z)^2 \quad (1.30)$$

1.1 Linear Perturbation Theory

Now if the Universe was perfectly homogeneous and isotropic, we would not see any structures today. To explain the presence and distribution of structures like galaxies we see today, we need some deviations from complete uniformity. We believe that tiny quantum fluctuations in the early Universe are the seeds to the perturbations that grow due to gravitational instability to eventually form the structures we see today.

The mass distribution in the Universe on cosmological scales can be treated as a perfect fluid of density $\rho(\vec{x}, t)$, pressure P , velocity field $\vec{v} = \vec{v}_H + \vec{v}_P$, and gravitational potential ϕ . Ideal fluids satisfy the following equations

1. Continuity equation:
$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot (\rho \vec{v}) = 0 \quad (1.31)$$

2. Euler equation:
$$\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \vec{\nabla}) \vec{v} = -\frac{1}{\rho} \vec{\nabla} P - \vec{\nabla} \phi \quad (1.32)$$

3. Poisson equation:
$$\nabla^2 \phi = 4\pi G \rho \quad (1.33)$$

We can define the density contrast as

$$\delta(\vec{x}, t) = \frac{\rho(\vec{x}, t) - \bar{\rho}}{\bar{\rho}} \quad (1.34)$$

where $\bar{\rho}$ is the mean density. Converting the fluid equations (eqn. 1.31, 1.32, and 1.33) in comoving coordinates and substituting ρ by δ , one obtains the fluid equations in expanding universe as:

$$\frac{\partial \vec{v}}{\partial t} + \frac{1}{a} \vec{\nabla} \cdot (1 + \delta) \vec{v} = 0 \quad (1.35)$$

$$\frac{\partial \vec{v}}{\partial t} + \frac{1}{a} (\vec{v} \cdot \vec{\nabla}) \vec{v} + \frac{\dot{a}}{a} \vec{v} = -\frac{1}{\rho a} \vec{\nabla} P - \frac{1}{a} \vec{\nabla} \phi \quad (1.36)$$

For a homogeneous Universe, $\vec{v}_p = 0$, $\phi = 0$, $P = 0$. Very small deviations from this homogeneity gives the following linear perturbation equation

$$\frac{\partial^2 \delta}{\partial t^2} + 2 \frac{\dot{a}}{a} \frac{\partial \delta}{\partial t} = 4\pi G \rho_b \delta \quad (1.37)$$

This equation describes how or whether the gravitational instability leads to the growth of perturbations in the expanding universe. Now the density fluctuation is a function of space and time, $\delta(\vec{x}, t)$, and since equation 1.37 does not contain any spatial derivatives, the solutions can be written as $\delta(\vec{x}, t) = D(t) \bar{\delta}(x)$. Here, $\bar{\delta}(x)$ is a function of the spatial coordinates, and $D(t)$ satisfies the following equation -

$$\frac{\partial^2 D}{\partial t^2} + 2 \frac{\dot{a}}{a} \frac{\partial D}{\partial t} = 4\pi G \rho_b D \quad (1.38)$$

Equation 1.38 has two linearly independent solutions - one of them decreases with time (decaying mode) and another increases with time (growing mode). We will consider only the growing mode, denoted by $D_+(t)$, as the solution of equation 1.38. $H(t)$ is the solution for decaying mode. Therefore, the density contrast is

$$\delta(\vec{x}, t) = D_+(t) \bar{\delta}(x) \quad (1.39)$$

Using the Wronskian with $\chi = \frac{8}{3}\pi G \rho_b a^3 \frac{1}{a} + \frac{\Lambda a^2}{3} - k$ [Heath, 1977], the *growth factor*, $D_+(t)$, can be expressed as

$$D_+(t) = \frac{1}{a} \chi^{\frac{1}{2}} \int^a \frac{da}{\chi^{\frac{3}{2}}} \quad (1.40)$$

1.2 Two Point Correlation Function

The two point correlation function (2PCF) is the excess probability of finding a pair of objects at a particular separation, compared with that expected for a random distribution. For a homogeneous Poisson distribution the probability of finding an object in volume elements dV_1 and dV_2 at a separation \vec{r} is $dP_{12} = n^2 dV_1 dV_2$, where n is the mean number density. When the distribution is clustered, this probability becomes $dP_{12} = n^2 dV_1 dV_2 (1 + \xi(\vec{r}))$. $\xi(\vec{r})$ is known as the two point correlation function. It can also be defined as the joint

ensemble average of the density contrast at two different locations,

$$\xi(\vec{r}) = \xi(\vec{r}, \vec{r}') = \langle \delta(\vec{r}) \delta(\vec{r}') \rangle \quad (1.41)$$

Since we assume homogeneity and isotropy on large scales, we can write the density contrast $\delta(x)$ in terms of its Fourier components as

$$\begin{aligned} \delta(\vec{r}) &= \int \frac{d^3 k}{(2\pi)^3} \delta(\vec{k}) \exp(-i\vec{k} \cdot \vec{r}) \\ \text{where,} \quad \delta(\vec{k}) &= \int d^3 r \delta(\vec{r}) \exp(i\vec{k} \cdot \vec{r}) \end{aligned} \quad (1.42)$$

We can compute the correlation function in Fourier space,

$$\begin{aligned} \langle \delta(\vec{k}) \delta^*(\vec{k}') \rangle &= \int \int d^3 r d^3 r' \langle \delta(\vec{r}) \delta(\vec{r}') \rangle \exp(i\vec{k} \cdot \vec{r}) \exp(-i\vec{k}' \cdot \vec{r}') \\ &= (2\pi)^3 \delta_D^3(\vec{k} - \vec{k}') P(\vec{k}) \end{aligned} \quad (1.43)$$

where $P(\vec{k})$ is the power spectrum of density fluctuations, defined as the Fourier transform of the 2-point correlation function

$$P(\vec{k}) = \int d^3 r \xi(\vec{r}) \exp(i\vec{k} \cdot \vec{r}) \quad (1.44)$$

For an isotropic Universe, $\xi(\vec{r}) = \xi(r)$ and $P(\vec{k}) = P(k)$. Therefore, equation 1.44 can be evaluated as

$$\begin{aligned} \xi(\vec{r}) &= \frac{1}{(2\pi)^3} \int P(k) \frac{\sin kr}{kr} 4\pi k^2 dk \\ &= \int \Delta^2(k) \frac{\sin kr}{kr} d \ln k \end{aligned} \quad (1.45)$$

where $\Delta^2(k) = \frac{k^3 P(k)}{2\pi^2}$, is the dimensionless power spectrum.

We can define the variance of the density field filtered on a scale R enclosing a mass M as

$$\begin{aligned} \sigma^2(R) &= \frac{\langle (M - \langle M \rangle)^2 \rangle}{\langle M \rangle^2} \\ &= \frac{1}{2\pi^2} \int_0^\infty dk k^2 P(k) \left| \tilde{W}(k, R) \right|^2 \end{aligned} \quad (1.46)$$

where $\tilde{W}(k, R)$ is the window function. For a spherical-top hat window function,

$$\sigma^2(R) = \frac{1}{2\pi^2} \int_0^\infty dk k^2 P(k) \left(3 \frac{\sin kR - kR \cos kR}{(kR)^3} \right)^2 \quad (1.47)$$

Rewriting eqn. 1.46, one can see that $\Delta^2(k)$ is related to the mass variance

$$\sigma^2(R) = \int \Delta^2(k) \left| \tilde{W}(k, R) \right|^2 d \ln k \quad (1.48)$$

For a power law model, $P(k) \propto k^n$, and hence

$$\begin{aligned} \sigma^2(R) &\propto R^{-(n+3)} \\ &= \left(\frac{R}{R_{nl}} \right)^{-(n+3)} \\ \sigma^2(M) &= \left(\frac{M}{M_{nl}} \right)^{-(n+3)/3} \end{aligned} \quad (1.49)$$

Here, R_{nl} is the scale of non-linearity, which is identified as the scale at which the mass variance becomes unity.

1.3 Mass Function : The Press-Schechter Formalism

The mass function is defined as the underlying number density of objects having masses within a range $[M, M + dM]$, expressed as

$$\phi(M) dM = \frac{dn}{dM} dM = \frac{dn}{d \log M} d \log M \quad (1.50)$$

The Press-Schechter Formalism [Press and Schechter, 1974] is an analytical approach to predict the mass function. It is based on the linear perturbation theory and the spherical collapse model. It estimates the number density by identifying the fraction of mass $F(M)$ in collapsed objects more massive than some mass M , at any time t , with the fraction of volume samples in which the smoothed initial density contrast are above some critical density contrast, i.e., $\delta > \delta_c$.

The spherical top-hat model describes how a spherical overdensity collapses by considering the evolution of a sphere with a constant density ρ in a uniform background of density $\bar{\rho}$ and $\rho > \bar{\rho}$. The overdensity expands initially with the expanding universe, then eventually pulls away from the background expansion, after reaching a maximum size when the inward gravitational pull dominates over the outward expansion. It turns around and starts to collapse. However it is assumed that the overdensity virializes instead of collapsing to a singularity. And the virialization occurs at half of the turn around radius value. In an Einstein-de Sitter Universe, the density contrast at virialization is $\delta_{vir} \approx 178$. The linear theory approximation of this value is $\delta_c \approx 1.686$, which is considered as the critical density contrast in the Press-Schechter formalism.

Considering the initial density field to be a Gaussian random field, the probability of finding a spherical region with density contrast in between δ and $\delta + d\delta$ can be expressed as

$$P(\delta, R) d\delta = \frac{1}{\sqrt{2\pi} \sigma^2(R)} \exp\left(-\frac{\delta^2}{2\sigma^2(R)}\right) d\delta \quad (1.51)$$

Now the fraction of mass $F(M)$ in the collapsed objects with mass greater than M is the probability that δ exceeds the critical density contrast δ_c , expressed as

$$\begin{aligned} F(M) &= \int_{\delta_c}^{\infty} P(\delta, R) d\delta \\ &= \int_{\delta_c}^{\infty} \frac{1}{\sqrt{2\pi} \sigma^2(R)} \exp\left(-\frac{\delta^2}{2\sigma^2(R)}\right) d\delta \end{aligned} \quad (1.52)$$

$$\begin{aligned} &= \frac{1}{2} \operatorname{Erfc}\left[\frac{\delta_c}{\sqrt{(2\sigma^2(R))}}\right] \\ &= \frac{1}{2} \operatorname{Erfc}\left[\frac{\nu}{\sqrt{2}}\right] \end{aligned} \quad (1.53)$$

where $\nu = \frac{\delta_c}{\sigma(R)}$, and $\operatorname{Erfc}(x)$ is the complimentary error function.

This formalism underestimates the mass function $F(M)$ by a factor of two as it does not account for the masses in underdense regions. Hence the correct mass function will be given

by multiplying eqn 1.53 by two

$$F(M) = \operatorname{Erfc} \left[\frac{\nu}{\sqrt{2}} \right] \quad (1.54)$$

The excursion set formalism for mass function [Bond et al., 1991] automatically takes care of this problem. The fraction of mass $g(M) dM$, in the objects with mass range $[M, M + dM]$, can be expressed as

$$\begin{aligned} g(M) dM &= \frac{\partial F(M)}{\partial M} dM \\ &= \sqrt{\frac{2}{\pi}} \left| \frac{d \ln \sigma}{dM} \right| \nu \exp \left(-\frac{\nu^2}{2} \right) dM \end{aligned} \quad (1.55)$$

And, then the number density becomes

$$\begin{aligned} N(M) dM &= \frac{\rho}{M} g(M) dM \\ &= \sqrt{\frac{2}{\pi}} \frac{\rho}{M} \left| \frac{d \ln \sigma}{dM} \right| \nu \exp \left(-\frac{\nu^2}{2} \right) dM \\ &= \sqrt{\frac{2}{\pi}} \frac{\rho}{M^2} \left| \frac{d \ln \sigma}{d \ln M} \right| \nu \exp \left(-\frac{\nu^2}{2} \right) dM \end{aligned} \quad (1.56)$$

For a power law model, using equation 1.49, the above equation reduces to

$$N(M) dM = \frac{1}{\sqrt{2\pi}} \frac{\rho}{M^2} \frac{n+3}{3} \left(\frac{M}{M_{nl}} \right)^{\frac{n+3}{6}} \exp \left(-\frac{1}{2} \left(\frac{M}{M_{nl}} \right)^{\frac{n+3}{3}} \right) \quad (1.57)$$

Refinements on the Press-Schechter theory have been done by Sheth and Tormen [1999], Sheth et al. [2001] considering the ellipsoidal collapse model. Although we do not have a complete theory for galaxy formation, observationally we see that the distribution function associated with observations like the galaxy stellar mass function, the galaxy luminosity function, etc., can be parametrized using a Schechter-like function.

1.4 Neutral Atomic Hydrogen - HI

The Universe is ~ 14 billion years old today and it has been expanding since the big bang. This expansion is governed by the relative contribution of radiation, matter, and dark en-

ergy present at that epoch. In the first phase the Universe is radiation dominated, and after redshift $z = z_{eq} \sim 3400$ the universe become matter dominated. Currently the Universe is dominated by dark energy, Λ .

The early Universe was very hot and dense, and the radiation in such an environment instantly ionizes neutral atoms. As it expands, it cools down and when it cools sufficiently radiation decouples from matter, and the Universe becomes neutral. This happens at redshift of $z = z_{rec} \sim 1000$. The decoupled photons then travel through space and we see them as the cosmic microwave background [Penzias and Wilson, 1965].

The Big-Bang Nucleosynthesis theory [Gamow, 1946, Alpher et al., 1948] predicts that about 3/4 of the total baryonic mass in the Universe is hydrogen, and about 1/4 is helium and the remaining small fraction is in light elements heavier than helium. Most of the elements heavier than lithium are produced in stellar nucleosynthesis. Currently about 1% – 2% are metals, where metals are defined as anything heavier than helium. Post-recombination the Universe is in a matter dominated phase. During this time density perturbations grow due to gravitational instability. At about redshift of $z = 30$ gas is able to cool and collapse and form the first stars in significant numbers [Bromm et al., 1999, Abel et al., 2000, Nakamura and Umemura, 2001, Yoshida et al., 2003]. Based on the Λ CDM model the first halos, which host galaxies, formed at redshift of $z \sim 12$ [Reed et al., 2003, 2007, Matteo et al., 2012]. The formation of first galaxies in turn leads to the emission of UV radiation. The radiation coming from these luminous structures ionizes their neighborhood (re-ionization phase) and eventually the Inter-Galactic medium (IGM). This process of reionization [Loeb and Barkana, 2001] starts around the most massive dark matter halos which host these galaxies and by redshift of $z = 6$ the Universe is re-ionized [Becker et al., 2001, Fan et al., 2006a,b]. From redshift of $z = 6$ to today the Universe is mostly ionized. One part in 10^3 is however neutral. The neutral gas is mostly locked up in galaxies since the recombination rate is greater than the ionization rate inside the galaxies due to their larger

densities. In this thesis we will be looking at how this neutral atomic hydrogen is distributed among different galaxy populations in the post-reionization epoch.

1.4.1 21 cm radiation from neutral Hydrogen (HI line)

The spin-flip transition from the hyper-fine levels of atomic hydrogen (HI) (energy difference $E = 5.9 \times 10^{-6}$ eV) corresponds to a wavelength of $\lambda_{21} = 21$ cm, frequency of $\nu_{21} = 1420$ MHz, or a temperature of $T_{21} = \frac{E}{k_B} \approx 0.068$ K. This radiation falls in the radio range of the electromagnetic spectrum. Though it is a very weak signal (half-life ~ 11 million years), hydrogen being the most abundant element in the Universe, 21 cm line is a useful probe of cool gas in galaxies [Oort, 1941, Ewen and Purcell, 1951, Muller and Oort, 1951].

The interaction between radiation and matter can be understood by the radiative transfer equation [Rybicki and Lightman, 1985]

$$\frac{d I_\nu}{ds} = j_\nu - \alpha_\nu I_\nu \quad (1.58)$$

This expresses the rate of change of brightness (specific intensity), which is defined as the energy radiated along a given direction per unit time per unit area per unit solid angle and per unit frequency (I_ν , unit: $\text{ergs s}^{-1} \text{cm}^{-2} \text{sr}^{-1} \text{Hz}^{-1}$), along the ray's path; the angle averaged specific intensity is the flux associated with a frequency ν , $F_\nu = \int I_\nu d\Omega$ and total flux is $F = \int F_\nu d\nu$. j_ν in eqn 1.58 is the emissivity or emission coefficient, the energy emitted spontaneously per unit frequency per unit time per unit mass and α_ν is the absorption coefficient, the loss of intensity in a beam as it travels a distance ds .

Considering the case where the radiation passes through a cloud where matter only absorbs, but does not emit, the radiative transfer equation (eqn 1.58) becomes

$$\frac{d I_\nu}{ds} = -\alpha_\nu I_\nu \quad (1.59)$$

Integration of this over the ray's path gives

$$I_\nu(\tau_\nu) = I_\nu(0) \exp(-\tau_\nu) \quad (1.60)$$

where $\tau_\nu = \int \alpha_\nu ds$ is the optical depth along the path of the ray. If the optical depth $\tau_\nu \gg 1$, the medium is called an optically thick medium. An average photon of frequency ν cannot pass over an optically thick medium without getting absorbed. On the other hand, a medium is known as optically thin or transparent if $\tau_\nu \ll 1$. In an optically thin medium a typical photon of frequency ν can traverse without being absorbed.

The specific intensity or brightness (I_ν) at a certain frequency can also be characterized as the temperature (T_b) of a blackbody having the same brightness at that frequency ($I_\nu = B_\nu(T_b)$), called the brightness temperature (T_b). Using the Rayleigh-Jeans low frequency limit relevant for the 21 cm line, the brightness temperature is expressed as

$$T_b(\nu) = \frac{c^2 I_\nu}{2\nu^2 k_B} \quad (1.61)$$

where c is the speed of light and k_B is the Boltzmann constant.

The radiative transfer equation can be rearranged as

$$\frac{dI_\nu}{d\tau_\nu} = S_\nu - I_\nu \quad (1.62)$$

where, $S_\nu = \frac{j_\nu}{\alpha_\nu}$, called the source function. Rewriting eqn 1.62 in terms of T_b , eqn 1.61, the radiative transfer equation becomes

$$\frac{dT_b}{d\tau_\nu} = T - T_b \quad (1.63)$$

where $T = \frac{c^2}{2\nu^2 k} S_\nu$, the temperature of the material.

Upon integration from 0 to τ_ν we obtain

$$T_b(\nu) = T_b(0)e^{-\tau_\nu} + T(1 - e^{-\tau_\nu}) \quad (1.64)$$

$T_b(0)$ is the background temperature which in our case is the brightness temperature of the CMB, T_{CMB} . And the temperature in the second term of eqn 1.64 is the excitation temperature or the spin temperature T_S for the case of the 21 cm line. The spin temperature quantifies the relative number densities of atoms in the two hyperfine levels

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} \exp\left(-\frac{E_{10}}{k_B T_S}\right) \quad (1.65)$$

where n_1 and n_0 are the number densities of atoms in the excited (triplet) and ground states, respectively and g_1 and g_0 are their statistical weights.

Eqn 1.64 becomes,

$$T_b(\nu) = T_{CMB} e^{-\tau_\nu} + T_S (1 - e^{-\tau_\nu}) \quad (1.66)$$

Here, $\exp(-\tau_\nu)$ gives the transmission probability of the background radiation, whereas $(1 - \exp(-\tau_\nu))$ represents the emission probability of the 21 cm photons. Hence, the change in brightness temperature is

$$\begin{aligned} \Delta T &= T_b - T_{CMB} \\ &= (T_S - T_{CMB}) (1 - e^{-\tau_\nu}) \end{aligned} \quad (1.67)$$

In the observer's frame of reference, the above equation reduces to

$$\Delta T = \frac{(T_S - T_{CMB}) (1 - e^{-\tau_\nu})}{1 + z} \quad (1.68)$$

It can be inferred from equation 1.68 that for $T_S > T_{CMB}$, $\Delta T > 0$ and so 21 cm signal will appear as an emission line; for $T_S < T_{CMB}$, it leads to a net absorption ($\Delta T < 0$); and for $T_S = T_{CMB}$, the spin state will be in thermal equilibrium with the CMB resulting in no 21 cm signal.

1.4.2 Einstein Coefficients

The relation between emission and absorption was explained by Einstein by considering the interaction of radiation with an atomic system. If we consider a discrete two-state system with energies E_0 and $E_1 = E_0 + h\nu_0$ then one can identify three different processes —

- *Spontaneous Emission* — the system transitions from a higher energy state (E_1) to a lower energy state (E_0) by emitting a photon of energy $h\nu_0 = E_1 - E_0$. The transition probability per unit time for spontaneous emission is known as the Einstein's A-coefficient (A_{10}). If n_1 is the number density of atoms in the higher energy state then $n_1 A_{10}$ is the number of spontaneous transitions per second per unit volume.
- *Absorption* — the system goes to a higher energy state (E_1) from the lower one (E_0) by absorbing a photon of energy $h\nu_0$. The line profile, $\phi(\nu)$, describes the distribution of absorption or emission around the transition frequency ν_0 and

$$\int_0^{\infty} \phi(\nu) d\nu = 1 \quad (1.69)$$

The line profile function represents the probability of getting a photon around ν_0 . The transition probability per unit time for absorption $\propto \int J_\nu \phi(\nu) d\nu = B_{01} \bar{J}$, and the number of absorbed photons is $n_0 B_{01} \bar{J}$ where \bar{J} is the average energy density of the radiation field, B_{01} is the Einstein's B-coefficient and n_0 is the number density of atoms in the lower energy state.

- *Stimulated Emission* — an incoming photon of energy $h\nu_0 = E_1 - E_0$ makes the system go down to a lower energy level (E_0) from a higher energy level (E_1). The transition probability per unit time for the stimulated emission is expressed as $B_{10} \bar{J}$, B_{10} is another B-coefficient. And the number of photons emitted by stimulated emission is $n_1 B_{10} \bar{J}$

If we assume thermodynamic equilibrium between the radiation bath and the two-level systems the number of excitations and de-excitations are in balance. Therefore,

$$n_1 A_{10} + n_1 B_{10} \bar{J} = n_0 B_{01} \bar{J} \quad (1.70)$$

$$\begin{aligned} \bar{J} &= \frac{n_1 A_{10}}{n_0 B_{01} - n_1 B_{10}} \\ &= \frac{A_{10}}{\frac{n_0}{n_1} B_{01} - B_{10}} \end{aligned} \quad (1.71)$$

According to the Boltzmann distribution

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} \exp\left(-\frac{h\nu_0}{k_B T}\right) \quad (1.72)$$

Eqn 1.70 becomes,

$$\bar{J} = \frac{A_{10}}{\frac{g_0}{g_1} \exp\left(\frac{h\nu_0}{k_B T}\right) B_{01} - B_{10}} \quad (1.73)$$

In thermal equilibrium \bar{J} must be given by the Planck function of the radiation bath:

$$\bar{J} = B_\nu(T) = \frac{2 h \nu_0^3}{c^2} \frac{1}{\exp\left(\frac{h\nu_0}{k_B T}\right) - 1} \quad (1.74)$$

which implies

$$B_{10} = \frac{g_0}{g_1} B_{01} \quad \text{and} \quad A_{10} = \frac{2 h \nu_0^3}{c^2} B_{10} \quad (1.75)$$

1.4.3 Absorption and Emission Coefficients in Terms of Einstein Coefficients

Assuming the emission to be distributed in accordance with the line profile $\phi(\nu)$, the energy emitted and distributed by the photon is $\frac{h\nu_0}{4\pi} n_1 A_{10} \phi(\nu) d\nu d\Omega dV dt$. On the other hand by the definition of emissivity, the amount of energy emitted in volume dV , solid angle $d\Omega$, frequency range $d\nu$ and time dt can be expressed as $j_\nu d\nu d\Omega dV dt$. Therefore the emission coefficient is

$$j_\nu = \frac{h\nu_0}{4\pi} n_1 A_{10} \phi(\nu) \quad (1.76)$$

Again the energy absorbed in time dt , volume dV , solid angle $d\Omega$ and frequency range $d\nu$ is $n_0 \frac{h\nu_0}{4\pi} B_{01} I_\nu \phi(\nu) d\nu d\Omega dV dt$. And the absorption coefficient is defined as

$$dI_\nu = -\alpha_\nu I_\nu dS \quad (1.77)$$

Therefore,

$$\alpha_\nu = n_0 \frac{h\nu}{4\pi} B_{01} \phi(\nu) \quad (1.78)$$

Considering the contribution from stimulated emission, which can be considered as a negative absorption, the absorption coefficient can be expressed as

$$\alpha_\nu = \frac{h\nu}{4\pi} \phi(\nu) (n_0 B_{01} - n_1 B_{10}) \quad (1.79)$$

1.4.4 Einstein Coefficients, Flux and HI mass

Flux is defined as the energy passing through per unit area per unit time per unit frequency, $dE = F_\nu dA dt d\nu$. In terms of emission coefficient $dE = j_\nu dV d\Omega dt d\nu$.

Therefore,

$$\begin{aligned} F_\nu d\nu &= \frac{dE}{dA dt} \\ F_\nu &= j_\nu \frac{dV d\Omega}{dA} \end{aligned}$$

Now $d\Omega = \frac{dA}{D_L^2(z)}$, where $D_L(z)$ is the luminosity distance at redshift z . Then the flux density is

$$S_\nu = \frac{F_\nu}{dV} = \frac{j_\nu}{D_L^2(z)} \quad (1.80)$$

Now for $\frac{h\nu_0}{k_B T_S} \ll 1$ eqn 1.65 reduces to

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} \quad (1.81)$$

The statistical weights for 21 cm radiation are $g_0 = 1$ (singlet state) and $g_1 = 3$ (triplet state). Hence $n_0 + n_1 = 4n_0$. The emission coefficient can then be written as

$$\begin{aligned}
 j_\nu &= \frac{h\nu_0}{4\pi} A_{10} \frac{n_1}{n_0 + n_1} \frac{(n_0 + n_1) m_H}{m_H} \phi(\nu) \\
 &= \frac{n_1}{n_0 + n_1} \frac{h\nu_0}{4\pi} A_{10} \frac{M_{HI}}{m_H} \phi(\nu) \\
 &= \frac{3}{4} \frac{h\nu(1+z)}{4\pi} A_{10} \frac{M_{HI}}{m_H} \phi(\nu)
 \end{aligned} \tag{1.82}$$

where, m_H is the mass of a hydrogen atom and $M_{HI} = (n_0 + n_1)m_H$ is the total HI mass.

Therefore, the flux density can be written as

$$S_\nu = \frac{3}{4} \frac{h\nu(1+z)}{4\pi D_L^2(z)} A_{10} \frac{M_{HI}}{m_H} \phi(\nu) \tag{1.83}$$

And the integrated flux becomes

$$\begin{aligned}
 S_{int} &= \int S_\nu d\nu \\
 &= \frac{3}{4} \frac{h\nu(1+z)}{4\pi D_L^2(z)} A_{10} \frac{M_{HI}}{m_H}
 \end{aligned} \tag{1.84}$$

Hence the HI mass is expressed as

$$M_{HI} = \frac{16\pi m_H}{3h\nu_0 A_{10}} D_L^2 S_{int} \tag{1.85}$$

For 21 cm transition $A_{10} = 2.85 \times 10^{-15} \text{s}^{-1}$ [Wild, 1952, Roberts, 1962, Furlanetto et al., 2006], which gives

$$M_{HI} = 2.356 \times 10^5 D_L^2 S_{int} M_\odot \tag{1.86}$$

1.5 Structure of the Thesis

In this chapter we have briefly discussed the models and formalism to describe the expansion as well as the present state of the Universe. HI is the most abundant element in the Universe and also represents the cold gas, which is the fuel for star formation. Thus HI

is a good tracer to understand the process of galaxy formation and evolution in the post-reionization era. The effective approaches to quantify the amount of HI in the Universe are - 1) absorption line studies, 2) emission line studies (both targeted and blind), and 3) stacking analysis. The abundance of HI at higher redshift has been derived from the damped Lyman- α Absorption systems (DLAs) [Prochaska et al., 2005, Noterdaeme et al., 2012, Neeleman et al., 2016, Rao et al., 2017]. In the local Universe, blind HI surveys [HIPASS, Meyer et al., 2004] [ALFALFA, Giovanelli et al., 2005] give an estimate of the HI mass function as well as HI density parameter analyzing HI emission data [Zwaan et al., 2003, Zwaan et al., 2005, Martin et al., 2010, Haynes et al., 2011, Jones et al., 2018]. Some targeted HI observations [GASS, Schiminovich et al., 2010, Catinella et al., 2013] [ATLAS^{3D} HI survey, Cappellari et al., 2011, Serra et al., 2012] have also been done in local Universe. For the intermediate redshift range, HI spectral stacking techniques are used [Lah et al., 2007, Rhee et al., 2013, Kanekar et al., 2016, Chowdhury et al., 2020] to give the estimate of HI content in the Universe.

HI gas directly traces the gravitational potential. Therefore, to understand the distribution of halo mass, HI rotation curves lead an important role [Rubin et al., 1978, Bosma, 1981, van Albada et al., 1985]. SPARC [Lelli et al., 2016] provide HI rotation curves for 175 disc galaxies which are observed over ~ 30 years using Westerbork Synthesis Radio Telescope (WSRT), Australian Telescope Compact Array (ATCA), Very Large Array (VLA), and Giant Metrewave Radio Telescope (GMRT). These 175 rotation curves have been extracted from the following works: Begeman et al. [1991, 9], de Blok et al. [1996, 7], Sanders [1996, 17], van Zee et al. [1997, 4], Verdes-Montenegro et al. [1997, 1], Walsh et al. [1997, 1], Sanders and Verheijen [1998, 30], Verheijen and de Blok [1999, 1], Côté et al. [2000, 3], de Blok et al. [2001, 10], de Blok and Bosma [2002, 7], Fraternali et al. [2002, 2011, 3], Begum et al. [2003, 1], Simon et al. [2003, 1] Begum and Chengalur [2004, 1], Blais-Ouellette et al. [2004, 1], Gentile et al. [2004, 2007, 4], Barbieri et al. [2005, 1],

Battaglia et al. [2006, 1], Chemin et al. [2006, 1], Spekkens and Giovanelli [2006, 5], Keping et al. [2007, 1], Noordermeer et al. [2007, 12], Kuzio de Naray et al. [2008, 6], Swaters et al. [2009, 32], Trachternach et al. [2009, 2], Elson et al. [2010, 1], Lelli et al. [2012a,b, 2014, 8], Hallenbeck et al. [2014, 2], Richards et al. [2015, 1]. The number in the parentheses denotes the number of rotation curves extracted from the corresponding publication. The SPARC catalog have used these rotation curves to estimate the halo masses [Li et al., 2019b].

In this dissertation we look at the statistical properties of HI in terms of the HI mass function, the HI width function, and the HI velocity function of HI-selected galaxies from a blind extragalactic HI survey *ALFALFA* [Giovanelli et al., 2005, Haynes et al., 2011]. We also look at the distribution of HI in dark matter halos in terms of scaling relations between HI properties and halo mass using observational constraints.

The thesis is organized as follows. In chapter 2 we present the estimate of the HI mass function (HIMF) of HI selected galaxies, from the ALFALFA survey (40% data release). We define different populations of galaxies based on their color($u - r$) and magnitude(M_r) values and find their contribution to the total HIMF. We also report the HI density parameter (Ω_{HI}) estimates for these sub-samples and their contribution to the total HI budget ($\Omega_{\text{HI}}^{\text{total}}$). In chapter 3 we show the conditional HIMFs conditioned on ($u - r$) color and/or M_r magnitude and using these conditional HIMFs we present the underlying distribution of Ω_{HI} in the color-magnitude plane of galaxies. Chapter 4 presents the HI velocity width function (HIWF) and the HI velocity function (HIVF) of HI selected galaxies. Also the contribution of red and blue populations to the total HIWF and HIVF is discussed here. We outline a model to populate HI in dark matter halos motivated and calibrated from our work and recent results. We present the scaling relations between HI-halo mass, velocity width - halo mass, rotational velocity - halo mass using abundance matching techniques. Finally in chapter 5 we summarize our results. In Appendix A we elaborately discuss the steps

to obtain the HIMF/HIWF with the statistical tool known as the 2-Dimensional Step-Wise Maximum Likelihood (2DSWML) method. We also point out the sources of uncertainty while estimating the HIMF and HIWF, and discuss the estimation of errors on the HIMF and HIWF.

Throughout this thesis we consider a flat Lambda-Cold Dark Matter (Λ CDM) cosmology with the matter density parameter $\Omega_m = 0.3$, the dark energy density parameter $\Omega_\Lambda = 0.7$ and the Hubble parameter $H_0 = 100 h \text{ km.s}^{-1}\text{Mpc}^{-1}$, We follow the notation of [Zwaan et al. \[2010\]](#) to define a normalized Hubble constant, $h_{70} = h/0.7 = H_0/(70\text{km.s}^{-1}\text{Mpc}^{-1})$ which takes the value of 1 for our choice of $h = 0.7$.

Chapter 2

The Population of Galaxies that Contribute to The HI Mass Function

2.1 Introduction

Understanding the formation and evolution of galaxies requires a clear picture of how dark matter halos, which host galaxies, are supplied with cold gas which may eventually lead to star formation, how the process of star formation in galaxies depends on the local and global conditions; and finally how various feedback processes self regulate galaxy formation. Galaxies can basically be described with three properties: i) the cold neutral hydrogen gas mass (atomic [M_{HI}] + molecular [M_{H_2}]), ii) the stellar mass (M_{star}) and iii) the star formation rate (SFR). The amount of cold gas quantifies the fuel available at present for future star formation. The SFR is the rate of forming stars with the available supply of cold gas and M_{star} is the SFR integrated over time. At sub-kpc scales, the observed correlation between surface density of molecular hydrogen, Σ_{H_2} , and the SFR surface density, Σ_{SFR} , is stronger than that of the HI surface density, Σ_{HI} , and Σ_{SFR} [Bigiel et al., 2008, Leroy et al., 2008]. The HI is more diffuse whereas H_2 and star formation are relatively clumpier and concentrated within the optical radius. However HI gas cools, becomes denser and transforms to molecular gas, which eventually leads to gravitational collapse to form stars. Therefore, it is common to correlate the total gas (HI+ H_2) to the SFR, the so-called Kennicutt-Schmidt

law [Schmidt, 1959, 1963, Kennicutt, 1998, 1989], where $\Sigma_{\text{SFR}} \propto \Sigma_{\text{gas}}^{1.4}$.

Data from optical, UV and IR surveys over the past decade give a clearer picture of how galaxies, on average, have formed and evolved over the past ~ 12.5 billion years. The cosmic stellar mass density, ρ_* has increased monotonically by nearly 2.5 decades from redshift $z = 6$ to today Madau and Dickinson [2014]. These observations also tell us that the cosmic SFR density (SFRD) increases from $z = 6$ to a peak value at $z \simeq 2$ and then drops by a decade today [Le Floc’h et al., 2005, Madau and Dickinson, 2014, Casey et al., 2014]. Whereas the surveys targeting gas content of galaxies do not give a consistent picture with these observations. The 21cm line of HI is a weak line; its detection in emission is limited only to the local Universe ($z \leq 0.05$). For $0.4 < z < 5$ the HI content of the Universe is derived from the absorption line studies of damped Lyman- α systems (DLA) [Prochaska et al., 2005, Noterdaeme et al., 2012, Neeleman et al., 2016, Rao et al., 2017]. Combining the emission and absorption data one sees that Ω_{HI} increases monotonically from $z = 0$ to $z = 5$ Rhee et al. [2018]. The HI density at $z = 0$ is only 1.5 times smaller than that at $z = 2$ with the data consistent with a no-evolution picture. While the cosmic SFRD has decreased 10 fold in this interval, this decrease in SFRD is not commensurate with the depletion in ρ_{HI} .

Accurate estimates of the HIMF have been carried out with blind HI surveys [HIPASS, Meyer et al., 2004] [ALFALFA, Giovanelli et al., 2005] [Zwaan et al., 2003, Zwaan et al., 2005, Martin et al., 2010, Haynes et al., 2011, Jones et al., 2018] in the local Universe, which can be integrated to obtain the HI density parameter $\Omega_{\text{HI}} = \rho_{\text{HI}}/\rho_c$. These data have been further analyzed to understand the dependence of the abundance and distribution of HI selected galaxies on different galaxy properties and environments [Moorman et al., 2014]. The clustering measurements of HI selected galaxies in the ALFALFA survey [Papastergis et al., 2013, Guo et al., 2017] suggest that ALFALFA galaxies cluster weakly. Jones et al. [2018] report the dependence of the HIMF on the environment for ALFALFA galaxies,

and find a decrease in the low mass slope with increasing environment density. Whereas [Said et al. \[2019\]](#) report an increase in the low mass slope with increasing density in the HI Zone of Avoidance survey with the Parkes telescope. The study of the HI velocity width function (HIWF) for wall and void galaxies as well as red and blue galaxies [Moorman et al. \[2014\]](#), indicates a strong dependence on both environment and galaxy color. Similarly, [Zwaan et al. \[2003\]](#) have looked at the dependence of the HIMF on the morphology of galaxies and on the early-late type classification. In this chapter we inspect the dependence of HIMF on different populations of galaxies, where we classify galaxy populations on the basis of their color-magnitude values. This is similar in spirit to the analysis of [Zwaan et al. \[2003\]](#), [Moorman et al. \[2014\]](#) on HI surveys and of [Baldry et al. \[2004\]](#), [Drory et al. \[2009\]](#) on optical surveys.

This chapter is organized as follows. In section [2.2](#) we describe the ALFALFA catalog and we define our sample as well as sub-samples corresponding to different galaxy populations. In section [2.4](#) we give a brief description of the estimators of HIMF. In section [2.5](#) we show the results of the HIMF estimates for different populations and we present the contribution of those populations to the total HI budget. In section [2.6](#) we discuss our HIMF results and their implications. Finally we conclude in section [2.7](#) by summarizing the results.

2.2 Data

2.2.1 ALFALFA Survey

The Arecibo Legacy Fast ALFA (ALFALFA) survey is a blind extragalactic HI survey, which uses the 305 meter telescope located at the Arecibo observatory in Arecibo, Puerto Rico. This survey utilizes the seven-feed receiver, Arecibo L-band Feed Array (ALFA),

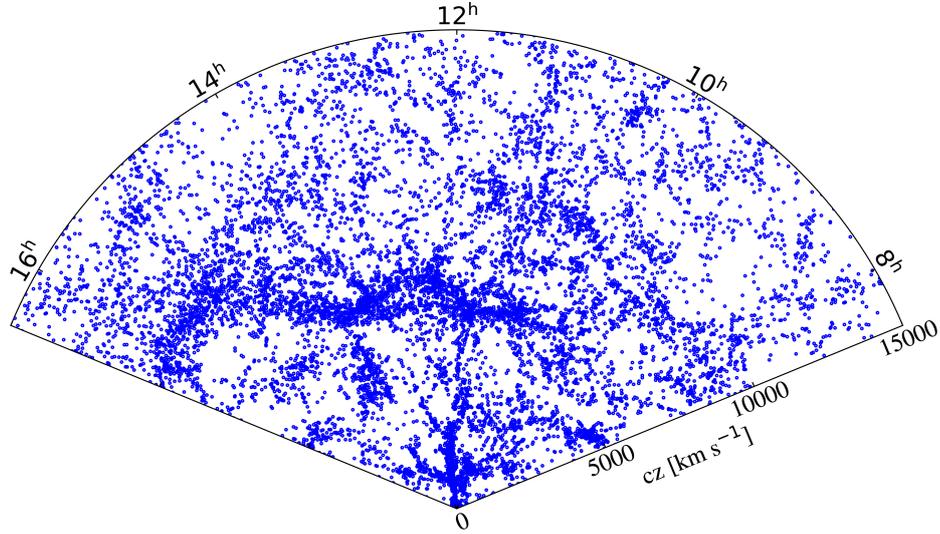


Figure 2.1: Cone-plot of $\alpha.40$ galaxies in Virgo direction region.

which detects signals at frequencies 1225 to 1425 MHz. The Arecibo telescope functions in a fixed azimuth drift scan mode in a two-pass strategy [Giovanelli et al., 2005]. Compared to the previous extragalactic HI surveys like HI Jodrell All Sky Survey (HIJASS) [Kilborn, 2002], HI Parkes All Sky Survey (HIPASS) [Meyer et al., 2004], ALFALFA delivers with 8 times better sensitivity and 4 times better angular resolution. At its final stage, ALFALFA has surveyed over $\sim 6500 \text{ deg}^2$ of the sky and detected ~ 30000 galaxies out to $cz = 18000 \text{ km.s}^{-1}$. Most of the ALFALFA galaxies have optical counterparts (OC) in Sloan Digital Sky Survey (SDSS) Data Release 7 (DR7) [Abazajian et al., 2009] and ALFALFA provides the DR7 objectID of those OCs. The ALFALFA sample contains some ($\sim 2\%$ of the total) HI sources, which do not have OCs in DR7, though they lie inside the DR7 footprint.

The ALFALFA galaxy catalog provides some observed quantities and some derived quantities. The observed quantities, contained in this catalog - i) an unique ID from Arecibo General Catalog (AGC); ii) right ascension (RA) and declination (dec) of the detected HI source and of the most probable OC from SDSS DR7; iii) the heliocentric velocity (cz_{helio}) in km.s^{-1} , the midpoint of the HI flux density profile,; iv) the velocity profile width (w_{50})

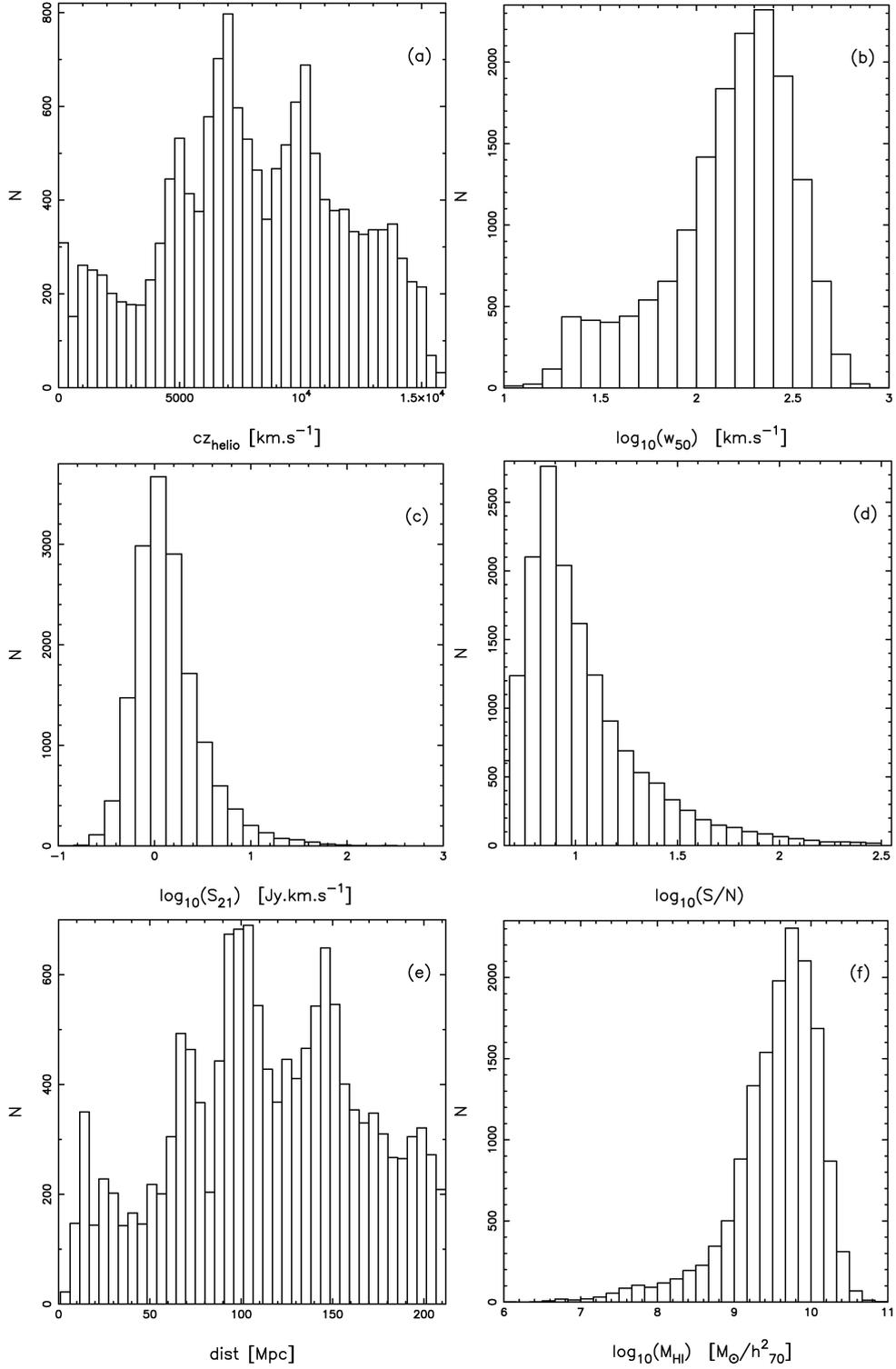


Figure 2.2: Number distribution of $\alpha.40$ galaxies as a function of a) heliocentric velocity (cZ_{helio}) b) velocity width (w_{50}) c) integrated flux (S_{21}) d) signal-to-noise ratio (SNR) e) distance (dist) f) HI mass (M_{HI}).

in km.s^{-1} , which is the full width at half maximum of the HI flux density profile; v) the integrated HI flux density (S_{21}) in Jy.km.s^{-1} . The catalog has the following derived quantities - i) the distance (D) to the object in Mpc. $D = cz_{cmb}/H_0$ for $cz_{cmb} > 6000 \text{ km.s}^{-1}$ where H_0 is the Hubble constant and cz_{cmb} is the velocity in reference frame of Cosmic Microwave Background (CMB), and for $cz_{cmb} < 6000 \text{ km.s}^{-1}$ D has been calculated using a local flow model [Masters, 2005]; ii) the HI mass (M_{HI}) in units of solar masses, computed as $M_{\text{HI}} = 2.356 * 10^5 \left(\frac{D}{\text{Mpc}} \right) \left(\frac{S_{21}}{\text{Jy.km.s}^{-1}} \right)$. The ALFALFA catalog also classifies the detections by their signal-to-noise ratio (SNR) and codes them as Code 1, 2 or 9 objects. Code 1 objects are the detections with $\text{SNR} > 6.5$, confirmed sources with $\text{SNR} < 6.5$ are referred as Code 2 objects and Code 9 objects are the high velocity clouds (HVC).

2.2.2 The Sample

As of doing this analysis the ALFALFA 100% catalog ($\alpha.100$) [Haynes et al., 2018] has been released, which includes the RA and dec of the OCs from SDSS DR7. But some of the $\alpha.100$ galaxies are affected by the presence of luminous foreground stars; because of which SDSS has masked the region and does not provide the photometric values. Due to the lack of the total optical information of $\alpha.100$ galaxies we restrict ourselves with the 40% catalog ($\alpha.40$).

ALFALFA has surveyed over a volume of $\sim 2.65 \times 10^6 \text{ Mpc}^3$ at the 40% data release, with 15855 galaxies over $\sim 2752 \text{ deg}^2$ of the sky. $\alpha.40$ covers the following four regions — $7^{\text{h}}30^{\text{m}} < \text{R.A.} < 16^{\text{h}}30^{\text{m}}$, $4^\circ < \text{dec.} < 16^\circ$, and $24^\circ < \text{dec.} < 28^\circ$ in the northern galactic hemisphere, known as the Virgo direction region (VDR) [shown in figure 2.1] and $22^{\text{h}} < \text{R.A.} < 3^{\text{h}}$, $14^\circ < \text{dec.} < 16^\circ$ and $24^\circ < \text{dec.} < 32^\circ$ in southern galactic hemisphere, named as the anti-Virgo direction region (aVDR). Some of the distributions of the objects are shown in figure 2.2, the histograms of a) cz_{helio} , b) w_{50} , c) S_{21} , d) SNR, e) distance in Mpc and f) M_{HI} are plotted. As expected most of these distributions are peaked due to the

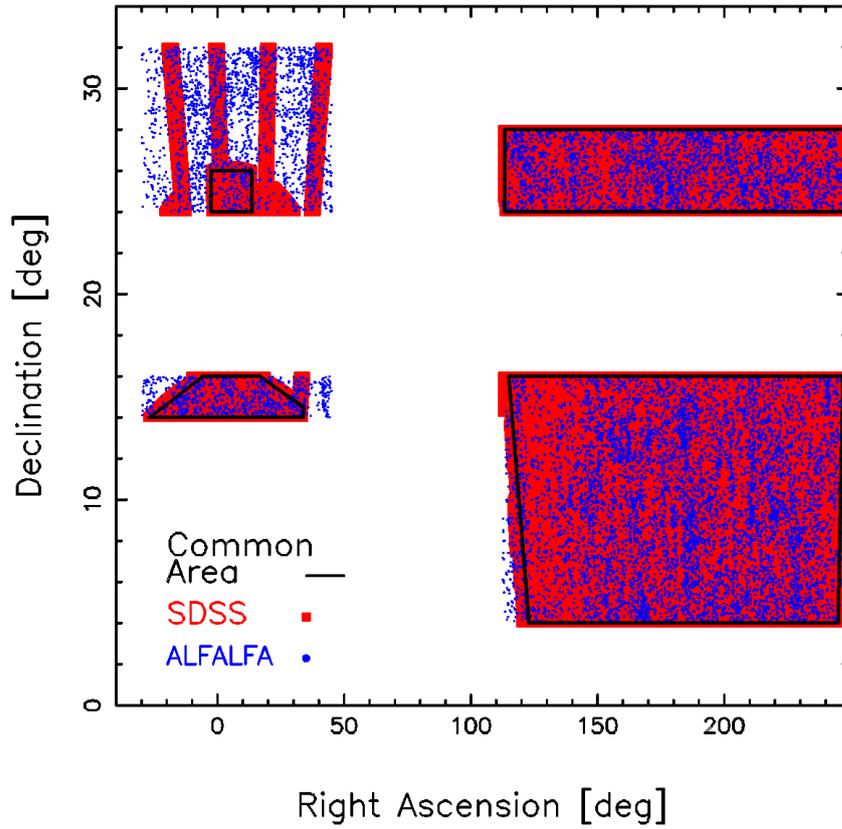


Figure 2.3: Common footprint of ALFALFA and SDSS used in this thesis. The red patches is that of SDSS DR7 overlapping with the ALFALFA survey region. The blue dots are the positions of ALFALFA galaxies. The common boundary used in this work is outlined by the thick black line.

selection function of ALFALFA. There are 11941 Code 1 galaxies, 3100 Code 2 galaxies and 814 HVCs in $\alpha.40$ catalog. We have considered only Code 1 galaxies with high SNR for our analysis. Another restriction we put on the sample at $cz_{\text{CMB}} = 15000\text{km.s}^{-1}$ to avoid radio frequency interference (RFI) caused by Federal Aviation Administration (FAA) radar at the San Juan airport [Martin et al., 2010, Haynes et al., 2011]. This limits the sample to redshift, $z \leq 0.05$ and the number of galaxies reduce to 10785.

We define a boundary, shown as the black line in figure 2.3, which is common to both the surveys, ALFALFA and SDSS, in this work. The final area in our analysis has four sub-regions (figure 2.3), whose vertices are given by the following (RA-Dec) values: (i)(123°, 4°), (245°, 4°), (247°, 16°), (115°, 16°); (ii) (113.31°, 24°), (247.5°, 24°), (247.5°, 28°), (113.31°, 28°); (iii) (-27°, 14°), (34°, 14°), (34°, 14.5°), (17°, 16°), (-6°, 16°); and (iv) (-2.6°, 24°), (13.6°, 24°), (13.6°, 26°), (-2.6°, 26°). These sub-regions sum to an angular area of $\sim 2093 \text{ deg}^2$, and total volume of these regions is $2.02 \times 10^6 \text{ Mpc}^3$. There are 8344 ALFALFA galaxies in these regions. We find 4 OH Megamasers [Suess et al., 2016] in our area of analysis, which have been excluded.

In figure 2.4 the distribution of Code 1 $\alpha.40$ galaxies from our area of analysis has been shown in the $S_{21} - w_{50}$ plane. We also divide the sample in 3 groups of HI mass and draw the 1σ contours of their distribution. These mass groups corresponds to the low-mass end (thick solid blue line), the high-mass end (thin-solid red line) and the knee (dot dashed black line) of the HIMF. These show that the velocity width on average increases with increasing HI mass, which has also been shown in Moorman et al. [2014]. Also it can be seen that low (high) mass galaxies are more likely to be detected with smaller(larger) velocity width, whereas the intermediate mass objects can be detected over the full range of velocity width. This conveys that the detection of sources depends not only on the integrated flux or HI mass but also on the HI velocity width. At a fix value of S_{21} , detection is more probable for

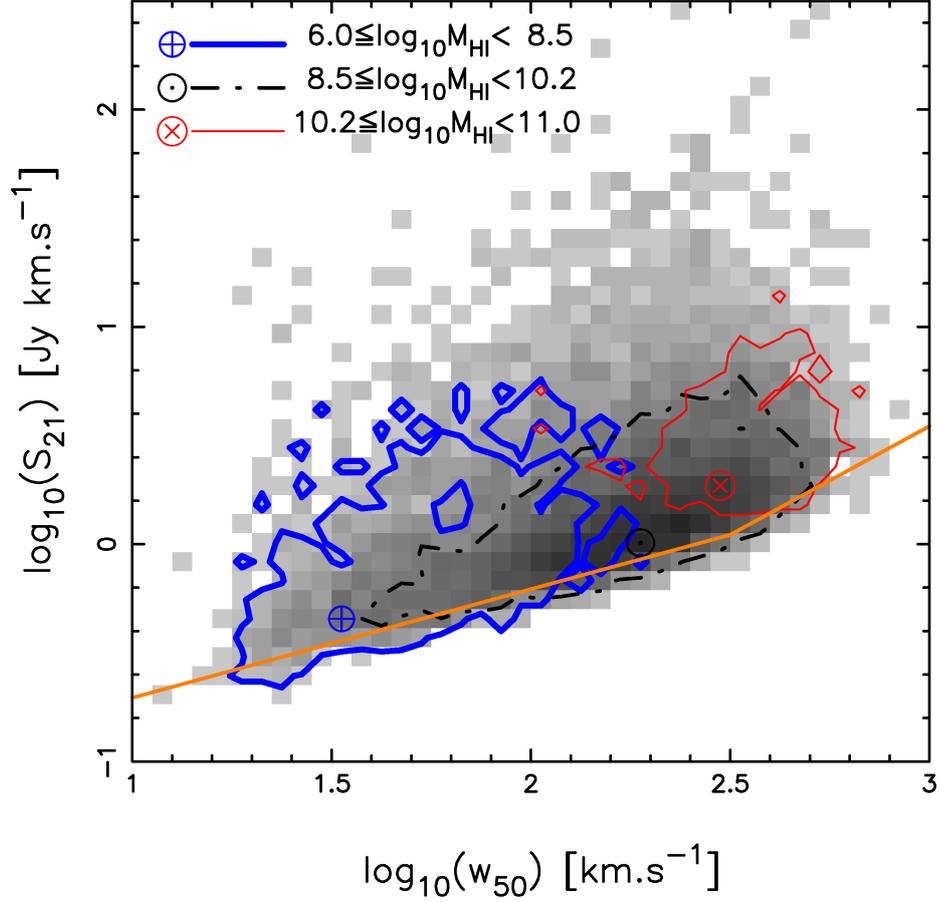


Figure 2.4: The distribution of galaxies in the $S_{21} - w_{50}$ plane is shown for all Code 1 ALFALFA galaxies. The broken orange solid line is the 50% completeness relation [Haynes et al., 2011]. It is given in eq. 2.1. We have further divided the sample into three mass bins: $\log_{10}[M_{\text{HI}}/M_{\odot}] \in [6.0, 8.5[, [8.5, 10.2[, [10.2, 11.0[$. The 1σ contours and the peaks of the distributions for these three populations are given by thick blue solid line (plus-circle), the dot-dashed black line (dot-circle) and the thin red solid line (cross-circle).

narrower HI velocity widths. The broken orange solid line in figure 2.4 is the sensitivity limit of the ALFALFA survey and is given by the 50% completeness relation of Code 1 objects, as expressed in eq. 2.1 [Haynes et al., 2011].

$$\log S_{21} = \begin{cases} 0.5 \log w_{50} - 1.207 & : \log w_{50} < 2.5 \\ \log w_{50} - 2.457 & : \log w_{50} \geq 2.5 \end{cases} \quad (2.1)$$

We apply the completeness cut (eq. 2.1) to our sample and consider the sources above the orange line shown in figure 2.4. This cut reduces the sample to 7857 galaxies. All these galaxies have photometric measurements in SDSS DR7, among them 6076 galaxies have spectroscopic measurements as well. The remaining 148 (2% of the total), are Dark galaxies, which are not being identified as potential galaxies in the SDSS pipeline. We have looked at the images of the dark galaxies in SDSS and do not find any visible object. Using the SDSS DR7 objectIDs we extract the photometric properties like model magnitudes (*ugriz* values) for these 7709 galaxies. We correct these magnitudes for the extinction due to our own galaxy [Schlegel et al., 1998] and *kcorrect* [Blanton and Roweis, 2007] them to obtain the rest frame magnitudes. We use the HI redshifts for the objects which do not have spectroscopic redshifts in DR7 to do the *k*correction. Additionally the *kcorrect* code provides some additional properties like stellar mass, metallicity and integrated star formation history of those objects.

Figure 2.5 shows the distribution of SDSS (left) and ALFALFA (right) galaxies, observed in our defined volume, in the color(*u* - *r*)-magnitude(*M_r*) plane. The 0.25 σ and 1 σ contours are drawn as the thick and thin lines. The observed distribution of SDSS galaxies shows a clear bimodality, whereas in the case of ALFALFA the presence of two peaks is not obvious. Baldry et al. [2004] describes a separator as expressed in eq. 2.2 (the solid black line in figure 2.5) to classify the red and blue galaxies. The number of galaxies detected in

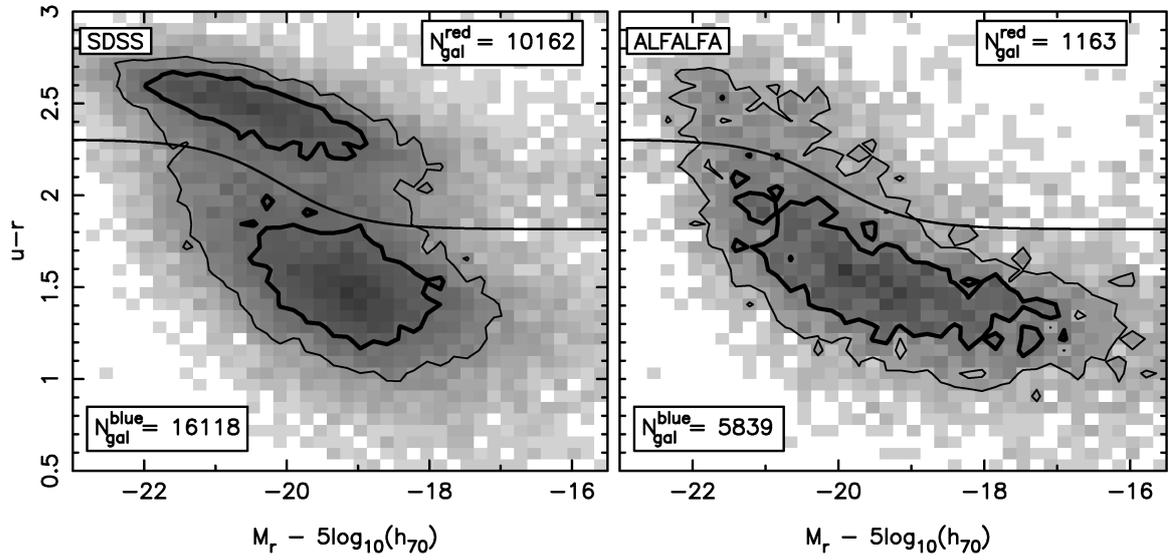


Figure 2.5: The observed distribution of galaxies in the color-magnitude plane for SDSS (left panel) and ALFALFA (right panel) in the common volume considered in this work. The 1σ and 0.25σ contours are given by the thin and thick lines. A bimodal distribution of galaxies is seen in SDSS. The solid black curve (eq. 2.2) is used to classify the galaxies into red (above curve) and blue (below curve) populations [Baldry et al., 2004]. The numbers in each panel indicate the observed counts of galaxies in this color-magnitude range for each of the populations.

SDSS and ALFALFA survey show that $\sim 98\%$ of ALFALFA galaxies have OCs in SDSS. We also see from the figure 2.5) that ALFALFA mainly samples the blue galaxy cloud - 38% of blue galaxies from SDSS have HI counterparts, whereas only 11% of the red SDSS galaxies are detected in ALFALFA.

2.3 Populations of Galaxies

We create sub-samples from our sample of 7857 galaxies, by putting conditions on the color($u - r$)-magnitude(M_r) values of the galaxies. We divide the color-magnitude plane into six disjoint regions, as shown in figure 2.6. We put the separator from Baldry et al. [2004] to demarcate red (above the separator) and blue (below the separator) populations. This optimal divider is expressed as

$$C'_{ur}(M_r) = 2.06 - 0.244 \tanh \left[\frac{M_r + 20.07}{1.09} \right] \quad (2.2)$$

The vertical solid line breaks up the full sample into luminous (leftward of the line) and faint (rightward to the line) populations. We choose this line in such a way that the fraction of luminous red galaxies over all red galaxies is 0.87, we refer to this as 1.5σ cut in magnitude. In a similar way we choose another line, parallel to the previous *tanh* line given by Baldry et al. [2004] (eqn. 2.2), such that the fraction of blue galaxies above that line over all blue galaxies is 0.87, 1.5σ cut in color. We call the galaxies above this curve as blue galaxies and the galaxies below this cut are referred as bluer galaxies. Similarly we define the 1σ sample (dashed lines in 2.6). In this way we get six disjoint samples of HI selected galaxies — 6 populations of galaxies — luminous red, faint red, luminous blue, faint blue, luminous bluer and faint bluer. Additionally we have the population of dark galaxies, having no OCs in SDSS DR7. Figure 2.6 shows the galaxies from different populations, based on 1.5σ definition, with different color and point types. Triangles, squares and circles represent the red, blue and bluer galaxies, respectively, and open and filled points represent faint and

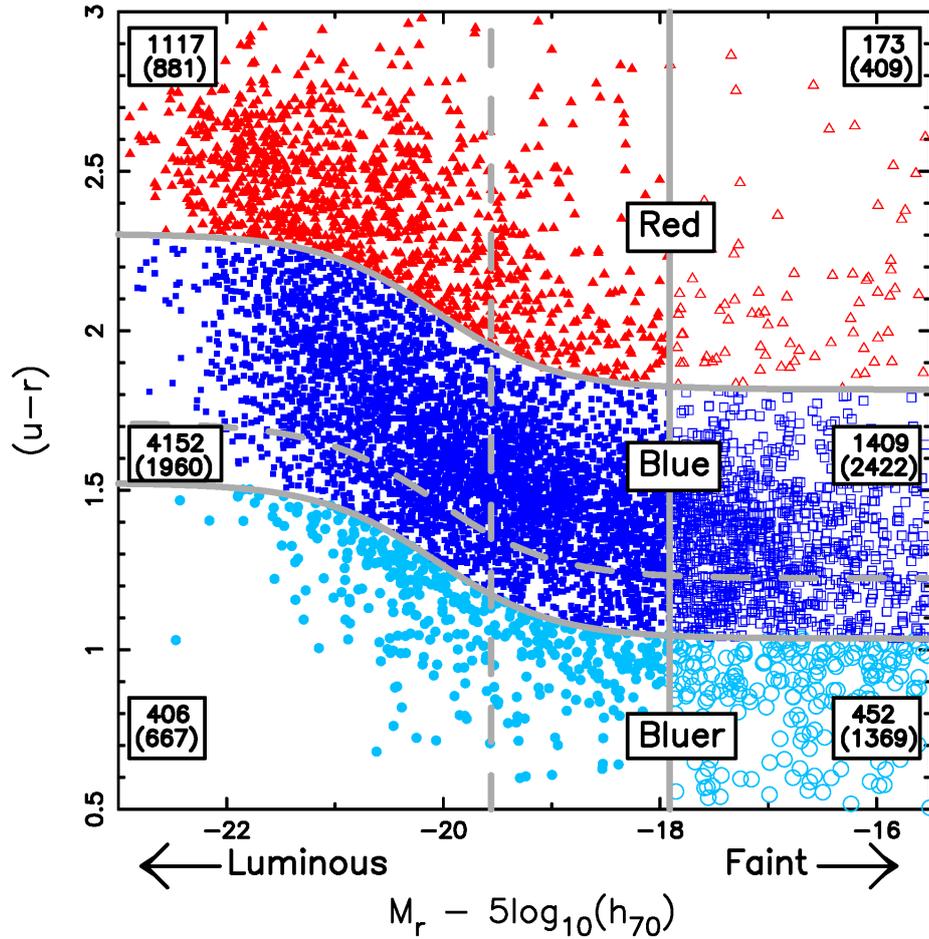


Figure 2.6: The distribution of HI selected galaxies in this work in the color-magnitude plane. The upper solid curve demarcates the red (above curve) from the blue (below curve) population as in Baldry et al. [2004]. The vertical solid (dashed) line divides the luminous (left) and faint (right) populations, referred as 1.5σ (1σ) cut in magnitude. The lower solid (dashed) line divides the blue population into bluer (below the curve) and blue (above the curve) populations, we call it 1.5σ (1σ) cut in color. This breaks the sample of HI selected galaxies into 6 disjoint sub-samples in the color-magnitude plane. The number for each population is quoted, where the numbers in brackets are for the 1σ sample.

luminous populations of galaxies, respectively. The number of galaxies for each of the population, in case of 1.5σ definition, is quoted in the figure 2.6, where the numbers in parentheses are that for the 1σ sample. Each population depends on the boundaries that define them, which we have named as the 1σ line or 1.5σ line; and the number of galaxies in a given population differs for the two different definitions ($1\sigma/1.5\sigma$) of populations. We consider two definitions of populations to show that the results do not change qualitatively with the sample definition. However looking at the differences in the observed counts, in 1σ and 1.5σ sample, changes can be expected for the HIMF of the luminous blue population.

2.4 The HI Mass Function (HIMF)

The HIMF, $\phi(M_{\text{HI}})$, is the underlying number density of galaxies in the mass range $[M_{\text{HI}}, M_{\text{HI}} + dM_{\text{HI}}]$, expressed as -

$$\phi(M_{\text{HI}}) = \frac{dN}{V dM_{\text{HI}}} \quad (2.3)$$

where, V is the survey volume of interest, dN is the number of galaxies in V , having HI masses in the range $[M_{\text{HI}}, M_{\text{HI}} + dM_{\text{HI}}]$. The HIMF can be parametrized as a three parameter Schechter Function

$$\phi(M_{\text{HI}}) dM_{\text{HI}} = \phi_* \left(\frac{M_{\text{HI}}}{M_*} \right)^\alpha \exp \left(-\frac{M_{\text{HI}}}{M_*} \right) dM_{\text{HI}} \quad (2.4)$$

where, ϕ_* is the amplitude, α is the faint end slope and M_* is the characteristic HI mass.

2.4.1 HIMF Estimators

$1/V_{\text{max}}$ Method

$1/V_{\text{max}}$ method Schmidt [1968] is the simplest way to calculate the HIMF. The basic assumption in this method is that the observed galaxies are being detected from a homoge-

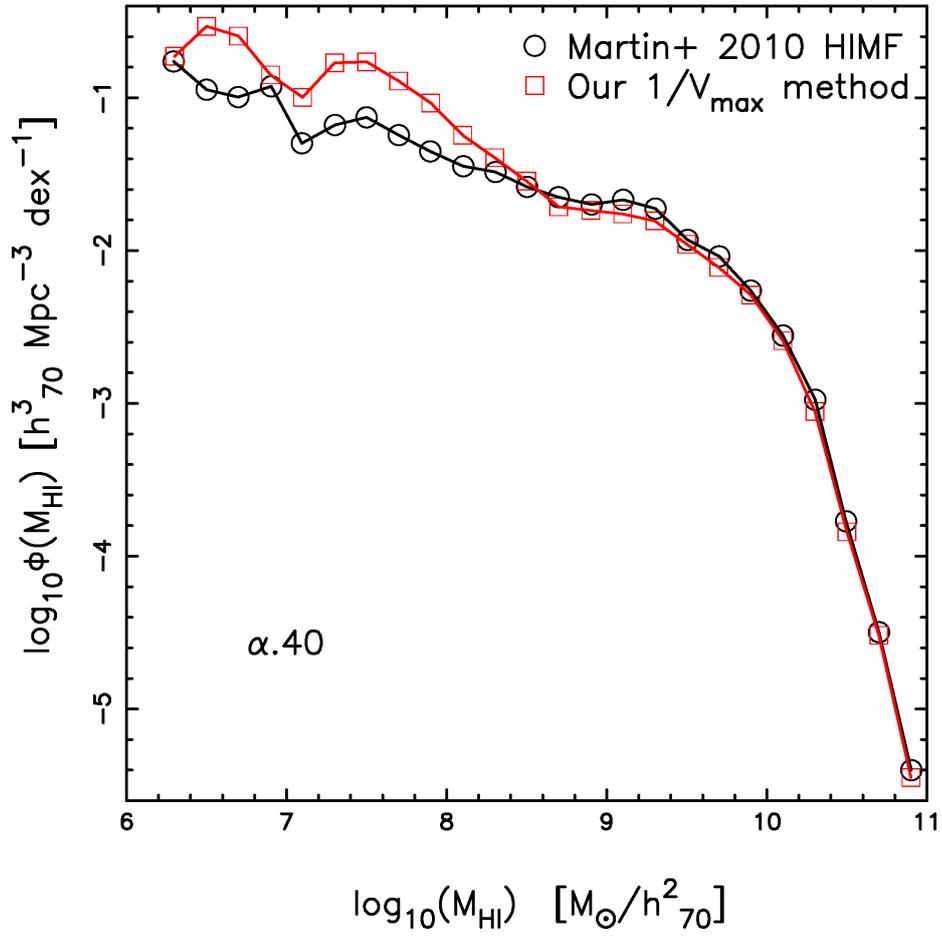


Figure 2.7: HIMFs estimated by $1/V_{\text{max}}$ method by [Martin et al., 2010](black circles) and our result of HIMF (red squares), where we do not use the flow model [Martin et al., 2010].

neous Universe. A maximum distance D_{max}^i for each of the observed galaxy 'i' is calculated, upto which the galaxy can be detected and still be inside the survey volume. These D_{max}^i s are estimated using the completeness relation 2.1, which determines the limiting flux S_{lim}^i for a galaxy of HI mass M_{HI}^i to be detected for a given velocity width w_{50}^i . D_{max}^i s are then converted to a V_{max}^i , and by construction $V_{max}^i \leq V_{survey}, \forall i$, where V_{survey} is the survey volume. Lastly the observed galaxies are binned with relative weights of $1/V_{max}^i$, which gives the desired mass function. The advantage of this method is — it is non-parametric, no prior knowledge is required to obtain the HIMF. But it is sensitive to the presence of large scale structure in the local volume, which can be corrected using local flow models [Martin et al., 2010]. Figure 2.7 shows the result we get from the $1/V_{max}$ method without correcting for the effects coming from the presence of large scale structure, in comparison with the Martin et al. [2010] HIMF where the local flow model has been applied. As will be seen later the features at lower masses in fig. 2.7 are less pronounced as compared to the maximum likelihood methods, which we discuss next.

Maximum Likelihood Method

On the other hand Maximum likelihood [Sandage et al., 1979] method is adapted to be insensitive to large scale structure. This works with the assumption that the galaxies are drawn from an underlying distribution function, and then the parameters of that chosen function are estimated by maximizing the likelihood. Although Schechter function is the most considered function, Baldry et al. [2012] and Drory et al. [2009] find that the stellar mass function (SMF) can not be described by a single Schechter function.

Step-Wise Maximum Likelihood Method

Step-Wise Maximum Likelihood (SWML) [Efstathiou et al., 1988] also assumes that the galaxy sample is drawn from an underlying distribution $\phi(M_{\text{HI}})$, which does not have a parametric form, but it is discretized in steps or grids of mass and for each mass grid a uniform distribution function is assumed. So ϕ_j , the value of ϕ in the j^{th} mass bin, becomes a parameter. Finally the maximization of the joint likelihood of detecting all galaxies with respect to the ϕ_j s gives the HIMF. This SWML method works for a flux-limited sample. When the detection of galaxies depends on other properties also, i.e., the survey selection function is multivariate, an underlying multivariate distribution has to be considered in that case. Loveday [2000] has presented a bivariate luminosity function $\phi(M_{\text{K}}, M_{\text{B}})$ and the K-Band luminosity function, $\phi(M_{\text{K}})$, which has been obtained by marginalizing over M_{B} , with a b_J -selected sample from the Stromlo-APM Redshift Survey.

2.5 Results

Figure 2.4 and eq. 2.1 shows that the detection of ALFALFA galaxies depends on the flux (or HI mass) as well as on the velocity width. Hence to obtain the HIMF one should apply a two-dimensional SWML (2DSWML) method, where the underlying bivariate distribution is $\phi(M_{\text{HI}}, w_{50})$, and marginalize over w_{50} [Zwaan et al., 2003, Martin et al., 2010, Haynes et al., 2011, Jones et al., 2018]. On the other hand $\phi(M_{\text{HI}}, w_{50})$ can be marginalized over M_{HI} to obtain the HI velocity width function (HIWF) [Zwaan et al., 2010, Moorman et al., 2014]. A detailed discussion on our implementation of the 2DSWML method can be found in appendix A.

Figure 2.8, 2.9, 2.10, and 2.11 shows the HIMF for our total sample (grey crosses) and the HIMFs for all the seven populations of galaxies, including the population of dark galax-

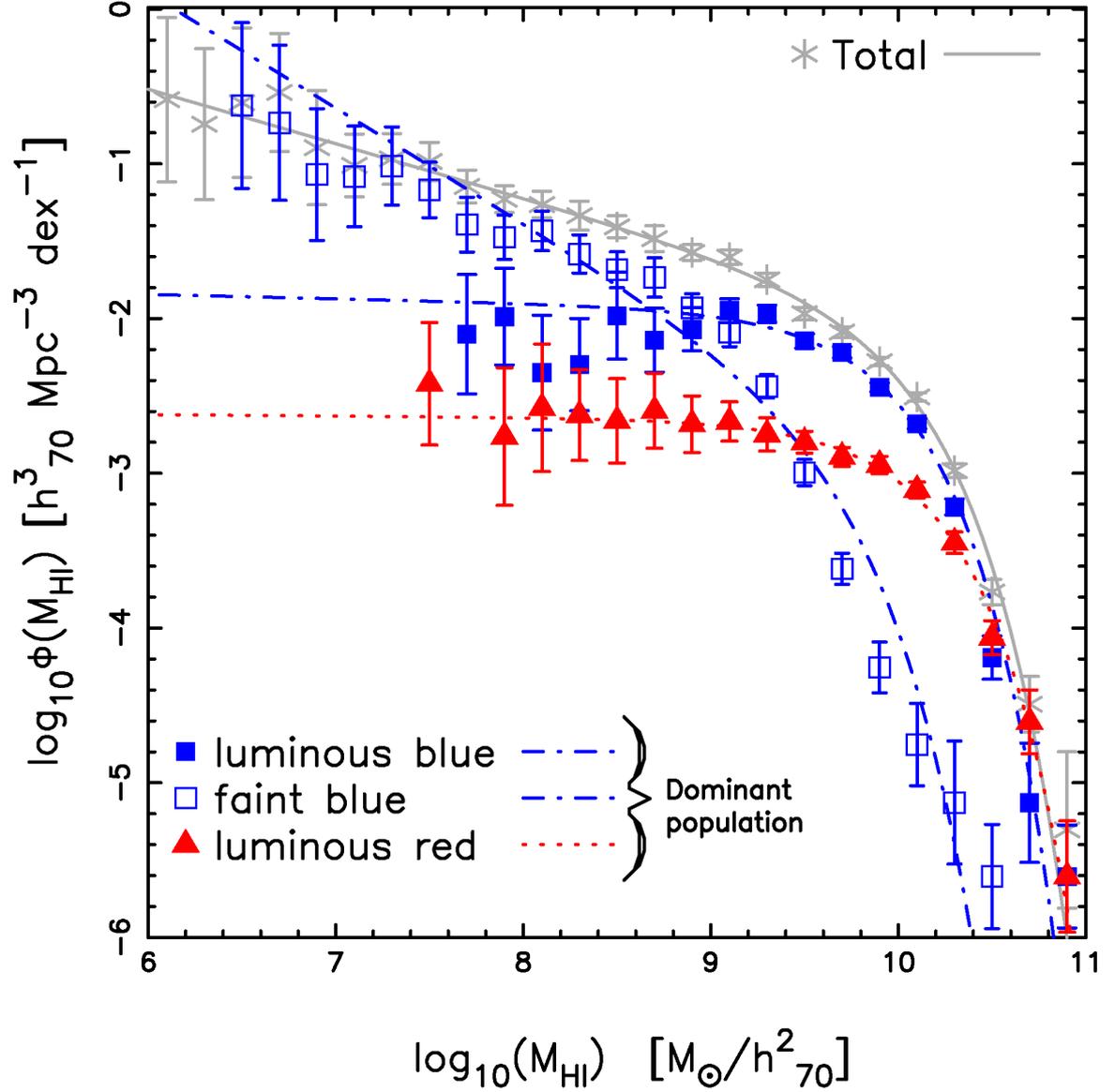


Figure 2.8: The HIMF for the luminous blue, luminous red, and faint blue populations (1.5σ definition). Data points and error bars were estimated using the 2DSWML method (see appendix A). The red dotted and blue dot-dashed curves are the Schechter function fits for red and blue populations, respectively. For comparison we have added the total HIMF (crosses) and its fit (solid line). The symbols for the populations are the same as in figure 2.6. The details of the fits are given in table 2.1.

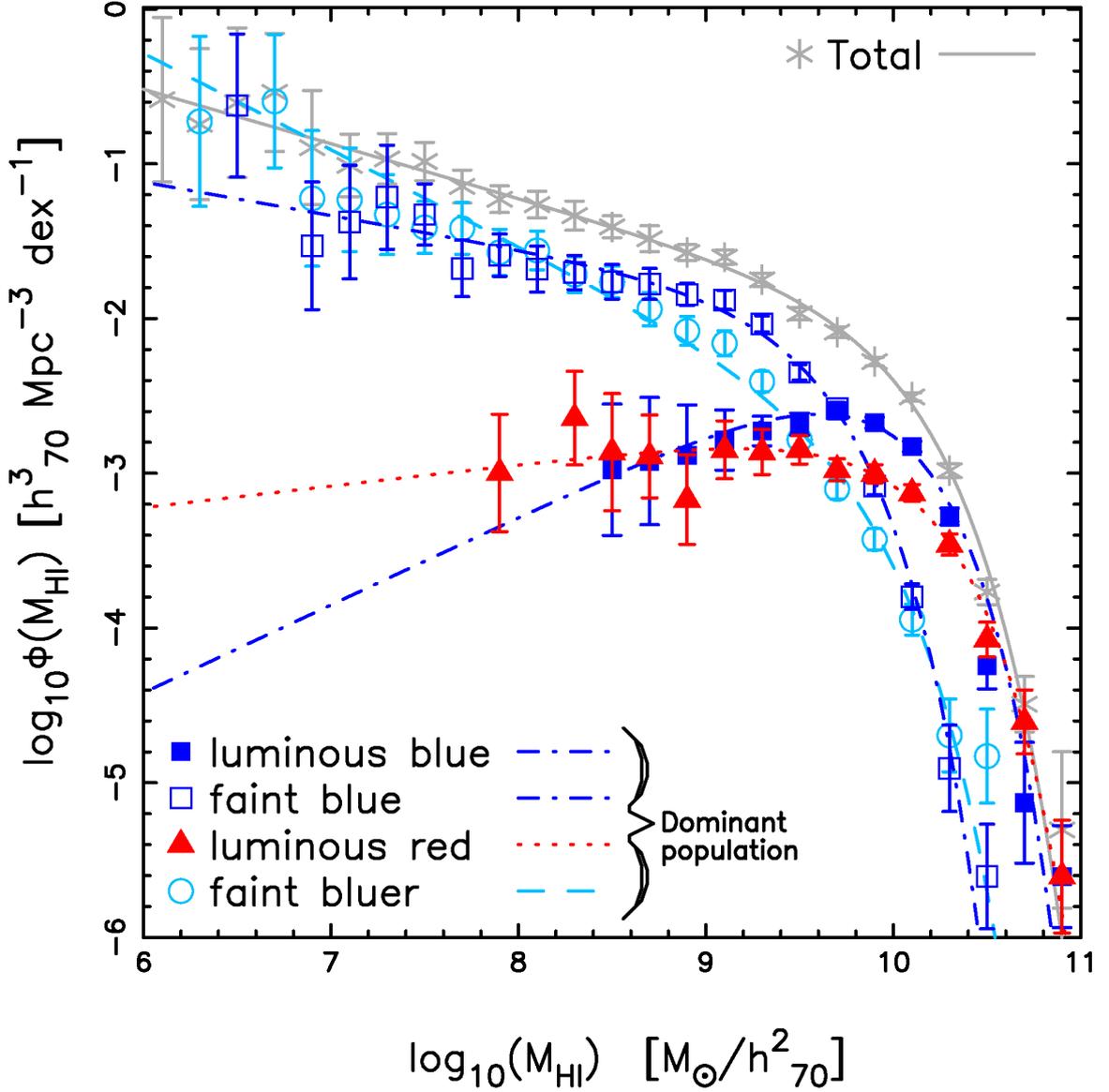


Figure 2.9: The HIMF for the luminous blue, luminous red, faint blue and faint bluer populations (1σ definition). Data points and error bars were estimated using the 2DSWML method (see appendix A). The red dotted, blue dot-dashed and cyan dashed curves are the Schechter function fits for red, blue, and bluer populations, respectively. For comparison we have added the total HIMF (crosses) and its fit (solid line). The symbols for the populations are the same as in figure 2.6. The details of the fits are given in table 2.1.

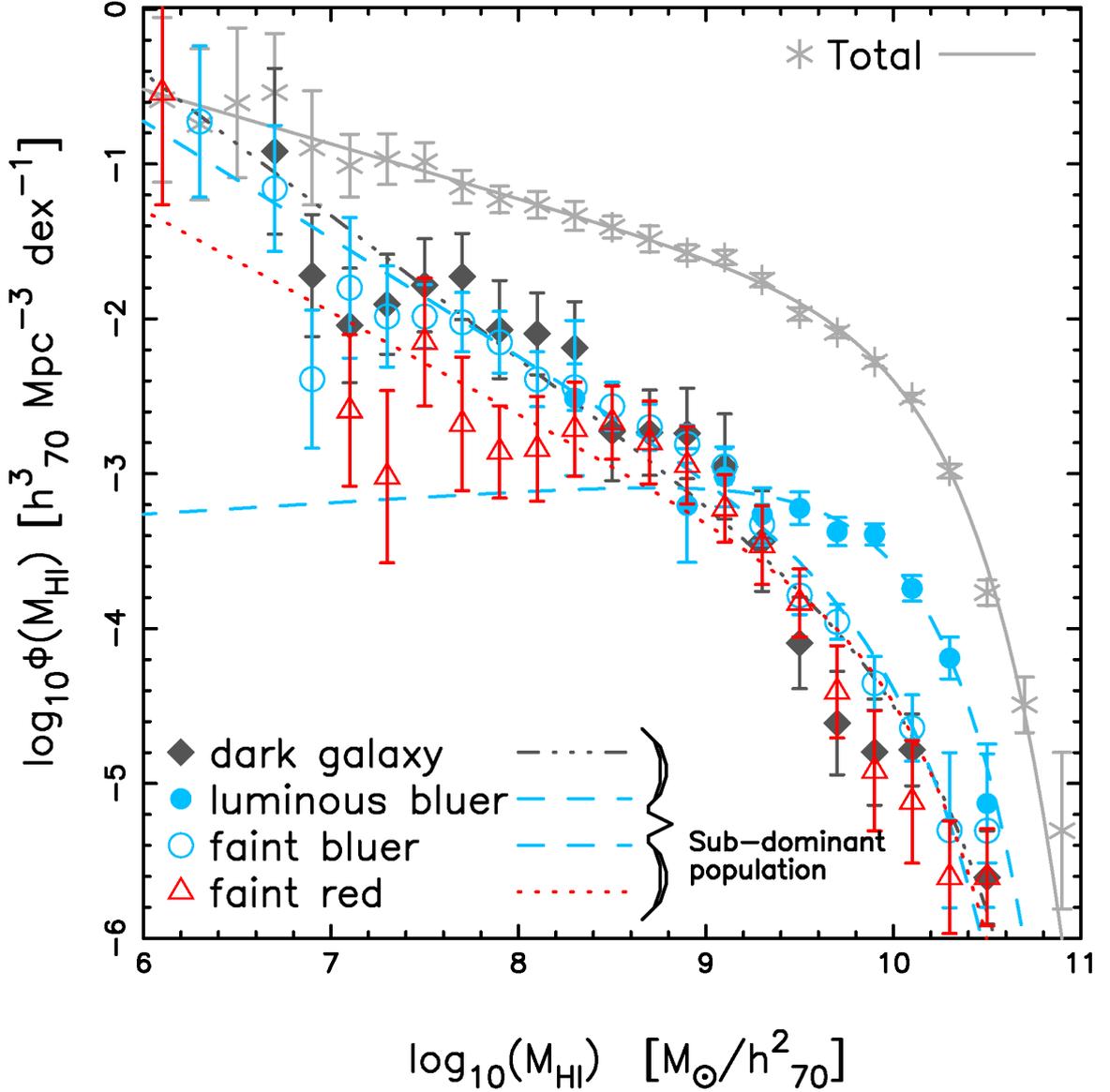


Figure 2.10: The HIMF for the luminous bluer, faint bluer, and faint red populations (1.5σ definition). Data points and error bars were estimated using the 2DSWML method (see appendix A). The red dotted and cyan dashed curves are the Schechter function fits for red and bluer populations, respectively. For comparison we have added the total HIMF (crosses) and its fit (solid line). 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 populations are the same as in figure 2.6. The details of the fits are given in table 2.1.

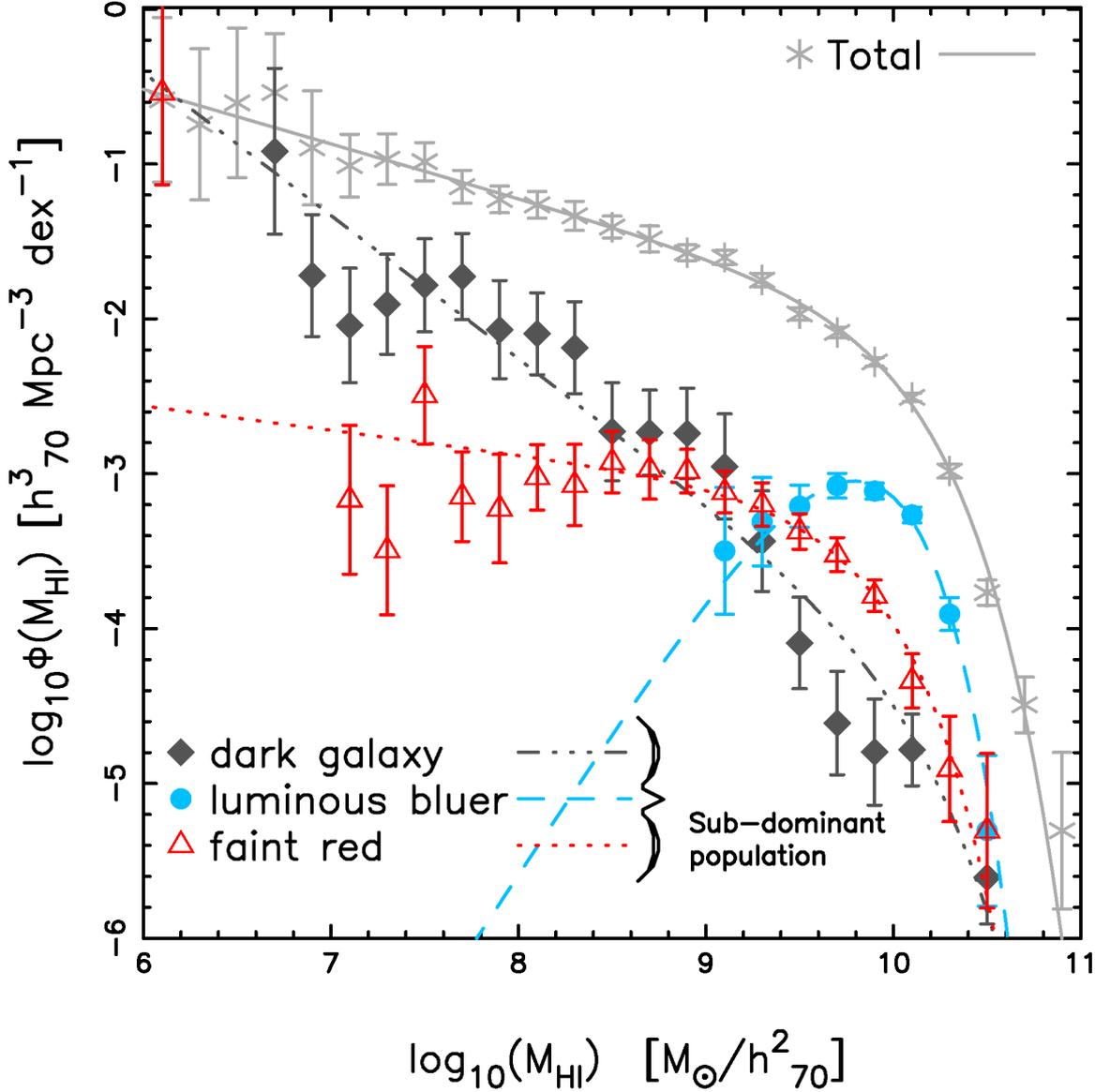


Figure 2.11: The HIMF for the luminous bluer and faint red populations (1σ definition). Data points and error bars were estimated using the 2DSWML method (see appendix A). The red dotted and cyan dashed curves are the Schechter function fits for red and bluer populations, respectively. For comparison we have added the total HIMF (crosses) and its fit (solid line). 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 populations are the same as in figure 2.6. The details of the fits are given in table 2.1.

ies (graphite diamonds). Compared to figure 2.7, the HIMF is smoother at lower masses. The symbols and color codes for the 6 populations are the same as in figure 2.6. Figure 2.8 and 2.10 shows the HIMFs for the 1.5σ populations; and figure 2.9 and 2.11 shows that for the 1σ populations. The results are shown in two parts — dominant populations (figure 2.8 and 2.9) and sub-dominant populations (figure 2.10, and 2.11). The population which dominates the total HIMF over the others in some mass range and contributes more than 10% to the total HI budget $\Omega_{\text{HI}}^{\text{tot}}$ (details in table 2.2) is called a dominant population. The curves going through the data points are the Schechter function fit to that data, where the Schechter function fit of the total, red, blue and dark populations are presented with solid line, dotted lines, dashed lines and dot-dashed line, respectively. The detail of the Schechter function fit is given in table 2.1. The values of the amplitude of the Schechter function ϕ_* and the characteristic mass M_* has been quoted in this thesis in the units of $(10^{-3}h_{70}^3 Mpc^{-3} dex^{-1})$ and $\log(M_*/M_{\odot}) + 2 \log h_{70}$, respectively. The faint end slope in figure 2.8, 2.9, 2.10, and 2.11 (α') is different from the α in equation 2.4 by 1, i.e., $\alpha' = \alpha + 1$, as we do the binning of HI mass and the velocity width in logarithmic grids.

We can compare our results with the HIMF from Martin et al. [2010], estimated for $\alpha.40$ code 1 objects, We find that our results of the best fit Schechter function parameters $(M_* \pm \sigma_{M_*}, \phi_* \pm \sigma_{\phi_*}, \alpha \pm \sigma_{\alpha}) = (9.96 \pm 0.02, 5.34 \pm 0.40, -1.35 \pm 0.02)$ are consistent at the 1σ level with the parameter values of $(9.96 \pm 0.02, 4.8 \pm 0.30, -1.33 \pm 0.02)$, quoted in Martin et al. [2010]. However the ϕ_* is just within 1σ of each other, which is because of the choice of our sample. The number of galaxies in our sample $\sim 25\%$ less than that of Martin et al. [2010]. In appendix A, it is shown that when we consider the full sample of ALFALFA, our result matches well (figure A6).

The reduced chisquare values (χ_{reduced}^2), given in the last column of table 2.1, represents the goodness of fits. χ_{reduced}^2 is of order unity for the dark, faint red and faint bluer popu-

region	$\log(M_*/M_\odot) + 2 \log h_{70}$	ϕ_* ($10^{-3} h_{70}^3 Mpc^{-3} dex^{-1}$)	α	$\chi^2_{reduced}$
total	9.96 ± 0.02	5.34 ± 0.40	-1.35 ± 0.02	0.79
luminous 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)
luminous 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)
luminous 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 ± 0.09	$[3.25 \pm 1.50] \times 10^{-2}$	-1.92 ± 0.09	1.17

Table 2.1: 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, $\chi^2_{reduced}$ is given in the last column.

lations of galaxies. In the faint red sub-sample of 1σ definition of populations, the lowest mass bin has the biggest contribution to the reduced chisquare value. Removal of that point flattens the slope α from -1.16 to -1.06 and increases ϕ_* by $\sim 22\%$, changing $\chi_{reduced}^2$ to 0.57 . The change in M_* is negligible. The luminous red and luminous bluer populations have lower $\chi_{reduced}^2$ for both the definition of population. Whereas luminous blue and faint blue populations have larger $\chi_{reduced}^2$. The faint blue population has a larger $\chi_{reduced}^2$ of 4.10 for 1.5σ definition, which reduces to 0.96 for the 1σ sample. This population has a larger number of luminous galaxies in the 1σ sample in comparison with 1.5σ sample (figure 2.6) and as M_{HI} is correlated to M_r (see figure 2.12), the 1σ sample represents the high mass end of the HIMF better, this is why the faint blue population is fitted better in the 1σ sample. This can be seen again in figure 2.9. The luminous blue population has $\chi_{reduced}^2 \sim 2$ no matter what sample definition you consider.

Since M_{HI} and M_r (figure 2.12) have a monotonic relation, we expect the high mass end of the HIMF to be dominated by the luminous population (figure 2.8, 2.9). The characteristic mass (M_*) increases from the luminous bluer sub-sample to the luminous red sub-sample. We see a little change in M_* for luminous red and luminous blue populations for different sample definition. For both of these populations, HIMFs also show little changes for $M_{HI} \geq 10.3$, with respect to the sample definition. On the other hand $M_* = 9.84$ for the luminous bluer population in 1.5σ definition, which reduces to $M_* = 9.52$ for the 1σ sample. The luminous populations have shallower slopes at the low mass end and this section of HIMF is dominated by the faint population, which is obvious from the fact that M_{HI} and M_r are correlated (figure 2.12). The sample definition affects the luminous bluer population the most at low mass end. α changes from -0.92 in 1.5σ sample to $+0.86$ in 1σ sample, whereas the luminous red population has the smallest change. In terms of amplitude, luminous blue population is affected the most, it changes from $\phi_* = 4.85$ (1.5σ sample) to $\phi_* = 2.54$ (1σ sample). The change for luminous bluer population is about $\sim 72\%$ and

for luminous red it is negligible. These changes can be better supported with the figure 2.6 and figure 2.12, as pointed out earlier. At the knee of the HIMF, the luminous blue is the dominant population with 1.5σ definition, however in case of 1σ definition it contributes almost equally as the faint blue population, which can be explained by the increase in the number of observed galaxies in the faint blue sub-sample from 1409 (1.5σ sample) to 2422 (1σ sample). And this is the main reason for the increase in ϕ_* to be increased by $\sim 4.8\times$. The interesting result is that the luminous red is the dominant population at $M_{\text{HI}} \geq 10.4$, irrespective of the sample definition. The luminous red population represents $\sim 60\%$ of the total detections at this mass range, where the other $\sim 40\%$ comes from the luminous blue population. Moving to the faint populations, all of them have steeper slopes at low mass end as compared to their luminous counterparts; they do not dominate the high mass range of the HIMF. The count of observed galaxies of the faint bluer population changes $\sim 3\times$ from 1.5σ sample to 1σ sample. This changes α from -1.76 to -1.62 and increases ϕ_* by $\sim 10\times$. In the 1.5σ definition faint bluer is a sub-dominant population, but in 1σ definition it becomes a dominant population for $M_{\text{HI}} \leq 8.1$. The faint blue population is the dominant population from low mass values to the knee of the HIMF, whereas the faint red population is a sub-dominant population, for both the definitions.

The HIMF of the dark galaxy population is portrayed with a very steep slope $\alpha = -1.92$, a very small amplitude $\phi_* = 3.25 \times 10^{-2}$ and a large characteristic mass $M_* = 10.03$, this is a sub-dominant population. Extrapolation of the HIMF of the dark galaxies below masses $M_{\text{HI}} \leq 6$ makes the dark population a dominant one. However the validation of this extrapolation is doubtful since it is very unlikely to have too many low mass HI rich galaxies, without stars and self-shielded from the photo-ionizing background.

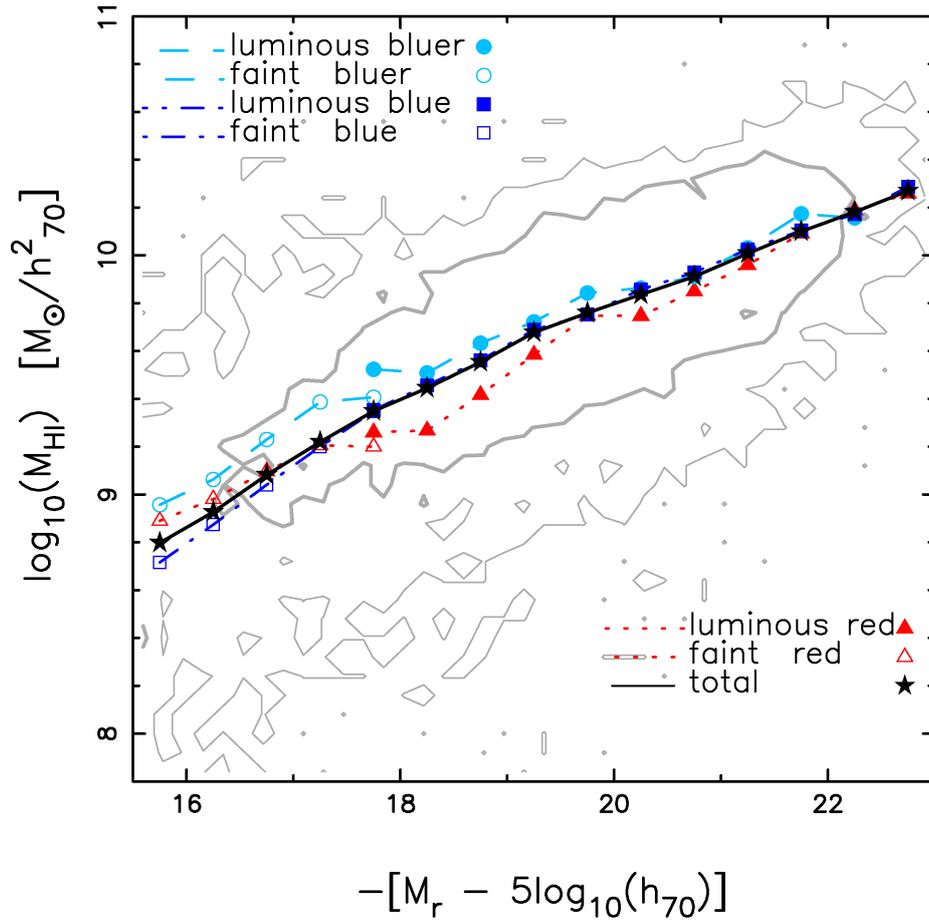


Figure 2.12: The $M_r - M_{\text{HI}}$ relation for the 1.5σ sample. The filled star represents the total sample excluding the dark galaxies. The other data points and line styles are the same as in figure 2.8, 2.9, 2.10, and 2.11.

2.5.1 The contribution of different galaxy populations to Ω_{HI}

Since we have the best-fit Schechter function to our HIMF, we can integrate it analytically to obtain the HI density parameter,

$$\Omega_{\text{HI}} = \frac{\rho_{\text{HI}}}{\rho_c} = \frac{M_* \phi_*}{\rho_c} \Gamma(\alpha + 2) \quad (2.5)$$

An alternate way is to sum the binned values of the HIMF. We find that both the methods give similar results, similar to [Haynes et al. \[2011\]](#), except for the dark population. As discussed in the previous section extrapolation of the dark HIMF to very small masses is not correct, we choose to present the Ω_{HI} values by the summation method instead of considering the analytical form as in eqn 2.5. Our results are summarized in table 2.2. Column 2 is the estimate of Ω_{HI} for different sub-samples and column 3 is the fractional contribution to the total HI budget ($\Omega_{\text{HI}}^{\text{tot}}$), for the 1.5σ definition. The values in parentheses are for the 1σ definition of populations.

We obtain $\Omega_{\text{HI}} = (4.894 \pm 0.469) \times 10^{-4} h_{70}^{-1}$ for the total sample, which is consistent with $\Omega_{\text{HI}} = (4.4 \pm 0.1) \times 10^{-4} h_{70}^{-1}$ of [Martin et al. \[2010\]](#) within 1σ . [Haynes et al. \[2011\]](#) report $\Omega_{\text{HI}} = (4.2 \pm 0.1) \times 10^{-4} h_{70}^{-1}$, including the code 2 objects also, which is consistent with our values at the 2σ level. The main reason for the discrepancy in the Ω_{HI} values in comparison to that of [Martin et al. \[2010\]](#) is due to the difference between theirs and our ϕ_* values. Though our M_* value is the same as [Martin et al. \[2010\]](#) and α is little less than the $\alpha.40$ result, our value of ϕ_* is $\sim 10\%$ higher than [Martin et al. \[2010\]](#), which translates to a 10% larger value in Ω_{HI} value (equation 2.5).

Table 2.2 shows that the total red cloud (luminous and faint) contributes $\sim 17\%$ to $\Omega_{\text{HI}}^{\text{tot}}$, and the dark population also has a non-negligible ($\sim 3\%$) contribution, which adds up to a fraction of $\sim 20\%$ (red and dark). Hence the full blue cloud (luminous and faint blue and bluer) has a contribution of $\sim 80\%$ of the total HI budget. The luminous blue and faint blue populations, which dominate the knee of the HIMF, contribute $\sim 73\%$ and $\sim 55\%$ to the

region	$\Omega_{HI}[10^{-4}h_{70}^{-1}]$	$\Omega_{HI}/\Omega_{HI}^{total}$
total	4.894 ± 0.469	1.00
luminous blue	2.543 ± 0.298 (1.099 ± 0.115)	0.520 (0.224)
faint blue	1.014 ± 0.455 (1.604 ± 0.196)	0.207 (0.328)
luminous red	0.764 ± 0.124 (0.653 ± 0.110)	0.156 (0.133)
luminous bluer	0.215 ± 0.135 (0.333 ± 0.165)	0.044 (0.068)
faint bluer	0.167 ± 0.126 (0.957 ± 0.233)	0.034 (0.196)
faint red	0.094 ± 0.050 (0.144 ± 0.038)	0.019 (0.029)
dark	0.162 ± 0.137	0.033

Table 2.2: The contribution of different populations to Ω_{HI} . Column 2 is the estimate of Ω_{HI} from a single population and column 3 is the fractional contribution to Ω_{HI}^{tot} . The estimated values for the 1σ sample is in brackets.

Ω_{HI}^{tot} for 1.5σ and 1σ samples, respectively.

Earlier in section 2.5 we defined a dominant population to be the one which dominates the total HIMF in some mass range and contributes $\geq 10\%$ to the Ω_{HI}^{tot} . According to this definition the dominant populations in the 1.5σ classification are the faint blue, luminous blue and luminous red populations. Whereas in the 1σ classification the faint blue, faint bluer, luminous blue and luminous red populations are the dominant populations. Irrespective of the sample definition, the dominant populations together have a contribution of 90% to Ω_{HI}^{tot} .

2.6 Discussion

As we are investigating the contribution of different population of galaxies to the total HIMF, it will be an interesting exercise to look at the scaling relations for these populations.

Catinella et al. [2010], Huang et al. [2012], Maddox et al. [2015] explore such scaling relations for ALFALFA galaxies and Parkash et al. [2018] do that for HI Parkes All-sky Survey catalog (HICAT). In this thesis we estimate the stellar masses with *kcorrect*, which uses the population synthesis code of Bruzual and Charlot [2003]. The estimation of the M_{star} values of those authors are different from our estimates. Figure 2.13 presents the observed scaling relations between M_{HI} and M_{star} for our sample.

Figure 2.13 shows the distribution of HI selected galaxies as a function of M_{HI} and M_{star} . 1σ and 2σ contours are drawn as thick and thin grey curves, respectively. The black line and the black stars represent the observed scaling relation between M_{HI} and M_{star} . The $M_{\text{HI}} - M_{\text{star}}$ scaling relations for different populations are shown using the same color code, line style and point type as in figure 2.6, 2.8, 2.9, 2.10, and 2.11. We also plot the scaling relation (crossed-circle, thin solid green line) obtained for the total ALFALFA sample [Huang et al., 2012], which compare well with our results within the scatter. Our scaling relation is also consistent with Maddox et al. [2015], Parkash et al. [2018] for HI selected samples but it differs from that of Catinella et al. [2010], which is obtained from an M_{star} selected sample. We have not corrected for dust in this work while Huang et al. [2012] have used UV bands in Galaxy Evolution Explorer (GALEX) to correct for it. Not correcting for dust therefore can be the reason for the discrepancy between these scaling relations. Also our stellar masses are underestimated at lower masses. The affect of dust can also be seen in the scaling relations for different populations. The faint bluer and faint blue have similar slopes, which is steeper than their luminous counterpart. But the faint red population has a shallower slope than the luminous red population and also that of the faint bluer and faint blue populations. We also find the HI fraction, $f_{\text{HI}} = M_{\text{HI}}/M_{\text{star}}$, to be increasing with decreasing ($u-r$) color.

A clear transition can be seen in the scaling relations when going from lower masses to higher masses. The HI fraction f_{HI} has a suppression at about $M_{\text{star}} \sim 9$, which is consistent with Huang et al. [2012], Maddox et al. [2015]. This transition is dependent on the galaxy

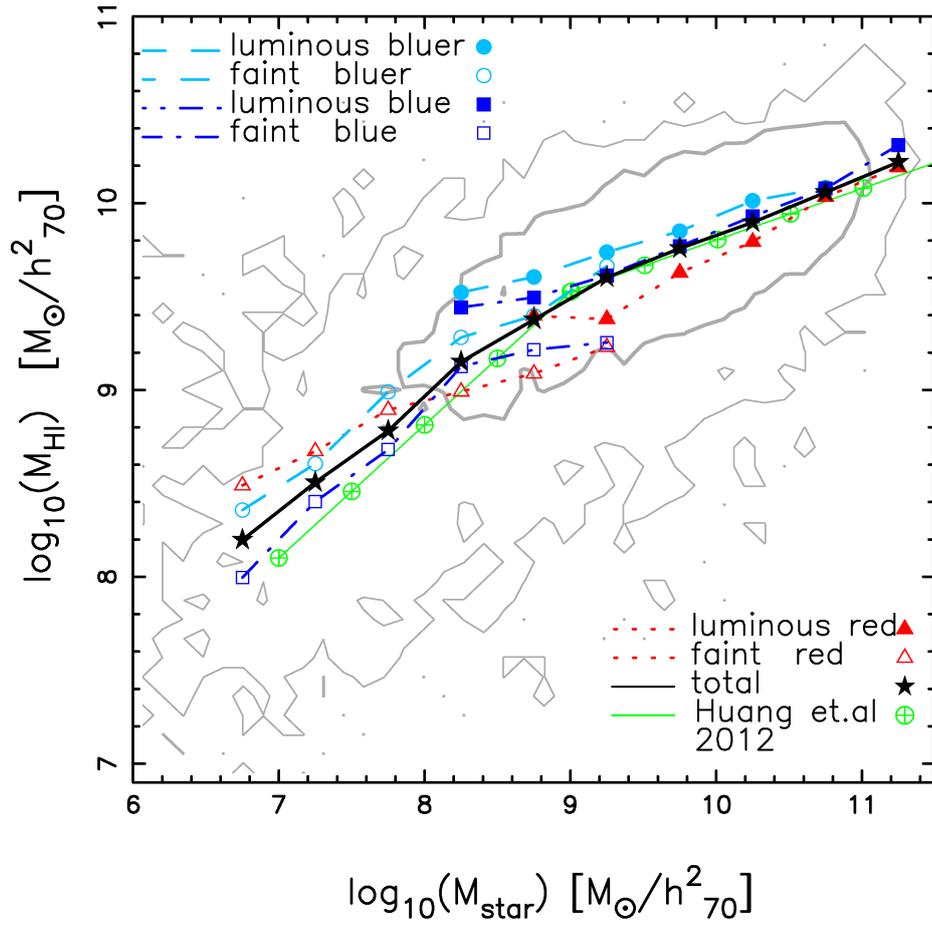


Figure 2.13: The $M_{star} - M_{HI}$ relation for the 1.5σ sample. The filled star (thick solid line) represents the total sample excluding the dark galaxies. The results are compared with Huang et al. [2012] (crossed-circle and thin solid line). other data points and line styles are the same as in figure 2.8, 2.9, 2.10, and 2.11.

population. For bluer, blue and red populations it occurs at $M_{\text{star}} = 9.4, 10.1, 8$ 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 to a transition between hot and cold mode accretion seen in cosmological hydrodynamical simulations [Kereš et al., 2009].

Important results, obtained in the last section, are — i) the luminous red population dominates the HIMF at higher masses, $M_{\text{HI}} \geq 10.4$, and ii) the red galaxies contribute a non-negligible (17%) fraction to the total HI content, $\Omega_{\text{HI}}^{\text{tot}}$. Paul et al. [2018] also find that the red population to dominate at higher masses, using the HI-HOD formalism. Looking at the morphology of the ALFALFA red galaxies, we find that these are mostly spiral and lenticular galaxies, but some of them are elliptical. Some of the spiral galaxies have prominent bulges, which puts them into an early-type population. Few of them have visible dust lanes on their disk plane. Some of the galaxies are edge-on or somewhere in between edge-on and face-on. HI detection has been reported in early type galaxies [Morganti et al., 2006, Oosterloo et al., 2007, Grossi et al., 2009, Serra et al., 2012], but with $M_{\text{HI}} \leq 10$. 47% of the total local SFR density is reported for $M_{\text{star}} > 10$ in the GALEX Arecibo SDSS Survey (GASS) sample [Schiminovich et al., 2010]. A fraction of the GASS sample have CO detections also [Saintonge et al., 2011], which is a tracer for H_2 and proxy for star formation. As GASS (and ALFALFA) detections are mostly in the blue cloud [Catinella et al., 2010], it is expected that these red gas-rich galaxies have a negligible contribution to the local SFR density. This does not imply that all individual red HI-rich galaxies have low star formation rates, but their total contribution is negligible because they are less in number. A fraction of them would be those which are transitioning from a blue star-forming phase to a red passive phase, another fraction will be the dusty star-forming galaxies and remaining will be passively evolving galaxies. Being edge-on enhances the amount of reddening, and the color would be reddened of disk galaxies which are either transitioning to the red phase or contain considerable dust on their disks [Graham and Worley, 2008, Tempel et al., 2011].

Although the luminous red population dominates the high mass end ($M_{\text{HI}} \geq 10.4$) of the HIMF, the observed count for luminous red galaxies are only 50% more than that of the luminous blue population. On a logarithmic scale the difference between the HIMFs of these two are not very large (figure 2.8, 2.9). If the reddening due to inclination of galaxies is important and we correct the magnitudes for this, then a fraction of red HI-rich galaxies will move to the luminous blue population, which will bring these two HIMFs closer to each other. Again if inclination increases the reddening, it would show prominent features in HI velocity width functions (HIWF). Moorman et al. [2014] find that the red and blue HIWFs are well separated at the high velocity end. Therefore the reason of red population becoming the dominant population at the high mass end of the HIMF can partly be explained by the reddening due to dust and inclination.

In figure 2.5 we show that only 11% of the red galaxies in SDSS have HI counterpart in ALFALFA, whereas for blue galaxies this number is 38% in the volume of analysis. On the other hand 98% of ALFALFA galaxies have optical counterparts. Since the detection in HI depends on HI flux as well as on HI velocity widths, we can argue that the non-detections in luminous red sample are due to the insufficiency of HI gas and larger velocity widths and the HI mass should decrease with increasing stellar mass. Figure 2.13 shows that HI-rich galaxies are massive (in terms of stellar mass) also. But this relation is biased as it is for an HI-selected sample. Looking at the GASS sample [Catinella et al., 2013], selected on stellar mass, the relation for $M_{\text{star}} \in [10.76, 11.30]$ has nearly a flat slope, being slightly negative if all the non-detections are assigned $M_{\text{HI}} = 0$ and slightly positive if a limiting HI mass is assigned to the non-detections. However GASS and ALFALFA will miss a significant number of galaxies, having larger stellar masses, as they are relatively shallow than the optical surveys like SDSS. The number of red galaxies, which dominates the tail of the stellar mass function (SMF), is $\sim 10\times$ more than that of the blue galaxies at $M_{\text{star}} = 11.3$ [Baldry et al., 2012]. These are most likely the central red galaxies [Drory et al., 2009]

and resides in halos of $\log_{10}(M_{\text{halo}}/\odot) \simeq 14 - 14.5$ [Behroozi et al., 2010] with virial temperatures $T_{\text{vir}} \sim 10^7\text{K}$ and circular velocities $V_{\text{circ}} \sim 6 - 9 \times 10^2\text{km.s}^{-1}$. Most of these will be the central galaxies of clusters of galaxies. But it would be very unlikely to detect a considerable amount of HI as these systems corresponds to large virial temperatures and large velocity widths. The $M_{\text{HI}} - M_{\text{star}}$ relation, would probably not asymptote to a constant as seem in Catinella et al. [2013] rather it would decrease with increasing M_{star} values as suggested in the results of Kim et al. [2017] and Spinelli et al. [2020].

HI abundance matching techniques will break down for a non-monotonic HI-stellar mass relation, and therefore between HI-halo mass. However Khandai et al. [2011], Padmanabhan and Kulkarni [2017] obtain $M_{\text{HI}} - M_{\text{halo}}$ relation using abundance matching methods. The HI-HOD models which also assume a step-like function [Guo et al., 2017, Paul et al., 2018] for the average occupation of centrals, may need some modifications. Shen et al. [2013] compared the step-like parametrization of the mean occupation function with their log normal form and found the quasar HOD parameters to be not well constrained. Only more direct observations will provide a better estimate of the HI content of these massive galaxies and hopefully constrain the HOD parameters better.

At the end of this chapter we will discuss about the sensitivity limits of ALFALFA and how they may affect our results. We will start by looking at the targeted HI observations, which look at the HI content of luminous massive early-type galaxies (ETGs) in the local Universe. The targets in the GASS survey [Catinella et al., 2010, 2013] are selected by stellar mass, $10 < M_{\text{star}} < 11.5$, from an area common to ALFALFA [Giovanelli et al., 2005], SDSS [Abazajian et al., 2009] and GALEX [Martin et al., 2005] within a redshift range of $0.025 < z < 0.05$. The detection limits of GASS, $f_{\text{HI, lim}} = M_{\text{HI, lim}}/M_{\text{star}} > 0.015$ for $M_{\text{star}} > 10.5$ and $M_{\text{HI}} = 8.7$ for $M_{\text{star}} < 10.5$, translates to upper HI mass limits of $8.7 \leq M_{\text{HI, lim}} \leq 9.7$ for the non-detections. The maximum limiting mass for GASS survey, which is more sensitive than ALFALFA, is below the characteristic mass, M_* , of

the population we have considered in this work (table 2.1), the large mass ($M_{\text{HI}} > M_*$) end of the HIMF therefore will not be affected by the ALFALFA sensitivity limit.

The non-detections of GASS [Catinella et al., 2013] span the entire M_{star} range targeted; these are mostly redder in color (NUV-r) and the detection fraction is $\sim 70\%$ for $M_{\text{star}} < 10.7$, which drops to 40% above that. The detection fraction of ALFALFA in SDSS (figure A1) is consistent with the trend seen in GASS, where non-detections are dominated by red galaxies. It can be expected then that there are a non-negligible number of massive galaxies, mostly from the luminous red population, which could host HI gas of mass upto $M_{\text{HI, lim}} = 9.7$ and will not be seen by ALFALFA because of their large velocity widths.

The ATLAS^{3D} HI survey [Serra et al., 2012] reports the HI masses of early-type galaxies, selected from a volume limited sample of 871 nearby ($D < 42\text{Mpc}$, $|\delta - 29^\circ| < 35^\circ$, $|b| > 15^\circ$) galaxies brighter than $M_K < -21.5$ which translates to stellar masses $M_{\text{star}} \geq 9.78$ [Cappellari et al., 2011]. The HI observation are done for 166 ETGs with the Westerbork Synthesis Radio Telescope. Among them 53(32%) are detections and 113(68%) are non-detections. 95% of the ETGs lie on the red sequence (as demarcated by eq. 2.2). The ATLAS^{3D} sample then represents massive, red, E or S0 type galaxies. The ETGs span the range of $M_{\text{HI}} \in [7.9, 5]$. The spirals from the parent sample have a narrower HI mass distribution, with a peak at $M_{\text{HI}} \sim 9.3$ and a tail at $M_{\text{HI}} \sim 8$ [Serra et al., 2012]. On the other hand the limiting masses of non-detections are distributed in the range $M_{\text{HI, lim}} \in [6.5, 8.5]$ with a peak at $M_{\text{HI, lim}} \sim 7.1$. There is a significant overlap between the HI distribution of detected ETGs and spirals which implies that a non-negligible fraction of ETGs contain HI as much as in spirals. The knee of the HIMF of ETGs is about $M_{\text{HI}} = 9.26$, which is a factor of 5 smaller than the ALFALFA total sample ($M_* = 9.96$) and a factor of 6 smaller than the luminous red population of ALFALFA ($M_* = 10.04$). The difference between HI column density distribution (N_{HI}) for spirals and ETGs in the ATLAS^{3D} survey is significant. The characteristic column density is $N_* = 1.03 \times 10^{21} \text{cm}^{-2}$ for spirals

and $N_* = 9.2 \times 10^{19} \text{cm}^{-2}$ for ETGs. Therefore the HI in gas rich ETGs is rarely as dense as the average column densities of spirals [Serra et al., 2012]. Given that the GASS and the ATLAS^{3D} HI surveys are targeted surveys, more sensitive than ALFALFA, and are specifically looking at the HI gas content of massive galaxies we can conclude that it is very unlikely that ALFALFA has missed out any galaxy with masses $M_{\text{HI}} > 10$ due to an unusually large velocity width.

Finally it is worth noting how different choices of the sensitivity limits of ALFALFA affect the HIMF. Haynes et al. [2011] addressed this question by considering Code 1 and 2 sources, which in turn lowers the sensitivity limit compared to Code 1 only objects [eqs. 6 and 7 and discussion in section 6 of Haynes et al., 2011]. Both the amplitude and characteristic mass of HIMF remained unchanged, however the faint-end slope α decreased from -1.33 ± 0.02 (Code 1) to -1.29 ± 0.02 (Code 1 and 2).

Papastergis et al. [2012] explored the differences in HIMF by considering an HI-selected sample and an optically-selected sample. The HI-selected sample is the $\alpha.40$ sample and the optically-selected sample consists of all the HI detections and non-detections. The non-detections are the SDSS galaxies in the same volume as ALFALFA which do not have HI detections. These galaxies are assigned a lower and upper limiting HI mass ($M_{\text{HI, lim}}^{\min}$, $M_{\text{HI, lim}}^{\max}$). The lower limit is $M_{\text{HI, lim}}^{\min} = 0$. The upper limit is computed by converting the detection limit, which is the 25% completeness limit [eqs. 5 and 7 in Haynes et al., 2011], to an HI mass. To estimate $M_{\text{HI, lim}}^{\max}$ we need a distance (which exists) and velocity width, w_{50} (which has to be estimated). Papastergis et al. [2012] used the average stellar mass Tully-Fisher relation, $M_{\text{star}} - V_{\text{rot}}$ for the $\alpha.40$ galaxies to estimate w_{50} for the non-detections, after accounting for inclination effects. The optically-selected sample then consists of ALFALFA detections and two estimates of HI masses for the non-detections. The optically-selected sample is an r-band flux limited sample which has a different selection function as compared to the HI-selected sample. The HIMF estimated from the optically-selected and HI-selected

sample should broadly be consistent with each other. A naive expectation is that the HIMF from the HI selected-sample should lie in between the two estimates of the HIMF from the optically-selected sample. In the limit that the assigned lower and upper limiting HI masses approach the *true* HI mass of the undetected source we expect the HIMF estimated from the optically-selected sample to approach that of the HI-selected sample. We also point out at the high mass end, all the estimates of the HIMF should be the same. This is demonstrated in figure 6 of [Papastergis et al. \[2012\]](#). Given the uncertainties associated in obtaining $M_{\text{HI,lim}}^{\text{max}}$ which use average scaling relations and inclination effects (which are prone to errors) we do not comment further on the differences between the HIMF from the two samples. Based on the arguments presented above, we believe that although the sensitivity of ALFALFA affects the individual detections as compared to more sensitive targeted surveys, the estimates of the HIMF for different populations considered here are robust.

2.7 Summary

In this chapter we present the HIMF of different populations of galaxies, considered from ALFALFA and SDSS, based on their color-magnitude values in the local Universe. After putting all the restrictions on the volume, quality and completeness our final sample contains 7857 galaxies, which we divide into seven populations: six based on their color-magnitude values and another one which is optically undetected (dark galaxies). We consider two definitions to get the six sub-samples from the color-magnitude plane. We have called them the 1σ and 1.5σ samples (section 2.3). This, we do, to illustrate that our sample definition does not change the qualitative results that we report. 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 2.9 and tables 2.1 and 2.2). We summarize our results below:

- For a fixed range in color, in the color-magnitude plane (e.g. red, blue or bluer), the

HIMF of the luminous 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.

- For a fixed magnitude interval, in the color-magnitude plane (e.g. luminous 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 luminous population, we see that the HIMF at the high mass end increases with increasing color. The luminous red population is the dominant population at this end.
- The luminous red population 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 Ω_{HI} budget. The dark population contributes $\sim 3\%$ to Ω_{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 $\sim 80\%$ of the Ω_{HI} budget.
- In the mass range about the knee, $M_{\text{HI}} \in [8, 10.4]$, the dominant populations are the faint and luminous blue populations with the latter dominating at larger masses in this interval. Their total contribution to Ω_{HI} is $\sim 55 - 70\%$ depending on sample definitions.
- The dominant populations contributing to the low mass end of the HIMF are the faint blue and faint bluer populations, the latter being dominant only for the 1σ sample definition.
- The luminous bluer and faint red populations are subdominant populations contributing a total of $\sim 6 - 10\%$ to Ω_{HI} . For the 1.5σ sample the fraction of luminous

bluer(faint red) galaxies in ALFALFA to that in SDSS is 41%(28%). In comparison the detection rate of ALFALFA in the red cloud is 11% and the blue cloud is 38%. This shows that although the detection rate of ALFALFA for the faint red population is higher than the luminous red population their contribution to Ω_{HI} is small. This is because the number density of these galaxies is small and the HI mass of these galaxies is also small. On the contrary the luminous blue galaxies are HI rich (figure 2.13) on average but due to their small number densities (figure 2.8,2.9) they too contribute little to the Ω_{HI} budget.

It would be interesting to see if simulations [Davé et al., 2017], SAMs [Kim et al., 2017] and HOD models [Paul et al., 2018] are able to reproduce our results which provide additional constraints on the population of HI selected galaxies. In section 2.6 we argued that the effect of dust and inclination are responsible for the luminous 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.

Chapter 3

The distribution of neutral hydrogen in the color–magnitude plane of galaxies

3.1 Introduction

The amount of neutral hydrogen (atomic and molecular) gas represents the fuel stocked for future star formation in the galaxies. Observations of late-type disk galaxies show that the surface density of star formation rate, Σ_{SFR} , is strongly correlated with the surface density of cold gas (HI + H₂), Σ_{gas} , known as the Kennicutt-Schmidt law [Schmidt, 1959, 1963, Kennicutt, 1998, 1989]. Targeted observations have detected HI in late-type (E and S0) galaxies [Morganti et al., 2006, Oosterloo et al., 2007, Serra et al., 2012], but their star formation rate is negligible to construct a corresponding KS-like law for them. Blind surveys on the other hand have constrained the HIMF in the local Universe [Zwaan et al., 2003, Martin et al., 2010, Haynes et al., 2011, Jones et al., 2018], but the HIMF does not reveal how HI is distributed amongst different galaxy populations.

Although the HIMF, and other one-dimensional functions (e.g. multiband luminosity functions, stellar mass functions, SFR function, to name a few) are important distributions which any theory of galaxy formation should reproduce, they only represent marginalized distributions of higher dimensional multivariate distribution functions of galaxies. These multivariate functions encode the effects and interplay of complex processes between vari-

ous baryonic components of galaxies. With the advent of ongoing and future large surveys which target different bands of the electromagnetic spectrum there is a need to go beyond one-dimensional functions. It is common to present bivariate or multivariate functions, when the observables are from different surveys, as conditional functions. The bivariate HI mass – B-band luminosity function was estimated from a sample of 61 galaxies in the blind Arecibo HI Strip Survey (AHISS) [Zwaan et al., 2001]. More recently Lemonias et al. [2013] presented the HI mass – stellar mass bivariate function using a parent sample of 480 galaxies from the GALEX Arecibo SDSS Survey (GASS) Data Release 2 [Catinella et al., 2010, 2012]

This chapter presents the conditional HIMFs conditioned on $(u - r)$ color and r-band rest frame magnitude (M_r). Section 3.2 describes the data we use for this work. We define conditional HIMF in section 3.3 and describe the steps to estimate the distribution of Ω_{HI} in the color-magnitude plane in section 3.4.1. In section 3.5 we discuss our results.

3.2 Data

For this part of the work we use the ALFALFA 40% data [Haynes et al., 2011], most of which have optical counterparts in SDSS DR7 [Abazajian et al., 2009]. This is the same dataset described in the previous chapter, but we nevertheless briefly describe it to keep the chapter self-contained. We put a limit on our sample at redshift $z_{\text{cmb}} = 0.05$ to avoid radio frequency interference (RFI). We also restrict ourselves to the best quality detections with signal-to-noise ratio (SNR) > 6.5 , known as *Code 1* objects. And finally, we define an area which is common to both ALFALFA and SDSS (DR7) surveys. A detailed description of this common area of analysis is given in chapter 2. Lastly we apply the 50% completeness cut [Haynes et al., 2011]. Our final sample of analysis thus consists of 7857 galaxies, detected over $\sim 2093 \text{ deg}^2$ of the sky, and the corresponding volume is $\sim 2.02 \times 10^6 \text{ Mpc}^3$.

Among these galaxies 7709 ($\sim 98\%$) have optical counterparts in SDSS DR7, and remaining 148 (2%) galaxies, referred to as *Dark galaxies*, do not have optical counterparts though they reside in the DR7 footprint. In this particular analysis we ignore these dark galaxies, and estimate the conditional mass functions with 7709 galaxies. As we have discussed in chapter 2, the contribution of dark galaxies is $\sim 3\%$ to the total HI budget, therefore our results should not be sensitive to this population of galaxies. We extract the *ugriz* model magnitudes (corrected for the extinction due to Milky Way [Schlegel et al., 1998]) and obtain the rest frame magnitudes, M_u, M_g, M_r, M_i, M_z , using *kcorrect* [Blanton and Roweis, 2007].

3.3 Conditional HIMF

The HIMF, $\phi(M_{\text{HI}})$, represents the underlying number density of galaxies in the Universe in an infinitesimal HI mass range dM_{HI} , as a function of their HI mass (M_{HI}), which is expressed as

$$\phi(M_{\text{HI}}) = \frac{dN}{V d \log_{10} M_{\text{HI}}} \quad (3.1)$$

where dN is the number of galaxies in the survey volume V with HI masses in the range $[M_{\text{HI}}, M_{\text{HI}} + dM_{\text{HI}}]$. This HIMF can be described reasonably well by a three parameter Schechter function (see chapter 2 sec 2.4).

We discuss in the previous chapter that the sensitivity limit of the ALFALFA survey depends on both the 21cm flux (or HI mass), S_{21} (or M_{HI}), and the profile width, w_{50} . We estimate this bivariate distribution, $\phi(M_{\text{HI}}, w_{50})$, using the two-dimensional step wise maximum likelihood (2DSWML) method. A detailed discussion on this method is given in appendix A. The advantages of using the 2DSWML method are - this is less sensitive to the effects of large scale structure, and this method does not assume any functional form of the

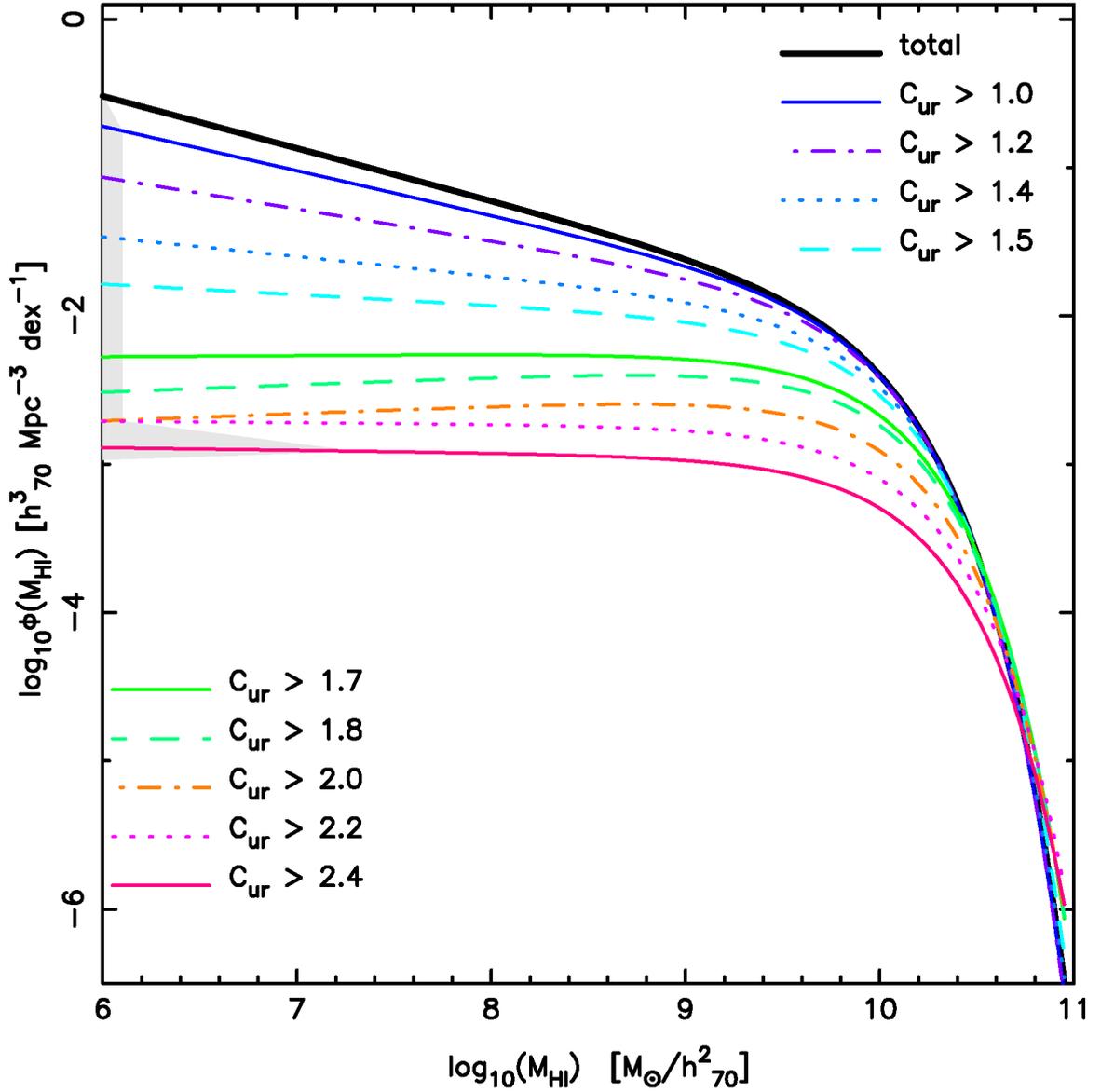


Figure 3.1: Conditional HIMF as a function of increasing color thresholds (top to bottom). The thick solid line is the HIMF for the full sample. The shaded gray region does not contain data, the conditional HIMF have however been extrapolated into this regime as well.

HIMF rather it considers the ϕ_{jk} s, discretized bivariate function $\phi(M_{\text{HI}}, w_{50})$ in the j^{th} HI mass bin and k^{th} velocity profile width bin. The disadvantage is that the normalization of the HIMF is lost in this method and it has to be fixed separately (see sec A1.2 of appendix A). Finally integrating the bivariate distribution $\phi(M_{\text{HI}}^j, w_{50}^k)$ over the velocity profile width, we get the HIMF $\phi(M_{\text{HI}}^j)$ [Zwaan et al., 2003, Martin et al., 2010, Haynes et al., 2011, Jones et al., 2018].

In this chapter we present the conditional HIMFs conditioned on $C_{ur} = u - r$ and/or M_r . The color-conditioned HIMF can be defined as the HIMF estimated for galaxies with color C_{ur} redder than a threshold value C_{ur}^t -

$$\phi(M_{\text{HI}}|C_{ur}^t) = \phi(M_{\text{HI}})|_{C_{ur} > C_{ur}^t} \quad (3.2)$$

Similarly we can define the luminosity-conditioned HIMF as the HIMF for a more luminous sample of galaxies than a threshold value M_r^t .

$$\phi(M_{\text{HI}}|M_r^t) = \phi(M_{\text{HI}})|_{M_r < M_r^t} \quad (3.3)$$

Starting with the full sample we create several sub-samples based on different threshold values of color (C_{ur}) or magnitude (M_r). We estimate the HIMF for these sub-samples and also compute the uncertainties on them following the method discussed in appendix A. We also fit a three parameter Schechter function to all the conditional HIMFs.

3.4 Results

Since 98% of the HI selected galaxies have optical counterparts it would be interesting to look at the conditional HIMF, conditioned on different optical properties. In this work, we have ignored the population of dark galaxies. As we have discussed earlier, the omission of dark galaxies (which are only 2% of the total sample) will not affect our results quantitatively, because these galaxies have only 3% contribution to the total HI density, $\Omega_{\text{HI}}^{\text{tot}}$ (see

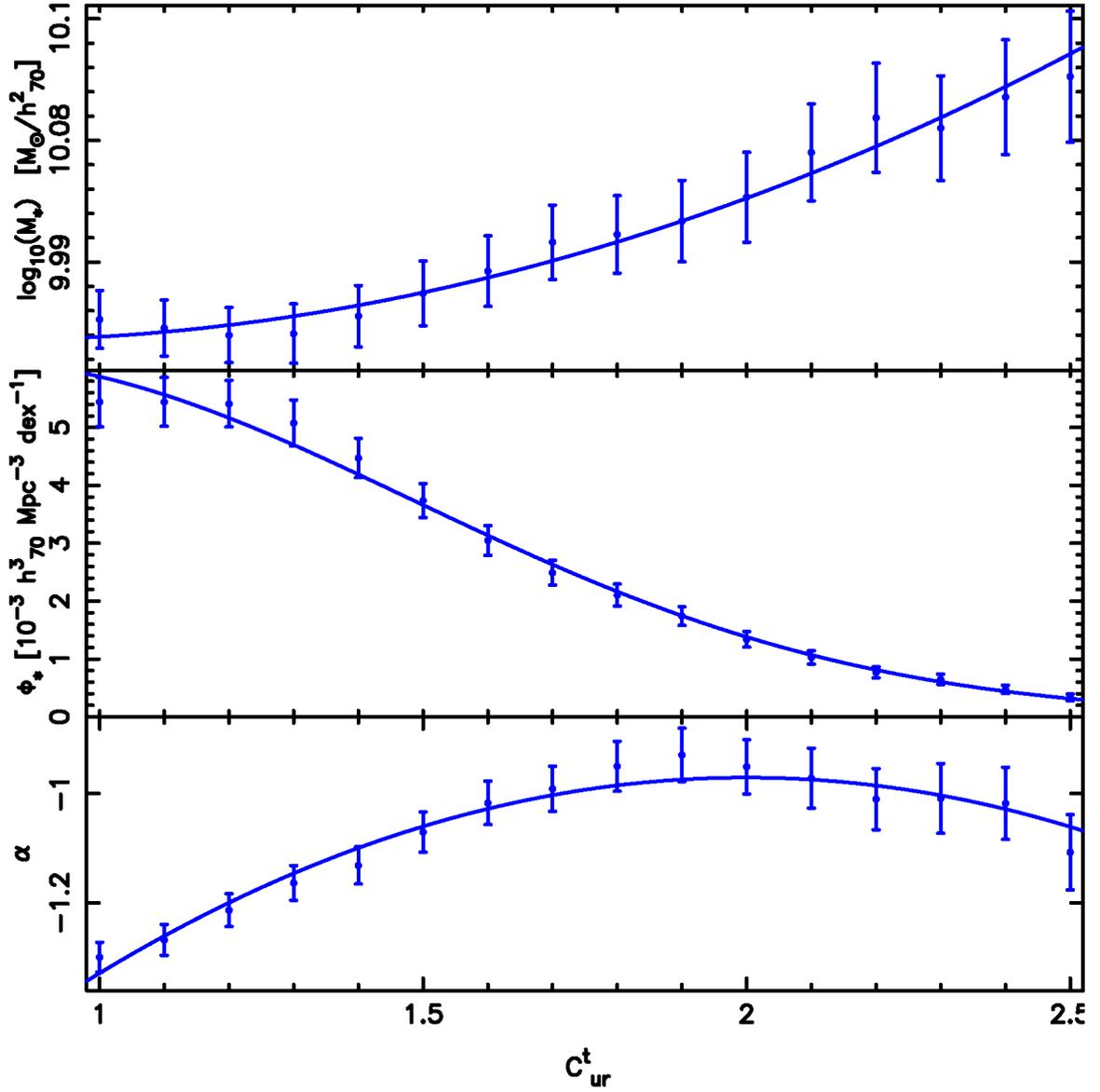


Figure 3.2: The Schechter function parameters of the conditional HIMF and their uncertainties as a function of color thresholds. The solid lines are fits to the data points with a quadratic function. The top, middle and bottom panels show the dependence of M_* , ϕ_* and α respectively, on the color threshold C_{ur}^t .

chapter 2). We emphasize that this is an HI selected sample for which optical properties exist for all galaxies. Therefore the computation of the conditional HIMF, conditioned on an optical property, will only take into consideration the ALFALFA selection function and volume.

Figure 3.1 shows the conditional HIMFs as a function of C_{ur}^t and figure 3.3 presents the conditional HIMFs as a function of M_r^t . We only draw the bestfit Schechter functions to the respective data sample rather than the data points to show the results clearly. The thick solid black line is the HIMF for the total sample and the other lines are the HIMFs with different threshold values on C_{ur} and M_r . The grey colored shaded area in both these two plots represents the region where we have no observational data points. However we have extrapolated the Schechter functions to this region also.

Figure 3.2 and 3.4 shows the best-fit Schechter function parameters (M_* , ϕ_* and α from top to bottom panel) as a function of the thresholds in C_{ur} and M_r , respectively. The solid lines in each of the panel represent a parametric fit to those values of the parameters. It can be noted that the sample at each threshold contains the sample of the previous neighboring threshold. Therefore the errorbars on the Schechter function parameters are correlated in neighboring bins, although they are representative of the sample.

Figure 3.1 and 3.2 presents the HIMFs conditioned on color and how it changes with different threshold values of color C_{ur}^t . The value of α is ~ -1 (bottom panel of figure 3.2) for the sub-samples with $2.0 \leq C_{ur}^t \leq 2.4$, i.e., the slope at the low mass end for those sub-samples are flat. For the smallest sub-samples or the sub-sample with the largest color threshold, the value of amplitudes (ϕ_* s) are smaller; $\sim 16\times$ smaller than that of the total sample, for which $\phi_* = 5.3 \times 10^{-3}$. But the value of the characteristic mass ($M_* = 10.13$) is $\sim 50\%$ larger than the value of M_* of the full sample.

A sub-sample with a larger value of color threshold (C_{ur}^t) implies that the sub-sample contains less number of blue galaxies and more red galaxies. Decrease in the threshold

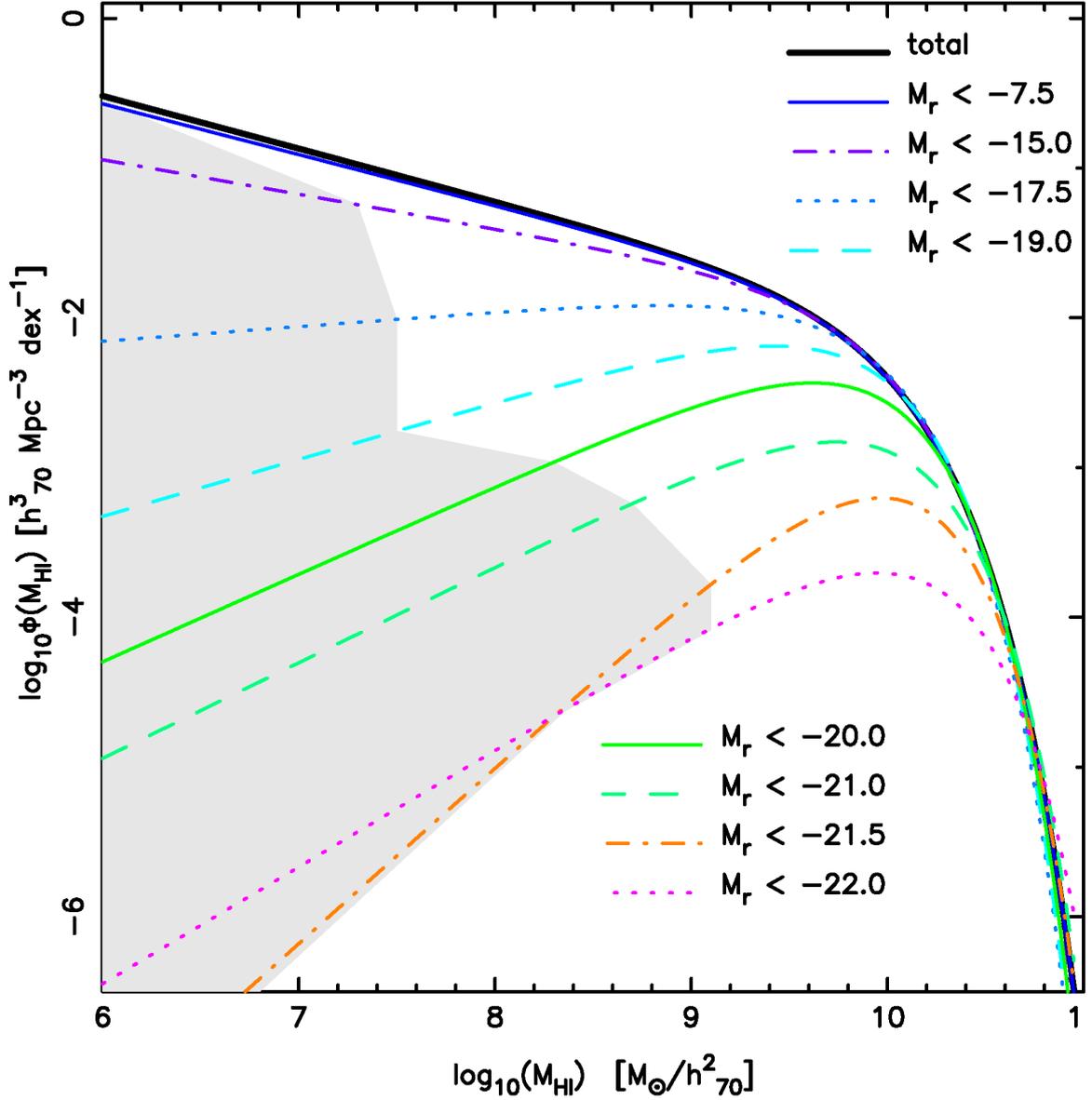


Figure 3.3: Conditional HIMF as a function of decreasing rest frame magnitude thresholds (top to bottom). The thick solid line is the HIMF for the full sample. The shaded gray region does not contain data, the conditional HIMF have however been extrapolated into this regime as well.

value represents addition of blue galaxies to the sample and then the conditional HIMF approaches the total HIMF in the limit $C_{\text{ur}}^t \Rightarrow -\infty$. In our sample this is achieved when $C_{\text{ur}}^t = 0$. Figure 3.2 shows a near monotonic change in the shape for M_* and ϕ_* with C_{ur}^t . Although α as a function of C_{ur}^t peaks at about $C_{\text{ur}}^t = 1.9$, the variation is still consistent with a constant value beyond that. Incidentally the peak in α occurs close to the value of the optimal divider of Baldry et al. [2004] at $C_{\text{ur}} = 2.3$ (refer to figure 3.5). The red population dominates the HIMF at the large mass end whereas decreasing the color threshold, C_{ur}^t , we progressively add bluer galaxies to our sample which start to dominate the knee and then the low mass end for even smaller values of C_{ur}^t .

Figure 3.3 and 3.4 shows the dependence of the M_r -conditioned HIMF on the thresholds, M_r^t . Unlike the color-conditioned case, the dependence of the conditional HIMF on M_r^t is not monotonic (figure 3.4). We see a dip (bump) in M_* (ϕ_*) at $M_r^t = -19$ ($M_r^t \sim -17.5$). Coincidentally the distribution of the blue (red) population of galaxies is centered at $M_r = -19$ ($M_r = -20$) (refer to figure 2.5 of chapter 2). As we move from the luminous ($M_r^t \leq -20$ dominated by the red sample), to the faint end, the conditional HIMF picks the contribution from the blue cloud at $M_r = -19$. The bimodality of the underlying optical galaxy sample is reflected more strongly in the luminosity-conditioned HIMF than the color-conditioned HIMF.

3.4.1 The Distribution of Ω_{HI} in the $C_{\text{ur}} - M_r$ plane

We generalize the definition of conditional HIMF for two-dimensions as:

$$\phi(M_{\text{HI}}|C_{\text{ur}}^t, M_r^t) = \phi(M_{\text{HI}})_{(C_{\text{ur}} > C_{\text{ur}}^t), (M_r < M_r^t)} \quad (3.4)$$

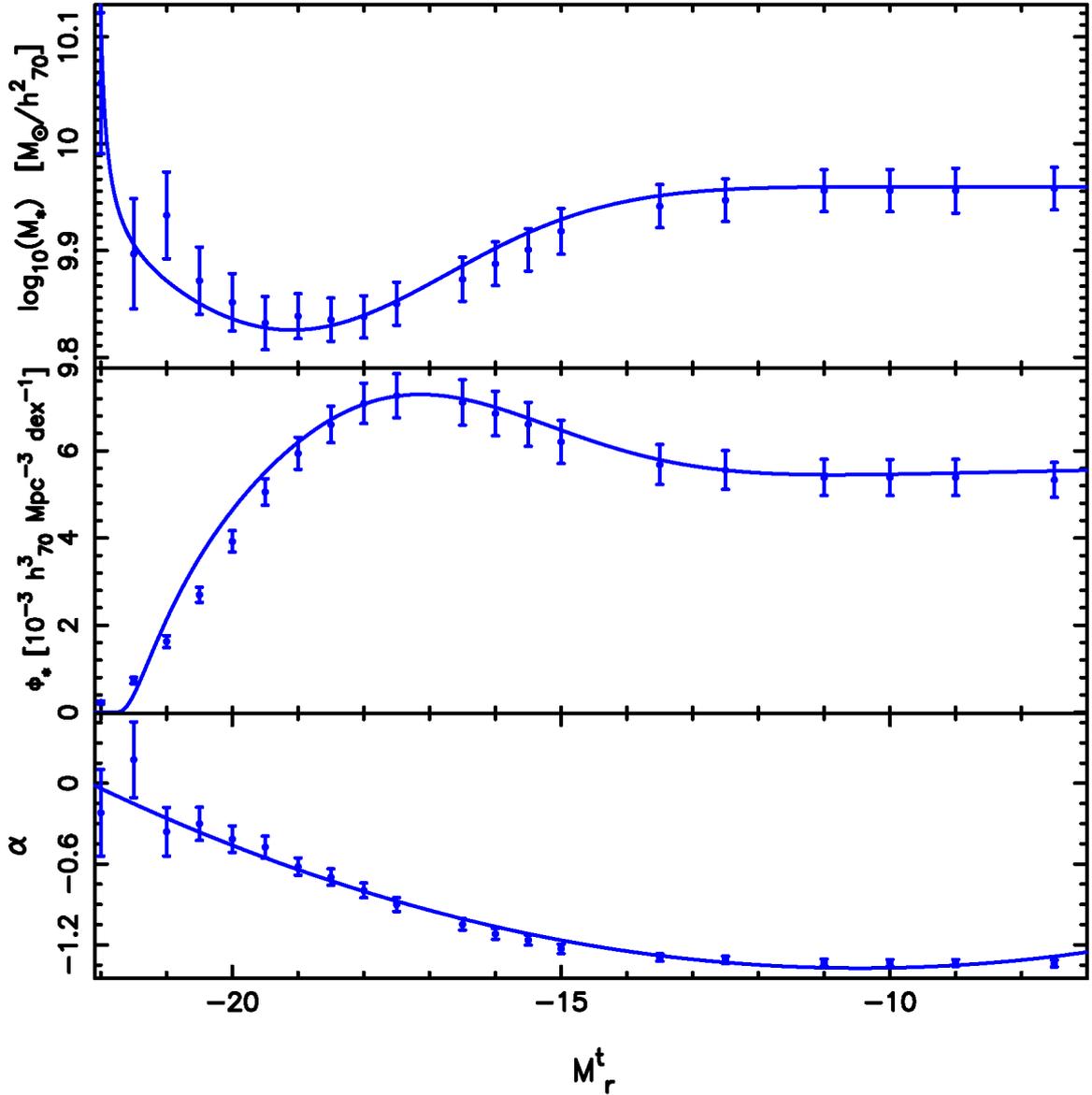


Figure 3.4: The Schechter function parameters of the conditional HIMF and their uncertainties as a function of magnitude thresholds M_r^t . The solid lines are fits to the data points. For α (bottom) we fit with a quadratic function. For M_* (top) and ϕ_* (middle) we fit with a function of the form: $y(x) = \left[a + b \exp\left(-\frac{(x+c)^2}{2d}\right) \right] \frac{f}{(x+e)}$

which is the HIMF for a sample of galaxies brighter than M_r^t , and redder than C_{ur}^t . The HI density parameter for this is:

$$\Omega_{\text{HI}}(C_{ur}^t, M_r^t) = \frac{1}{\rho_c} \int_0^\infty M_{\text{HI}} \phi(M_{\text{HI}} | C_{ur}^t, M_r^t) dM_{\text{HI}} \quad (3.5)$$

The value of HI density parameter for the full sample, i.e., $\Omega_{\text{HI}}^{\text{tot}}$, is 4.894×10^{-4} (see chapter 2 table 2.2). The full sample can be defined with the color threshold value $C_{ur}^t = 0.0$ and magnitude threshold value $M_r^t = -6.0$. We compute 2500 conditional HIMFs and their associated errors in the CM plane by dividing $M_r^t \in [-23.0, -6.0]$ (increasing magnitude threshold) and $C_{ur}^t \in [3.0, 0.0]$ (decreasing color threshold). into 50 bins each.

From equation 3.5 we see that the variation of $\Omega_{\text{HI}}(C_{ur}^t, M_r^t)$ is that of a cumulant in the two-dimensional CM plane. If we define the normalized conditional HI density parameter as $\Omega_{\text{HI}}^{\text{norm}}(C_{ur}^t, M_r^t) = \frac{\Omega_{\text{HI}}(C_{ur}^t, M_r^t)}{\Omega_{\text{HI}}^{\text{tot}}}$, then $\Omega_{\text{HI}}^{\text{norm}}(C_{ur}^t, M_r^t)$ is bounded and varies from 0 (luminous-red, top left corner of figure 3.5) and 1 (faint-blue, bottom right corner of figure 3.5).

The distribution function of the HI density parameter $\Omega_{\text{HI}}(C_{ur}, M_r)$ in the CM plane can be defined as

$$p(\Omega_{\text{HI}}(C_{ur}, M_r)) = \left. \frac{\partial^2 \Omega_{\text{HI}}^{\text{norm}}(C_{ur}^t, M_r^t)}{\partial C_{ur}^t \partial M_r^t} \right|_{[C_{ur}^t=C_{ur}, M_r^t=M_r]} \quad (3.6)$$

This distribution function is normalized by the way it has been constructed, i.e.,

$$\int \int p(\Omega_{\text{HI}}(C_{ur}, M_r)) dC_{ur} dM_r = 1.0 \quad (3.7)$$

Therefore the cosmological HI density in the ij^{th} pixel of color-magnitude plane is

$$(\Omega_{\text{HI}}^{ij})^{\text{norm}} = \int_{M_r^i}^{M_r^{i+1}} \int_{C_{ur}^j}^{C_{ur}^{j+1}} p(\Omega_{\text{HI}}(C_{ur}, M_r)) dC_{ur} dM_r \quad (3.8)$$

Figure 3.5 presents the distribution of the cosmological HI density parameter, $p(\Omega_{\text{HI}})$, in the color-magnitude $(C_{ur} - M_r)$ plane. the pixels are color-coded to the $(\Omega_{\text{HI}}^{ij})^{\text{norm}}$ value. The

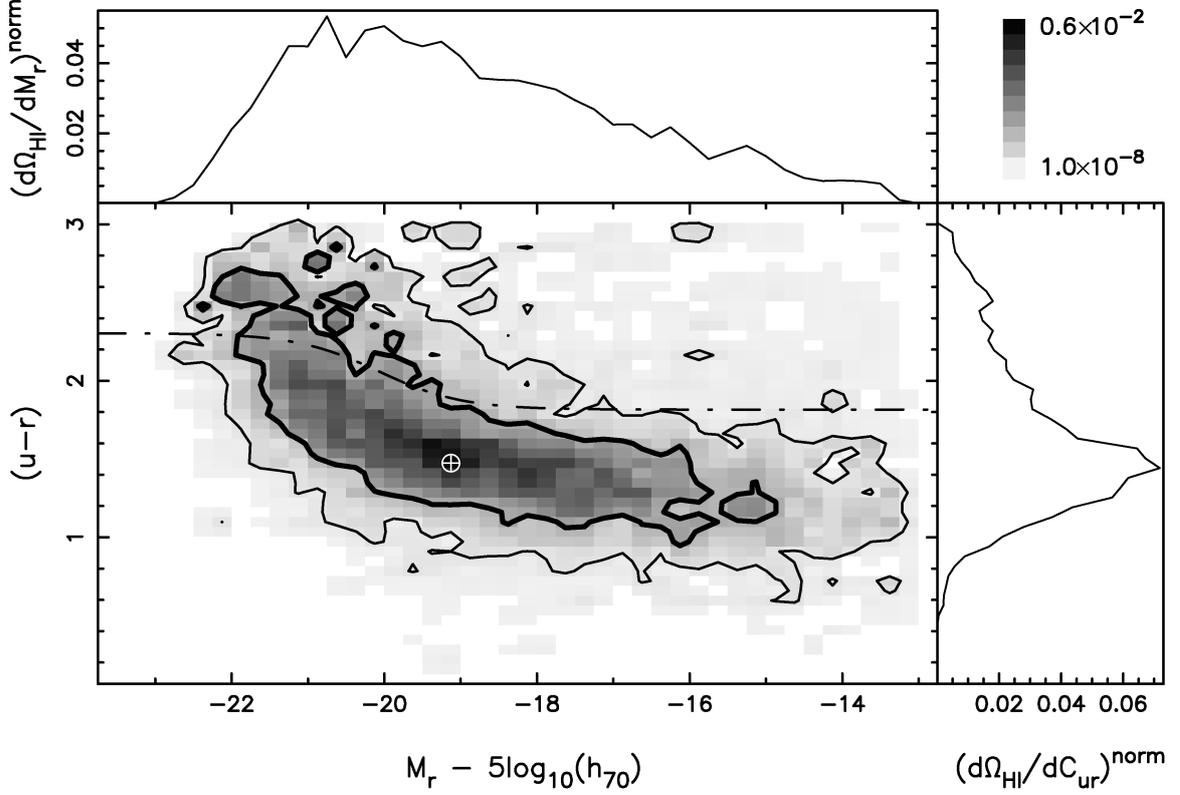


Figure 3.5: The bottom left panel shows the distribution function $p(\Omega_{HI})$ (see eq. 3.6) in the CM plane color coded by $(\Omega_{HI}^{ij})^{\text{norm}}$ (eq. 3.8). The thick (thin) line represent the 1σ (2σ) widths of $p(\Omega_{HI})$. The dash-dot line separates the optical red (above) and blue (below) populations [Baldry et al., 2004]. The top left (bottom right) panel is the marginalized distribution of Ω_{HI} as a function of M_r (C_{ur}). The crossed circle represents the peak of the two-dimensional distribution function, $p(\Omega_{HI})$.

bottom right panel shows the distribution of Ω_{HI} as a function of $(u-r)$ color and the top left panel shows that as a function of M_r . The demarcation of the red and blue populations has been shown by the dot-dashed line, as given in Baldry et al. [2004] (eqn. 2.2). The 1σ (68%) and 2σ (95%) width of the distribution function are marked by the thick and thin contours, respectively. The peak of the distribution is denoted by the white crossed-circle. Since the distribution, $p(\Omega_{HI}(C_{ur}, M_r))$ is skewed, the peak in the 2D distribution is different from the peak of the marginalized distributions.

3.5 Discussion

This chapter presents the conditional HIMFs, conditioned on color and/or magnitude. We estimate the distribution of Ω_{HI} in CM plane, $p(\Omega_{\text{HI}}(C_{\text{ur}}, M_r))$, based on the conditional HIMFs. Our results for $\phi(M_{\text{HI}})|_{M_r < -21}$ and even brighter thresholds is similar to those obtained for the conditional HIMF, $\phi(M_{\text{HI}})|_{M_{\text{star}} \geq 10}$, for massive galaxies [Lemonias et al., 2013] from the GASS survey [Catinella et al., 2010, 2012]; this is because the stellar mass of galaxies is correlated with its luminosity.

The two-dimensional distribution and the marginalized distribution, both, show that they have a long tail towards faint blue galaxies and luminous red galaxies. The peak of $p(\Omega_{\text{HI}})$ in the CM plane occurs at $C_{\text{ur}}^{\text{max}} = 1.44$, $M_r^{\text{max}} = -19.25$ in the blue cloud, which is about 1.36 magnitude fainter than the characteristic luminosity of blue galaxies in SDSS [Baldry et al., 2004]. The width of $p(\Omega_{\text{HI}})$ is also fairly broad in both color and magnitude. The 1σ and 2σ widths in $(u - r)$ color are 0.8 and 1.1, respectively; and 3.0 and 4.8 in M_r . At the fainter end $M_r > -16$, $\sim 10\%$ of $\Omega_{\text{HI}}^{\text{tot}}$ is locked in gas rich low surface brightness galaxies. On the other hand, the red population contributes $\sim 18\%$ to the HI budget.

The CM plane can be thought of as a coordinate system in which we can plot distributions of other cosmological density parameters (related to galaxies), $p(\Omega_X)$ where X denotes a property, e.g. stellar mass M_{star} , SFR , molecular hydrogen mass M_{H_2} , which in turn are computed from $\phi(X|C_{\text{ur}}^t, M_r^t)$. We therefore have all the information needed to obtain the mean relation between different galaxy properties by discarding the common coordinate system, i.e., $C_{\text{ur}} - M_r$. We emphasize that this relation is unbiased and represents the underlying relation since the distributions have folded in the survey selection. The blind nature of the survey is also important since there is no selection bias in estimating $\phi(X)$. This can be repeated for different galaxies populations (blue or red) and for other bands as well. The methods outlined in this chapter are statistical in nature and provide a powerful

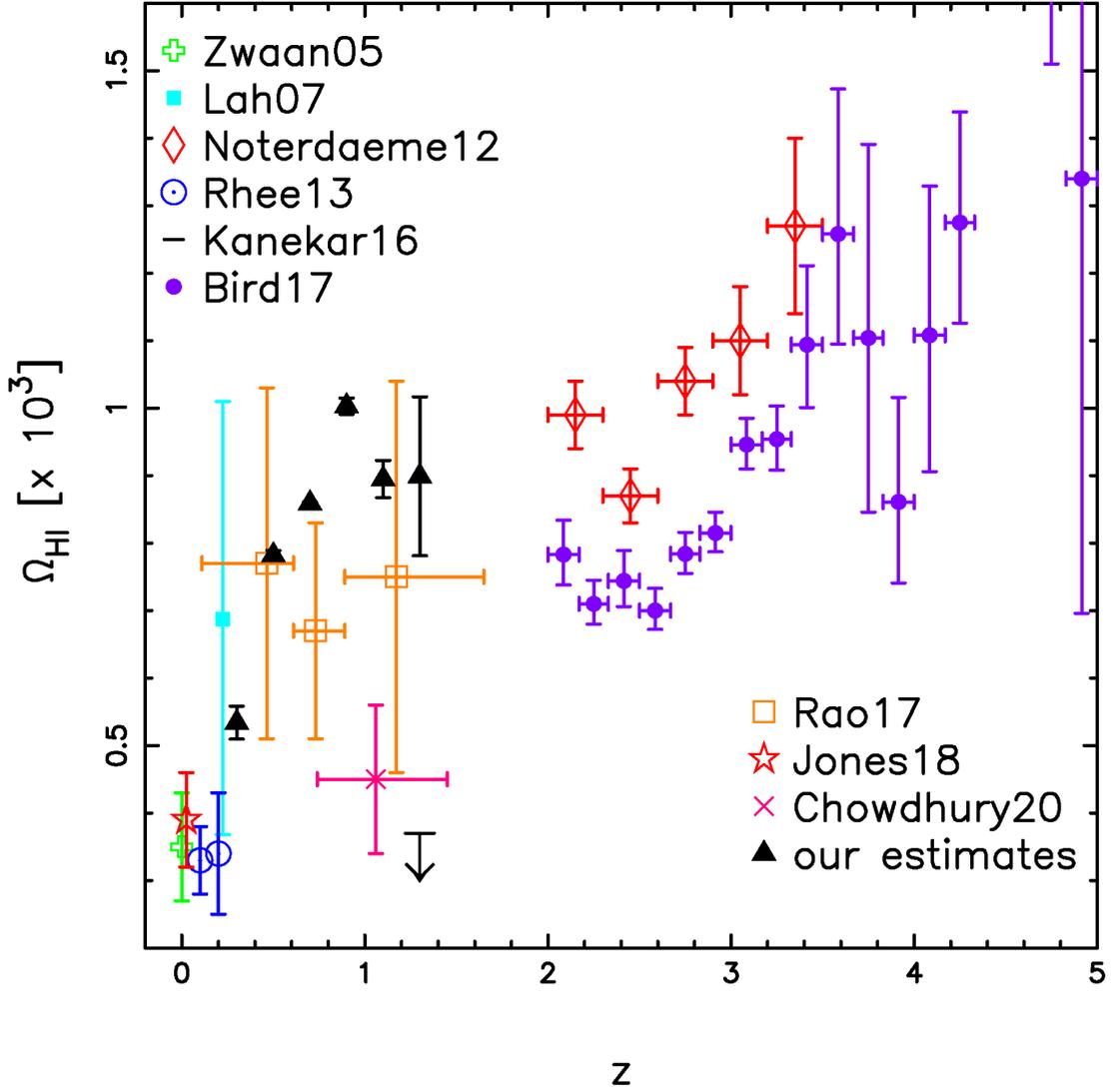


Figure 3.6: HI density, Ω_{HI} , as a function of redshift, z . The black up-triangles are our estimates of Ω_{HI} . The green plus sign and red star are the estimates from Zwaan et al. [2005] for HIPASS galaxies and Jones et al. [2018] for ALFALFA galaxies, respectively, at $z \sim 0$, using 21 cm emission line. The cyan filled square is measured by Lah et al. [2007] by stacking analysis using the Giant Metrewave Radio Telescope (GMRT). The black arrow line is the upper limit estimated by Kanekar et al. [2016] using the GMRT HI stacking. The red open diamonds, violet filled circles, and orange open squares are the measurements from damped Lyman- α from the Hubble Space Telescope (HST) and the SDSS by Noterdaeme et al. [2012], Bird et al. [2017], and Rao et al. [2017], respectively. The blue dot circles are the estimates by Rhee et al. [2013] using the Westerbork Synthesis Radio Telescope (WSRT) and HI stacking. The magenta cross is the estimate of Ω_{HI} by Chowdhury et al. [2020] using upgraded GMRT(uGMRT) and DEEP2 galaxy redshift survey.

and unbiased way to probe the multivariate distributions of galaxy populations.

We end this section by discussing how stacking of HI is used to estimate Ω_{HI} and discuss possible biases in this approach. Figure 3.6 shows the observational estimates of Ω_{HI} across different redshifts. The Ω_{HI} values obtained in the local Universe ($z < 0.05$) are using data from blind HI surveys [HIPASS, Zwaan et al., 2005, green plus point], [ALFALFA, Jones et al., 2018, red star point]. The measurements at higher redshifts ($z > 2$) are based on large samples of damped Lyman- α (DLA) systems [Noterdaeme et al., 2012, red open diamonds], [Bird et al., 2017, violet filled circles]. Rao et al. [2017] (orange open squares) also estimate the Ω_{HI} from DLAs at redshifts $0.4 < z < 1.4$. Other low redshift measurements [Lah et al., 2007, cyan filled square], [Kanekar et al., 2016, upper limit, black downward arrow], [Rhee et al., 2013, blue dot circle], [Chowdhury et al., 2020, magenta cross] are estimated by stacking HI on known optically selected galaxies. These stacking analysis target the bright blue cloud of galaxies to obtain the HI density parameter under the assumption that HI is only associated with these galaxies. Our result (figure 3.5) suggests, $\sim 10\%$ and $\sim 17\%$ of $\Omega_{\text{HI}}^{\text{tot}}$ is coming from the low surface brightness galaxies and the red galaxies, respectively. Therefore the Ω_{HI} estimates from HI stacking which consider only the bright blue population will underestimate the Ω_{HI} . We present a case of Ω_{HI} estimation considering the B-band luminosity function, $\phi(L_B)$, at different redshifts, $z = 0.3, 0.5, 0.7, 0.9, 1.1, 1.3$, from Willmer et al. [2006]

$$\phi(L_B) = \frac{\phi_*}{L_{B*}} \left(\frac{L_B}{L_{B*}} \right)^\alpha \exp\left(-\frac{L_B}{L_{B*}}\right) \quad (3.9)$$

and assuming a non-evolving scaling relation between M_{HI} and M_B [Dénes et al., 2014]

$$\frac{M_{\text{HI}}}{L_B} = K \left(\frac{L_B}{L_{B*}} \right)^\beta \quad (3.10)$$

where, ϕ_* is the amplitude of the luminosity function, L_{B*} is the characteristic luminosity, and α is the faint end slope. We use the value of the coefficient $\beta = 0.15$, obtained in the

local Universe [Dénes et al., 2014], and $K = 0.73 \frac{M_{\odot}}{L_{B\odot}}$ is fixed based on stacking of HI in these galaxies at $z = 0.3$ [Bera et al., 2019]. The black up-triangles in figure 3.6 show our estimates of Ω_{HI} . Though these estimates are within the uncertainty levels of previous results, it show a rapid increase in the values of Ω_{HI} from redshift $z = 0.3$ to $z = 0.9$ and flattens thereafter. We emphasize that this method uses stacking data at $z = 0.3$ to calibrate the $z = 0$ scaling relation (eq. 3.10). Finally, even under these assumptions we would miss out cold gas in low surface brightness galaxies and red galaxies. We need a better understanding of the relation between HI and optical properties of galaxies and our work presented in this chapter is a first step in this direction.

Chapter 4

The Dark Matter Halos of HI Selected Galaxies

4.1 Introduction

As per our current understanding, gas cools down and collapses within the potential wells provided by dark matter halos. This leads to the formation of galaxies inside dark matter halos. We still need a clear picture of the process of star formation and the evolution of galaxies inside these halos. To comprehend the formation and evolution of galaxies, we need to understand how these galaxies are distributed inside the dark matter halos. HI represents the cold gas which eventually leads to gravitational collapse to form stars. Therefore, by exploring the HI-galaxy-halo connection we can probe the star formation processes inside galaxies. In chapter 2 and 3 we looked at the HI-galaxy relation. As a next logical step we look into how HI and halos are connected.

Cosmological hydro-dynamical simulations are a direct but expensive way to understand the process of galaxy formation [[Vogelsberger et al., 2014](#), [Khandai et al., 2015](#), [Schaller et al., 2015](#), [Davé et al., 2017](#)]. Another popular way of investigating the formation of galaxies and clusters within dark matter halos is semi-analytic models [[White and Frenk, 1991](#), [Lacey and Silk, 1991](#), [Kauffmann et al., 1993](#), [Lacey et al., 2016](#), [Zoldan et al., 2016](#), [Kim et al., 2017](#)]. The other approach is the halo occupation distribution (HOD) models

[Cooray and Sheth, 2002], which have been used to model HI in halos [Guo et al., 2015, Paul et al., 2018].

Abundance matching technique is an alternate self-consistent method to connect the halos with galaxy properties. Unlike HOD models this technique does not require any fitting parameters. Simply using the SMF and the HMF, and equating the abundances for stellar mass and halo mass it gives a scaling relation between M_{star} and M_h . This abundance matching technique can be used for any two properties as long as they have a monotonic relation. The observed distribution of SDSS (optical) and ALFALFA (HI) galaxies (figure 2.5) shows that the HI detection rate in the red galaxies ($\sim 11\%$) are quite less compared to that of the blue galaxies (38%). The smaller HI detection rate in the bright red galaxies suggests that the bright red population of galaxies does not host considerable amount of HI. Serra et al. [2012] also report that HI detection in the early-type galaxies (which are generally associated with the red population) are less compared to that of the late-type galaxies (i.e. blue galaxies) in the Atlas^{3D} survey [Cappellari et al., 2011]. M_{star} , and hence M_r , is strongly correlated with M_h [Behroozi et al., 2010]; which implies that the halo mass and HI mass are not correlated. However the monotonic characteristic in the observed distribution of the ALFALFA galaxies (figure. 2.13), in the $M_{\text{HI}} - M_{\text{star}}$ plane, suggests that for an HI-selected sample M_{HI} is correlated to M_{star} , and so to M_h . Hence, for an HI-selected sample the abundance matching method can be used to connect M_{HI} with M_h .

In this chapter, we use the abundance matching technique to obtain the scaling relations between HI and halo properties. We define an HI-selected HMF, $\phi^{\text{HI}}(M_h)$, to carry out this method using an observational non-monotonic mean HI-halo mass relation from Guo et al. [2020]. Section 4.2 describes the data we use for this work. We briefly discuss the steps to estimate the HIWF in section 4.3. In section 4.4 we present the estimates of HIMF, HIWF, and HIVF for the total sample and red and blue populations of galaxies. We describe the abundance matching method and present the scaling relations between HI and

halo properties in section 4.5. Finally in section 4.6 we discuss our results.

4.2 Data

For this work, we use the HI-selected galaxies from the 40% ($\alpha.40$) data catalog of the ALFALFA survey [Haynes et al., 2011]. Most of these galaxies have been cross-matched with an optical detection in SDSS DR7 [Abazajian et al., 2009]. This is the same subsample as discussed in chapter 2. For completeness, we briefly describe it again in this chapter. We consider only the code 1 objects with $S/N > 6.5$ for our analysis and restrict our sample to $cz_{\text{cmb}} = 15000 \text{ km.s}^{-1}$ to avoid Radio Frequency Interference (RFI) [Martin et al., 2010, Haynes et al., 2011]. In this work we consider an area ($\sim 2093 \text{ deg}^2$) common to both ALFALFA and SDSS (as discussed in chapter 2), which contains 8344 HI-selected galaxies.

The distribution of the observed HI-selected galaxies in the $S_{21} - w_{50}$ plane (see figure 2.4 in chapter 2) shows that at fixed value of integrated flux, galaxies with narrower velocity profile widths are more probable to be detected. The sensitivity limit depends therefore on both flux and velocity width. We apply the sensitivity limit given by the 50% completeness limit [Haynes et al., 2011], which reduces the sample to 7857 galaxies. Among these, 148 galaxies (referred to as *dark galaxies*) do not have any optical counterparts in SDSS DR7 although they belong to the DR7 footprint. We exclude these galaxies in this work as they will not affect our results (see chapter 2). For the remaining 7709 galaxies we extract the extinction corrected (for our galaxy) *ugriz* magnitudes and ages from SDSS. We use *kcorrect* [Blanton and Roweis, 2007] to obtain rest frame magnitudes (M_u, M_g, M_r, M_i, M_z) and stellar masses (M_{star}) for these HI-selected galaxies. We also obtain galaxy age estimates from SDSS based on the Granada Flexible Stellar Population Synthesis (FSPS) models [Conroy et al., 2009, Ahn et al., 2014]. We point out that while

using the estimates from *kcorrect* we have not corrected for internal reddening due to dust, but as shown in chapter 2 our results do not change considerably with those of Huang et al. [2012], who have corrected for reddening by using two additional UV bands of GALEX.

A clear bimodality is seen in the distribution of SDSS galaxies (see figure 2.5) in the color ($u - r$) - magnitude (M_r) plane. We use the optimal divider [Baldry et al., 2004] (eq. 2.2) to further classify our sample as red or blue galaxies. Red(blue) galaxies lie above(below) the optimal divider defined in the color-magnitude plane. We have $N_{\text{gal}}^{\text{red}} = 1290$ red galaxies and $N_{\text{gal}}^{\text{blue}} = 6419$ blue galaxies in our observed sample.

4.3 Estimating the HI Velocity Width Function

Similar to the HI mass function (HIMF), the HI velocity width function (HIWF) can be defined as the underlying number density of galaxies with velocity widths in the range $[w_{50}, w_{50} + dw_{50}]$,

$$\phi(w_{50}) = \frac{1}{V} \frac{dN}{dw_{50}} \quad (4.1)$$

where, dN is the total number of galaxies in volume V having velocity widths within w_{50} and $(w_{50} + dw_{50})$. The HIWF is well described by a modified Schechter function [Zwaan et al., 2010, Papastergis et al., 2011, Moorman et al., 2014]

$$\begin{aligned} \phi(w_{50}) &= \frac{dn}{d \log_{10} w_{50}} \\ &= \ln(10) \phi_* \left(\frac{w_{50}}{w_*} \right)^\alpha \exp \left[- \left(\frac{w_{50}}{w_*} \right)^\beta \right] \end{aligned} \quad (4.2)$$

where ϕ_* is the amplitude, α is the slope at the low velocity width end, w_* is the characteristic velocity width, or the knee of the Schechter function, and β modifies the exponential suppression at high velocity widths. For the rest of the chapter we will quote the amplitude of the Schechter function, ϕ_* , and the knee of the HIWF, w_* , in units of $(10^{-3} h_{70}^3 \text{Mpc}^{-3} \text{dex}^{-1})$ and $\log_{10}[w_*/(\text{km.s}^{-1})]$ respectively. Later on we will also look at the HI velocity function

(HIVF), described by a modified Schechter function whose knee, V_* , will similarly be in the same units as w_* . Finally the values of the HI mass, M_{HI} , (and the knee of the HIMF, M_*) will be in units of $\log(M_{\text{HI}}/M_{\odot}) + 2 \log h_{70}$.

The stepwise maximum likelihood (SWML) method [Efstathiou et al., 1988] is used to estimate the underlying velocity width function. This model-independent method does not assume any functional form of the width function, rather it estimates a discretized (or binned) width function, $\phi(w_{50}^k)$. Being a maximum likelihood method it is insensitive to effects of local variations in galaxy densities due to clustering. For HI-selected galaxies the detection probability depends both on M_{HI} and w_{50} . We implement a two-dimensional SWML method (2DSWML) to first estimate the bivariate HI mass-velocity width function, $\phi(M_{\text{HI}}, w_{50})$ and then integrate over M_{HI} to obtain the HIWF. A common feature of maximum likelihood methods is that the normalization of $\phi(M_{\text{HI}}, w_{50})$ needs to be fixed separately since it gets lost in the process. We do this by computing the selection function and matching the observed number density to the underlying number density convolved by the selection function [Davis and Huchra, 1982, Martin et al., 2010]. A detailed description of the 2DSWML method is given in appendix A.

4.3.1 Uncertainties on HIWF

We consider The following four sources of error and add them in quadrature to quantify the uncertainty on the estimated HIWF, similar to Moorman et al. [2014].

1. **Velocity width errors:** Errors due to the measurement of w_{50} can change the occupation of one galaxy in $M_{\text{HI}} - w_{50}$ plane. To account for this we consider 300 realizations of w_{50} (Gaussian random realizations using w_{50} as the mean and $\sigma_{w_{50}}$ as the variance) and estimate HIWF for each of these realizations. The distribution of ϕ_k for a fixed k gives an estimate of the uncertainty on ϕ_k .

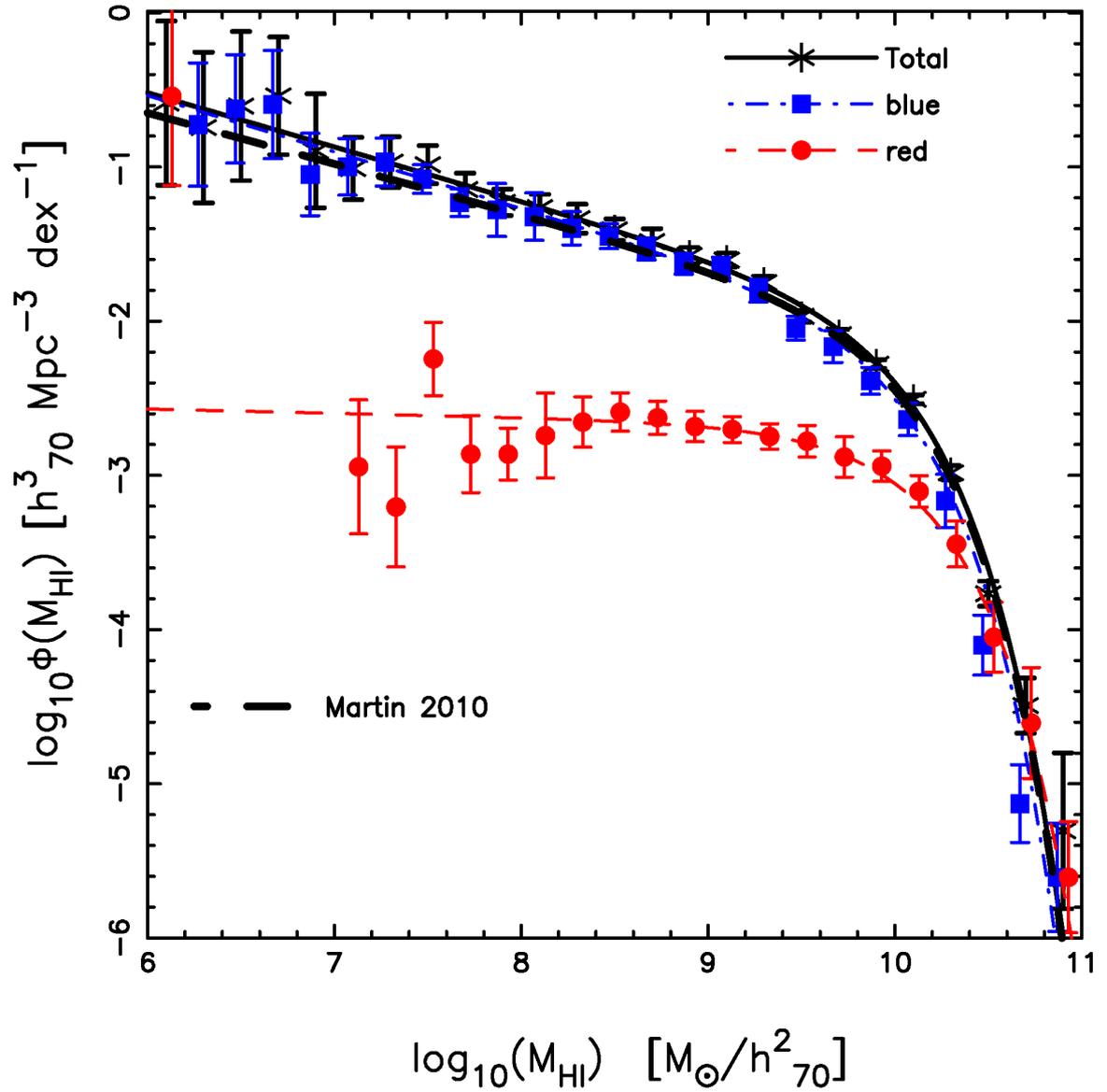


Figure 4.1: Our estimates of the HIMFs are shown for the total (black cross), blue (blue filled square) and red (red filled circle) samples. The Schechter function fits are plotted for the total (black thick solid line), red (red thin dashed line) and blue (blue thin dot-dashed line) samples for the HIMFs. The black thick dashed line is the estimate of the HIMF from [Martin et al. \[2010\]](#) using the $\alpha.40$ sample. To better display the data for the red and blue samples has been horizontally shifted with respect to the total sample.

2. **Distance errors:** HI mass (M_{HI}) depends on distance, and since we estimate the HIWF from a bivariate function of M_{HI} and w_{50} , the measurement errors on distance is also a source of uncertainty. We follow a similar approach as above to get the errors on ϕ_k , by generating 300 random realizations.
3. **Sample variance:** We split our survey area into 26 regions of approximately equal angular area and estimate the HIWF by eliminating one region at one time (the Jackknife sample). We then estimate the Jackknife uncertainty as $\sigma_{\phi_k} = \frac{N-1}{N} \sum_{i=1}^N (\bar{\phi}_k - \phi_k^i)^2$ where N is the number of Jackknife samples, $\bar{\phi}_k$ is the Jackknife mean and ϕ_k^i is the value for the i^{th} Jackknife sample.
4. **Poisson errors:** The observed counts at the two ends of $\phi(w_{50})$ are small, it is therefore important to consider Poisson errors.

4.4 HI Velocity Width Function

In this section we present the HIMF and the HIWF for the full sample as well as for the red and blue samples. We then use a deconvolution (or inversion) method [Papastergis et al., 2011] to estimate the HIFV from the HIWF.

In figure 4.1 and 4.2 we present our estimates of the HIMF and the HIWF, respectively, for the total (black cross), red (red filled circle) and blue (blue filled square) samples. Error estimates on the HIMF are described in appendix A and are similar to the error estimates for the HIWF. To better display our data we have horizontally offset data for the red and blue samples with respect to the total sample. The black thick solid, red thin dashed and blue thin dot-dashed lines are our Schechter (modified Schechter) function fits to the HIMF (HIWF) for the total, red and blue samples. We compare our results for the HIMF (HIWF) with Martin et al. [2010] [Moorman et al., 2014] (black thick dashed line). We find that our estimates compare well with these authors for the total sample. Moorman et al. [2014]

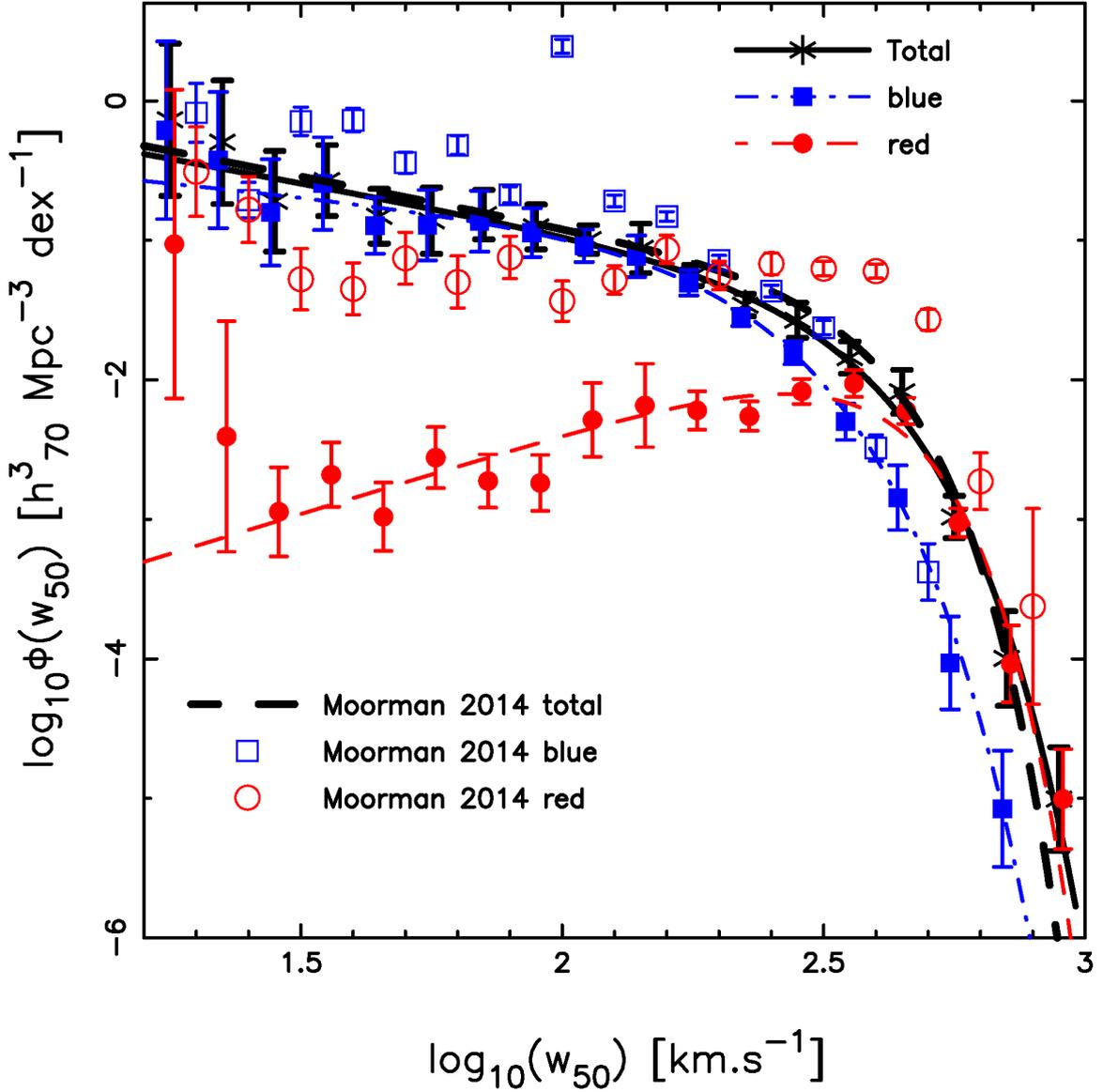


Figure 4.2: Our estimates of the HIWFs for the total (black cross), blue (blue filled square) and red (red filled circle) samples. The modified Schechter function fits are plotted for the total (black thick solid line), red (red thin dashed line) and blue (blue thin dot-dashed line) samples for the HIWFs. To better display the data for the red and blue samples has been horizontally shifted with respect to the total sample. The black thick dashed line is the estimate of the HIWF from Moorman et al. [2014] using the $\alpha.40$ sample. The open red circles (blue squares) are the unnormalized estimates of the HIWF for the red (blue) sample from Moorman et al. [2014].

also estimated the HIWF for the red and blue samples (red open circles and blue open squares in figure 4.2) but they are most likely unnormalized since they do not add up to give the HIWF for the total sample. The overall shape of the red sample compares well with [Moorman et al. \[2014\]](#) however since the binning is different it is difficult to make a point by point comparison. [Moorman et al. \[2014\]](#) have an outlier at $w_{50} \approx 100 \text{km.s}^{-1}$. We find that it comes from a single galaxy which is extremely gas poor with $M_{\text{HI}} = 6.78$. We have removed this object in our analysis.

Galaxies have random inclinations, therefore we relate the observed HI profile width, w_{50} , to the intrinsic HI rotational velocity of the galaxy, $V_{\text{rot}}^{\text{HI}}$, by

$$w_{50} = 2V_{\text{rot}}^{\text{HI}} \sin i + w_{\text{nr}} \quad (4.3)$$

This is similar to the approach taken by [Zwaan et al. \[2010\]](#), [Papastergis et al. \[2011\]](#). Here i is the inclination angle and w_{nr} is an additional term which captures broadening by turbulence and other non-rotational motion.

Although it is tempting to relate $V_{\text{rot}}^{\text{HI}}$ to rotational velocities, V_{rot} , associated with rotation curve measurements by radio synthesis observations, one needs to be careful in this regard. [Verheijen \[2001\]](#) classified rotation curves broadly as – i. rising (R-type) ii. flat (F-type), and iii. declining (D-type) rotation curves. R-type rotation curves [[Li et al., 2020](#)] are generally associated with low-surface brightness or dwarf galaxies and the observed maximum rotational velocity, V_{max} , is the last point of the rotation curve measurement. It represents a lower bound to the rotational velocity, V_{rot} , associated with the gravitational potential. F-type rotation curves [[Li et al., 2020](#)] are the classical flat rotation curves that extend beyond the optical radius. The rotation curves initially rise, reach a maximum value V_{max} , and taper off to a slightly lower, constant value, $V_{\text{flat}} \lesssim V_{\text{max}}$, at large radii. V_{flat} is associated with the maximum rotational velocity, V_{rot} , that would be induced by the mass of the halo. F-type curves are associated with late-type spirals; and $V_{\text{rot}}^{\text{HI}}$, in equation 4.3, can

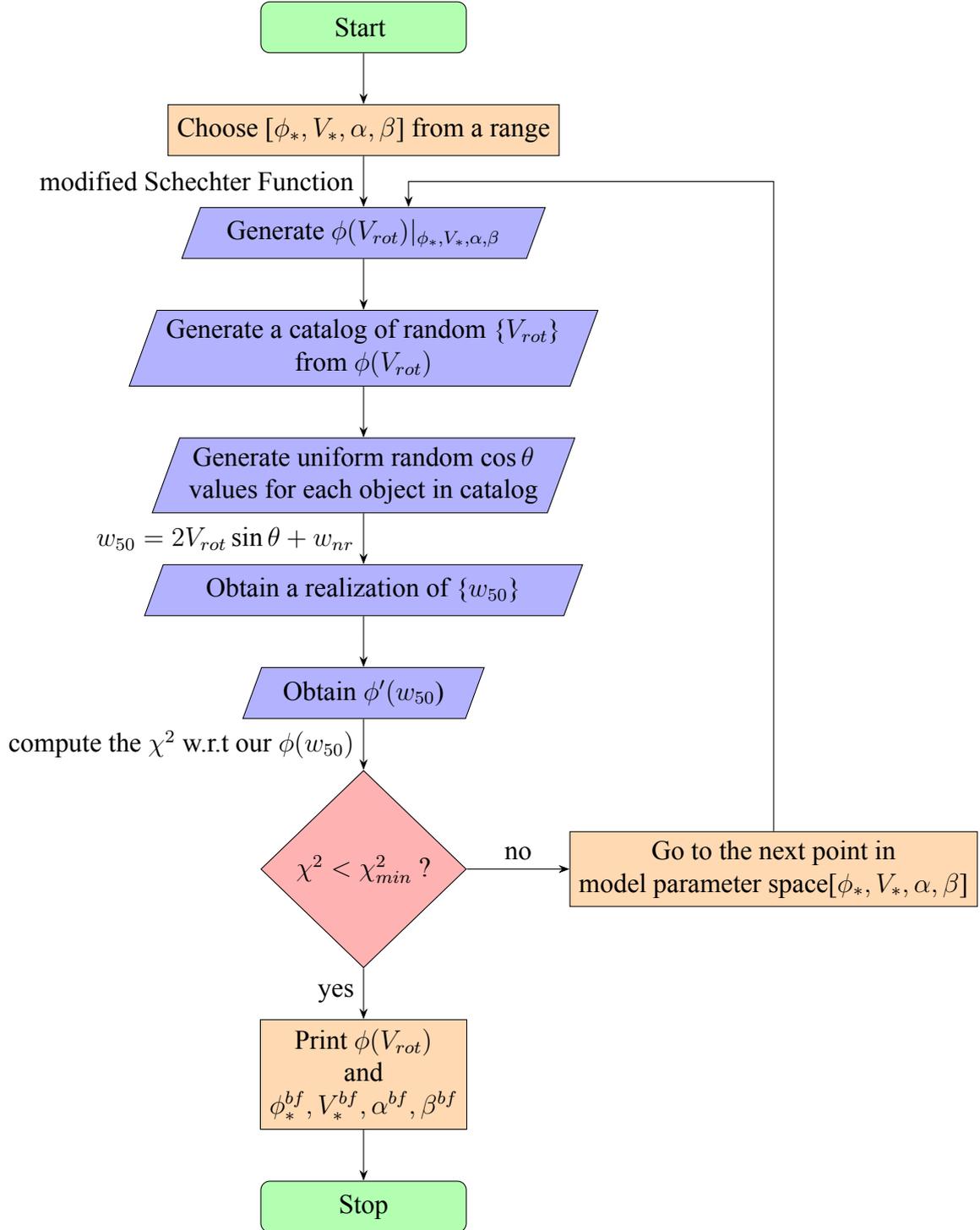


Figure 4.3: This flowchart summarizes the steps to obtain the HIVE.

be associated with $V_{\text{flat}} = V_{\text{rot}} = V_{\text{rot}}^{\text{HI}}$ [Verheijen, 2001]. Finally the D-type curves [Lelli et al., 2017], seen in early-type galaxies (ETGs), are characterized by an increasing rotation curve reaching to V_{max} within the stellar disk and declining thereafter beyond the optical radius to either a constant value, V_{flat} , or to a smaller value, V_{min} . In the D-type curves, $V_{\text{max}} > V_{\text{flat}}, V_{\text{min}}$.

Among the small fraction of ETGs detected in HI, in the ATLAS^{3D} survey [Serra et al., 2012] most have settled HI disk configurations which are rotating. However while creating dynamical mass models based on the rotation curves [Lelli et al., 2016, 2017] one has to also consider stellar velocity dispersion and pressure support of ionized X-ray gas in such systems [Lelli et al., 2017]. The corresponding V_{rot} based on these models is not the same as $V_{\text{rot}}^{\text{HI}}$. Pressure support is also important in dwarfs described by R-type rotation curves.

Given these considerations we will interpret $V_{\text{rot}}^{\text{HI}}$ in equation 4.3 as the HI profile width if the HI disk were seen edge-on, corrected for non-rotational broadening. In figure 4.13 we see that the $M_{\text{HI}}-V_{\text{rot}}^{\text{HI}}$ relation in our sample is broadly consistent with the $M_{\text{HI}}-V_{\text{flat}}$ measurements from the Spitzer Photometry and Accurate Rotation Curves (SPARC) [Lelli et al., 2016] sample.

Broadly there are two approaches in obtaining $V_{\text{rot}}^{\text{HI}}$ from w_{50} . One can correct for inclination effects [Zwaan et al., 2010] by estimating inclination angles from identified optical counterparts. This method however is accurate for inclination angles $i \leq 45^\circ$. Because of the ambiguity of interpreting $V_{\text{rot}}^{\text{HI}}$ with V_{rot} for all morphological types, Zwaan et al. [2010] estimated the HI velocity function (HIVF) with the 2DSWML method only for late-type galaxies with this method, after applying a correction in the abundance for galaxies with inclination angles $i > 45^\circ$. The second approach is a statistical method implemented by Papastergis et al. [2011]. We follow this approach. We assume that the HIVF is also a modified Schechter function described by equation 4.2 with parameters $(\phi_*, V_*, \alpha, \beta)$. We generate a realization of this model Schechter function and randomize their inclinations

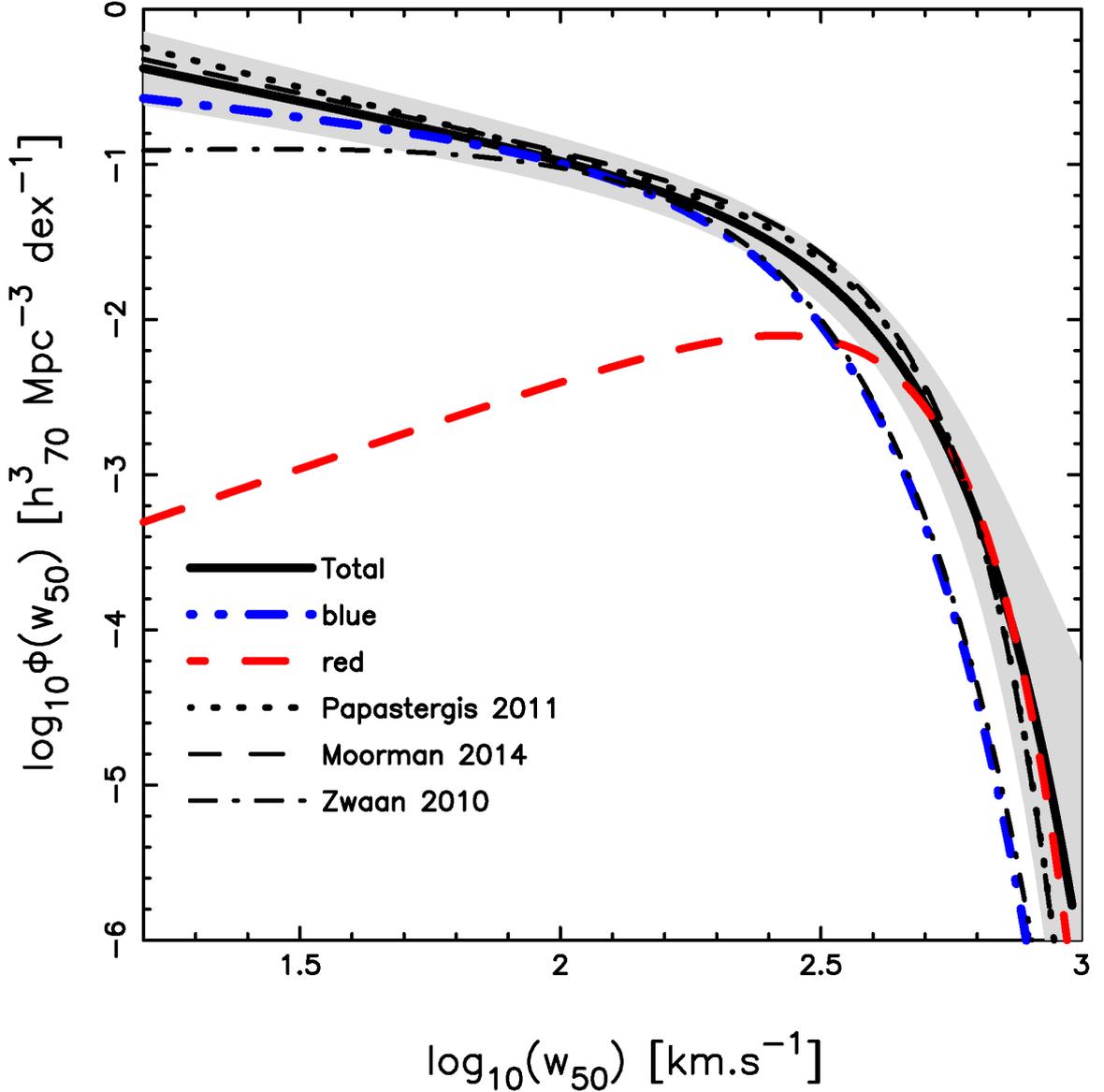


Figure 4.4: A comparison of our estimates with previously published results of the HIWF are shown in this figure. Our modified Schechter function fits and their 1σ uncertainty (shaded region for the total sample only) are plotted for the total (black thick solid line), red (red thin dashed line) and blue (blue thin dot-dashed line) samples. The black thick dotted line represents the estimates of HIWF from [Papastergis et al. \[2011\]](#). The black thin dot-dashed line is the estimates of [Zwaan et al. \[2010\]](#) from the HIPASS data for the full sample. The black thin dashed line is the estimate of the HIWF of [Moorman et al. \[2014\]](#).

($\cos i$ uniformly distributed) and use equation 4.2 to then obtain a realization of the HIWF. We use $w_{\text{nr}} = 5\text{km.s}^{-1}$ in equation 4.3 [Verheijen and Sancisi, 2001, Papastergis et al., 2011], and add it linearly for galaxies $V_{\text{rot}} > 50\text{km.s}^{-1}$ and in quadrature for galaxies with smaller velocities [Papastergis et al., 2011]. The model HIWF is compared to our estimated binned HIWF (data points with errors in figure 4.2) and a χ^2 is computed using the model, data and associated errors. Finally the best fit model parameters of the HIFV are obtained by minimizing the χ^2 with the model parameters. Figure 4.3 summarizes these steps in the form of a flowchart. We do this for the total, red, and blue samples.

Our results are shown in figure 4.4 and 4.5; and summarized in table 4.1. In figure 4.4 and 4.5, we compare our modified Schechter function fits with those obtained earlier with HIPASS [Zwaan et al., 2010] and the $\alpha.40$ sample of ALFALFA [Papastergis et al., 2011, Moorman et al., 2014]. Papastergis et al. [2011] and Moorman et al. [2014] have considered all the code 1 galaxies from the $\alpha.40$ sample, whereas we have a subsample of $\alpha.40$ (78% of the sample of Papastergis et al. [2011]) which shares a common area with SDSS DR7. The black thick solid, red thick dashed and blue thick dot-dashed lines are our modified Schechter function fits for the total, red and blue samples. The grey region is the 1σ uncertainty for the total sample. The black dotted line is the result of Papastergis et al. [2011]. The black thin dashed line in figure 4.4 is the estimate of the HIWF by Moorman et al. [2014]. The black thin dot-dashed line in figure 4.4 and 4.5 are the result of Zwaan et al. [2010] for the HIWF and the HIFV, respectively, in HIPASS.

As can be seen in figure 4.4 our result, for the HIWF agrees well, (within errors) with Papastergis et al. [2011], Moorman et al. [2014]. There is a factor of ~ 2 discrepancy in ϕ_* between Papastergis et al. [2011] ($\phi_* = 0.011 \pm 0.002$) and Moorman et al. [2014] ($\phi_* = 0.021 \pm 0.002$). Our value of $\phi_* = 0.023 \pm 0.008$ is more in agreement with Moorman et al. [2014]. However given the correlation between the parameters of the Schechter function the differences in other parameters compensate for each other and keep the overall shape

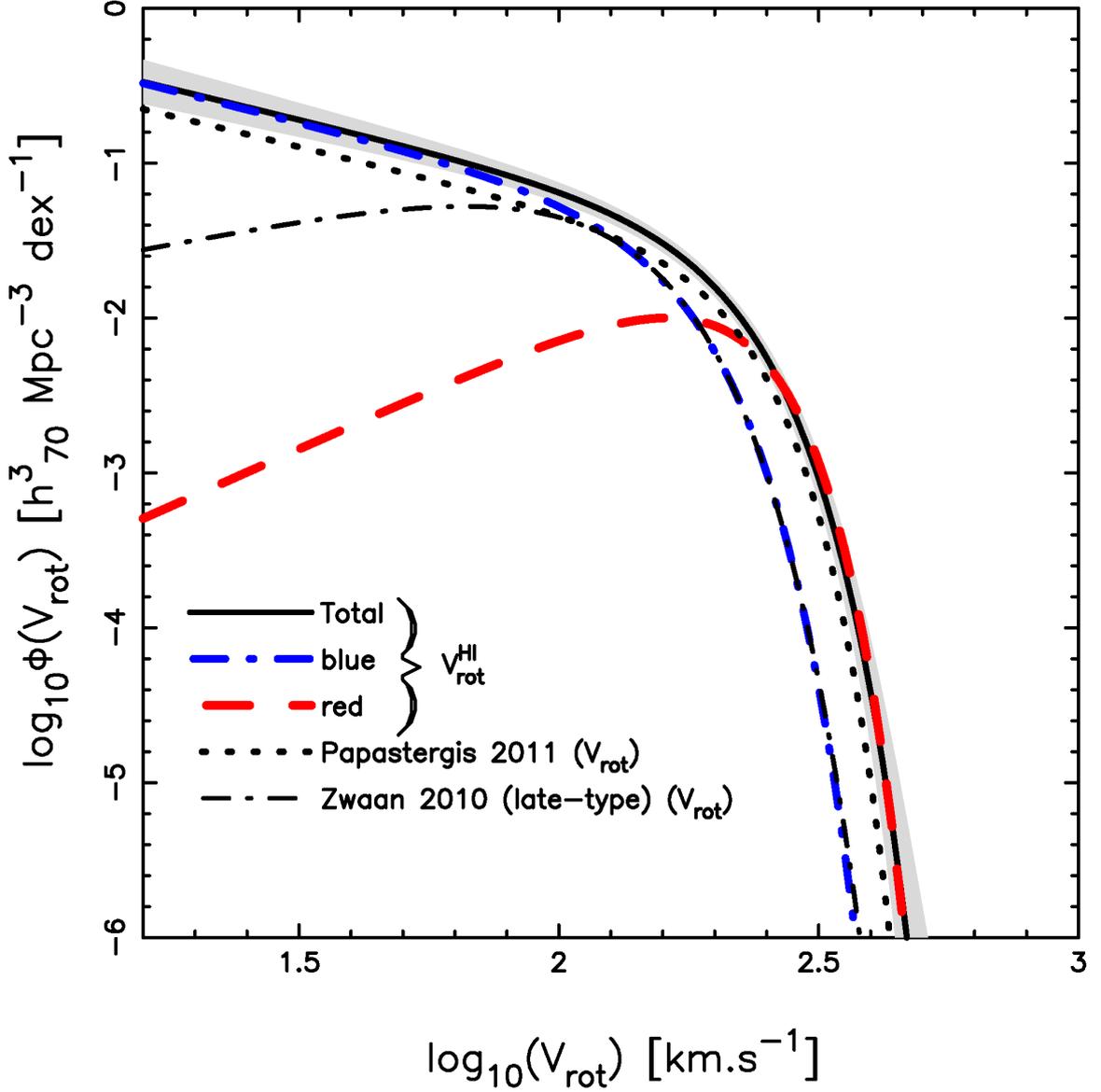


Figure 4.5: A comparison of our estimates of HIVF with previously published results are shown in this figure. Our modified Schechter function fits and their 1σ uncertainty (shaded region for the total sample only) are plotted for the total (black thick solid line), red (red thin dashed line) and blue (blue thin dot-dashed line) samples. The black thick dotted line represents the estimates of Papastergis et al. [2011] for HIVF. The black thin dot-dashed line is the estimate of Zwaan et al. [2010] for late-type galaxies.

within the 1σ uncertainty. The difference between the ALFALFA [Papastergis et al., 2011, Moorman et al., 2014, this work] results and HIPASS [Zwaan et al., 2010] are much starker. Although the full HIPASS sample has been taken to estimate the HIWF, the results Zwaan et al. [2010] seems to match only with the HIWF of blue galaxies in ALFALFA, at the high velocity end and has a shallower slope at lower velocity widths. The reason is that due to its higher sensitivity ALFALFA is able to detect larger velocity widths compared to HIPASS. As is evident from the figure, the large velocity end is dominated by red galaxies which one can associate with early-type galaxies (see figure 4.8), HIPASS would therefore miss this population altogether. The observed fraction of early type galaxies in HIPASS is 11% compared to 20% for red galaxies in our sample.

In figure 4.5 we show our estimates of the HIVF for the total, red and blue samples. Comparison is made with the total sample of ALFALFA [Papastergis et al., 2011] and the late-type galaxies in HIPASS [Zwaan et al., 2010]. Similar to what we saw in the HIWF, the high velocity end of the HIVF for late type galaxies in HIPASS compares well with that of the blue galaxies in our sample. This suggests that associating $V_{\text{rot}}^{\text{HI}}$ with V_{rot} is correct for the late type galaxies. This also suggests that associating blue galaxies with late type galaxies with the age criterion given in figure 4.8 is reasonable. Finally it confirms our earlier argument that HIPASS galaxies sample the blue cloud more frequently as compared to ALFALFA.

For the total sample our results of the HIVF is systematically offset compared to Papastergis et al. [2011] although our HIWF agree with each other. However Papastergis et al. [2011] have estimated the rotational velocity, V_{rot} , function for HI-selected galaxies, whereas we have estimated the HI rotational velocity, $V_{\text{rot}}^{\text{HI}}$, function. To estimate the rotational velocity function Papastergis et al. [2011] estimated the HIVF for late type galaxies, based on the inversion method outlined above. For early type galaxies they used the velocity dispersion function, $\phi^{\text{early}}(\sigma)$, of early type galaxies obtained from SDSS and the

	$\phi(w_{50})$		$\phi(V_{\text{rot}}^{\text{HI}})$				
	ϕ_*	w_*	α	β	ϕ_*		
					V_*		
total	0.023 ± 0.008	2.50 ± 0.06	-0.70 ± 0.16	2.02 ± 0.23	2.3 ± 0.01	α	β
blue	0.046 ± 0.014	2.29 ± 0.06	-0.37 ± 0.18	1.74 ± 0.16	2.2 ± 0.01	-0.81 ± 0.13	2.7 ± 0.17
red	0.008 ± 0.002	2.56 ± 0.07	1.14 ± 0.29	2.51 ± 0.45	2.31 ± 0.01	-0.84 ± 0.16	2.7 ± 0.18
						1.5 ± 0.18	2.96 ± 0.14

Table 4.1: Best fit values of modified Schechter parameters for $\phi(w_{50})$ and $\phi(V_{\text{rot}}^{\text{HI}})$ for total, blue and red samples.

2dFGRS Chae [2010]. $\phi^{\text{early}}(\sigma)$ was then converted to $\phi^{\text{early}}(V_{\text{rot}})$ assuming $V_{\text{rot}} = \sqrt{2}\sigma$ for an isothermal profile. Finally $\phi(V_{\text{rot}})$ for ALFALFA galaxies was constructed by smoothly interpolating the HIVF for late-type galaxies at low velocities to $\phi^{\text{early}}(V_{\text{rot}})$ at large velocities assuming a modified Schechter function. These results suggest that although $V_{\text{rot}}^{\text{HI}}$ is not the same as V_{rot} for early type galaxies, a systematic correction to $V_{\text{rot}}^{\text{HI}}$ can be added to bring it closer to the 'true' rotational velocity.

We end this section with a final observation. We find that the HIVF and HIWF for red and blue galaxies are well separated beyond the knee of the velocity functions. The red population dominates the velocity function at larger velocities. The red population also dominates the HIMF at larger masses but the difference between the red and blue mass functions is much smaller compared to the velocity functions. The blue population dominates the abundances at lower masses and velocities, and the red population has a slope which is much shallower compared to the blue population at this end.

4.5 Scaling Relations for HI Selected Galaxies

In the last section we have observationally derived three distributions which describe the abundances of HI-selected galaxies. These are the HIMF, HIWF, and HIVF. These abundances have been estimated based on their HI signal and the ALFALFA selection function. The next step is to relate HI properties to either galaxies or halos. In chapter 2 and 3 we have presented the conditional HI mass functions, conditioned on optical color and/or magnitude for HI-selected galaxies. In what follows we will build a model to populate halos with HI.

The Halo Abundance Matching (HAM) technique is a powerful and elegant method to match halo properties to galaxy properties [Conroy and Wechsler, 2009, Behroozi et al., 2010]. It assumes that every dark matter halo above a certain mass threshold hosts one galaxy. In its simplest form it assumes that the most massive (or luminous) galaxy is hosted

in the most massive halo, or a monotonic relation between galaxy mass and halo mass. To derive the stellar-halo mass relation of the galaxy one matches the spatial abundances of galaxies to that of halos.

$$n(> M_{\text{star}}) = n(> M_{\text{h}}) \quad (4.4)$$

Here n is the number density of objects above a certain mass threshold, M_{star} (galaxy stellar mass) or M_{h} (halo mass). To further obtain the $M_{\text{star}}-M_{\text{h}}$ relation one needs the stellar mass function (SMF) of galaxies (which are observationally constrained) and the halo mass function (which can be obtained from N -body simulations). The relation can include scatter between M_{star} and M_{h} , and has been shown to reproduce the clustering of galaxies [see [Behroozi et al., 2010](#), and references therein]. HAM techniques provide a starting point for parametrizing the $M_{\text{star}}-M_{\text{h}}$ scaling relation. This relation can be refined further and include redshift evolution and uncertainties so that it reproduces observed galaxy abundances and clustering for different galaxy types [[Behroozi et al., 2013a, 2019](#)] across cosmic time. [Behroozi et al. \[2019\]](#) find that the $M_{\text{star}}-M_{\text{h}}$ is a tight monotonic relation consistent with various observations and cosmological hydrodynamical simulations of galaxy formation.

HAM techniques have also been used to obtain the $M_{\text{HI}} - M_{\text{h}}$ relation [[Khandai et al., 2011, Padmanabhan and Kulkarni, 2017](#)]. However the basic assumption in HAM – a monotonic relation between halo mass and the galaxy property – may not hold if we consider the galaxy property to be the HI mass. Unlike the stellar mass the HI content of galaxies is very sensitive to its local environment and may decrease with increasing halo mass. In massive halos, e.g. clusters and groups, the virial temperature is large and feedback processes, e.g. from supermassive blackholes keep the gas ionized. It is therefore less likely, on average, to have a considerable amount of HI associated with such massive systems. These arguments are observationally supported. We find that in the sample and survey volume considered here, 11% (38%) of detections in HI are in the red (blue) cloud (see figure 2.5). The lu-

minous red galaxies dominate the high mass end of the galaxy SMF and at $M_{\text{star}} = 11.3$ the number of red galaxies is $\sim 10\times$ larger than the blue galaxies [Baldry et al., 2012]. These galaxies are mostly centrals [Drory et al., 2009] and would be hosted in halos of mass $M_h \sim 14$ [Behroozi et al., 2019] with virial temperatures $T_{\text{vir}} \sim 10^7\text{K}$. It would be very rare to see large amounts of HI in such systems. Direct and more sensitive observations targeting massive galaxies, the GASS survey [Catinella et al., 2013], and massive ETGs, the ATLAS^{3D} survey [Serra et al., 2012], confirm these arguments. They find that massive galaxies are dominated by non-detections in HI and the limiting HI masses (upper bound based on survey sensitivity) is well below the knee of the HIMF ($M_* = 9.95$). We finally note that the high mass end of the HIMF is dominated by luminous red galaxies which represent a small fraction of all luminous red galaxies. As mentioned earlier luminous red galaxies dominate the SMF (hence the halo mass function) over their blue counterparts by at least a factor of 10. It is not surprising, therefore, to see that the HIMF is dominated by red galaxies since a small fraction of gas rich red galaxies is all that is needed to boost their abundances over that of their blue counterparts.

We have argued that HAM cannot be applied to obtain the $M_{\text{HI}}-M_h$ relation since we do not expect a monotonic correlation between HI masses and stellar or halo masses. In what follows, given the tight monotonic relation between stellar and halo masses, we will use the halo and stellar masses as proxies for each other. Although a stellar mass selected sample should not have a monotonic relation with HI mass, we can turn this around and ask if a HI-selected sample like ours has a monotonic relation with stellar mass. This is indeed true and has been seen both in ALFALFA [Huang et al., 2012, D20] and the HI Parkes All-Sky Survey Catalog (HICAT) [Maddox et al., 2015]. One can therefore expect that gas rich galaxies will, on average, be found in more massive halos as compared to gas poor halos. A subsample of all halos should host HI in such a manner that the relation between M_{HI} and M_h is monotonic.

Recently [Guo et al. \[2020\]](#) used ALFALFA with an optical group catalog from SDSS to estimate the mean HI mass - halo mass, $\langle M_{\text{HI}} \rangle - M_{\text{h}}$, relation by stacking HI on the optically selected group catalog. The HI was stacked for the full group and separately for centrals in the catalog, the difference between the two therefore represents the contribution of total HI mass in satellites in the group. Due to confusion, stacking of centrals is contaminated by nearby satellites within the group. Therefore the $\langle M_{\text{HI}}^{\text{cen}} \rangle$ for centrals are upper limits, the $\langle M_{\text{HI}}^{\text{sat,tot}} \rangle$ for satellites are lower limits and the $\langle M_{\text{HI}}^{\text{group}} \rangle$ is the average stacked mass in groups. The stacking procedure was carried out as a function of group or halo mass. In figure 4.6 the result of [Guo et al. \[2020\]](#) $\langle M_{\text{HI}} \rangle - M_{\text{h}}$ is shown for centrals (red open circles).

At this stage it is important to clarify how we associate HI in galaxies and halos. Galaxies can be in all kinds of environments. They can be isolated field galaxies, central galaxies in groups and clusters or satellite galaxies in such systems. These are distinct objects in observations. The halo on the other hand can be defined as a concentration of mass within a radius such that it is virialized inside this radius (central halo). Simulations show that halos have substructure and subsubstructure (and so on) and these objects (satellite halo) are not only self-bound but bound to the halo. In what follows we will refer to a halo as either a central halo or satellite halo and will not distinguish between them. Most of the mass inside the radius of a halo is associated with the central halo (unless we have mergers of nearly equal masses), due to which the HMF is close to the HMF for centrals. The HMF for satellites contributes little to the HMF.

A similar result is seen observationally for the SMF. The SMF of the centrals dominates the total SMF, whereas satellites contribute little to the total SMF [[Drory et al., 2009](#)]. In our definition field halos are centrals without satellites. We will assume that every halo (central or satellite) hosts a single galaxy. Similarly HI is associated with a halo via its galaxy. Although it is rare to find considerable amounts of HI in massive central galaxies of clusters, a significant fraction of HI is locked up in satellite galaxies as seen in observations

[Lah et al., 2009]. The results of Guo et al. [2020] corroborate this observation. As seen in figure 3 of Guo et al. [2020] the total HI content in satellites can be at least 60% in a group of mass $M_h \sim 13$. However groups and clusters are less abundant as compared to lower mass objects. We therefore expect (as in the case of the galaxy SMF) that satellites contribute little to the HIMF. This expectation is consistent with semi-analytical models of HI [Kim et al., 2017].

We use a very large publicly available dark matter simulation – *MultiDark Planck 2* (MDPL2) [Klypin et al., 2016]. MDPL2 is run with the publicly available code GADGET-2 [Springel, 2005]. It evolves the matter density field, sampled by 3840^3 dark matter particles in a comoving box of side $L_{\text{box}} = 1\text{Gpc}/h$, to $z = 0$. This corresponds to a dark matter mass resolution $m_{\text{DM}} = 1.51 \times 10^9 M_\odot/h$. The force resolution is $\epsilon = 5\text{kpc}/h$. The MDPL2 was run with a flat Λ CDM cosmology with $(\Omega_\Lambda, \Omega_m, h, \sigma_8, n_s) = (0.693, 0.307, 0.678, 0.823, 0.96)$ consistent with the Planck15 results [Planck Collaboration et al., 2016]. The data products include halo (central+satellite) catalogs and merger trees which were obtained with the help of the ROCKSTAR [Behroozi et al., 2013b] and CONSISTENT TREES [Behroozi et al., 2013c] codes. The MultiDark-Galaxies [Knebe et al., 2018] is a catalog of galaxies built on the data products of MDPL2 with the semi-analytical model – Semi-Analytic Galaxy Evolution (SAGE). The SAGE model is tuned to reproduce the SMF in the local Universe. Galaxy colors and r -band luminosities are not well reproduced in SAGE. SAGE also provides the mean age of the stellar population of the galaxy, t_{age} which we use as a proxy for color. The stellar mass of the galaxy, M_{star} and t_{age} is available for our HI sample, so that we can choose a galaxy population based on t_{age} .

We use the MDPL2 halo catalog to obtain the HMF. This is shown as the black thick solid line in figure 4.7. The downward arrow at $M_h \sim 11$ is the mass resolution of the MDPL2 catalog. We have extrapolated the power-law to obtain the HMF at lower masses. We will next define the HI-selected halo mass function, $\phi^{\text{HI}}(M_h)$, which reproduces the

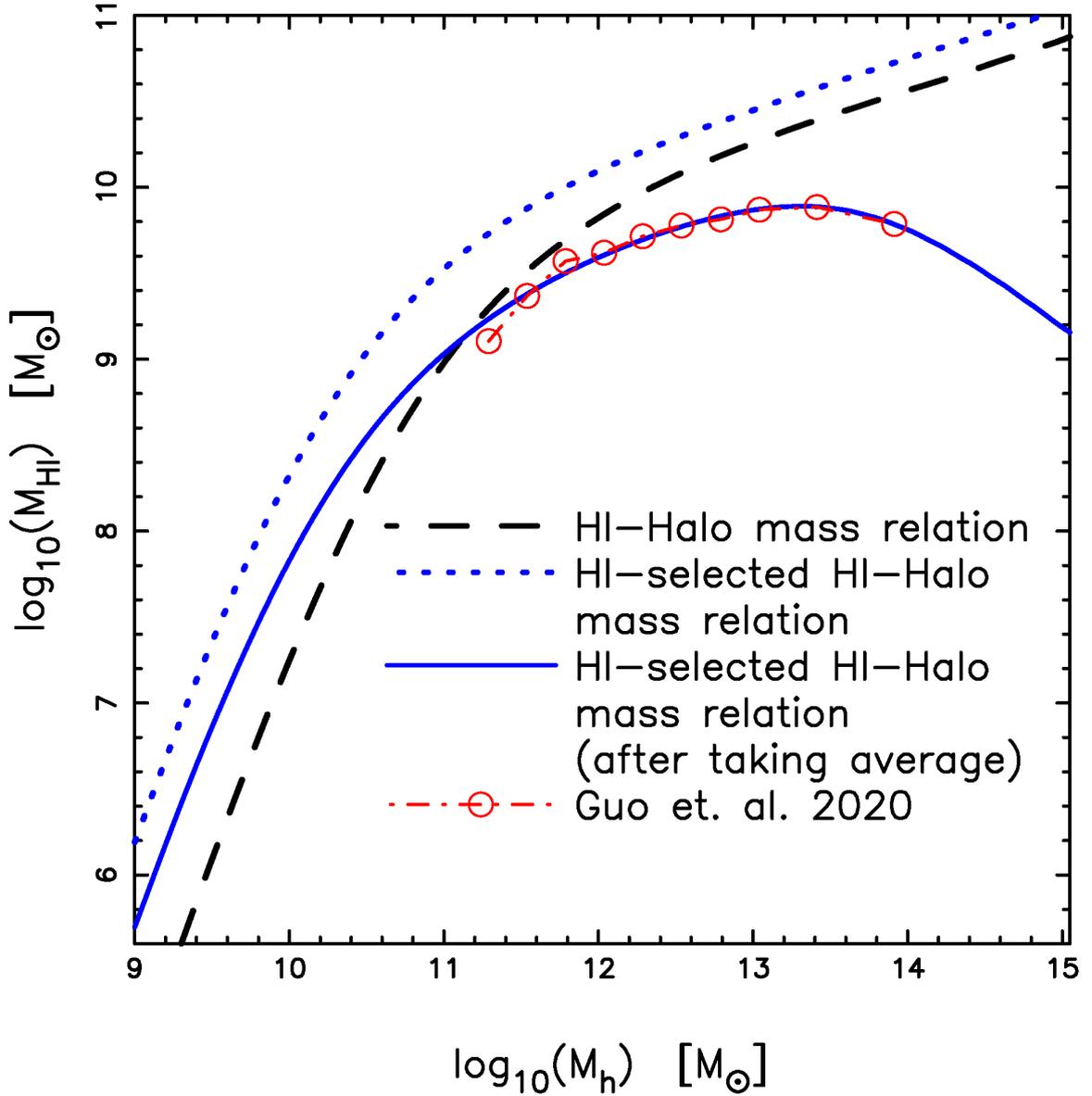


Figure 4.6: Average HI mass - halo mass relation from Guo et al. [2020] is shown as red open circles. The black dashed line is the $M_{\text{HI}} - M_{\text{h}}$ scaling relation obtained by abundance matching our HIMF to the HMF. The blue dotted line is the $M_{\text{HI}} - M_{\text{h}}$ relation obtained by abundance matching our HIMF to the HI-selected HMF (see equation 4.5 and discussion). The blue solid line is the mean $\langle M_{\text{HI}} \rangle - M_{\text{h}}$ relation that is obtained by averaging the blue dotted line over all the halos at any given mass $[M_{\text{h}}, M_{\text{h}} + dM_{\text{h}}]$. The HI-selected HMF has been defined so that the solid line matches the observed points of Guo et al. [2020].

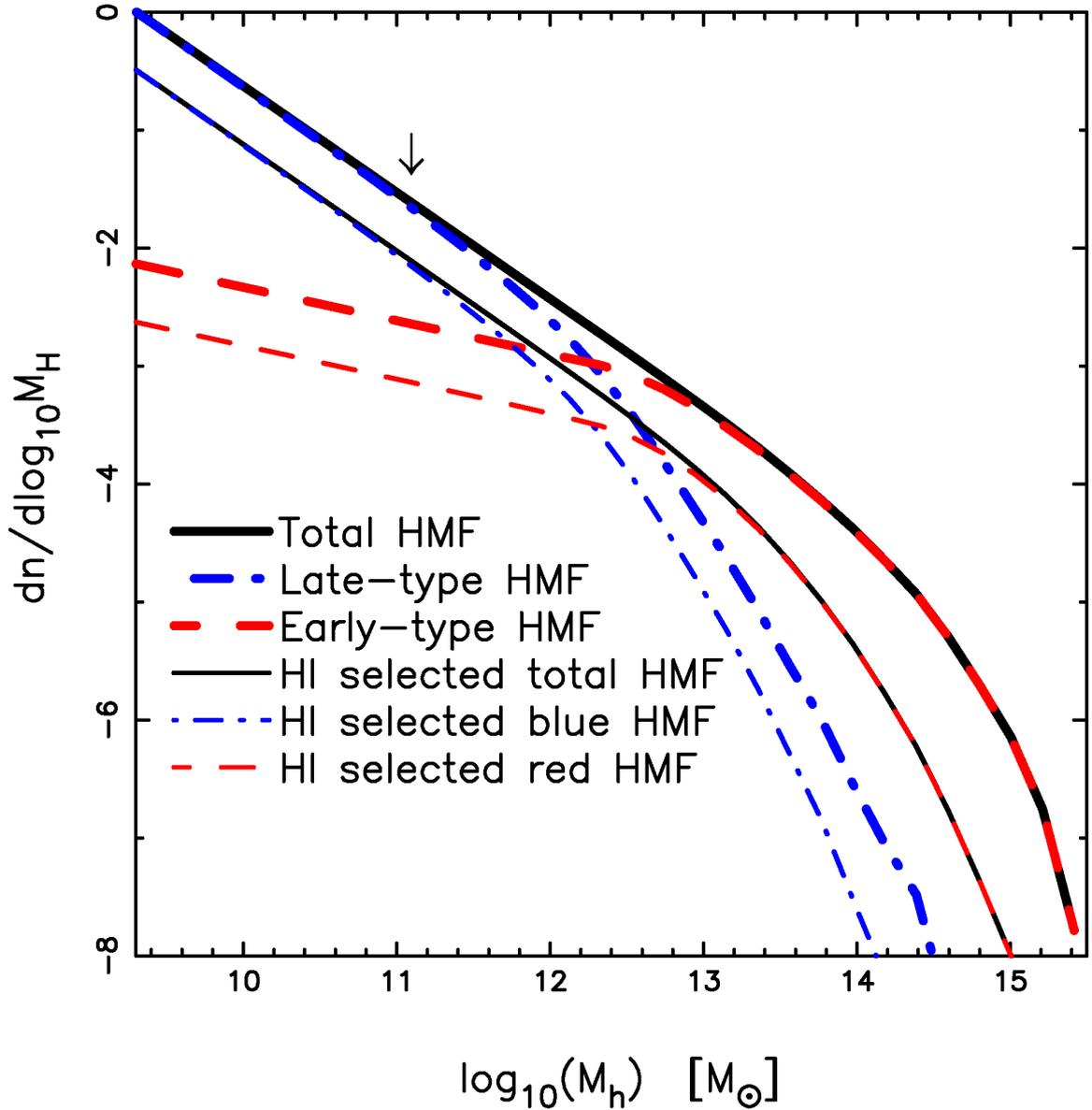


Figure 4.7: The HMF (HI-selected HMF) is shown as the solid thick (thin) line. The arrow at $M_h \simeq 11$ is the resolution of the simulation catalog. For masses below that we have linearly extrapolated the HMF. The red thick dashed (blue thick dot-dashed) line represents the HMF hosting early (late) type galaxies. (See figure 4.8 and corresponding text in section 4.5 for the definition of early (late) type galaxies. The corresponding HI-selected HMF are shown in thin lines.

observations of Guo et al. [2020] (red open circles in figure 4.6) by defining it with the parametric form:

$$\phi^{\text{HI}}(M_h) = \phi(M_h) \times \frac{f}{1 + \left(\frac{M_h}{M_h^*}\right)^\gamma} \quad (4.5)$$

where $\phi(M_h)$ is the halo mass function. f is an overall fraction that reduces the abundances $\phi^{\text{HI}}(M_h)$ with respect to $\phi(M_h)$ and the abundance is further reduced by $\left(\frac{M_h}{M_h^*}\right)^\gamma$ for $M_h \gtrsim M_h^*$. This is a three-parameter functional form that is justified by the various HI observations described earlier. However we point out that it is by no means unique. One can further suppress the HI mass at lower halo masses [Bagla et al., 2010, Padmanabhan et al., 2017] since these low mass halos would host negligible amounts of HI due to the ionizing background. However the suppression is expected to happen at circular velocities smaller than 30 km/s, which correspond to the smallest HI detections in ALFALFA ($M_{\text{HI}} \sim 7.0 - 7.5$) and is naturally taken care of by the ALFALFA selection function.

If we abundance match the HIMF to the HI-selected HMF we obtain a scaling relation (blue dotted line) in figure 4.6. The black dashed line is obtained by abundance matching the HIMF to the HMF. Since $\phi^{\text{HI}}(M_h)$ represents a subsample of all halos (described by $\phi(M_h)$), HI is now distributed in a smaller number of halos, thereby increasing the HI mass at fixed halo mass. This results in an HI-selected scaling relation (blue dotted line) above the halo mass selected scaling relation (dashed line). The HI selected scaling relation (dotted line) is well described by a double power law

$$M_{\text{HI}}(M_h) = M_{\text{HI}}^{\text{A}} \frac{\left(\frac{M_h}{M_{\text{ht}}}\right)^\alpha}{\left[1 + \left(\frac{M_h}{M_{\text{ht}}}\right)^\beta\right]} \quad (4.6)$$

Here M_{HI}^{A} is the amplitude, α is the slope of the scaling relation at lower masses which gets suppressed to a slope of $\alpha - \beta$ at masses greater than a transition halo mass, $M_{\text{ht}} \ll M_h$. We find $(M_{\text{HI}}^{\text{A}}, \alpha, \beta, M_{\text{ht}}) = (9.59, 2.10, 1.76, 10.62)$ describes well the HI selected $M_{\text{HI}} - M_h$ scaling relation in figure 4.6.

We finally need to fix $\phi^{\text{HI}}(M_h)$ described by the three parameters $\{f, M_h^*, \gamma\}$ so as to match the mean observed scaling relation $\langle M_{\text{HI}} \rangle - M_h$ for centrals of Guo et al. [2020]. As discussed earlier we will not distinguish between centrals and satellites while abundance matching. HI is assigned based on halo mass and does not depend whether the halo is a central or satellite. Guo et al. [2020] do not have a corresponding estimate of $\langle M_{\text{HI}}^{\text{sat}} \rangle - M_h$ but rather have $\langle M_{\text{HI}}^{\text{sat,tot}} \rangle - M_h$ where $\langle M_{\text{HI}}^{\text{sat,tot}} \rangle$ is the mean total HI mass in satellites hosted in centrals of mass M_h . The parameters $\{f, M_h^*, \gamma\}$ are fixed by taking the HI-selected scaling relation $M_{\text{HI}}(M_h)$ (blue dotted line, figure 4.6) at fixed halo mass M_h and averaging it over all halos in the mass range $[M_h, M_h + dM_h]$. The choice of $\{f = 0.320, M_h^* = 13.661, \gamma = 0.996\}$ (obtained by this minimization procedure) results in the blue solid line which reproduces the observed points of Guo et al. [2020]. The corresponding HI-selected HMF is shown as the black thin solid line in figure 4.7.

Since we are working with the red and blue populations amongst the HI-selected galaxies we would also like to obtain corresponding scaling relations for these populations as well. This is where the SAGE catalog becomes useful. However, the SAGE catalog does not have accurate estimates of colors and magnitudes that are needed to determine the red and blue populations. We therefore need to come up with an approximate proxy for the red and blue populations. We extract the mean age of the stellar population in galaxies, t_{age} , from SDSS. In figure 4.8 we show the age distribution of the red and blue samples in ALFALFA. The age distribution of the blue population has a pronounced peak at $t_{\text{age}} \sim 6.8\text{Gyr}$ and drops rapidly beyond $t_{\text{age}} > 7.2\text{Gyr}$. The age distribution of the red population has a peak at $t_{\text{age}} \sim 8.5\text{Gyr}$ and the distribution is broad. Although, bimodal, the distribution is suppressed for red galaxies since ALFALFA primarily samples the blue cloud. In spite of this we can see that the red population in ALFALFA is an older, early-type population compared to the blue population. The distributions intersect at $t_{\text{age}} = 8.38\text{Gyr}$. We can therefore use t_{age} , as a rough proxy for colors, with $t_{\text{age}} > 8.38\text{Gyr}$ ($t_{\text{age}} < 8.38\text{Gyr}$) representing the

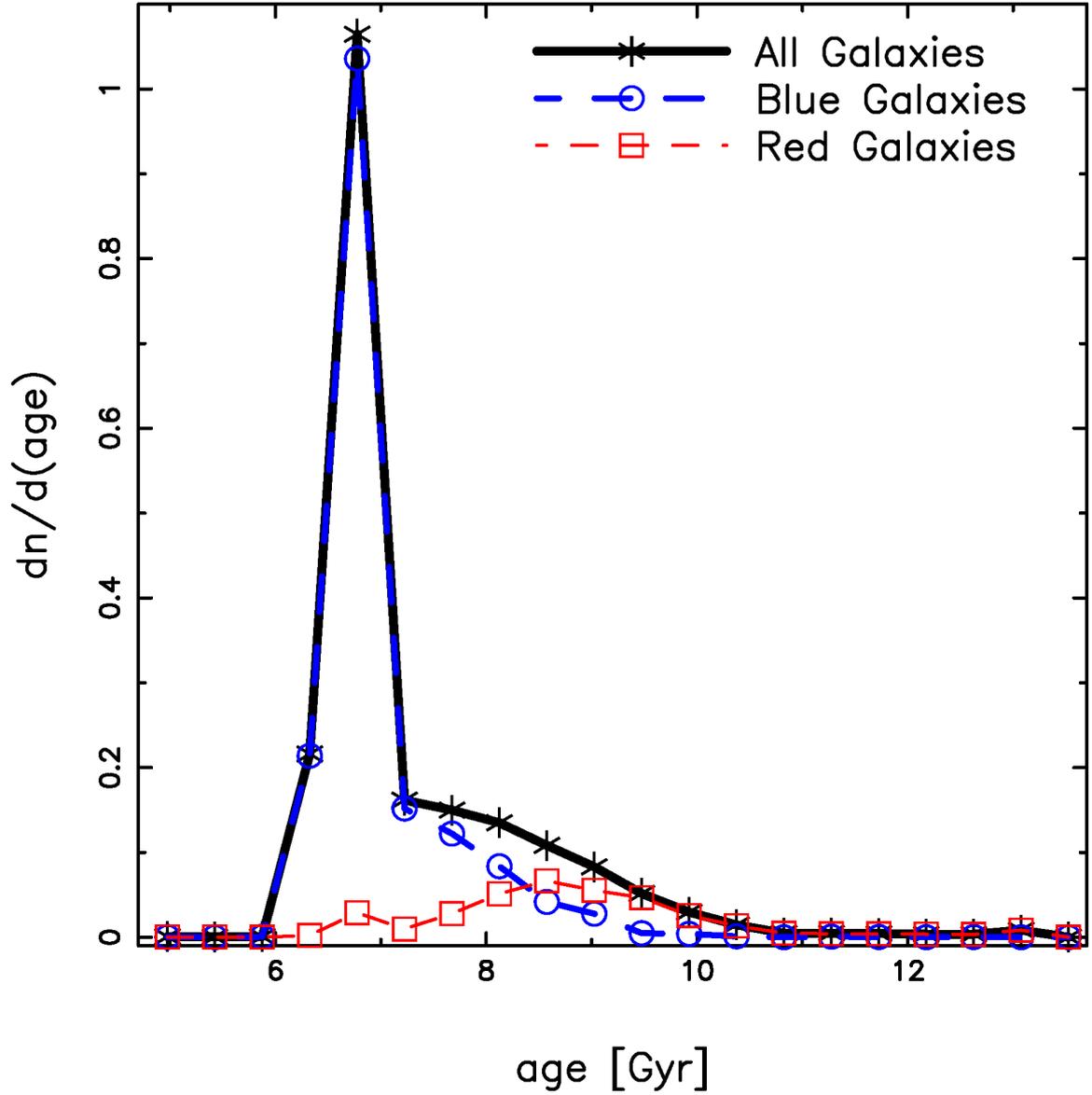


Figure 4.8: This figure shows the observed age (mean stellar age in Gyr) distribution of the total (black crosses, thick solid line), blue (blue open circles, thick dashed line), and red (red open squares, thin dashed line) sample of galaxies in ALFALFA. The intersection of the age distribution of blue and red galaxies at $t_{\text{age}} = 8.38\text{Gyr}$ is used to classify the blue and red populations as late-type and early-type galaxies respectively.

red (blue) populations. This definition has been made on the basis of the ALFALFA (HI-selected) sample. A similar definition could be made on the basis of a stellar mass selected sample, from SDSS. However we wish to use the HI-selected HMF to abundance match to the HI distributions therefore we will stick with this definition.

We use this criterion to identify galaxies in SAGE as early-type (red) or late-type (blue) galaxies. The SAGE SMF (black solid line) is plotted in figure 4.9. It compares well with the observed SMF (grey open diamonds, thin line) from SDSS-GALEX [Moustakas et al., 2013]. The red dashed (blue dot-dashed) line is the contribution from red, early-type (blue, late-type) galaxies to the SMF from SAGE. The red population dominates the high mass end of the SMF whereas the blue population dominates the SMF at lower masses. The bimodality is however not as distinct since the classification was done based on an HI-selected ALFALFA sample. If it were done on a stellar mass selected sample a clear bimodality is seen [Baldry et al., 2012].

In figure 4.7 we plot the HMF corresponding to red (early-type) and blue (late-type) galaxies as red thick dashed and blue thick dot-dashed lines respectively. The corresponding HI-selected HMF are plotted with thin lines. One can see that the early-type galaxies dominate the HMF, and the HI-selected HMF, at $M_h \gtrsim 12.2$ and the late-type galaxies dominate below this mass.

We are now in a position to obtain scaling relations between various HI properties by abundance matching the HIMF, HIWF, and HIVF to each other. Having defined and constrained the HI-selected HMF, $\phi^{\text{HI}}(M_h)$, we can also abundance match the HIMF, HIWF, and HIVF to the HI-selected HMF.

In figure 4.10 we show the scaling relations $M_{\text{HI}} - w_{50}$, $M_{\text{HI}} - V_{\text{rot}}^{\text{HI}}$, and $w_{50} - V_{\text{rot}}^{\text{HI}}$ in the left, middle and right panels respectively. These relations were obtained by abundance matching the HIMF-HIWF, HIMF-HIVF and HIWF-HIVF. This is done for the total (black solid line), red (red dashed line) and blue (blue dot-dashed line) samples. The scaling re-

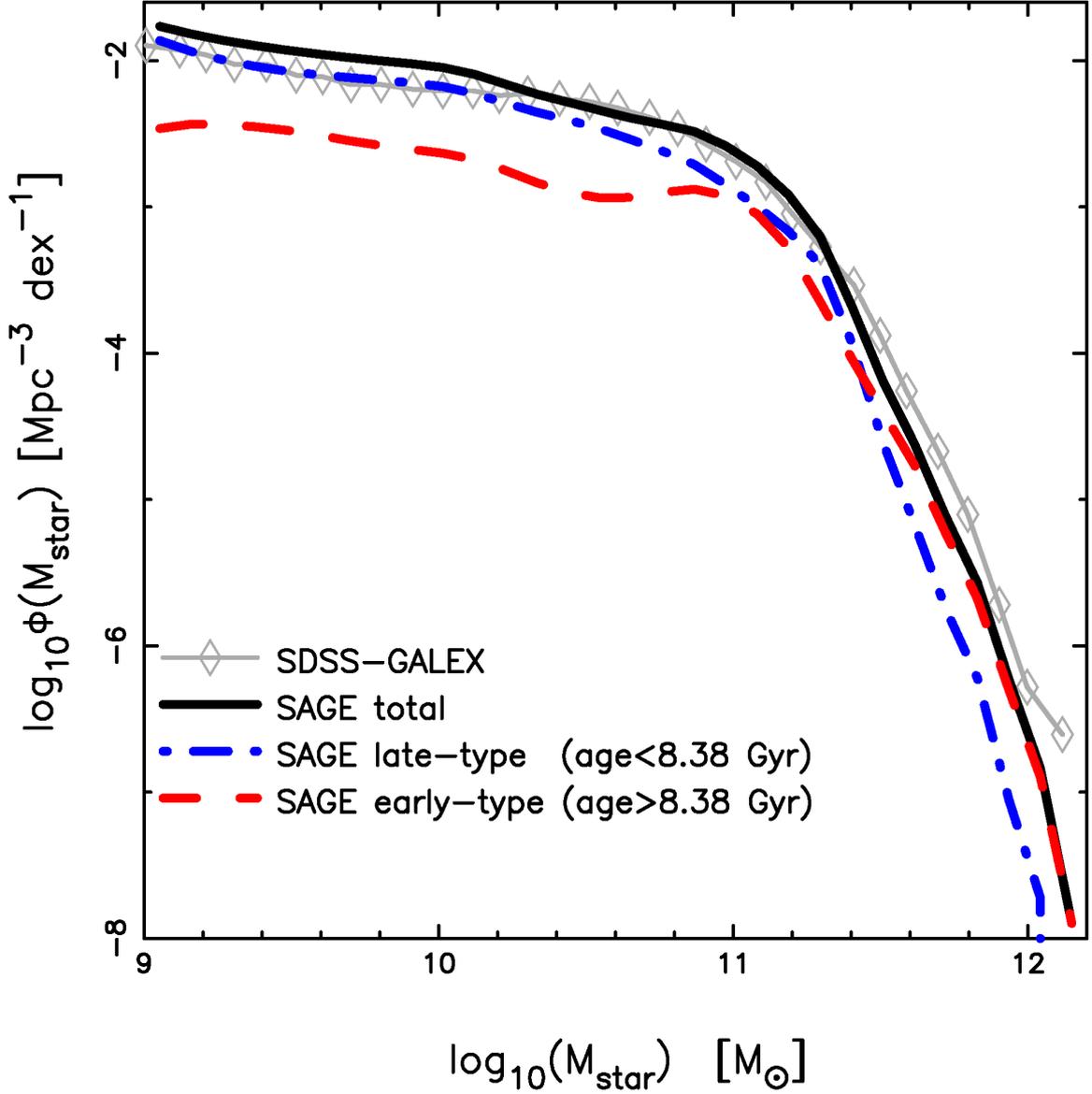


Figure 4.9: This figure shows the SMF. The grey open-diamonds with thin solid line represents observational estimates of the SMF from SDSS-GALEX [Moustakas et al., 2013]. The black thick solid line is the SMF from the SAGE catalog [Knebe et al., 2018] which is based on the MDPL2 simulation [Klypin et al., 2016]. The red dashed (blue dot-dashed) line is the SMF for early-type (late-type) galaxies defined as $t_{\text{age}} > 8.38$ ($t_{\text{age}} < 8.38$) Gyr in the context of HI-selected galaxies (see figure 4.8).

lation for the total sample can be thought as a galaxy-count weighted sum of the scaling relations of the red and blue samples. Since HI is primarily sampled by the blue cloud, the scaling relation for the full sample is closer to the scaling relation of blue galaxies. The scaling relations $M_{\text{HI}} - w_{50}$ and $M_{\text{HI}} - V_{\text{rot}}^{\text{HI}}$ are different for the red and blue samples. At lower masses the HI detections are primarily in the blue cloud, we therefore see a rapid drop in the scaling relations below $M_{\text{HI}} = 7.6$ for the red sample. For $M_{\text{HI}} \in [8.5, 10]$ we find that at fixed HI mass the red sample has larger velocity profile widths. This suggests that in this range the red sample is on average hosted in larger halos because the profile width (or rotational velocity) is a good proxy for the halo mass. Although we have not invoked the HI-selected HMF at this stage, we see that this explanation is consistent with figure 4.7 and figure 4.11. In the right panel of figure 4.10 we see the $w_{50} - V_{\text{rot}}^{\text{HI}}$ scaling relation for the red sample is above that of the blue sample at lower velocities, but they asymptote to each other at larger velocities.

In figure 4.11 we show the scaling relations $M_{\text{h}} - M_{\text{HI}}$ (left panel), $M_{\text{h}} - w_{50}$ (middle panel) and $M_{\text{h}} - V_{\text{rot}}^{\text{HI}}$ (right panel) which were obtained by abundance matching the HI-selected HMF-HIMF, HI-selected HMF-HIWF, and HI-selected HMF-HIVF respectively. This was done separately for the total (black solid line), red (red dashed line), and blue (blue dot-dashed line) samples. The $M_{\text{h}} - M_{\text{HI}}$ scaling relation is qualitatively similar in shape to the $M_{\text{h}} - M_{\text{star}}$ scaling relation [Behroozi et al., 2010, 2019] and is described by a double power-law. A steep power law with slope ~ 2.1 at lower masses transitioning to a shallower power law with slope ~ 0.34 above masses $M_{\text{h}} \gtrsim 10.62$. The transition mass $M_{\text{ht}} = 10.62$ for $M_{\text{h}} - M_{\text{HI}}$, is lower than the transition mass of $M_{\text{h}} = 12$ [Behroozi et al., 2019] for $M_{\text{h}} - M_{\text{star}}$. This suggests that baryonic processes like heating and feedback in larger mass halos suppress HI gas on a shorter time scale compared to star-formation. A double power-law is also seen in the scaling relations for $M_{\text{h}} - w_{50}$ and $M_{\text{h}} - V_{\text{rot}}^{\text{HI}}$ with the transition in slopes occurring at $M_{\text{h}} \sim 11.5 - 12.0$. As in figure 4.10, the scaling relation

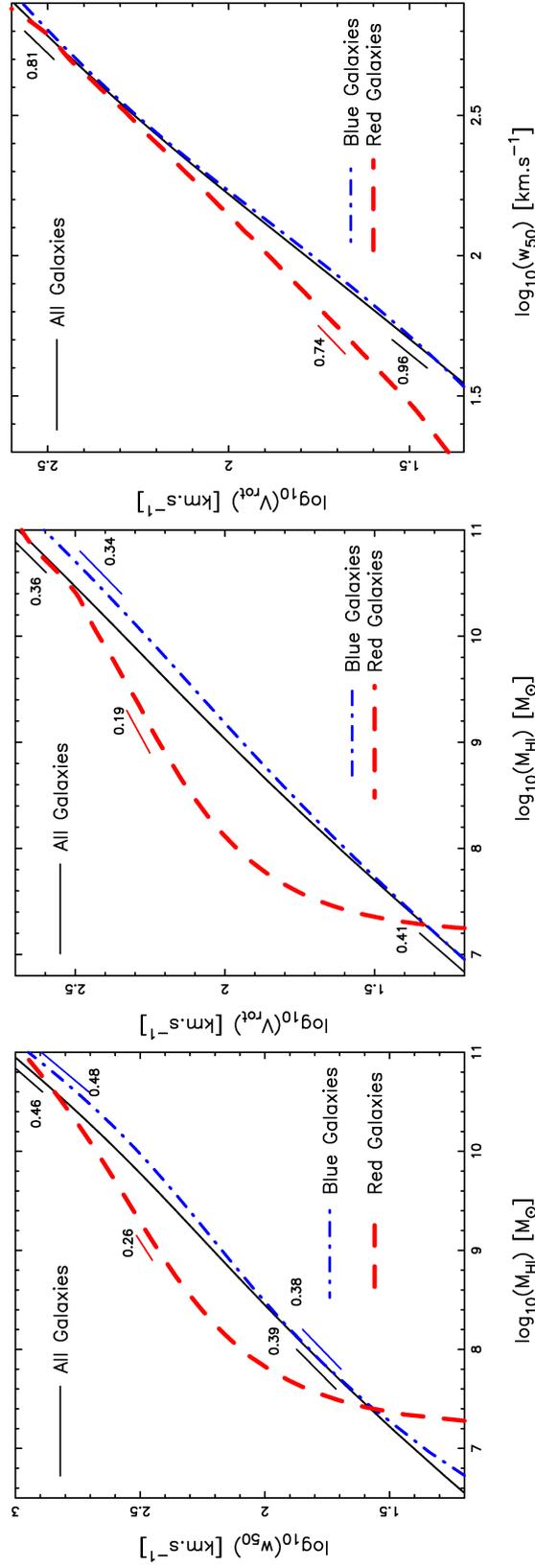


Figure 4.10: The figure shows the scaling relations obtained by abundance matching the three HI distributions, the HIMF, HIWF, and HIWF, amongst themselves. The left, middle and right panels show the scaling relation $M_{\text{HI}} - w_{50}$, $M_{\text{HI}} - V_{\text{rot}}^{\text{HI}}$, and $w_{50} - V_{\text{rot}}^{\text{HI}}$ respectively for the total (black solid), red (red dashed), and blue (blue dot-dashed) samples.

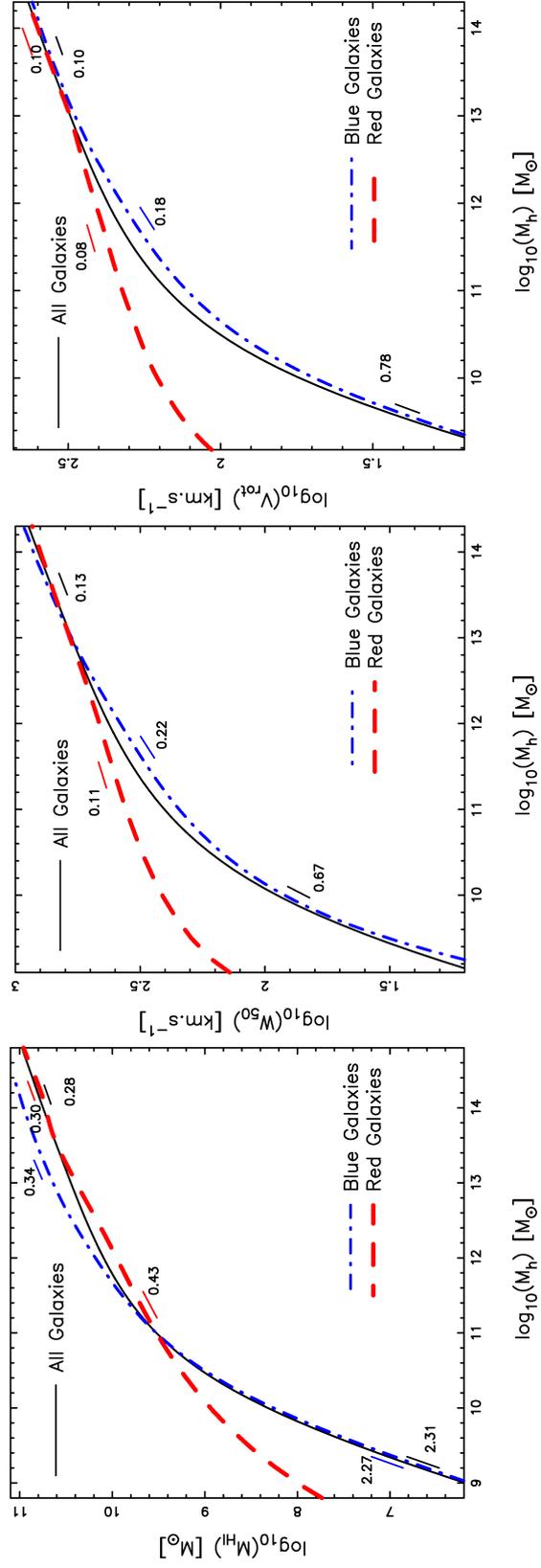


Figure 4.11: This figure shows the scaling relations by abundance matching the three HI distributions, namely the HIMF, HIWF, and HIWF with the HI-selected HIMF (equation 4.5). The left, middle, and right panels show the scaling relation $M_h - M_{\text{HI}}$, $M_h - w_{50}$ and $M_h - V_{\text{rot}}^{\text{HI}}$ respectively for the total (black solid), red (red dashed) and blue (blue dot-dashed) samples.

of the total sample is close to that of the blue sample.

At the low mass end ($M_h \leq 11$) of the $M_h - M_{\text{HI}}$ scaling relation we find that at fixed halo mass the red sample is richer in HI compared to their blue counterparts. The situation is reversed at the high mass end. At the high mass end, we can turn this around. We find at fixed M_{HI} the halo mass is larger for the red sample compared to the blue sample. In the middle and right panels we see a similar trend for $M_h - w_{50}$ and $M_h - V_{\text{rot}}^{\text{HI}}$ at lower masses. However at larger masses the relations of the blue and red samples asymptote to each other, suggesting that the velocity profile is a good descriptor of the halo mass irrespective of galaxy type.

We end this section by comparing the $M_{\text{HI}} - M_h$ relation obtained with ALFALFA data. We create five volume limited samples in equal bins of mass from ALFALFA, in the mass range $M_{\text{HI}} \in [8.0, 10.5[$. The five volume limited samples are disjoint sets. We combine them to create a final volume limited sample. We use the stellar masses of these galaxies and convert them into halo masses using the tight $M_{\text{star}} - M_h$ scaling relation of [Behroozi et al., 2010]. This is shown in figure 4.12. The solid line is the scaling relation that we obtain by abundance matching the HIMF to the HI-selected HMF (equation 4.5) and the dashed line is the scaling relation obtained by abundance matching the HIMF to the HMF. Clearly the scaling that we obtain (black solid line) is in better agreement with the data as compared to the scaling relation obtained by abundance matching the HIMF and HMF (black dashed line).

4.6 Discussion and Summary

In this work we use data from the ALFALFA survey to obtain the HIMF, HIWF, and HIVF for HI-selected galaxies. The survey volume that we consider overlaps with SDSS and allows us to also look at these abundances for the red and blue population of galaxies. We then

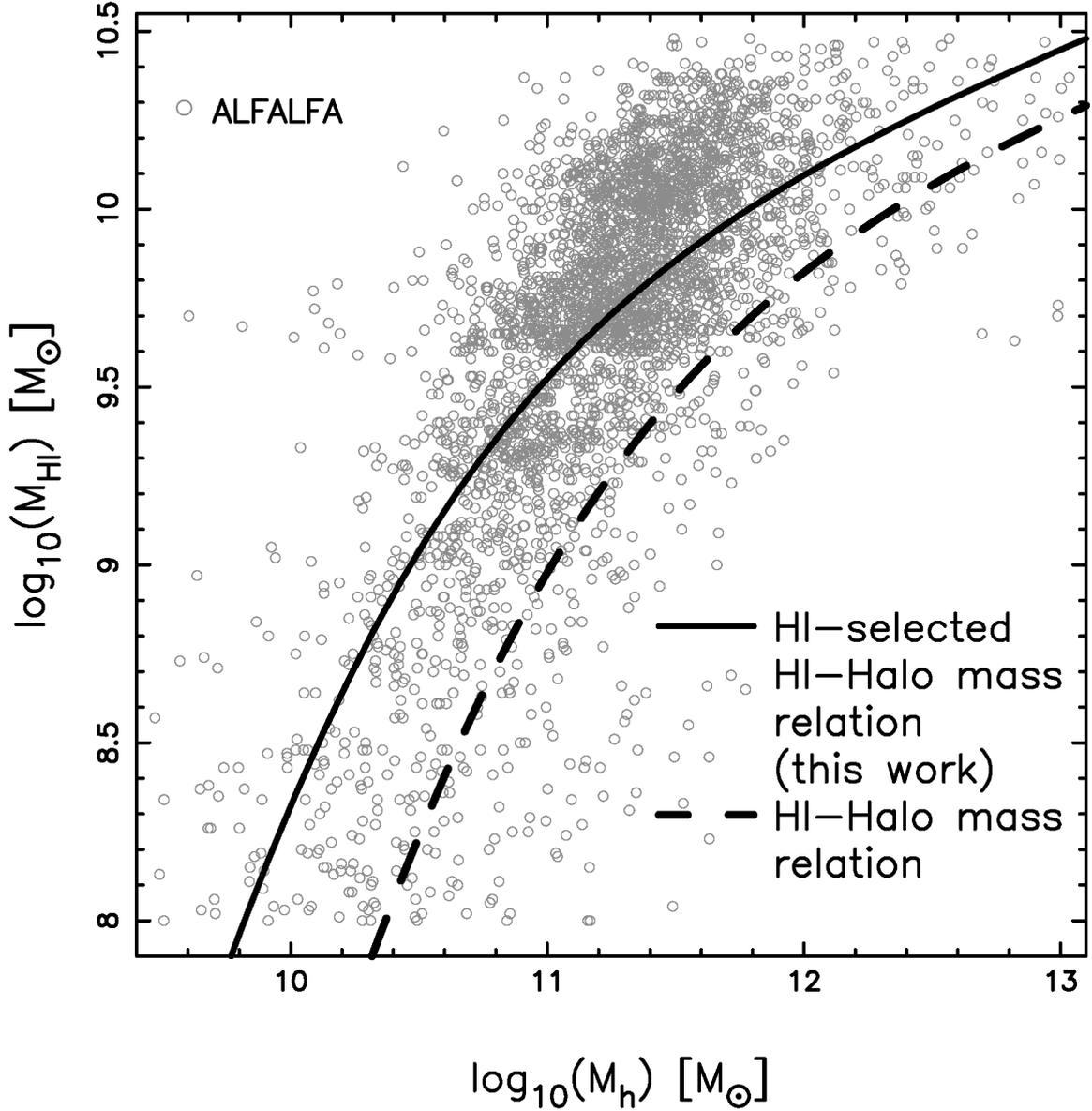


Figure 4.12: The $M_{\text{HI}} - M_{\text{h}}$ scatter plot (grey open circles) for a volume-limited subsample generated from the ALFALFA survey. The halo masses have been estimated using the $M_{\text{star}} - M_{\text{h}}$ relation of Behroozi et al. [2010]. The black solid (dashed) line is the $M_{\text{HI}} - M_{\text{h}}$ relation obtained by abundance matching the HIMF with the HI-selected HMF (HMF).

use recent observations from ALFALFA which estimate $\langle M_{\text{HI}} \rangle - M_{\text{h}}$ relation in massive centrals [Guo et al., 2020] to finally estimate an HI-selected HMF, $\phi^{\text{HI}}(M_{\text{h}})$ (equation 4.5). $\phi^{\text{HI}}(M_{\text{h}})$ is parametrized by three parameters which are fixed to match the observed $\langle M_{\text{HI}} \rangle - M_{\text{h}}$ relation. Although an upper bound, this relation explains the $M_{\text{HI}} - M_{\text{h}}$ relation for an HI-selected sample (figure 4.12). We then use a semi-analytic galaxy catalog, SAGE, which was generated from a large simulation, MDPL2, to further obtain the HI-selected HMF for red (early-type) and blue(late-type) galaxies.

There are a number of assumptions while obtaining the HI-selected HMF for red and blue galaxies. We have assumed that age is a proxy for color, justified observationally (figure 4.8). Although it gives a qualitatively similar bimodal behaviour to that seen in the observed [Drory et al., 2009, Baldry et al., 2012] SMF (figure 4.9), it is by no way an exact proxy. The second assumption is that the stellar ages from SAGE are accurate. The stellar ages from SAGE are based on the halo merger trees (or growth histories) and the various assumptions of their model. In spite of this it gives a qualitatively and physically reasonable bimodal distribution for age. Finally, we have used the same relation (equation 4.5) to obtain an HI selected halo mass function for both the red and blue galaxies. This may not be true. In order to distinguish between them we would need the stacking results of [Guo et al., 2020] to be made for red and blue galaxies separately which is not available. With these assumptions in mind we stress that the HI-halo scaling relations for the red and blue sample may not be completely accurate, but should be thought of as a result which should be revisited once more data (both from observations and simulations) becomes available in the future. However the various HI scaling relations (figure 4.10) are robust for the total, red and blue samples since there are no model assumptions. Similarly the HI-halo scaling relations for the total sample are also robust.

Recently Li et al. [2019b] presented the HI-selected HMF for late-type galaxies. They used a scaling relation $w_{50} - M_{\text{h}}$ from 175 late-type galaxies in the SPARC [Lelli et al.,

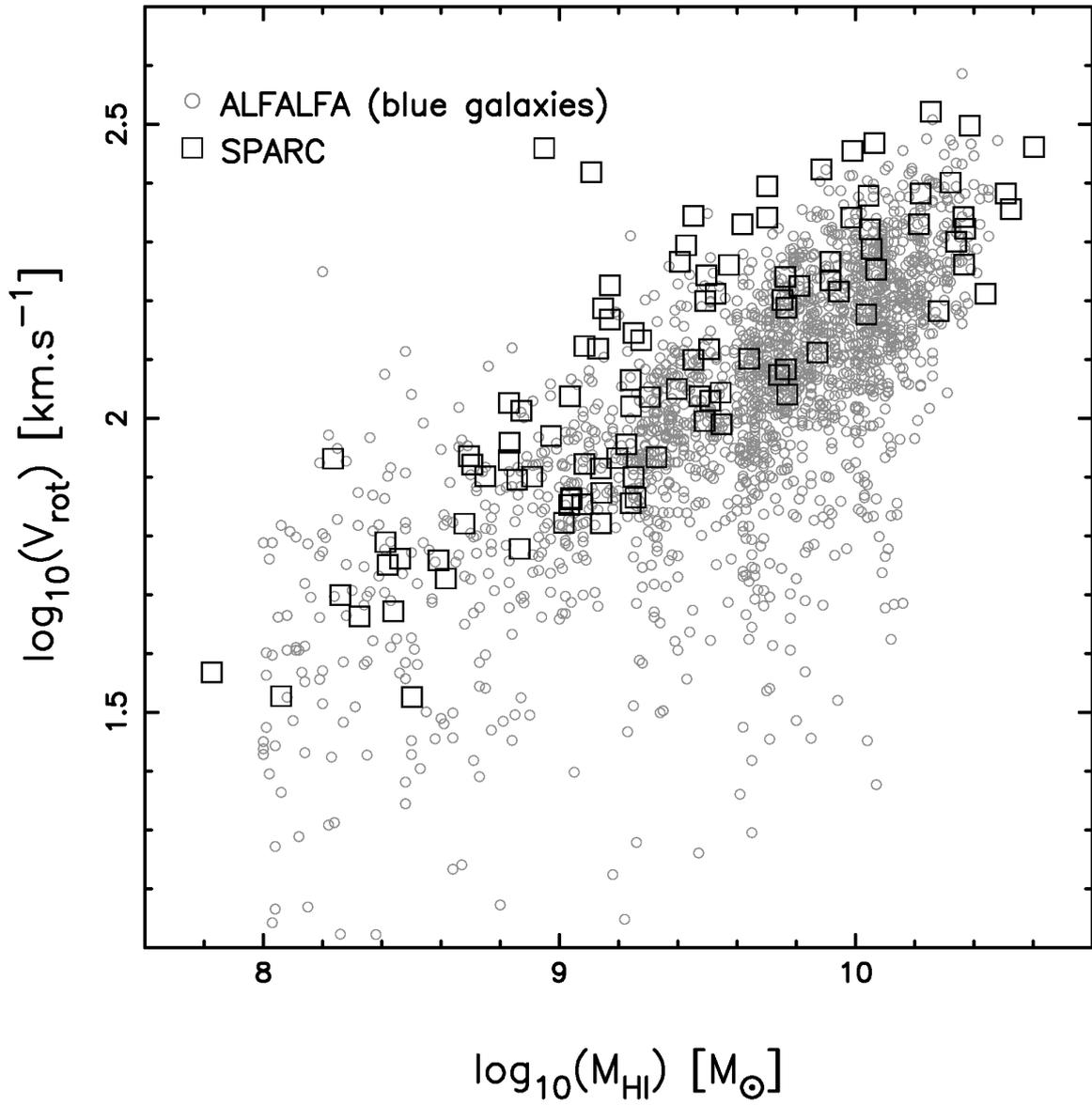


Figure 4.13: A comparison of HI properties, $M_{\text{HI}}-V_{\text{rot}}$, for a volume limited sample of blue galaxies in ALFALFA (grey open circle) and SPARC (black open square).

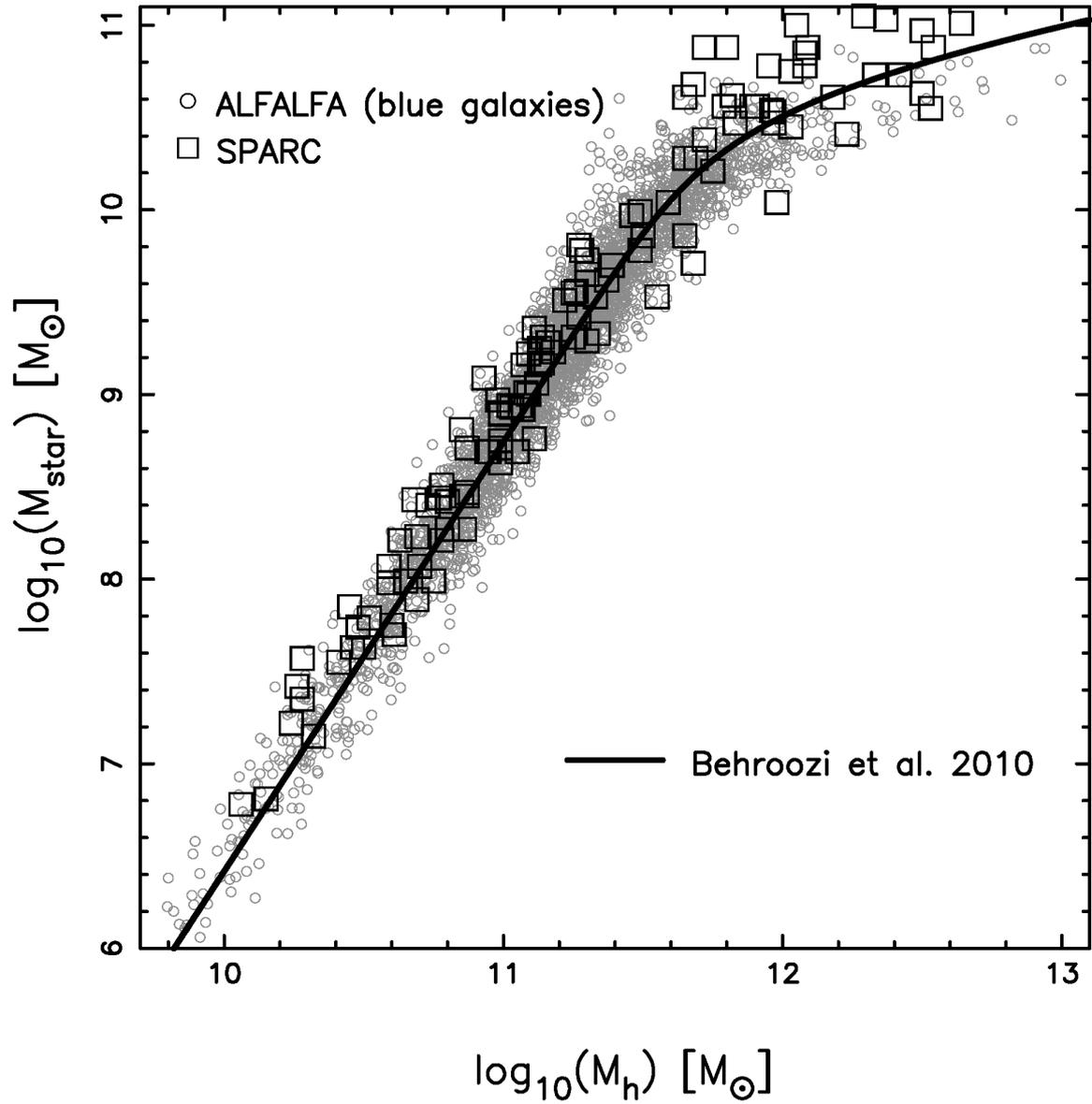


Figure 4.14: A comparison of HI properties, $M_{\text{h}}-M_{\text{star}}$, for a volume limited sample of blue galaxies in ALFALFA (grey open circle) and SPARC (black open square). The solid line is the $M_{\text{star}}-M_{\text{h}}$ relation of Behroozi et al. [2010]

2016] catalog to determine halo masses of late-type HI-selected galaxies in HIPASS. The 2DSWML method was then used to obtain the HI-selected HMF for early type galaxies by binning in M_h instead of w_{50} or V_{rot} . The SPARC catalog uses near-infrared (NIR) Spitzer photometry ($3.6\mu\text{m}$) to trace the stellar mass distribution in galaxies. This is important to break the star-halo degeneracy [Lelli et al., 2016] when mass modeling galaxies. Additionally it relies on HI/H α rotation curve measurements over the past 3 decades. The rotation curves are finally fitted assuming a halo profile, which results in a halo mass $M_{200} \equiv M_h$ estimate of the galaxy [Li et al., 2019a]. Although the halo mass estimates exist for many profiles, we will only consider the estimate based on the NFW [Navarro et al., 1996] profile. Finally SPARC extracts profile widths, w_{50} of these galaxies from the *Extragalactic Distance Database* [Tully et al., 2009, Courtois et al., 2009]. The galaxy properties e.g. M_{star} , M_h , M_{HI} , V_{rot} , w_{50} in SPARC forms a near homogeneous data set [Lelli et al., 2016].

In figure 4.13 and 4.14 we compare the $M_{\text{HI}} - V_{\text{rot}}$ and the $M_h - M_{\text{star}}$ relations, respectively, between a volume limited sample of blue galaxies in ALFALFA (grey open circle) and SPARC (black open square). We take the same volume limited sample as shown in figure 4.12 and choose galaxies with inclinations $i < 45^\circ$ to reliably obtain V_{rot} [Zwaan et al., 2010] from w_{50} after correcting for inclination. The volume limited sample is the same in figure 4.15 and 4.16. However in the figures which involve V_{rot} the sample is smaller since it excludes galaxies with $i > 45^\circ$. The scaling of HI properties (figure 4.13) between ALFALFA and SPARC agree with each other. The SPARC sample is also homogeneously sampling the range of $M_{\text{HI}} - V_{\text{rot}}$ covered by ALFALFA. In figure 4.14 we compare the $M_{\text{star}} - M_h$ relation between the ALFALFA and SPARC samples. For ALFALFA we do not have an independent measure of M_h , we have therefore used the $M_{\text{star}} - M_h$ scaling relation from Behroozi et al. [2010] (black solid line) to convert M_{star} into M_h after accounting for scatter. The SPARC sample agrees with the Behroozi et al. [2010] relation and therefore with the ALFALFA points, confirming again that this relation is tight. Here too we see that

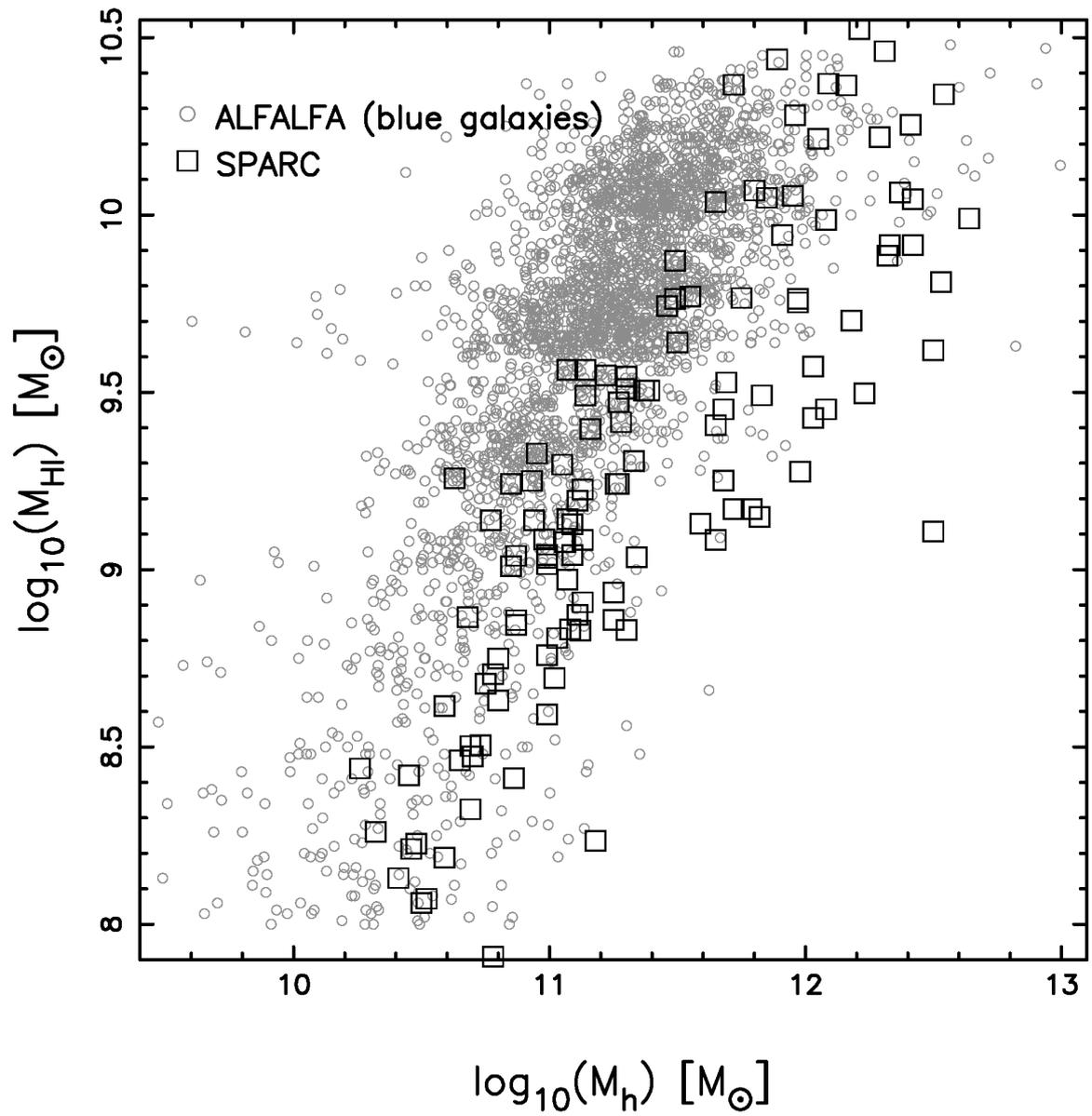


Figure 4.15: A comparison of $M_{\text{HI}}-M_h$, for a volume limited sample of blue galaxies in ALFALFA (grey open circle) and SPARC (black open square).

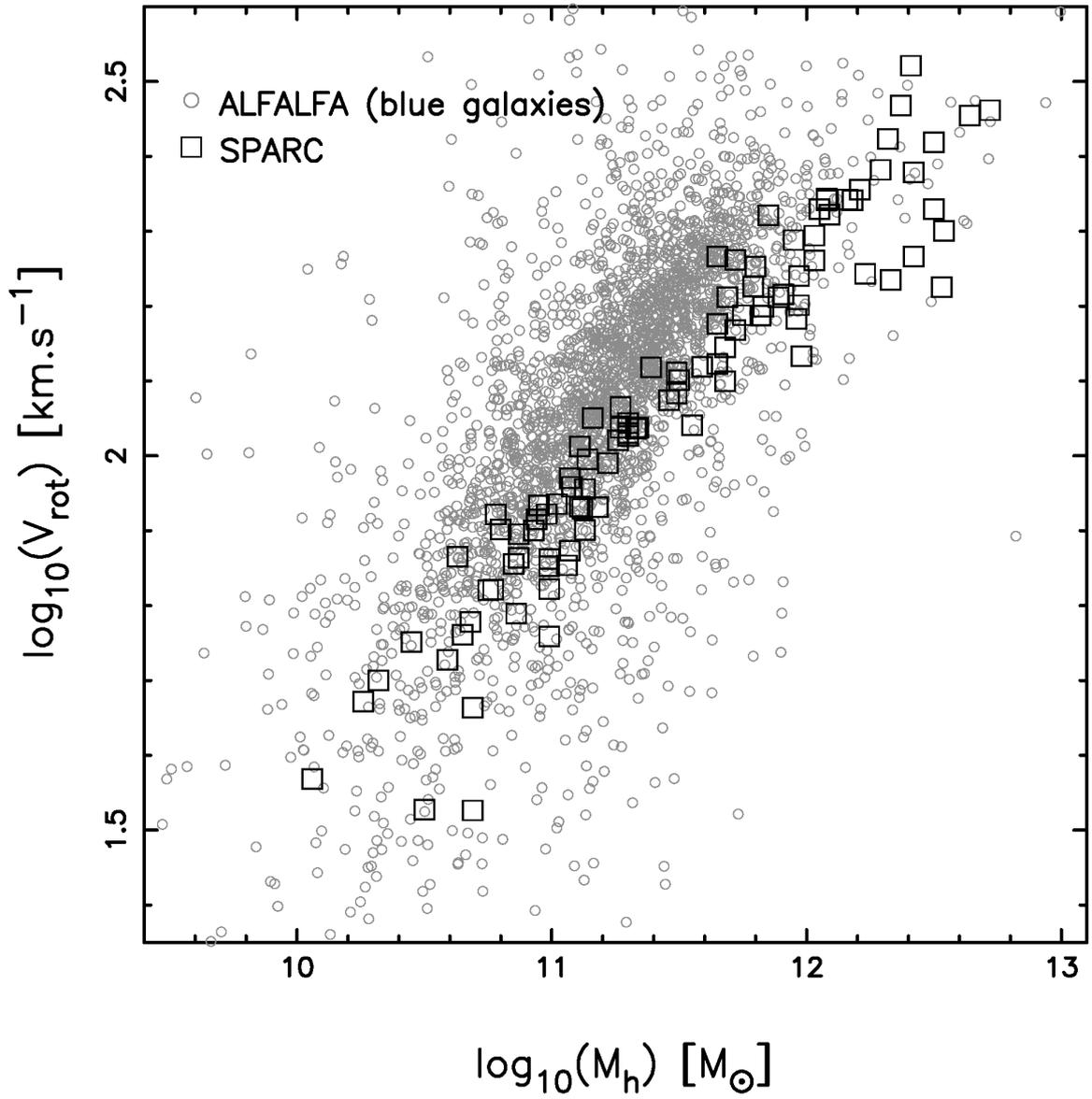


Figure 4.16: A comparison of $V_{\text{rot}}-M_h$, for a volume limited sample of blue galaxies in ALFALFA (grey open circle) and SPARC (black open square).

SPARC samples homogeneously the $M_{\text{star}}-M_{\text{h}}$ range.

We now compare the $M_{\text{HI}} - M_{\text{h}}$ and the $V_{\text{rot}} - M_{\text{h}}$ relation between ALFALFA and SPARC in figures 4.15 and 4.16, respectively. We see a marked difference between ALFALFA and SPARC in both these figures. We point out that if we replaced M_{h} with M_{star} the systematic differences remain. Clearly the joint distribution between an HI property (M_{HI} or V_{rot}) and optical property (M_{star} which is a proxy for M_{h}) is different between ALFALFA and SPARC. Although SPARC homogeneously samples individual properties, it is clearly biased compared to ALFALFA. For ALFALFA the selection function is well understood, however SPARC does not have a corresponding selection function since it relies on individual objects. At fixed, M_{HI} or V_{rot} SPARC predicts a larger halo mass. We would therefore expect the HI-selected HMF for late-type galaxies using the SPARC scaling ($M_{\text{h}} - V_{\text{rot}}$ or $M_{\text{h}} - M_{\text{HI}}$), to be offset towards larger halo masses when compared to our result.

This is shown in figure 4.17. The estimate by Li et al. [2019b] is offset by about 0.7 dex towards the right at lower masses and has a sharper exponential drop compared to our results (blue thin line). The offset is consistent with the offset of 0.77 dex which we estimate from figure 4.15. However the sharp drop at larger masses cannot be reconciled with our results by simple scaling arguments. We leave the investigation of such issues to future work.

We summarize our results below.

- We have shown that the HIWF and HIVE are described by modified Schechter functions. Both these abundances are well separated for the red and blue populations at larger velocities. The red population dominates the high velocity end and the blue population dominates the velocity function at the knee and lower velocities.
- A qualitatively similar result is seen for the HIMF. However unlike the HIWF and HIVE we find that the differences in the HIMF for the red and blue populations, at

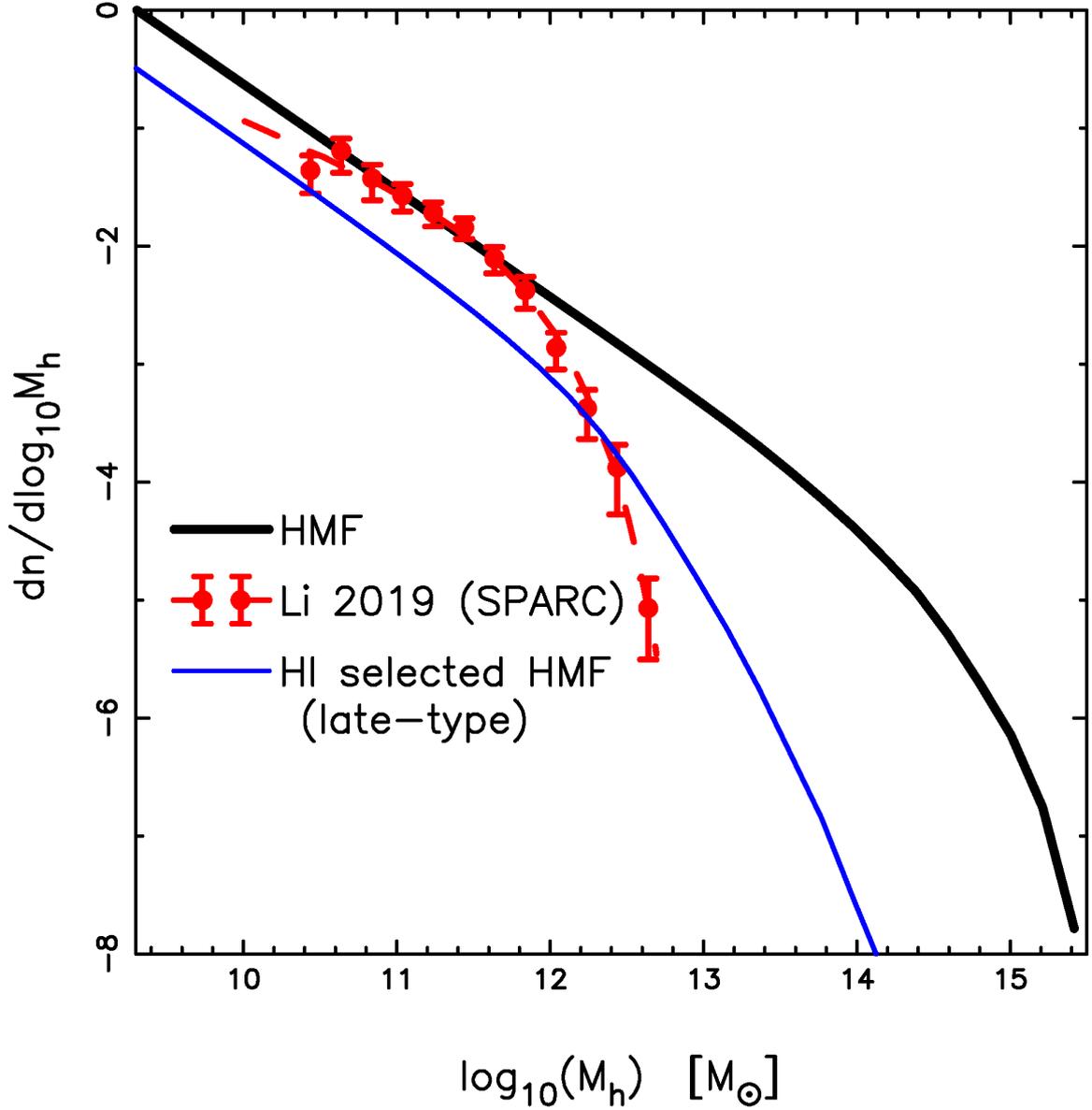


Figure 4.17: The black thick line is the HMF. The blue thin line is our estimate of the HI-selected HMF for late-type (or blue) galaxies based on ALFALFA data. The red dashed line is the HI-selected HMF for late-type galaxies by Li et al. [2019b] based on SPARC and HIPASS data.

the high mass end, is less pronounced.

- Using the recent observational $\langle M_{\text{HI}} \rangle - M_{\text{h}}$ relation of [Guo et al. \[2020\]](#) we have estimated the HI-selected HMF (equation 4.5) which represents the abundances of halos based on HI mass. Using a semi-analytic model based galaxy catalog (SAGE), we also estimate the HI-selected HMF for the red and blue galaxies.
- Using these six abundances which describe HI rich galaxies we have obtained scaling relations between HI properties ($M_{\text{HI}} - w_{50} - V_{\text{rot}}$) and HI-halo properties $M_{\text{HI}} - M_{\text{h}}$, $w_{50} - M_{\text{h}}$ and $V_{\text{rot}} - M_{\text{h}}$ (See figures 4.10 and 4.11).
- The $M_{\text{HI}} - M_{\text{h}}$ scaling relation is robust and consistent with a volume limited sample in ALFALFA. It is described by a steep power law slope ~ 2.1 at small masses and transitions to a shallow power slope ~ 0.3 at masses larger than $M_{\text{h}} \gtrsim 10.6$. It has a shape similar to the $M_{\text{star}} - M_{\text{h}}$ [[Behroozi et al., 2019](#)] but the transition scale is smaller by about 0.5 dex compared to that of the $M_{\text{star}} - M_{\text{h}}$ relation. This suggests that baryonic processes like heating and feedback suppress the HI content in large mass halos on a shorter timescale as compared to star-formation. The star formation in massive red galaxies is suppressed as they are in large halos with virial temperatures $T_{\text{vir}} \sim 10^7$ in the feedback dominated phase. The star formation is strongly correlated to the clumping of the cold gas. As seen in the ATLAS^{3D} HI survey [[Cappellari et al., 2011](#), [Serra et al., 2012](#)], the HI in the gas rich early-type red galaxies are not as dense as the gas rich late-type spiral galaxies. Which means, any small amount of feedback will easily ionize the HI gas in these systems in a shorter time scale. Whereas, feedback will suppress star formation on a longer time scale due to self-shielding in the dense star forming regions.

Appendix A

A1 Estimation of HIMF using 2DSWML Method

As discussed earlier, the HIMF ($\phi(M_{\text{HI}})$) is the number density of galaxies with HI masses in $[M_{\text{HI}}, M_{\text{HI}} + dM_{\text{HI}}]$. For a volume-limited sample, which contains all the galaxies above a threshold mass ($M_{\text{HI}}^{\text{th}}$) upto a limiting redshift (z_{lim}), the HIMF can be estimated just by counting the number of galaxies (N_j) in bins of HI mass (M_{HI}^j), and normalizing the counts by the total number of galaxies (N) in the sample and the volume of interest (V). But in the case of flux-limited surveys, which is associated with a threshold flux value (S^{th}), all the galaxies with flux $S \geq S^{\text{th}}$ are selected. Since flux is directly proportional to the luminosity of the galaxy and inversely proportional to the square of the distance to the galaxy, a galaxy appears to be fainter either because of it has a lower intrinsic luminosity value or due to its large distance, i.e., higher redshift value. Therefore counting the number of galaxies in each mass bin will not produce the true HIMF in a flux-limited survey. The counts should be weighted properly so that it represents the true number of galaxies in a bin.

In this appendix, we give a precise overview of the 2DSWML method, which is used in Chapters 2, 3, and 4 of this thesis. We start with the analytical definition in sec A1.1 and illustrate the steps to be followed. In sec A1.2 we define a method to normalize the distribution function. Finally we describe the measurement of uncertainties on the HIMF in sec A1.3.

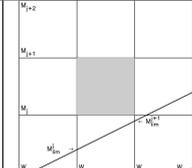
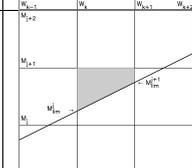
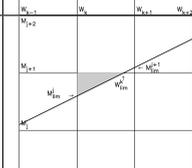
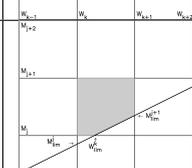
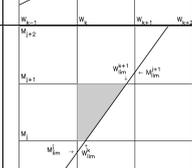
No.	Conditions	H_{ijk}
C1		1.0
C2		0.0
C3		$\frac{2M_{j+1} - M_{lim}^{j+1} - M_{lim}^j}{2\Delta M}$
C4		$\frac{(M_{j+1} - M_{lim}^j)(W_{lim}^{k+1} - W_k)}{2\Delta M \Delta W}$
C5		$1.0 - \frac{(M_{lim}^{j+1} - M_j)(W_{k+1} - W_{lim}^k)}{2\Delta M \Delta W}$
C6		$\frac{W_{lim}^k - W_k}{\Delta W} + \frac{W_{lim}^{k+1} - W_{lim}^k}{2\Delta W}$

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_{lim}^j, M_{lim}^k if it does not exactly intersect on the bin edges W_k, M_j .

A1.1 2DSWML Method : Definition

The Step-Wise Maximum Likelihood (SWML) method is a non-parametric way to estimate the HIMF by maximizing the joint likelihood of galaxies to be detected. For radio observations like ALFALFA the detection of galaxies depend not only on the HI flux (S_{21}), i.e., the HI mass (M_{HI}), but it depends also on the velocity profile width (w_{50}) [figure 2.4]. Therefore the underlying distribution is a bivariate distribution of M_{HI} and w_{50} , and we use a 2-dimensional generalization of the SWML method to obtain the HIMF/HIWF.

We follow the implementation by Zwaan et al. [2003], Martin et al. [2010] for the 2DSWML method. Assuming the sample is drawn from a bivariate distribution $\phi(M_{\text{HI}}, w_{50})$, the probability of detecting a galaxy i at a distance D^i with HI mass M_{HI}^i and velocity width w_{50}^i is

$$p_i = \frac{\phi(M_{\text{HI}}^i, w_{50}^i)}{\int_{w_{50}=0}^{\infty} \int_{M_{\text{HI}}=M_{\text{HI},lim}(D^i, w_{50}^i)}^{\infty} \phi(M_{\text{HI}}, w_{50}) dM_{\text{HI}} dw_{50}} \quad (\text{A1})$$

where, $M_{\text{HI},lim}$ is the limiting HI mass and depends on survey sensitivity. It can be detected with velocity width w_{50}^i at a distance D^i in the survey.

The above equation can be discretized as

$$p_i = \frac{\sum_j \sum_k V_{ijk} \phi_{jk}}{\sum_j \sum_k H_{ijk} \phi_{jk} \Delta M \Delta W} \quad (\text{A2})$$

Here we consider the distribution of galaxies in bins of $M = \log_{10} [M_{\text{HI}}/M_{\odot}]$ and $W = \log_{10} [w_{50}/(\text{km.s}^{-1})]$. N_M and N_W are the number of bins and ΔM and ΔW are the widths of the bins, respectively. The two-dimensional distribution can therefore be parametrized by ϕ_{jk} , where $j = 0, 1, \dots, N_M - 1$ and $k = 0, 1, \dots, N_W - 1$. V_{ijk} is a binary value which ensures that galaxy 'i' is only populated in its corresponding 'j-k' bin.

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

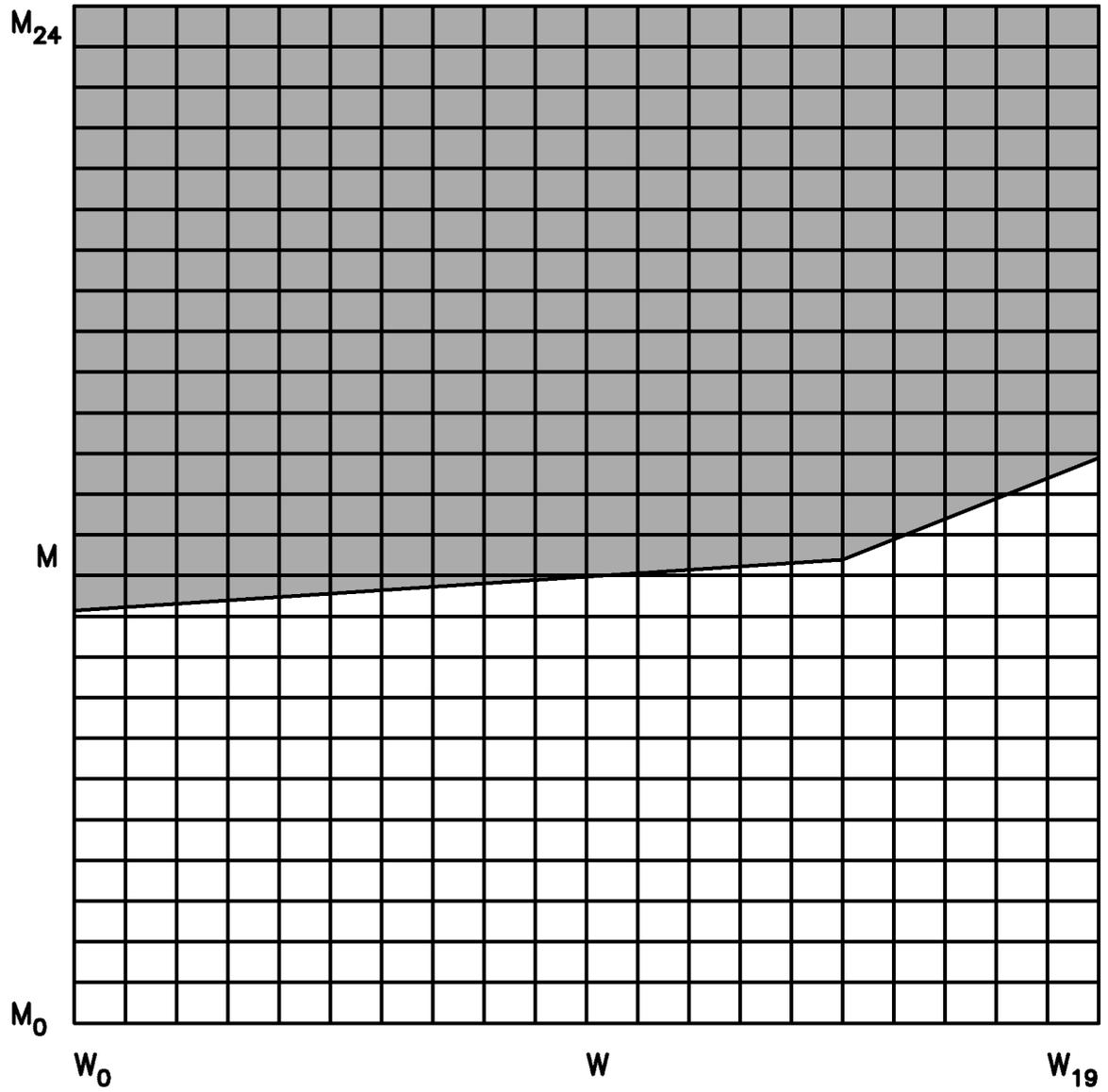


Figure A1: The completeness relation (*broken solid line*) is shown in the M - W plane. This particular galaxy 'i' has the following properties: $\log_{10} [M_{\text{HI}}^i/M_{\odot}] = 8.73$, $[w_{50}^i/\text{km.s}^{-1}] = 93$, $[D/\text{Mpc} = 27]$. The shaded(white) region is the area accessible(inaccessible) to this galaxy in the M - W plane.

H_{ijk} is a weight associated to every galaxy 'i' in their 'jk' plane. It takes values from 0 to 1, based on the survey completeness condition. H_{ijk} s are evaluated such that the denominator in the eq. A1 can be written as an integration over the entire plane of M - W after convolving with the survey's completeness function C^i —

$$\begin{aligned} & \int_{w_{50}=0}^{\infty} \int_{M_{\text{HI}}=M_{\text{HI},\text{lim}}(D^i, w_{50}^i)}^{\infty} \phi(M_{\text{HI}}, w_{50}) dM_{\text{HI}} dw_{50} \\ &= \int_{W=W_0}^{W=W_{N_W-1}} \int_{M=M_0}^{M=M_{N_M-1}} C^i(M, W) dM dW \end{aligned} \quad (\text{A4})$$

And this is given by

$$\Delta M \Delta W \sum_{k=0}^{N_W-1} \sum_{j=0}^{N_M-1} H_{ijk} = \int_{W=W_0}^{W=W_{N_W-1}} \int_{M=M_0}^{M=M_{N_M-1}} C^i(M, W) dM dW \quad (\text{A5})$$

An example is shown in figure A1 for one of the galaxies in our sample. The solid broken line is the 50% completeness relation as given in eqn 2.1, and the shaded region is the accessible region for that galaxy, which is given by eq. A5.

Table A1 explains how we compute the H_{ijk} s for different cases, where the shaded area in column 2 represents the 'j-k' bin of interest. For case 1 and case 2, the completeness curve does not intersect the shaded bin of interest and depending on whether that curve is below the bin or above the bin H_{ijk} is 1 or 0, respectively. For cases 3, 4 and 5, the completeness curve intersects the bin ($W_{\text{lim}}^k, M_{\text{lim}}^j$ denote the intersection point if the intersection does not happen exactly on the edges (W_k, M_j)) and divide the shaded square in two parts; and H_{ijk} is the fraction of area above the completeness line to the total. Column 3 gives the formula to estimate H_{ijk} for these three cases. Case 6 assumes a completeness slope ≥ 1 , which is not the case for ALFALFA. Also the completeness relation eq. 2.1 has a change in slope at $\log_{10} [w_{50}/(\text{km.s}^{-1})] = 2.5$, our choice of W bins are such that this value coincides with the edge of the bin. Table A1 assumes this so that no further cases are considered.

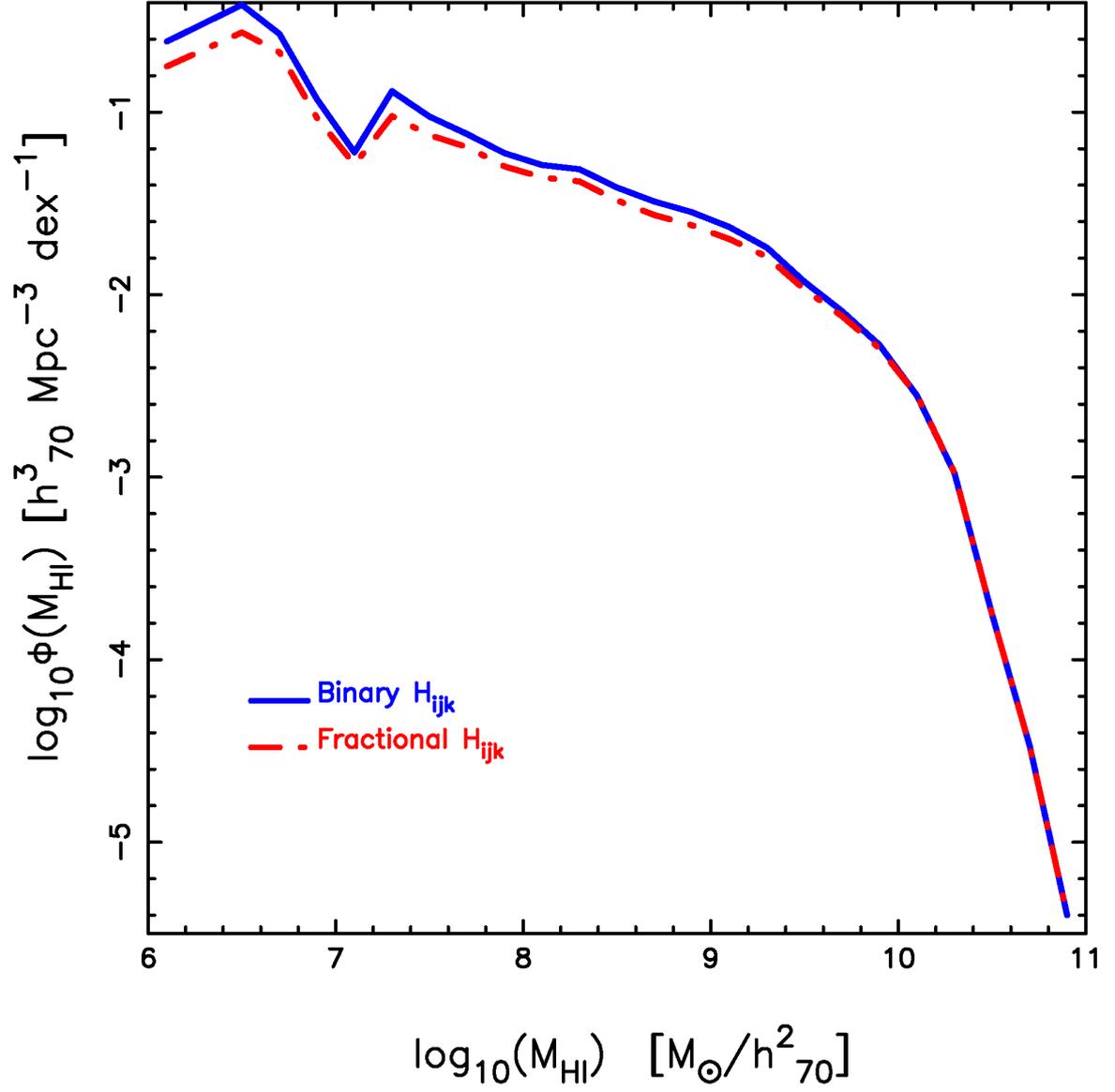


Figure A2: Comparison of the results from 2DSWML method with considering H_{ijk} s as 0 or 1 (blue solid line) and with fractional values of H_{ijk} s (red dot-dashed line).

Efstathiou et al. [1988] consider H_{ijk} s as binary values :

$$H_{ijk} = \begin{cases} 1, & \text{if galaxy } i \text{ satisfies the completeness condition to be in mass} \\ & \text{bin } j \text{ and profile width bin } k. \\ 0, & \text{otherwise.} \end{cases} \quad (\text{A6})$$

If the midpoint of the j - k bin of interest is above the completeness curve H_{ijk} is '1', in the other case it will be '0'. This may be seen as a limiting case when the number of galaxies is large and may be thought of as an approximation. Figure A2 shows a comparison between the two calculations of H_{ijk} s. The blue solid line represents the HIMF by considering binary values of H_{ijk} s and the red dashed line is the HIMF obtained using the fractional values of H_{ijk} s. Binary H_{ijk} s lead to an overestimation of the HIMF in the smaller HI mass values, while the high mass end of the HIMF is the same with both methods.

With the 2DSWML method, we obtain ϕ_{jk} s for which the joint probability or likelihood of detecting all the galaxies in the sample has a maximum value :

$$\mathcal{L} = \prod_{i=1}^{N_g} p_i \quad (\text{A7})$$

Using eq. A2 the joint likelihood is

$$\mathcal{L} = \prod_{i=1}^{N_g} \frac{\sum_j \sum_k V_{ijk} \phi_{jk}}{\sum_j \sum_k H_{ijk} \phi_{jk} \Delta M \Delta W} \quad (\text{A8})$$

We rather maximize the log-likelihood with respect to the ϕ_{jk}

$$\ln \mathcal{L} = \sum_{i=1}^{N_g} \ln \left(\frac{\sum_j \sum_k V_{ijk} \phi_{jk}}{\sum_j \sum_k H_{ijk} \phi_{jk} \Delta M \Delta W} \right) \quad (\text{A9})$$

This gives us the ϕ_{jk} s

$$\begin{aligned} \phi_{jk} &= [\sum_i V_{ijk}] \left[\sum_i \frac{H_{ijk}}{\sum_m \sum_n H_{imn} \phi_{mn}} \right]^{-1} \\ &= n_{jk} \left[\sum_i \frac{H_{ijk}}{\sum_m \sum_n H_{imn} \phi_{mn}} \right]^{-1} \end{aligned} \quad (\text{A10})$$

where, $n_{jk} = \sum_i V_{ijk}$ is the number of galaxies in j - k bin. ϕ_{jk} s are determined iteratively from eq. A10. We consider the initial value of ϕ_{jk} to be $n_{jk} / [V_{\text{surv}} \Delta M \Delta W]$ to start the

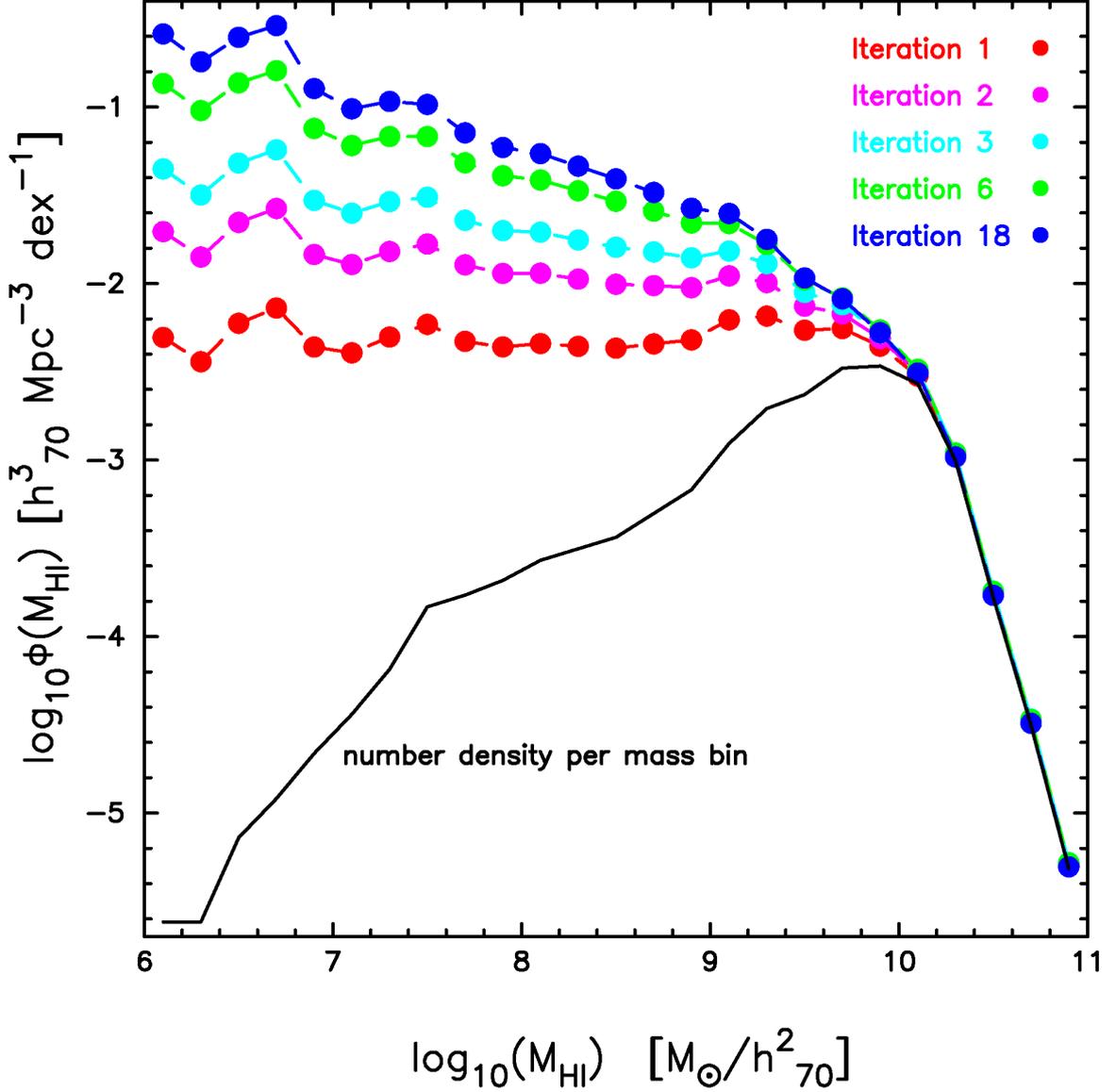


Figure A3: HIMF at different iterations. The black solid line is the observed number density. The blue filled circles connected by blue dashed line represents the true HIMF obtained for our sample. The red, pink, cyan and green filled circles are the ϕ_j s after 1, 2, 3 and 6 iterations, respectively.

iteration. We set a relative tolerance of 10^{-3} for all ϕ_{jk} s to converge. In this analysis we have taken 10 bins per dex in velocity width and 5 bins per dex in mass. And for our sample ϕ_{jk} converges within 20 iterations, illustrated in fig. A3.

Figure A4 shows our result from 2DSWML method. The color-coded pixels in the bottom left panel are the ϕ_{jk} values in the $M_{\text{HI}} - w_{50}$ plane. The orange dots are the observed ALFALFA galaxies. Finally the HIMF is obtained by summing over the velocity width bins k (see top left panel of figure A4).

$$\phi_j = \sum_k \phi_{jk} \Delta W \quad (\text{A11})$$

And the HIWF is obtained by summing over the HI mass bins j (see bottom right panel of figure A4).

$$\phi_k = \sum_j \phi_{jk} \Delta M \quad (\text{A12})$$

A1.2 Normalization of HIMF and HIWF

The likelihood methods have one disadvantage regarding the normalization of the distribution function, which is obvious from eq. A1. There are a number of ways to fix the normalization [Davis and 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, which can be checked by comparing the ratio of the observed counts per unit volume and the normalized mass function from the 2DSWML method. In the top panel of figure A5, the black filled pentagons connected with solid black line represent the observed HIMF and the red filled triangles connected with red dashed line represent the un-normalized HIMF from the 2DSWML method, for the ALFALFA 100% data ($\alpha.100$) [Haynes et al., 2018].

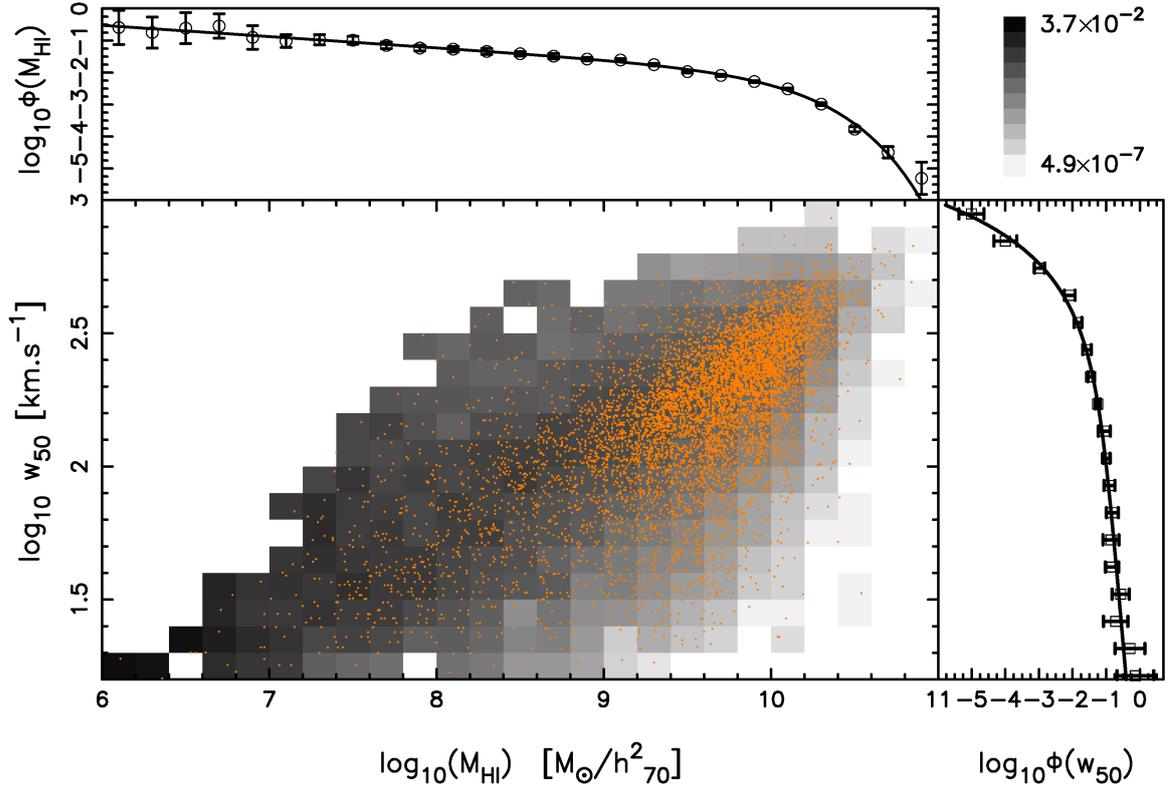


Figure A4: Results from 2DSWML method. Bottom left panel: the true two-dimensional distribution of ALFALFA galaxies in $M_{\text{HI}} - w_{50}$ plane. Orange dots are the observed galaxies from our sample of analysis. Top left panel: HIMF for our total sample, obtained by integrating the 2D distribution at the bottom panel over w_{50} bins. Bottom right panel: HIWF for our total sample, obtained by integrating the 2D distribution over M_{HI} bins. The error bars and the normalization of the distribution functions are discussed in the next sections.

The lower panel of figure A5 shows (blue squares and blue solid line) the ratio of these for the last 7 HI mass bins. The flat feature of the ratio curve at the high HI mass range supports the fact that the selection function does not affect the high mass end of the HIMF. We find that in this particular example of $\alpha.100$ 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 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 progress sequentially to 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 and Huchra [1982], Martin et al. [2010]. We find that they match at the relative level of $\sim 0.4\%$. Finally we compare our result with that of $\alpha.100$ HIMF in figure A6 [Jones et al., 2018]. We can see that our implementation reproduces the HIMF of Haynes et al. [2018] extremely well.

In case of the HIWF we can not apply the argument that the observed distribution at higher velocity widths will not be affected by the selection function. Thus it is not feasible to compare the observed number density per velocity width bin with the unnormalized HIWF at higher velocity width and estimate the normalization factor. We therefore use the method as described in Martin et al. [2010] to normalize the HIWFs. Or one can use the normalization factors from HIMF estimations and apply those factors in case of HIWFs, if the ranges and the number of bins in HI mass and velocity width are same for both the cases.

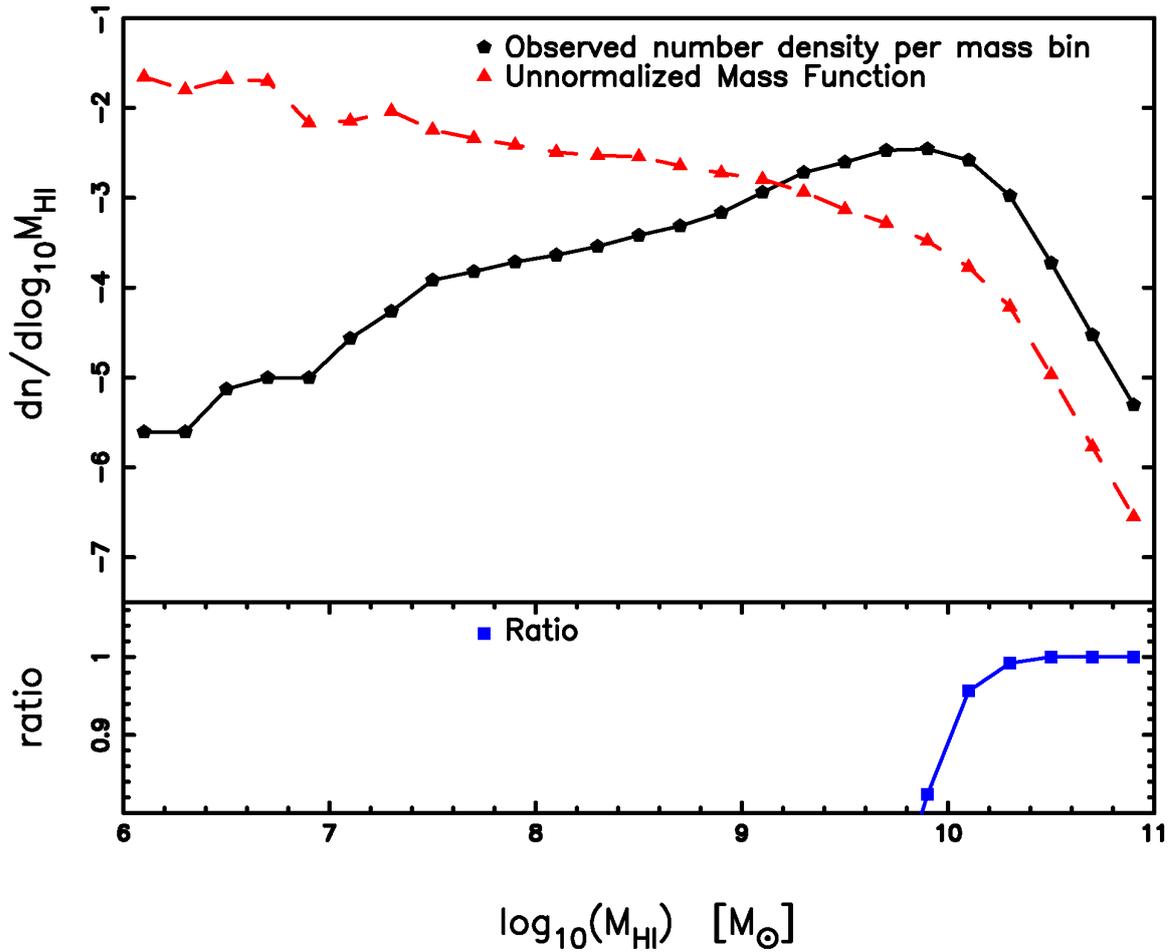


Figure A5: Upper panel: Red dashed line is the mass function estimated using 2DSWML method, which is not normalized. The black solid line is the number density per mass bin calculated for the same sample of HI galaxies. Lower panel: The blue solid line shows the ratio of the un-normalized mass function and the number density, multiplied by 100.

A1.3 Error Analysis

(i) *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. Therefore we consider Poisson errors for both the distribution functions — HIMF and HIWF, to obtain the error on their estimates.

(ii) *Mass Errors*: Since HI mass is derived from integrated flux and distance of the galaxy, $M_{\text{HI}} = 2.356 \times 10^5 S_{21} D^2$, the uncertainties in M_{HI} is related to the uncertainties in S_{21} and in D . The distance estimation of the galaxies are affected by the peculiar velocities. In the local volume it has a larger effect and at higher redshifts Hubble flow dominates over peculiar velocities resulting in a smaller effect on the estimates of distances. The ALFALFA catalog [Haynes et al. \[2011\]](#) quotes the distances of galaxies as cz_{CMB}/H_0 for $cz_{\text{cmb}} > 6000$ km.s⁻¹, and $cz_{\text{cmb}} < 6000$ km.s⁻¹ the distances are estimated using a local flow model [Masters \[2005\]](#). This model has an estimated local velocity dispersion of $\sigma_v = 163$ km.s⁻¹. For these galaxies we take the error in the distance to be the maximum of σ_v and 10% of the distance. For galaxies $cz_{\text{CMB}} > 6000$ km.s⁻¹ we take the error on distances to be 10%. We generate 300 Gaussian realizations using the errors on distance and fluxes and their observed values and recompute M_{HI} for every object. We then apply the 2DSWML method to each of the realizations and find out the width of the distribution for every ϕ_j which we quote as an error on ϕ_j . We follow this part of error measurements only for the HIMF.

(iii) *Velocity width errors*: The ALFALFA catalog lists the uncertainties on the observed velocity width. We generate 300 Gaussian realizations using the observed values of w_{50} and errors on w_{50} and perform the 2DSWML method for each of them to find out a spread on ϕ_k . Also since the completeness relation is related to the distance of the galaxies, uncertainties on distances also have a contribution to the errors on HIWF. Therefore we find the distance uncertainties as discussed in the previous point above and generate 300 realizations again

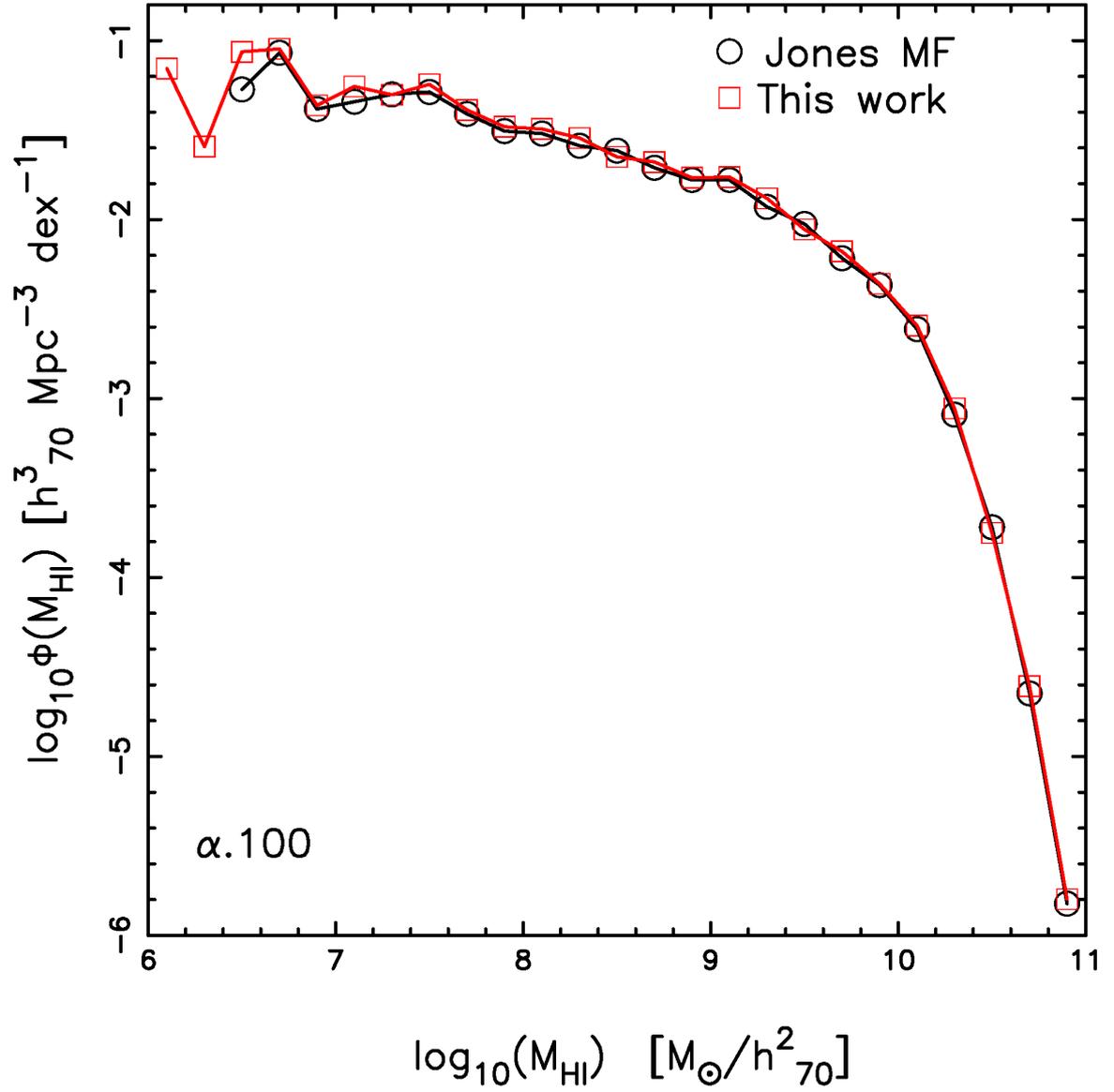


Figure A6: Comparison of the HIMF in the $\alpha.100$ sample. The open circles is the HIMF by Jones et al. [2018]. The open squares is the HIMF estimated by our implementation of the 2DSWML method.

to obtain the spread around the ϕ_k . These steps are followed only for the HIWF.

(iv) *Sample Variance:* We estimate this error by splitting the total area into 26 contiguous regions of approximately equal area each. The HIMFs/HIWFs are computed by removing one region at a time. Finally the jackknife uncertainty for ϕ_j is computed as $\sigma_{\phi^j} = \frac{N-1}{N} \sum_{i=1}^{N=26} (\bar{\phi}^j - \phi_i^j)^2$ where $\bar{\phi}^j$ is the jackknife mean and ϕ_i^j is the value for the i^{th} jackknife sample.

(v) *Other Errors:* Jones et al. [2018] discuss many other sources of errors. E.g. the error associated with the 2DSWML method which one can either estimate using the information matrix [Efstathiou et al., 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 in the HIMF calculation is the error in the velocity width w_{50} . Since the peak flux, S_{peak} , is inversely proportional to w_{50} we expect their errors to be correlated. As we integrate over w_{50} to obtain the HIMF, we do not consider errors in w_{50} .

We find the final errorbars on HIMF by adding the poisson errors, mass errors and sample variance in quadrature, and for HIWF we add the poisson errors, velocity width errors and sample variance. 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 $\chi_{\text{red}}^2 < 1$ which means that we may be overestimating the errors. Our HIMF and HIWF errors are comparable to Martin et al. [2010], Haynes et al. [2011] and Papastergis et al. [2011], Moorman et al. [2014], respectively.

A2 List of Constants

Gravitational constant	$G = 6.674 \times 10^{-11} \text{ m}^3\text{kg}^{-1}\text{s}^{-2}$
Planck constant	$h_p = 6.626 \times 10^{-34} \text{ m}^2\text{kg s}^{-1}$
Speed of light	$c = 3 \times 10^5 \text{ km s}^{-1}$
Boltzmann constant	$k_B = 1.381 \times 10^{-16} \text{ erg K}^{-1}$
Proton mass	$m_p = 1.673 \times 10^{-27} \text{ kg}$
Neutron mass	$m_n = 1.675 \times 10^{-27} \text{ kg}$
Electron mass	$m_e = 9.109 \times 10^{-31} \text{ kg}$
Solar mass	$M_\odot = 2 \times 10^{30} \text{ kg}$
Parsec	$pc = 3.086 \times 10^{16} \text{ m}$

References

- K. N. Abazajian et al. The Seventh Data Release of the Sloan Digital Sky Survey. *Astrophysical Journal, Supplement*, 182(2):543–558, June 2009. doi: 10.1088/0067-0049/182/2/543.
- T. Abel, G. L. Bryan, and M. L. Norman. The Formation and Fragmentation of Primordial Molecular Clouds. *Astrophysical Journal*, 540(1):39–44, Sept. 2000. doi: 10.1086/309295.
- N. Aghanim et al. Planck 2018 results. VI. Cosmological parameters. *Astronomy and Astrophysics*, 641:A6, Sept. 2020. doi: 10.1051/0004-6361/201833910.
- C. P. Ahn et al. The Tenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Apache Point Observatory Galactic Evolution Experiment. *Astrophysical Journal, Supplement*, 211(2):17, Apr. 2014. doi: 10.1088/0067-0049/211/2/17.
- R. A. Alpher, H. Bethe, and G. Gamow. The Origin of Chemical Elements. *Physical Review*, 73(7):803–804, Apr. 1948. doi: 10.1103/PhysRev.73.803.
- J. S. Bagla, N. Khandai, and K. K. Datta. HI as a probe of the large-scale structure in the post-reionization universe. *Monthly Notices of the RAS*, 407(1):567–580, Sept. 2010. doi: 10.1111/j.1365-2966.2010.16933.x.
- I. K. Baldry et al. Quantifying the Bimodal Color-Magnitude Distribution of Galaxies. *Astrophysical Journal*, 600(2):681–694, Jan. 2004. doi: 10.1086/380092.
- I. K. Baldry et al. Galaxy And Mass Assembly (GAMA): the galaxy stellar mass function at $z < 0.06$. *Monthly Notices of the RAS*, 421(1):621–634, Mar. 2012. doi: 10.1111/j.1365-2966.2012.20340.x.
- C. V. Barbieri, F. Fraternali, T. Oosterloo, G. Bertin, R. Boomsma, and R. Sancisi. Extraplanar gas in the spiral galaxy NGC 4559. *Astronomy and Astrophysics*, 439(3):947–956, Sept. 2005. doi: 10.1051/0004-6361:20042395.
- G. Battaglia, F. Fraternali, T. Oosterloo, and R. Sancisi. H I study of the warped spiral galaxy NGC 5055: a disk/dark matter halo offset? *Astronomy and Astrophysics*, 447(1):49–62, Feb. 2006. doi: 10.1051/0004-6361:20053210.
- R. H. Becker et al. Evidence for Reionization at $z \sim 6$: Detection of a Gunn-Peterson Trough in a $z=6.28$ Quasar. *Astronomical Journal*, 122(6):2850–2857, Dec. 2001. doi: 10.1086/324231.

- K. G. Begeman, A. H. Broeils, and R. H. Sanders. Extended rotation curves of spiral galaxies : dark haloes and modified dynamics. *Monthly Notices of the RAS*, 249:523, Apr. 1991. doi: 10.1093/mnras/249.3.523.
- A. Begum and J. N. Chengalur. Kinematics of two dwarf galaxies in the NGC 6946 group. *Astronomy and Astrophysics*, 424:509–517, Sept. 2004. doi: 10.1051/0004-6361:20041210.
- A. Begum, J. N. Chengalur, and U. Hopp. The little galaxy that could: kinematics of [10.1016/S1384-1076\(02\)00238-5](https://doi.org/10.1016/S1384-1076(02)00238-5) Camelopardalis B. *New Astronomy*, 8(4):267–280, May 2003. doi: 10.1016/S1384-1076(02)00238-5.
- P. Behroozi, R. H. Wechsler, A. P. Hearin, and C. Conroy. UNIVERSEMACHINE: The correlation between galaxy growth and dark matter halo assembly from $z = 0-10$. *Monthly Notices of the RAS*, 488(3):3143–3194, Sept. 2019. doi: 10.1093/mnras/stz1182.
- P. S. Behroozi, C. Conroy, and R. H. Wechsler. A Comprehensive Analysis of Uncertainties Affecting the Stellar Mass-Halo Mass Relation for $0 < z < 4$. *Astrophysical Journal*, 717(1):379–403, July 2010. doi: 10.1088/0004-637X/717/1/379.
- P. S. Behroozi, R. H. Wechsler, and C. Conroy. The Average Star Formation Histories of Galaxies in Dark Matter Halos from $z = 0-8$. *Astrophysical Journal*, 770(1):57, June 2013a. doi: 10.1088/0004-637X/770/1/57.
- P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu. The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. *Astrophysical Journal*, 762(2): 109, Jan. 2013b. doi: 10.1088/0004-637X/762/2/109.
- P. S. Behroozi et al. Gravitationally Consistent Halo Catalogs and Merger Trees for Precision Cosmology. *Astrophysical Journal*, 763(1):18, Jan. 2013c. doi: 10.1088/0004-637X/763/1/18.
- A. Bera, N. Kanekar, J. N. Chengalur, and J. S. Bagla. Atomic Hydrogen in Star-forming Galaxies at Intermediate Redshifts. *Astrophysical Journal, Letters*, 882(1):L7, Sept. 2019. doi: 10.3847/2041-8213/ab3656.
- F. Bigiel, A. Leroy, F. Walter, E. Brinks, W. J. G. de Blok, B. Madore, and M. D. Thornley. The Star Formation Law in Nearby Galaxies on Sub-Kpc Scales. *Astronomical Journal*, 136(6):2846–2871, Dec. 2008. doi: 10.1088/0004-6256/136/6/2846.
- S. Bird, R. Garnett, and S. Ho. Statistical properties of damped Lyman-alpha systems from Sloan Digital Sky Survey DR12. *Monthly Notices of the RAS*, 466(2):2111–2122, Apr. 2017. doi: 10.1093/mnras/stw3246.

- S. Blais-Ouellette, P. Amram, C. Carignan, and R. Swaters. Accurate determination of the mass distribution in spiral galaxies. III. Fabry-Perot imaging spectroscopy of 6 spiral galaxies. *Astronomy and Astrophysics*, 420:147–161, June 2004. doi: 10.1051/0004-6361:20034263.
- M. R. Blanton and S. Roweis. K-Corrections and Filter Transformations in the Ultraviolet, Optical, and Near-Infrared. *Astronomical Journal*, 133(2):734–754, Feb. 2007. doi: 10.1086/510127.
- J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser. Excursion Set Mass Functions for Hierarchical Gaussian Fluctuations. *Astrophysical Journal*, 379:440, Oct. 1991. doi: 10.1086/170520.
- A. Bosma. 21-cm line studies of spiral galaxies. II. The distribution and kinematics of neutral hydrogen in spiral galaxies of various morphological types. *Astronomical Journal*, 86:1825–1846, Dec. 1981. doi: 10.1086/113063.
- V. Bromm, P. S. Coppi, and R. B. Larson. Forming the First Stars in the Universe: The Fragmentation of Primordial Gas. *Astrophysical Journal, Letters*, 527(1):L5–L8, Dec. 1999. doi: 10.1086/312385.
- G. Bruzual and S. Charlot. Stellar population synthesis at the resolution of 2003. *Monthly Notices of the RAS*, 344(4):1000–1028, Oct. 2003. doi: 10.1046/j.1365-8711.2003.06897.x.
- M. Cappellari et al. The ATLAS^{3D} project - I. A volume-limited sample of 260 nearby early-type galaxies: science goals and selection criteria. *Monthly Notices of the RAS*, 413(2):813–836, May 2011. doi: 10.1111/j.1365-2966.2010.18174.x.
- C. M. Casey, D. Narayanan, and A. Cooray. Dusty star-forming galaxies at high redshift. *Physics Reports*, 541(2):45–161, Aug 2014. ISSN 0370-1573. doi: 10.1016/j.physrep.2014.02.009. URL <http://dx.doi.org/10.1016/j.physrep.2014.02.009>.
- B. Catinella et al. The GALEX Arecibo SDSS Survey - I. Gas fraction scaling relations of massive galaxies and first data release. *Monthly Notices of the RAS*, 403(2):683–708, Apr. 2010. doi: 10.1111/j.1365-2966.2009.16180.x.
- B. Catinella et al. The GALEX Arecibo SDSS Survey. VI. Second data release and updated gas fraction scaling relations. *Astronomy and Astrophysics*, 544:A65, Aug. 2012. doi: 10.1051/0004-6361/201219261.
- B. Catinella et al. The GALEX Arecibo SDSS Survey - VIII. Final data release. The effect of group environment on the gas content of massive galaxies. *Monthly Notices of the RAS*, 436(1):34–70, Nov. 2013. doi: 10.1093/mnras/stt1417.

- K.-H. Chae. Galaxy evolution from strong-lensing statistics: the differential evolution of the velocity dispersion function in concord with the Λ cold dark matter paradigm. *Monthly Notices of the RAS*, 402(3):2031–2048, Mar. 2010. doi: 10.1111/j.1365-2966.2009.16073.x.
- L. Chemin, C. Carignan, N. Drouin, and K. C. Freeman. H I Studies of the Sculptor Group Galaxies. VIII. The Background Galaxies: NGC 24 and NGC 45. *Astronomical Journal*, 132(6):2527–2538, Dec. 2006. doi: 10.1086/508859.
- A. Chowdhury, N. Kanekar, J. N. Chengalur, S. Sethi, and K. S. Dwarkanath. H I 21-centimetre emission from an ensemble of galaxies at an average redshift of one. *Nature*, 586(7829):369–372, Oct. 2020. doi: 10.1038/s41586-020-2794-7.
- C. Conroy and R. H. Wechsler. Connecting Galaxies, Halos, and Star Formation Rates Across Cosmic Time. *Astrophysical Journal*, 696(1):620–635, May 2009. doi: 10.1088/0004-637X/696/1/620.
- C. Conroy, J. E. Gunn, and M. White. The Propagation of Uncertainties in Stellar Population Synthesis Modeling. I. The Relevance of Uncertain Aspects of Stellar Evolution and the Initial Mass Function to the Derived Physical Properties of Galaxies. *Astrophysical Journal*, 699(1):486–506, July 2009. doi: 10.1088/0004-637X/699/1/486.
- A. Cooray and R. Sheth. Halo models of large scale structure. *Physics Reports*, 372(1):1–129, Dec. 2002. doi: 10.1016/S0370-1573(02)00276-4.
- S. Côté, C. Carignan, and K. C. Freeman. The Various Kinematics of Dwarf Irregular Galaxies in Nearby Groups and Their Dark Matter Distributions. *Astronomical Journal*, 120(6):3027–3059, Dec. 2000. doi: 10.1086/316883.
- H. M. Courtois et al. The Extragalactic Distance Database: All Digital H I Profile Catalog. *Astronomical Journal*, 138(6):1938–1956, Dec. 2009. doi: 10.1088/0004-6256/138/6/1938.
- R. Davé, R. Thompson, and P. F. Hopkins. MUFASA: galaxy formation simulations with meshless hydrodynamics. *Monthly Notices of the RAS*, 462(3):3265–3284, Nov. 2016. doi: 10.1093/mnras/stw1862.
- R. Davé, M. H. Rafieferantsoa, R. J. Thompson, and P. F. Hopkins. MUFASA: Galaxy star formation, gas, and metal properties across cosmic time. *Monthly Notices of the RAS*, 467(1):115–132, May 2017. doi: 10.1093/mnras/stx108.
- M. Davis and J. Huchra. A survey of galaxy redshifts. III. The density field and the induced gravity field. *Astrophysical Journal*, 254:437–450, Mar. 1982. doi: 10.1086/159751.

- W. J. G. de Blok and A. Bosma. High-resolution rotation curves of low surface brightness galaxies. *Astronomy and Astrophysics*, 385:816–846, Apr. 2002. doi: 10.1051/0004-6361:20020080.
- W. J. G. de Blok, S. S. McGaugh, and J. M. van der Hulst. HI observations of low surface brightness galaxies: probing low-density galaxies. *Monthly Notices of the RAS*, 283(1): 18–54, Nov. 1996. doi: 10.1093/mnras/283.1.18.
- W. J. G. de Blok, S. S. McGaugh, and V. C. Rubin. High-Resolution Rotation Curves of Low Surface Brightness Galaxies. II. Mass Models. *Astronomical Journal*, 122(5): 2396–2427, Nov. 2001. doi: 10.1086/323450.
- H. Dénes, V. A. Kilborn, and B. S. Koribalski. New H I scaling relations to probe the H I content of galaxies via global H I-deficiency maps. *Monthly Notices of the RAS*, 444(1): 667–681, Oct. 2014. doi: 10.1093/mnras/stu1337.
- B. Diemer et al. Atomic and molecular gas in IllustrisTNG galaxies at low redshift. *Monthly Notices of the RAS*, 487(2):1529–1550, Aug. 2019. doi: 10.1093/mnras/stz1323.
- N. Drory et al. The Bimodal Galaxy Stellar Mass Function in the COSMOS Survey to $z \sim 1$: A Steep Faint End and a New Galaxy Dichotomy. *Astrophysical Journal*, 707(2): 1595–1609, Dec. 2009. doi: 10.1088/0004-637X/707/2/1595.
- A. Durbala, R. A. Finn, M. Crone Odekon, M. P. Haynes, R. A. Koopmann, and A. A. O’Donoghue. The ALFALFA-SDSS Galaxy Catalog. *Astronomical Journal*, 160(6): 271, Dec. 2020. doi: 10.3847/1538-3881/abc018.
- G. Efstathiou, R. S. Ellis, and B. A. Peterson. Analysis of a complete galaxy redshift survey. II. The field-galaxy luminosity function. *Monthly Notices of the RAS*, 232:431–461, May 1988. doi: 10.1093/mnras/232.2.431.
- E. C. Elson, W. J. G. de Blok, and R. C. Kraan-Korteweg. The dark matter content of the blue compact dwarf NGC 2915. *Monthly Notices of the RAS*, 404(4):2061–2076, June 2010. doi: 10.1111/j.1365-2966.2010.16422.x.
- H. I. Ewen and E. M. Purcell. Observation of a Line in the Galactic Radio Spectrum: Radiation from Galactic Hydrogen at 1,420 Mc./sec. *Nature*, 168(4270):356, Sept. 1951. doi: 10.1038/168356a0.
- X. Fan, C. L. Carilli, and B. Keating. Observational Constraints on Cosmic Reionization. *Annual Review of Astron and Astrophys*, 44(1):415–462, Sept. 2006a. doi: 10.1146/annurev.astro.44.051905.092514.

- X. Fan et al. Constraining the Evolution of the Ionizing Background and the Epoch of Reionization with $z \sim 6$ Quasars. II. A Sample of 19 Quasars. *Astronomical Journal*, 132(1):117–136, July 2006b. doi: 10.1086/504836.
- F. Fraternali, G. van Moorsel, R. Sancisi, and T. Oosterloo. Deep H I Survey of the Spiral Galaxy NGC 2403. *Astronomical Journal*, 123(6):3124–3140, June 2002. doi: 10.1086/340358.
- F. Fraternali, R. Sancisi, and P. Kamphuis. A tale of two galaxies: light and mass in NGC 891 and NGC 7814. *Astronomy and Astrophysics*, 531:A64, July 2011. doi: 10.1051/0004-6361/201116634.
- A. Friedmann. Über die Krümmung des Raumes. *Zeitschrift für Physik*, 10:377–386, Jan. 1922. doi: 10.1007/BF01332580.
- S. R. Furlanetto, S. P. Oh, and F. H. Briggs. Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe. *Physics Reports*, 433(4-6):181–301, Oct. 2006. doi: 10.1016/j.physrep.2006.08.002.
- G. Gamow. Expanding Universe and the Origin of Elements. *Physical Review*, 70(7-8):572–573, Oct. 1946. doi: 10.1103/PhysRev.70.572.2.
- G. Gentile, P. Salucci, U. Klein, D. Vergani, and P. Kalberla. The cored distribution of dark matter in spiral galaxies. *Monthly Notices of the RAS*, 351(3):903–922, July 2004. doi: 10.1111/j.1365-2966.2004.07836.x.
- G. Gentile, P. Salucci, U. Klein, and G. L. Granato. NGC 3741: the dark halo profile from the most extended rotation curve. *Monthly Notices of the RAS*, 375(1):199–212, Feb. 2007. doi: 10.1111/j.1365-2966.2006.11283.x.
- R. Giovanelli et al. The Arecibo Legacy Fast ALFA Survey. I. Science Goals, Survey Design, and Strategy. *Astronomical Journal*, 130(6):2598–2612, Dec. 2005. doi: 10.1086/497431.
- A. W. Graham and C. C. Worley. Inclination- and dust-corrected galaxy parameters: bulge-to-disc ratios and size-luminosity relations. *Monthly Notices of the RAS*, 388(4):1708–1728, Aug. 2008. doi: 10.1111/j.1365-2966.2008.13506.x.
- M. Grossi et al. The Hi content of early-type galaxies from the ALFALFA survey. II. The case of low density environments. *Astronomy and Astrophysics*, 498(2):407–417, May 2009. doi: 10.1051/0004-6361/200810823.
- H. Guo, M. G. Jones, M. P. Haynes, and J. Fu. Direct Measurement of the H I-halo Mass Relation through Stacking. *Astrophysical Journal*, 894(2):92, May 2020. doi: 10.3847/1538-4357/ab886f.

- H. Guo et al. Redshift-space clustering of SDSS galaxies – luminosity dependence, halo occupation distribution, and velocity bias. *Monthly Notices of the Royal Astronomical Society*, 453(4):4368–4383, 09 2015. ISSN 0035-8711. doi: 10.1093/mnras/stv1966. URL <https://doi.org/10.1093/mnras/stv1966>.
- H. Guo et al. Constraining the H I-Halo Mass Relation from Galaxy Clustering. *Astrophysical Journal*, 846(1):61, Sept. 2017. doi: 10.3847/1538-4357/aa85e7.
- G. Hallenbeck et al. HIghMass—High H I Mass, H I-rich Galaxies at $z \sim 0$ High-resolution VLA Imaging of UGC 9037 and UGC 12506. *Astronomical Journal*, 148(4):69, Oct. 2014. doi: 10.1088/0004-6256/148/4/69.
- M. P. Haynes et al. The Arecibo Legacy Fast ALFA Survey: The $\alpha.40$ H I Source Catalog, Its Characteristics and Their Impact on the Derivation of the H I Mass Function. *Astronomical Journal*, 142(5):170, Nov. 2011. doi: 10.1088/0004-6256/142/5/170.
- M. P. Haynes et al. The Arecibo Legacy Fast ALFA Survey: The ALFALFA Extragalactic H I Source Catalog. *Astrophysical Journal*, 861(1):49, July 2018. doi: 10.3847/1538-4357/aac956.
- D. J. Heath. The growth of density perturbations in zero pressure Friedmann-Lemaître universes. *Monthly Notices of the RAS*, 179:351–358, May 1977. doi: 10.1093/mnras/179.3.351.
- P. F. Hopkins. A new class of accurate, mesh-free hydrodynamic simulation methods. *Monthly Notices of the RAS*, 450(1):53–110, June 2015. doi: 10.1093/mnras/stv195.
- S. Huang, M. P. Haynes, R. Giovanelli, and J. Brinchmann. The Arecibo Legacy Fast ALFA Survey: The Galaxy Population Detected by ALFALFA. *Astrophysical Journal*, 756(2): 113, Sept. 2012. doi: 10.1088/0004-637X/756/2/113.
- E. Hubble. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15(3):168–173, Mar. 1929. doi: 10.1073/pnas.15.3.168.
- M. G. Jones, M. P. Haynes, R. Giovanelli, and C. Moorman. The ALFALFA H I mass function: a dichotomy in the low-mass slope and a locally suppressed ‘knee’ mass. *Monthly Notices of the RAS*, 477(1):2–17, June 2018. doi: 10.1093/mnras/sty521.
- N. Kanekar, S. Sethi, and K. S. Dwarakanath. The Gas Mass of Star-forming Galaxies at $z \approx 1.3$. *Astrophysical Journal, Letters*, 818(2):L28, Feb. 2016. doi: 10.3847/2041-8205/818/2/L28.

- G. Kauffmann, S. D. M. White, and B. Guiderdoni. The formation and evolution of galaxies within merging dark matter haloes. *Monthly Notices of the RAS*, 264:201–218, Sept. 1993. doi: 10.1093/mnras/264.1.201.
- J. Kennicutt, Robert C. The Star Formation Law in Galactic Disks. *Astrophysical Journal*, 344:685, Sept. 1989. doi: 10.1086/167834.
- J. Kennicutt, Robert C. The Global Schmidt Law in Star-forming Galaxies. *Astrophysical Journal*, 498(2):541–552, May 1998. doi: 10.1086/305588.
- A. A. Kepley, E. M. Wilcots, D. A. Hunter, and T. Nordgren. A High-Resolution Study of the H I Content of Local Group Dwarf Irregular Galaxy WLM. *Astronomical Journal*, 133(5):2242–2257, May 2007. doi: 10.1086/513716.
- D. Kereš, N. Katz, M. Fardal, R. Davé, and D. H. Weinberg. Galaxies in a simulated Λ CDM Universe - I. Cold mode and hot cores. *Monthly Notices of the RAS*, 395(1):160–179, May 2009. doi: 10.1111/j.1365-2966.2009.14541.x.
- N. Khandai, T. Di Matteo, R. Croft, S. Wilkins, Y. Feng, E. Tucker, C. DeGraf, and M.-S. Liu. The MassiveBlack-II simulation: the evolution of haloes and galaxies to $z \sim 0$. *Monthly Notices of the RAS*, 450(2):1349–1374, June 2015. doi: 10.1093/mnras/stv627.
- N. Khandai et al. Detecting neutral hydrogen in emission at redshift $z \lesssim 1$. *Monthly Notices of the RAS*, 415(3):2580–2593, Aug. 2011. doi: 10.1111/j.1365-2966.2011.18881.x.
- V. A. Kilborn. The H I Jodrell All-Sky Survey (HIJASS). In A. R. Taylor, T. L. Landecker, and A. G. Willis, editors, *Seeing Through the Dust: The Detection of HI and the Exploration of the ISM in Galaxies*, volume 276 of *Astronomical Society of the Pacific Conference Series*, page 80, Dec. 2002.
- H.-S. Kim et al. The spatial distribution of neutral hydrogen as traced by low H I mass galaxies. *Monthly Notices of the Royal Astronomical Society*, 465(1):111–122, 10 2017. ISSN 0035-8711. doi: 10.1093/mnras/stw2779. URL <https://doi.org/10.1093/mnras/stw2779>.
- A. Klypin, G. Yepes, S. Gottlöber, F. Prada, and S. Heß. MultiDark simulations: the story of dark matter halo concentrations and density profiles. *Monthly Notices of the RAS*, 457(4):4340–4359, Apr. 2016. doi: 10.1093/mnras/stw248.
- A. Knebe et al. MULTIDARK-GALAXIES: data release and first results. *Monthly Notices of the RAS*, 474(4):5206–5231, Mar. 2018. doi: 10.1093/mnras/stx2662.
- R. Kuzio de Naray, S. S. McGaugh, and W. J. G. de Blok. Mass Models for Low Surface Brightness Galaxies with High-Resolution Optical Velocity Fields. *Astrophysical Journal*, 676(2):920–943, Apr. 2008. doi: 10.1086/527543.

- C. Lacey and J. Silk. Tidally Triggered Galaxy Formation. I. Evolution of the Galaxy Luminosity Function. *Astrophysical Journal*, 381:14, Nov. 1991. doi: 10.1086/170625.
- C. G. Lacey et al. A unified multiwavelength model of galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 462(4):3854–3911, 08 2016. ISSN 0035-8711. doi: 10.1093/mnras/stw1888. URL <https://doi.org/10.1093/mnras/stw1888>.
- P. Lah et al. The HI content of star-forming galaxies at $z = 0.24$. *Monthly Notices of the RAS*, 376(3):1357–1366, Apr. 2007. doi: 10.1111/j.1365-2966.2007.11540.x.
- P. Lah et al. The HI gas content of galaxies around Abell 370, a galaxy cluster at $z = 0.37$. *Monthly Notices of the RAS*, 399(3):1447–1470, Nov. 2009. doi: 10.1111/j.1365-2966.2009.15368.x.
- E. Le Floc’h et al. Infrared Luminosity Functions from the Chandra Deep Field-South: The Spitzer View on the History of Dusty Star Formation at $0 < z < 1$. *Astrophysical Journal*, 632(1):169–190, Oct. 2005. doi: 10.1086/432789.
- F. Lelli, M. Verheijen, F. Fraternali, and R. Sancisi. Dynamics of starbursting dwarf galaxies: I Zw 18. *Astronomy and Astrophysics*, 537:A72, Jan. 2012a. doi: 10.1051/0004-6361/201117867.
- F. Lelli, M. Verheijen, F. Fraternali, and R. Sancisi. Dynamics of starbursting dwarf galaxies. II. UGC 4483. *Astronomy and Astrophysics*, 544:A145, Aug. 2012b. doi: 10.1051/0004-6361/201219457.
- F. Lelli, M. Verheijen, and F. Fraternali. Dynamics of starbursting dwarf galaxies. III. A HI study of 18 nearby objects. *Astronomy and Astrophysics*, 566:A71, June 2014. doi: 10.1051/0004-6361/201322657.
- F. Lelli, S. S. McGaugh, and J. M. Schombert. SPARC: Mass Models for 175 Disk Galaxies with Spitzer Photometry and Accurate Rotation Curves. *Astronomical Journal*, 152(6):157, Dec. 2016. doi: 10.3847/0004-6256/152/6/157.
- F. Lelli, S. S. McGaugh, J. M. Schombert, and M. S. Pawlowski. One Law to Rule Them All: The Radial Acceleration Relation of Galaxies. *Astrophysical Journal*, 836(2):152, Feb. 2017. doi: 10.3847/1538-4357/836/2/152.
- G. Lemaître. Expansion of the universe, A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae. *Monthly Notices of the RAS*, 91:483–490, Mar. 1931. doi: 10.1093/mnras/91.5.483.
- J. J. Lemonias, D. Schiminovich, B. Catinella, T. M. Heckman, and S. M. Moran. The GALEX Arecibo SDSS Survey. VII. The Bivariate Neutral Hydrogen-Stellar Mass Function for Massive Galaxies. *Astrophysical Journal*, 776(2):74, Oct. 2013. doi: 10.1088/0004-637X/776/2/74.

- A. K. Leroy et al. The Star Formation Efficiency in Nearby Galaxies: Measuring Where Gas Forms Stars Effectively. *Astronomical Journal*, 136(6):2782–2845, Dec. 2008. doi: 10.1088/0004-6256/136/6/2782.
- P. Li, F. Lelli, S. S. McGaugh, N. Starkman, and J. M. Schombert. A constant characteristic volume density of dark matter haloes from SPARC rotation curve fits. *Monthly Notices of the RAS*, 482(4):5106–5124, Feb. 2019a. doi: 10.1093/mnras/sty2968.
- P. Li, F. Lelli, S. McGaugh, and J. Schombert. A Comprehensive Catalog of Dark Matter Halo Models for SPARC Galaxies. *Astrophysical Journal, Supplement*, 247(1):31, Mar. 2020. doi: 10.3847/1538-4365/ab700e.
- P. Li et al. The Halo Mass Function of Late-type Galaxies from H I Kinematics. *Astrophysical Journal, Letters*, 886(1):L11, Nov. 2019b. doi: 10.3847/2041-8213/ab53e6.
- A. Loeb and R. Barkana. The Reionization of the Universe by the First Stars and Quasars. *Annual Review of Astron and Astrophys*, 39:19–66, Jan. 2001. doi: 10.1146/annurev.astro.39.1.19.
- J. Loveday. The K-band luminosity function of nearby field galaxies. *Monthly Notices of the Royal Astronomical Society*, 312(3):557–566, 03 2000. ISSN 0035-8711. doi: 10.1046/j.1365-8711.2000.03179.x. URL <https://doi.org/10.1046/j.1365-8711.2000.03179.x>.
- P. Madau and M. Dickinson. Cosmic star-formation history. *Annual Review of Astronomy and Astrophysics*, 52(1):415–486, Aug 2014. ISSN 1545-4282. doi: 10.1146/annurev-astro-081811-125615. URL <http://dx.doi.org/10.1146/annurev-astro-081811-125615>.
- N. Maddox, K. M. Hess, D. Obreschkow, M. J. Jarvis, and S. L. Blyth. Variation of galactic cold gas reservoirs with stellar mass. *Monthly Notices of the RAS*, 447(2):1610–1617, Feb. 2015. doi: 10.1093/mnras/stu2532.
- F. Marinacci et al. First results from the IllustrisTNG simulations: radio haloes and magnetic fields. *Monthly Notices of the RAS*, 480(4):5113–5139, Nov. 2018. doi: 10.1093/mnras/sty2206.
- A. M. Martin et al. THE ARECIBO LEGACY FAST ALFA SURVEY. x. THE h i MASS FUNCTION AND Ω_{HI} FROM THE 40% ALFALFA SURVEY. *The Astrophysical Journal*, 723(2):1359–1374, oct 2010. doi: 10.1088/0004-637x/723/2/1359. URL <https://doi.org/10.1088/0004-637x/723/2/1359>.
- D. C. Martin et al. The Galaxy Evolution Explorer: A Space Ultraviolet Survey Mission. *Astrophysical Journal, Letters*, 619(1):L1–L6, Jan. 2005. doi: 10.1086/426387.

- K. L. Masters. *Galaxy flows in and around the Local Supercluster*. PhD thesis, Cornell University, New York, USA, Jan. 2005.
- T. D. Matteo et al. COLD FLOWS AND THE FIRST QUASARS. *The Astrophysical Journal*, 745(2):L29, jan 2012. doi: 10.1088/2041-8205/745/2/L29. URL <https://doi.org/10.1088/2041-8205/745/2/L29>.
- M. J. Meyer et al. The HIPASS catalogue - I. Data presentation. *Monthly Notices of the RAS*, 350(4):1195–1209, June 2004. doi: 10.1111/j.1365-2966.2004.07710.x.
- E. A. Milne. World-Structure and the Expansion of the Universe. Mit 6 Abbildungen. *Zeitschrift fuer Astrophysik*, 6:1, Jan. 1933.
- C. M. Moorman et al. The H I mass function and velocity width function of void galaxies in the Arecibo Legacy Fast ALFA Survey. *Monthly Notices of the RAS*, 444(4):3559–3570, Nov. 2014. doi: 10.1093/mnras/stu1674.
- R. Morganti et al. Neutral hydrogen in nearby elliptical and lenticular galaxies: the continuing formation of early-type galaxies. *Monthly Notices of the RAS*, 371(1):157–169, Sept. 2006. doi: 10.1111/j.1365-2966.2006.10681.x.
- J. Moustakas et al. PRIMUS: Constraints on Star Formation Quenching and Galaxy Merging, and the Evolution of the Stellar Mass Function from $z = 0-1$. *Astrophysical Journal*, 767(1):50, Apr. 2013. doi: 10.1088/0004-637X/767/1/50.
- C. A. Muller and J. H. Oort. Observation of a Line in the Galactic Radio Spectrum: The Interstellar Hydrogen Line at 1,420 Mc./sec., and an Estimate of Galactic Rotation. *Nature*, 168(4270):357–358, Sept. 1951. doi: 10.1038/168357a0.
- J. P. Naiman et al. First results from the IllustrisTNG simulations: a tale of two elements - chemical evolution of magnesium and europium. *Monthly Notices of the RAS*, 477(1):1206–1224, June 2018. doi: 10.1093/mnras/sty618.
- F. Nakamura and M. Umemura. On the Initial Mass Function of Population III Stars. *Astrophysical Journal*, 548(1):19–32, Feb. 2001. doi: 10.1086/318663.
- J. F. Navarro, C. S. Frenk, and S. D. M. White. The Structure of Cold Dark Matter Halos. *Astrophysical Journal*, 462:563, May 1996. doi: 10.1086/177173.
- M. Neeleman et al. The H I Content of the Universe Over the Past 10 GYRS. *Astrophysical Journal*, 818(2):113, Feb. 2016. doi: 10.3847/0004-637X/818/2/113.
- D. Nelson et al. First results from the IllustrisTNG simulations: the galaxy colour bimodality. *Monthly Notices of the RAS*, 475(1):624–647, Mar. 2018. doi: 10.1093/mnras/stx3040.

- E. Noordermeer, J. M. van der Hulst, R. Sancisi, R. S. Swaters, and T. S. van Albada. The mass distribution in early-type disc galaxies: declining rotation curves and correlations with optical properties. *Monthly Notices of the RAS*, 376(4):1513–1546, Apr. 2007. doi: 10.1111/j.1365-2966.2007.11533.x.
- P. Noterdaeme et al. Column density distribution and cosmological mass density of neutral gas: Sloan Digital Sky Survey-III Data Release 9. *Astronomy and Astrophysics*, 547:L1, Nov. 2012. doi: 10.1051/0004-6361/201220259.
- J. H. Oort. Note on the structure of the inner parts of the galactic system. *Bulletin Astronomical Institute of the Netherlands*, 9:193, June 1941.
- T. A. Oosterloo, R. Morganti, E. M. Sadler, T. van der Hulst, and P. Serra. Extended, regular HI structures around early-type galaxies. *Astronomy and Astrophysics*, 465(3):787–798, Apr. 2007. doi: 10.1051/0004-6361:20066384.
- H. Padmanabhan and G. Kulkarni. Constraints on the evolution of the relationship between H I mass and halo mass in the last 12 Gyr. *Monthly Notices of the RAS*, 470(1):340–349, Sept. 2017. doi: 10.1093/mnras/stx1178.
- H. Padmanabhan, A. Refregier, and A. Amara. A halo model for cosmological neutral hydrogen : abundances and clustering. *Monthly Notices of the RAS*, 469(2):2323–2334, Aug. 2017. doi: 10.1093/mnras/stx979.
- E. Papastergis, A. M. Martin, R. Giovanelli, and M. P. Haynes. The Velocity Width Function of Galaxies from the 40% ALFALFA Survey: Shedding Light on the Cold Dark Matter Overabundance Problem. *Astrophysical Journal*, 739(1):38, Sept. 2011. doi: 10.1088/0004-637X/739/1/38.
- E. Papastergis, A. Cattaneo, S. Huang, R. Giovanelli, and M. P. Haynes. A Direct Measurement of the Baryonic Mass Function of Galaxies and Implications for the Galactic Baryon Fraction. *Astrophysical Journal*, 759(2):138, Nov. 2012. doi: 10.1088/0004-637X/759/2/138.
- E. Papastergis, R. Giovanelli, M. P. Haynes, A. Rodríguez-Puebla, and M. G. Jones. The Clustering of ALFALFA Galaxies: Dependence on H I Mass, Relationship with Optical Samples, and Clues of Host Halo Properties. *Astrophysical Journal*, 776(1):43, Oct. 2013. doi: 10.1088/0004-637X/776/1/43.
- V. Parkash, M. J. I. Brown, T. H. Jarrett, and N. J. Bonne. Relationships between HI Gas Mass, Stellar Mass, and the Star Formation Rate of HICAT+WISE (H I-WISE) Galaxies. *Astrophysical Journal*, 864(1):40, Sept. 2018. doi: 10.3847/1538-4357/aad3b9.

- N. Paul, T. R. Choudhury, and A. Paranjape. Halo models of H I selected galaxies. *Monthly Notices of the Royal Astronomical Society*, 479(2):1627–1637, 06 2018. ISSN 0035-8711. doi: 10.1093/mnras/sty1539. URL <https://doi.org/10.1093/mnras/sty1539>.
- A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *Astrophysical Journal*, 142:419–421, July 1965. doi: 10.1086/148307.
- A. Pillepich et al. First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies. *Monthly Notices of the RAS*, 475(1):648–675, Mar. 2018. doi: 10.1093/mnras/stx3112.
- Planck Collaboration et al. Planck 2015 results. XIII. Cosmological parameters. *Astronomy and Astrophysics*, 594:A13, Sept. 2016. doi: 10.1051/0004-6361/201525830.
- W. H. Press and P. Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *Astrophysical Journal*, 187:425–438, Feb. 1974. doi: 10.1086/152650.
- J. X. Prochaska, S. Herbert-Fort, and A. M. Wolfe. The SDSS Damped Ly α Survey: Data Release 3. *Astrophysical Journal*, 635(1):123–142, Dec. 2005. doi: 10.1086/497287.
- A. Rahmati, A. H. Pawlik, M. Raičević, and J. Schaye. On the evolution of the H I column density distribution in cosmological simulations. *Monthly Notices of the RAS*, 430(3): 2427–2445, Apr. 2013. doi: 10.1093/mnras/stt066.
- S. Rana, S. Dutta, and N. Khandai. The Effect of HI Halo Mass Scatter on the Clustering of HI selected Galaxies. *Manuscript in preparation*, 2021.
- S. M. Rao, D. A. Turnshek, G. M. Sardane, and E. M. Monier. The statistical properties of neutral gas at $z < 1.65$ from UV measurements of Damped Lyman Alpha systems. *Monthly Notices of the RAS*, 471(3):3428–3442, Nov. 2017. doi: 10.1093/mnras/stx1787.
- D. Reed et al. Evolution of the mass function of dark matter haloes. *Monthly Notices of the RAS*, 346(2):565–572, Dec. 2003. doi: 10.1046/j.1365-2966.2003.07113.x.
- D. S. Reed, R. Bower, C. S. Frenk, A. Jenkins, and T. Theuns. The halo mass function from the dark ages through the present day. *Monthly Notices of the RAS*, 374(1):2–15, Jan. 2007. doi: 10.1111/j.1365-2966.2006.11204.x.
- J. Rhee et al. Neutral atomic hydrogen (H I) gas evolution in field galaxies at $z \simeq 0.1$ and $\simeq 0.2$. *Monthly Notices of the Royal Astronomical Society*, 435(3):2693–2706, 09 2013. ISSN 0035-8711. doi: 10.1093/mnras/stt1481. URL <https://doi.org/10.1093/mnras/stt1481>.

- J. Rhee et al. Neutral hydrogen (H I) gas content of galaxies at $z \approx 0.32$. *Monthly Notices of the RAS*, 473(2):1879–1894, Jan. 2018. doi: 10.1093/mnras/stx2461.
- E. E. Richards et al. Baryonic distributions in the dark matter halo of NGC 5005. *Monthly Notices of the RAS*, 449(4):3981–3996, June 2015. doi: 10.1093/mnras/stv568.
- M. S. Roberts. The neutral hydrogen content of late-type spiral galaxies. *Astronomical Journal*, 67:437–446, Jan. 1962. doi: 10.1086/108752.
- H. P. Robertson. Kinematics and World-Structure. *Astrophysical Journal*, 82:284, Nov. 1935. doi: 10.1086/143681.
- V. C. Rubin and J. Ford, W. Kent. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *Astrophysical Journal*, 159:379, Feb. 1970. doi: 10.1086/150317.
- V. C. Rubin, J. Ford, W. K., and N. Thonnard. Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa \rightarrow Sc. *Astrophysical Journal, Letters*, 225:L107–L111, Nov. 1978. doi: 10.1086/182804.
- G. B. Rybicki and A. P. Lightman. *FUNDAMENTALS OF RADIATIVE TRANSFER*, chapter 1, pages 1–50. John Wiley & Sons, Ltd, 1985. ISBN 9783527618170. doi: <https://doi.org/10.1002/9783527618170.ch1>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527618170.ch1>.
- K. Said, R. C. Kraan-Korteweg, and L. Staveley-Smith. The H i mass function in the Parkes H i Zone of Avoidance survey. *Monthly Notices of the Royal Astronomical Society*, 486(2):1796–1804, 04 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz956. URL <https://doi.org/10.1093/mnras/stz956>.
- A. Saintonge et al. COLD GASS, an IRAM legacy survey of molecular gas in massive galaxies - I. Relations between H₂, H I, stellar content and structural properties. *Monthly Notices of the RAS*, 415(1):32–60, July 2011. doi: 10.1111/j.1365-2966.2011.18677.x.
- A. Sandage, G. A. Tammann, and A. Yahil. The velocity field of bright nearby galaxies. I. The variation of mean absolute magnitude with redshift for galaxies in a magnitude-limited sample. *Astrophysical Journal*, 232:352–364, Sept. 1979. doi: 10.1086/157295.
- R. H. Sanders. The Published Extended Rotation Curves of Spiral Galaxies: Confrontation with Modified Dynamics. *Astrophysical Journal*, 473:117, Dec. 1996. doi: 10.1086/178131.
- R. H. Sanders and M. A. W. Verheijen. Rotation Curves of Ursa Major Galaxies in the Context of Modified Newtonian Dynamics. *Astrophysical Journal*, 503(1):97–108, Aug. 1998. doi: 10.1086/305986.

- M. Schaller et al. The EAGLE simulations of galaxy formation: the importance of the hydrodynamics scheme. *Monthly Notices of the RAS*, 454(3):2277–2291, Dec. 2015. doi: 10.1093/mnras/stv2169.
- D. Schiminovich et al. The GALEX Arcicibo SDSS Survey - II. The star formation efficiency of massive galaxies. *Monthly Notices of the RAS*, 408(2):919–934, Oct. 2010. doi: 10.1111/j.1365-2966.2010.17210.x.
- D. J. Schlegel, D. P. Finkbeiner, and M. Davis. Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. *Astrophysical Journal*, 500(2):525–553, June 1998. doi: 10.1086/305772.
- M. Schmidt. The Rate of Star Formation. *Astrophysical Journal*, 129:243, Mar. 1959. doi: 10.1086/146614.
- M. Schmidt. The Rate of Star Formation. II. The Rate of Formation of Stars of Different Mass. *Astrophysical Journal*, 137:758, Apr. 1963. doi: 10.1086/147553.
- M. Schmidt. Space Distribution and Luminosity Functions of Quasi-Stellar Radio Sources. *Astrophysical Journal*, 151:393, Feb. 1968. doi: 10.1086/149446.
- P. Serra et al. The ATLAS^{3D} project - XIII. Mass and morphology of H I in early-type galaxies as a function of environment. *Monthly Notices of the RAS*, 422(3):1835–1862, May 2012. doi: 10.1111/j.1365-2966.2012.20219.x.
- Y. Shen et al. Cross-correlation of SDSS DR7 Quasars and DR10 BOSS Galaxies: The Weak Luminosity Dependence of Quasar Clustering at $z \sim 0.5$. *Astrophysical Journal*, 778(2):98, Dec. 2013. doi: 10.1088/0004-637X/778/2/98.
- R. K. Sheth and G. Tormen. Large-scale bias and the peak background split. *Monthly Notices of the RAS*, 308(1):119–126, Sept. 1999. doi: 10.1046/j.1365-8711.1999.02692.x.
- R. K. Sheth, H. J. Mo, and G. Tormen. Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes. *Monthly Notices of the RAS*, 323(1):1–12, May 2001. doi: 10.1046/j.1365-8711.2001.04006.x.
- J. D. Simon, A. D. Bolatto, A. Leroy, and L. Blitz. High-Resolution Measurements of the Dark Matter Halo of NGC 2976: Evidence for a Shallow Density Profile. *Astrophysical Journal*, 596(2):957–981, Oct. 2003. doi: 10.1086/378200.
- K. Spekkens and R. Giovanelli. The Structure of Rapidly Rotating Late-Type Spiral Galaxies. I. Photometry, H I, and Optical Kinematics. *Astronomical Journal*, 132(4):1426–1444, Oct. 2006. doi: 10.1086/506177.

- M. Spinelli, A. Zoldan, G. De Lucia, L. Xie, and M. Viel. The atomic hydrogen content of the post-reionization Universe. *Monthly Notices of the Royal Astronomical Society*, 493(4):5434–5455, 03 2020. ISSN 0035-8711. doi: 10.1093/mnras/staa604. URL <https://doi.org/10.1093/mnras/staa604>.
- V. Springel. The cosmological simulation code GADGET-2. *Monthly Notices of the RAS*, 364(4):1105–1134, Dec. 2005. doi: 10.1111/j.1365-2966.2005.09655.x.
- V. Springel et al. First results from the IllustrisTNG simulations: matter and galaxy clustering. *Monthly Notices of the RAS*, 475(1):676–698, Mar. 2018. doi: 10.1093/mnras/stx3304.
- K. A. Suess, J. Darling, M. P. Haynes, and R. Giovanelli. Identifying OH Imposters in the ALFALFA Neutral Hydrogen Survey. *Monthly Notices of the RAS*, 459(1):220–231, June 2016. doi: 10.1093/mnras/stw666.
- R. A. Swaters, R. Sancisi, T. S. van Albada, and J. M. van der Hulst. The rotation curves shapes of late-type dwarf galaxies. *Astronomy and Astrophysics*, 493(3):871–892, Jan. 2009. doi: 10.1051/0004-6361/200810516.
- E. Tempel et al. Galaxy morphology, luminosity, and environment in the SDSS DR7. *Astronomy and Astrophysics*, 529:A53, May 2011. doi: 10.1051/0004-6361/201016196.
- C. Trachternach, W. J. G. de Blok, S. S. McGaugh, J. M. van der Hulst, and R. J. Dettmar. The baryonic Tully-Fisher relation and its implication for dark matter halos. *Astronomy and Astrophysics*, 505(2):577–587, Oct. 2009. doi: 10.1051/0004-6361/200811136.
- R. B. Tully et al. The Extragalactic Distance Database. *Astronomical Journal*, 138(2):323–331, Aug. 2009. doi: 10.1088/0004-6256/138/2/323.
- T. S. van Albada, J. N. Bahcall, K. Begeman, and R. Sancisi. Distribution of dark matter in the spiral galaxy NGC 3198. *Astrophysical Journal*, 295:305–313, Aug. 1985. doi: 10.1086/163375.
- L. van Zee, M. P. Haynes, J. J. Salzer, and A. H. Broeils. A Comparative Study of Star Formation Thresholds in Gas-Rich Low Surface Brightness Dwarf Galaxies. *Astronomical Journal*, 113:1618–1637, May 1997. doi: 10.1086/118379.
- L. Verdes-Montenegro, A. Bosma, and E. Athanassoula. The ringed, warped and isolated galaxy NGC 6015. *Astronomy and Astrophysics*, 321:754–764, May 1997.
- M. Verheijen and E. de Blok. The HSB/LSB Galaxies NGC 2403 and UGC 128. *Astrophysics and Space Science*, 269:673–674, Dec. 1999. doi: 10.1023/A:1017015229229.

- M. A. W. Verheijen. The Ursa Major Cluster of Galaxies. V. H I Rotation Curve Shapes and the Tully-Fisher Relations. *Astrophysical Journal*, 563(2):694–715, Dec. 2001. doi: 10.1086/323887.
- M. A. W. Verheijen and R. Sancisi. The Ursa Major cluster of galaxies. IV. HI synthesis observations. *Astronomy and Astrophysics*, 370:765–867, May 2001. doi: 10.1051/0004-6361:20010090.
- F. Villaescusa-Navarro et al. Ingredients for 21 cm intensity mapping. *The Astrophysical Journal*, 866(2):135, oct 2018. doi: 10.3847/1538-4357/aadba0. URL <https://doi.org/10.3847/1538-4357/aadba0>.
- M. Vogelsberger et al. Properties of galaxies reproduced by a hydrodynamic simulation. *Nature*, 509(7499):177–182, May 2014. doi: 10.1038/nature13316.
- A. G. Walker. On milne’s theory of world-structure*. *Proceedings of the London Mathematical Society*, s2-42(1):90–127, 1937. doi: <https://doi.org/10.1112/plms/s2-42.1.90>. URL <https://londmathsoc.onlinelibrary.wiley.com/doi/abs/10.1112/plms/s2-42.1.90>.
- W. Walsh, L. Staveley-Smith, and T. Oosterloo. The Giant, Gas-Rich, Low-Surface-Brightness Galaxy NGC 289. *Astronomical Journal*, 113:1591–1606, May 1997. doi: 10.1086/118377.
- S. D. M. White and C. S. Frenk. Galaxy Formation through Hierarchical Clustering. *Astrophysical Journal*, 379:52, Sept. 1991. doi: 10.1086/170483.
- J. P. Wild. The Radio-Frequency Line Spectrum of Atomic Hydrogen and its Applications in Astronomy. *Astrophysical Journal*, 115:206, Mar. 1952. doi: 10.1086/145533.
- C. N. A. Willmer. Estimating galaxy luminosity functions. *Astronomical Journal*, 114: 898–912, Sept. 1997. doi: 10.1086/118522.
- C. N. A. Willmer et al. The Deep Evolutionary Exploratory Probe 2 Galaxy Redshift Survey: The Galaxy Luminosity Function to $z \sim 1$. *Astrophysical Journal*, 647(2):853–873, Aug. 2006. doi: 10.1086/505455.
- N. Yoshida, A. Sokasian, L. Hernquist, and V. Springel. Early structure formation and reionization in a cosmological model with a running primordial power spectrum. *The Astrophysical Journal*, 598(1):73–85, nov 2003. doi: 10.1086/378852. URL <https://doi.org/10.1086/378852>.
- A. Zoldan, G. De Lucia, L. Xie, F. Fontanot, and M. Hirschmann. H i-selected galaxies in hierarchical models of galaxy formation and evolution. *Monthly Notices of the Royal Astronomical Society*, 465(2):2236–2253, 11 2016. ISSN 0035-8711. doi: 10.1093/mnras/stw2901. URL <https://doi.org/10.1093/mnras/stw2901>.

-
- M. A. Zwaan, F. H. Briggs, and D. Sprayberry. The luminosity function and surface brightness distribution of H I selected galaxies. *Monthly Notices of the RAS*, 327(4):1249–1259, Nov. 2001. doi: 10.1046/j.1365-8711.2001.04844.x.
- M. A. Zwaan, M. J. Meyer, L. Staveley-Smith, and R. L. Webster. The hipass catalogue: h i and environmental effects on the h i mass function of galaxies. *Monthly Notices of the Royal Astronomical Society: Letters*, 359(1):L30–L34, May 2005. ISSN 1745-3933. doi: 10.1111/j.1745-3933.2005.00029.x. URL <http://dx.doi.org/10.1111/j.1745-3933.2005.00029.x>.
- M. A. Zwaan, M. J. Meyer, and L. Staveley-Smith. The velocity function of gas-rich galaxies. *Monthly Notices of the RAS*, 403(4):1969–1977, Apr. 2010. doi: 10.1111/j.1365-2966.2009.16188.x.
- M. A. Zwaan et al. The 1000 Brightest HIPASS Galaxies: The H I Mass Function and Ω_{HI} . *Astronomical Journal*, 125(6):2842–2858, June 2003. doi: 10.1086/374944.
- F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6: 110–127, Jan. 1933.