NEUTRAL HYDROGEN IN THE LOCAL UNIVERSE: DISTRIBUTIONS OF GAS-RICH GALAXIES FROM ALFALFA

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Ann Marie Martin

May 2011

NEUTRAL HYDROGEN IN THE LOCAL UNIVERSE: DISTRIBUTIONS OF GAS-RICH GALAXIES FROM ALFALFA

Ann Marie Martin, Ph.D.

Cornell University 2011

We make use of the sensitivity and large sample size of the Arecibo Legacy Fast ALFA (ALFALFA) survey to investigate statistical properties of HI-selected galaxies in the local universe. This ongoing survey will eventually produce a census of over 30,000 gas-rich galaxies. The partial sample available for this work contains over 10,000 galaxies, making it twice as large as the previous generation blind HI survey, HIPASS. We have characterized the survey sensitivity and the selection function for use in the analyses presented in this dissertation, and are therefore able to make robust estimates of the characteristics of the galaxy population sampled by ALFALFA.

From the available sample, we derive the HI mass function at redshift z = 0 via both the $1/V_{max}$ and 2DSWML methods, which are in extremely close agreement with each other. These measurements are influenced by the survey design and the cosmological volume sampled, and we investigate and describe the magnitude and characteristics of the imprint of large-scale structure, survey sensitivity, and distance errors on the HI mass function. Our finding that there are more galaxies with high HI masses on the extreme end of the distribution will alter predictions for future large-scale 21 cm line surveys with such instruments as the Square Kilometer Array (SKA). The large volume and sample size of ALFALFA also allows study of the dependence of the HI mass function on galaxy environments at various scales by splitting into subsamples. This work

confirms that the distribution of HI mass does change with environment and indicates that there is some threshold density beyond which the HI reservoirs of dwarf galaxies are efficiently removed, thereby flattening the HI mass function.

The scale and degree of clustering for HI-selected galaxies is probed through the galaxy-galaxy two-point correlation function. Again, the large size of the ALFALFA galaxy catalog allows the sample to be split. The correlation function can thus be investigated as a function of HI mass, resolving a discrepancy from the HIPASS analysis, and as a function of optical luminosity and color. We find that HI-selected galaxies, and low HI mass galaxies in particular, are the most weakly clustered population of galaxies known to exist. For HI-selected galaxies, hydrogen mass more reliably distinguishes between environments on large scales than optical properties do, but on small scales these galaxies cluster similarly regardless of HI mass. The clustering scale of HI-selected galaxies at low redshift will help to disentangle the influence of environment on galaxy evolution, and to constrain the populations to which future large 21 cm line surveys will be sensitive.

The robust measurement of the statistical properties of gas-rich galaxies at low redshifts provided in this dissertation will be used to improve simulations and models that predict the characteristics of this population of galaxies. In the case of the correlation function, we show that the clustering observed in ALFALFA can distinguish between popular models of star formation that predict the present-day distribution of gas in simulated galaxies. We suggest future projects that will further our understanding of the interplay between HI and optical properties and galaxy environment, while also providing robust observables for comparison to models.

BIOGRAPHICAL SKETCH

Ann Marie Martin was born in Binghamton, New York, on May 31, 1983. She grew up in Johnson City, New York, along with her younger brother John. After graduating from high school there, she attended the University at Buffalo. While at UB, Ann studied both physics and English literature. Summer opportunities to work in Cornell's ExtraGalactic Group and at the Green Bank Telescope in West Virginia solidified her interest in studying large-scale structure through galaxy surveys in the 21 cm line. Following her graduation from UB in 2005, Ann returned to Cornell and began her graduate studies as a National Defense Science and Engineering Graduate fellow. Her time in Ithaca was mainly spent working on this dissertation, contributing to the Astronomy Grads Network, developing outreach and science communication programs, and watching *Lost*. After graduating from Cornell in the Spring of 2011, she will begin a NASA postdoctoral fellowship at Langley Research Center, helping to evaluate science education and outreach programs.

To my friends, who are the family I choose, and to my family, who are my oldest friends.

Here are we, on this starry night, staring into space.

And I must say –

I feel as small as dust lying down here.

- Dave Matthews Band

ACKNOWLEDGEMENTS

Martha and Riccardo – I will be eternally grateful for the chance you gave me to come to Cornell as a summer student in 2003, and for all of the opportunities since then. Under your guidance, I have learned, traveled, grown, and watched many fantastic films. You are wonderful teachers and mentors, and I know that I am speaking for every student who has ever passed through your offices. I especially appreciate your support in my education and outreach efforts, and I promise that I will put all of that time to good use. I don't know who or where I would be today if not for my incredible luck in working with you, and your incredible patience in working with me. I will always treasure the time I was able to spend as a member of EGG, and will miss you.

Every member of the ALFALFA collaboration – Thank you for observing, flagging, baselining, flatfielding, and extracting signals for the thousands and thousands of galaxies that I was able to use in this work. This has been a wonderful team to be a part of.

Rebecca Koopmann and Luigi Guzzo – Your guidance and advice have made a big mark on this work. Becky, I am so grateful to have your support and your friendship.

Amélie, Brian, Chris, Karen, Kristine and all other EGGs – I thank you for all of your help over the years. You all left behind some pretty big shoes to fill and I've learned so much from your scholarship and your example.

The Brinson Foundation, the National Defense Science and Engineering Graduate Fellowship, and the Cornell Graduate School offered me support in my research, my travels, and my professional life.

Nancy Schaff, Ben Ortiz, and the Public Service Center – I am leaving graduate school with a diverse set of experiences and a greater sense of community

responsibility. For that, your mentorship, your guidance, and your support, I thank you.

Betsey, Manolis, Babs, and Carl, and all of AGN – It has been great fun working (and playing!) with you. I've learned so much from you, not just about astronomy but about friendship, principles, leadership, and who I want to be in this world. I think this department is a better place because of all of you and that we are certainly leaving our mark behind us. I know I will be hearing great things about all of you and reading your names in press releases for the rest of your careers. The work you do is exciting and amazing, but the *way* you do it is equally impressive. Think of me if you ever find yourself running a project in need of an outreach guru! Betsey - the Camuy Cave would not have been the same without you. I can't wait to see what your work turns up and to hear about the incredible journey I know you are going to embark upon after you graduate. Wherever your path may lead you, don't forget to pack Little Umi. Babs - you are an incredible woman and a fabulous scientist. I wish you all of the very best and most beautiful things in life.

Sabrina – The fact that, two years down the road, I still refer to the desk hatillo sits on as "Sabrina's desk" tells you everything you need to know. When I look back on this time, I know my foremost association with my time at Cornell will be you and all the good times I had working with you. I think there is a point in every acknowledgments section I've ever seen that says something along the lines of "I couldn't have done it without you." But in my case, I really, truly, and literally could not have done it without you, and would not be here today if I had not been fortunate enough to have you as a mentor, a teacher, and most importantly a friend. Thank you for introducing me to the best canine friend I will ever have, for giving Chewie a reasonable name instead of 'Panini',

for making sure I was never gluten-poisoned on an observing trip (now that I think about it, that is nothing short of a miracle), for talking me into giving *The Soup* a chance, for noticing when Josh had a throwing-related anecdote at the ready, for taking me on the epic 5 B's Across America tour, for helping me figure out twilight flats, for exposing me to the sheer joy that is focusing the Palomar 5m, and for not burying me at the beach. Despite all of your help and all of the hours we spent trying to puzzle out the HI mass function, low-mass samples, and Palomar observing plans, there is still one thing I just cannot figure out: what IS jelly?

Mom, Dad, Johnny, Minnie, Paco and all the rest of my family – Many, many thanks are owed! I'm so glad that I was able to have you in the audience at my defense, and I am extremely fortunate to have the family I do. Mom and Dad - do you remember the night before I left for Space Camp, when we tried to pop the movie in and it had finally broken? Well, thank you for letting me watch it dozens of times, and thank you for *Apollo 13*, and *The Right Stuff*, and *From the Earth to the Moon*, and going to Roberson and the Air and Space Museum, and every other little thing that led me down this interesting and unusual path. Johnny - best of luck in your own education, and I know you are going to do great things. Minnie and Paco - you have always been my biggest cheerleaders. Thank you for checking in on me and for reminding me to write, write, write!

Sarah, Emily, Amy, Sam, and Manju – It turns out that having sisters is really awesome. Who knew? No matter how far away we are now or will be in the future, I know I can count on you and hope you know that I am always thinking of you. You were all there for me, rallying support, when I wasn't sure I would be able to make it through this process or who I was going to be afterward. And, of course, during a very tough year you made endless treks to my apartment to

keep me busy, distracted, happy, and well-fed. If there is ever anything I can do to repay your favors, love, and support, you know that I would be there in a split second. Emily - I don't know that a single sentence can do justice to our history, so instead I will say that I look forward to knowing you for the next 22 years (and beyond), and the first 22 were pretty great. Sarah - With the completion of this dissertation, I am one step closer to making the compound dream a reality. You know me and get me in ways that nobody else does, and can therefore continue with your nefarious plans to steal my identity. You know me so well that I am confident you are able to pick up on the secret shmoop undertones in this message. Sam - I'm really glad that we both had a terrible day freshman year and that I was invited to commiserate with you over Ying's pizza, or I may never have known someone who is an absolute treasure to me and has an everlasting place in my heart. Your grace and your strength are awe-inspiring. Manju - I miss you! - but I know that we are always friends (and that you are always my brown-eyed girl) even when we are too busy and too far apart. I'm so proud of you, and I look up to you and your successes with admiration. Amy - I'm especially grateful for all the time we were able to spend together after you came back to the mainland. Thanks for all of your kind words and questions, and for believing in me. I always believe in you.

Last but certainly not, by any definition, least, Josh – I don't know what I could say that would ever be adequate, but I hope you know everything I could ever write and more, and that you see your many contributions reflected in these pages. You have been an excellent partner in crime during our time in Ithaca, and I can't wait to see what's in store for us next.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The 21 cm hyperfine transition of neutral hydrogen (HI), and its distribution at various redshifts, has been called the richest data set available on the sky [Barkana and Loeb, 2005a]. The 21 cm line will become increasingly important to cosmology as arrays with large fields of view, high spatial resolution, and large bandwidth make possible the detection of cold gas in millions of galaxies, tracing the evolution of gas density and environment with redshift [Myers et al., 2009, Abdalla et al., 2010]. HI may be used to probe cosmology at all redshifts where the universe is either neutral or only partly ionized [Barkana and Loeb, 2005b], as well as at extremely low redshifts ($z \lesssim 0.1$) when the Universe has been reionized but the signatures from individual galaxy gas reservoirs are detectable.

The next generation of 21 cm instruments will be designed for cosmological parameter estimation, to trace the evolution of gas density, and to probe the epoch of reionization through a variety of statistical techniques. Intensity mapping, unlike the galaxy survey approach that has been used thus far, will probe the distribution of neutral gas with poor angular resolution corresponding to scales $\sim$10 Mpc [Peterson et al., 2009], efficiently and cheaply addressing such questions as the nature of dark energy through statistical measures like the baryon acoustic oscillation (BAO) signature. This simple statistical approach becomes even more straightforward in a 21 cm survey, since the detection of a line feature eliminates the need for costly optical followup to determine redshift, and even more feasible given recent apparent successes in the proposed intensity mapping technique [Pen et al., 2009, Chang et al., 2010]. The power

spectrum of 21 cm fluctuations, best understood by analogy to the CMB, will measure the fluctuations in the cold gas caused by gas density, temperature, neutral fraction, and Lya flux from galaxies [Barkana and Loeb, 2005a,b]. These effects are separable and will evolve with redshift, pinning down the redshift extent of the epoch of reionization and the sources that contributed to it.

1.1 21 cm Cosmology with the ALFALFA Survey

This dissertation will investigate the characteristics of the HI-rich galaxies in the local universe via a survey of the 21 cm line at low redshift. This survey, the Arecibo Legacy Fast ALFA (ALFALFA) survey, is thus both the first survey of its kind, in that its sensitivity is unprecedented, and the last survey of its kind, in that future HI surveys will not be conducted in the same way. The findings of ALFALFA will pave the way for the next generation of 21 cm cosmologists.

As a low-redshift ($z \lesssim 0.06$) survey, ALFALFA makes a key contribution to the understanding of 21 cm cosmology and how future efforts will be undertaken. ALFALFA's results at low redshift will provide constraints on the HI mass function and the density of HI in the present-day universe, which provides a baseline to study the evolution of gas as a function of redshift. Models for the evolution of gas density are otherwise not well constrained, given conflicting results that either indicate that Ω_{HI} has evolved very little since $z \sim 2$ [Prochaska and Wolfe, 2009] or that a factor of 2 increase in gas density is required by $z \sim 2$ in order to match observations of damped Lya systems [Rao et al., 2006]. This unknown evolution of gas density therefore leads to an inability to make strong predictions of the telescope sensitivity required to make

high-redshift cosmological measurements in the 21 cm line.

With increased sensitivity, higher data quality, and a significantly larger source catalog than previous blind HI surveys, ALFALFA will contribute to this and other cosmological problems by accurately and robustly measuring baseline statistical properties of the neutral gas distribution at z = 0. Additionally, ALFALFA will contribute to 21 cm cosmology by allowing detailed examination of an HI-selected sample of galaxies unbiased by their stellar content and sensitive to very low-surface-brightness systems. In addition to the large-scale properties of the universe over cosmic time, we are interested in understanding the details of galaxy formation and evolution, all of which is tied into gas content (which contributes to future star formation potential), star formation history, and environment. These cosmological 'details' are key to completing our understanding of cosmic evolution.

1.2 HI-Rich Galaxies in the Local Universe

To date, targeted surveys in the local universe have revealed details about the relationship between HI gas and other characteristics of galaxies. HI rich galaxies tend to be blue [West et al., 2010], of late type, and less clustered than their gas-poor counterparts, and the gas mass-to-luminosity ratio is a well-described function of morphological type [Roberts and Haynes, 1994]. Furthermore, the star formation rate in normal galaxies is known to relate strongly, via the Kennicutt-Schmidt law [Kennicutt, 1998] to the HI gas surface density Σ_{HI}, indicating that there may be a global cutoff for star formation.

Hindering this understanding is the fact that blind HI surveys have, to date,

been severely limited in their sample sizes. What can be learned from a large-scale survey that selects galaxies by their neutral hydrogen rather than by their optical and stellar properties?

1.2.1 Baryon and Cold Gas Fractions of Dwarf Galaxies

Extremely low-mass dwarf galaxies tend to be more dominated by gas than their higher-mass counterparts [Geha et al., 2006], and a population of low surface brightness dwarf galaxies exists that is difficult to find in optical surveys and that displays curious star formation properties [van Zee et al., 1997]. HI surveys are more sensitive to the very dwarf galaxy samples that are necessary in order to fully explain the distribution of low-mass galaxies at the present epoch. Studies such as ALFALFA are ideal for developing large, statistically complete samples of dwarf galaxies that are not biased by optical properties. West et al. [2010] found that HI selection finds nearby and dwarf galaxies with lower surface brightnesses and fainter luminosities compared to optical selection. Those authors also found that compared to an optical sample selected from the same volume, their HI-based sample selection identified galaxies with higher star formation rates, indicating that surveys like ALFALFA provide better insight to the sites of star formation in the low-redshift universe.

Dwarf galaxies are not only the most numerous class of galaxy, but also provide important constraints on our understanding of cosmology and galaxy evolution. It has been known for over a decade [Klypin et al., 1999] that the population of known low-mass galaxies is in conflict with the predictions of small dark matter halos from cosmological simulations. This so-called 'missing satel-

lites problem' reveals the tension between the limits of observations and theory. Extreme dwarf galaxies are notoriously difficult to detect in galaxy redshift surveys, but it also remains difficult to simulate the details of baryonic mass evolution, star formation, and gastrophysics in galaxies, particularly low-mass galaxies that push the boundaries of a given simulation's mass resolution [Governato et al., 2010]. The solution to the missing satellites problem will likely involve a decrease in simulated halos predicted to host galaxies at z = 0 in tandem with an increase in observed dwarf galaxy counts.

The ALFALFA survey's science case includes a search for potential 'missing satellites' that may be gas-rich but completely, or mostly, lacking stars, but thus far the number counts of ALFALFA alone have not made a dent in the problem. However, extremely low mass 'minihalos' in the vicinity of the Milky Way may still be detectable, though disguised as very compact high velocity clouds with systemic recessional velocities barely distinguishable from Galactic HI. Their extremely narrow 21 cm line profile velocity widths demand the application of new algorithms to the existing dataset [Giovanelli et al., 2010, Ricotti, 2009a].

At present, the missing satellites problem may be on the verge of solution, due to the discovery of several new 'ultrafaint' dwarf galaxies in the Sloan Digital Sky Survey, new simulations which show that photoheating during the epoch of reionization can severely deplete low-mass halos of baryons [Hoeft et al., 2006, Hoeft and Gottloeber, 2010], and renewed attention to low surface brightness dwarfs or 'stealth galaxies' [Bullock et al., 2010]. Current understanding of mechanisms that can disrupt star formation in the lowest-mass galaxies early in the Universe indicate that the difference between the dark matter and baryon mass functions can be explained through baryon processes

[Baldry et al., 2008, Martin et al., 2010].

1.2.2 The Role of Environment

Large HI surveys also probe important cosmological questions by relating the characteristics of cold gas to galaxy environment. For instance, the distribution of HI masses, quantified through the HI mass function, changes with environmental density. ALFALFA covers a broad variety of galaxy environments, from void galaxies [Saintonge et al., 2008] to medium-density groups [Stierwalt et al., 2009] to rich clusters [Giovanelli et al., 2007, Kent et al., 2008]. What is already known about the interaction of gas and environment?

HI in Clusters

The HI deficiency of galaxies in clusters, compared to their counterparts in the field, provides the foundation for our understanding of the impact of environment on gas characteristics. As the closest rich cluster, Virgo is the most well-studied high-density environment and serves as a laboratory for testing the interactions of galaxies, their cold gas, and the hot ICM. However, findings regarding HI deficiency have been confirmed in many other clusters [Solanes, 2001].

HI in Voids

Related to the missing satellites problem is the 'void problem'; while CDM simulations generally overpredict the number of dark matter halos small enough

to be dwarf galaxy hosts, they severely overpredict the number when voids are considered in particular [Tikhonov and Klypin, 2009]. We would expect the term 'void' to be, instead, a misnomer, and for these regions to host many dwarf galaxies. While the cluster environment may have been adequately addressed by targeted surveys, voids have largely been neglected due to the paucity of interesting sources in those regions. A blind survey like ALFALFA covering a void region brings new light to bear on the question of HI content in that environment. Koribalski et al. [2004] used HI-selected galaxies from the HIPASS Bright Galaxy Catalog (BGC) to more strictly define the boundaries of the Local Void, and found that HI's sensitivity to low surface brightness galaxies enhanced their ability to fully probe void environments.

While HI-selected dwarf galaxies do prefer underdense regions [Basilakos et al., 2007] they avoid the most extreme end of the distribution, that is, the true voids. This may be due to their particular susceptibility to photoheating during reionization [Hoeft et al., 2006, Hoeft and Gottloeber, 2010], an area of active research in solving the void problem. The Void Galaxy Survey (VGS) project has reported indications that void galaxies in their sample may currently be accreting cold gas along filaments [van de Weygaert et al., 2011]. Other factors may contribute, and little is known about the process of reionization, which may have occurred first in voids or first in high-density regions; such differences influence the results of simulations.

HI in Groups

Medium-density groups (both loose and compact) bridge the gap between understanding galaxies in voids and galaxies in clusters. HI has served as a tracer

of dynamics in groups, due to lower velocities within that environment that leave persistent tidal bridges, tails and remnants [Haynes et al., 1984]. Further linking environment to HI content, galaxies in more compact groups are HI deficient compared to loose groups [Oosterloo and Iovino, 1997].

In the Local Group, HI content is known to correlate strongly with group-centric distance; gas-poor dwarf spheroidals are found near massive galaxies, while the gas-rich dwarf irregulars are more widely distributed [Grebel, 2005]. While this morphological segregation inspires tantalizing questions about the role of environment, it also makes blind HI surveys particularly well suited for probing the group environment at all distances from the massive, dominant group members. Koribalski et al. [2004] also found that some new groups that were obvious in their HI content had not previously been discovered in optical surveys, likely due to the faintness of optical galaxies within them. Given the different biases of HI and optical surveys, it is likely that a combination of them would best identify groups in the nearby universe.

1.3 ALFALFA and an HI-Selected Galaxy Sample

The Arecibo Legacy Fast ALFA (ALFALFA) survey takes advantage of the sensitivity and resolution of the 305m Arecibo Observatory and its new 7 pixel ALFA receiver to conduct a census of neutral hydrogen in the local universe. The survey, nearing completion and estimated to conduct its final observations in Spring 2012, will cover nearly 7,000 square degrees of sky, and its catalogs will contain an estimated 30,000 galaxies. Published catalogs to date include Martin et al. [2009], Stierwalt et al. [2009], Kent et al. [2008], Saintonge et al. [2008], and

Giovanelli et al. [2007].

ALFALFA is sensitive to the redshift range -2,000 to 18,000 km s^{-1}, allowing the survey to probe everything from gas-rich minihalo candidates in the dark matter halo of the Milky Way to clusters at 250 Mpc. As a survey searching for a single line feature, ALFALFA galaxy detections include a redshift, without the need for costly optical followup. In part because of this advantage, ALFALFA's contribution in the anti-Virgo direction towards the Pisces-Perseus foreground void doubled the number of available galaxy redshifts. ALFALFA is sensitive to galaxies with 10^6 $M_\odot$ of HI, and has found extremely massive galaxies up to $10^{10.8}$ (6.3 x 10^{10}) $M_\odot$ of neutral hydrogen. A benchmark of ALFALFA's sensitivity is its ability to study HI sources down to 10^7 $M_\odot$ at the distance of the Virgo Cluster. The velocity resolution of the survey allows us to detect dwarf galaxies with rotational velocities of only $\sim$15 km s^{-1}.

As a blind survey, ALFALFA is sweeping its sky area without regard to targeting known optical galaxies. Its catalog thus represents a purely HI-selected sample and is not biased by optical magnitude and/or surface brightness. Roughly 70% [Kent et al., 2008, Giovanelli et al., 2007] of galaxies detected in ALFALFA had not been previously targeted for HI studies, even in regions like the Virgo Cluster that have been very well-studied. This indicates that HI selection provides a completely different view of the hydrogen in the local universe, as well as challenging the 'conventional wisdom' regarding the distribution and characteristics of gas-rich galaxies.

1.3.1 Previous HI Surveys

The first generation of blind HI surveys were plagued by small source counts, including AHISS, ADBS, and HIJASS. In the last decade, the HIPASS survey produced the first catalog with thousands of blindly selected HI sources.

The Arecibo HI Strip Survey (AHISS; Zwaan et al. [1997]) was a blind survey covering two strips of the sky visible to the Arecibo Observatory. AHISS yielded 66 extragalactic sources covering 65 deg^2, out to a redshift of 7,400 km s^{-1}. This survey provided the first measurement of the energy density of neutral hydrogen in the present day universe, $\Omega_{HI} \sim 2.0 \pm 0.05 \times 10^{-4} h^{-1}$, but only covered 3 orders of magnitude in HI mass. The Arecibo Dual-Beam Survey (ADBS; Rosenberg and Schneider [2002]) made a large improvement over this sample size, detecting 265 galaxies over 430 deg^2.

A planned all-sky survey, the HI Jodrell All Sky Survey (HIJASS; Lang et al. [2003]) using the Lovell Telescope at Jodrell Bank has published a catalog of 222 confirmed sources representing 1115 deg^2 of sky out to 10,000 km s^{-1}. The first true all-sky blind HI survey, however, was the HI Parkes All Sky Survey (HIPASS), which used a 13-beam receiver on the Parkes Telescope to fully cover the Southern sky. Their all-sky catalog [Meyer et al., 2004] included 4,315 sources, and an extension in the Northern hemisphere [Wong et al., 2006] added an additional 1,002 sources, for a total of 5,317. The HIPASS observing scheme had a sharp redshift cutoff at 12,000 km s^{-1}but a median redshift of only 3,000 km s^{-1}.

1.3.2 ALFALFA as a Cosmological Probe

As a second-generation blind survey, ALFALFA improves upon the first generation and takes advantage of a large source catalog filling a cosmologically significant volume to probe a new set of cosmological questions. In particular, this dissertation will focus on two statistical properties that relate HI rich galaxies in the local universe to the underlying cosmology. The first, the HI mass function (HIMF), determines the global distribution of gas, measuring the energy density of HI (Ω_{HI}) at the current epoch and determining the abundance of low-mass galaxies. The second statistical measure, the HI correlation function, explores the tendency of HI-rich galaxies to cluster, and measures the bias between HI-selected galaxies and the underlying dark matter mass distribution.

1.3.3 Survey Status

Four large chunks of the ALFALFA area, where source extraction and catalog production have been completed, are available for the work presented in this dissertation. In the spring (toward Virgo) portion of the survey, $07^h30^m < \alpha$ (J2000) $< 16^h30^m$, we cover the declination ranges $4^\circ < \delta$ (J2000) $< 16^\circ$ and $24^\circ < \delta$ (J2000) $< 28^\circ$. In the fall (anti-Virgo) portion of the survey, $22^h00^m < \alpha$ (J2000) $< 03^h00^m$, we cover the declination ranges $14^\circ < \delta$ (J2000) $< 16^\circ$ and $24^\circ < \delta$ (J2000) $< 32^\circ$.

This corresponds to a sky fraction of $\sim 6.7\%$, or 2700 deg^2. Because the survey extends to a redshift of 18,000 km s^{-1}, the volume, assuming a Hubble constant H$_0$ = 70 km s^{-1} Mpc^{-1}, of this fraction of the survey is 4.74×10^6 Mpc^3. In practice, specific regions of volume are eliminated from the survey. Because of

radio frequency interference (RFI) due to contaminating signals from the FAA radar at the San Juan airport, the volume between 15,000 and 16,000 km s^{-1} is inaccessible to the survey. For most statistical analyses, then, ALFALFA is considered to extend only to 15,000 km s^{-1}. Other RFI contamination makes the survey blind to cosmic emission in some fraction of the three-dimensional volume, approximately 2.5%. Therefore, the nominal volume available for this dissertation is 2.67×10^6 Mpc^3.

Published ALFALFA catalogs include sources in three different categories, labeled as Code 1, Code 2, and Code 9 sources. Code 1 sources are of high signal-to-noise (SNR $>$ 6.5) and are included as confident detections. Code 2 sources are less confident (4.5 $<$ SNR $<$ 6.5) but coincide with a known optical galaxy with a previously-measured redshift that matches the ALFALFA measurement. For most statistical analyses, Code 2 sources are excluded, and this is discussed in detail in Chapter 4. Code 9 sources are high velocity clouds in the vicinity of the Milky Way and are not extragalactic detections. Code 9 sources may be subcomponents of large HVC complexes and may be accompanied by a second measurement of the full complex.

The current catalog includes 11,942 Code 1 sources, 3,101 Code 2 sources, and 812 Code 9 sources. When both Code 1 and Code 2 sources are included, this catalog contains 431 low-mass HI dwarfs ($M_{HI} < 10^8\ M_\odot$) and 83 extremely high-mass galaxies ($M_{HI} > 10^{10.5}\ M_\odot$). When only Code 1 sources are included, these figures change to 366 and 81, respectively. Compared to the HIPASS source counts (44 and 25), this sample reflects the superior sensitivity of the Arecibo Observatory and the ALFALFA observing scheme.

1.4 The HI Mass Function

A major goal of this dissertation is to explore the distribution of HI gas in the local universe through the HI mass function (HIMF). Primarily, this statistic probes the relationship between cold gas mass, baryonic mass, and cold dark matter mass. The Press-Schechter formalism [Press and Schechter, 1974] predicts that dark matter halos follow a mass function giving a low-mass slope $\alpha \sim$ -1.8, but both stellar and HI mass functions are distinctly shallower, around -1.2 to -1.3. Various processes have been invoked to explain the discrepancy, such as photoevaporation of gas from dwarf galaxies during reionization and the suppression of star formation, but it is unclear to what extent these various processes are responsible for the observed lack of dwarf galaxies at z=0 and to what extent each functions to remove cold gas and/or to quench star formation.

1.4.1 HIMF Evolution

Measuring the HI mass function also determines the density of cold gas, Ω_{HI}. While Ω_{HI} contributes a negligible amount to the overall baryon budget, roughly 5×10^{-4} [Fukugita and Peebles, 2004], its importance in understanding galaxy evolution is clear. One open question in galaxy evolution is how Ω_{HI} is related to the global star formation density, and how that relationship evolves with redshift. To date, blind HI surveys have only been possible out to the ALFALFA redshift limit z = 0.06. The ALFA Ultra Deep Survey (AUDS) will detect galaxies out to 0.16 [Freudling et al., 2008, 2011]. Thus, surveys of HI in galaxies can probe the evolution of Ω_{HI} to (cosmologically) insignificant redshifts.

On the high-redshift side, Ω_{HI} is constrained by studies of damped Lyman α absorption (DLA) systems. However, the observations are in conflict and either are consistent with constant co-moving gas density [Prochaska and Wolfe, 2009] or suggest evolution over $0 < z < 2$ [Noterdaeme et al., 2009]. This evolution is modest enough, regardless, to be inconsistent with the change in the star formation rate over the same period. Over the same time period when Ω_{HI} appears to be stagnant in co-moving gas density, the star formation rate appears to undergo strong evolution. The global star formation rate is well-established out to z=1 [Hopkins, 2004] with observations out to z = 8 [Kistler et al., 2009] beginning to constrain the full history of galaxies. Between redshifts of 0 and 1, the global star formation rate rises steeply (by an order of magnitude). How can this be reconciled with the non-evolution of cold gas over that same period? It is clear that the relationship between HI and star formation is more complex than current theories and simulations are able to explain. The ALFALFA measurement of the HI mass function will provide a robust value at z = 0 for comparison with other findings.

Understanding the evolution of HI in galaxies is also important in order to make better predictions of what future surveys can expect to find. Predictions of the utility and design requirements of future instruments such as the Square Kilometer Array [Rawlings et al., 2004, Abdalla et al., 2010] depend strongly on models of the HI mass function evolution. These are very poorly constrained by observations at high redshift, but until recently even the z = 0 HI mass function was not robustly determined.

1.4.2 Environmental Dependence of the HIMF

Given the unknown intricacies of the relationship between HI and star formation, it is important to understand how the HI mass function varies as a function of environment. It has been well-established that gas rich spirals tend to be found in the field, while gas-poor ellipticals are found in cluster environments. Until recently, however, little has been understood about how the shape of the HIMF itself changes in those environments. HIPASS [Zwaan et al., 2005] found a significant difference between galaxies in low density environments (with flat faint-end slopes) and in high density environments (with steeper faint-end slopes), and further determined that this difference is exacerbated when environmental differences are defined on larger scales.

By contrast, other studies have found that flatter faint-end slopes occur in high-density environments. Springob et al. [2005] found a low-significance difference between the HIMF in high and low density environments, with the HIMF faint-end slope α significantly flattened and M_* shifted to lower characteristic masses in the denser environments. Studies of Virgo [Briggs and Rao, 1993, Davies et al., 2004] and Ursa Major [Verheijen et al., 2001], each with approximately 3 dozen galaxies, have measured flattened low-mass HIMF slopes in the cluster environment. This flattening has been attributed to HI deficiency in denser environments [Gavazzi et al., 2005]. Both Pisano et al. [2007] and Freeland et al. [2009] have found a flat HIMF in surveys of 6 and 5 Local Group analogs, respectively.

Why are detailed studies of rich environments claiming flattened HIMFs while blind HI surveys report steepening in those environments? Zwaan et al. [2005] acknowledge the discrepancy between their finding and previous stud-

ies of the HIMF in the cluster environment, but point out that their blind HI survey is biased against the highest-density environments; in an HI-selected study, 'high' density has a different meaning than in an optically-selected sample. In the highest densities, it's possible that ram-pressure stripping and other gas-depleting processes completely eliminate low-mass galaxies, flattening the faint-end slope, while medium density environments may shift originally high-mass objects into lower mass bins, boosting the faint-end slope. Those authors also suggest that the gas-consumption ratio between low-mass and high-mass galaxies is greater in low-density environments than in high density environments, which could flatten the HIMF in low-density regions.

This discrepancy is one of the major issues to be explored in this dissertation, with unprecedented precision due to the volume, sensitivity and sample size of ALFALFA.

1.5 The Correlation Function ξ(r)

While the HI mass function can be used to explore how the HI properties of galaxies change with environment, we can also understand the typical environment of HI galaxies and how that changes with the HI properties of those galaxies through the correlation function ξ(r). The two-point galaxy-galaxy correlation function ξ measures the scales over which galaxy populations cluster in three-dimensional space, and also measures the degree to which various populations cluster.

Galaxy clustering is known to depend strongly on optical properties. Morphology is the most obvious of these, with gas-rich spirals tending to avoid

cluster centers, but clustered galaxies are also known to be redder, more massive, more concentrated, and more luminous [Zehavi et al., 2005, Li et al., 2006a, Swanson et al., 2008] than their counterparts in the field. Since these properties are also correlated with HI mass, we expect HI-selected galaxies to be less clustered and they are, in reality, the least clustered class of galaxies.

1.5.1 Galaxy Bias

Cold dark matter simulations predict a distribution of dark matter halos, inside of which galaxies form and evolve. Because baryons undergo processes like ram pressure stripping, star formation, supernova feedback, and photoevaporation, it is expected that galaxies will generally, but not perfectly, trace the underlying mass distribution. We are therefore faced with the challenge of using imperfect tracers to understand our universe and its cosmic evolution. One part of the solution is to better understand the difference between the mass we observe and the underlying dark matter; this difference in galaxy density fluctuations compared to dark matter density fluctuations is known as the galaxy bias. It is parametrized by a bias parameter, b, via δ_{galaxy} = b δ_{matter}. The correlation function of galaxies depends on the square of the bias parameter, with ξ_{galaxy}(r) = b^2 ξ_{matter}(r).

Because HI galaxies avoid clusters, they are expected to be strongly biased with respect to the CDM. ALFALFA can robustly measure the correlation function of galaxies in the local Universe that are HI-rich, and can determine the bias compared to dark matter. This is an interesting question in its own right, but the clustering and bias of HI-selected galaxies has practical applications in the

coming decade of cosmology with the 21 cm line. Large galaxy surveys with the next generation radio telescopes, such as the proposed Square Kilometer Array (SKA), propose to detect enormous numbers of galaxies through the 21 cm line [Myers et al., 2009]. Their understanding of the universe, its structure, and its gas characteristics as a function of redshift will be based on a strongly biased sample. As a specific example, the clustering of galaxies as a function of redshift will be used as a tracer of cosmic structure formation, but it is unclear to what degree HI-rich galaxies at various redshifts will reliably trace the overall structure. To some extent, that bias, particularly at high redshift, will require constraints from simulations rather than observations. However, a firm understanding of the galaxy bias at z=0 is a necessary component, in particular because it provides the best observational constraint for the development and improvement of such simulations.

1.5.2 The Clustering of Gas-Rich Galaxies

Despite the importance of the correlation function of HI-selected galaxies, little work has been done in this field. Meyer et al. [2007] presented their analysis of ξ for the HIPASS sample, and Basilakos et al. [2007] provided a second look external to the HIPASS collaboration. Generally, the results are in agreement with expectations and with each other. HI galaxies were found to cluster weakly relative to optical galaxies, with a correlation length r_0 of ~ 3.5 h^{-1} Mpc and a slope γ $\sim$1.5. Meyer et al. [2007] reported the correlation function of HIPASS galaxies to indicate the weakest clustering of any published sample of galaxies.

These two works disagreed, however, regarding the dependence of cluster-

ing on HI mass. Meyer et al. [2007] found no statistically significant difference between high and low mass galaxy clustering scales, while Basilakos et al. [2007] found that high HI mass galaxies clustered more strongly than low mass galaxies, and found the low-mass galaxies to be the weakest clustered population studied. ALFALFA's large sample size and volume will allow this survey to settle these questions.

The ALFALFA sample will measure the correlation function of HI rich galaxies over a large, cosmologically significant volume for the first time. Another major goal of this dissertation is to explore how the correlation function depends on galaxy properties, including HI mass and optical luminosity, in order to better understand how HI interacts with galaxy environments.

1.6 Overview of Dissertation

This dissertation takes advantage of the increased sensitivity and large sample size of ALFALFA to make more robust statistical claims about the nature of gas-rich galaxies in the local Universe. The ALFALFA survey is able to break new ground in this area compared to the first-generation blind HI surveys, with the data available for this work more than doubling what was previously available.

In Chapter 2 we present a catalog of galaxies in the anti-Virgo region of the ALFALFA survey, covering $22^h00^m < \alpha$ (J2000) $< 03^h00^m$, $24^\circ < \delta$ (J2000) $< 26^\circ$.

The focus then turns to the distribution of HI masses in the ALFALFA survey and its application to near-field cosmology. Chapter 3 presents the global HI mass function derived from the ALFALFA survey at roughly 40% completion,

and Chapter 4 evaluates the impact on the HI mass function of the survey's sensitivity, mass uncertainties in the local Universe, and the large scale structure in the ALFALFA volume. Chapter 5 quantifies the dependence of the HI mass function on galaxy environment on various scales.

By exploring both the HI mass function and the clustering characteristics of the ALFALFA sample of HI-selected galaxies, we can begin to understand the relative weight of nature vs. nurture in the evolution of gas-bearing galaxies. Chapter 6 presents the two-point (galaxy-galaxy) correlation function from the ALFALFA survey at 40% completion, along with further discussion of the dependence of HI-selected galaxy clustering on such properties as mass and luminosity in Chapter 7.

We summarize our findings and suggest future directions for work with HI selected galaxies in the Conclusion (Chapter 8).

The Appendices include the full catalog of galaxies described in Chapter 2, provide details on the methodology used in the HIMF analysis of Chapter 3, and finally describe a technique for reducing images from the Vatican Advanced Technology Telescope (VATT) in preparation for nebular abundance measurements on the Palomar 5m telescope.

CHAPTER 2

HI SOURCE CATALOG OF THE ANTI-VIRGO REGION AT δ = +25°

2.1 Introduction

First-generation blind HI surveys conducted in recent years (HI Parkes All Sky Survey, HIPASS: Barnes et al. [2001]; Arecibo Dual-Beam Survey, ADBS: Rosenberg and Schneider [2000]) have uncovered populations of HI-rich galaxies. Such surveys have made attempts at measuring the distribution of neutral hydrogen in the local Universe through such statistical measures as the HI mass function [Zwaan et al., 2005, Rosenberg and Schneider, 2002] and the galaxy-galaxy correlation function [Meyer et al., 2007] of HI-selected objects. However, it is necessary for second-generation surveys to develop a larger sample that covers a cosmologically significant volume and a variety of environments in order to truly probe the statistical nature of this distribution without bias. Furthermore, the first-generation attempts suffer from small sample size, especially for low-mass objects, and low median sample redshifts. The Arecibo Legacy Fast ALFA Survey (ALFALFA), by contrast, with larger bandwidth, superior sensitivity, and high spatial resolution, is expected to make a significant contribution to the local HI census.

ALFALFA is an ongoing project at the Arecibo Observatory 305m telescope, taking advantage of the seven-beam ALFA receiver to conduct a large-scale

*This chapter is published in Martin et al. (2009)

blind survey of neutral hydrogen in the local Universe. The use of HI as a cosmological probe depends critically upon blind surveys, in order to prevent bias toward highly luminous objects. As a second-generation project, ALFALFA proposes to improve upon the results of HIPASS and ADBS and pave the way for large galaxy redshift surveys using the 21 cm line. As the first survey of this kind to probe a cosmologically significant volume, approximately 2.7×10^7 Mpc^3, ALFALFA will provide a complete picture of the distribution of hydrogen in nearby galaxies. Additionally, the large sample size will allow studies of subsamples to investigate environmental influence; by design, ALFALFA will include galaxies from a range of environments, from the Pisces-Perseus foreground void [Saintonge et al., 2008], to medium-density groups like Leo [Stierwalt et al., 2009], to the turbulent high-density Virgo Cluster [Giovanelli et al., 2007, Kent et al., 2008].

ALFALFA will make neutral hydrogen measurements of $> 25,000$ galaxies [Giovanelli et al., 2005b]. Tracing the distribution of HI out to $z \sim 0.06$, ALFALFA samples have a median redshift of about 7,000 km s^{-1}, compared to the median redshift of HIPASS at 3,000 km s^{-1} [Meyer et al., 2004]. Furthermore, ALFALFA probes deeper into the mass distribution, sensitive to neutral hydrogen masses of $M_{HI} \sim 2 \times 10^6 M_\odot$ out to the distance of the Virgo Cluster, ∼16 Mpc. We expect to detect on the order of hundreds of objects with HI masses less than $10^{7.5} M_\odot$, a population of galaxies that has never before been adequately sampled. ALFALFA will thus provide a unique dataset for HI cosmology, including the measurement of the HI mass function (HIMF), especially at the low-mass end, and the correlation function of HI-selected galaxies at $z \sim 0$. Additionally, the ALFALFA footprint has significant overlap with galaxy samples from SDSS, 2MASS, and GALEX. Combining an HI survey with these datasets

provides a more complete view of the relationship between star formation, environment, and gas. The small beam size and high sensitivity of the Arecibo Telescope, as compared with the Parkes Telescope used for HIPASS, reduces centroiding error, so that HI detections can be matched and cross-catalogs produced with great confidence in order to take full advantage of publicly available datasets at other wavelengths.

When complete, the ALFALFA survey will cover 7000 deg^2 of high galactic latitude sky out to $cz_\odot \sim 18,000$ km s^{-1}. This area is divided into two regions: a "spring" sky as viewed from Arecibo, covering $07^h30^m < R.A. < 16^h30^m$ and $0^\circ < \delta < +36^\circ$, and a "fall" sky covering the same region in declination and $22^h00^m < R.A. < 03^h00^m$. The spring sky includes the Virgo Cluster overdensity [Giovanelli et al., 2007, Kent et al., 2008]; here, we report on a 2°-thick slice of the fall sky, described in Section 2.

This work is the fourth catalog of ALFALFA sources published since observations commenced in February 2005. Previous catalogs [Giovanelli et al., 2007, Kent et al., 2008, Saintonge et al., 2008], and a fifth catalog paper currently submitted to the AJ [Stierwalt et al., 2009], have contained 706, 439, 564, and 546 extragalactic HI detections in regions covering 132, 135, 132 and 118 deg^2, respectively. Including the catalog presented here, ALFALFA has covered 653 deg^2, $\sim 9\%$ of its total area, and recovered 2706 HI-rich galaxies, or roughly 4.1 sources deg^{-2}. For comparison, HIPASS detected 5317 extragalactic sources over 29,000 deg^2 [Meyer et al., 2004, Wong et al., 2006] for an average of ~ 0.2 sources deg^{-2}.

The remainder of this paper is organized as follows. In Section 2, we provide an overview of the ALFALFA data collection methods, including our observing

strategy and data reduction processes. A catalog of 541 HI detections at $\delta = +25^\circ$ is presented in Section 3. In Section 4, we discuss several objects of note, followed by a discussion of the catalog's statistical properties and a summary of findings in Section 5. Throughout, a value for the Hubble constant of 70 km s^{-1} Mpc^{-1} has been assumed.

2.2 Data

The catalog presented here covers 136 deg^2 of contiguous area in a 2°-wide strip of the ALFALFA sky centered on $\delta = +25^\circ$. As part of ALFALFA's "fall sky" described in Section 1, the region stretches from 22^h00^m to 03^h00^m in right ascension. This region includes two key features: a portion of the Pisces-Perseus Supercluster (PPS), including the southernmost region of the prominent main ridge of the cluster, and a portion of a void in the foreground of the PPS (first reported in Haynes and Giovanelli [1986b]) around $cz \lesssim 2500$ km s^{-1}.

ALFALFA observations are conducted at the Arecibo telescope, using the 7-pixel ALFA receiver installed in 2004. The observing strategy is minimally invasive, employing a fixed-azimuth drift scan mode. As a result of this strategy, many nights of observing time are required to fill in the full declination range. Additionally, because two passes are made over every point in the survey, a given region of sky will typically not be completed until many months after it is first observed. Since this second pass occurs at a different point in the Earth's orbit around the Sun, and Doppler tracking is not employed, this strategy more clearly reveals spurious signals resulting from radio frequency interference (RFI) while confirming true cosmic sources. ALFA's beams are el-

liptical in shape, resulting in an angular resolution of the survey of $3.3' \times 3.8'$. For a galaxy at a distance of 10 Mpc, this corresponds to a physical scale of 10 kpc; for a galaxy at the very edge of the ALFALFA volume, at 250 Mpc, this corresponds to a physical scale of 250 kpc. We cover a bandwidth of 100 MHz (-1,600 km s^{-1} to 18,000 km s^{-1}) with a spectral resolution, before spectral smoothing, of 5.5 km s^{-1}.

Once the data is acquired for a given region, individual 600-second drift scans are combined into three-dimensional data cubes known as "grids." These cubes cover 2.4° in both R.A. and declination, with 1 arcmin sampling, and are designed to overlap so that sources on the edges of grids are easily recoverable. The sources for this catalog were obtained from 38 such grids. Each grid is also broken down into four "subgrids" along the spectral direction, again with substantial overlap. Each of these subgrids covers 1024 channels along the spectral axis, respectively covering the ranges -2,000 to 3,300 km s^{-1}, 2,500 to 7,900 km s^{-1}, 7,200 to 12,800 km s^{-1}, and 12,100 to 17,900 km s^{-1}. The grids also provide useful information on the positions of continuum sources. By comparing the positions measured by ALFALFA to those given in the NVSS catalog [Condon et al., 1998], we are able to account for Arecibo's systematic pointing errors. This process results in excellent position reliability for the sources extracted from each grid. The processing undergone by each grid is discussed in detail in Giovanelli et al. [2007] and Saintonge [2007a].

Source detection and extraction uses a combination of automated techniques, which reliably identify a set of possible signals, and individual assessment and measurement of confirmed sources by the authors. Candidate detections are identified using a matched-filtering protocol with galaxy profile tem-

plates built from combinations of Hermite functions [Saintonge, 2007a]. Each of the resulting candidates is examined individually, and for those selected for inclusion in the final catalog, a set of source parameters is interactively measured. These parameters are presented in Section 3. Each source is also compared with optical images from DSS2 and cross-correlated with NED and the AGC ("Arecibo General Catalog," a private database of galaxies maintained by M.P.H. and R.G.). Some Sloan Digital Sky Survey coverage is now available for this region, following the seventh data release of SDSS imaging and spectroscopy (DR7). The SDSS data was used to confirm DSS2 optical information in a few ambiguous cases. While most galaxies in this catalog have a high signal-to-noise ratio and can be confidently included on the basis of their HI profile alone, 21 of these sources are labeled as "Code 2" objects (see Section 3), which have slightly lower S/N (typically $5.0 < S/N < 6.5$) but are well-corroborated by previous redshift measurements. Each galaxy is confidently matched with an optical counterpart with only one exception (see Section 4). These matches are mostly unambiguous, and the positions of the HI centroid and optical center match very well, although the pointing offsets are dependent on S/N. For all detections in the catalog, the median pointing offset between the measured HI coordinates and the assigned optical counterpart, is 18 arcsec, with the error reduced to 14 arcsec for detections with $S/N > 12$. The absolute HI positions themselves are calculated using a fit from continuum maps [Kent et al., 2008]; continuum sources detected in ALFALFA are compared to VLA radio continuum sources (NVSS; Condon et al. [1998]) to correct Arecibo pointings, thus removing this source of systematic error from ALFALFA coordinates. The typical RMS for all sources is ~ 2.3 mJy.

Our point source mass sensitivity as a function of distance, and related scal-

ing relations, are detailed in Giovanelli et al. [2005b]; for a particular signal–to–noise ratio S/N our mass sensitivity is:

$$\frac{M_{HI}}{M_\odot} = 2.356 \times 10^5 \, D_{Mpc}^2 \, F_c = 2.356 \times 10^5 \, D_{Mpc}^2 \, \frac{W_{50}}{w_{smo}^{1/2}} \, \frac{(S/N) \, \sigma_{rms}}{1000} \qquad (2.1)$$

where F_c is the integrated flux density in Jy km s^{-1}, W_{50} is the velocity width of the line profile at the 50% level, w_{smo} is a smoothing width (either $W_{50}/20$ for $W_{50} < 400$ km s^{-1} or 20 for $W_{50} \geq 400$ km s^{-1}), and σ_{rms} is the r.m.s. noise figure measured in mJy at 10 km s^{-1} resolution. These parameters are more thoroughly discussed in Section 3, as is the relationship between integrated flux and mass sensitivity.

2.3 Catalog Presentation

We present in Appendix A the measured parameters for 541 detections, 451 of which are associated with extragalactic objects, while the remaining 90 are HVC features. The contents of the different columns are:

- Col. 1: an entry number for this catalog
- Col. 2: the source number in the Arecibo General Catalog, a private database of extragalactic objects maintained by M.P.H. and R.G. AGC numbers, along with all other parameters, will be made available on our public digital archive site[1] and are listed on the NASA Extragalactic Database[2](NED). Therefore, the AGC, while a private database, provides

[1]http://arecibo.tc.cornell.edu/hiarchive/alfalfa/
[2]http://nedwww.ipac.caltech.edu/

a single, unique identifier for these HI sources and their optical counterparts.

- Cols. 3 & 4: center (J2000) of the HI source, after correction for systematic telescope pointing errors, which are on the order of 20 arcsec and depend on declination. The accuracy of the HI positions, compared to the positions of the optical counterparts, depends on source strength. On average, the positional accuracy, estimated as the difference between the HI position and the optical counterpart, is about 18 arcsec.

- Cols. 5 & 6: center (J2000) of the optical counterpart matched with the source. No optical counterpart is listed for High Velocity Clouds. The position has been checked for each listed object and assessed using tools provided through the *SkyView* website, in addition to NED and the AGC. The quality of centroids is estimated to be 2 arcsec or better. One extragalactic object has no identified optical counterpart. In the case of several other objects with a listed optical counterpart, a note explains an ambiguity, as alerted by an asterisk in Col. 14.

- Col. 7: heliocentric velocity of the HI source, $cz_{\odot}$, measured as the midpoint between the channels at which the flux density drops to 50% of each of the two peaks (or of one, if only one is present) at each side of the spectral feature. Units are km s^{-1}. The error on $cz_{\odot}$ to be adopted is half the error on the width, tabulated in Col. 8.

- Col. 8: velocity width of the source line profile, W_{50}, measured at the 50% level of each of the two peaks, as described for Col. 7. This value is corrected for instrumental broadening. No corrections due to turbulent motions, disk inclination or cosmological effects are applied. In parentheses we show the estimated error on the velocity width, ϵ_w, in km s^{-1}. This

error is the sum in quadrature of two components: the first is a statistical error; the second is a systematic error associated with the subjective guess with which the observer estimates the spectral boundaries of the feature, flagged during the interactive assessment of candidate detections. In the majority of cases, the systematic error is significantly smaller than the statistical error; thus the former is ignored.

- Col. 9: integrated flux density of the source, F_c, in Jy km s^{-1}. This is measured on the integrated spectrum, obtained by spatially integrating the source image over a solid angle of at least 7 arcmin $\times$7 arcmin and dividing by the sum of the survey beam values over the same set of image pixels (see Shostak and Allen [1980]). Integrated fluxes for very extended sources with large spatial asymmetries can be misestimated by our algorithm, which is not optimized for that category of sources. A special catalog with parameters of extended sources will be produced after completion of the survey.

  The issue is especially severe for extended High Velocity Clouds that exceed in size that of the ALFALFA data cubes. In these specific cases, concentrations of emission are identified and the flux in these knots of emission are measured in the same way as extragalactic sources and included as separate entries in Table 1. In general, this meant applying the same kind of S/N selection threshold as for the extragalactic signals, with the exception of the southern extension of Wright's cloud [Wright, 1979]. This cloud extends into the region covered by Saintonge et al. [2008] and was discussed in that previous ALFALFA data release. The cloud has a significant velocity gradient over its full extent, from about -350 km s^{-1}/ to -475 km s^{-1}, and is thought to potentially be associated with M33. Here,

we have made a bulk measurement of the region of the cloud extending South of +26° in declination, in addition to separate measurements of a selection of the brightest knots.

Generally, our measurements of HVCs are likely to be underestimates of the total size and flux, since we are not very sensitive to diffuse emission that may connect filaments and fill the most extended clouds. Detailed study of extended HI features and HVCs will therefore take place in a future paper. See Column 14 and the corresponding comments for individual objects.

- Col. 10: signal–to–noise ratio S/N of the detection, estimated as

$$S/N = \left(\frac{1000F_c}{W_{50}}\right)\frac{w_{smo}^{1/2}}{\sigma_{rms}} \tag{2.2}$$

where F_c is the integrated flux density in Jy km s^{-1}, as listed in Col. 9; the ratio $1000F_c/W_{50}$ is the mean flux across the feature in mJy; w_{smo} is either $W_{50}/(2\times 10)$ for $W_{50} < 400$ km s^{-1} or $400/(2\times 10) = 20$ for $W_{50} \geq 400$ km s^{-1}(w_{smo} is a smoothing width expressed as the number of spectral resolution bins of 10 km s^{-1} bridging half of the signal width); and σ_{rms} is the r.m.s noise figure across the spectrum measured in mJy at 10 km s^{-1} resolution, as tabulated in Col. 11.

- Col. 11: noise figure of the spatially integrated spectral profile, σ_{rms}, in mJy. The noise figure as tabulated is the r.m.s. as measured over the signal– and RFI-free portions of the spectrum, after Hanning smoothing to a spectral resolution of 10 km s^{-1}. The regions of the spectrum affected by RFI are identified, based on visual inspection by a member of the ALFALFA collaboration, in an early stage of data processing. These identifications are then tracked through the data processing pipeline, allowing

affected spectral channels to be excluded from the calculation of the noise figure, as discussed in Giovanelli et al. [2007]. Prominent sources of RFI include the Federal Aviation Administration radar operating at 1350 MHz from the San Juan Airport.

- Col. 12: adopted distance in Mpc, D_{Mpc}. For objects with $cz_{\odot} > 6000$ km s^{-1}, the distance is simply estimated as $cz_{cmb}/H_{\circ}$; cz_{cmb} is the recession velocity measured in the Cosmic Microwave Background reference frame [Lineweaver et al., 1996] and $H_{\circ}$ is the Hubble constant, for which we use a value of 70 km s^{-1}Mpc^{-1}. For objects of lower cz_{cmb}, we use the local Universe peculiar velocity model of Masters [2005], which is based on data from the SFI++ catalog of galaxies [Springob et al., 2007]. The Masters [2005] peculiar velocity model results from analysis of the peculiar motions of galaxies, groups, and clusters, using a combination of primary distances from the literature and secondary distances from the Tully-Fisher relation. The resulting model includes two attractors, with infall onto the Virgo Cluster and the Hydra-Centaurus Supercluster, as well as a quadrupole and a dipole component. The transition from one distance estimation method to the other is selected to be at $cz_{\odot} = 6000$ km s^{-1} because the uncertainties in each method become comparable at that point. In cases where a galaxy has a known primary distance, that distance will be adopted; when the galaxy is a known member of a group [Springob et al., 2007], that group's recessional velocity cz_{cmb} is used to determine the distance estimate according to the prescription just described.

- Col. 13: logarithm in base 10 of the HI mass, in solar units. That parameter is obtained by using the expression $M_{HI} = 2.356 \times 10^5 D^2_{Mpc} F_c$.

- Col. 14: object code, defined as follows:

Code 1 refers to sources of high S/N and general qualities that make it a reliable detection. These signals exhibit a good match between the two independent polarizations observed by ALFALFA, a spatial extent consistent with the telescope beam, a spectral profile clean of RFI features, and an approximate S/N threshold of 6.5. These criteria lead to the exclusion of some candidate detections with $S/N > 6.5$; likewise, some features with S/N slightly below this soft threshold are included, due to optimal overall characteristics of the feature, such as well-defined spatial extent, broad velocity width, and obvious association with an optical counterpart. We estimate that the detections with code 1 will be confirmed with follow-up observations in greater than 95% of cases [Saintonge, 2007a].

Code 2 refers to the sources we call 'priors.' These are sources of low S/N (< 6.5), which would ordinarily not be considered reliable detections by the criteria set for Code 1, but have been matched with optical counterparts with known optical redshifts which corroborate that measured in the HI line. We include them in our catalog because they are very likely to be real. There are 21 such sources in the present catalog.

Code 9 refers to objects assumed to be HVCs; no estimate of their distances is made.

Notes flag. An asterisk in this column indicates a comment is included for this source in the text below.

Only the first few entries of Table 1 are listed in the printed version of this paper. The full content of Table 1 is available in Appendix A and is accessible through the electronic version of the paper. The catalog will also be made

available also through our public digital archive site[3]. The comments for those sources marked with an asterisk in column 14 are given in Appendix A.

2.4 Objects of Note

The catalog contains two objects worthy of some special attention. We report one non-HVC detection that cannot be confidently matched with an optical counterpart, as well as a peculiar High Velocity Cloud candidate detected at $+84$ km s^{-1}. Footnotes were made for each of these objects.

The first of these, catalog object number 449 (HI021617.2+252616; AGC 122913), has no discernible optical emission in publicly available images. Its spectrum is shown in the top panel of Figure 2.1; the dashed line represents the measured heliocentric velocity. Point-source HI emission is detected at 9,768 km s^{-1} with a width of 73 km s^{-1}. The object's profile is very well defined, and its S/N is 10.8. There is a good match between the two polarizations recorded in ALFALFA, and we are very confident that the detection is real. Nevertheless, there are no known galaxies within 2 arcmin, no identifiable optical emission to the depth of the DSS2 Blue images (this source is outside the SDSS footprint), and no continuum source or IRAS source counterpart, although the estimated B-band extinction in this area is ~ 0.2 magnitudes [Schlegel et al., 1998, Burstein and Heiles, 1982]. This source is a prime candidate for follow-up. Assuming that it is, indeed, an extragalactic HI source at this redshift, we estimate object 449 to be at a distance of 136 Mpc with an HI mass $\log(M_{HI}) = 9.56 M_{\odot}$. Given the velocity width of the profile and this HI mass, we would expect the optical

[3]http://arecibo.tc.cornell.edu/hiarchive/alfalfa/

images to show a face-on spiral system. Despite the lack of counterpart, there is no compelling evidence of the 1665.401/1667.358 MHz doublet of an OH megamaser. ALFALFA expects to detect several dozen OH megamasers [Giovanelli et al., 2005b], which may explain some of the detections without optical counterparts; followup on candidate objects will be conducted by J. Darling (U. of Colorado).

Object number 362 in the catalog (HI011032.7+250559; AGC113843) is identified as an HVC, but it is found at a velocity of $+84$ km s^{-1}while the HVCs in this region are more typically found at heliocentric velocities near -400 km s^{-1}. Its spectrum is displayed in the bottom panel of Figure 2.1, again with the dashed line representing the measured heliocentric velocity of the source. The strong source near 0 km s^{-1} is the HI emission of the Galaxy. Detected with a S/N of 11.2, and with no potential optical counterpart visible in DSS2 Blue or SDSS DR7, this source's identification as an HVC is further evidenced by its spatial extent of ~ 10 arcmin, though its narrow velocity width of 17 km s^{-1} is on the border of ALFALFA's detection limit. There is significant polarization squint, so we caution that this signal may be spurious and followup will be necessary to determine the true nature of this source.

2.5 Statistical Properties of the Catalog

We first compare our catalog to previous HI and optical measurements. A portion of this region was covered in the Northern HIPASS catalog [Wong et al., 2006], which added an extension in the range $+2^\circ < \delta < +25^\circ 30'$. HIPASS finds only 6 sources in this partial region of overlap. A more fair comparison can

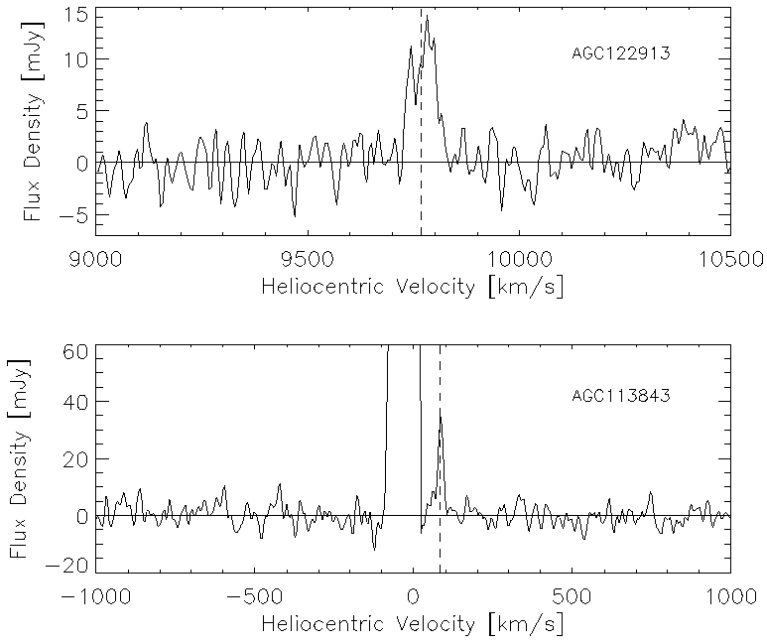

Figure 2.1 Top panel: The spectrum for AGC122913 (object 449 in the catalog, HI021617.2+252616). The dashed line represents the heliocentric velocity measured for this source, 9,768 km s^{-1}. The source, which is not aligned with any discernible optical emission in publicly available images, has a velocity width of 73 km s^{-1} and a S/N of 10.8, so we are very confident that the detection is real. Followup work may better determine its nature. Bottom panel: The spectrum for AGC113843 (object 362 in the catalog, HI011032.7+250559). The dashed line represents the heliocentric velocity measured for this source, 84 km s^{-1}. This object is identified as an HVC but its velocity is not consistent with other HVCs in the region. The source is extremely compact and has a narrow velocity width of 17 km s^{-1}, placing it on the border of ALFALFA's ability to detect it. Substantial polarization squint suggests that the source may be spurious, and followup observations will be necessary.

be found in Giovanelli et al. [2007], since the northern HIPASS extension fully surveyed the area of the Virgo cluster. While ALFALFA found a total of 730 HI detections in that region, HIPASS reports only 40. Of the 451 detections that correspond to extragalactic objects, 296 (65%) are new HI detections and 266 (58%) are altogether new redshifts. This result, which is similar to that found by Giovanelli et al. [2007] and Kent et al. [2008] in the highly-targeted Virgo Clus-

ter region, demonstrates the failure of the "conventional wisdom" of targeted HI surveys. Blind surveys are key to uncovering the true distribution of HI in the local Universe, and the sensitivity improvements of this second-generation survey further reveals the distribution of HI-rich galaxies.

Of the 541 objects here, 90 (16%) are High Velocity Clouds, 430 (80%) are Code 1 extragalactic detections, and 21 (4%) are Code 2 extragalactic detections. The median distance of extragalactic sources in the catalog is ∼114 Mpc, with a median heliocentric recessional velocity of ∼7,300 km s^{-1} for all sources and ∼8,300 km s^{-1} for all extragalactic sources. The High Velocity Clouds in this region possess highly negative velocities, ranging from -460 km s^{-1} to a single, unusual positive-velocity cloud candidate found at 84 km s^{-1}; -350 km s^{-1} is the median. The velocity width of clouds measured here is typically 25 km s^{-1}. Beyond the High Velocity Clouds, for which no optical match is made, there is one source with no optical counterpart listed.

Relative to the overall ALFALFA sample, this catalog contains a large number of high velocity clouds, totaling 16% of detections. This includes some large filaments of high velocity gas, as well as compact and isolated knots. Paper V [Saintonge et al., 2008], which surveys a region adjacent to this one, contains a similar proportion, with 49 of 488 objects (10%) identified as high velocity clouds of as bright knots of emission within a larger cloud complex. The anti-Virgo region was previously known to contain many HVCs, such as Wright's Cloud [Wright, 1979]. Braun and Thilker [2004] completed a study of gas in this area, and found high velocity emission covering a total of 29% of an 1800 deg^2 region centered on $(\alpha, \delta) = (10^\circ, 35^\circ)$ (J2000 coordinates). They identified ∼100 peaks of HVC emission with $cz_\odot$ ∼0 to -350 km/s. Their map of the regions ex-

plored here agree with our finding that there is a great deal of gas and structure at high velocities. By contrast, in the Virgo and Leo regions [Giovanelli et al., 2007, Kent et al., 2008, Stierwalt et al., 2009], the proportion of HVCs is smaller than 5% with a total of only $\sim$40 HVCs. A more thorough examination of the HVCs detected by ALFALFA and their distribution across the sky and in velocity space will be presented in a future paper. In particular, large and bright cloud complexes may require special data reduction, and compact, low-velocity width clouds may require a different source detection scheme.

In Figure 2.2, we show the sky distribution of the catalog, in three different velocity bins, revealing hints of large scale structure. The foreground underdensity discussed in Saintonge et al. [2008] can be seen in the central panel, which includes sources with $0 < cz_{\odot} < 3,000$ km s^{-1}, although there are some sources which appear at the low-declination end of the surveyed region. The source detection rate for this catalog is 3.3 sources per square degree. This is lower than the rate reported for previous ALFALFA catalogs in the spring sky [Giovanelli et al., 2007, Kent et al., 2008, Stierwalt et al., 2009] since these papers cover the Virgo cluster and Leo region, and thus we expect an overdensity in those catalogs. The bottom panel of Figure 2.2 shows the distribution of the High Velocity Clouds.

In Figure 2.3 we present the redshift distribution of the extragalactic sources as a cone diagram. In the top panel, we show the distribution for this catalog; the bottom panel displays the previously known optical redshifts for the same region of sky. ALFALFA's inability to observe sources near 15,000 km s^{-1} is evident in the top panel of this diagram. Note that the decrement of sources at high velocity is due both to this broad source of RFI around 15,000 km s^{-1}

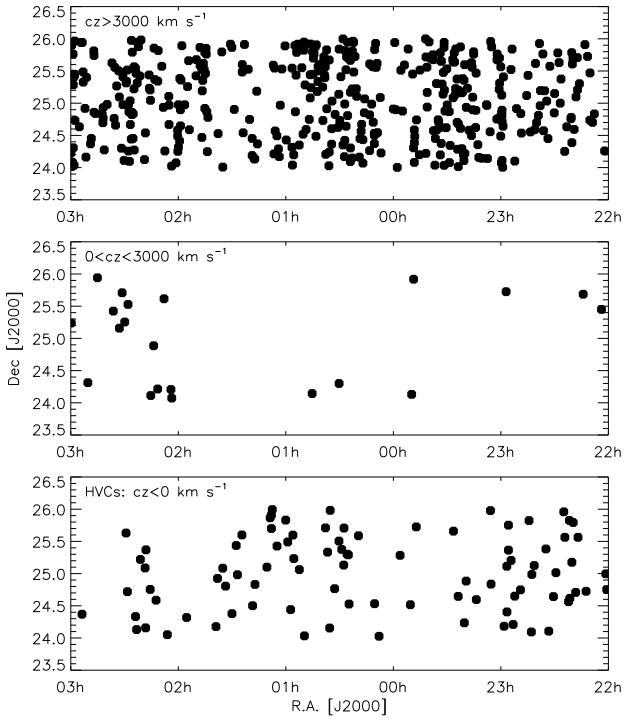

Figure 2.2 Distribution of all the sources on the sky, in three different recessional velocity bins. The areas $\delta < 24^{\circ}$ and $\delta > 26^{\circ}$ are outside the catalog limits. The bottom panel represents the High Velocity Clouds.

that makes ALFALFA blind to cosmic emission there, as well as to decreased sensitivity to sources with distance. While there is a clear relationship between the structure probed through previous work in the optical and the ALFALFA detections, it is also clear that ALFALFA's detections in this region provide a substantial new dataset of redshift information.

Figs. 2.2 and 2.3 also provide insight to the prominent local void in the foreground of the Pisces-Perseus Supercluster (PPS). Saintonge et al. [2008] report no detections in ALFALFA or in previous optical surveys between $cz \sim 1,000$ and $cz \sim 2,500$ km s^{-1} in a 2° region centered on $\delta = +27^{\circ}$ in the range from

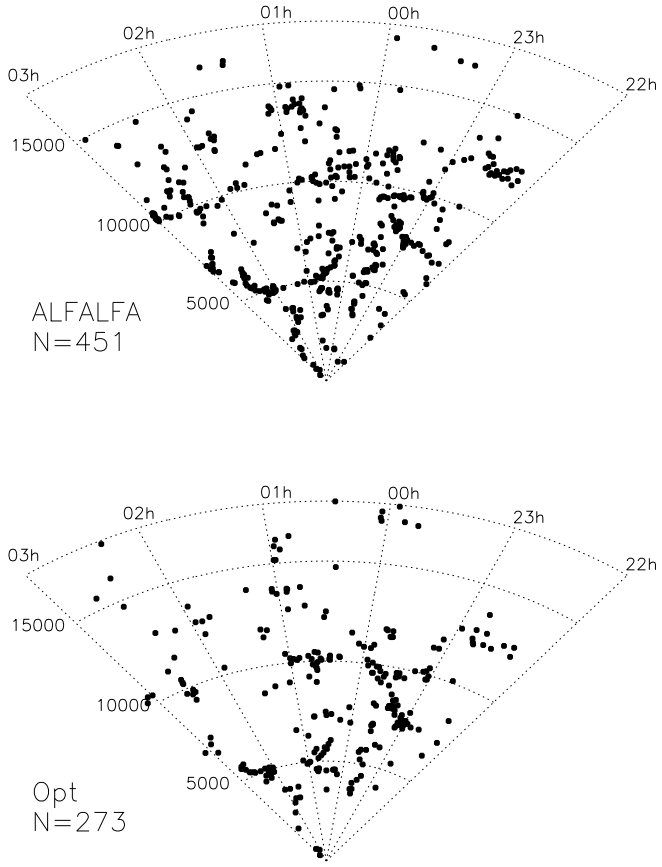

Figure 2.3 Distribution of all the sources in the 2° declination strip. The top panel shows all the ALFALFA detections presented here, while the bottom panel represents the galaxies with measured optical redshifts in the same volume of space. In each case, the number of plotted galaxies is reported. Note that due to RFI, ALFALFA is blind to cosmic emission between about 15,000 and 16,000 km s^{-1}.

22^h00^m to 02^h00^m. In the analogous region here, we do find several detections in ALFALFA, although the previous optical detections in the bottom panel of Figure 2.3 agree with the findings of Saintonge et al. [2008]. The void in the foreground of the PPS was reported by Haynes and Giovanelli [1986b] (see their Figure 2), and while it is most prominent at higher declinations ($+30^\circ < \delta < +50^\circ$), the underdensity also stretches into $+20^\circ < \delta < +30^\circ$ and the regions reported

in this paper and in Saintonge et al. [2008].

Figure 2.4 shows histograms displaying the main statistical properties of the catalog, including extragalactic detections of type 1 and 2. From top to bottom, the histograms represent the heliocentric velocity, velocity width, flux integral, S/N, and HI mass distributions. Figure 2.4a reiterates the few signals found near 15,000 km s^{-1}, as well as the large number of sources found in this catalog beyond the HIPASS velocity limit of 12,700 km s^{-1} [Meyer et al., 2004]. As shown in the last panel, Figure 2.4e, there are 8 sources with $M_{HI} < 10^8 M_\odot$, but all have masses $M_{HI} > 10^{7.5} M_\odot$. This result is expected, since we find few sources very nearby, where ALFALFA is most sensitive to the lowest mass HI clouds.

Including this work, published ALFALFA catalogs to date contain 2,706 extragalactic HI detections over 653 deg^2, or 9% of ALFALFA's total survey area. Other ALFALFA catalogs include the Virgo Cluster region [Giovanelli et al., 2007, Kent et al., 2008], the anti-Virgo region at a declination of 27° [Saintonge et al., 2008], and the Leo region [Stierwalt et al., 2009]. A series of histograms representing the total ALFALFA survey thus far are shown in Figure 2.5. This '20% ALFALFA Survey' sample includes this work as well as previous publications in the Virgo region, Leo region, and anti-Virgo region at $\delta = +27°$ [Giovanelli et al., 2007, Saintonge et al., 2008, Kent et al., 2008, Stierwalt et al., 2009] and reduced data to be published in future data releases. Figure 2.5 thus represents all ALFALFA sources detected within the region of the sky having $22\mathrm{h} < \alpha < 03\mathrm{h}$ and $24° < \delta < 28°$ in addition to $07\mathrm{h}30\mathrm{m} < \alpha < 16\mathrm{h}30\mathrm{m}$ and $4° < \delta < 16°$. This combined catalog of published and soon-to-be published sources includes 7,166 extragalactic objects coded as type 1, and 1,462 coded as

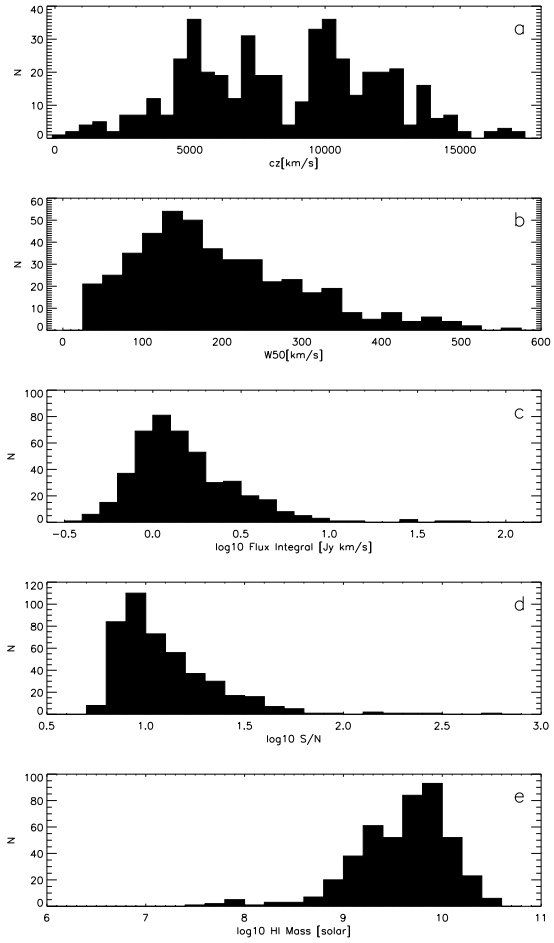

Figure 2.4 Histograms of the HI detections with Code 1 and 2 (i.e. excluding High Velocity Clouds): (a) heliocentric recession velocity in km s^{-1}; (b) HI line width at half power (W_{50}) in km s^{-1}; (c) logarithm of the flux integral in Jy km s^{-1}; (d) logarithm of the signal-to-noise ratio; and (e) logarithm of the HI mass in solar units.

type 2. Soon, the total publicly available ALFALFA catalog will surpass HIPASS in source counts.

Generally speaking, the sample presented in this work is consistent with the ALFALFA catalog as a whole, as summarized by Figure 2.5. The velocity distribution in Figure 2.5a has smoothed out much of the large scale structure, but in particular the Virgo Cluster at $cz_{\odot} \leq 3000$ km s^{-1} remains an obvious

feature. Once again, the lower density of sources at large velocities is caused by a combination of our reduced sensitivity as a function of distance, as well as sources of radio frequency interference such as the FAA radar, which essentially leaves us blind to emission at 15,000 km s^{-1}. ALFALFA's sensitivity to sources at the Virgo Cluster's distance, where we can detect HI masses down to $10^7 M_\odot$, also extends the HI mass distribution (Figure 2.5e) to lower masses than the complementary histogram for this catalog alone (Figure 2.4e). Overall, the 20% ALFALFA Survey includes ∼300 sources with neutral hydrogen masses below $10^8 M_\odot$. The paucity of sources at low heliocentric velocity (and therefore small distances) in the +25° catalog reduces the number of the lowest-mass sources detected, but as a whole ALFALFA has clearly demonstrated an ability to probe the distribution below $10^8 M_\odot$ in the local Universe.

A comparison of Figure 2.4d with Figure 2.5d shows that the overall ALFALFA sample includes more sources with lower S/N than the catalog in this work. This can be explained by the difference in objects coded as 1 and those coded as 2, since objects with Code 2 may have a lower signal-to-noise ratio as long as they have a corroborating redshift from optical data. In the Virgo and Leo regions, which have been well-studied and are covered by the Sloan Digital Sky Survey, there are therefore more objects with lower S/N and code 2, while the catalogs in the fall sky (this work and Saintonge et al. [2008]) are mainly composed of objects coded as type 1.

With this catalog and its predecessors, ALFALFA continues to demonstrate its utility and success as a second-generation blind HI survey. With coverage now totaling roughly 9% of the final survey area and including ∼2,700 extragalactic HI detections and many High Velocity Clouds, ALFALFA will soon out-

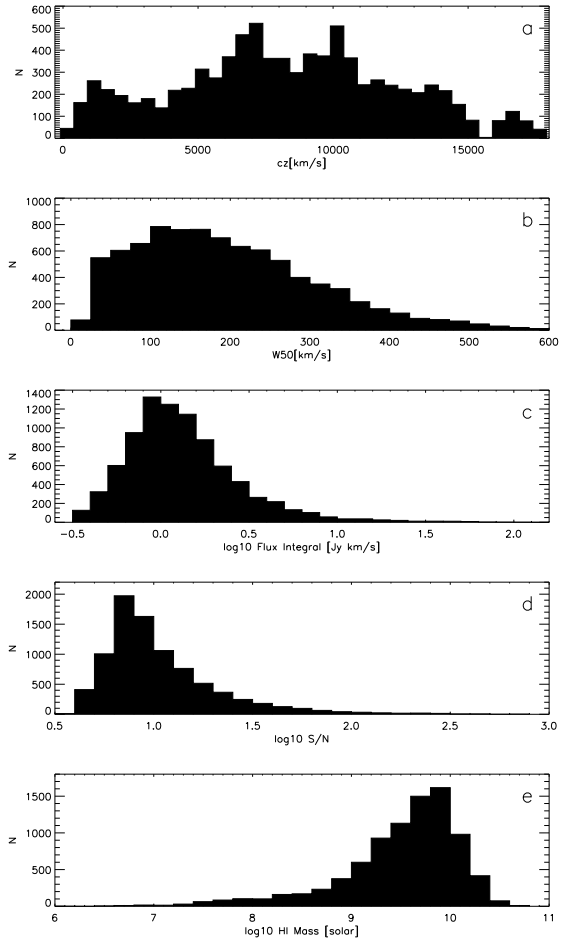

Figure 2.5 Histograms of the total ALFALFA sample to date, including this work as well as previous publications in the Virgo region, Leo region, and anti-Virgo region at $\delta = +27°$ [Giovanelli et al., 2007, Saintonge et al., 2008, Kent et al., 2008, Stierwalt et al., 2009] and reduced data to be published in future data releases. This total sample includes all ALFALFA sources detected within the region of the sky having $22\mathrm{h} < \alpha < 03\mathrm{h}$ and $24° < \delta < 28°$ in addition to $07\mathrm{h}30\mathrm{m} < \alpha < 16\mathrm{h}30\mathrm{m}$ and $4° < \delta < 16°$. Plot includes HI detections with code 1 and 2 (i.e. excluding High Velocity Clouds): (a) heliocentric recession velocity in km s^{-1}; (b) HI line width at half power (W_{50}) in km s^{-1}; (c) logarithm of the flux integral in Jy km s^{-1}; (d) logarithm of the signal-to-noise ratio; and (e) logarithm of the HI mass in solar units.

pace the first-generation efforts of HIPASS. These results will contribute to our understanding of the distribution of neutral hydrogen in the local Universe, the mass function, and the dependence on environment of galaxy parameters. For

the first time, a cosmologically significant sample of HI sources, unbiased with respect to the stellar populations of host galaxies, will soon be available.

CHAPTER 3

THE HI MASS FUNCTION AND Ω_{HI} FROM THE 40% ALFALFA SURVEY

3.1 Introduction

The disagreement between predictions of the number of low-mass dark matter halos and the observations of low-luminosity dwarf galaxies, commonly characterized as the 'missing satellite problem,' is reflected in the faint-end slopes of galaxy luminosity functions and neutral hydrogen (HI) mass functions. Current dark matter simulations and models [Boylan-Kolchin et al., 2009, Jenkins et al., 2001] imply that the faint-end slope of the underlying mass function is $\alpha \sim -1.8$, in agreement with the Press-Schechter analysis of cosmic structure formation [Press and Schechter, 1974], but observational evidence is consistent with a significantly shallower slope.

There is hope of resolving this discrepancy by investigating physical effects on the observed baryons that would not influence the underlying dark matter distribution. For example, photoheating by the UV background can deplete baryons from low mass halos, reducing the number of luminous galaxies observable today. There appears to be a characteristic halo mass, below which severe baryon depletion could eliminate the abundance of dwarf galaxies [Hoeft et al., 2008]; Hoeft and Gottloeber [2010] find this halo mass to be $\approx 6 \times 10^9 h^{-1}$ $M_\odot$, and that it is robust against assumed UV background flux density and simulation resolution effects. Other processes related to star formation, such as

*This chapter is published in Martin et al. (2010)

supernova feedback [Efstathiou, 2000] can remove gas from galaxies, preferentially removing baryons from those early galaxies residing in weak potential wells. Understanding these baryonic processes has the potential to resolve the missing satellite problem [Simon and Geha, 2007], but it remains difficult to fully simulate baryons in forming and evolving galaxies [Governato et al., 2007, Mayer et al., 2008, Ceverino and Klypin, 2009, Gnedin et al., 2009], and it is therefore important to develop other observational constraints.

Since low-mass dark matter halos are the most likely to suffer from baryon depletion, these effects may cause the shallow faint-end slopes observed in luminosity, circular velocity [Zwaan et al., 2009], and HI mass functions (HIMFs). Detailed study of these influences in the lowest-mass galaxies are only possible very nearby, and the dwarf galaxies in the Local Group have been shown to have great diversity in their star formation histories and metallicities [Tolstoy et al., 2009, Grebel and Gallagher, 2004], with some galaxies losing gas and ceasing star formation early while others have undergone this process only recently. Recently, Ricotti [2009a] has suggested that these halos may be able to re-accrete cold gas at late times, and proposes that the gas-bearing ultrafaint dwarf Leo T [Irwin et al., 2007, Ryan-Weber et al., 2008] may be an example of this process. Such galaxies may then be observable in HI line surveys like the ALFALFA survey [Giovanelli et al., 2010].

Blind HI surveys are ideal for probing these questions surrounding the lowest-baryon systems. HI line surveys are unbiased by properties like optical surface brightness, and ALFALFA in particular is designed to detect systems with lower HI masses than the blind surveys of the previous generation, down to $\sim 3 \times 10^7\ M_\odot$ at the distance of the Virgo cluster with SNR ~ 6.5 [Giovanelli

et al., 2007]. Since neutral gas fractions become large for dwarf galaxies, dominating the stellar mass, HI surveys are efficient at finding the extremely low-baryon-mass systems locally [Schombert et al., 2001, Geha et al., 2006], and the HIMF is a better measure of baryon content at the lowest masses. Furthermore, environment is well known to have an impact on gas reservoirs, with galaxies in clusters tending to be HI deficient compared to those in the field [Haynes et al., 1984]. The results of this bias as seen in the ALFALFA survey catalogs and in HI mass functions of various environments may provide insights to the relationship between HI gas densities, tidal and ram pressure stripping, and star formation.

Surveys like ALFALFA which probe a cosmologically fair sample also provide a wealth of information on the rare galaxies at the highest masses. High-mass gas-rich galaxies constrain the cosmic density of neutral gas in the local universe, Ω_{HI}. HI contributes only about 1% of the baryon budget at z=0 [Prochaska and Tumlinson, 2009, Fukugita and Peebles, 2004, Fukugita et al., 1998]. The HI mass function is necessary to estimate this with great precision in order to trace the evolution of the neutral gas fraction, measured through damped Lyα systems at higher redshifts.

HI surveys also have the advantage of combining a galaxy detection, a redshift, and a mass estimate in a single observation without followup. This is particularly important given that about 70% of galaxies in the blind ALFALFA catalog are new HI detections and many are altogether new redshifts, indicating that the conventional wisdom guiding targeted HI surveys toward galaxies expected to contain large reservoirs was severely limited. Finally, as simulations and semianalytic models of warm and cold gas in evolving galaxies improve,

the HIMF can be used as a test of these results, as done in Obreschkow et al. [2009] through a comparison of modeled cold hydrogen gas in Millennium Run galaxies to the Zwaan et al. [2005] mass function (see Section 3.6.3).

The first generation of blind HI surveys resulting in a measurement of the local HIMF contained few galaxies: Henning et al. [2000] detected 110 galaxies in the Southern Zone of Avoidance, and the Arecibo Dual Beam Survey (ADBS) HIMF was based on a sample of 265 galaxies [Rosenberg and Schneider, 2002]. Both found a faint-end slope $\alpha \sim -1.5$, significantly steeper than what is found in other larger blind HI surveys. The published HIPASS HIMFs were based on more galaxies than previous blind surveys; the function extracted from the 1000 brightest detections [Zwaan et al., 2003] had a faint-end slope -1.3 and the later paper, with a fuller catalog of 4315 sources [Zwaan et al., 2005], found -1.37. At the low-mass end of the HIMF, there is clearly severe disagreement, and previous data did not include enough low-mass objects to robustly constrain masses $< 10^8\ M_\odot$. Springob et al. [2005] investigated a complete sample of 2771 optically-selected galaxies and found a shallow slope, $\alpha \sim -1.24$. Improving the number of sources by, for example, increasing the area of a shallow survey is not enough, on its own, to resolve the issue; rather, increasing the volume over which low-mass sources are detectable has the largest impact. Distance uncertainties are largest nearby, so a shallower survey will tend to base its low-mass slope on more uncertain objects [Masters et al., 2004].

The ALFALFA survey catalogs, including those previously published [Giovanelli et al., 2007, Saintonge et al., 2008, Kent et al., 2008, Stierwalt et al., 2009, Martin et al., 2009] and those about to be published [Haynes et al. 2010, in prep], now represent $\sim 40\%$ of the final survey area, and the HI mass function

presented here considers a sample of $\sim 10,000$ HI-selected galaxies. In the following section, we will discuss the ALFALFA dataset (Section 3.2). In Section 3.3.3 and Section 3.3.4, respectively, we describe the $1/V_{max}$ method of estimating the HIMF from corrected galaxy counts, and the two-dimensional Stepwise Maximum Likelihood (2DSWML) method. Details of these methods are discussed in Appendices B.1 and B.2. After presenting the results of the global measurement of the HIMF along with Ω_{HI} in Section 3.4 and 3.5, we will discuss the results as compared to the expectations of dark matter simulations and those including cold gas, addressing the divergence between HIMF slopes and that predicted by the Press-Schechter formalism (Section 3.6).

3.2 ALFALFA Dataset

3.2.1 The ALFALFA Survey

The ongoing ALFALFA survey takes advantage of the new multipixel ALFA receiver at the Arecibo Observatory. When complete, the survey will have measured $> 30,000$ galaxies in the 21 cm line out to $z \sim 0.06$ with a median redshift of $\sim$ 8,000 km s^{-1}. The survey is more sensitive than HIPASS, with a 5σ detection limit of 0.72 Jy km s^{-1} for a source with profile width 200 km s^{-1} in ALFALFA compared to a 5σ sensitivity 5.6 Jy km s^{-1} for the same source in HIPASS [Giovanelli et al., 2005a]. Narrow profile widths, down to $\sim$ 15 km s^{-1}, allow us to probe extremely small objects. ALFALFA detects objects with neutral hydrogen masses $M_{HI} \sim 3 \times 10^7\ M_\odot$ out to the distance of the Virgo cluster. In addition to greater sensitivity, ALFALFA probes gas-rich galaxies in the local

universe with greater velocity resolution (11 km s^{-1}after Hanning smoothing vs. 18 km s^{-1}) and a deeper limiting redshift (18,000 km s^{-1} vs. 12,700 km s^{-1}) than HIPASS. Our significantly improved survey depth for low-mass objects allows the ALFALFA survey to better constrain the low-mass slope of the HI mass function.

ALFALFA survey data are acquired in a minimally-invasive drift scanning mode, in two passes ideally separated by several months, and individual 600 second drift scans are combined into three-dimensional data grids covering 2.4° in both right ascension and declination; it therefore takes many nights of observations to complete a grid from which extragalactic sources can be extracted.

Confidently detected sources are assigned one of three object codes, where Code 1 refers to a reliable extragalactic detection with a high S/N (> 6.5), Code 2 refers to extragalactic sources with marginal S/N ($4.5 < S/N < 6.5$) confirmed by an optical counterpart with known optical redshift matching the HI measurement, and Code 9 refers to High Velocity Clouds (HVCs). For this analysis, we consider only objects designated Code 1, since we are interested in extragalactic objects with well-known selection criteria. Code 1 objects have a reliable S/N, a good match between the two polarizations that are independently observed by ALFALFA, a clean spectral profile and, in almost every case, a confident match with an optical counterpart. The signal detection pipeline, discussed at length in Saintonge [2007a], combines a matched-filtering technique for identifying source candidates with an interactive process for source confirmation and parameter measurement. This technique is estimated to result in a reliability of candidate detections $\sim 95\%$ for Code 1 objects, with a completeness better than 90% for the narrowest galaxies above the prescribed S/N threshold. The

subsample of Code 1 objects provides a robust sample for the HIMF.

3.2.2 Derived Parameters

Published ALFALFA catalogs contain a set of measured parameters (including coordinates, heliocentric velocity, line profile velocity width W_{50} measured at the 50% level of two profile peaks, integrated flux density S_{int}, S/N, and noise figure σ_{rms}) in addition to a distance estimate and a derived HI mass in solar units, obtained from the expression $M_{HI} = 2.356 \times 10^5 D^2_{Mpc} S_{int}$. Our distance estimates are subject to errors due to each galaxy's unknown peculiar velocity, which translate into mass errors. The fractional distance error due to peculiar velocity decreases with increasing distance (the so-called 'Eddington effect'), so the lowest-mass galaxies which are only found nearby are most prone to this error, our treatment of which is discussed in detail in Section 3.3.2.

3.2.3 Profile Width-Dependent Sensitivity

ALFALFA's ability to detect a signal depends not only on the integrated flux, but also on the profile width W_{50} (km s^{-1}). Figure 3.1 displays the distribution of sources detected by ALFALFA. Rather than a single flux limit, the ALFALFA detection threshold is dependent on both S_{int} and profile width W_{50}, and we find that this relationship changes above $W_{50} \sim 400$ km s^{-1}. We fit the $S_{int,th}$ relationship empirically to the data, rather than using the assumed expression above. Due to differences in the two methods we employ to calculate the HIMF, we consider two different threshold cuts, described separately in Section 3.3.3

and 3.3.4.

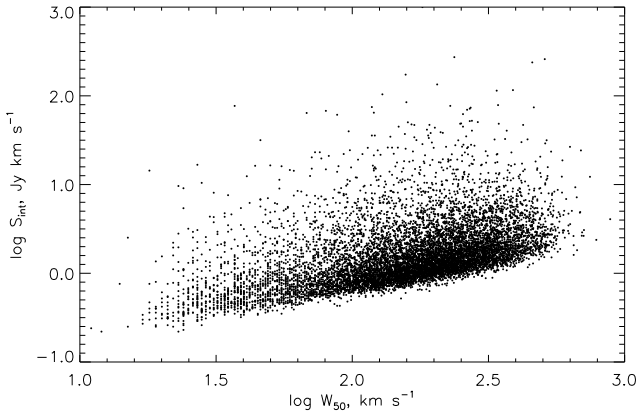

Figure 3.1 The distribution of sources detectable by ALFALFA, which is dependent on both flux S_{int} in Jy km s^{-1} and profile width W_{50} in km s^{-1}.

3.2.4 The 40% ALFALFA Survey Sample

ALFALFA catalogs have been extracted for a large contiguous region in the southern Galactic hemisphere (i.e. anti-Virgo direction) ($22^h < \alpha < 03^h$, $24^\circ < \delta < 32^\circ$), and two regions in the northern Galactic hemisphere (i.e. Virgo direction) ($16^h30^m < \alpha < 07^h30^m$, $4^\circ < \delta < 16^\circ$ and $24^\circ < \delta < 28^\circ$), with coverage totaling 2,607 deg^2 or $\sim 40\%$ of the final ALFALFA volume. This includes the previously published catalogs with a total of 2,706 extragalactic source measurements [Martin et al., 2009, Stierwalt et al., 2009, Kent et al., 2008, Saintonge et al., 2008, Giovanelli et al., 2007] in addition to an upcoming large online data release [Haynes et al. 2010, in prep][1]. This primary dataset includes both Code 1 (n = 10,452) and Code 2 (n = 2,750) galaxies in addition to Code 9 (n = 629) HVCs, where this figure includes measured subcomponents of larger cloud complexes.

[1]This data release includes an additional strip of coverage, $22^h < \alpha < 03^h$, $14^\circ < \delta < 16^\circ$, which is excluded here in favor of large contiguous areas.

From the primary dataset, we have selected the 40% ALFALFA Survey sample, hereafter $\alpha.40$. This sample has been selected to include only Code 1 objects, and the total sample size is further reduced by the exclusion of galaxies found beyond 15,000 km s^{-1}, where radio frequency interference from FAA radar makes ALFALFA blind to cosmic emission in a spherical shell ~ 10 Mpc wide. The final $\alpha.40$ sample contains 10,119 Code 1 galaxies, for a detection rate 3.9 deg^{-2} compared with the HIPASS detection rate ~ 0.2 deg^{-2} (5,317 extragalactic sources over 29,000 deg^2; Meyer et al. [2004] and Wong et al. [2006]). While rich in absolute number, HIPASS does not extend deep enough in redshift to sample a cosmologically fair volume.

In Figs. 3.2 and 3.3 we present the redshift distribution of the 10,119 Code 1 objects in $\alpha.40$ as a set of cone diagrams by region in the survey. The two most obvious features in Figure 3.2 are the prominent void in the foreground of the Pisces-Perseus supercluster, leading to the dearth of detections out to about 3,000 km s^{-1}, and the portion of the main ridge of that supercluster that cuts across the diagram. In the top panel of Figure 3.3, the nearby Virgo cluster is prominent, as is the Coma supercluster. ALFALFA probes a wide variety of environments in the local universe, and will soon study the overall properties of HI-selected galaxies as a function of environment (Saintonge et al. 2010, in prep).

Figure 3.4 displays histograms of the statistical properties of the $\alpha.40$ sample. From (a) to (d), these histograms represent the heliocentric velocity, velocity width W_{50}, integrated flux S_{int}, and S/N properties. In particular, note that the S/N is high for all detections, since Code 2 objects have been excluded from this analysis. For clarity, the histogram of the HI masses of galaxies in

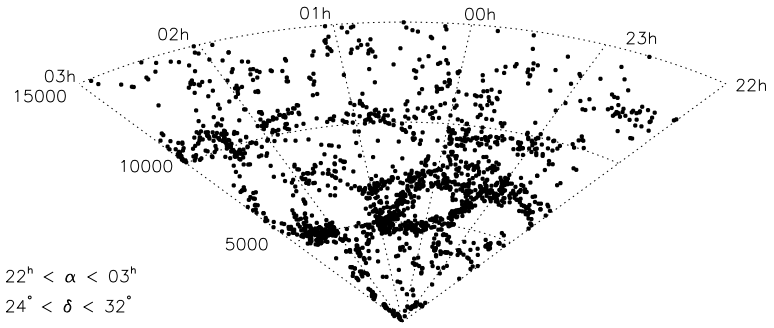

Figure 3.2 Distribution of 2,004 sources in the $22^h < \alpha < 03^h$, $24^\circ < \delta < 32^\circ$ portion of the $\alpha.40$ sample, plotted as R.A. vs. observed heliocentric recession velocity in km s^{-1}.

the sample is plotted separately, in Figure 3.5. On the low mass end, where ALFALFA can place strong constraints on the faint-end slope of the HIMF, the $\alpha.40$ sample contains $\sim$ 340 galaxies with $\log(M_{HI}/M_\odot) < 8.0$ and $\sim$ 114 with $\log(M_{HI}/M_\odot) < 7.5$; on the high mass end, which is best probed by surveys with deep redshift limits, there are $\sim$ 50 galaxies with $\log(M_{HI}/M_\odot) > 10.5$.

The large sample size of ALFALFA, extending over a range of HI masses, is one of its key strengths in relation to the problem of characterizing the density of neutral gas in the present-day universe. With such a large number of galaxies, we can approach our calculation of the HIMF in two distinct ways. First, using the entire sample and a well-known characterization of our sensitivity, we can apply corrections and obtain the overall function without excluding sources. Second, however, we can make stringent integrated flux cuts and use only those galaxies bright enough to be detectable irrespective of other properties (e.g. profile width). The sample contains $\sim$ 3500 galaxies with an integrated flux > 1.8 Jy km s^{-1}, which provides a strict cut above which our objects are de-

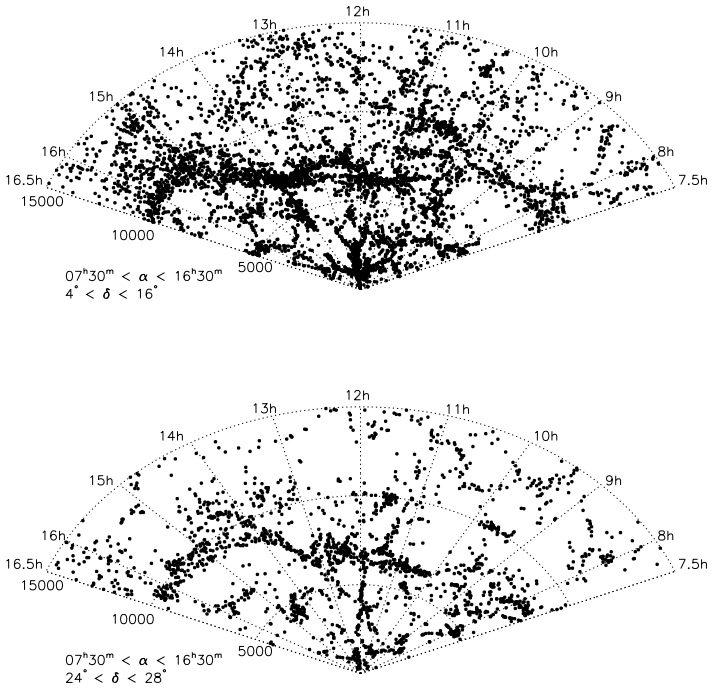

Figure 3.3 Top panel: Distribution of 5,960 sources in the $07^h30^m < \alpha < 16^h30^m$, $4^\circ < \delta < 16^\circ$ portion of the $\alpha.40$ sample, plotted as R.A. vs. observed heliocentric recession velocity in km s^{-1}. Bottom panel: 2,155 sources over the same R.A. range as above, with $24^\circ < \delta < 28^\circ$.

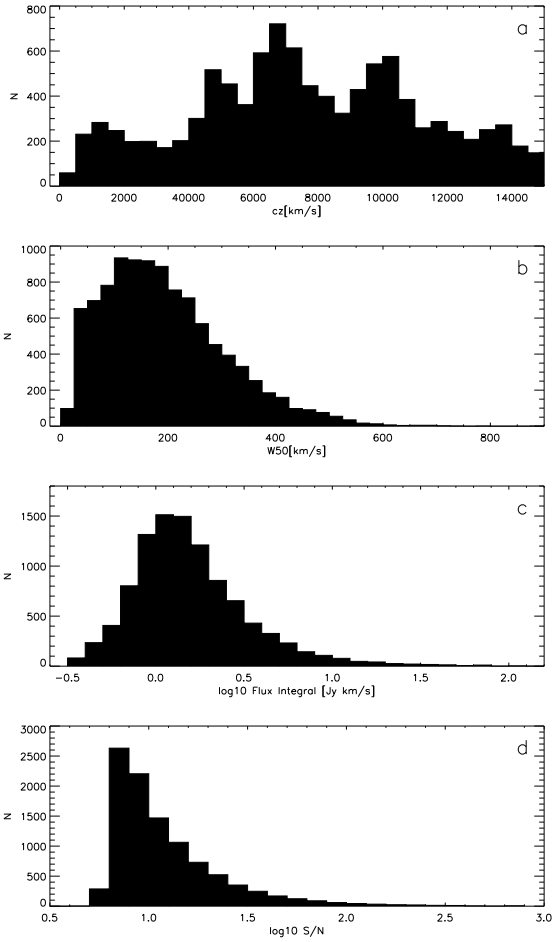

Figure 3.4 Histograms of the galaxy properties within $\alpha.40$: (a) heliocentric recession velocity in km s^{-1}; (b) HI line width at half power (W_{50}) in km s^{-1}; (c) logarithm of the flux integral in Jy km s^{-1}; (d) logarithm of the S/N.

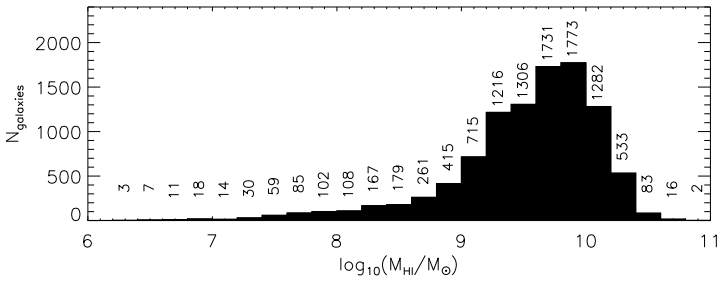

Figure 3.5 Histogram of the distribution of HI masses in the sample, plotted as logarithm of the HI mass in solar units.

tected regardless of profile width. This subsample size is comparable to the full sample size for previously-published HIMFs such as HIPASS, but samples a fair cosmological volume. This subsample, referred to hereafter as $\alpha.40_{1.8}$, provides a test case for analyzing the quality of the HIMF measurement for the full $\alpha.40$ sample. The precise details of the calculation, of ALFALFA's sensitivity, and of the corrections applied to the HIMF calculated from $\alpha.40$, make up the bulk of the following sections and of Appendices B.1 and B.2.

3.3 Determination of the HIMF

3.3.1 The HI Mass Function

The HI mass function, like galaxy luminosity functions, is usually parametrized as a Schechter function of the form

$$\phi(M_{HI}) = \frac{dn}{d\log M_{HI}} = \ln 10\ \phi_*\ \left(\frac{M_{HI}}{M_*}\right)^{\alpha+1}\ e^{-\frac{M_{HI}}{M_*}} \tag{3.1}$$

The parameters of interest are the faint-end slope α, the characteristic mass $\log M_*$, and the scaling factor ϕ_*.

$\phi(M_{HI})$ has historically been calculated in one of two ways. The $\Sigma 1/V_{max}$ method [Schmidt, 1968] can be understood by analogy to a purely volume-limited sample, in which case the HIMF would be obtained by the galaxy counts divided by the total volume of the survey. The $\Sigma 1/V_{max}$ method treats each individual galaxy in this way, by weighting the galaxy counts by the maximum volume $V_{max,i}$ within which a given source could have been detected. This weighting strategy allows the inclusion of low-mass galaxies, visible only in the nearby

Universe, in the same sample as rare high-mass galaxies, found only in larger volumes. Additionally, the weights may be adjusted in order to correct for a variety of selection effects, large-scale structure effects, and missing volume within the dataset, so that a well-characterized survey can robustly measure the HIMF.

An alternative method, the Two Dimensional Step-Wise Maximum Likelihood (2DSWML) approach, was applied to the HIPASS measurements of the HIMF [Zwaan et al., 2003, 2005]. This method is designed to make the calculation of the HIMF less sensitive to local large-scale structure, since shallow blind HI catalogs are contaminated by the richness of the Local Supercluster. If $1/V_{max}$ is used without correction for this overdensity, the resulting HIMF will overestimate the contribution by low-mass galaxies and steepen the faint-end slope α. Stepwise maximum likelihood methods, by contrast, are designed to reduce this bias, by assuming that the shape of the HIMF is the same everywhere and then obtaining the $\phi(M_{HI})$ that maximizes the probability of the observed distribution [Efstathiou et al., 1988]. Given the dependence of the ALFALFA survey's sensitivity on both mass and profile width (Section 3.2.3), a Two-Dimensional Step-wise Maximum Likelihood (2DSWML) approach is necessary to calculate the HIMF for the full sample [Loveday, 2000]. 2DSWML maintains the main advantages of the SWML method, which are its robustness against density fluctuations in the survey volume and its model-independent approach.

In this work, we apply both the $1/\mathrm{V}_{max}$ and the 2DSWML method for various reasons. Given our knowledge of our sample's characteristics and sensitivity, the $1/\mathrm{V}_{max}$ method is simple to apply and straightforward to assess for potential bias. We can account for large-scale structure and other selection effects by applying well-motivated corrections (discussed in Section 3.3.3). Perhaps

more significantly, this method also allows us to quantify and understand those effects on the ALFALFA survey. In particular, a goal of ALFALFA is to further probe the differences between HI mass functions in different environments; the 2DSWML assumption that the shape of the function is the same throughout a sample may not be valid. By contrast, the 2DSWML method is designed to be more resistant to effects from large-scale structure, and also results in a calculation of the selection function which can be used in future analysis of the sample via, for example, the two-point correlation function. A comparison of the $1/V_{max}$ and 2DSWML methods as applied to $\alpha.40$ is considered in Section 3.6 .

In both the $1/V_{max}$ and 2DSWML analyses, we have used 5 mass bins per dex, and have found that the HIMF is not strongly affected by choice of bin size. In the case of 2DSWML, we also bin by profile velocity width, and find no significant difference for bin sizes between 2 and 20 bins per dex. The two main sources of error are counting statistics within the bins and mass errors.

3.3.2 Errors on Distances and Masses

Minimizing and taking into account distance errors is key to robust estimation of luminosity and mass functions, in particular at the faint end. Masters et al. [2004] considered how strongly distance uncertainties will tend to affect a given local volume survey's estimate of the faint-end slope of the mass function. In that work, the authors accounted for distance errors by constructing a mock catalog, with masses assigned from a chosen HIMF and with the spatial distribution determined from the density field of the IRAS Point Source Catalog Redshift survey (PSCz; Branchini et al. [1999]). They concluded that a survey

toward the Virgo cluster, like a portion of the sample considered here, will overestimate distances to those galaxies if pure Hubble flow is used, since objects in that field are falling into Virgo. Since the HI mass depends on distance as D^2, this has serious consequences for the faint-end slope of the HIMF. Therefore, work in this region relies both on the development of well-constrained local velocity models from primary and secondary distance catalogs and on a careful consideration of the effects of distance uncertainties. We consider the Virgo cluster as a special case of this general problem in Section 3.6.1.

These difficulties arise precisely because the lowest mass objects can be detected only at small distances, so that the fractional distance errors due to deviations from Hubble flow most strongly affect the most interesting bins of the mass function. The best distance estimates, primary distances based on, e.g., Cepheids or the tip of the red giant branch, can only estimate distances to within $\sim 10\%$ error, so beyond cz $\sim$ 6,000 km s^{-1} the uncertainties on distances obtained via a primary method and those obtained assuming pure Hubble flow become comparable, and the latter is typically used for simplicity. Within that distance, however, the distance uncertainties can have a very strong influence, up to 100% in the case of the Virgo cluster.

To minimize distance uncertainties, the ALFALFA survey has adopted a distance estimation scheme that makes use of a peculiar velocity flow model for the local Universe [Masters, 2005]. This parametric multiattractor model, based on the SFI++ catalog of galaxies with Tully-Fisher distances [Springob et al., 2007], includes two attractors (Virgo and a Great Attractor) along with a dipole and quadrupole component. Distances to almost all $\alpha.40$ galaxies within 6,000 km s^{-1} are estimated from the flow model. Beyond cz$_{CMB}$ = 6,000 km s^{-1}, the model

is not well-constrained, so distances are estimated from Hubble flow (H_0 = 70 km s^{-1} Mpc^{-1}). Within 6,000 km s^{-1}, some galaxies have measured primary distances, which are applied in our scheme, and other galaxies are known to belong to a group, in which case the group's mean velocity is used for distance estimation. The Masters [2005] flow model also provides error estimates, constrained by the fit of the model to the observed velocity field and with a minimal error based on the local velocity dispersion 163 km s^{-1}. When distances are estimated using pure Hubble flow, the error is estimated to be $\sim 10\%$ via the assumption that peculiar velocities are $\sim$ a few hundred km s^{-1}.

Mass errors for individual galaxies in our sample are calculated from the measured error on the integrated flux and an estimated error on the distance, which is the larger of the local velocity dispersion 163 km s^{-1}, the distance error estimate of the Masters [2005] flow model, or 10% of the distance. Because the mass error shifts galaxies into different bins of the HIMF, the relationship between these errors and the final HIMF parameter errors is complex. We deal with these errors by calculating several hundred realizations of the HIMF after randomly assigning flux and distance errors to each galaxy to find the spread in each mass bin.

There is a complication on the high-mass end of the sample, as well. Arecibo's relatively large beam size at 21 cm ($\sim$ 3.5 arcmin) can cause source confusion at large distances, where we also find our largest-mass objects. When this occurs, ALFALFA may be detecting more than one individual gas-rich galaxy as a single source, but in cases of interaction it's also possible that the galaxies involved are part of a single, large HI envelope. While higher-resolution followup would be required to fully resolve this issue, we have in-

vestigated optical images and redshift catalogs for the highest mass ($\log M_{HI} > 10.5$) ALFALFA detections, and have found that the majority of these objects are not likely to be blends of HI emission from an interacting system and some others are close pairs that are likely to share a single gas envelope.

3.3.3 1/V$_{max}$ Method

For each galaxy in $\alpha.40$, $V_{max,i}$ is calculated based on that galaxy's HI mass M$_i$, the minimum integrated flux S$_{min,i}$ at which such a galaxy is detected in ALFALFA, and finally the distance D$_{max,i}$ corresponding to that limit. The calculated $V_{max,i}$, corresponding to the effective search volume for that galaxy, excludes volume that is not covered by ALFALFA, including volumes where detection ability, and therefore effective search volume, is reduced by the appearance of radio frequency interference at the corresponding frequency. Galaxies are binned by mass and $\phi(M_{HI})$ is calculated by summing the reciprocals of V_{max}.

By weighting the count for each galaxy, the $1/V_{max}$ method can be corrected for a variety of known systematic effects. The major corrections applied to the HIMF for this sample address (1) missing volume, (2) the profile width-dependent sensitivity of the survey, and (3) the known large-scale structure in the local volume.

Sources of radio frequency interference contaminate the signal in regions of frequency space corresponding to spherical shells in the survey volume. This effectively reduces the search volume of the overall survey. Figure 3.6 shows the average relative weight, compared to 100% coverage, within the $\alpha.40$ sur-

vey volume as a function of velocity. The large dip between 15,000 and 16,000 km s^{-1} is due to the FAA radar at the San Juan airport, and because of this extreme loss of volume at large distances we restrict the $\alpha.40$ sample to only those galaxies within 15,000 km s^{-1}. Given our knowledge that the relative weight is less than 1.0 at specific distances, the V_{max} value calculated for a specific galaxy is reduced to reflect the loss of effective search volume. This correction is not significant for the lowest-mass galaxies, but more generally, the correction is very small. The effect on the final Schechter parameters for the HIMF is on the order of 2%.

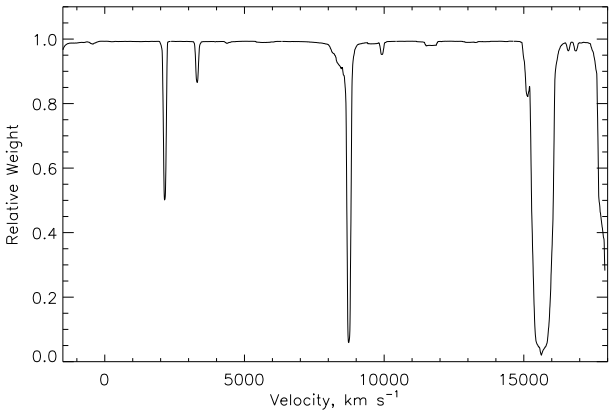

Figure 3.6 The average relative weight within the 40% ALFALFA survey volume as a function of observed heliocentric velocity. Where the relative weight is near 1.0, nearly the entire surveyed volume was accessible for source extraction, and the regions of lower relative weight correspond to manmade radio frequency interference. These sources are not always present, and do not always result in a complete loss of signal, so there are regions where the average weight is reduced only modestly. The large dip between 15,000 and 16,000 km s^{-1} is due to the FAA radar at the San Juan airport, and because of this extreme loss of volume at large distances we restrict our sample to only those galaxies within 15,000 km s^{-1}.

As discussed in 3.2.3, ALFALFA's detection ability is dependent on the profile velocity width of the signal, W_{50}, in km s^{-1}, rather than strictly on the integrated flux of the signal. To obtain an expression for this detection limit, we

used the data itself, as displayed in Figure 3.1. The dependence of ALFALFA sensitivity on both flux and profile width, described in Section 3.2.3, has the further complication of affecting the survey's completeness, and this must be accounted for in order to extract the underlying HIMF. The distribution in Figure 3.1 indicates that ALFALFA finds many galaxies with low fluxes and narrow widths, but there is a deficiency of galaxies with low fluxes and large widths. Because we have no knowledge of the true distribution below ALFALFA's detection capability, we have developed a completeness correction that takes advantage only of the data, making no assumptions about the potentially intrinsically small unobserved population. The profile width completeness correction most strongly affects galaxies with $\sim 9.0 < \log(M_{HI}/M_{\odot}) < 10.0$, and has a very small influence ($< 2\%$) on both the faint-end slope α, since low-mass (i.e. narrow velocity width) galaxies aren't affected, and $\log(M_*)$, since the counts in the high-mass bins are large enough to robustly constrain this. This is essentially a galaxy counting correction, so its primary influence is on ϕ_*, increasing that parameter by a factor of 20%. The full details of this completeness correction are described in Appendix B.1. The validity of this completeness correction, which we have applied to the full sample, is tested in Section 3.4.1 and 3.5.1 by calculating the HIMF using an integrated flux cut, which allows us to neglect the biased sensitivity dependence on width. By comparing the resulting HIMF in both cases, we assess the impact of this correction.

The most significant bias in the $1/\mathrm{V}_{max}$ calculation of the HIMF is that due to the large-scale structure of the galaxy distribution. Blind HI surveys tend to be relatively shallow and are thus biased by the overdensity of the local volume, which particularly affects the lowest-mass HI-rich galaxies that are only found nearby. If a correction for large-scale structure is not applied, we overestimate

the impact of low-mass galaxies on the overall HIMF, therefore boosting the faint-end slope α artificially. We discuss this correction in Appendix B.1. The large-scale structure volume correction has only a very weak effect on $\log(M_*)$, but the effects on α ($\sim 10\%$) and ϕ_* ($\sim 30\%$) are large. Since this correction is so significant, it is sensitive to the details of the density reconstruction used. Agreement between the $1/\mathrm{V}_{max}$ and 2DSWML results provide the best indication of the quality of this correction.

However, large-scale structure introduces the further bias of selectively reducing counts in mass bins that are primarily detectable in void volumes, and the weighting scheme correction cannot account for that. The voids in the Pisces-Perseus region between 3,000 and 8,000 km s^{-1}, visible in Figure 3.2, in particular, bias that portion of the $\alpha.40$ sample against galaxies with $8.5 < \log(M_{HI}/M_{\odot}) < 9.0$, leading to a systematic undercounting in those bins. Because the $1/V_{max}$ method is sensitive to large-scale structure, this undercounting introduces a spurious 'bump' feature into the HIMF, describe in detail in Section 3.4.1.

3.3.4 2DSWML Method

As discussed in Section 3 and 3.3.3, the main disadvantage of the $1/V_{max}$ method is its potential sensitivity to large-scale structure. If large-scale structure corrections were not adopted, the density of low HI-mass galaxies would be systematically overestimated, since most of these galaxies are detectable only in the very local, substantially overdense universe, including the Virgo Cluster and the Local Supercluster. This would bias the low-mass slope of the Schechter fit

to the HIMF (α), weakening one of the major strengths of the ALFALFA dataset, which is its ability to probe the population of extremely low HI-mass galaxies over a wide solid angle for the first time.

The original SWML method is applicable to samples selected by integrated flux. It assumes that the observed galaxy sample is drawn from a common HI mass function throughout the survey volume, denoted by $\phi(M_{HI})$. Unlike most Maximum Likelihood methods, which assume a functional form for $\phi(M_{HI})$ [Sandage et al., 1979], SWML splits the distribution in bins of $m = \log(M_{HI}/M_{\odot})$ and assumes a constant distribution within each logarithmic bin. In this way, the value of the distribution in each of the bins becomes a parameter, ϕ_j ($j = 1, 2, ..., N_m$), which is adjusted in order to maximize the joint likelihood of detecting all galaxies in the sample, hence yielding a Maximum Likelihood estimate of the mass distribution. Since the values of the parameters are free to vary independently, the procedure above is completely general and does not assume any functional form for the distribution *a priori*.

In the case where the sample is not integrated flux-limited and the selection function depends on additional observables, the SWML technique has to be extended to take into account the underlying galaxy distribution in all the physical properties that enter the calculation of the selection function. In the case of α.40, the limiting integrated flux depends on the galaxy profile width W_{50} and thus the method needs to consider the joint two-dimensional distribution of galaxies in both HI mass and observed velocity width, $\phi(M_{HI}, W_{50})$. 2DSWML relies on the assumption that the sample is statistically complete. Since ALFALFA's sensitivity to a source is dependent on both its integrated flux and its velocity width W_{50} (Section 3.2.3), we fit a strict completeness threshold to the observed

relationship as seen in Figure 3.1 and exclude galaxies falling below this completeness cut.

The details of the 2DSWML method and its application to α.40 are given in Appendix B.2.

3.3.5 HIMF Error Analysis

The simplest source of error in the estimate of the HIMF is from Poisson counting errors in the bins, which is added to the other sources of error considered next. The relationship between errors on corrections applied to individual galaxies and errors on the final HIMF points and measured parameters is complex. Mass errors, for example, may shift galaxies in the sample from one bin to another as discussed in Section 3.3.2, so it is not possible to analytically calculate the error on a particular bin. In order to treat these errors appropriately, we create > 250 realizations of the HIMF for each of the results shown in Section 3.4 and 3.5. The error on S_{int} is measured in the ALFALFA source extraction pipeline, and we have estimated errors on the distance for each galaxy in the sample. Each of these contributes to the mass error, and we apply a Gaussian random error to each galaxy's mass in each realization. The spread in the bin values across the ensemble of realizations contributes to the overall error in each point. We consider errors due to uncertain parameter estimation in the relationship between $\log(M_{HI}/M_{\odot})$ and the Gumbel distribution parameters μ and β in the same way. This results in a HIMF that has taken known sources of error into consideration.

Sources of systematic bias remain, particularly for the $1/V_{max}$ measurement

which is sensitive to the large-scale structure in the galaxy distribution. The effects of large-scale structure and of cosmic variance will be reduced as the survey continues, increasing its volume and coverage of varied cosmological environments.

In order to account for errors that are more difficult to quantify, we follow the example of Zwaan et al. [2005] and jackknife resample 21 equal-area regions. The resampling technique will help account for residual large-scale structure beyond that which we have corrected, and also for any systematic survey effects that change spatially across the sky or temporally throughout the survey's observations.

2DSWML Error Estimates

The 2DSWML approach introduces another source of error. We assign errors on the parameters ϕ_{jk}, introduced in Section 3.3.4, via the inverse of the information matrix following Loveday [2000] and Efstathiou et al. [1988]. The general form of the information matrix for a likelihood function $\mathcal{L}$ that depends on a set of parameters θ is given by

$$\mathbf{I}(\theta) = - \begin{bmatrix} \frac{\partial^2}{\partial\theta_m \partial\theta_n} \ln \mathcal{L} + \frac{\partial}{\partial\theta_m} g \frac{\partial}{\partial\theta_n} g & \frac{\partial}{\partial\theta_n} g \\ \frac{\partial}{\partial\theta_m} g & 0 \end{bmatrix} \tag{3.2}$$

where g is a constraint of the form $g(\theta) = 0$. We choose to apply the constraint $g = \sum_j \sum_k (\frac{M_{HI,j}}{M_{HI,ref}})^{\beta_1} (\frac{W_{50,k}}{W_{50,ref}})^{\beta_2} \phi_{jk} \Delta m \Delta w - 1 = 0$, with $\beta_1 = \beta_2 = 1$ and reference values for the HI mass and W_{50} equal to the $\alpha.40$ sample mean. The result is an error estimate for the parameters ϕ_{jk}, i.e. the value of the HIMF in

each mass bin, and is added in quadrature to the other sources of error described above.

3.4 1/V$_{max}$ Method: Results

3.4.1 Global HI Mass Function and Ω_{HI}

The global HI mass function derived from the $\alpha.40$ sample via the 1/V$_{max}$ method is presented in the top panel of Figure 3.7. Overplotted error bars have been derived as described above; mass errors due to errors on flux and distance estimates are reflected in the errors on the HIMF points, rather than on the mass-axis bin positions, since these errors change the bin counts.

The best-fit Schechter function describing this HIMF is overplotted as a dashed line. The derived parameters are ϕ_* (h$_{70}^3$ Mpc^{-3}dex^{-1}) = 6.0 $\pm$ .3 $\times$ 10^{-3}, $\log$(M$_*$/M$_\odot$) + 2 $\log$ h$_{70}$ = 9.91 $\pm$ 0.01 and α = -1.25 $\pm$ 0.02. However, the large-scale structure in the ALFALFA survey regions has introduced a 'bump' into this measurement of the HIMF. The feature visible in Figure 3.7 at $\log(M_{HI}/M_\odot) \sim 9.0$ does not appear to be intrinsic to the HI-rich galaxy population. However, previous work on luminous galaxies has suggested that the shape of luminosity and mass functions may be more complex than single Schechter functions. Luminosity functions in clusters, such as Coma and Fornax, are inconsistent with single values of α; Trentham [1998] has recommended a 'composite' luminosity function that steepens for both bright and faint objects and flattens out in between, which provides a 'dip' feature. Single Schechter functions provided a poor fit to 2dFGRS luminosity functions [Madgwick et al.,

2002], and results from the Sloan Digital Sky Survey also suggest that a second [Baldry et al., 2004] or third [Li and White, 2009] Schechter function component best describes the underlying population of galaxies at low redshift. While, given these findings, it is possible that the feature in Figure 3.7 suggests a complex shape in the HIMF, it is more likely that the feature is spurious, as we discuss below.

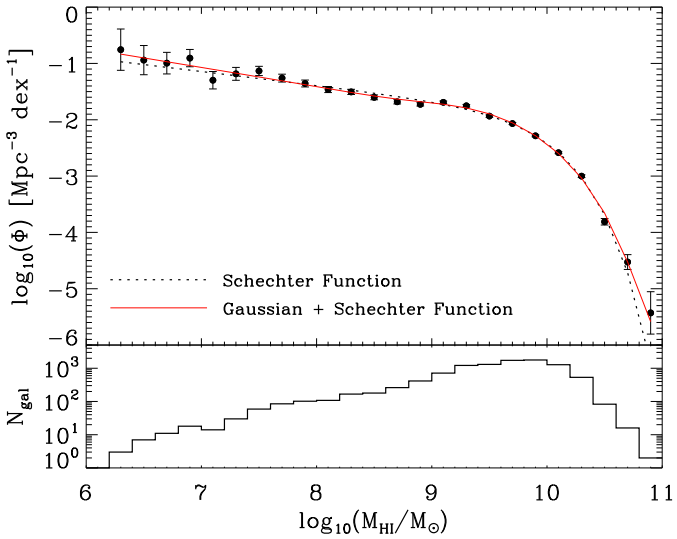

Figure 3.7 The global HI mass function derived from $\alpha.40$ via the $1/V_{max}$ method. Points are the HIMF value, per dex, in each mass bin, with errors as described in the text overplotted. The black dotted line is the Schechter function fit to the points, and the red solid line is the sum of a Schechter function and a Gaussian fit to the points. The histogram, bottom panel, shows the logarithm of the bin counts.

Such features occur because the $1/V_{max}$ method is sensitive to large-scale structure. Because the survey's HI mass sensitivity varies with distance (i.e., $\alpha.40$ is not a volume-limited sample), each mass bin in the HIMF corresponds to some preferred distance at which ALFALFA is most sensitive to galaxies in that mass bin. Extended large-scale structures can therefore change the shape of

the HIMF in bins corresponding to the distance of those features. Because of the large sample size of α.40, it is possible to separately investigate the three survey regions represented by the cone diagrams in Figs. 3.2 and 3.3 and to isolate the structures that contribute to such features. Specifically, the 'bump' feature in Figure 3.7 is due to a lack of sources in the foreground of the Great Wall and an overabundance within the Great Wall, clearly evident in Figure 3.3. The large-scale structure correction (Section 3.3.3 and Appendix B.1) reduces this feature, but cannot totally eliminate it, in part because density maps used to correct for large-scale structure are smoothed to $\sim$ a few Mpc scales and can underestimate extremes in the density contrast. Features such as this one will be reduced as the ALFALFA survey continues and the sample grows. The 2DSWML method is not sensitive to large-scale structure and does not produce this feature (Section 3.5).

This feature appears significant in part because our statistical errors on the HIMF points are so small, but it leads to a poor fit and an underestimate for the faint-end slope α. This is clear in Figure 3.8, which displays the residual between the $1/V_{max}$ HIMF points and the derived best-fit Schechter function in the top panel and shows that the Schechter function systematically over- and under-estimates the HI mass function due to this feature.

While this feature is well-understood, it has the undesirable effect of artificially reducing the faint-end slope α. In an effort to reduce the effect of this spurious feature and to better fit the points, we fit the sum of a Schechter function and a Gaussian; the Gaussian component serves to filter out the feature, leading to a better estimate of α. The results are shown as the solid line in Figure 3.7 with the residuals shown in the bottom panel of Figure 3.8. This

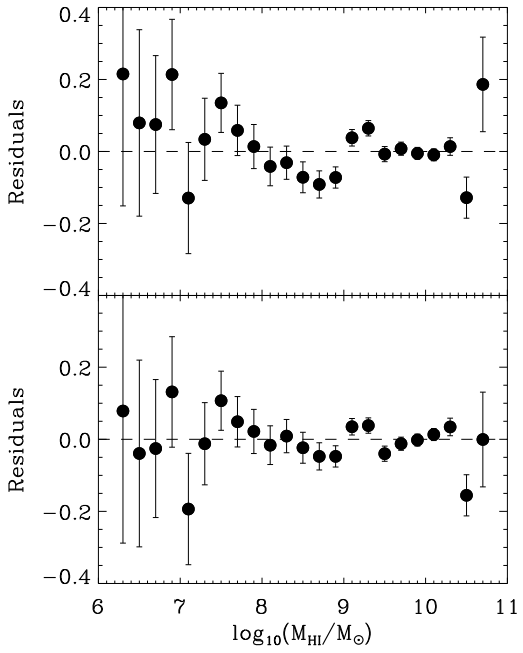

Figure 3.8 The residuals between the $1/V_{max}$ HIMF points and the derived best-fit Schechter function (top panel) and the best-fit sum of a Schechter and a Gaussian (bottom panel). Bars represent the errors on each point, to show the significance of the residual in each case. The Schechter function provides a poor fit to the spurious 'bump' feature, and this effect is reduced by the addition of a Gaussian component. The highest-mass bin, which has a large error value, is excluded from this plot.

fit significantly improves the reduced χ^2, and the residuals are small and, near $\log(M_{HI}/M_\odot) \sim 9.0$, more randomly scattered about 0 in contrast to the top panel of Figure 3.8. However, there is larger uncertainty in the parameters in this case, since each function is constrained by fewer points. The Schechter function parameters, displayed in Table 3.1, are $\log(M_*/M_\odot)$ + 2 $\log$ h_{70} = 9.95 $\pm$ 0.04 and α = -1.33 $\pm$ 0.03. The Schechter function measurement of ϕ_* (h_{70}^3 $Mpc^{-3}dex^{-1}$) = 3.7 $\pm$ .6 $\times$ 10^{-3}, however, has been affected by the addition of the second component to the fit, and we therefore defer to the 2DSWML measurement of that parameter.

The Gaussian parameters are not included in Table 3.1, since they are used to filter out the 'bump' feature and are not expected to have physical meaning. The best-fit Gaussian has peak height (h_{70} Mpc^{-3}) 5 $\pm$ 1 $\times$ 10^{-3}, mean $\log(M_\mu/M_\odot)$ + 2 $\log$ h_{70} 9.28 $\pm$ 0.06 and spread in $\log(M_\mu/M_\odot)$ + 2 $\log$ σ = 0.41 $\pm$ 0.03.

We conclude that the proper values of α and $\log(M_*/M_\odot)$ extracted from the $1/V_{max}$ method are -1.33 $\pm$ 0.03 and 9.95 $\pm$ 0.04, respectively. Table 3.1 lists both the spurious $1/V_{max}$ Schechter function parameters as well as the parameters found when a Gaussian is added to fit the spurious feature. The addition of the Gaussian brings the $1/V_{max}$ results for the parameters α and M_* into excellent agreement with the 2DSWML method and the flux-limited $\alpha.40_{1.8}$ subsample results.

As an additional test of our corrections for profile width sensitivity, we have derived the $1/V_{max}$ HIMF from the integrated flux-limited subsample $\alpha.40_{1.8}$ (described in Section 3.2.4). This mass function is corrected for large-scale structure and include mass errors, but is not subject to the same bias against broad HI profiles. The $\alpha.40_{1.8}$ HIMF is well-fit by a pure Schechter function. The re-

sults are listed in Table 3.1. The $\alpha.40_{1.8}$ HIMF is consistent with those derived from the full $\alpha.40$ sample. We therefore conclude that our survey sensitivity is well-characterized and that our measurements based on the full sample are complete and representative. However, since this limited sample does not probe the galaxies at the extremes of the mass function, it is subject to larger errors on the points and in the parameters.

Measurement of Ω_{HI}

The density Ω_{HI} of neutral hydrogen in the local Universe, expressed in units of the critical density, can be calculated in two ways from the derived HI mass function. Integrating analytically over the best fit Schechter function gives $\Omega_{HI} = \phi_* M_* \Gamma(2+\alpha)$= 4.4 $\pm$ 0.3 $\times 10^{-4}$ h_{70}^{-1}, slightly (16%) higher than the final HIPASS value 3.7 $\times 10^{-4}$ h_{70}^{-1} [Zwaan et al., 2005]. Using the binned points directly, we find the same result: Ω_{HI} = 4.4 $\pm$ 0.1 $\times 10^{-4}$ h_{70}^{-1}. This agreement is an indication that our findings are well-represented in the high-mass bins by our Schechter function fit, despite the spurious feature. Ω_{HI} carries a small error since it is negligibly affected by the mass and distance errors on the faint end.

In Figure 3.9, we show the contribution of each $1/V_{max}$ mass bin to Ω_{HI} as filled circles. The total density of neutral hydrogen in the local Universe is dominated by galaxies with $9.0 < \log(M_{HI}/M_\odot) < 10.0$, and in these bins we measure the HIMF to be larger than Zwaan et al. [2005] do, thus finding a larger value of Ω_{HI}. The ALFALFA survey extends further in redshift than HIPASS, with a median redshift $\sim$ 8,000 km s^{-1} compared to $\sim$ 3,000 km s^{-1}, allowing us to detect significantly more high-mass objects (Section 3.6.2).

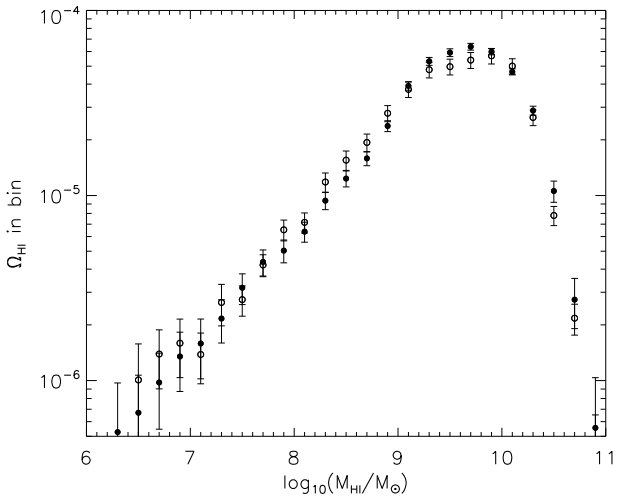

Figure 3.9 The contribution to Ω_{HI} by the galaxies in each bin in $\alpha.40$. Filled circles have been calculated via the $1/V_{max}$ method, and open circles are from the 2DSWML method. The total density of neutral hydrogen in the local Universe is dominated by galaxies with $9.0 < \log(M_{HI}/M_\odot) < 10.0$.

3.5 2DSWML Method: Results

3.5.1 Global HI Mass Function and Ω_{HI}

The HIMF derived from $\alpha.40$ through the 2DSWML method is shown in Figure 3.10. The derived parameters are ϕ_* (h_{70}^3 Mpc^{-3}dex^{-1}) = 4.8 $\pm$ .3 $\times$ 10^{-3}, $\log(M_*/M_\odot)$ + 2 $\log$ h_{70} = 9.96 $\pm$ 0.02 and α = -1.33 $\pm$ 0.02. To test the robustness of this HIMF estimate, we also applied a one-dimensional SWML approach to the flux-limited $\alpha.40_{1.8}$ sample, and found results consistent with the global, two-dimensional result (ϕ_* = 4.5 $\pm$ .9 $\times$ 10^{-3}, $\log(M_*)$ = 9.96 $\pm$ 0.04 and α = -1.36 $\pm$ 0.06).

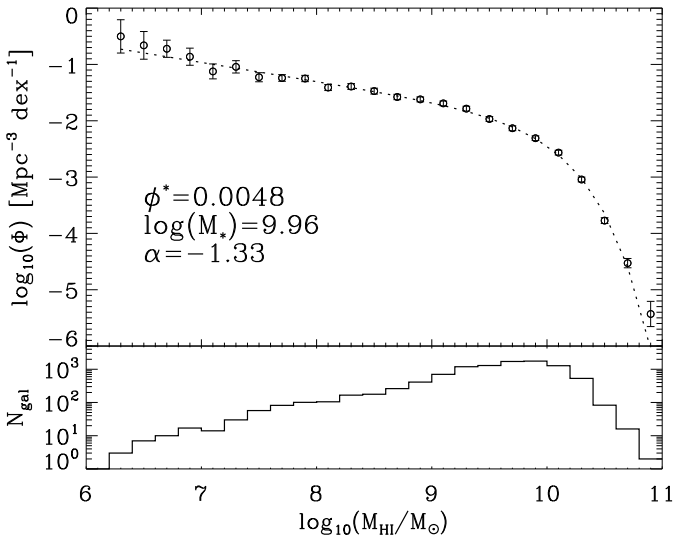

Figure 3.10 The global HI mass function derived from $\alpha.40$ via the 2DSWML method. As in Figure 3.7, points are the HIMF value, per dex, in each mass bin, with errors as described in the text overplotted. The dotted line is the Schechter function fit to the points and the Schechter function parameters are listed. The histogram, bottom panel, shows the logarithm of the bin counts.

Table 3.1. HI Mass Function Fit Parameters

Sample and Fitting Function	α	ϕ_* (10^{-3} h_{70}^3 Mpc^{-3} dex^{-1})	log ($M_*/M_\odot$) + 2 log h_{70}	Ω_{HI}, fit ($\times 10^{-4}$ h_{70}^{-1})	Ω_{HI}, points ($\times 10^{-4}$ h_{70}^{-1})
$1/V_{max}$	-1.25 (0.02)	6.0 (0.3)	9.91 (0.01)	4.4 (0.2)	4.4 (0.1)
Schechter + Gaussian[a]	-1.33 (0.03)	3.7 (0.6)[b]	9.95 (0.04)		
$1/V_{max}$, Non-Virgo	-1.20 (0.02)	6.1 (0.3)	9.90 (0.01)	4.1 (0.2)	4.3 (0.1)
Schechter + Gaussian[a]	-1.33 (0.04)	3.1 (0.6)[b]	9.95 (0.05)		
2DSWML	-1.33 (0.02)	4.8 (0.3)	9.96 (0.02)	4.3 (0.3)	4.4 (0.1)
2DSWML, Non-Virgo	-1.34 (0.02)	4.7 (0.3)	9.96 (0.01)	4.3 (0.3)	4.4 (0.1)
$1/V_{max}$, $\alpha.40_{1.8}$	-1.30 (0.03)	4.6 (0.3)	9.96 (0.02)	4.0 (0.3)	4.0 (0.1)
1DSWML, $\alpha.40_{1.8}$	-1.36 (0.06)	4.5 (0.9)	9.96 (0.04)	4.4 (0.9)	4.3 (0.3)
HIPASS [Zwaan et al., 2005][c]	-1.37 (0.06)	5 (1)	9.86 (0.04)	3.7 (0.5)	
Leo Group [Stierwalt et al., 2009][d]	-1.41 (0.2)				

[a] In the $1/V_{max}$ case, pure Schechter functions provide a poor fit to the faint-end slope α, which explains the difference in α for two fitting functions. The Gaussian component parameters are not shown in the table, given that they are not expected to be physical.

[b] We defer to the 2DSWML measurement of ϕ_*, due to the spurious feature in the $1/V_{max}$ results.

[c] Reported statistical and systematic errors combined in quadrature.

[d] The excluded parameters ϕ_* and M_* in the Leo Group are highly uncertain due to the lack of high-mass galaxies in its small volume.

Measurement of Ω_{HI}

As in the case of the $1/V_{max}$ method, we calculate the neutral hydrogen density Ω_{HI} from an analytical integration of the best-fit Schechter function and from a summation over the points themselves. From the Schechter function we find $\Omega_{HI} = 4.3 \pm 0.3 \times 10^{-4}$ h_{70}^{-1} and from the binned points we find 4.4 ± 0.1 $\times 10^{-4}$ h_{70}^{-1}. In both cases, our result is consistent with the $1/V_{max}$ method and is slightly higher than the HIPASS result. The contribution by each bin is shown in Figure 3.9 as open circles.

3.6 Discussion

Figure 3.11 compares the $\alpha.40$ HIMF derived via the $1/V_{max}$ method (filled circles) and the SWML method (open circles), and shows the difference between

them in the bottom panel. The bin-by-bin differences between the SWML and $1/V_{max}$ methods are small, and do not affect the measurement of Ω_{HI}, though the faintest, most error-prone bins are found to be more populated in the SWML analysis. After we have corrected for the feature introduced to the $1/V_{max}$ result by large-scale structure, we find excellent agreement between all measurements of α (-1.33 $\pm$ 0.02) and Ω_{HI} (4.3 $\pm$ 0.2).

In the case of $1/V_{max}$, large-scale structure and the correction we estimate to deal with it have the largest impact on the final result.The 2DSWML method is designed to be insensitive to density fluctuations, and the agreement between the two measurements indicates that the large-scale structure correction is successful.

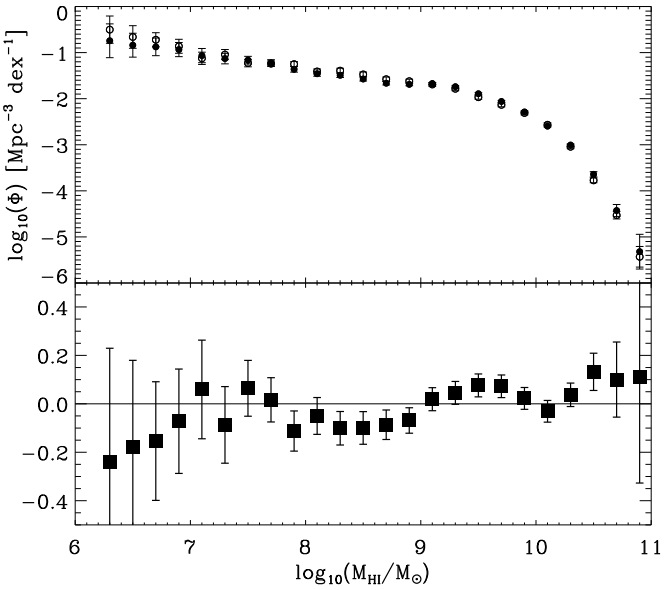

Figure 3.11 Top panel: The HIMF derived from $\alpha.40$ with the $1/V_{max}$ method (filled circles) and the 2DSWML method (open circles), with error bars. Bottom panel: The difference between the HIMF points, shown above, derived from the $1/V_{max}$ and 2DSWML methods.

3.6.1 Impact of the Virgo Cluster

Measurements of the HI mass function can be sensitive to large-scale structure in the survey volume. As discussed above, we correct for large-scale structure in the $1/V_{max}$ method to ameliorate this effect, but our 2DSWML measurement could also be sensitive to this large nearby overdensity. To test the robustness of the $1/\mathrm{V}_{max}$ correction and of our derived HIMF, we consider the result obtained when we exclude the portion of $\alpha.40$ that crosses the Virgo cluster. Many of our low-mass objects are contributed by this nearby overdensity, and our large-scale structure correction mechanism is the largest in this region; if we are correcting appropriately, we should obtain the same result regardless of the inclusion of the Virgo sources. This test is imperfect, given that the local volume generally is overdense. We exclude all galaxies lying within our adopted Virgo field, covering $12^h < \alpha < 13^h$ and the full declination extent of the $\alpha.40$ survey [Trentham and Hodgkin, 2002], reducing the sample size to ∼9,200 for $1/\mathrm{V}_{max}$ and ∼ 8,600 for 2DSWML. Errors are measured as described above, but in this case we jackknife resample over only 18 subregions.

Our results, within the errors, are the same whether or not we exclude the Virgo overdensity. This is true both for parameters and for our measurement of Ω_{HI}. In the case of $1/\mathrm{V}_{max}$, we again find that a Schechter summed with a Gaussian provides a better fit to the data by accounting for features introduced by large-scale structure in the foreground of the Pisces-Perseus supercluster In Table 3.1, we compare our findings for samples inclusive and exclusive of Virgo. Additionally, we list the HIPASS HI mass function and the Stierwalt et al. [2009] HIMF of ALFALFA sources in the Leo group. In the case of the $\alpha.40$ and $\alpha.40_{1.8}$ samples, we also list the value of Ω_{HI} found by integrating the Schechter func-

tion fit or using the HIMF bin points. Each table entry is accompanied by 1σ errors in parentheses.

3.6.2 Comparison with Previous Work

We find a value of Ω_{HI} that is 16% higher than the complete HIPASS survey value [Zwaan et al., 2005]. That HIPASS result is excluded by our 2σ errors, but the more preliminary HIPASS result [Zwaan et al., 2003] is in agreement with our result while carrying significantly larger error than we find. We also find $\log$ ($M_*/M_\odot$) = 9.96, so that the break in our HIMF occurs at masses 0.1 dex higher than was found in either of the HIPASS analyses. Since the high-mass end of the HIMF is sensitive to M_*, HIPASS significantly undercounts the highest-mass gas-rich galaxies. When our Schechter function is extrapolated to $\log(M_*/M_\odot)$ = 11.0, we predict an order of magnitude more galaxies than HIPASS. At more modest values, $\log$ ($M_*/M_\odot$) = 10.75, this is reduced to a factor of ~ 5.

In Figure 3.12, we show the mass of $\alpha.40$ detections as a function of their distance in Mpc, and compare that to the HIPASS completeness and detection limits. The dashed vertical line shows the 12,700 km s^{-1} redshift cutoff of HIPASS assuming H_0 = 70 km s^{-1} Mpc^{-1}), demonstrating the ALFALFA survey's ability to probe the rare highest-mass galaxies at large redshifts. While the $\alpha.40$ sample extends only to 15,000 km s^{-1}/ in order to avoid rfi, the full ALFALFA bandwidth allows us to probe to 18,000 km s^{-1}. Given that the survey was designed to be sensitive at those greater redshifts, we are still able to observe many galaxies at the limit of $\alpha.40$, while the Zwaan et al. [2005] sample becomes very sparse near the survey's redshift limits.

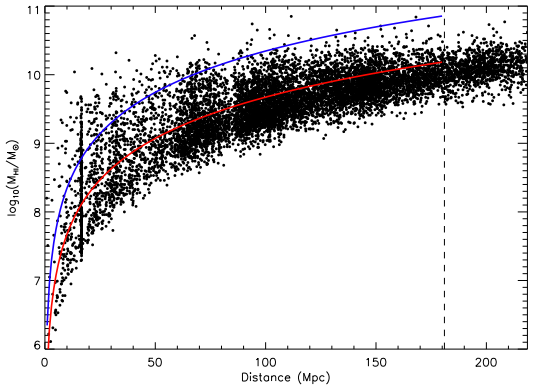

Figure 3.12 $\alpha.40$ detections plotted as $\log(M_{HI}/M_{\odot})$ vs. distance in Mpc. The upper (blue) solid line is the HIPASS completeness limit, and the lower (red) solid line is the HIPASS detection limit. The dashed vertical line shows the redshift limit of HIPASS assuming the ALFALFA adopted value H_0 = 70 km s^{-1} Mpc^{-1}.

This improved measurement of the HIMF has implications for work that relied upon the HIPASS results. Present-day HI surveys are limited in their ability to probe redshift space, even when they are targeted ($z < 0.5$), so models of evolution of the HI mass function rely on the measurement at z = 0. Higher-precision measurements provide better constraints for evolutionary models. Numerical models of galaxy formation and evolution [Power et al., 2010] depend on the z = 0 HIMF to assess the success of the models and to extrapolate that result to predictions for future HI surveys. For example, Abdalla et al. [2010] predicted the ability of future HI line surveys with an instrument like the Square Kilometer Array (SKA) to constrain dark energy through measurements of the baryon acoustic oscillation scale. Those authors consider models of the HIMF evolution that are sensitive to the value M_*. Typically, these galaxy models also depend on the assumed H_2/HI ratio to convert simulated cold gas into atomic and molecular components (e.g. Baugh et al. [2004]), so updated estimates of either Ω_{HI} or Ω_{H_2} affect our ability to produce realistic models of

gas-rich galaxies.

We confirm previous findings that Ω_{HI} at z = 0 is inconsistent with the value inferred from damped Lyman absorber (DLA) systems at z $\sim$ 2 and that significant evolution is required to reconcile measurements in the two epochs [Noterdaeme et al., 2009, Rao et al., 2006, Prochaska and Wolfe, 2009, Prochaska et al., 2005], while providing a tighter constraint on the present-day energy density of cold gas.

3.6.3 Comparison with Simulations

Obreschkow et al. [2009] (hereafter O09) used the Millennium Simulation catalog, the De Lucia and Blaizot [2007] virtual catalog of galaxies, and a physically-motivated prescription to assign realistic gas (HI, H_2 and He) masses at a range of redshifts. While this catalog has a limited ability to realistically trace detailed galaxy evolution and limited mass resolution – down to about $10^{8.0} M_\odot$ of neutral hydrogen, which is comparable to the particle size in the Millennium run [Springel et al., 2005] – it serves as the best currently available comparison of observed gas-rich disks with the underlying theory of dark matter halos.

Simulated HI Mass Function

O09 derive an HI mass function that is, in its gross properties, consistent with HIPASS [Zwaan et al., 2005], ignoring spurious features near the mass resolution limit of the simulation. The O09 gas masses are obtained by combining the cold particle masses from the Millennium Run with a model to split the cold gas

into molecular hydrogen and atomic hydrogen and helium components. Figure 3.13 compares the O09 HIMF, including only galaxies with $\log(M_{HI}/M_{\odot}) > 8.0$ and at redshift z = 0, with the $1/V_{max}$ and 2DSWML HIMFs derived in this work. Ω_{HI}^{sim} = 3.4 $\times 10^{-4}$ inferred from the O09 HIMF is in good agreement with this work and with HIPASS. While it is clear that the overall statistical distribution of the cold gas prescription generally recovers the overall density and the gross properties of the statistical distribution, the details of the O09 HIMF disagree with observations, particularly at the extreme low-mass end where the Millennium Run work suffers from poor resolution and inadequate merger histories.

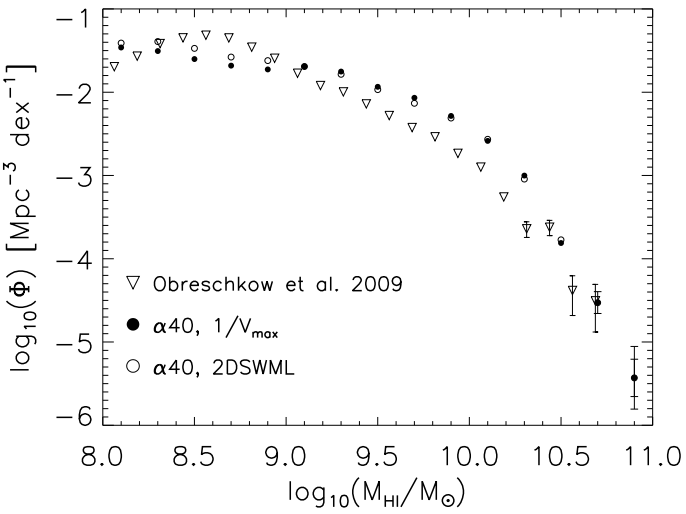

Figure 3.13 The HIMF of the Obreschkow et al. [2009] analysis of cool gas in simulated galaxies from the Millennium run (open triangles), compared to the $\alpha.40$ $1/V_{max}$ (filled circles) and 2DSWML (open circles) HIMFs. The ALFALFA sample is divided to 5 mass bins per dex, and the simulated galaxies to 8 bins per dex. Only the mass range $\log(M_{HI}/M_{\odot}) > 8.0$ is displayed, due to poor mass resolution in O09, and the simulated galaxy sample includes only galaxies at redshift z=0. For the ALFALFA HIMF, error bars represent both counting and mass estimate errors, but errors on the O09 HIMF are based on Poisson counting only. Where not visible, error bars are smaller than the plotted symbol size.

It is also worth noting that O09 report that they overpredict the number of high-mass sources in comparison to HIPASS, and suggest that this may be due to opacity in observed disks at these masses. However, we find that they underpredict high mass galaxies at z=0, the opposite effect. This is likely due to the O09 analysis of the HIMF, which is not limited to the final galaxies evolved to z=0; rather, their HIMF also includes galaxies at higher-redshift simulation snapshots which are presumably more gas-rich than their present-day counterparts. This would therefore overpredict the abundance of high-mass galaxies.

Faint-End Slope

As has been found in previous work, the faint-end slope of the $\alpha.40$ HIMF is significantly shallower than the Press-Schechter prediction of $\alpha \sim -1.8$ [Press and Schechter, 1974]. Potentially, this difference can be linked to baryon loss and the suppression of accretion via photoheating in the low-mass dark matter halos. Simulations suggest that dark matter halos with masses below $\sim$ 6.5 $\times$ 10^9 h^{-1} $M_\odot$ result in baryon-poor galaxies in present-day voids and other environments [Hoeft et al., 2008, 2006, Hoeft and Gottloeber, 2010]. In principle, the discrepancy could be explained by an argument invoking the mass scale at which photoheating becomes important.

A fitting function has been proposed [Gnedin, 2000] to describe the behavior of baryon fraction as a function of underlying halo mass:

$$f_b = f_{b0}\left[1 + (2^{\gamma/3} - 1)\left(\frac{M_c}{M_{tot}}\right)^{\gamma}\right]^{-3/\gamma} \tag{3.3}$$

where the parameters f_{b0} and M_c are, respectively, the baryon fraction in large

halos and the characteristic halo mass where f_b=f_{b0}/2.

If decreasing baryon fraction with decreasing halo mass explains the difference between low-mass slopes in baryonic (stellar and HI) and halo mass functions, then this fitting function should consistently predict baryonic and cold gas mass functions with values of $\alpha \sim -1.3$. In the low-mass limit, the first term of Equation 3.3 can be dropped and the total mass in a halo can be assumed to be dominated by the dark matter, $M_{tot} \approx M_D$. Via the definition f_B=M_B/M_D we have

$$M_D = \frac{M_B}{f_B} = \frac{M_B}{f_{b0}}(2^{\gamma/3} - 1)^{3/\gamma} \left(\frac{M_D}{M_c}\right)^{-3} \tag{3.4}$$

Compressing all constants gives the relation $M_D \propto M_B^{1/4}$, which can then be used to relate the low-mass ends of the baryonic and dark matter mass functions. On the faint end of the dark matter mass function, the exponential term of the Schechter function can be dropped. From $\frac{d\log M_D}{d\log M_B} = \frac{1}{4}$ we can, finally, conclude that

$$\phi(M_B) = \frac{dn}{d\log M_B} \propto \phi_* \left(\frac{M_B}{M_{*,B}}\right)^{(\alpha_D+1)/4} \tag{3.5}$$

where $(\alpha_D + 1)/4 = \alpha_B + 1$. Starting from the Press-Schechter prediction of a faint-end slope $\alpha_D \sim -1.8$, the consideration of baryon fraction leads to $\alpha_B \sim -1.2$, which is more consistent with HI and stellar mass functions. In principle, the discrepancy between dark matter simulations and observed baryon mass functions could be explained by the photoheating simulations of Hoeft et al. [2006] and Hoeft and Gottloeber [2010].

The baryon fraction of O09's simulated galaxies loosely follows this descriptive baryon fraction function (Equation 3.3). However, the halo mass scale at which the baryon loss starts to drop steeply is about two orders of magnitude larger than the scale found by the detailed hydrodynamical simulations of Hoeft et al. [2008]. Additionally, there is large scatter in the mass interval of interest, since the simulation's resolution is poor for the halo masses where baryon loss becomes important. The level of agreement between O09 and Hoeft et al. [2008] is therefore difficult to quantify, and we use the latter's determination of f_b in what follows.

Equation 3.3 suggests that the baryonic content of low-mass galaxies in $\alpha.40$ may be severely biased with respect to the underlying halo mass distribution. If simulations accurately predict the relationship between initial halo masses and resulting baryon fractions after reionization and photoheating, then the application of f_b should provide an estimate of the resulting baryon mass function at z=0. This depends on the extremely naive assumption that the cold HI gas content is depleted in the same fraction as the baryons overall.

The publicly available GENMF code[2] produces halo mass function fits to the Reed et al. [2007] N-body simulations at high resolution, from 10^5 to 10^{12} h^{-1} $M_\odot$. We adopt their mass function at z=0, with their suggested parameters $\Omega_M \approx 0.238$, $\Omega_\Lambda \approx 0.762$, and $\sigma_8 = 0.74$ (at z=0), and apply Equation 3.3 to extract the predicted baryon mass function and fit the faint-end slope. The results are displayed in Table 3.2 for an exemplary set of values for $f_{b,0}$, M_c and γ.

Through this approach, it is possible to modify the underlying halo mass function ($\alpha \approx$ -1.8) to meet our observations ($\alpha \approx$ -1.3). The suggestion that low-

[2]http://icc.dur.ac.uk/Research/PublicDownloads/genmf_readme.html

Table 3.2. Faint-End Slopes of Modeled Baryon Mass Functions

$f_{b,0}$	M_c	γ	α
0.20	9.0	1.0	-1.30
0.20	9.5	1.0	-1.27
0.16	9.0	1.0	-1.31
0.16	9.5	1.0	-1.28
0.16	9.0	1.5	-1.24
0.16	9.5	1.5	-1.22
0.16	9.0	2.0	-1.21
0.16	9.5	2.0	-1.19
0.15	9.0	1.0	-1.31
0.15	9.5	1.0	-1.28
0.15	9.0	2.0	-1.21
0.15	9.5	2.0	-1.19

mass halos may re-accrete cold gas at late times [Ricotti, 2009a], if substantiated, could further change the shape of the resulting baryon mass function. While this approach indicates we may be close to resolving the missing satellites problem and the discrepancy between predicted and observed faint-end mass function slopes, the precise requirements of baryon depletion mechanisms are not well-constrained by available simulations.

3.7 Conclusions

We have derived the HI mass function from a sample of ∼10,000 extragalactic sources comprising the ALFALFA 40% Survey, and have adapted the $1/V_{max}$ method to fully account for survey sensitivity, large-scale structure, and mass errors. We have demonstrated the robustness of this method by testing flux-limited samples and by calculating the HIMF via a second approach,

the structure-insensitive 2DSWML method. Our major result, the derivation of the global HIMF, indicates a Schechter function with parameters ϕ_* (h_{70}^3 $Mpc^{-3}dex^{-1}$) = 4.8 $\pm$ .3 $\times$ 10^{-3}, $\log(M_*/M_\odot)$ + 2 $\log$ h_{70} = 9.96 $\pm$ 0.02 and α = -1.33 $\pm$ 0.02.

We find Ω_{HI}= 4.3 $\pm$ 0.3 $\times 10^{-4}$ h_{70}^{-1}, a robust constraint that is 16% higher than the complete HIPASS survey value 3.7 $\times 10^{-4}$ h_{70}^{-1} [Zwaan et al., 2005], which we exclude at the 2σ level. The more preliminary HIPASS result [Zwaan et al., 2003] is in agreement with our result, but carries a significantly larger error. When we exclude the Virgo cluster from our analysis, the Ω_{HI} value remains stable, indicating that our measurements are robust against large-scale structure. In each case, we find the same value Ω_{HI} whether derived from the binned HIMF points themselves or from the best-fit Schechter parameters.

The larger values of Ω_{HI} and of M_* that we find in comparison to HIPASS demonstrate ALFALFA's advantage in detecting high-mass galaxies at large distances. On the extreme high-mass end of the HI mass function, our measurement and the accompanying Schechter function predict an order of magnitude more galaxies at $\log(M_{HI}/M_\odot) \sim 11.0$, and we find a factor of $\sim$5 more galaxies at $\log(M_{HI}/M_\odot)$ = 10.75. This has implications for previous estimates of the detection rate of future large-scale HI line surveys with the SKA.

We confirm previous findings that significant evolution in cold gas reservoirs must occur between z $\sim$ 2 and z = 0 given that Ω_{HI} is a factor of $\sim$ 2 smaller in the former epoch compared with the latter [Noterdaeme et al., 2009, Rao et al., 2006]. Further, we suggest that work on photoheating and other processes that prevent low-mass dark matter halos from accreting gas may be coming close to explaining the so-called 'missing satellite problem' at low redshift. Further

numerical work, particularly at resolutions capable of recovering low densities of cold gas at z=0, is required in this area of research.

Future work will consider the variation of the HI mass function with environment, and will include larger numbers of galaxies across a full range of extragalactic environments as the ALFALFA survey continues and new data products are released.

CHAPTER 4

THE IMPACT OF SURVEY SENSITIVITY, MASS UNCERTAINTIES, AND LARGE SCALE STRUCTURE ON THE 40% ALFALFA HI MASS FUNCTION

4.1 Introduction

The $\alpha.40$ sample of gas-rich galaxies, described in the previous chapter, is not a perfect sample. In particular, the HIMF derived from $\alpha.40$ is subject to biases due to the survey's sensitivity, uncertainties in fluxes and distance leading to uncertainties in masses, and the presence of large-scale structures in the survey volume. On the other hand, however, our detailed understanding and consideration of these biases has led us to a robust determination of the HIMF that takes into account our knowledge of the survey's weaknesses as well as its strengths. One key finding is that the impact of these characteristics is usually different depending on the method uses to calculate the HIMF, i.e. $1/\mathrm{V}_{max}$ or 2DSWML.

Martin et al. [2010] (Chapter 3) provided an overview of important effects. Here, we will discuss in greater detail the magnitude and character of survey properties that influence the final HI mass function. We briefly discussed a correction factor based on the distribution of velocity widths and other systematics related to the survey sensitivity in Chapter 3, and consider that more fully in Section 4.2. Section 4.3 explores various schemes for the assignment of distances, justifying the choice of the mass and mass uncertainty estimation scheme described in Chapter 3. Finally, in Chapter 3, we discussed our chosen scheme for weighting the $1/\mathrm{V}_{max}$ method to account for large-scale structure using the PSCz survey, and here we explore other possibilities for large-scale structure

volume weighting in Section 4.4.

It is particularly important to understand these effects now because we anticipate the '100% ALFALFA survey' to be available in the near future. The large increase in the number of galaxies available for that analysis will decrease the statistical uncertainties on the measurements, thus amplifying the relative impact of systematics and biases. Additionally, at that stage it will be less practical to create thousands of realizations to help understand the various effects. The results presented in this chapter will provide a baseline and dictate procedure for the final measurement of the HI mass function from the completed ALFALFA survey.

4.2 Survey Sensitivity

In the following sections, we consider the ALFALFA source detection scheme and its impact on the final catalog, including the properties of HI-selected galaxies toward which $\alpha.40$ is biased. We also compare our width-dependent sensitivity to the same phenomenon in HIPASS, in order to demonstrate ALFALFA's superior ability to probe the extreme-mass HI-selected galaxy population. The width-dependent sensitivity correction for the $1/V_{max}$ method, described in Chapter 3 [Martin et al., 2010], will be more fully explored. Finally, we will consider the absolute limits of ALFALFA's sensitivity by including both Code 2 sources and the full (18,000 km s^{-1} or z ~ 0.06) redshift extent of the survey. How do these galaxies at the edge of ALFALFA's sensitivity influence the HIMF, and has our previously strict cutoff produced the desired effect, that is, a robust and less-biased HIMF?

4.2.1 ALFALFA Source Detection Scheme

The ALFALFA survey source detection scheme [Saintonge, 2007a] is more robust than a simpler peak-finding algorithm, which, while simple to program and independent of human interaction or bias, will tend to bias a survey toward the brightest and/or narrowest galaxies in the survey volume. As Saintonge [2007a] points out, these algorithms often miss detections that the human eye can pick up in even a casual examination of data cubes. Instead, ALFALFA applies a matched-filtering technique using galaxy templates, which vary in shape as a function of profile width. Source candidates are then visually inspected by members of the ALFALFA collaboration and source parameters are interactively measured.

The adopted scheme results in a reliable catalog of detections that is, nevertheless, dependent on *both* the integrated flux of a given source and the profile width, W_{50} (defined in Chapter 2 as the velocity width of the profile at the 50% flux level). ALFALFA is more sensitive to narrow velocity profiles than to the equivalent integrated flux in a broader profile. Giovanelli et al. [2005a] predicted, from the precursor survey observations, the specific relationship between the integrated flux detection threshold ($S_{int,th}$ in Jy km s^{-1}) and the profile width (W_{50}, in km s^{-1}) of a source in terms of the signal-to-noise ratio required for inclusion in the catalog:

$$S_{int,th} = \begin{cases} 0.15\, S/N\, (W_{50}/200)^{1/2},\ W_{50} < 200 \\ 0.15\, S/N\, (W_{50}/200),\ \ W_{50} \geq 200 \end{cases} \tag{4.1}$$

In practice, the data themselves can be used rather than this prediction. This

is motivated both by a desire to use the observables, when possible, rather than predictions of the performance of the observing equipment and signal extraction pipeline, and by the fact that the ALFALFA survey has since outperformed the prediction. In Figure 4.1, we show the integrated flux detections as a function of profile width, as in Chapter 3, and show fits to the integrated flux threshold to attain strict completeness (red) and a looser criterion for detectability (green). The subsample shown includes only the Code 1 objects in the catalog. Such objects meet the strict requirements for inclusion in $\alpha.40$ source catalogs, i.e., they have SNR > 6.5, a consistent line profile shape, and agreement between the two independently observed polarizations.

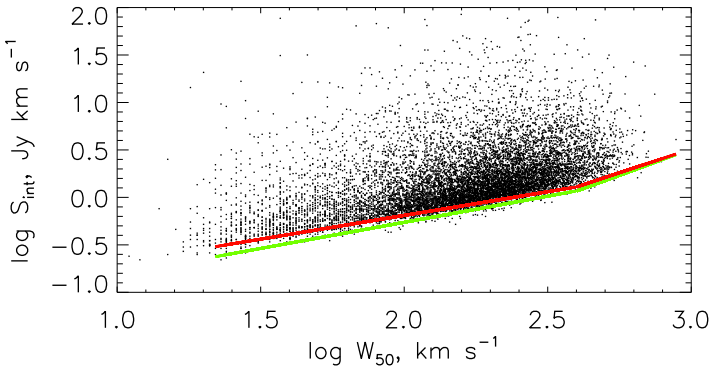

Figure 4.1 The distribution of sources detectable by ALFALFA, dependent on both integrated flux, S_{int}, and profile width W_{50}. Fits to the integrated flux threshold as a function of profile width are overplotted in red (strict completeness limit, necessary for the 2DSWML method) and green (less stringent detection limit, used for the $1/V_{max}$ method).

In Figure 4.2, we show the same information, but instead we plot the Code 2 objects. These catalog entries did not meet the ALFALFA criteria for inclusion in the main catalog (typically, that is because these sources had SNR < 6.5), but were well-matched to an optical galaxy with known redshift corresponding to the redshift of the ALFALFA detections. These objects, therefore, are expected

to have a lower detection threshold, which is seen in Figure 4.2. Excluding these objects from the HIMF analysis guarantees that more confident detections with well-understood selection criteria are used, but may be detrimental if a significant portion of the local HI density is found in such objects and thereby left uncounted. This is discussed further in Section 4.2.4.

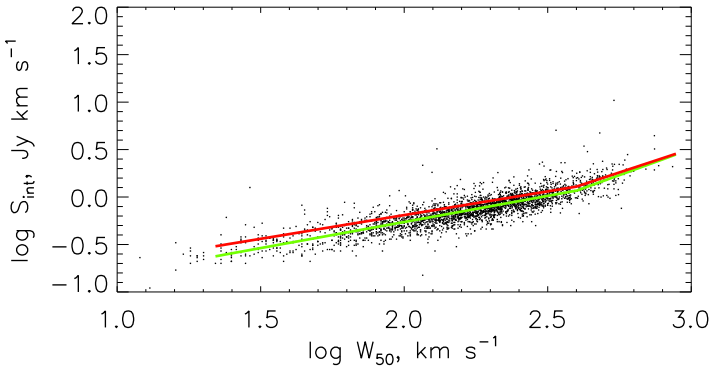

Figure 4.2 The distribution of Code 2 sources in the ALFALFA catalog, plotted with the integrated flux threshold fits described in Figure 4.1.

4.2.2 ALFALFA Sensitivity Compared to HIPASS

The first-generation blind survey that set the standard for future work like ALFALFA was the HI Parkes All-Sky Survey (HIPASS; Meyer et al. [2004]). The HIPASS detection scheme and survey parameters differed from ALFALFA, which benefits from the superior sensitivity and angular resolution of the Arecibo telescope and the new ALFA receiver. HIPASS was thus limited in its ability to probe low-mass and narrow-width objects. The narrowest objects included in the HIPASS catalog have W_{50} = 30 km s^{-1}. In contrast, ALFALFA's velocity resolution is 15 km s^{-1}, after smoothing is applied, and the catalog

includes sources with extremely narrow velocity widths. This improves ALFALFA's ability to probe the lowest HI masses, which in turn robustly constrains the faint-end slope of the HI mass function, α.

A reasonable comparison of the impact of the different source detection schemes (including the absolute level of flux sensitivity) may be made by determining the distribution of the highest-mass galaxies in HIPASS and ALFALFA. The HIPASS detection scheme could bias the catalog against edge-on (extremely wide) profiles at the highest masses, which may explain the finding of Martin et al. [2010] that the HIPASS HIMF underestimated the number density of the highest mass galaxies (Chapter 3).

Such comparisons as Figure 4.3, however, which compares the profile width distributions for HIPASS and ALFALFA galaxies with $M_{HI}/M_{\odot} > 10.0$, show no obvious difference between the profile width distributions of the highest-mass galaxies in $\alpha.40$ and HIPASS. HIPASS does not appear to be particularly biased against objects with very large velocity widths. While the peak-flux threshold detection could introduce such a bias, it is apparent that the matched filtering technique subsequently applied to the HIPASS dataset recovers high-width objects as does the technique used in ALFALFA. Instead, we attribute the lack of extremely high-mass sources in the HIPASS catalog to that survey's limited redshift extent and weakened detection ability near the redshift limit. HIPASS analyses of sample statistics (both the HIMF and the correlation function) include only galaxies with S(cz$_i$) > 0.001 where S is the selection function. This indicates that HIPASS is, at its redshift limit, becoming insensitive. At the redshift limit of $\alpha.40$, which is artificially cut at 15,000 km s^{-1}, the minimum value of the selection function is S(15,000 km s^{-1}) = 0.002. It is likely that this differ-

ence explains ALFALFA's ability to detect statistically significantly more of the highest-mass galaxies.

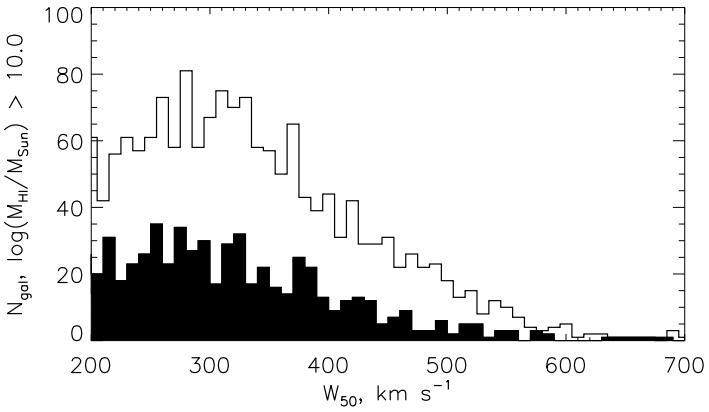

Figure 4.3 The distribution of profile widths in $\alpha.40$ (open histogram) and HIPASS (filled histogram), for objects with $M_{HI}/M_{\odot} > 10.0$.

At the opposite extreme, we consider the HIPASS survey's sensitivity to low mass galaxies. Given that ALFALFA is more sensitive, we expect to detect more low-mass galaxies with weaker fluxes, but our velocity resolution and HIPASS's cutoff at W_{50} = 30 km s^{-1} also add to our ability to probe the lowest mass gas-rich systems. We adopt a cutoff of $\log M_{HI}/M_{\odot}$ = 8.0 to define low-mass; in this regime, the ALFALFA catalog includes 339 galaxies (101 of which would be undetectable to HIPASS based on their narrow velocity width alone), while HIPASS contains 40. Figure 4.4 displays these low-mass systems and their profile velocity widths in the two surveys, demonstrating ALFALFA's ability to probe the lowest mass systems. The $\alpha.40$ HIMF is based on a more populous sample of low-mass galaxies, yet this more robust analysis results in a lower faint-end slope α than reported by Zwaan et al. [2005].

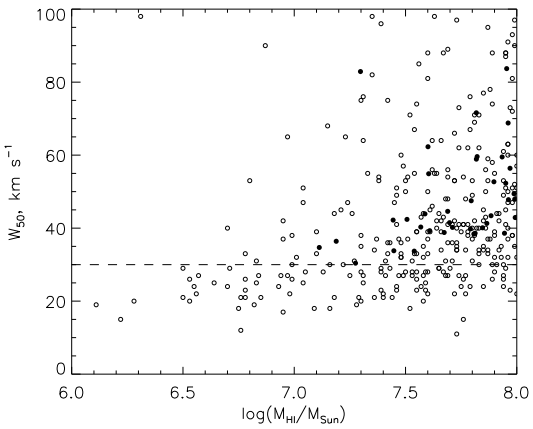

Figure 4.4 The distribution of profile widths W_{50} in ALFALFA (open circles) and HIPASS (filled circles, enlarged for visual clarity), for objects with $M_{HI}/M\odot < 8.0$. The overplotted dashed line shows the profile width cutoff at 30 km s^{-1} for inclusion in the HIPASS catalog.

4.2.3 Accounting for Width-Dependent Sensitivity in the 1/V$_{max}$ Method

In Martin et al. [2010] (Chapter 3), we developed a correction for the ALFALFA detection scheme's bias against galaxies with large velocity widths. Physically, we would not expect there to be a large population of galaxies with wide profiles but small gas masses (observed as fluxes), but the correction does not rely on a model based on this expectation. Rather, the correction takes advantage of the $\alpha.40$ catalog itself, binned by HI mass, to infer the underlying profile width distribution of HI-rich galaxies. This correction is therefore conservative, since we are unlikely to overcorrect high-width galaxies using this scheme. In the sections that follow, we will discuss this correction in greater detail, as well as considering the measured errors on the profile velocity width, which were not included in the previous error analysis for the HIMF.

Width-Dependent Sensitivity Correction

HI profile widths and masses are correlated [Minchin, 2001, Briggs and Rao, 1993], and we use that fact to develop a completeness correction based on observed profile velocity width. This correction is used in the $1/\mathrm{V}_{max}$ derivation of the HIMF, whereas the 2DSWML method includes the bias against large velocity widths in an automatic way. We bin the $\alpha.40$ galaxies by $\log(M_{HI}/M_{\odot})$ and then fit a Gumbel distribution to a histogram of W_{50} in each mass sub-sample (Figure 4.5). The parameters of interest in a Gumbel fit, μ and β, indicate the center of the distribution and the breadth, respectively. We find that μ increases and β decreases linearly with mass, and from the discrete values we derive a relationship between the HI mass and these parameters.

The probability of detecting a galaxy in ALFALFA depends on the mass, the profile width distribution for the bin into which that mass falls, and the limiting profile width $\mathrm{W}_{50,lim}$ beyond which that galaxy would not be detectable by ALFALFA. The correction factor, C, that corrects the count for each galaxy based on its detection probability, is described fully in Appendix B.2.1. The sum over effective search volume, $\Sigma 1/V_{max}$, becomes $\Sigma C/V_{max}$. While this correction has only a modest effect on the HIMF parameters α and $\log(M_*)$, it increases the parameter ϕ_* by 20%, effectively scaling the number density of HI sources in the local universe. In Chapter 3, we assessed the validity of this correction and found that a flux-limited subsample (thus unaffected by the profile width bias) was in agreement with the corrected $\alpha.40$ $1/\mathrm{V}_{max}$ HIMF.

As Figure 4.5 makes clear, the correction for the most populated bins is the most confidently determined, while extrapolation is necessary to correct galaxies falling into the lowest and highest mass bins. By comparing the relationship

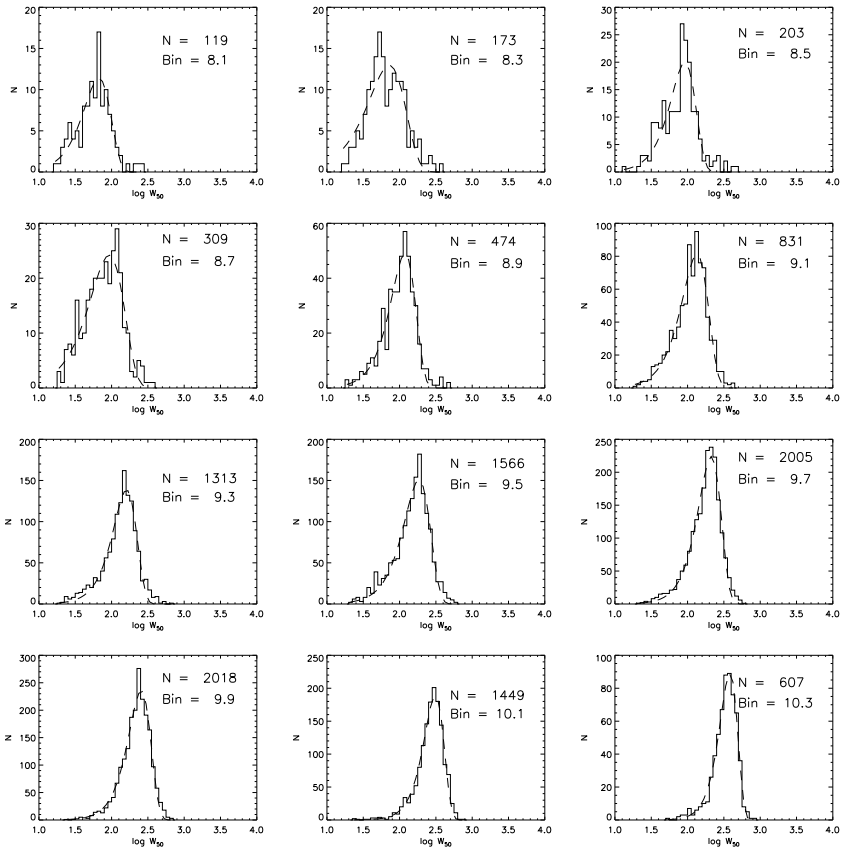

Figure 4.5 The distribution of profile widths $\log W_{50}$ in ALFALFA, binned by HI mass (labeled in the upper-right corner of each box) with Gumbel distribution fits overplotted.

between profile width and galaxy HI mass observed in ALFALFA to that predicted in the O09 cold gas samples from the Millennium Run, we can further test the validity of the profile width correction applied to the $\alpha.40$ sample while also probing O09's ability to reproduce physical characteristics of gas-rich galaxies at z=0.

In Figure 4.6, the relationship between assigned HI mass and line profile velocity width for objects in the O09 sample, taking inclination effects into account, is displayed. This calculation does bring the results of the simulation into closer agreement with the $\alpha.40$ observations, but not completely. This fur-

ther implies that the non-inclined galaxy parameters, i.e. the true modeled parameters from the Millennium Simulation, do not reflect the distribution of true parameters in real galaxies in the local universe. The distribution of O09 simulated galaxies in each mass bin is still narrower than what is observed, and the peak is shifted to a higher width. O09 thus predicts more HI-rich galaxies with large velocity widths and a tighter clustering around a central value dependent on the galaxy's mass. Part of the discrepancy can be explained by ALFALFA's interest in probing galaxies with extremely low HI masses, while O09 is constrained by its mass resolution. We should, therefore, not expect agreement below $\log(M_{HI}/M_{\odot}) \sim 9.0$.

Beyond that, the $\alpha.40$ sample is, certainly, biased toward galaxies with narrow profiles, but the fact that this discrepancy lessens but persists for galaxies with extremely high masses, which are detectable in ALFALFA regardless of velocity width, indicates that the characteristics of the prescription used in O09 to predict gas masses and line widths may not fully represent the complications of real galaxies. Specifically, O09 model the velocity profile for each simulated galaxy by summing contributions from a halo, disk, and bulge component, which may poorly represent the gas distribution particularly in the lowest-mass galaxies in $\alpha.40$, which tend to be irregular rather than disk galaxies.

4.2.4 The Limits of ALFALFA: Code 2 Sources and the Full Redshift Extent

In this section, we will probe those objects at the very limit of ALFALFA's sensitivity. Code 1 objects, with consistent profile shapes, matched polarizations, and

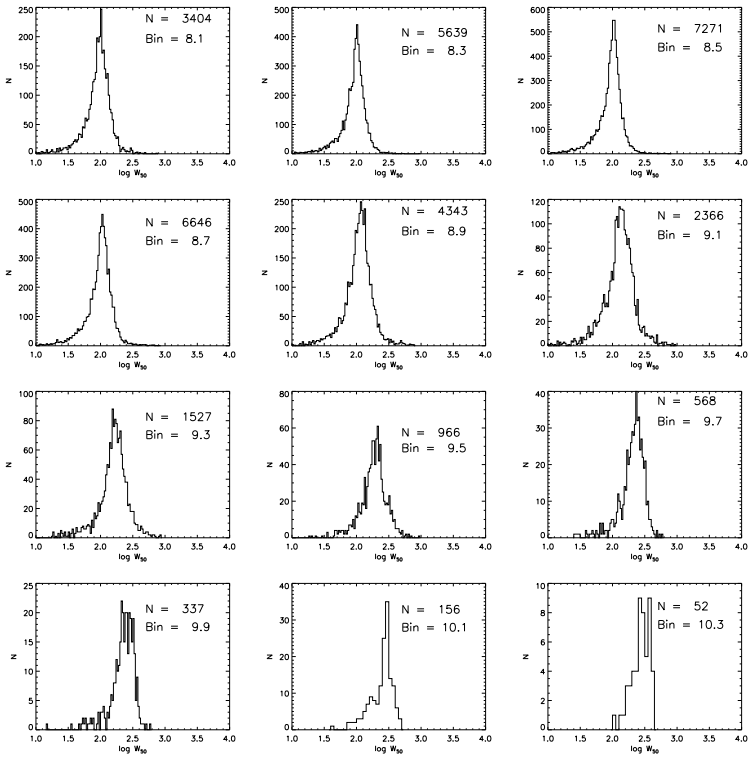

Figure 4.6 The distribution of profile widths $\log W_{50}$, including the effect of inclination of the galaxy with respect to the observer, in the Millennium Run (O09), binned by HI mass (labeled in the upper-right corner of each box).

signal-to-noise parameters above 6.5, have well-described criteria for inclusion in ALFALFA. Code 2 objects, known as "priors", which have signal-to-noise parameters above 4.5 but are matched to an optical counterpart of known redshift, on the other hand, are included in the catalog based on less-strict criteria. We know, however, that the gas in Code 2 objects will also contribute to the overall density in the local universe. Additionally, ALFALFA's sensitivity as a function of distance is strongly affected by radio frequency interference (RFI) near the redshift limits of the survey. For these reasons, Martin et al. [2010] included only objects with Code 1, detected within 15,000 km s^{-1}. We therefore want to consider how the HI mass function is affected when less stringent criteria are applied to the catalog.

Code 2 Sources

Code 2 sources are on the edge of ALFALFA's 6.5 signal-to-noise ratio limit, but are included in the catalog because they coincide with an optical counterpart of known redshift. They are, therefore, biased toward overdensities, toward those regions of the local volume that have been included in targeted or wide-area redshift surveys, such as the Virgo Cluster, and in particular toward those regions of the sky that have been covered in the spectroscopic catalogs of the Sloan Digital Sky Survey (SDSS).

First, we ask whether the exclusion of Code 2 sources in the previous chapter led to an exclusion of a significant amount of energy density in the form of neutral hydrogen. We account for mass and flux errors by creating 500 realizations of an HI mass function that includes sources of both Code 1 and Code 2, and compare those to 100 realizations of the fiducial HI mass function published in Martin et al. [2010] which contained only the Code 1 objects. We use the 2DSWML method, but do not jackknife resample. The inclusion of Code 2 sources increases the sample size from 10,021 to 11,177.

Figure 4.7 displays the HI mass function found when Code 2 sources are included in the analysis. The parameters of the function are not strongly affected by the inclusion of these sources. We find ϕ_* (h_{70}^3 $Mpc^{-3}dex^{-1}$) = 4.8 $\pm$ .3 $\times$ 10^{-3}, $\log(M_*/M_\odot)$ + 2 $\log$ h_{70} = 9.96 $\pm$ 0.02 and α = -1.29 $\pm$ 0.02. These correspond to Ω_{HI} = 4.1 $\pm$ 0.3 $\times 10^{-4}$ h_{70}^{-1} found by integrating the Schechter function, or Ω_{HI} = 4.2 $\pm$ 0.1 $\times 10^{-4}$ h_{70}^{-1} when summing the binned measurements directly. As in the previous chapter, the fiducial HI mass function, which only includes Code 1 objects, finds ϕ_* = 4.8 $\pm$ 0.3, $\log(M_*/M_\odot)$ = 9.96 $\pm$ 0.02, α = -1.33 $\pm$ 0.02, Ω_{HI} (analytical) = 4.3 $\pm$ 0.3, and Ω_{HI} (summed) = 4.4 $\pm$ 0.1, all in the same units

as expressed for the Code 1+2 case.

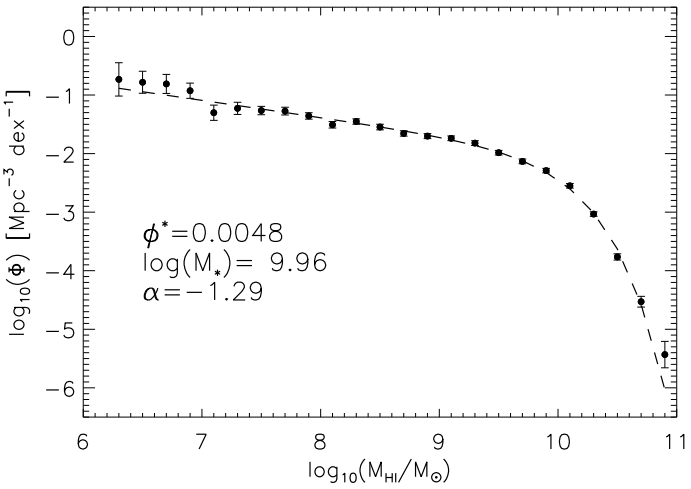

Figure 4.7 The HI mass function found when Code 2 sources are included. The best-fit Schechter function is overplotted as a dashed line, with the best-fit parameters displayed. While Ω_{HI} and the overall Schechter function shape are not changed, the inclusion of the additional sources does slightly flatten the faint-end slope.

Encouragingly, these results indicate that the ALFALFA survey's coding process does not systematically exclude significant sources of HI gas energy density in the local universe. Rather, the agreement between the Code 1 and the Code 1/2 HIMFs suggest that our robust understanding of the survey's sensitivity extends to those sources we identify as Code 2 objects.

The only potentially significant impact is on the faint-end of the HI mass function, influencing both the slope and the points there. The slope parameter α is flattened in the Code 1/2 case, though the two values are just barely within 1σ of each other. In Figure 4.8, we compare the residuals (the best-fit, fiducial HI mass function Schechter model, subtracted from the binned data) for the case where we consider only Code 1 objects (top panel) and the Code 1/2 case (bottom panel). The Figure clearly demonstrates that the Code 1/2 HI mass

function measured fewer low-mass objects per unit volume, which flattens the slope.

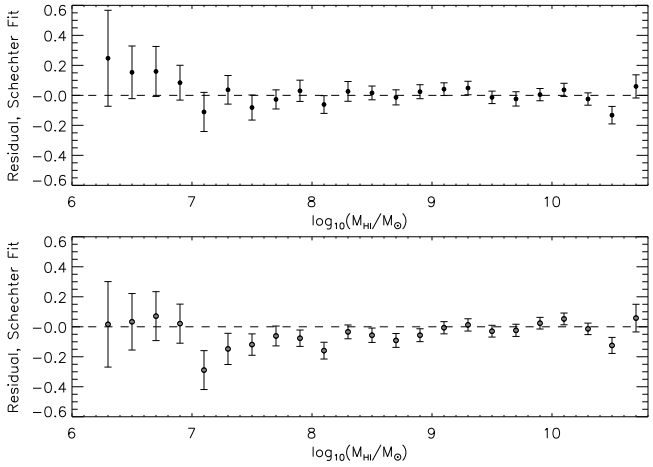

Figure 4.8 Residuals (best-fit Schechter model subtracted from binned data) of HI mass functions calculated using only Code 1 sources (top) and both Code 1 and 2 sources (bottom). In both cases, the comparison model is the fiducial, Code 1-only Schechter function. The zero-residual reference line is overplotted as a dashed line.

Naively, one would expect that the inclusion of the Code 2 objects could only increase the faint-end slope, since it would lead only to the inclusion of more (or the same) number of objects in those bins. However, while this makes sense in the $1/V_{max}$ method where object counts directly impact each bin, the shape of the HIMF is determined in the 2DSWML method via ratios weighted by survey sensitivity, and the impact of a few extra objects in low-mass bins is small compared to the changes in the higher mass bins. Thus, the change in shape in the high and intermediate mass bins causes a decrement in the low mass bins.

Further, the HIMF measured using Code 1 and 2 objects is in very good agreement with the model found using the Code 1 objects only; in fact, at low

masses, its residuals are smaller. This serves as a reminder that while the faint-end slope is extremely interesting, it can be poorly constrained due to small object counts. The inclusion of the Code 2 objects, and the better fit to the Schechter function on the faint end, suggests that exclusion of the Code 2 objects may have led to an overestimate of the faint-end slope. This question will have to be resolved with the full ALFALFA survey, which will contain many more objects, potentially enough to compare the shape of a Code 1 HIMF to that of a HIMF that uses Code 2 objects exclusively.

The Full Redshift Extent of the ALFALFA Survey

Figure 6 of the previous chapter displayed the relative spectral weight within the 40% ALFALFA survey volume as a function of observed heliocentric velocity. A relative weight of 1.0 indicates that the entire surveyed volume was accessible for source extraction and produced high-quality data. As displayed in the Figure, the FAA radar at the San Juan Airport causes a loss of detection ability between 15,000 and 16,000 km s^{-1}. Beyond 16,000 km s^{-1}, where ALFALFA's sensitivity recovers, the survey is sensitive only to the most massive of galaxies and this volume therefore contributes only a small number of galaxies to the overall sample.

For these reasons, the analysis of the HI mass function in the previous chapter neglected galaxies beyond 15,000 km s^{-1}, so that the results would not be influenced by the large spectral weight gap. This exclusion was especially important in the case of the 2DSWML method, since the $1/V_{max}$ method allows the inclusion of explicit corrections for known missing volumes. 2DSWML, by contrast, determines the shape of the HIMF by comparing counts in mass bins

to a built-in description of ALFALFA's flux sensitivity as a function of distance and width. The large gap, which is not anticipated by this approach, may have caused problems in the analysis. It was safer to limit the first measurement of the HIMF to regions where the spectral weights are relatively smooth, that is, to galaxies within 15,000 km s^{-1}.

In this section, we will consider the influence, if any, of including the full redshift extent of the survey in the HIMF analysis. Figure 4.9 displays a Spanhauer diagram for the full redshift extent of ALFALFA, and it is clear that the contribution by galaxies beyond the FAA radar dip will be limited.

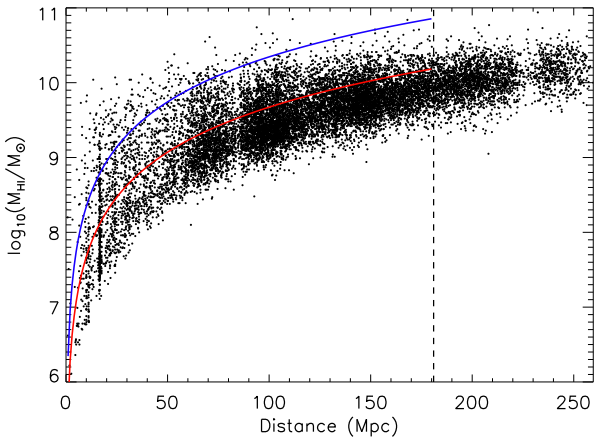

Figure 4.9 Spanhauer diagram for the full redshift extent of the ALFALFA survey, with the gap at 15,000 km s^{-1} or $\sim$ 220 Mpc clearly visible. The vertical dashed line indicates the redshift limit of the HIPASS survey; that survey's sensitivity limit is plotted as a red solid line and the completeness limit as a blue solid line.

In particular, we would expect that the increased bin counts at the very highest masses may increase the statistical significance of our measurement there. This is of interest because Martin et al. [2010] determined that ALFALFA is more sensitive to high-mass galaxies than HIPASS was, with HIMF results indicating that previous blind HI surveys have have missed a significant percentage of the

Table 4.1. HI Mass Function Fit Parameters by Redshift Extent

Sample and Fitting Function	α	ϕ_* (10^{-3} h_{70}^3 Mpc^{-3} dex^{-1})	log ($M_*/M_\odot$) + 2 log h_{70}	Ω_{HI}, fit ($\times 10^{-4}$ h_{70}^{-1})	Ω_{HI}, points ($\times 10^{-4}$ h_{70}^{-1})
$1/V_{max}$, 15,000 km s^{-1} [a]	-1.33 (0.04)	3.1 (0.6)	9.95 (0.05)		4.4 (0.1)
$1/V_{max}$, 18,000 km s^{-1} [a]	-1.34 (0.03)	3.8 (0.6)	9.92 (0.04)		4.3 (0.1)
2DSWML, 15,000 km s^{-1}	-1.34 (0.02)	4.7 (0.3)	9.96 (0.01)	4.3 (0.3)	4.4 (0.1)
2DSWML, 18,000 km s^{-1}	-1.26 (0.02)	3.4 (0.2)	10.00 (0.01)	3.0 (0.2)	3.1 (0.1)

[a] In the $1/V_{max}$ case, pure Schechter functions provide a poor fit to the faint-end slope α, and the sum of a Schechter and a Gaussian function are used to complete the fit. The Gaussian component parameters are not shown in the table, given that they are not expected to be physical.

highest-mass galaxies. Given that the most distant portions of the survey will increase our counts and reduce Poisson errors in the highest-mass bins, this is an interesting problem to consider.

To test this, we calculated the HI mass function using both methods and using all of the Code 1 sources out to 18,000 km s^{-1}. For each method, we created 250 realizations of the survey to account for flux, distance, and mass errors.

The fit parameters and Ω_{HI} values are displayed in Table 4.1 for both the $1/V_{max}$ and 2DSWML methods. It is worth noting that the 2DSWML result is distorted, likely because of the influence of the inaccessible volume and the inability of this method to correct for it. This method drastically underestimates Ω_{HI}, shifts $\log (M_*/M_\odot)$ to a higher value, and flattens out the low-mass slope α. On the other hand, the $1/V_{max}$ method continues to function as expected and to result in a reliable measurement. Both of the $1/V_{max}$ results are consistent with each other, including the measured values for Ω_{HI}, but the 2DSWML method performs poorly.

This provides further evidence of the relative strengths and weaknesses

of the two available methods for estimation of the HI mass function. While 2DSWML provides a powerful statistical tool, it functions as a 'black box' method that cannot be manipulated by additional knowledge of the survey. In some cases, this may be an advantage, but in the case of ALFALFA where we have detailed information about the survey volume, the survey sensitivity, and other factors contributing to the HIMF, the $1/V_{max}$ method provides a clear path and an understandable answer.

4.3 Mass Uncertainties

On the low-mass end, mass uncertainties are the primary source of error on the HI mass function, and the ALFALFA error analysis has taken this explicitly into account, unlike the HIPASS survey [Martin et al., 2010]. Because the HIMF is based on binning galaxies by mass and then considering each bin as an independent data point, it is not straightforward to carry mass uncertainties through analytically. Instead, the ALFALFA HIMF's uncertainties due to mass errors are calculated through the creation of many hundreds of realizations, each with randomly assigned mass (i.e., distance and flux) errors. In the following sections, we elaborate further on the distance estimate scheme used in ALFALFA, the biases that would be introduced by using alternative schemes (i.e., a pure Hubble flow model) and the overall impact of distance and flux errors on the mass estimates used to construct the HIMF.

4.3.1 Distance Estimates as a Source of Mass Uncertainty

The following analysis makes use of additional data not available in the previous chapter. In addition to the regions used for the initial HIMF estimate, we will hereafter use an additional region in the anti-Virgo direction ($22^h < \alpha < 03^h$, $14° < \delta < 16°$). Once again, the sample is limited to those galaxies that are Code 1 and within 15,000 km s^{-1}.

Distance Estimates in ALFALFA

The distance estimation scheme in ALFALFA was described in the previous chapter [Martin et al., 2010]. When distances are based on the flow model, we employ the model's error estimates, constrained by the fit of the model to the observed velocity fit. When distances are estimated using pure Hubble flow, the error is estimated to be $\sim$10%. We fix a minimum error of 163 km s^{-1}, based on the local velocity dispersion measured by Masters [2005]. To demonstrate the importance of using the full suite of available information to estimate distances, Figure 4.10 compares the primary distances (used in $\alpha.40$) to the values that would be obtained assuming pure Hubble flow.

Out of 11,547 galaxies, primary distances from the literature are used to estimate the distance to only 40. The remainder are estimated using Hubble flow (68%), the flow model (25%), or information on group membership (6%).

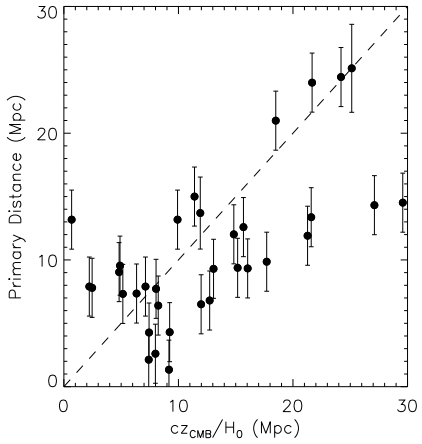

Figure 4.10 Primary distances from the literature vs. estimates based only on pure Hubble flow, with the ALFALFA distance uncertainty estimates overplotted. The dashed line indicates a one-to-one correlation.

Alternative Distance Estimates

In their estimate of galaxy masses for the HI mass function, HIPASS assumed Hubble flow. This is not a safe assumption, particularly in the regions of the sky surveyed in $\alpha.40$. The Virgo Cluster, in particular, represents a strong deviation from any assumed relationship between distance and recessional velocity. Masters et al. [2004] showed the danger of assuming pure Hubble flow, especially because of the small redshifts accessible to blind HI surveys. The authors concluded that the low-mass slope of the HIPASS HIMF was underestimated due to neglecting peculiar velocities, and predicted that a survey in the direction of Virgo could severely underestimate the low-mass slope.

Given the large-scale structure in the $\alpha.40$ volume, we would epect the HI mass function to vary strongly if a poor choice of distance estimate were made. In order to test this, we have re-calculated the 2DSWML estimate of the HI mass function using pure Hubble flow to estimate distances. That is, we con-

verted the observed heliocentric velocities into the CMB rest frame, and then assumed $D_{Mpc} = cz_{cmb}/H_0$, where we adopt the ALFALFA standard H_0 = 70 km $s^{-1}Mpc^{-1}$. In this case, we have no ideal estimate of the distance error, and therefore use 10% of the Hubble flow distance or the local dispersion value 163 km s^{-1}, whichever is greater. As usual, flux errors are also folded into the mass uncertainties. Once again, we create 250 realizations to estimate uncertainties.

The resulting HI mass function and Schechter fit parameters are displayed in Figure 4.11. As anticipated [Masters et al., 2004], the use of Hubble flow has caused a serious underestimate of the faint-end slope α. ALFALFA's success at robustly measuring the HIMF has depended not only on large sample size over a cosmologically significant volume, but also on the selection of a reasonable model for distance estimation.

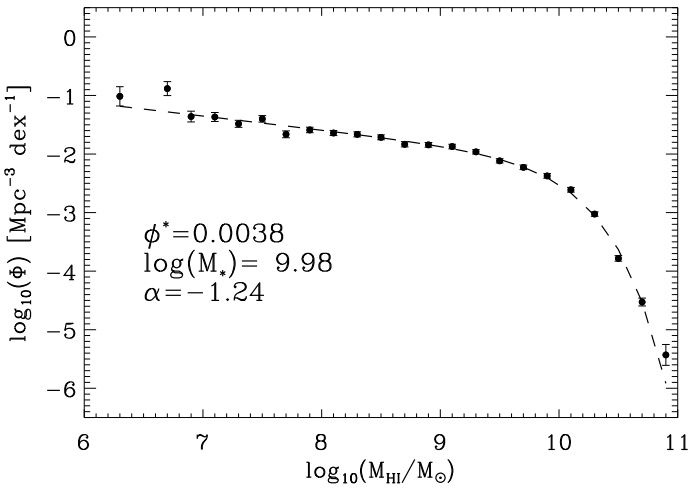

Figure 4.11 The HI mass function obtained when distances, and therefore masses, are obtained assuming pure Hubble flow with H_0 = 70 km $s^{-1}Mpc^{-1}$.

4.3.2 Impact of Distance and Flux Errors on Mass Estimates

Given Masters et al. [2004] emphasis on distance errors and their large impact on the HI mass function and its uncertainties, it is reasonable to question the overall scale of the mass uncertainties in ALFALFA. How large are the mass errors when both distance and flux errors in the $\alpha.40$ sample are taken into account?

To obtain robust estimates of mass errors, we created many thousand realizations of each galaxy in the $\alpha.40$ sample and applied distance and flux errors. The result, displayed in Figure 4.12, compares the HIMF mass bin galaxies would nominally fall into assuming a perfect measurement of distance and flux (along the abscissa) to the mean mass of the galaxies assigned to that bin once realistic uncertainties are taken into account. The horizontal uncertainties indicate the 1σ spread of potential 'true' masses falling into nominally assigned mass bins. From the Figure, it is clear that ALFALFA's measurement of the HIMF and Ω_{HI} is not prone to large uncertainties above $10^{9.0} M_{\odot}$. In the mass range of interest to the missing satellites problem, dwarf galaxy studies, and the low-mass slope of the HIMF, that is below $10^{8.0} M_{\odot}$, galaxies can easily be mis-assigned to bins, even when a realistic distance model is being used. Depending on the large-scale structure in the survey volume, this effect would lead to either an over- or under-estimate of α.

4.4 Large-scale Structure

Because blind HI surveys are relatively shallow, with ALFALFA probing the local Universe only out to z $\sim$ 0.06, inhomogeneity in the survey volume has a

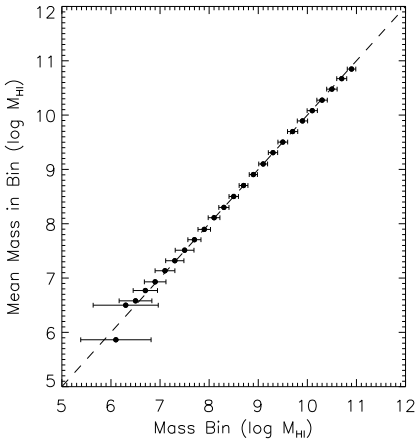

Figure 4.12 The average (mean) mass falling into each HIMF bin. The estimated 1σ uncertainty of a galaxy's HI mass is overplotted as error bars, along with a dotted line indicating a one-to-one relationship.

strong impact on the derived HI mass function from $\alpha.40$. This is particularly true in the case of the $1/V_{max}$ method, which is not as robust against large-scale structure, but the 2DSWML method isn't completely immune from these effects. To test the homogeneity of a sample, the usual statistical test applied is the V/V_{max} test [Schmidt, 1968]. Much like the $1/V_{max}$ method, this test considers the maximum volume out to which each source in a survey can be detected. By comparing the actual volume the source was detected in to the accessible volume, homogeneity in the sample can be evaluated; the expectation value $< V/V_{max} >$ is 0.5 in a homogeneous volume.

In the case of the $\alpha.40$ volume, $< V/V_{max} >$ = 0.45. This indicates that, at 40% completion, the survey does not yet contain enough volume to fully 'smooth out' the effects of large-scale structure. This is reflected in Figure 4.13 where $< V/V_{max} >$ is shown for each mass bin. The most obvious feature in this Figure, the dip near $\log (M_{HI}/M_{\odot})$, is due to overdensities in the sample volume, primarily the Virgo cluster. Galaxies in those overdensities are found, prefer-

entially, in those regions, rather than filling the full volume where ALFALFA's sensitivity could detect them, causing this dip.

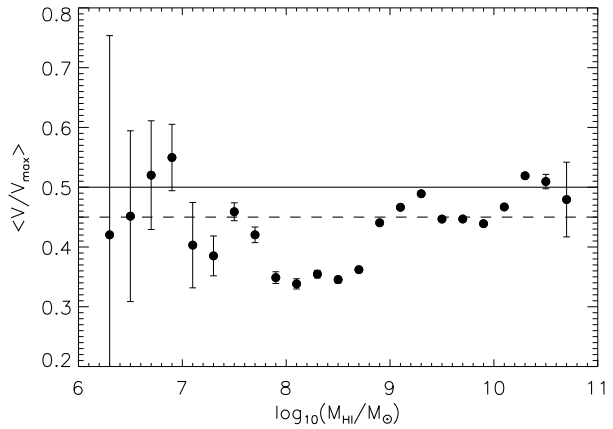

Figure 4.13 The typical (mean) value of V/V$_{max}$, binned by masses. Error bars are Poisson counting uncertainties. The solid line indicates $< V/V_{max} >$ = 0.5, while the dashed line indicates $< V/V_{max} >$ = 0.45 for the $\alpha.40$ sample.

It is clear that $\alpha.40$ does not, yet, constitute a fully representative slice of the Universe; as the survey progresses, we anticipate that the full sample will pass the V/V_{max} test. Another, perhaps more intuitive, way to view the impact of voids and clusters in $\alpha.40$ is to compare the redshift distribution of cataloged galaxies to the prediction based on the survey's selection function (i.e., the percentage of galaxies at a given distance that are detectable in ALFALFA). The selection function is determined by the 2DSWML analysis of the HIMF, and when combined with the measurement of the HI mass function, predicts the redshift distribution for a homogeneously distributed set of galaxies selected from the HIMF.

Figure 4.14 compares this expectation to the observations in $\alpha.40$. The bumps and dips in the histogram represent under- and overdensities, respectively, in the survey volume. For example, the Virgo Cluster explains the enhancement

near 1,000 km s^{-1}. The Pisces-Perseus supercluster and its foreground void also make clear imprints on this figure.

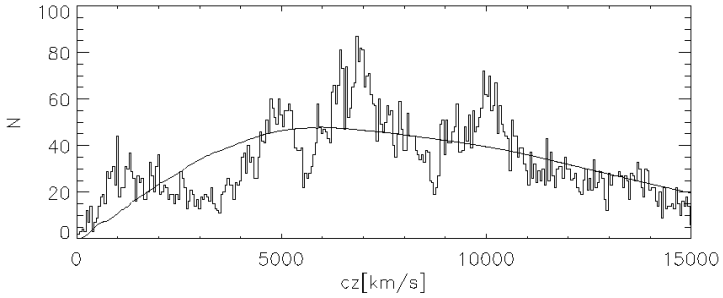

Figure 4.14 The observed redshift distribution of $\alpha.40$ galaxies (histogram) compared to the expected distribution obtained via the survey's selection function.

If $\alpha.40$ does not represent a representative sampling of the Universe, then statistical studies of the sample's characteristics, like the HI mass function, may be subject to biases from large-scale structure. In the following sections, we further explore the impact of structure on the HI mass function derived from the current ALFALFA dataset. First, we consider sub-regions in $\alpha.40$, then we explore the large-scale structure corrections employed in the last chapter given that they depend on datasets other than our own.

4.4.1 Subregions of $\alpha.40$

The $\alpha.40$ sample is made up of 3 large, contiguous areas, as shown in the cone diagrams in Chapter 3. In the Northern Galactic hemisphere, $\alpha.40$ covers $07^h30^m < \alpha < 16^h30^m$ in two chunks, $4^\circ < \delta < 16^\circ$ and $24^\circ < \delta < 28^\circ$. We refer to these subregions as $\alpha.40.North1$ and $\alpha.40.North2$, respectively. In the Southern Galactic hemisphere, $\alpha.40$ covers $22^h < \alpha < 03^h$, $24^\circ < \delta < 32^\circ$, referred to

as $\alpha.40.South$. The entire $\alpha.40$, combined together, covers enough cosmological volume for the effects of large-scale structure to become minimal, but these individual regions are strongly affected by over- and under-densities within their volume.

Figures 4.15, 4.16, and 4.17 display the HIMF for $\alpha.40.North1$, $North2$, and $South$, respectively. The fit parameters and values of Ω_{HI} are shown in Table 4.2 along with the fiducial 2DSWML HIMF for the entire $\alpha.40$ sample for comparison. The largest by a significant fraction is $\alpha.40.North1$, and it contributes over 50% of the 10,000 galaxies in $\alpha.40$. As expected, the HIMF for this region, when isolated, follows the HIMF for the sample as a whole. Because of the large volume in this region, the HIMF displayed in Figure 4.15 is also smooth and featureless.

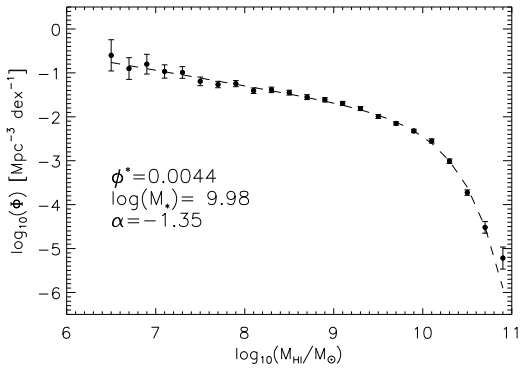

Figure 4.15 The HIMF estimated via the 2DSWML method in the $\alpha.40.North1$ region, with Schechter fit parameters.

In the case of the smaller samples in Figures 4.16 and 4.17, features due to large-scale structure are clearly visible. Because of the inhomogeneity of the surveyed volume, the HIMFs don't follow the prescribed Schechter function. In the case of $\alpha.40.North2$, the faint-end slope is better fit on its own, in which case

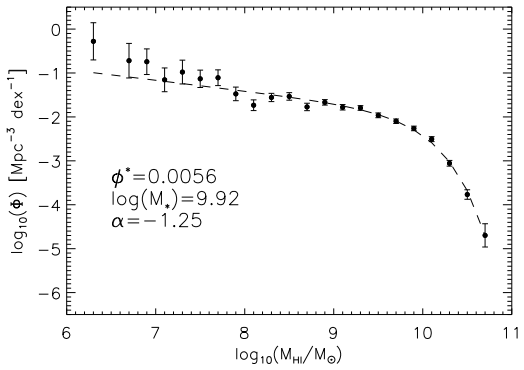

Figure 4.16 The HIMF estimated via the 2DSWML method in the $\alpha.40.North2$ region, with Schechter fit parameters.

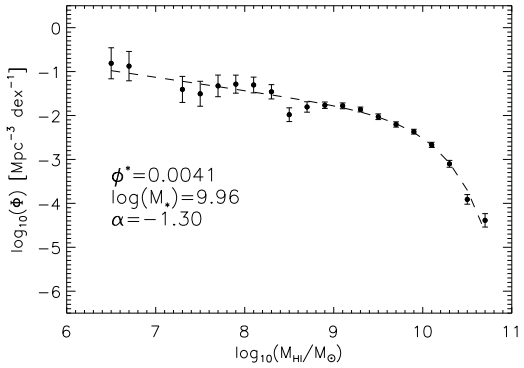

Figure 4.17 The HIMF estimated via the 2DSWML method in the $\alpha.40.South$ region, with Schechter fit parameters.

it is measured to be α = -1.4 $\pm$ 0.1.

In every case, the 'bumps' and wiggles in the sub-region HIMFs correspond to the cone diagram distributions in Martin et al. [2010]. In essence, the combination of the survey's sensitivity and the scaling of survey volume with redshift leads to preferred distances for each of the mass bins in the HIMF (or preferred masses for every distance in the survey). A dip, for example, in the HIMF corresponds to an overdensity at the preferred distance for those mass scales. While

Table 4.2. 2DSWML HIMF Schechter Parameters by Region

Sample and Fitting Function	α	ϕ_* (10^{-3} h_{70}^3 Mpc^{-3} dex^{-1})	log ($M_*/M_\odot$) + 2 log h_{70}	Ω_{HI}, fit ($\times 10^{-4}$ h_{70}^{-1})	Ω_{HI}, points ($\times 10^{-4}$ h_{70}^{-1})
North1	-1.35 (0.02)	4.4 (0.3)	9.98 (0.02)	4.3 (0.4)	4.4 (0.1)
North2	-1.25 (0.04)	5.6 (0.6)	9.92 (0.02)	4.2 (0.5)	4.3 (0.2)
South	-1.30 (0.04)	4.1 (0.5)	9.96 (0.3)	3.6 (0.5)	3.5 (0.2)
Whole $\alpha.40$	-1.34 (0.02)	4.7 (0.3)	9.96 (0.01)	4.3 (0.3)	4.4 (0.1)

the 2DSWML method has been designed to be less sensitive to large-scale structure, the volumes of these subregions are too small for these effects to average out.

Such techniques can only work with the data they are given, but the $1/V_{max}$ approach allows for explicit correction for known structures. When these corrections are included in the $1/V_{max}$ analysis of these subregions, the (unshown) results are very similar to those provided here. These corrections are based on the PSCz density correction (see Section 4.4.2), but imperfections in this correction lead to the same bump and dip features. An additional weakness of the $1/V_{max}$ density correction is that the counts can only be increased for galaxies that do end up in the sample, making the correction significantly less useful in voids. By contrast, 2DSWML essentially 'self-corrects' for over- and underdense regions. Rather than looking at volumes and scaling counts by $1/V_{max}$, 2DSWML constructs the relationship between bins by scaling the counts themselves and therefore automatically scales the HIMF downward for regions that are overdense and upward for regions that are underdense.

This consideration of subregions within $\alpha.40$ makes clear the impact that large-scale structure can have on blind HI surveys and the importance of cosmologically significant volumes before drawing global conclusions.

4.4.2 Large-scale Structure Correction from Previous Surveys

As described in Martin et al. [2010], the $1/\mathrm{V}_{max}$ method can be corrected to account for large-scale structure in the survey volume. Essentially, overdense regions are considered to represent more effective volume ($\Sigma 1/V_{eff}$, rather than $\Sigma 1/V_{max}$) and vice verse for underdense regions, so that galaxies in various environments are weighted appropriately.

While this correction is successful, it does rely on datasets outside of the ALFALFA survey. In the case of the previous chapter, a IRAS Point Source Catalog redshift (PSCz; Branchini et al. [1999]) density map was used to correct for large-scale structure. However, other options exist, in particular other PSCz maps (smoothed to different levels) and the 2MASS Redshift Survey (2MRS; Erdoğdu et al. [2006]) density reconstruction map. The large-scale structure correction used has a large influence on the final HI mass function estimate; a $\sim$ 20% effect on the Schechter parameters, compared to neglecting the density correction, was reported in Martin et al. [2010]. Given the magnitude of the effect, it is important to consider the impact that a different choice would make. In particular, since this portion of HIMF analysis is likely to always rely on external information, examining it here may be helpful in the future for the 100% ALFALFA sample.

The parameter of interest reported by PSCz and 2MRS is the overdensity δ, defined relative to the average number density of galaxies found in those surveys:

$$\delta = \frac{n - \bar{n}}{\bar{n}} \tag{4.2}$$

In the case of the ALFALFA survey and the HIMF, we are primarily interested in the average value of δ interior to each source's maximum detectable distance or, in other words, the average value of δ in the volume over which the source could have been detected. Both the 2MRS and PSCz density maps report the value of δ in equal-volume cells throughout their survey volume.

Figure 4.18 displays the average interior overdensities from the density reconstructions published by PSCz and 2MRS. The overall structure is dominated by the galaxies in the Virgo direction, as that portion of the survey dominates $\alpha.40$ in both source counts and total area on the sky. In both cases, the average tends toward 0 as larger and larger volumes are included. The peaks near 0 Mpc indicate large-scale structure in the Local Supercluster overdensity, such as the Virgo cluster.

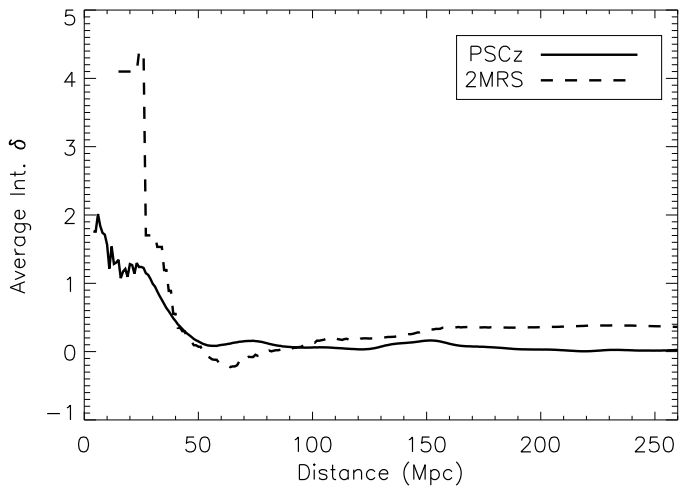

Figure 4.18 The average interior overdensity δ as a function of distance in the $\alpha.40$ survey region, obtained by the density reconstructions from the PSCz (solid) and 2MRS (dashed) surveys.

It is clear from the plot that the 2MRS reconstruction indicates a dramatic overdensity at small volumes, and this is, in part, why the PSCz reconstruction

was chosen for Martin et al. [2010]. The 2MRS overdensity was an estimate based on only one point in the entire ALFALFA volume and was obviously a poor estimate for the survey's average density given the void in the anti-Virgo portion of the survey. The 2MRS reconstruction also did not include estimates of the density for the most nearby objects in ALFALFA. Given that the lowest-mass objects in ALFALFA, which determine the faint-end slope α of the HIMF, are only found nearby, this made the PSCz map preferable.

The PSCz reconstruction was also preferable because of its agreement with an independent analysis of the number density field within the ALFALFA volume. Saintonge [2007b] has used a reference sample of galaxies in the Sloan Digital Sky Survey (SDSS) to estimate the average density in the ALFALFA Virgo-direction volume; those results are in general agreement with the PSCz findings, although a full analysis is forthcoming. Because the absolute magnitude and surface brightness limits of SDSS fall off dramatically with distance, Saintonge [2007b] limited the reference sample to 80 Mpc. The average interior δ for this reference sample is displayed in Figure 4.19.

While PSCz was a good choice for the analysis of the $\alpha.40$ HIMF, there are actually several choices of maps available from Branchini et al. [1999], with the primary differences being the smoothing size of each volume cell and the maximum distance out to which the density fields were reconstructed. In the previous chapter, the chosen map extended to 240 h^{-1} Mpc and was smoothed with a Gaussian kernel of width 3.2 h^{-1} Mpc. The alternative options include a map that extends to only 120 h^{-1} Mpc with a 3.2 h^{-1} Mpc kernel, and one that extends to 240 h^{-1} Mpc with a larger Gaussian kernel of 7.7 h^{-1} Mpc. The smoothing scale of PSCz maps can lead to underestimation of density contrasts. Because

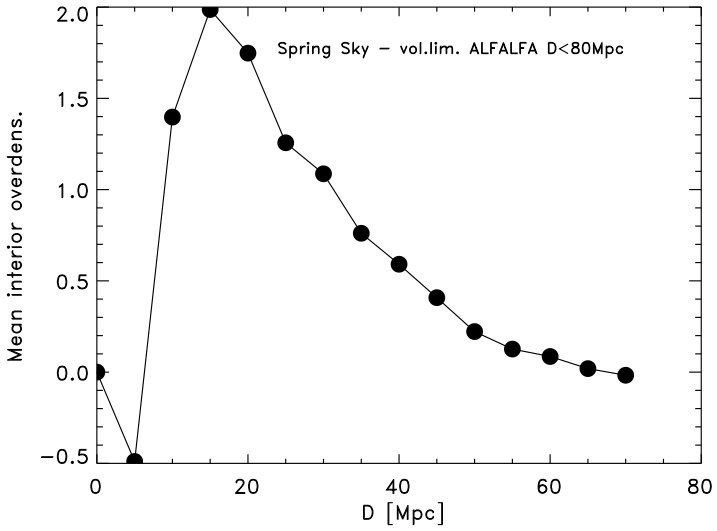

Figure 4.19 The average interior overdensity δ as a function of distance in the Virgo-direction (spring) $\alpha.40$ survey region, obtained by Saintonge [2007b].

the primary effect of the large-scale structure correction is on the lowest-mass bins of the HIMF, it is important to explore and understand the influence of this outside dataset.

In Figure 4.20, we display the average interior δ for each of these density maps, truncated at 150 Mpc for visual clarity. The PSCz.240.G3.2 map used in the previous chapter for the $1/V_{max}$ analysis of the HIMF represents an extreme estimate of the impact of large-scale structure within the survey volume. It is the most conservative option, given that it attaches lower weight to those galaxies found in nearby overdensities, particularly the Virgo cluster, to prevent them from artificially boosting the faint-end slope.

In order to quantify the effect of these options on the resulting HIMF, we use PSCz.120.G3.2 and PSCz.240.G7.7 to re-analyze the $\alpha.40$ HIMF. In the case of PSCz.120.G3.2, which does not reach the full redshift extent of the $\alpha.40$ sample,

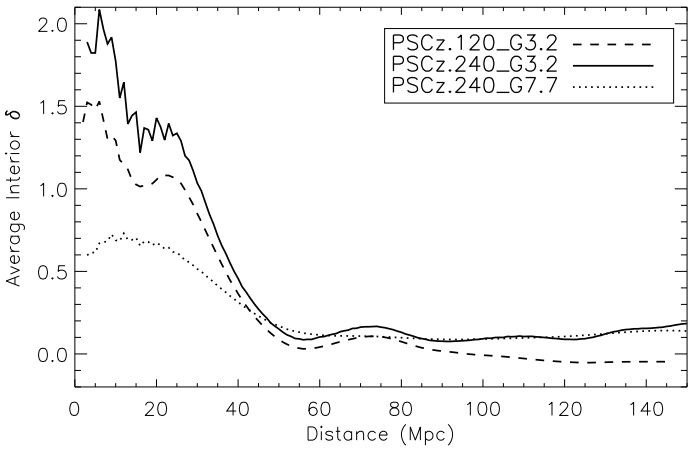

Figure 4.20 The average interior overdensity δ as a function of distance in the $\alpha.40$ survey region, obtained via three different density reconstruction maps from the PSCz survey. The solid line represents the PSCz map used in the previous chapter for the measurement of the HIMF, which extends to 240 h^{-1} Mpc and is smoothed with a Gaussian kernel of 3.2 h^{-1} Mpc, indicated with the label PSCz.240.G3.2. The dashed and dotted lines, respectively, indicate maps with smaller extent and a larger smoothing kernel.

we set the average interior δ to 0 for galaxies beyond the distance limit for that map. In order to fit Schechter function parameters in each case, we use the same uncertainty estimates for each mass bin point as presented in the previous chapter, as the PSCz map applied would only change the magnitude of each point and not its fluctuation due to mass uncertainties.

Figure 4.21 shows the results, focusing on the low-mass end of the HIMF, since mass bins with $M_{HI} > 10^{8.0} M_{\odot}$ are not affected by the large-scale structure volume correction. The different large-scale structure corrections function effectively as a scaling in each bin, so that each option follows the fiducial case closely. Both PSCz.120.G3.2 and PSCz.240.G7.7 boost the faint-end slope, indicating that they are overcounting galaxies in the nearby overdensities, namely the Virgo cluster.

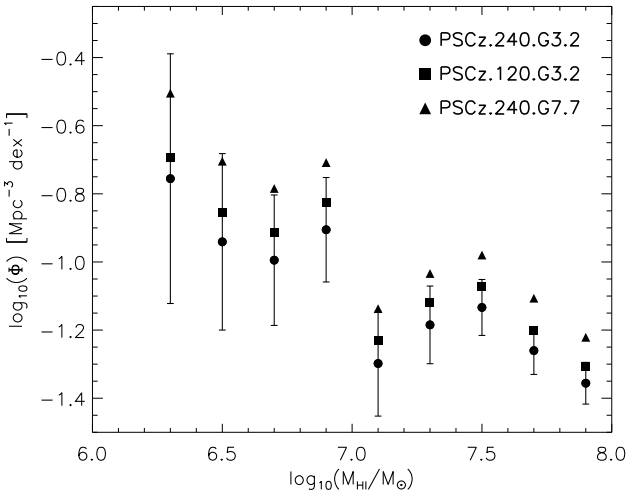

Figure 4.21 The low-mass end of the HIMF, showing dependence on the chosen PSCz density reconstruction map. The fiducial $1/V_{max}$ HIMF reported in the previous chapter is shown as a filled circle, with two other maps represented by squares and triangles.

This verifies that PSCz.240.G3.2 was the most conservative choice for correcting the $1/V_{max}$ HIMF for the effects of large-scale structure. The changes to the low-mass slope α and the turnover mass M_* are displayed in Table 4.3, along with the measured 2DSWML parameters for reference. It is clear that the PSCz map with the greatest extent and the smallest smoothing radius is most appropriate for estimating the $\alpha.40$ HI mass function.

4.5 Summary

In this chapter, we have considered various biases and systematic and statistical effects that can influence the outcome of a HIMF measurement. By probing the $\alpha.40$ sample in detail, the eventual '$\alpha.100$' sample provided by the full and complete ALFALFA survey can best be exploited, so that optimal results can be

Table 4.3. $1/V_{max}$ HIMF Schechter Parameters by PSCz Map

PSCz Map	α	log ($M_*/M_\odot$)
2DSWML Result	-1.33 (0.02)	9.96 (0.02)
PSCz.240.G3.2	-1.33 (0.03)	9.95 (0.04)
PSCz.120.G3.2	-1.39 (0.03)	9.96 (0.05)
PSCz.240.G7.7	-1.44 (0.04)	9.98 (0.06)

achieved quickly and efficiently.

The ALFALFA detection scheme is found to be more capable than HIPASS of probing extreme-mass systems, and to differ slightly for Code 1 compared to Code 2 sources. A comparison to the distribution of cold gas sources in O09 showed that the Millennium Run provides only a limited basis for comparing HI observations to simulated gas reservoirs. Finally, we determined that the inclusion of Code 2 sources and the full redshift extent of the survey has only a major effect on the HIMF, although the 2DSWML method is not well-equipped to handle the missing volume between 15,000 and 16,000 km s^{-1}.

Section 4.3 probed the impact of mass uncertainties on the HIMF. This analysis confirms that using a pure Hubble flow model to determine distances severely biases the HI mass function, and in particular the faint end slope. We also showed the magnitude of the influence of mass errors on the final estimate of the HI mass function. Throughout this chapter and the previous chapter, it has been made clear that a careful treatment of mass (and therefore flux and distance) uncertainties is absolutely critical for a robust determination of the HI mass function.

Finally, we more fully explored the control and correction of large-scale structure effects in the analysis of the HI mass function (Section 4.4). The $\alpha.40$ sample approaches homogeneity but is not a perfectly representative sample of the nearby volume. It is anticipated that as the ALFALFA survey continues, this situation will improve along with estimates of the HIMF. We demonstrated the influence that large-scale structure has on the HI mass function derived from various subregions of the $\alpha.40$ survey region. For the $1/V_{max}$ method, these findings motivated the use of a large-scale structure correction, and we have determined that the PSCz.240.G3.2 density reconstruction map provided the best estimate of large-scale structure within the ALFALFA survey volume.

CHAPTER 5

DEPENDENCE OF THE HI MASS FUNCTION ON GALAXY ENVIRONMENT

5.1 Introduction

The variety of gas processes, present to lesser and greater extent depending on environment, result in an HI mass function that varies with local density. Many observational results confirm that environment-dependent processes influence galaxies and, in particular, their gas reservoirs, but ALFALFA provides the first blindly HI-selected sample that is large enough to begin to probe this dependence with statistical significance. First, we provide an overview of what is already known about the impact of environment on the general HI characteristics of galaxies.

5.1.1 Clusters

The Virgo cluster's richness and proximity has made it a classic target for studies of the impact of a high-density environment on galaxy properties, and observations of HI deficiency and disk truncation indicate that HI is strongly influenced by the processes in this kind of environment. Much of what we understand about HI in clusters comes from detailed studies of this particular 'laboratory.'

HI deficiency is defined in terms of a cluster galaxy's morphology and gas mass, compared to the typical gas mass of a field galaxy with the same morphological structure [Haynes and Giovanelli, 1986a, Solanes et al., 2001] and there-

fore depends on the availability of large samples of normal field galaxies with known gas masses; given the availability of ALFALFA's sample, these HI standards may need to be adjusted [Toribio et al., 2011a,b]. Spirals near the Virgo core contain less cold hydrogen gas than their field counterparts, by factors of $\sim$ a few, due primarily to ram-pressure stripping. HI deficiency correlates with the distance from the center of the cluster [Chung et al., 2008], but continues at distances where the ICM density cannot ram-pressure strip gas, suggesting that tidal effects or a lumpy ICM may also contribute to deficiency. Such 'suburban' galaxies may have also been stripped of their gas during earlier encounters with other cluster members or passage through the hot ICM at an earlier epoch. More general work on the cluster environment, extending beyond the cores of clusters, indicates that extreme-environment processes like ram-pressure stripping cannot be accountable for all of the environmental variation of such properties as star formation rate [Lewis et al., 2002]. Alternatively, distance errors may be partially to blame for the existence of gas-deficient galaxies, making them only appear to be on the outskirts of the Virgo cluster [Sanchis et al., 2004].

Furthermore, the HI disks of galaxies close to the Virgo core are truncated compared to their optical disks [Chung et al., 2009, 2008], and are often displaced from the optical center of the galaxy, indicating an ongoing stripping event. There are indications that HI deficiency is accompanied by molecular hydrogen (H_2) deficiency, potentially impeding the star formation potential of such galaxies [Fumagalli et al., 2009]. Truncated disks in Virgo can show ongoing star formation near the center of the disk, with a passively evolving stellar population on the galaxy's outskirts [Crowl and Kenney, 2008].

5.1.2 Groups

The ALFALFA survey region includes many groups, and special attention has thus far been paid to the medium-density Leo Group environment [Stierwalt et al., 2009]. Leo I was found to be dominated by low-mass HI-rich galaxies, and the number of sources found in ALFALFA's blind search in the 21 cm line suggests the power of this strategy for probing the group environment. The HI mass function in the group was found to be steeper than global estimates, and therefore in better agreement with cold dark matter simulations of galaxy formation. By contrast, Pisano et al. [2007] studied nearby Local Group analogs and determined that the typical HI mass function in these medium-density environments was flat across several orders of magnitude.

In the Local Group, HI content is known to correlate strongly with group-centric distance; gas-poor dwarf spheroidals are found near massive galaxies, while the gas-rich dwarf irregulars are more widely distributed [Grebel, 2005]. While this morphological segregation inspires tantalizing questions about the role of environment, it also makes blind HI surveys particularly well suited for probing the group environment at all distances from the massive, dominant group members. Koribalski et al. [2004] also found that some new groups that were obvious in their HI content had not previously been discovered in optical surveys, likely due to the faintness of optical galaxies within them. Given the different biases of HI and optical surveys, it is likely that a combination of them would best identify groups in the nearby universe.

5.1.3 Voids

Voids are of particular interest to ALFALFA because of their potential as 'pristine' environments, in which galaxies could have evolved independently of major interactions and mergers with neighbors. It was long suspected that HI selected galaxies would, generally, trace the mass distribution of more massive galaxies despite HI deficiency in clusters; that is, there is no expected analogous HI enrichment in low-density environments such as voids [Weinberg et al., 1991]. While HI-selected dwarf galaxies do prefer underdense regions [Basilakos et al., 2007] they avoid the most extreme end of the distribution, the true voids. This may be due to their particular susceptibility to photoheating during reionization [Hoeft et al., 2006, Hoeft and Gottloeber, 2010], an area of active research in solving the 'void problem'. The Void Galaxy Survey (VGS) project has reported indications that void galaxies in their sample may currently be accreting cold gas along filaments [van de Weygaert et al., 2011]. Other factors may contribute, and little is known about the process of reionization, varying flavors of which have been shown in simulations to effect the population of observable dwarf galaxies at z = 0.

One goal of ALFALFA was to settle, with increased sensitivity, the question of whether a population of gas-rich galaxies could fill voids and go undetected in optical surveys. ALFALFA has already been able to explore the void environment in the Pisces-Perseus foreground [Saintonge et al., 2008] and found no detections within the void. Instead, the HI population of galaxies followed the optical population, and every HI detection had an optical counterpart, ruling out the possibility of massive starless HI clouds in that void. Because this void is so nearby, ALFALFA was sensitive to HI masses of 10^8 $M_\odot$. The population of

HI-rich, potentially optically dark galaxies that ALFALFA is sensitive to cannot, in sheer numbers alone, explain the void problem.

5.1.4 Chapter Overview

After discussing the literature on the environmental dependence of the HI mass function, this chapter will discuss the datasets used to determine galaxy environments (Section 5.3) and the methods used to quantify environment (Section 5.4). We will present findings in Section 5.5, focusing first on the tendency of low-mass galaxies to exist in low density environments in Section 5.5.1 and next on the HI mass functions found in different environments when density is quantified on different scales in Section 5.5.2. In order to check those results, we consider in Section 5.5.2 a comparison between the Full $\alpha.40$ and Spring $\alpha.40$ datasets. Finally, we discuss the implications of the results and suggest further work that could be done to better understand the relationship between gas and environment in Sections 5.6 and 5.7.

5.2 Previous Environmental Studies

The first-generation blind HI survey, HIPASS, investigated the faint-end slope of the HIMF as a function of environment. Zwaan et al. [2005] found a significant difference between galaxies in low density environments (with flat faint-end slopes α) and in high density environments (with steep faint-end slopes α), and further determined that this difference is exacerbated when environmental differences are defined on larger scales. They found no dependence of the

characteristic Schechter function turnover mass M_* on density. However, the statistical significance of their result was low.

Springob et al. [2005], on the other hand, found a low-significance difference between the HIMF in high and low density environments, but found that α was significantly flattened and M_* shifted lower in denser environments, the opposite of what HIPASS found.

The Springob et al. [2005] result is consistent with the expectation of HI deficiency in denser environments, since stripping and other gas processes would more severely affect gas reservoirs in dwarf galaxies, reducing their overabundance, and consistent with Gavazzi et al. [2005] and with other small-scale environmental studies of the HIMF. Surveys of Virgo [Briggs and Rao, 1993, Davies et al., 2004] have found a flattened HI mass function, and Ursa Major is known to have a flat faint-end HI mass function and luminosity function [Verheijen et al., 2001]. Such studies, however, contain only a few tens of galaxies.

Zwaan et al. [2005] acknowledge the discrepancy between their findings and the general understanding of the HIMF in cluster environments, but offer several possible explanations. First, the authors point out that blind HI surveys are biased against the highest-density environments. What they call 'high density' environments may therefore be immune from cluster effects like ram-pressure stripping, and may correspond better to what would be considered a medium-density or group environment in studies based on optical samples. Second, they suggest that the rate of HI gas consumption in low- vs. high-mass galaxies is environment-dependent, with smaller galaxies in low-density regions forming stars faster and, therefore, flattening the HIMF.

There is a third possibility, related to the bias of blind HI surveys against extremely high-density environments. High-mass objects that evolve in medium density environments are, over time, stripped and shifted into lower and lower mass bins, steepening the HIMF in such environments. It is possible that low-mass objects in high density environments are not simply shifted but entirely destroyed, flattening the HIMF. This could explain the discrepancy between HIPASS's observations in its so-called 'high' density environments compared to studies based on an optical definition of density.

We propose a final possibility: the trend reported in Zwaan et al. [2005] may not have the statistical significance to claim any particular relationship between density and the shape of the HIMF.

These discrepancies, and the question of how environment truly shapes the HIMF, are among the major issues to be explored in this dissertation, with unprecedented precision due to the volume, sensitivity and sample size of ALFALFA.

5.3 The Dataset

5.3.1 ALFALFA $\alpha.40$ Sample

For this analysis, we use the $\alpha.40$ sample from ALFALFA as described in Martin et al. [2010], with some minor changes. The version of $\alpha.40$ used here covers 4 subregions. In the southern Galactic hemisphere, $\alpha.40$ includes $22^h < \alpha < 03^h$, $24^\circ < \delta < 32^\circ$ and $14^\circ < \delta < 16^\circ$; this second region represents a new addition not

included in the cited work. In the northern Galactic hemisphere, the regions are $16^h30^m < \alpha < 07^h30^m$, $4^\circ < \delta < 16^\circ$ and $24^\circ < \delta < 28^\circ$. To eliminate systematic uncertainties due to distance uncertainties, we exclude galaxies nearer than 20 Mpc, for a total sample size of ∼10,200.

In order to make use of outside datasets to quantify galaxy environment in our sample, we must consider a restricted subsample of $\alpha.40$. Henceforth we will refer to this restricted region as Spring $\alpha.40$, as it covers the right ascension ranges visible above the Arecibo Observatory during spring ALFALFA observations. Spring $\alpha.40$ covers $16^h00^m < \alpha < 08^h30^m$, $4^\circ < \delta < 16^\circ$ and $16^h20^m < \alpha < 07^h50^m$, $24^\circ < \delta < 28^\circ$. This reduces the sample size to ∼ 6,800.

In all analyses, we use only Code 1 objects within 15,000 km s^{-1}.

5.3.2 PSCz Density Reconstruction

One method used to quantify environment makes use of cosmic density maps provided by Branchini et al. [1999], based on the distribution of IRAS galaxies in the Point Source Catalog Redshift survey (PSCz). Hereafter, we will refer to methods using these density maps as the PSCz Model method.

Two PSCz density field maps are available, which smooth the density fields on different distance scales and therefore quantify environment on different scales. We refer to these maps and their associated scales as the PSCz Fine Grid and the PSCz Coarse Grid (Gaussian smoothed with a kernel of ∼ 4 Mpc and ∼ 10 Mpc).

Because these maps are all-sky, we can use the PSCz Model method on the

full $\alpha.40$ dataset as well as on Spring $\alpha.40$. This approach has several advantages; the full $\alpha.40$ dataset is significantly larger and the results therefore have greater statistical significance. At the same time, including Spring $\alpha.40$ as a separate analysis allows for more direct comparison with other methods that can only use this restricted catalog. Since the overall dataset is less biased with respect to large-scale structure than Spring $\alpha.40$, it is important to confirm that $\alpha.40$ and Spring $\alpha.40$ agree with each other in the PSCz Model method before extending the comparison to other methods.

5.3.3 SDSS Spectroscopic Sample

This chapter will also present analyses using a second method to quantify environment, the Nearest Neighbors method, developed for this application in collaboration with Saintonge [2007b]. Saintonge [2007b] used the Arecibo General Catalog (AGC, a privately maintained database) as the reference catalog to calculate the distance between an ALFALFA galaxy and its nearest neighbors in the AGC. In that work, a flux-limited sample was used and corrected for the dropoff of sensitivity with distance. Here we are able to use a spectroscopic, volume-limited sample of galaxies from the Sloan Digital Sky Survey (SDSS[1]; Abazajian et al. [2009]). The Nearest Neighbor method is applied only to the Spring $\alpha.40$ sample because of the coverage of SDSS.

[1]Funding for the SDSS has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the US Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS Web site is http://www.sdss.org. The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, the Johns Hopkins University, Los Alamos National Laboratory, the Max Planck Institute for Astronomy, the Max Planck Institute for Astrophysics, New Mexico State University, the University of Pittsburgh, Princeton University, the US Naval Observatory, and the University of Washington.

The Nearest Neighbors method may be preferable due to a low density of sources in the nearby universe, reducing the sensitivity of density maps like PSCz to nearby environmental fluctuations. A key comparison in this chapter will consider the differences between the two methods.

From SDSS, first we select objects from the spectroscopic catalog, which is complete to an SDSS r-band Petrosian magnitude r = 17.7. We have extended the SDSS search area to slightly outside the limits of Spring $\alpha.40$ to combat edge effects. To calculate absolute magnitudes and 3D positions, we estimate distances to SDSS galaxies using both the flow model described in the previous chapter and an assumption of pure Hubble flow (D_{Mpc} = cz_{CMB} / 70.0 km s^{-1} Mpc^{-1}). No statistically significant differences in the final result are seen, likely because we exclude sources within 20 Mpc. In what follows, we present the results found using the flow model distance estimate. From the calculated absolute magnitudes, we are able to construct a volume-limited subsample of SDSS and conduct the Nearest Neighbors analysis using this catalog as a reference. No correction is made for SDSS fiber collisions.

5.4 Quantifying Environment

In this and other works, environment is quantified as the local over- or underdensity relative to the average, through the parameter δ where

$$\delta = \frac{\rho - \bar{\rho}}{\bar{\rho}} = \frac{\rho}{\bar{\rho}} - 1.0 \tag{5.1}$$

In presenting results, the parameter $\delta + 1.0$ is usually displayed, in order to

Table 5.1. Scales on which Environment is Measured

Density Measure	Scale (Mpc)	Std. Dev. σ
NN1	1.4	1.4
NN3	2.8	1.7
NN5	3.6	2.0
NN7	4.2	2.2
NN10	5.0	2.4
NN13	5.6	2.5
PSCz Fine Grid	4.0	
PSCz Coarse Grid	10.0	

straightforwardly represent the fractional density compared to average.

In the case of the PSCz Model method, the density reconstruction maps provide us with the value of δ in a map of cells Gaussian smoothed to some scale of several Mpc. As described above, the Nearest Neighbors method involves calculating the distance between an ALFALFA galaxy and its N^{th} nearest neighbor in an SDSS reference catalog. We find it appropriate to use N = 1, 3, 5, 7, 10 and 13. In Table 5.1, we present the Nearest Neighbor (NN) and PSCz Model methods along with the median scale over which each measurement probes local density. In the case of the Nearest Neighbor method, a standard deviation is provided in order to give a sense of the broad range of distances within which the N^{th} nearest neighbor can be found.

Figures 5.1 and 5.2 show the maximum and typical values, respectively, over the density parameters δ as a function of the scales presented in Table 5.1. It is clear that, as the method moves from NN1 to the PSCz Coarse Grid, the dynamic range of probed densities becomes smaller. On larger scales, the parameter δ becomes less sensitive to density contrasts.

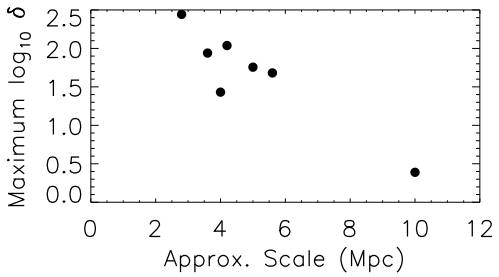

Figure 5.1 The maximum value of the density parameter δ as a function of the approximate scales on which we estimate environment.

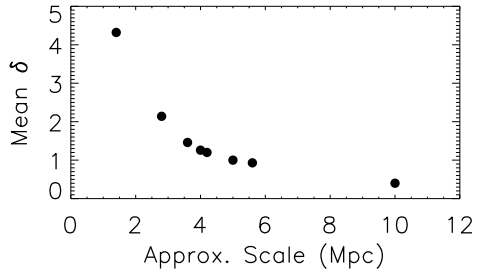

Figure 5.2 The mean (typical) value of the density parameter δ as a function of the approximate scales on which we estimate environment.

Observational (summarized in Blanton and Moustakas [2009], Kauffmann et al. [2004], Blanton [2006]) and theoretical [Haas et al., 2011] work has found that scales of $\sim$1-2 Mpc show the strongest environmental influences on galaxy properties, though larger-scale influences have also been found [Park et al., 2007]. The numerical simulation work in Haas et al. [2011] points out that since environment correlates with halo mass, it is difficult to observationally disentangle the effects of external environment on galaxies from the internal effects due to underlying halo mass. For those reasons, we expect that the range of scales we probe may tell a complicated story about the interplay of HI and small- and large-scale environment.

Once δ on each scale is calculated for each $\alpha.40$ galaxy, the sample is split into five density regimes. The splitting is done in two different ways, in order to probe the full range of observed environments. In the first case, the sample is split into quintiles so that each of five resulting HIMFs has a statistically significant number of galaxies. This has the disadvantage of not accurately representing the environmental variations, given that there are more galaxies in denser environments. In the second case, the sample is split into five bins of δ that fairly represent the full range of density variation, grouping the sample into two underdense bins and three overdense bins. This approach has the disadvantage that some of the resulting HIMFs carry extreme uncertainties due to low sample size.

When considering the results in this analysis, it is important to note that each quintile-binned sample results in an independent measurement of the HIMF; that is, no galaxies are found in more than one quintile. However, any other way of subsampling or binning would not produce a set of HIMFs that are independent of the quintile-binned sample, and vice-versa. For that reason, figures in this chapter will indicate each set of independently binned HIMF measurements with a separate symbol.

5.5 Results: The 2DSWML HIMF as a Function of Environment

The ALFALFA galaxies, binned by environment, are used to construct the HIMF using the Two-Dimensional Step-Wise Maximum Likelihood (2DSWML) method. 150 realizations of each HIMF are calculated in order to capture flux

and distance (i.e. mass) errors. The Schechter parameters for each HIMF are calculated taking into account uncertainties due to sample size and mass errors. The parameters of interest are the faint-end slope α and the characteristic turnover mass M_*, expressed in $\log_{10}$ solar units. We considered the degeneracy in the parameters α and M_* and that potential that underlying changes in one parameter could appear as changes in the other. In order to confirm the findings presented here, we checked the faint-end slope with a simple linear fit to the HIMF for masses below $10^{9.5}\ M_\odot$, and found that our analysis accurately reflected true changes in the HIMF shape as a function of environment.

In the following sections, we will first present the main findings of our analysis and then discuss them in the context of what is already known about environmental dependence of gas properties.

5.5.1 Low HI Mass Galaxies in Low Density Environments

The analysis of the environments in which $\alpha.40$ galaxies are found confirms previous findings that low-mass HI-selected galaxies are preferentially found in underdense regions, and that the highest density regions are populated by more massive galaxies. Figure 5.3 displays δ values for all galaxies in the Spring $\alpha.40$ sample, using the tenth nearest neighbor as the environment metric. The lowest mass galaxies are below the overplotted $\delta = 0$ line and more intermediate mass galaxies avoid the highest density regions.

This is also reflected in the minimum and mean masses of galaxies falling into each density quintile. Table 5.2 presents these values for the tenth nearest neighbor. Both the minimum and the average masses grow as the environment

becomes denser. Because of the relatively few sources on the extreme low-mass end, the increase of the average mass provides stronger support for the changing distribution with environment.

This preference of low-mass galaxies for low-density environments is not statistically extreme, and does not provide any evidence on a low-mass galaxy solution for the so-called 'void problem.' It is significant, however, because ALFALFA would, if anything, be biased against this result, given that the lowest-mass galaxies are only detectable nearby and that the survey area covers nearby overdensities.

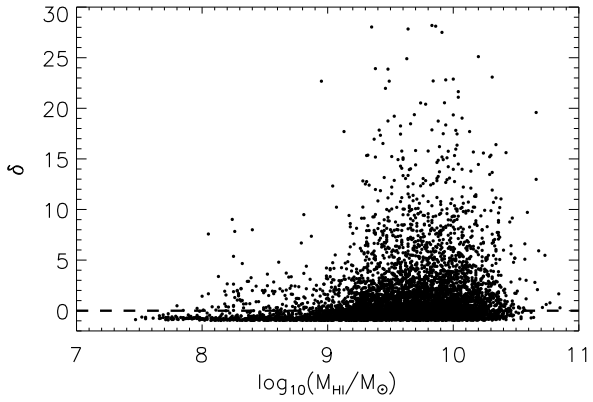

Figure 5.3 The density parameter δ (10^{th} nearest neighbor) for all Spring $\alpha.40$ galaxies beyond 20 Mpc and within 15,000 km s^{-1}, as a function of HI mass. The dashed line indicates δ=0, the dividing line between under- and over-dense regions.

5.5.2 Environmental Dependence in $\alpha.40$

The effect of environment, measured on scales from 1.5 to 10 Mpc, appears to tell a complicated story, and this analysis is inconclusive due to low statistical

Table 5.2. Typical Mass in 10th NN Density Quintiles

Quintile	Minimum Mass ($\log(M_{HI}/M_{\odot})$)	Mean Mass
Least Dense	7.47	9.62
2	7.52	9.75
3	7.73	9.83
4	7.80	9.85
Most Dense	8.05	9.89

significance. If taken at face value, however, this analysis implies a middle road between the known flattening of the HIMF in extremely high-density environments and the HIPASS finding that α increases with increasing density. Before discussing these results in context, we'll first examine the influence of environment on the HIMF when defined on small and large scales.

Environment on Small Scales

The first, third, and fifth nearest neighbor (NN) smooth the density field on relatively small scales of 1 to several Mpc. The 1^{st} nearest neighbor, NN1, is not typically used as a measure of environment because of its greater noise and its potential for contamination by merging galaxies or an offset between the optical positions reported by ALFALFA and the SDSS. Interpretation of the NN1 results should keep this caveat in mind. In this case, NN1 shows a clear trend with α becoming significantly flattened in the highest-density environments. This is inconsistent with Zwaan et al. [2005], which found that the HIMF was steepened at high densities. The ALFALFA result is, on the other hand, consistent with the expectation that cluster-scale processes would reduce the number of low-mass HI reservoirs, thus flattening the HIMF. However, the NN3 and NN5 results

show the trend weakening, rather than strengthening with scale as found in the HIPASS case. Given this and the problems associated with NN1 as an environmental measure, it is unclear what the data are indicating. It is possible that the density parameter δ, which, for a given Nth nearest neighbor, will be measuring density on a range of different scales for the different galaxies in the sample, is muddying the waters by blending the impact of small- and large-scale processes on HI-selected galaxies.

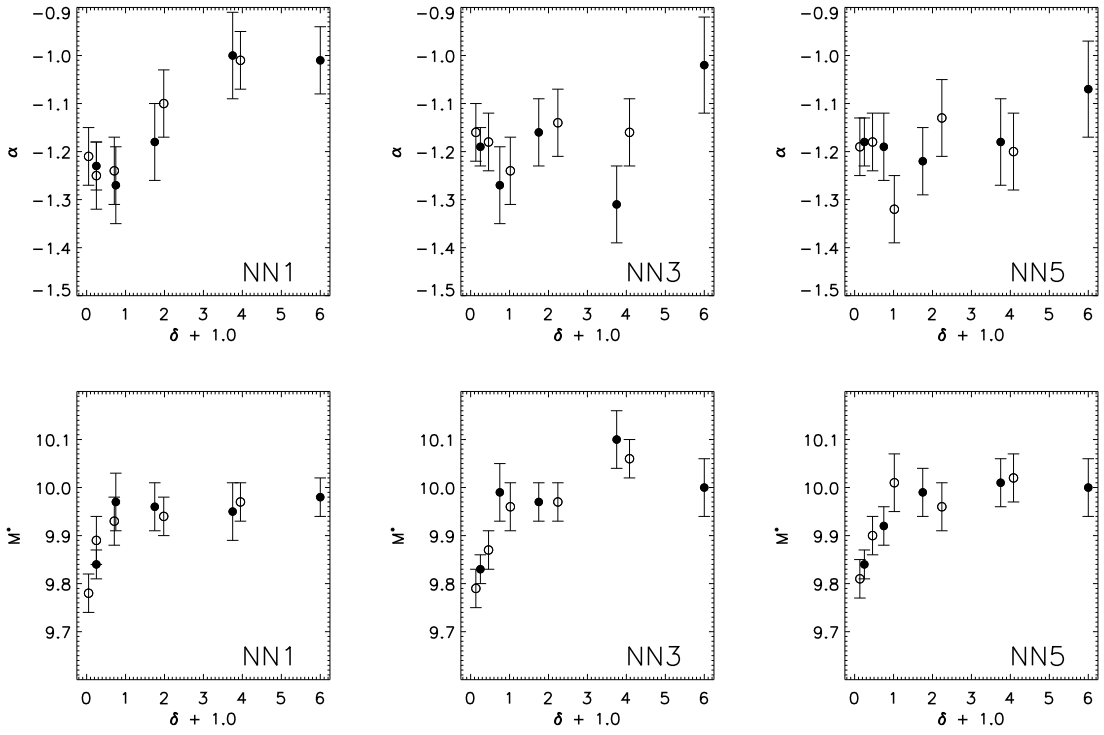

Figure 5.4 The variation of the HIMF parameters α (top row) and M_* (bottom row) as a function of the relative overdensity parameter $\delta + 1.0$ for first, third, and fifth nearest neighbors in Spring $\alpha.40$. The set of filled symbols, and the set of unfilled symbols, are independent points, but a nearby pair consisting of a filled and an unfilled symbol do not represent an independent assessment of the HIMF.

The variation of the HIMF parameters on these small scales as a function of density is shown in Figure 5.4. The filled and unfilled symbols represent two

different ways of subsplitting Spring $\alpha.40$ by environment, as explained previously. As expected, the characteristic mass increases with increased density, reflecting the abundance of more massive dark matter halos in more dense environments. The faint-end slope α, on the other hand, indicates a flattening with increased density when environment is quantified on the smallest scales, with that trend flattening out, and becoming consistent with no HIMF dependence on density, as the scale increases.

Environment on Large Scales

Figures 5.5 and 5.6 are the same as Figure 5.4, but they display results on larger scales for seventh, tenth and thirteenth nearest neighbors and the fine and coarse PSCz density reconstruction maps, respectively.

On mid-scales, as these figures show, M_* increases with increasing density as expected, but α appears to first steepen with density. Then, when some density threshold is reached, it appears that low-mass HI-rich galaxies are effectively stripped of their gas, causing the HIMF to flatten out. The results are only tenuous and inconclusive, but they are consistent across the Nearest Neighbors and PSCz Model methods. As noted previously, these results are independent of the distance estimation method used to calculate δ from nearest neighbor distances. It is possible that a similar trend is seen in the NN1 panel of Figure 5.4, though with no statistical significance. The finding that medium-density environments show a steepened HIMF is consistent with the HIMF for the Leo group derived from ALFALFA by Stierwalt et al. [2009].

On the largest scales, probed by the fine and coarse PSCz density recon-

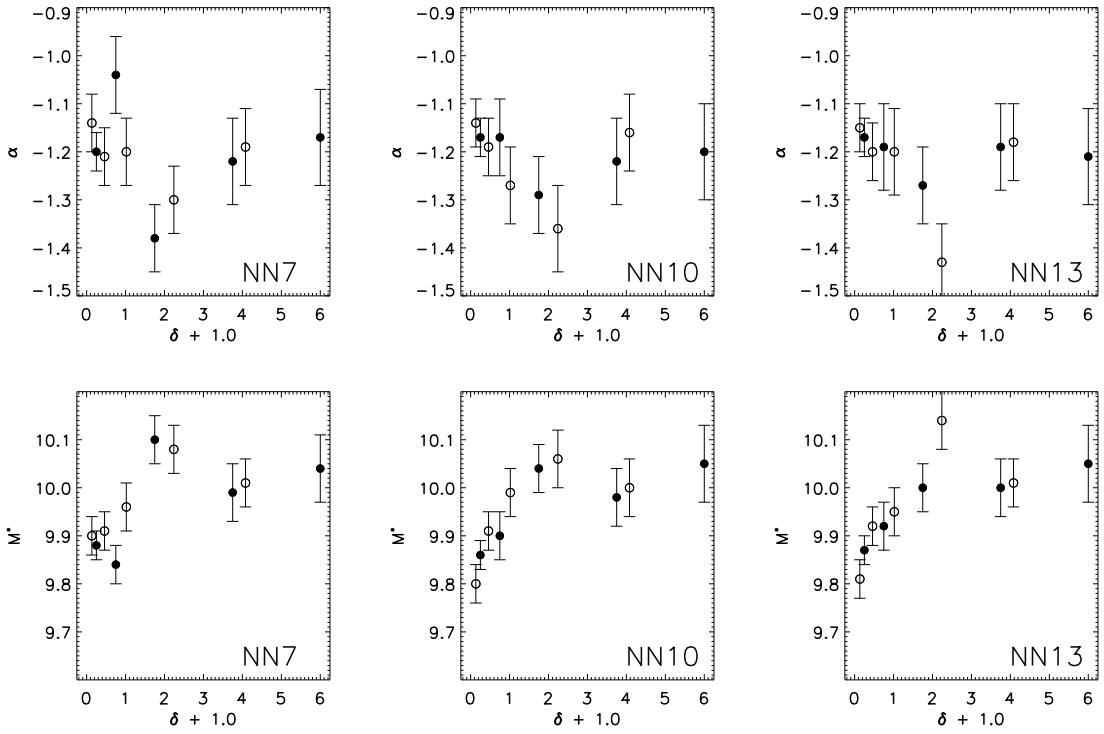

Figure 5.5 The variation of the HIMF parameters α and M_* as a function of the relative overdensity parameter $\delta + 1.0$ for seventh, tenth and thirteenth nearest neighbors in Spring $\alpha.40$.

struction map, these trends weaken. In the coarsest case, where environment is defined on scales $\sim$10 Mpc, both α and M_* show no trend except that the most overdense bins show a steepened α and a larger M_*. This result is also expected, since smoothing on such large scales will tend to reduce the dynamic range and smear out density contrasts, so that each density bin is contaminated by galaxies from a variety of true local environments.

Comparison of Full $\alpha.40$ and Spring $\alpha.40$

Using the PSCz model method, we can test whether the results presented in this section may be due to systematic biases in the Spring $\alpha.40$ sample. This

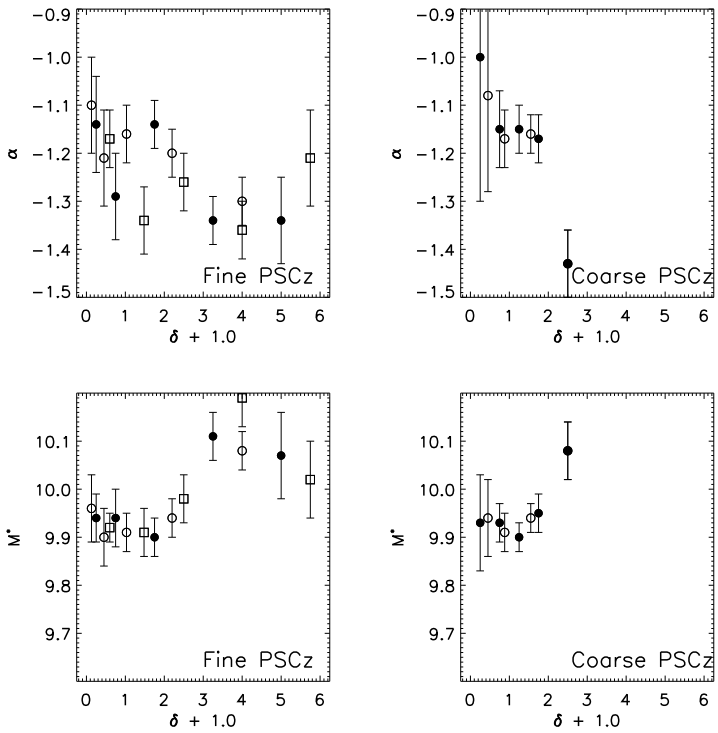

Figure 5.6 The variation of the HIMF parameters α and M_* as a function of the relative overdensity parameter $\delta + 1.0$ from the PSCz fine and coarse maps covering Spring $\alpha.40$.

is of particular concern because this region points toward the Virgo cluster, an overdensity that could influence our results. Such effects should not be important because we have chosen to exclude galaxies closer than 20 Mpc for this analysis.

In Figures 5.7 and 5.8, we display the trends of the parameters α and M_* with relative density using both the PSCz fine and coarse maps. The left-hand columns in each figure display the results when the entire $\alpha.40$ dataset is used, taking full advantage of PSCz's full-sky coverage, while the right-hand columns display the results only for Spring $\alpha.40$ (also displayed in Figure 5.6).

The Full and Spring $\alpha.40$ findings in the case of both the fine and coarse

grid agree with each other and, in the case of the fine PSCz density reconstruction map, agree with the findings for the large-scale environment quantified via Nearest Neighbors method. This bolsters our confidence in the findings presented in Section 5.5.2. Once again, environment defined on the largest scales of $\sim$10 Mpc does not show a trend in HIMF parameters with density, although the highest densities feature both a sharply steepened faint-end slope α and an increase in the turnover mass M_*.

Finally, there are slight indications from the fine PSCz map that the value of M_* may slightly decrease when environment measured on these scales is 5 to 6 times more dense than the typical value. Although the dark matter halo distribution would continue to shift toward higher masses, baryon processes in these extreme environments would lead to observable galaxies with smaller neutral gas reservoirs, and we may be seeing indications of that in the top panels of Figure 5.8.

5.6 Discussion

The findings presented in Section 5.5 tell a complicated story about the relationship between the HI mass function and the environment in which HI selected galaxies live and evolve. In some senses, these results are aligned with expectations. We do find that M_* rises with density on all measured scales. We also expected, based on previous observational and theoretical results, that the smallest scales would show the strongest relationships, though the findings of Zwaan et al. [2005] challenged that. Given our uncertain finding of a strong relationship on the smallest scales we probe, and a different relationship on the

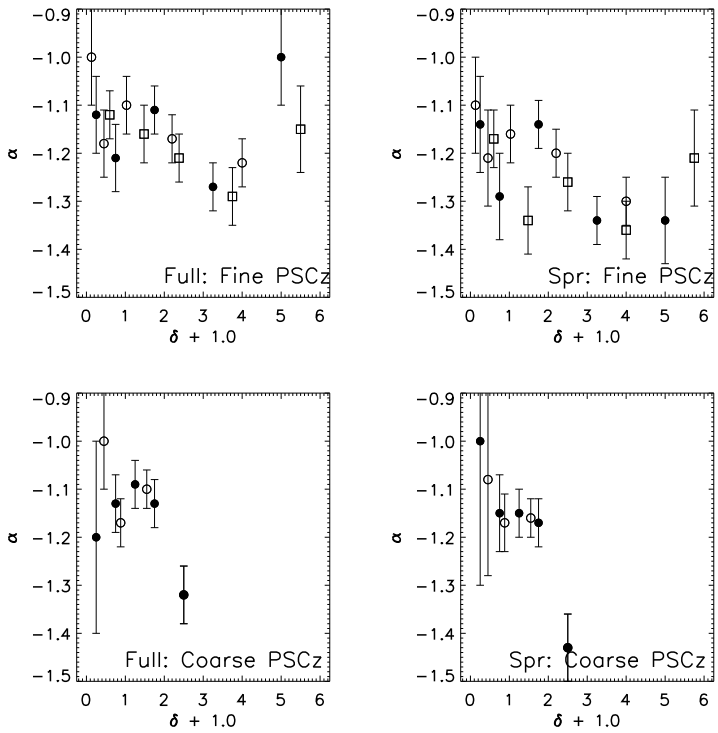

Figure 5.7 A comparison of the variation of the HIMF parameter α as a function of the relative overdensity parameter $\delta + 1.0$ from the PSCz fine (top) and coarse (bottom) maps covering Full $\alpha.40$ (left) and Spring $\alpha.40$ (right).

larger scales, these findings imply a 'middle path' between these two conflicting expectations. Since the coarse grid used in the PSCz Model method shows a very weak relationship between density and the shape of the HIMF, we have confirmed that overdensities on the largest scales (10 Mpc or more) do not appear to have an influence on the evolution of gas in galaxies.

As mentioned previously, the previous findings in this field of study were muddled. While some groups found that the HIMF should be flattened in high density regions – which is completely consistent with our understanding of clusters like Virgo – others found that the HIMF should be steepened with density. It is possible that the $\alpha.40$ HIMF is resolving these two claims by, essen-

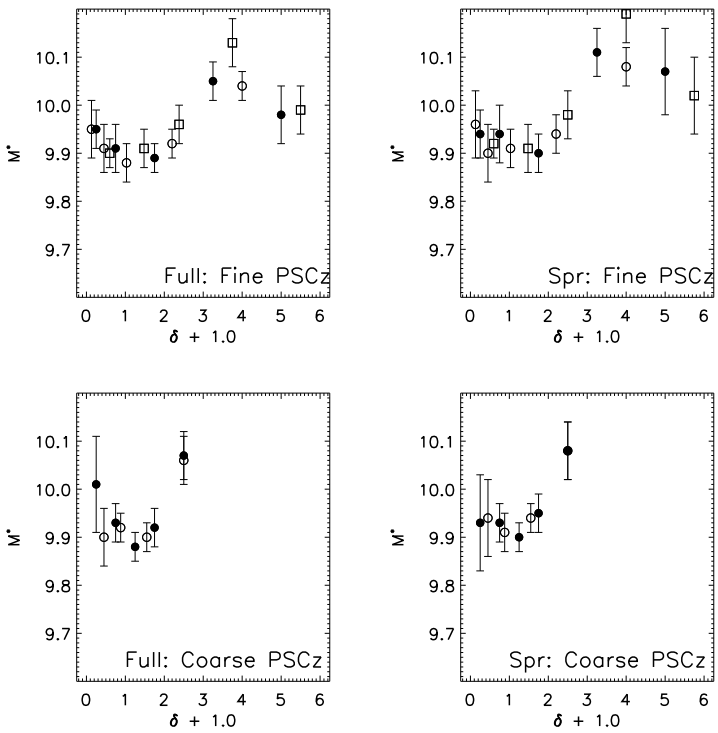

Figure 5.8 A comparison of the variation of the HIMF parameter M_* as a function of the relative overdensity parameter $\delta + 1.0$ from the PSCz fine (top) and coarse (bottom) maps covering Full $\alpha.40$ (left) and Spring $\alpha.40$ (right).

tially, combining them. If our results are confirmed, they would imply that, up to some point, higher-density regions are richer in low-mass galaxies. Underdense voids ($\delta < 0$) would, in that scenario, balance out their abundance of low-mass galaxies with an abundance of intermediate-mass galaxies, potentially a result of increased sensitivity to reionization in void regions [Hoeft et al., 2006]. Higher-density groups would be rich in low-mass galaxies (like the Leo group HIMF of Stierwalt et al. [2009]) and exhibit steep faint-end HIMF slopes. However, at some threshold or 'trigger' density, low-mass galaxies become stripped, leading to a flattened slope. The recent work of Ricotti [2009a] suggests that late cold gas accretion onto dwarf galaxies may be occurring, and because more massive and more concentrated halos can accrete gas more efficiently, the prop-

erties of halos in medium-density environments (compared to low-density environments: Haas et al. [2011]) may suggest a mechanism for gas enrichment and faint-end slope steepening in those environments.

Since it is not well-understood which baryonic processes contribute most to the altering of cluster galaxies that are far from the cluster core (Section 5.1.1), it remains unclear how to connect these findings to our understanding of other cluster galaxy properties, and whether both gas enrichment and gas depletion can take place near/within clusters. The 2dFGRS survey [Lewis et al., 2002] found that galaxies within about 6 Mpc of a cluster core differ systematically from field galaxies, challenging previous indications that environment only has an influence on scales of about 1-2 Mpc. A better understanding of the complex processes and their varying dominance on different scales makes these ALFALFA results more in line with the emerging picture of galaxy evolution. The 2dFGRS work dealt directly with the quenching of star formation, which could be reflected in the HIMF either as stripping (loss of HI and a flattened faint-end slope) or retaining of an HI reservoir (enhanced HI and a steepened faint-end slope) depending on the dominant mechanism of quenching.

While it is possible that $\alpha.40$ is revealing this complex trend in the shape of HIMF with varying environment, it is also possible that the results for the faint-end slope α are consistent with no trend, in which case the observed steepening in medium densities could be a statistical anomaly. Within a particular methodological quantification of environment (e.g. NN10 and the PSCz fine grid), each $\delta + 1.0$ bin represents an independent sample. However, when we are comparing results across different methods, the points are highly correlated so that anomaly could be present in several measurements. The agreement between

the Nearest Neighbors and the PSCz fine grid methods tends to support the suggestion that there is a pattern emerging.

5.7 Summary and Suggestions for Future Efforts

In this Chapter, we investigated the dependence of the $\alpha.40$ HI mass function on galaxy environment. We quantified environment using a density parameter, δ, both from the PSCz density reconstruction maps (large scales) and a nearest neighbors approach with a volume-limited spectroscopic sample of galaxies in SDSS (small to large scales).

We find that the characteristic Schechter function turnover mass, M_*, increases with density. We also determined that we can rule out the low-confidence HIPASS result [Zwaan et al., 2005], which indicated that faint-end slopes α steepen with density and that this relationship grows stronger when environment is defined on larger and larger scales. Instead, we find that the HIMF is increasingly flattened with increasing local density when environment is defined on small scales. On more intermediate scales (NN3 and NN5), this relationship is not observed. This potentially occurs because using δ as a quantifier of environment leads to a mixing of scales. HIPASS also found no relationship between M_* and environment. These differences are most likely due to HIPASS using their HI-selected, flux-limited galaxy sample to define environment rather than an optically-selected, volume-limited sample.

However, the emerging overall picture of the dependence of the HIMF on environment is both complex and inconclusive. The analysis of the 40% ALFALFA Survey ($\alpha.40$) sample indicates that, on the smallest (cluster) scales indicated by

the first nearest neighbor, known processes like ram-pressure stripping flatten the HIMF in the highest-density environments. On larger scales, however, a combination of 'nature' and 'nurture' effects cause void HIMFs to be relatively flattened, medium-density galaxies to be HI enriched, and galaxies in extremely high-density environments to be stripped and the HIMF flattened. Any analysis of this question is complicated by the difficulties of reliably defining galaxy environment.

While the results of this analysis are inconclusive, the scenario they suggest is highly interesting, and has the potential to open new understanding of the relationship between environment, star formation, and gas-loss mechanisms. Better understanding of the HIMF's changing shape in environments of varying densities will emerge as more ALFALFA data becomes available, reducing the statistical uncertainties. This understanding may also be bolstered by using different environmental definitions, such as the identification of true voids, groups and clusters through structure-finding algorithms. Selecting truly over- and under-dense environments could eliminate some of the systematic uncertainties while potentially isolating the various effects that are influencing the faint-end slope of the HIMF.

CHAPTER 6

CLUSTERING CHARACTERISTICS OF HI-SELECTED GALAXIES FROM THE 40% ALFALFA SURVEY

6.1 Introduction

Galaxies selected by their neutral hydrogen are known to be less clustered than their optically-selected counterparts (Basilakos et al. [2007], Meyer et al. [2007] for HIPASS) and less likely to be found in such dense environments. Given the cosmological utility of 21 cm galaxy redshift surveys, it is important to understand the clustering characteristics of this population of galaxies. Specifically, 21 cm line surveys obtain detections and redshift concurrently, reducing their expense and eliminating the need for follow-up observations. Such surveys are also able to probe galaxy populations irrespective of luminosity, stellar mass, or dust extinction. Additionally, such surveys are able to probe low-luminosity dwarf systems, which tend to be gas-dominated [Geha et al., 2006]. Conversely, such surveys are biased against clusters, the most luminous galaxies, and the 'red and dead' galaxy population.

Given the lack of large samples of HI-selected galaxies to date (the HIPASS main catalog and its northern extension contain, respectively, 4,315 and 1,002 galaxies; Meyer et al. [2004], Wong et al. [2006]), this population, its evolution, and its bias compared to dark matter are poorly understood. The selection of these galaxies is strongly limited in redshift, and targeted observations can only extend to $z \sim 0.2$ [Catinella et al., 2008, Freudling et al., 2011], while the ALFALFA survey is limited to $z < 0.06$. At the same time, this population is poised to become the standard for cosmological measurements based on observations

of resolved galaxies as well as intensity mapping. For example, galaxy redshift surveys taking advantage of the 21 cm transition of neutral hydrogen undertaken with instruments like the Square Kilometer Array (SKA) would potentially provide constraints on the dark energy equation of state and its variation with redshift [Abdalla et al., 2010, Myers et al., 2009].

The differences in neutral hydrogen distribution between galaxies in clusters and those in the field are not well understood, with proposed solutions spanning from 'nature' (i.e., gas-rich galaxies form in low-concentration dark matter halos and/or in underdense environments) to 'nurture' (i.e., processes that occur after formation deplete the HI gas from halos or enrich HI reservoirs through cold accretion). The reality is a combination of many processes and initial conditions. Probing the relationship between cold gas mass and other properties known to be anticorrelated with clustering (such as spiral morphology, late type [Norberg et al., 2002], active star formation [Kauffmann et al., 2004], and blue colors [Zehavi et al., 2005]) may help to better articulate the influence of environment on galaxy evolution while also constraining the populations to which future large 21 cm line surveys will be sensitive.

Since previously available HI-selected galaxy samples were small, some of our understanding and expectations of clustering in gas-rich galaxies is based on optical work that studied clustering effects due to other galaxy properties that are correlated with HI. A cross-correlation between two different populations of galaxies, for example, can be used to measure their relative bias. Wang et al. [2007] used cross-correlation analysis of galaxies of varying colors and luminosities to provide a quantitative view of what was already known, that is, that red galaxies cluster more than blue galaxies but that the largest enhance-

ment in the clustering of red galaxies is on small, cluster-sized ($r \lesssim 1\ h^{-1}$ Mpc) scales. Intriguingly, those authors also found that the brightest blue galaxies are clustered in a way that most closely resembles intermediate-luminosity red galaxies and are unlike fainter blue galaxies. These results suggest that there is a significant population of very bright, potentially gas-rich, blue galaxies in denser environments.

Most work directly related to the clustering of gas-rich galaxies came out of the HIPASS survey. Meyer et al. [2007] and Basilakos et al. [2007] both identified the HIPASS HI-selected sample as the weakest clustering population of galaxies known, but their results regarding the mass dependence of the clustering were in conflict. While the HIPASS team found a statistically insignificant difference between 'high' and 'low' mass galaxies, Basilakos et al. [2007] found that high-mass galaxies clustered more strongly. More recently, Passmoor et al. [2011] compare the ALFALFA and HIPASS projected correlation function and angular correlation function, and find that they are similar but that ALFALFA's sensitivity to low-mass galaxies makes that sample more strongly anti-biased relative to dark matter. However, Passmoor et al. [2011] use only the ALFALFA catalogs published in Giovanelli et al. [2007], Saintonge et al. [2008], and Kent et al. [2008] ($\sim$ 1,800 galaxies) despite several other ALFALFA catalogs being available at time of publication, and are therefore severely limited in their sample size.

The excellent sensitivity and large sample size of the $\alpha.40$ sample allows us to probe the clustering characteristics of HI-selected galaxies through the two-point galaxy-galaxy correlation function. Given the size of the available sample, we are also able to sub-divide in order to investigate the dependence of

clustering on galaxy properties while still maintaining a robust measurement.

In the following sections, we describe our dataset (Section 6.2) and the methodology used to measure the galaxy-galaxy correlation function (Section 6.3). We then estimate the real-space correlation function, both assuming a power law and by direct inversion, and investigate the impact of methodology choices in Section 6.4. We compare the ALFALFA clustering results to those found in simulations that have, for the first time, attempted to assign reasonable cold HI gas masses to simulated galaxies, in Section 6.5, while also discussing the results in context, before concluding in Section 6.6.

6.2 Dataset

6.2.1 ALFALFA $\alpha.40$ Sample

The ongoing ALFALFA survey is completing a census of galaxies in the local universe, out to z $\sim$0.06, using the seven-pixel ALFA receiver at the Arecibo Observatory. Compared to previous blind neutral hydrogen surveys (e.g. HIPASS), ALFALFA's enhanced sensitivity, detection centroiding, volume, and sample size make it ideally suited for an accurate measurement of the correlation function of gas-rich galaxies.

The sample used here is similar to that in Martin et al. [2010], referred to as $\alpha.40$ in that work, except for the addition of galaxies in an area defined by $22^h00^m <$ RA $< 03^h00^m$, $14^\circ <$ Dec. $< 16^\circ$ (J2000). From the available volume, we select ALFALFA Code 1 detections, therefore including only those reliable

extragalactic detections (SNR $>$ 6.5). We exclude Code 2 objects, which are marginally detected (4.5 $<$ SNR $<$ 6.5) but are confirmed by the existence of an optical galaxy with previously measured redshift corresponding to that measured by ALFALFA. The sample is further limited to those Code 1 objects with an integrated flux measurement above a threshold defined as a function of the 21 cm profile velocity width, described in detail in Martin et al. [2010]. As in that work, we exclude galaxies with $cz_{helio} > 15{,}000$ km s^{-1}, where our sensitivity is drastically decreased by radio frequency interference from the Federal Aviation Administration radar at the San Juan airport.

6.2.2 Selection Function

The distance dependence of the selection function of $\alpha.40$ was determined using the 2DSWML method, described fully in Chapter 3 and Appendix B. Here, we summarize how the selection function is calculated and then applied to the sample. Because we have an expression for the survey sensitivity expressed in terms of source mass (integrated flux) and profile velocity width, the selection function S is written in terms of the measured HI mass function mass (m, in bins of width Δm) and width (w, in bins of width Δw) bin values, ϕ_{jk}:

$$S(D_{Mpc}) = \frac{\sum_{all\ w} \sum_{m_{det}} \phi_{jk} \Delta m \Delta w}{\sum_{all\ w} \sum_{all\ m} \phi_{jk} \Delta m \Delta w} \tag{6.1}$$

where m_{det} is the limiting detectable mass bin at D_{Mpc}. However, the calculation of the HIMF via 2DSWML has already provided the coefficients H_{ijk}, which indicate whether or not a particular galaxy falls into a bin (or, in practice, a pair of bins in m and w) of interest. Therefore, for a galaxy i:

$$S(D_i) = \frac{\sum_{all\ w} \sum_{all\ m} H_{ijk} \phi_{jk} \Delta m \Delta w}{\sum_{all\ w} \sum_{all\ m} \phi_{jk} \Delta m \Delta w} \tag{6.2}$$

or, simplified,

$$S(D_i) = \frac{\sum_j \sum_k H_{ijk} \phi_{jk}}{\sum_j \sum_k \phi_{jk}} \tag{6.3}$$

While the numerator in Equation 6.3 must be calculated independently for every galaxy i, the denominator can be calculated once, along with the HIMF, and applied to every individual galaxy.

For application to the correlation function, S_{D_i} is calculated for every galaxy in the sample and then S_D is smoothed. This smoothed selection function can additionally be combined with an HI mass function to make predictions regarding the number of galaxies of a given mass that are expected to be found in the survey, or used to predict the redshift distribution of the survey under an assumption of homogeneity. Figure 6.1 shows a histogram of the $\alpha.40$ redshift distribution, with peaks and dips representing clusters and voids, respectively, along with an overplotted prediction based on the selection function.

Disagreements between the prediction and the observations are due both to the existence of large scale structure in the survey volume and to the loss of survey sensitivity at certain velocities due to radio frequency interference. This contamination is quantified as a percentage of survey coverage at a given heliocentric velocity, or spectral weighting, as discussed in Martin et al. [2010]. For the purposes described here, the weights map has been translated into the CMB reference frame in order to most accurately model the predicted ALFALFA galaxy distribution (see Section 6.3.2). The selection function is used both for the

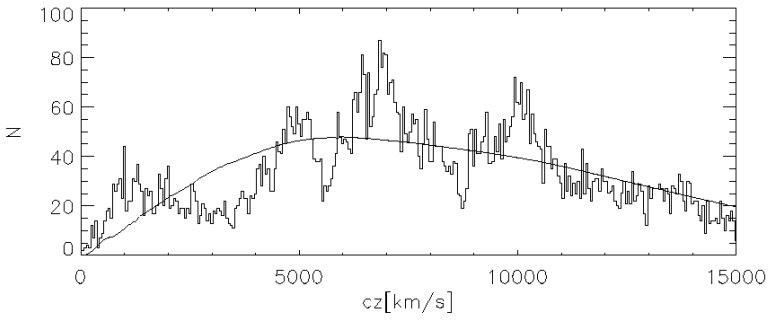

Figure 6.1 The observed redshift distribution of $\alpha.40$ galaxies (histogram) compared to the expected distribution obtained via the survey's selection function.

creation of the random samples for estimation of the correlation function, and for the weighting of pair counts in that estimate.

6.3 Method: Estimation of ξ(r) and Error Analysis

We measure the correlation function, $\xi(\sigma, \pi)$ in bins of on-sky (σ) and radial (π) redshift-space separations, using their observed velocities. Given the redshift extent of $\alpha.40$, we have translated measured galaxy velocities from the heliocentric frame of reference to the CMB frame of reference using Lineweaver et al. [1996]. For two galaxies i and j, these separations are:

$$\sigma = \frac{v_i + v_j}{H_0} \times \tan\theta/2 \tag{6.4}$$

and

$$\pi = \frac{|v_i - v_j|}{H_0} \tag{6.5}$$

where θ is the angular separation of the two galaxies on the sky and H_0 is expressed in units of h (H_0 = 100h).

Our ultimate goal is to measure ξ(r), the real space correlation function, through the observables actually available to us, that is, $\xi(\sigma, \pi)$. In particular, we are interested in modeling the power-law shape of the correlation function up to $\sim$10 h^{-1} Mpc, beyond which point the correlation function is known to diverge from a simple power-law.

Since ξ measures not simply the probability distribution of galaxy separations in a sample, but the *excess* probability compared to a homogeneously distributed sample, estimators compare the observed galaxy distribution to a random distribution designed to reflect the survey's observational limitations but to exclude the effects of large scale structure. This is straightforwardly accomplished by comparing the number of pairs in (σ, π) separation bins from the observed sample to the pair counts from the random sample. In the sections that follow, we will describe this method and the corresponding error analysis in greater detail.

6.3.1 Pairwise Estimation

We adopt the Landy-Szalay pairwise estimator [Landy and Szalay, 1993] for the correlation function. First, we define a set of pair counts P_{DR}, P_{DD} and P_{RR} that consider data-random pairs, data-data pairs, and random-random pairs, respectively. As an example, the P_{DR} pair count function is:

$$P_{DR} = \sum_{x \in D} \sum_{y \in R} \Phi_{\sigma,\pi}(\vec{x}, \vec{y}) \tag{6.6}$$

where $\Phi_{\sigma,\pi}$ = 1 when the separation δ between $\vec{x}$ and $\vec{y}$ falls on $\sigma < \delta_\sigma < \sigma + \Delta\sigma$, $\pi < \delta_\pi < \pi + \Delta\pi$ and $\Phi_{\sigma,\pi}$ = 0 otherwise. Bins in π and σ ($\Delta\pi$ and $\Delta\sigma$) are designed for reasonable spacing depending on the number density of a particular galaxy sample. Effectively, P_{DR} counts the number of pairs of points falling into each (σ, π) separation bin. These are then normalized to the galaxy and random sample counts, N_D and N_R, to give normalized versions D_{DR}, D_{DD} and D_{RR}:

$$D_{DR} = \frac{P_{DR}}{N_D \times N_R} \tag{6.7}$$

$$D_{DD} = \frac{P_{DD}}{N_D \times (N_D - 1)} \tag{6.8}$$

$$D_{RR} = \frac{P_{RR}}{N_R \times (N_R - 1)} \tag{6.9}$$

The normalization procedure allows us to construct a random catalog that contains many more objects than the observed data catalog, thereby reducing the introduction of shot noise from the random set. The Landy-Szalay estimator is constructed from these normalized counts:

$$\hat{\xi}_{LS} = \frac{D_{DD} - 2(D_{DR}) + D_{RR}}{D_{RR}} \tag{6.10}$$

Because $\alpha.40$ is not volume-limited, the pair counts must be weighted so that the measurement is not dominated by galaxies at the peak of the selection function. Following Meyer et al. [2007] and Hawkins et al. [2003], we apply a weighting $w_{ij} = w_i \times w_j$ for the contribution of each pair i, j to P_{DR}, P_{DD} and P_{RR}, given by:

$$w_i = \frac{1.0}{1.0 \; + \; 4\pi \; N_D \; S(r_i) \; J_3(s)} \tag{6.11}$$

where N_D is the number of galaxies in $\alpha.40$, $S(r_i)$ is the selection function measured for $\alpha.40$ at $r_i = cz_{CMB,\,i}/H_0$, and

$$J_3(s) = \int_0^s s'^2 \; \xi(s') \; ds' \tag{6.12}$$

defined in terms of the redshift space coordinate $s = \sqrt{\sigma^2 + \pi^2}$.

This expression for J_3 requires an assumed model for $\xi(s)$, but the final measurement of the correlation function is not sensitive to this assumed input for object weighting; we assume a power-law form:

$$\xi(s) = \left(\frac{s}{s_0} \right)^{-\gamma} \tag{6.13}$$

and we test our robustness by first assuming a fiducial value found for optically-selected samples, s_0 = 5.0 h^{-1} Mpc and γ = 1.8 and, after that, iterating to the value s_0 and γ measured for $\alpha.40$. No statistically significant difference is observed through this iterative process, and we therefore proceed as other authors have, using the fiducial optical values reported here in our J_3 weighting.

Following Fisher et al. [1994a,b], we apply an artificial cutoff with a maximum value of s = 30 h^{-1} Mpc in the expression for J_3.

6.3.2 Random Samples

We construct random samples that contain 20 times the number of objects in the $\alpha.40$ dataset. These random samples are carefully designed to include survey selection effects while excluding correlations due to large-scale structure. This is accomplished by predicting the distribution of cz$_{CMB}$ from the survey selection and HI mass functions (see Figure 6.1) and then folding in the loss of volume as a function of velocity due to radio frequency interference, measured from the spectral weights map in Martin et al. [2010]. Objects in the random set are randomly assigned a sky position from the parameters of $\alpha.40$ and a redshift from this predicted distribution. The resulting redshift distribution for one example instance of the random sample procedure is shown in Figure 6.2.

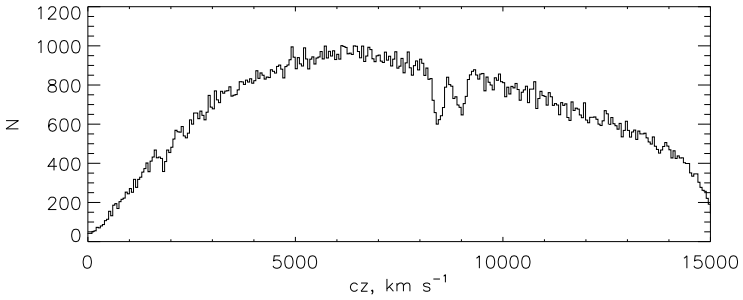

Figure 6.2 The redshift distribution of the constructed random sample. The dips in the distribution at $\sim$ 8,000 km s^{-1} are due to radio frequency interference. When data at these frequencies is flagged, it leads to a reduction in the effective search volume at the corresponding velocities, which translates into a reduction in counts in the random samples. See Chapter 3, Figure 6 for a plot of the average relative weight as a function of velocity in $\alpha.40$.

6.3.3 Error Analysis

The correlation function is measured in bins of separation. While the correlation function is expressed as a function of several different coordinates while iterating towards the real-space correlation function ξ(r), the bin counts and thus the measured correlation functions are correlated with one another in every such coordinate system. Because structures, such as clusters, will contribute an overabundance of pairs to a set of several bins, the measurement in each bin is not independent of the others. In plots of the correlation function shown in this chapter, we display the on-diagonal elements of the covariance matrix (i.e. the standard deviations) as plotted uncertainties on each point. However, in order to work with our measurement to estimate the power-law shape of the correlation function of gas-rich galaxies, we must construct a full covariance matrix and take off-diagonal elements into account.

To construct the covariance matrix C, we carry out our pair-counting routine on more than 500 bootstrap resamplings of the data, and a single catalog of random objects is reused in each case. Each of the bootstrap measurements of $\xi(\sigma,\ \pi)$ contains N_g galaxies selected at random from $\alpha.40$, with replacement. From this set of realizations, we construct the covariance matrices for ξ(s), $\Xi(\sigma)$ and ξ(r). The covariance between two correlation function bins b_l and b_m is given by:

$$C(l,\ m) = \sum_{i=1}^{N_{realizations}} \frac{(b_{l,\,i} - \bar{b_l}) - (b_{m,\,i} - \bar{b_m})}{N_{realizations} - 1} \tag{6.14}$$

The significant off-diagonal elements of the covariance matrix make it difficult to obtain a power-law model fit by minimizing the χ^2 values weighted by

the variance. The covariance matrix, however, is not an inescapable quality of the data, but is actually dependent on the basis in which the data are projected. In this case, we have some number of bins N_b representing a set of variables b (bin centers in h^{-1} Mpc), and can choose to work in an orthonormal basis with N_b coordinate axes in which the covariance matrix C is diagonalized. This basis is defined by the principal component eigenvectors of the measurement, and we borrow elements of principal component analysis in order to obtain model parameter fits and uncertainty estimates.

The principal components are linear combinations of the original N_b variables arranged such that the first principal component corresponds to an orthonormal axis through N_b-dimensional space that explains the largest proportion of variance in the dataset. If desired, those axes that explain a very small fraction of the variance can be excluded from the analysis, but for our purposes that provides no advantage and could lead to a loss of critical information, so we keep the full matrix intact. These principal component vectors are defined by the eigenvectors of the covariance matrix of the original dataset.

Following Fisher et al. [1994a], we make use of the Karhunen-Loeve transform to calculate the principal eigenvectors and construct a diagonalizing matrix, R, the columns of which are these eigenvectors, and a new covariance matrix, $\tilde{C}$, projected in the new basis set. Since all of the covariance has been accounted for in the definition of the principal components, $\tilde{C}$ has no off-diagonal elements, and the variance is captured in the on-diagonal elements $\tilde{\sigma}$.

Given $\tilde{C}$ and R, a set of models with varying values for s_0 and γ can be projected into the principal component basis via $\tilde{b}_{model} = R^T\, b_{model}$, for comparison to the measured $\tilde{b}$. We find the value of each parameter that minimizes the

expression

$$\chi^2 = \frac{1}{N_b - 2} \sum_{1}^{N_b} \frac{\tilde{b}_l - \tilde{b}_{l,model}}{\tilde{\sigma}_l} \tag{6.15}$$

Finally, we construct error ellipses to fully describe the likely parameter space of the power-law model for the correlation function [Press et al., 1992].

6.3.4 Precursor Calculations, ξ(s) and $\xi(\sigma, \pi)$

The simplest measurement of clustering that can be made, ξ(s), is also the least robust and the most prone to the influence of redshift-space distortions. To make this measurement, two-dimensional information is disregarded and the only measurement of interest is pairwise separations in the redshift space coordinate $s = \sqrt{\sigma^2 + \pi^2}$. The result, out to separation scales of $\sim$30 h^{-1} Mpc, is shown in Figure 6.3, with the on-diagonal elements of the covariance matrix plotted as error bars as described previously. The redshift space correlation function's curvature results in a poor fit to any power-law model, but the small scale of the uncertainties gives strong confidence in the robustness that can be anticipated for the full correlation function.

If we include the two-dimensional information available to us, we can then consider $\xi(\sigma, \pi)$. The resulting two-dimensional image, in Figure 6.4 with contours overplotted, clearly reveals the redshift space distortions that lead to the curvature of the redshift-space correlation function shown in Figure 6.3. The radial coordinate, π, appears stretched at small angular separation σ because of the Eddington effect in clusters, while π is flattened on large scales because of

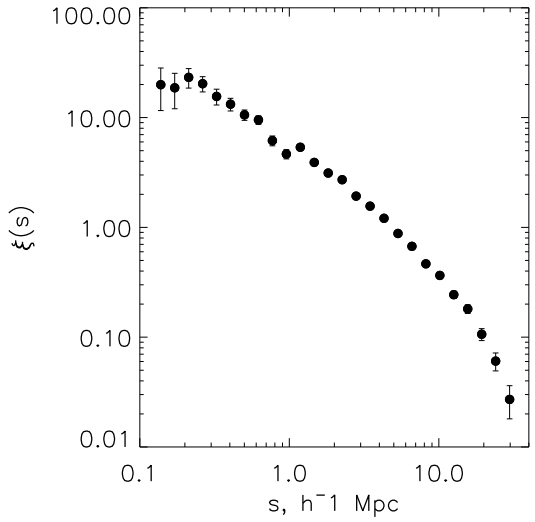

Figure 6.3 The redshift-space correlation function from $\alpha.40$ with on-diagonal elements of the covariance matrix plotted as error bars.

the coherent motion of galaxies towards attractors.

6.3.5 Obtaining the Real-Space Correlation Function

This calculation of $\xi(\sigma, \pi)$ is the fundamental measurement upon which the results presented in the rest of this chapter are based. In order to work with the real-space correlation function ξ(r), we must take the intermediate step of projecting $\xi(\sigma, \pi)$ along the π axis (in practice, using the discrete bins of size $\Delta\pi$), resulting in what is known as the 'projected correlation function' and symbolized as $\Xi(\sigma)/(\sigma)$:

$$\frac{\Xi(\sigma)}{\sigma} = \frac{2}{\sigma} \sum_{0}^{\pi_{max}} \xi(\sigma,\, \pi)\, \Delta\pi \tag{6.16}$$

The maximum value of the integration, π_{max}, is selected so that the summa-

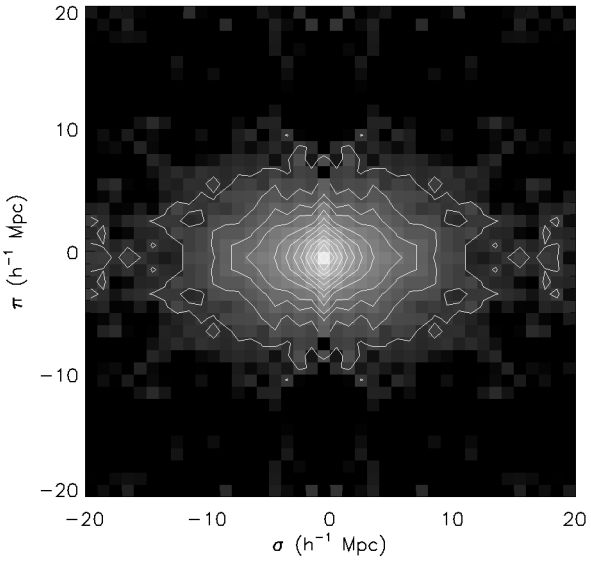

Figure 6.4 The two-dimensional correlation function from $\alpha.40$ in σ and π, measured in h^{-1} Mpc. Brighter colors indicate stronger clustering.

tion is convergent but is kept as low as possible to avoid the introduction of noise from poorly-measured scales. We carry the sum up to a scale of $\sim$30 h^{-1} Mpc.

$\Xi(\sigma)/(\sigma)$ is closely related to the function in which we are truly interested, ξ(r) where r is the real-space distance, via:

$$\frac{\Xi(\sigma)}{\sigma} = \frac{2}{\sigma}\int_{\sigma}^{\infty} \xi(r)\,\frac{r\,dr}{(r^2\,-\sigma^2)^{1/2}} \tag{6.17}$$

In order to evaluate the real space correlation function, then some assumptions must be made about its form. Two options are usually explored in the literature: a power-law form, or a stepwise-function form which makes no assumptions about shape but does assume that the binning used well-represents an underlying smooth correlation function. If we assume a power law of the

form ξ(r) = (r/r$_0$)$^{-\gamma}$, we find:

$$\frac{\Xi(\sigma)}{\sigma} = (\frac{r_0}{\sigma})^{\gamma} \frac{\Gamma(1/2)\,\Gamma((\gamma-1)/2)}{\Gamma(\gamma/2)} \tag{6.18}$$

In Equation 6.18, the function Γ is the well-known Gamma function. Equation 6.18 can be recast in terms of fitting parameters:

$$\frac{\Xi(\sigma)}{\sigma} = \left(\frac{r_0}{\sigma}\right)^{\gamma} A(\gamma) \tag{6.19}$$

Following Meyer et al. [2007], we rearrange Equation 6.19, obtain the best-fit power-law of the form $\xi(\sigma) = a_1\ \sigma^{a_2}$ using the χ^2 minimization given by Equation 6.15, and then relate those parameters to r_0 and γ which represent the best-fit power law for ξ(r).

In the next section, we derive and discuss ξ(r) using the mechanisms described in this section.

6.4 Results: Clustering in $\alpha.40$

6.4.1 $\Xi(\sigma)/\sigma$ and ξ(r) Assuming Power-Law Model

The projected correlation function $\Xi(\sigma)/\sigma$ (recast for the figure and the fitting as $\Xi(\sigma)$) is displayed in Figure 6.5, along with error bars reflecting the on-diagonal elements of the full covariance matrix. The dashed line is the best-fit model obtained by χ^2 minimization using the full covariance matrix. In Table 6.1, we list

the parameters for the fit and their uncertainties, along with the fits obtained if only the the on-diagonal elements (the standard deviations, σ) are used to carry out the standard least-squares fit. For comparison, we also include the clustering reported by HIPASS (which ignored off-diagonal elements), the clustering found by other groups using the HIPASS dataset, and the clustering of several optically-selected samples. We also display the Passmoor et al. [2011] results, which used a small, publicly available subset of the ALFALFA data.

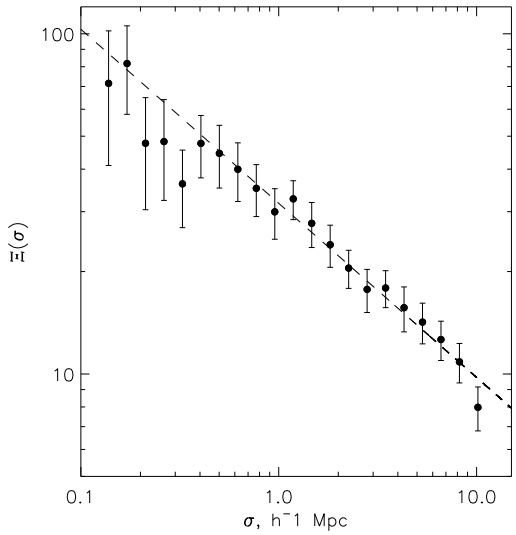

Figure 6.5 The projected correlation function $\Xi(\sigma)$ from $\alpha.40$. Error bars reflect the on-diagonal elements of the full covariance matrix. The overplotted dashed line is the fit from the full covariance analysis, with γ = 1.51 $\pm$ .09; $r_0 = 3.3^{+0.3}_{-0.2}$ (h^{-1} Mpc).

Error ellipses are displayed in Figure 6.6, with the dashed contour giving the 1σ single-parameter uncertainties listed in Table 6.1.

While both the full covariance analysis and the assumption of bin independence give similar results, the larger uncertainties on the full covariance analysis give an indication of the need to be conservative. Parameter uncertainties previously reported in the literature (i.e., Meyer et al. [2007]) significantly underes-

Table 6.1. Best-Fit Correlation Function Power Law Models

Fitting Method	r_0 (h^{-1} Mpc)	γ
Full Covariance	3.3 (+0.3, -0.2)	1.51 ($\pm$ .09)
On-Diagonal Only	3.2 ($\pm$ 0.1)	1.48 ($\pm$ .03)
Passmoor ALFALFA [a]	2.3 ($\pm$ 0.6)	1.6 ($\pm$ .1)
HIPASSa (2007) [b]	3.5 ($\pm$ 0.3)	1.47 ($\pm$ .08)
HIPASSb [c]	3.3 ($\pm$0.3)	1.4 ($\pm$ 0.2)
2dFGRS late-type faint[d]	3.7 ($\pm$0.8)	1.8 ($\pm$ 0.1)
SDSS Bright[e]	6.2 ($\pm$0.2)	1.85 ($\pm$ 0.03)
SDSS Faint[e]	3.5 ($\pm$0.3)	1.92 ($\pm$ 0.05)

[a]Passmoor et al. [2011]

[b]Meyer et al. [2007]

[c]Basilakos et al. [2007]

[d]We include the second-faintest sample due to a warning in Norberg et al. [2002] that the faintest (and smallest) sample provides poorly-constrained fits.

[e]Zehavi et al. [2005]

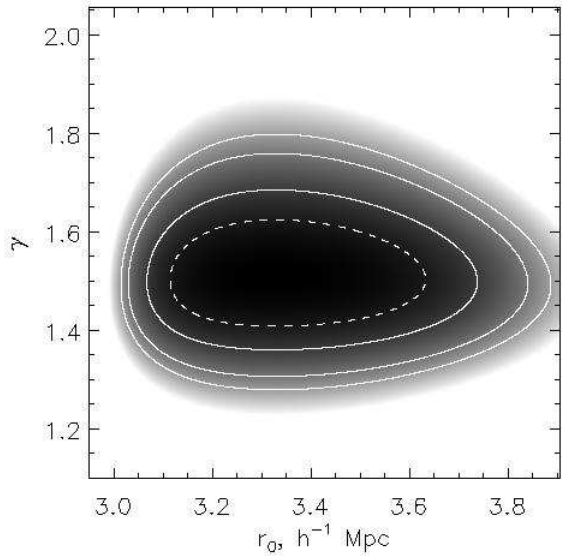

Figure 6.6 χ^2 contours for γ and r_0 (h^{-1} Mpc). The dashed contour gives the 1σ projected uncertainties on each parameter, and the solid contours give 1, 2 and 3 σ fits, respectively, to the pair of parameters.

timate their reported statistical uncertainties. Even with the greater sensitivity, larger sample size, and deeper redshift range of $\alpha.40$, the correlation function analysis allows for quite a large range of clustering scenarios.

We confirm the HIPASS result that HI-selected galaxies are among the most weakly clustered known class of galaxies, most comparable to, but still less clustered than, the faint late-type subsamping in 2dFGRS. This is in agreement with previous results, particularly with our understanding that ALFALFA galaxies tend to be blue, spiral, and late type galaxies which are already known to be weakly clustered [Norberg et al., 2002, Kauffmann et al., 2004, Zehavi et al., 2005]. Apart from the estimates of uncertainties, the clustering of $\alpha.40$ is in agreement with the HIPASS findings but not with Passmoor et al. [2011], which is not unexpected giving the weaknesses – in particular extremely small sample size – of the latter.

6.4.2 ξ(r) via the Inversion Method

ξ(r) is only tidily related to $\Xi(\sigma)/\sigma$ if we assume that the underlying physics of galaxy formation dictates a power-law form for ξ(r). In the previous section, we calculated the correlation function of gas-rich galaxies under that assumption, but it is also possible to avoid that assumption and obtain ξ(r) by direct inversion of the projected correlation function Ξ. The inversion method tests the power-law assumption, though it is a noisy measurement and results in large scatter.

Following Meyer et al. [2007], Hawkins et al. [2003] and Saunders et al. [1992], we take our measurement of $\Xi(\sigma)$ to represent an underlying step function form with values Ξ_l in intervals with centers σ_l, rearrange Equation 6.17, and interpolate between bins to give:

$$\xi(r = \sigma_i) = -\frac{1}{\pi} \sum_{j \geq i} \frac{\Xi(\sigma_{j+1}) - \Xi(\sigma_j)}{\sigma_{j+1} - \sigma_j} \times \ln \frac{\sigma_{j+1} + \sqrt{\sigma_{j+1}^2 - \sigma_i^2}}{\sigma_j + \sqrt{\sigma_j^2 - \sigma_i^2}} \tag{6.20}$$

The sum in Equation 6.20 is truncated so that $\sigma_{max} = \pi_{max}$ for the value of π_{max} used in Equation 6.16.

The projected correlation function and ξ(r) obtained by the inversion method are in excellent agreement, as shown in Figure 6.7, where the points are the inversion ξ(r) with on-diagonal uncertainties and the overplotted dashed line is the best-fit model for the projected correlation function. It is also clear that the inversion method is more vulnerable to noise, motivating its use as a check on the assumed shape of the correlation function.

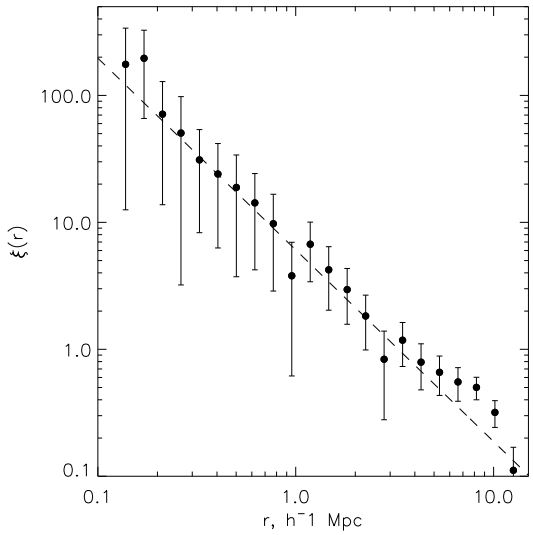

Figure 6.7 ξ(r) obtained via the inversion method, with the fit to the projected correlation function overplotted as a dashed line to demonstrate agreement.

6.4.3 Systematics and Methodology

Following Hogg et al. [2010], we do not fit a linear regression model to the logarithm of our power law correlation function, but instead compute least squares between the full value range for our observations and models. We find that this reduces the scatter in our estimated parameters, beyond being more statistically sound. We did, however, check for consistency between this fitting method, linear regression to the logarithm of the power law, and linear regression to the logarithm using only the on-diagonal elements of the covariance matrix, and find that there is no statistically significant difference between these three methods. Our findings are robust.

We consider whether extreme redshift distortions nearby, for small values of cz_{CMB}, could be contaminating our results. To test this possibility, we repeat the measurement, this time excluding galaxies within $cz_{CMB} < 2{,}000$ km s^{-1},

as well as 3,000 km s^{-1}. This reduced the sample size to $\sim$ 9,300 and $\sim$ 8,900, respectively, but we found no difference in the final fitting parameters r_0 and γ. We conclude that there is no advantage to be gained in eliminating nearby galaxies from $\alpha.40$ for the correlation function analysis.

The expression for J_3 in Equation 6.12 requires an expression for the shape of the correlation function. In calculating $\xi(\sigma, \pi))$ there is therefore a presumably small dependence on the as-yet unknown parameters s_0 and γ. As briefly mentioned in Section 6.3.1, one possible way to avoid any potential problems is to use the fiducial optical sample parameters s_0=5.0 and γ=-1.8 to calculate the parameters for an HI-selected sample, and then iterate towards a stable solution. In attempts to do this, we find that there is no significant difference between the parameters estimated via these two methods, and confirm that ξ is not dependent on the precise form of J_3. Such iteration does not provide any advantage. We demonstrate this in Figure 6.8, which displays the error contours on the power-law fit parameters for a sample limited to cz$_{CMB} >$ 2,000 km s^{-1}, and for J_3 using the approximate parameters estimated in Section 6.4.1. The results are very close to those in Figure 6.6 and the 1σ estimated parameters are identical.

6.5 Discussion

6.5.1 Comparison with Mock Catalogs

The correlation function of gas-rich galaxies has implications for the improvement of galaxy simulations, by providing an observational constraint for the results of simulations. This work will allow a better match between simulations

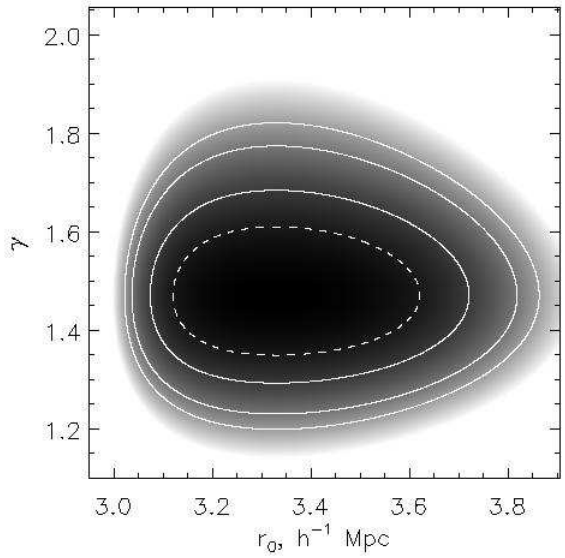

Figure 6.8 χ^2 contours for γ and r_0 (h^{-1} Mpc) excluding all galaxies with cz_{CMB} $<$ 2,000 km s^{-1}. The parameters used in J_3 are approximations of the HI-selected r_0 (3.4 h^{-1} Mpc) and γ (-1.5). The dashed contour gives the 1σ projected uncertainties on each parameter, and the solid contours give 1, 2 and 3 σ fits, respectively, to the pair of parameters.

and the observed relationship between gas mass and clustering properties. Simulations are just now progressing to the point where reasonable, realistic cold HI gas masses can be assigned to galaxies. In this section, we will compare the results of the correlation analysis of $\alpha.40$ with presently available cold dark matter simulations.

We are limited in our ability to compare our observations to simulations by what is available publicly. In Chapter 3 [Martin et al., 2010], we took advantage of the Obreschkow et al. [2009] (hereafter O09) simulation, which assigned cold gas to galaxies from the De Lucia and Blaizot [2007] catalog of Millennium Simulation galaxies. We found that O09's simulation provided a reasonable fit to the observed HI mass function. However, this catalog may not be ade-

quate for comparison with an observed correlation function, due to some of its limitations. In particular, O09 caution that their work applies well to regular spiral galaxies but not to elliptical galaxies; the mass resolution of the simulation prevents them from applying their findings to faint, low surface brightness, or low-mass galaxies. Given the known correlations between galaxy type and clustering, and between HI mass and luminosity, it would be difficult to use this catalog to explore the relationship between current simulations and ALFALFA's observations. Furthermore, O09 did not themselves carry out this analysis.

As an illustration of potential problems with the correlation function analysis in O09, Figure 6.9 shows some properties of galaxies at z=0 in the simulation. We would expect gas-rich galaxies to have low luminosities and high specific star formation rates, or, approaching the problem differently, we would not expect high luminosity galaxies with little to no star formation to be gas rich. Figure 6.9 illustrates that the simulation includes an abundance of objects that would be considered extreme based on observations of the local Universe. For these reasons, we do not here compare ALFALFA to the Millennium Simulation.

Kim et al. [2010] have provided another option for comparison, using a set of four GALFORM semi-analytical models that treats a range of processes which influence gas reservoirs, including cooling, ram-pressure stripping, mergers, star formation, and supernova feedback. They report results over a range of redshifts, but for this work only their results at z = 0 are of interest. They find that the galaxy-galaxy correlation function of the simulations is consistent with those found for HIPASS, and confirm that their simulation shows gas-rich galaxies as being significantly less clustered than dark matter. Differences between the models and the scales at which those differences are important can be used

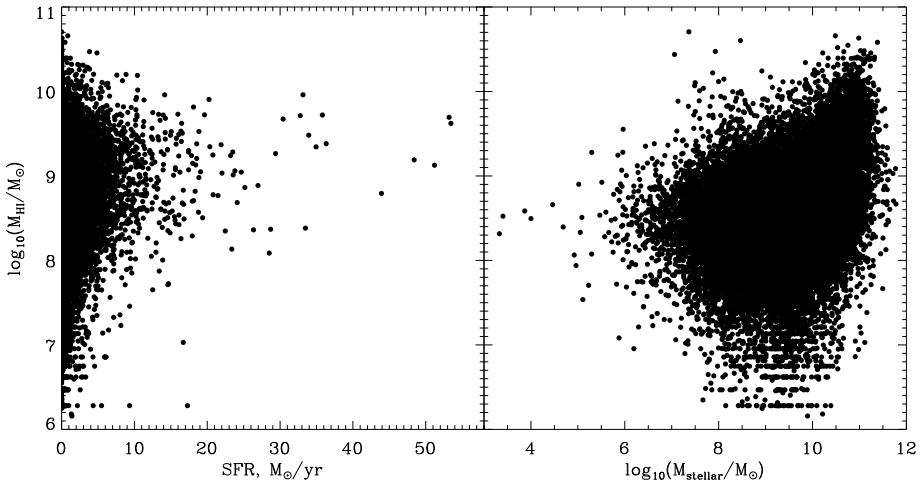

Figure 6.9 Left: HI mass vs. star formation rate ($M_\odot$ per year) in O09. Right: HI mass vs. stellar mass in O09.

to highlight potential problems in the assumptions, such as models that over-predict the gas richness of satellites.

In Figure 6.10, we compare the models presented in Kim et al. [2010] with the observed $\alpha.40$ correlation function for HI-selected galaxies. The models include that of Bower et al. [2006], labeled as Bow06; a modified version of the same, labeled MHIBow06; a version that uses a slightly different background cosmology that is in better agreement with the WMAP parameters, labeled GpcBow06; and, finally, the model of Font et al. [2008], labeled as Font08. In all models, only galaxies with $M_{cold} > 10^{9.5}$ h^{-2} $M_\odot$, where M_{HI} = 0.76 M_{cold}/(1+.04), are included, which matches the HIPASS galaxy selection but may be more massive than would be ideal for matching $\alpha.40$. The models are described in detail in Kim et al. [2010], and here we only discuss the main differences that may be relevant for a comparison to $\alpha.40$.

Bow06, MHIBow06, and Font08 all use the Millennium Simulation to track

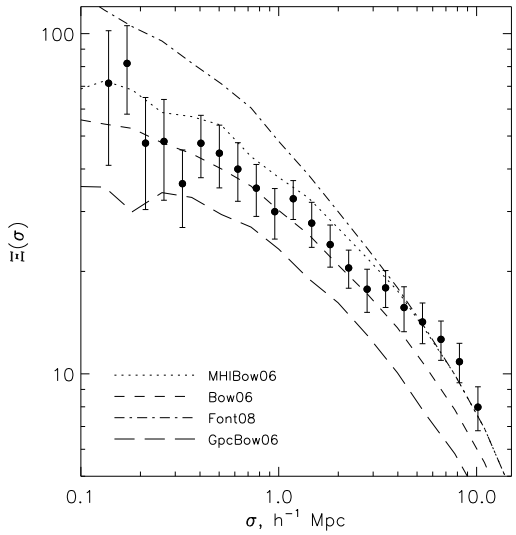

Figure 6.10 Models for $\xi(\sigma)$ from Kim et al. [2010], compared with $\alpha.40$ (filled points with error bars).

galaxies and halos, while GpcBow06 uses a different method involving merger trees and a large box size. Bow06 and Font08 are able to match optical luminosity functions, but both overpredict the abundance of HI in low-mass galaxies. MHIBow06 was created by adjusting the star formation timescale in Bow06, thereby fixing this excess while maintaining the agreement with optical properties of galaxies. GpcBow06, finally, also has a modified star formation prescription which better fits the HI mass function compared to Bow06.

In Figure 6.10, it is clear that MHIBow06 and Bow06 fit the observed HI correlation function well on small scales, while Font08 drastically overpredicts and GpcBow06 drastically underpredicts the strength of clustering for gas-rich galaxies on those scales. At large scales, both GpcBow06 and Bow06 underpredict the clustering strength, while Font08 and MHIBow06 follow it closely. The MHIBow06 model appears to be most consistent with the clustering of gas-rich galaxies over the full range of accessible scales.

Part of these differences may be due to the mass resolution of the models, given that $\alpha.40$ probes to significantly lower masses than the HIPASS survey for which these models were designed. What is clear, however, is that $\alpha.40$ already provides constraints that can begin to differentiate between successful and unsuccessful models, and the full ALFALFA sample should be able to provide very robust constraints for testing simulations that take HI into account.

6.5.2 The Bias Parameter for HI-Selected Galaxies

The bias between any two classes of objects indicates their relative clustering strength. For cosmological purposes, we are interested in comparing the clustering of types of galaxies with the underlying dark matter halo distribution, in order to understand how well future surveys would probe the true (baryonic + dark) mass distribution. The comparison is achieved through the linear bias parameter b_0:

$$\xi_{gal}(r) = b_0^2 \, \xi_{DM}(r) \tag{6.21}$$

If $b > 1.0$, as is the case for red galaxies which tend to be found in clusters [Seljak, 2000], then the distribution is positively biased with respect to the dark matter. For galaxies like HI-selected populations, $b < 1.0$ and they are said to be antibiased.

As a proxy for the underlying dark matter distribution, we use the correlation function of dark matter halos from the Millennium Simulations, given in Springel et al. [2005] as a function over the same scales that we are interested

in. In Figure 6.11 we compare that correlation function to the $\alpha.40$ observation of ξ(r) for HI-selected galaxies. The dark matter correlation function deviates strongly from a power law on small scales, an indication of the well-known fact that bias is scale-dependent. Dark matter is, as expected, significantly more strongly clustered than this particular population of galaxies.

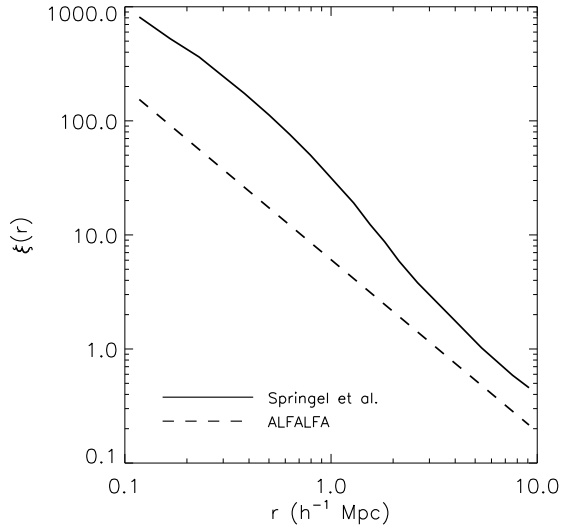

Figure 6.11 The real-space correlation function ξ(r) for dark matter from the Millennium Simulation (solid line) for for HI selected galaxies from $\alpha.40$ (dashed line; model fit given in Table 6.1).

In Figure 6.12, we display the bias parameter as a function of scale. The error bars are based only on the $\alpha.40$ uncertainties and they assume that there is no uncertainty in the Millennium Simulation's measurement of the correlation function. This Figure reflects what we already understand about the clustering properties of HI selected galaxies, taking into account the results of Chapter 5: on small scales, the clustering of gas-rich galaxies is weaker, and on ever-larger scales the distribution of gas-rich galaxies begins to more closely reflect the underlying matter distribution. Basilakos et al. [2007] measured the linear bias parameter on large scales for the HIPASS sample using a different technique.

They explored modeled dark matter power spectra for different values of b_0, including bias and assuming a concordance cosmology, and identified the most likely bias parameter. They found b_0 = 0.7 $\pm$ 0.1, in general agreement with the findings for $\alpha.40$. The preliminary work of Passmoor et al. [2011] is generally consistent with this result, though that work does not capture our finding that the sample becomes significantly less biased on large scales.

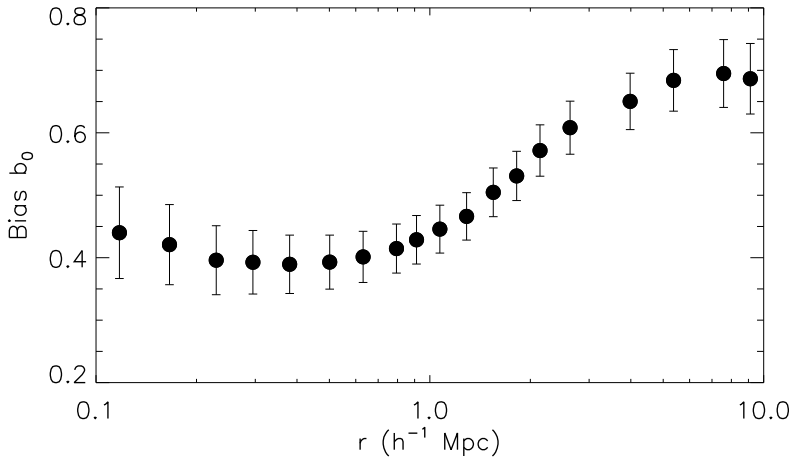

Figure 6.12 The bias parameter b_0(r).

Recently, Marín et al. [2010] have used a simple model relating M_{HI} to M_{DM} to estimate HI masses of Millennium Simulation halos and then investigated the bias of HI with respect to the halo distribution. At z=0, they estimate that the overall linear bias parameter on large scales is $\sim$0.8, in general agreement with our findings. Their Figure 6 is more comparable to our Figure 6.12, and shows the same overall rise of the bias with increasing scale. Their models also predict that the bias will rise sharply with redshift, with the linear bias parameter reaching $b_0 \sim 2$ by z $\sim$ 4.

The $\alpha.40$ observations and Marín et al. [2010] predictions have implications

for large-scale 21 cm galaxy surveys and intensity mapping projects with such instruments as the Square Kilometer Array (SKA). If the theoretical results reflect the true evolution of HI gas in the universe, then these projects can expect strong 21 cm signals at a range of redshifts. Perhaps more importantly, the $\alpha.40$ observation of the correlation function at low redshift provides a robust baseline constraint for the development of SKA model predictions, and future simulations will need to match both the HI mass function and the correlation function at z $\sim$ 0.

Because the HI-selected galaxy bias is likely to be strongly dependent on HI mass, with low-mass objects severely antibiased with respect to the underlying dark matter distribution, high-redshift surveys which are sensitive only to the high-mass end of the HIMF should expect to be mildly antibiased at low redshifts and increasingly positively biased at intermediate to high redshifts. We explore the mass dependence of the correlation function in Chapter 7.

6.6 Summary and Conclusions

We have used the $\sim$10,000 galaxy $\alpha.40$ sample to measure the correlation function of HI-selected galaxies in the local universe. We use bootstrap resampling and a full covariance analysis in order to model the real-space correlation function on scales < 10 h^{-1} Mpc as a power law, $\xi(r)=(r/r_0)^{-\gamma}$. We find that γ = 1.51 $\pm$ 0.09 and that the clustering scale length is r_0 = 3.3 + 0.3, -0.2 h^{-1} Mpc. Furthermore, we show using a direct inversion method that the observed $\alpha.40$ real-space correlation function closely follows this power law. Our findings are shown to be robust against the precise form of the weighting used in the pair-

wise estimation of $\xi(\sigma, \pi)$ and the $\alpha.40$ sample selection criteria.

The clustering of HI-selected galaxies is significantly weaker than the clustering of general populations of optically-selected galaxies, and is most closely comparable to samples of faint, late-type and/or blue galaxies found in optical surveys. Available models of HI in simulated galaxies are in general agreement with our observations, and the $\alpha.40$ measurement of the correlation function is robust enough to begin constraining these models.

Finally, we measure the bias parameter for $\alpha.40$, using the correlation function of dark matter haloes from the Millennium Simulation, and find that the small-scale clustering of HI galaxies is severely antibiased with respect to the underlying dark matter distribution. On large scales the antibiasing becomes more moderate. We suggest that isolating the high-mass galaxies in $\alpha.40$ will show that this population more closely follows the true mass distribution and that an abundance of low-mass galaxies in underdense voids partially explains the strong antibiasing observed. We explore this idea in the following Chapter.

CHAPTER 7

THE DEPENDENCE OF ξ(R) ON GALAXY PROPERTIES

7.1 Introduction

Galaxies that are of early type, brighter [Norberg et al., 2002], and redder [Li et al., 2006a] cluster more strongly. HI-selected galaxies tend to be blue and to extend to lower surface brightnesses than optically-selected samples, and they bear out the expectation that they would therefore be weakly clustered relative to other populations. Using the SDSS g-r color for a sample of $\sim$280,000 galaxies, Wang et al. [2007] cross-correlated galaxies with different properties, which informs our expectation for the $\alpha.40$ sample. The cross-correlation analysis finds that small-scale differences (on scales smaller than 1 h^{-1} Mpc) are important in understanding galaxy clustering, and that for any given luminosity red galaxies are more strongly clustered than blue galaxies irrespective of scale.

Some work has previously investigated the relationship between clustering and galaxy properties for the HIPASS HI-selected sample. Meyer et al. [2007] and Basilakos et al. [2007] both identified this sample as the weakest clustering population of galaxies known, but their results regarding the mass dependence of the clustering were in conflict. While the HIPASS team itself found a statistically insignificant difference between 'high' and 'low' HI mass galaxies, Basilakos et al. [2007] found that high-mass galaxies clustered more strongly. The HIPASS team also found that HI-selected galaxies, like optically selected galaxies, tend to cluster more strongly when they are more optically luminous, but the observed difference was very weak. The preliminary study of the ALFALFA correlation function carried out by Passmoor et al. [2011] did not see a trend

with mass, but cautioned that their small sample carried large uncertainties into their power-law fits.

Given the results of Chapter 5, which investigated the dependence of HI mass on galactic environment, we can make some predictions for the clustering dependence of galaxies in $\alpha.40$. We expect more HI-massive galaxies to be more strongly clustered, captured in the tendency of the HIMF parameter M_* to shift to higher and higher masses as density increases. This is due to two effects. First, higher density environments have larger numbers of high-mass halos which would attract larger masses of baryons; secondly, we also expect that higher-mass galaxies would be more immune to processes that may strip or alter gas reservoirs in high-density environments and that galaxies with large HI masses would preferentially be found in these regions. The findings of Chapter 5 complicate the issue of the influence of gas processes as a function of environment, but lead us to the same conclusion regarding the clustering of high-mass galaxies.

Accordingly, we anticipate that low-mass galaxies will be less clustered, because they prefer void environments. The results of Basilakos et al. [2007], in particular, lead us to expect to find extremely weak clustering for the lowest-mass HI-selected galaxies. While we expect that the optical properties will confirm previous findings (that is, that red and bright galaxies are more strongly clustered), it is important that we understand the influence of various properties to truly understand the properties of HI-selected galaxies, their clustering, and their environmental dependence.

In exploring the dependence of the real-space correlation function ξ(r) on galaxy properties, ALFALFA has a distinct advantage given its high sensitiv-

ity across 5 orders of magnitude in mass, its blind ability to detect both low surface brightness and extremely large, bright spirals, and its overall sample size. Throughout this chapter, we will directly compare the observed correlation functions, rather than model fits, due both to the large uncertainties introduced by subsampling $\alpha.40$ by galaxy properties and to the availability of this more direct comparison, which is not possible when comparing work from various authors as in Chapter 6.

7.2 Method

The methods for measuring the correlation function and the uncertainties follow those presented in Chapter 6, with one exception. For the weighting applied to each galaxy, we once again use

$$w_i = \frac{1.0}{1.0 \; + \; 4\pi \; N_D \; S(r_i) \; J_3(s)} \tag{7.1}$$

where the expression J_3 includes an assumed form for the real-space correlation function ξ(s). In this case, however, we used the previously-derived HI quantities for this weighting: r_0 = 3.4 h^{-1} Mpc and γ = -1.5. We make this choice to be conservative, given that each subsample will have a different redshift distribution. Since this will lead to different selection functions and distributions of weights, another choice of the weighting function could lead to an artificial bias between subsamples.

Once again, we used the full covariance matrix to estimate uncertainties, which were quite large due to the small subsample sizes. To be careful, we also

checked all model fits using only the on-diagonal errors (standard deviations) and found that the results are consistent in every case. For display purposes, our comparisons only include the on-diagonal elements of the covariance matrix.

7.2.1 Subsampling by HI Mass

Previously, small sample sizes in blindly HI-selected catalogs of galaxies made it difficult to make reasonable cuts in certain properties, such as mass. In order to investigate the dependence of clustering on HI mass while keeping sample sizes large enough for a reasonable estimate of ξ(r), Meyer et al. [2007] had to make a single cut, choosing the HIPASS value of $M_*/4$. They characterized everything above that mass, $\sim 10^{9.56}\ M_\odot$, as high mass, and everything below it as low mass. They found no significant difference between the two samples and claimed that their HI mass limits do not help to select gas-rich galaxies on the basis of environment. Basilakos et al. [2007] also made a single cut at a slightly different mass scale, $\sim 10^{9.70}\ M_\odot$.

In Figure 7.1, we show how this single mass cut would relate to the full distribution of masses in $\alpha.40$. The histogram shows the mass distribution, with M_* overplotted as a dashed line and with the HIPASS mass cut (corresponding to their $M_*/4$) overplotted as a dotted line. It is clear that the galaxy counts in the 'low mass' HIPASS sample would be dominated by galaxies near the peak of the distribution with masses that would be considered typical, rather than by true dwarf galaxies.

Such choices in a strict, single mass-cut are not justified by the physics, only by statistical requirement. Taking M_*, the turnover mass for the Schechter func-

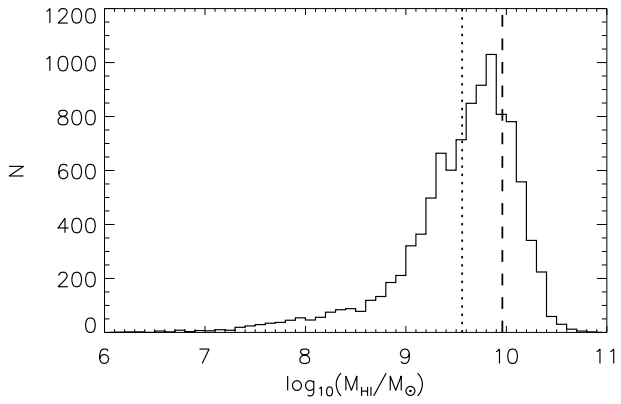

Figure 7.1 The mass distribution of galaxies in $\alpha.40$ compared to the HIPASS mass cut (leftmost dotted line) with the $\alpha.40$ M_* value indicated by the dashed line.

tion, as a 'typical' HI mass, there is no substantive difference between galaxies with M_{HI} $\sim$0.25 M_* compared to galaxies masses of, for example, M_{HI} $\sim$0.75 M_*. Taken as an ensemble, such galaxies should be qualitatively similar, and these weak mass cuts will not distinguish between various galaxy environments or evolutionary tracks. Such cuts are especially egregious because of the magnitudes of mass errors, particularly in the shallow HIPASS survey, which is limited to objects within 6,000 km s^{-1} for the analysis of the correlation function. Without correction for peculiar velocity, strict, single mass cuts will in effect smudge two categories of galaxies together.

The larger sample size afforded by ALFALFA allows us to make mass cuts that are more indicative of low-mass and high-mass populations of galaxies. Our mass cuts, and the number of galaxies contained within each subsample, are listed in Table 7.1; it is worth noting that gaps are left between each subsample to try to identify truly different samples of galaxies. The low-mass sample includes only galaxies with masses smaller than 10% M_*, while the high-mass

Table 7.1. Mass Subsamples

Sample	Minimum $\log_{10}$ M_{HI}	Maximum $\log_{10}$ M_{HI}	N_{gal}	N_{random}
Low Mass	6.0	8.96	1302	60000
Medium Mass	9.44	9.80	2973	60000
High Mass	9.96	11.0	2355	60000

sample includes galaxies with masses larger than M^*; the medium-mass sample includes an intermediate set of galaxies with masses 30 - 70% M_*.

The regions delineated by each subsample are shown as shaded boxes in Figure 7.2. Both the high- and medium-mass subsamples are dominated by galaxies near the peak of the distribution, and thus we would not expect to find a stark difference between the clustering characteristics. The low-mass sample, however, represents a truly distinct set of objects. When the full ALFALFA catalog is available, it is likely that these cuts can be further refined, with a low-mass sample including only objects with $\log M_{HI} < 8.0$ and a high mass sample with $\log M_{HI} > 10.2$ (a factor of 2 increase over M_*).

Figure 7.3 shows the expected redshift distribution of the low-mass sample (dashed line) and the high-mass sample (solid line), based on the selection function of the ALFALFA survey. Due to large-scale structure effects, the observed redshift distributions (histograms) have peaks corresponding to overdensities and dips corresponding to underdensities. While the correlation function method is designed to take large-scale structure in the survey volume into account, it is possible that the large nearby overdensities, such as the Virgo cluster, may bias the $\alpha.40$ clustering results and the mass subsamples are particularly vulnerable to such effects.

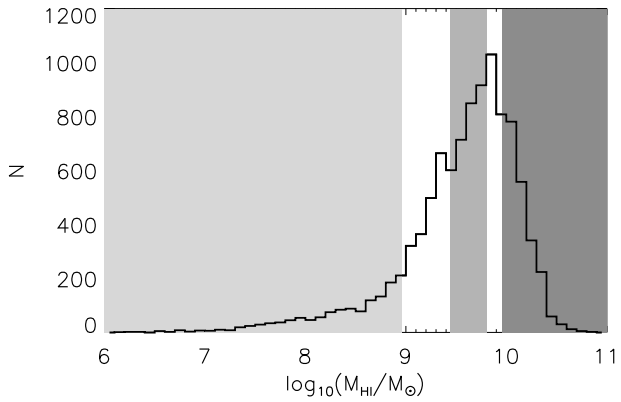

Figure 7.2 The mass distribution of galaxies in $\alpha.40$ with regions marking the low-mass sample (lightest tone), medium-mass sample (midtone), and high-mass sample (darkest tone).

7.2.2 Subsampling by Optical Luminosity

Attempts to examine the optical properties of ALFALFA galaxies are plagued by difficulties, due to the limitations of available optical catalogs in the survey region, of which SDSS is the best. Using the SDSS immediately limits the sky area we are able to use to the Spring $\alpha.40$ region described in Chapter 5. This reduces the potential sample to only $\sim$ 6,800 objects. For all of those objects, we used an internal cross-match catalog by Haynes that matches each ALFALFA galaxy to its best counterpart in SDSS. All optical properties reported in this chapter use these cross-match identifications.

Beyond that, however, determining the optical properties of HI-selected galaxies is particularly difficult for objects at the extreme ends of the HI mass function. Photometry is difficult to obtain for the lowest-mass galaxies due to SDSS's well-known 'shredding' issue, and high-mass galaxies contend with the SDSS pipeline's trouble with extended objects. Even assuming well-measured

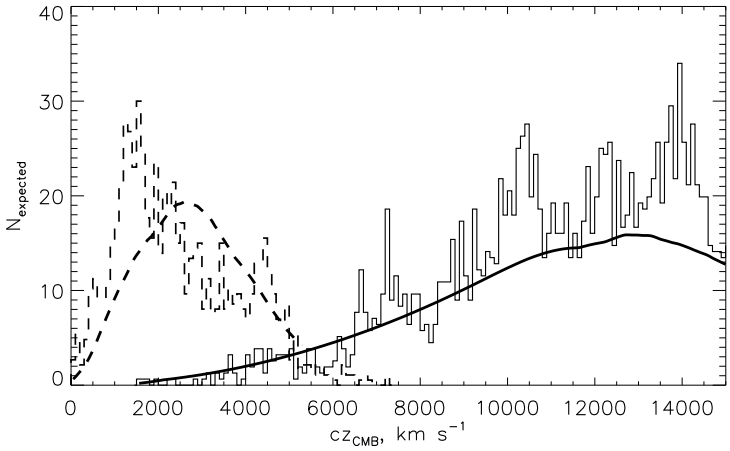

Figure 7.3 The smoothed anticipated redshift distribution, based on the selection function of the ALFALFA survey, for the low-mass (dashed) and high-mass (solid) subsamples, along with the observed distribution for each plotted as a histogram.

apparent magnitudes, correcting for internal extinction is especially difficult for low-mass galaxies, which may not have disks allowing the usual estimates of the inclination angle, and which may not follow extinction laws derived from non-dwarf galaxy samples.

For those reasons, we sought to use a definition of optical luminosity that would allow us to split $\alpha.40$ into a bin of confidently faint galaxies and a bin of confidently bright galaxies. The simplest facet of this strategy was to exclude a large central bin, which would be likely to include some true medium-luminosity galaxies as well as contamination from improperly measured galaxies that should truly fall into the faint or the bright bin.

As an additional precaution, we estimate luminosities using two different extinction corrections in the literature for SDSS galaxy samples, and keep only those that fall into the faint or bright sample according to both definitions. We

Table 7.2. Optical Luminosity Subsamples

Sample	M_r Limit	N_{gal}	N_{random}	Min cz_{CMB}, km s^{-1}	Max cz_{CMB}, km s^{-1}
Bright	$<$ -20.5	2074	42000	4,000	15,000
Faint	$>$ -18.5	1876	40000	0	12,000

selected previous work which focused on galaxy samples with similarities to $\alpha.40$. The extinctions are estimated using Cho and Park [2009], which used late-type galaxies with $0.025 < z < 0.044$, and Shao et al. [2007], which includes only spiral galaxies but extends up to $z < 0.22$. We use flow model distances, the same as are applied to all ALFALFA galaxies, to convert to absolute magnitudes, and used the SDSS Galactic extinction values to modify both colors and apparent magnitudes.

The luminosity cuts on absolute Petrosian r-band magnitudes for the bright and faint samples are given in Table 7.2, along with the number of galaxies and the redshift range for each sample. These cuts were selected on the basis of the distribution of values for the Spring $\alpha.40$ subsample, but they also result in $\alpha.40$ luminosity subsamples that are similar in character to the faint and bright bins in Zehavi et al. [2005]. While contamination of the bright and faint galaxy samples is possible and, at least to some degree, likely, the steps taken to reduce contamination should have produced ensembles that are statistically distinct in their true luminosity.

7.2.3 Subsampling by Optical Color

In order to subdivide $\alpha.40$ into a red and a blue subsample, we use the SDSS $g - r$ color, following examples in the literature including Wang et al. [2007] (where $g - r$ was used to study clustering characteristics), Zehavi et al. [2005], and West et al. [2010] (where $g - r$ was used specifically to explore the optical properties of HI-selected galaxies). We take advantage of Galactic extinction values provided by SDSS to correct the colors for Galactic reddening, but, as in Baldry et al. [2004], do not attempt to correct the colors for internal extinction. Again, we attempt to be conservative by leaving a large gap between our red and blue subsamples.

The color distribution is displayed in Figure 7.4, with dashed lines indicating our cuts in color space; we leave a large range of color values out of the analysis in order to be conservative. Compared to Baldry et al. [2004], which investigated the color-magnitude relationship for a large sample of low-redshift (0.004 $<$ z $<$ 0.08) galaxies in SDSS based on $u - r$ color, ALFALFA galaxies have a significantly bluer distribution. This is not unexpected, but we also observe the same bimodal distribution, with distinct red and blue Gaussian contributions, in proportions like those seen for the faint SDSS galaxies. Our color cuts were selected to exclude the region of significant overlap between these two populations. Using a small (N = 195) sample of HI-selected galaxies, West et al. [2009] found no significant bimodality nor red sequence, but did note the existence of red galaxies with $g - r > 0.5$. These authors also noted that those red galaxies in their sample exhibited emission lines, suggesting that the red colors of these galaxies indicate a period of star formation in the distant past while their lines and HI content indicate recent accretion of gas and a new period of star

Table 7.3. Optical Color Subsamples

Sample	$g-r$ Limit	N_{gal}	N_{random}
Red	≥ 0.65	1408	40000
Blue	≤ 0.35	2032	42000

formation.

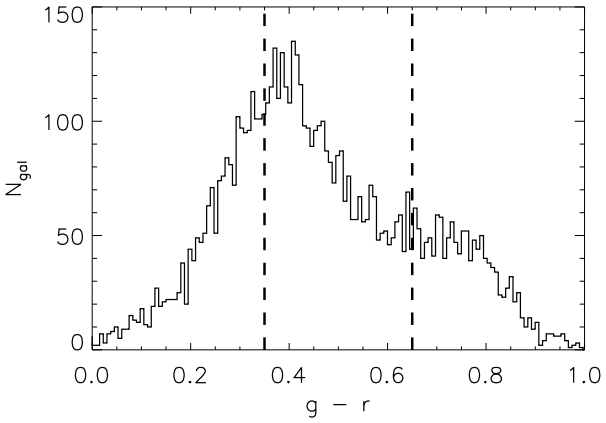

Figure 7.4 The $g-r$ color distribution in Spring $\alpha.40$ based on SDSS Petrosian magnitudes. The color cuts are indicated by the dashed line, with blue galaxies falling leftward of the leftmost line, and with red galaxies falling rightward of the rightmost line.

The adopted cuts on apparent, Galactic reddening-corrected, Petrosian $g-r$ color are given in Table 7.3.

7.3 Results

7.3.1 Clustering Dependence on HI Mass

In Figure 7.5, we show the two-dimensional $\xi(\sigma, \pi)$ obtained for the low-mass subsample (left) and the high-mass subsample (right); we use the same color scale in each image, for a more direct comparison, though it is clear that little information is contained in most pixels for the low-mass case. In each case, the smaller sample size results in a distinctly noisier image than that seen for the full sample in the previous chapter, and it is clear that the low-mass sample's results are only informative out to scales $\sim$10 h^{-1} Mpc.

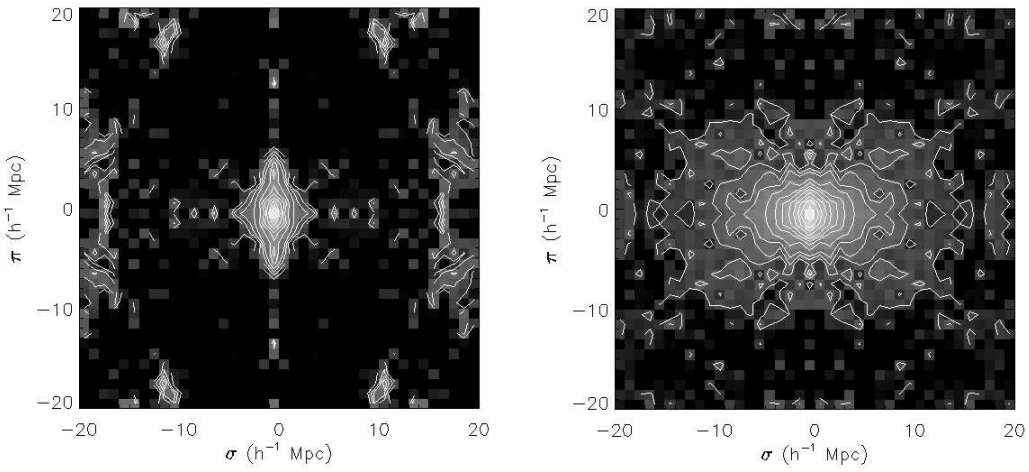

Figure 7.5 The two-dimensional $\xi(\sigma, \pi)$ for the low HI mass (left panel) and high HI mass (right panel) subsamples.

$\xi(\sigma, \pi)$ reveals that the low-mass sample is, due to structures like the very nearby Virgo cluster, more susceptible to the 'stretching' effect that is due to redshift distortions (peculiar velocities), while the high-mass sample maintains

the character of the overall sample. The medium-mass subsample (not shown) is similar to the high-mass subsample. If the distortions are not adequately corrected by integrating over the π dimension, then the low-mass sample may appear more clustered in this analysis than it truly is.

The projected correlation function for the low HI mass and high HI mass cases are displayed in Figure 7.6, and we do observe a significant difference between the clustering. Particularly on larger scales, high-mass HI-selected galaxies cluster more strongly than their low-mass counterparts. The modest redshifts of the low-mass catalog only allow us to extend our conclusions to modest scales. Because of the contamination by the Virgo cluster, the result shown here may be biased in favor of stronger clustering for low-mass galaxies. When the full ALFALFA sample is available, this bias may be ameliorated by the addition of more nearby volume or the ability, due to larger sample sizes, to excise the cluster region altogether while maintaining statistically significant galaxy counts.

Taken with the results of Chapter 5, the clustering dependence on mass indicates that small-scale processes in dense environments have a relatively weak influence on the HI masses of constituent galaxies, but that on larger and large scales, environmental processes become important.

The distinct differences between the low- and high-mass samples make it possible to determine reliable power law ($\xi(r)=(r/r_0)^{-\gamma}$) parameters, which are listed in Table 7.4 along with their uncertainties. The two extreme-mass subsamples clearly cluster differently. The medium-mass sample is included for comparison, and while it is more strongly clustered than the low-mass and more weakly clustered than the high-mass subsample, the model parameters are con-

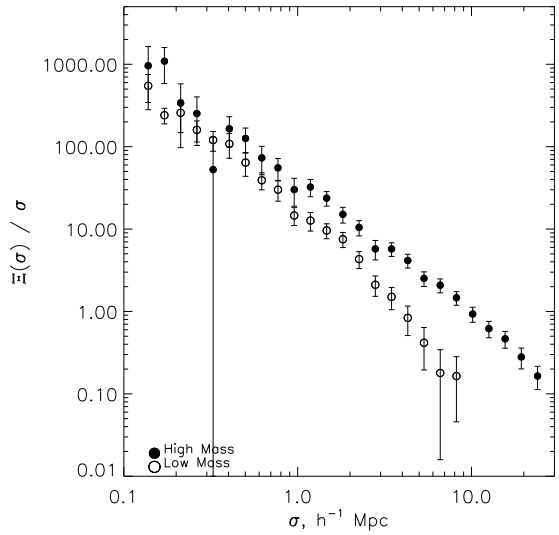

Figure 7.6 The projected correlation function $\Xi(\sigma)/\sigma$ for the low HI mass (open circles) and high HI mass (filled circles) subsamples.

sistent within the uncertainties. The model fits are in all cases highly uncertain.

7.3.2 Clustering Dependence on Optical Luminosity and Color

In the case of optical properties, the clustering differences are small and model fits, which reduce the total information, are therefore uninformative [Li et al.,

Table 7.4. Mass Subsample Power-Law Models

Sample	r_0 (h^{-1} Mpc)	γ
Low Mass	$2.42^{+.24}_{-.18}$	$1.76^{.15}_{.15}$
Medium Mass	$2.80^{+.47}_{-.29}$	$1.45^{+.23}_{-.17}$
High Mass	$3.50^{+.40}_{-.55}$	$1.59^{+.32}_{-.21}$

2006a, Zehavi et al., 2005]; throughout this section, we will instead directly compare the data.

The projected correlation function for the faint and bright subsamples are displayed in Figure 7.7. The dependence on optical luminosity is weaker than the dependence on HI mass, and is less dependent on scale. However, small-scale processes appear to potentially have a stronger effect on optical luminosity than they do on gas mass. As in optically-selected samples, the fainter galaxies show weaker clustering. On the largest scales, there appears to be no distinction between the two populations.

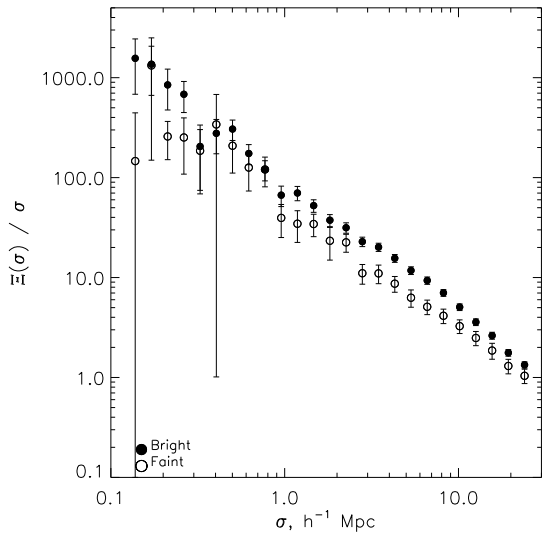

Figure 7.7 The projected correlation function $\Xi(\sigma)/\sigma$ for the faint (open circles) and bright (filled circles) subsamples.

Given that the overall $\alpha.40$ sample is primarily blue, these results are in agreement with Swanson et al. [2008], who found that the clustering of galaxies with blue colors is only weakly dependent on luminosity.

The projected correlation function for the red and blue subsamples are dis-

played in Figure 7.8, and, as in the previous cases, the subsample that would be expected to cluster the most strongly (red) is indicated with a filled circle. Once again, the dependence on this optical property is weaker than the dependence on HI mass, and as expected, the red galaxies cluster more significantly at all scales. However, the correlation between blue color and faintness results in no ability to compare the two populations at the smallest scales ($r \lesssim 0.8\ h^{-1}$ Mpc). Despite the large uncertainties on small scales, it appears probable that, like luminosity, color is strongly influenced by the processes occurring on small scales in environments ranging from pairs to groups to clusters.

Interestingly, previous work has shown that galaxy color is more predictive of environment than other properties [Kauffmann et al., 2004, Blanton et al., 2005, Li et al., 2006a], while structural parameters driven by small-scale processes are not observed to drive this color dependence. We find that color is certainly predictive of clustering on small scales, which probe so-called 'local' environments, but that on larger scales, the HI mass dependence of the correlation function is more significant. These previous works have found that properties like morphology, concentration, and surface brightness, hallmarks of the well-known morphology-density relation (e.g. Dressler et al. [1997]), may be nuisance parameters for optically selected galaxies. Work specifically addressing the spiral sequence, however, has shown that clustering is significantly weaker for late-type spirals than for the whole spiral population [Guzzo et al., 1997]. Future work to explore this with an HI-selected sample which is severely biased with regards to these structural properties will help determine the relative strength of nature vs. nurture in galaxy evolution.

The approach described here did not correct for the reddening of the galax-

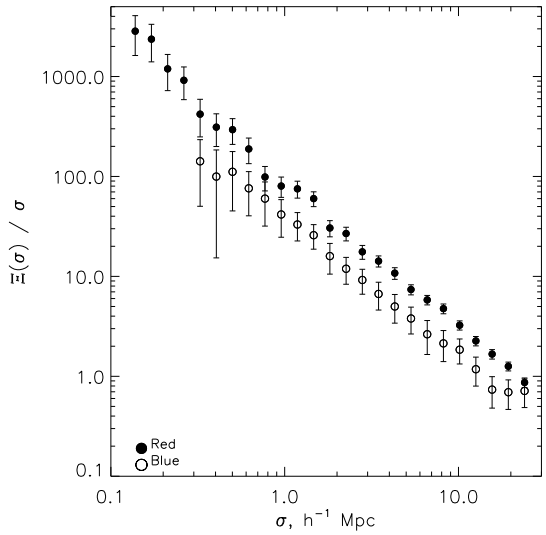

Figure 7.8 The projected correlation function $\Xi(\sigma)/\sigma$ for the blue (open circles) and red (filled circles) subsamples.

ies, and the result obtained may inadvertently include some level of blending between the red and blue populations. There is a very weak correlation between HI mass and optical color, shown in Figure 7.9. There is no substantial difference between the distribution of masses for the blue and red galaxies, but redder galaxies may tend to be slightly more massive. Thus, the observed clustering dependence is likely to be due, in part, to the different mass distributions of the subsamples.

Furthermore, while redder galaxies are generally known to cluster more strongly, the red galaxies in $\alpha.40$ are still typically bluer than the distribution seen in optically-selected samples. For comparison, the truly red galaxies in Baldry et al. [2004] and Blanton et al. [2003] show a strong peak at $g - r \sim 0.9$, where we see few galaxies. This is likely to explain the weakness of the clustering dependence on optical color for HI-selected galaxies compared to optically-selected samples.

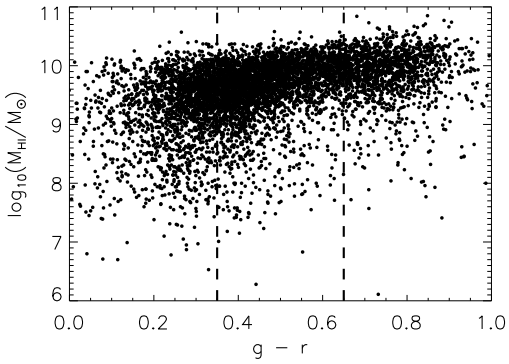

Figure 7.9 The relationship between the optical $g - r$ color obtained from SDSS and the HI mass for the Spring $\alpha.40$ sample. The red and blue color cuts are displayed as overplotted dashed lines.

Brighter galaxies are known to cluster more strongly in optically selected samples, but more luminous galaxies also tend to be red [Swanson et al., 2008, Zehavi et al., 2005], so it remains difficult to understand which properties are causative and which are merely correlated. Spring $\alpha.40$ could not reliably be subsampled into faint and bright samples by color, which would be one way to explore the dependence of the observed clustering on these properties. As the catalog becomes larger, this will be an important area of study in order to better understand the relationship between HI properties, optical properties, and environment.

7.4 Discussion & Conclusions

We find that high HI mass HI-selected galaxies are significantly more clustered than low-mass HI-selected galaxies, and find that this dependence is more pronounced on larger scales. This is counter to the findings of the HIPASS group [Meyer et al., 2007], though an independent analysis Basilakos et al. [2007] did

find a similar dependence on HI mass, albeit at a statistically weaker level than found here. Given the large sample size, we are also able to consider a medium-mass sample. Though it is not completely distinct from the high-mass sample, we confirm that clustering increases with mass consistently through the three subsamples.

This dependence implies that future 21 cm line surveys, which will be sensitive to extremely high HI mass galaxies beyond the M_* turnover of the HIMF, will be significantly less antibiased with regard to the underlying dark matter halo distribution. Furthermore, this observation implies that the influence of different processes that act upon galaxies in various environments can be determined by comparing high- and low-mass galaxies. Alternatively, the $\alpha.40$ finding suggests that the intensity mapping approach to 21 cm cosmology [Peterson et al., 2009, Pen et al., 2009] will continue to find success since large HI halos better trace the underlying dark matter on large scales.

Using optical data from the SDSS matched to galaxies in Spring $\alpha.40$, we estimate colors and extinction-corrected luminosities and investigate the dependence of clustering on these optical properties. Both optical luminosity and color are less significant to clustering characteristics than HI mass, but due to correlations between these properties we are unable to determine which optical properties are most closely tied to environment for HI-selected galaxies. We find that faint galaxies are less clustered than bright galaxies irrespective of scale, but that on the largest scales they may begin to cluster similarly; larger and contiguous volumes in future ALFALFA survey catalogs will allow us to extend this analysis to larger scales and determine whether the two populations are equally biased on those scales. The clustering difference based on color is

more pronounced than that based on luminosity, likely because redder galaxies of any luminosity are more tightly clustered than bluer galaxies [Zehavi et al., 2005].

Wang et al. [2007] found, in their cross-correlation analysis, that the most luminous galaxies with blue colors are strongly correlated on small scales with red galaxies. These authors suggest that some very luminous blue galaxies may be central, gas-enriched and actively star forming galaxies in groups with relatively low gas masses, which would explain the correlation function dependence on both HI mass and color at small scales and our finding that medium-density environments appear to be enhanced in their HIMF.

Ultimately, it is clear that HI-selected galaxies display the same general clustering dependences as optically-selected galaxies. The lowest-mass, reddest, least-luminous galaxies in ALFALFA are likely to represent the least-clustered class of galaxies in the local universe. Because of sample size, we cannot yet determine the relative role of HI mass, stellar mass/luminosity, and color in the observed clustering, and future work will involve these investigations as well as focusing on morphology and, possibly, emission lines indicative of current star formation.

The results of Chapter 5, which explored the dependence of the HI mass function on environment, along with the results of this Chapter, indicate that the processes that take place on small scales ($\sim$ 1 Mpc), such as ram-pressure stripping, are not as important to gas reservoirs and the characteristics of gas-bearing galaxies as previously thought. On larger scales, the clustering of high- and low-mass galaxies becomes significantly different, while the HI mass function at these scales shifts to an increasing proportion of high-mass galaxies in

denser regions.

The relationship between the faint-end slope α and the clustering dependence on mass, however, is less simple to understand. On small scales, high- and low-mass galaxies cluster similarly, while on the smallest scale considered in Chapter 5, the HIMF was found to be increasingly flattened with increasing density. The situation appears to change when environment is considered on larger scales, indicating that small-scale processes such as ram-pressure stripping are not the only processes that matter. In medium-density environments, HI reservoirs appear enriched (leading to steep faint-end HIMF slopes) and in high-density environments, HI-bearing galaxies appear to be stripped or depleted (leading to shallow faint-end slopes). In the correlation function ξ(r), that may appear as a break or deviation from a power law, but this analysis provides no evidence that such a break exists. When more volume is available and statistical uncertainties are reduced, this situation may change. The idea that different processes are necessary to explain the clustering dependence of different populations of galaxies and on different scales is not new [Li et al., 2006a] but our ability to explore this has been limited for HI-selected galaxies.

Recent work [Watson et al., 2011] suggests that the power-law shape of the galaxy-galaxy correlation function, generally, and the specific parameters r_0 and γ for any particular populations, depend primarily on processes that deplete the number of surviving satellite galaxies in a halo. Therefore, the relationship between the HIMF dependence on environment and clustering's dependence on HI mass (that is, on the abundance of dwarf galaxies) are important ingredients in a full understanding of which processes, including gas enrichment via late cold accretion, result in the population of gas-rich galaxies observed at z = 0.

CHAPTER 8

CONCLUSION AND FUTURE WORK

Current work using galaxies and their evolution to probe cosmology range from the discovery and characterization of ultrafaint dwarf galaxies [Ryan-Weber et al., 2008, Simon and Geha, 2007] to the impact of reionization on dwarf galaxies [Hoeft and Gottloeber, 2010, Ricotti, 2009b, Hoeft et al., 2008, 2006] to the relationship between galaxy properties and environment [Blanton and Moustakas, 2009, Park et al., 2007, Blanton et al., 2005, Zwaan et al., 2005]. These approaches all attempt to tease out the relationship between the universe we observe today and the evolution of dark and baryonic matter over cosmic time. As large-scale surveys improve our observational and statistical understanding of galaxies, and new simulations include important baryonic effects in attempts to match these observations, we are able to approach questions with a greater level of detail. To date, however, the astronomical community has been extremely limited in its ability to relate gas properties to other baryonic properties related to, e.g., galactic structure, star formation history, and interactions.

With the understanding of galaxy evolution stronger than ever, and with tantalizing questions about the details of galaxy evolution beginning to arise, the time is ripe for HI properties of galaxies to begin fully informing our understanding of how gas, stars, and dark matter evolve together. Are there populations of gas-rich but starless galaxies ('minihalos') in the Local Group? How does galaxy environment determine time scales over which HI reservoirs can be maintained, ensuring future star formation fertility? What is the gas content of early-type galaxies? How does HI gas, particularly in large reservoirs and tidal tails, function to spur star formation in the group environment?

8.1 ALFALFA as a Cosmological Probe

For the first time, data is available to answer such questions and more. ALFALFA provides a sample of blindly HI-selected galaxies large enough to begin probing cosmology through the 21 cm line in earnest. The greatest strength of ALFALFA may be in its ability to provide constraints for simulations that try to take cold gas masses into account, which have been, up to this point, severely limited in their ability to reproduce the gas characteristics of the observed local universe.

The main goal of this dissertation was to understand the statistical characteristics of the ALFALFA sample, particularly those that relate to cosmology and the future of 21 cm observations, and to characterize the sensitivity of the survey in order to reduce systematic errors. ALFALFA is an ongoing, blind HI survey that is providing a census of local gas-rich galaxies out to z $\sim$ 0.06. To date, the survey has provided $\sim$11,900 Code 1 sources, $\sim$3,100 Code 2 sources, and 812 Code 9 sources which represent high velocity cloud complexes and components thereof. A catalog of isolated, ultra-compact high velocity cloud components identified as minihalo candidates is underway.

ALFALFA probes 5 orders of magnitude in HI mass. The current catalog contains 431 low-mass dwarfs with HI masses $< 10^{8.0}\ M_{\odot}$, including both Code 1 and Code 2 sources, or 366 such dwarfs when only Code 1 objects are included. On the extreme high-mass end, $> 10^{10.5}\ M_{\odot}$, the catalog contains 83 (81 Code 1). The superior sensitivity of Arecibo and of the ALFALFA observing scheme is allowing us to probe in great detail the statistical distribution of gas-rich galaxies. The data available for this dissertation also covers a variety of environments;

when complete, the ALFALFA catalog will contain ∼30,000 galaxies in environments ranging from voids to superclusters.

All of this combines to make ALFALFA's sample of gas-rich galaxies by far the best for statistically understanding the characteristics of this population of galaxies and for probing cosmology with them. Throughout this work, we take advantage of the available catalog for the survey at approximately 40% completion, referred to as $\alpha.40$.

In Chapter 2, we presented a contribution to the ALFALFA catalog of galaxies, as δ = 25°, covering 136 contiguous square degrees and including 90 high velocity cloud components, 430 Code 1 extragalactic detections, and 21 Code 2 extragalactic detections. 65% of these detections represent new HI detections, and 58% are altogether new redshift measurements. This catalog also includes a peculiar detection of a strong 21 cm source (AGC122913) with profile characteristics indicative of a galaxy but with no optical counterpart, as well as a detection of a compact source like a high velocity cloud (AGC113843) with a velocity (84 km s^{-1}) inconsistent with other HVCs in the region. This latter object has been targeted for optical followup by Elizabeth Adams to search for a stellar component. The full catalog, along with notes for a subset of relevant objects, is presented in Appendix A.

8.2 The HI Mass Function from a Large, Blind HI Survey

Chapter 3 presented a robust determination of the HI mass function faint-end slope α from the 40% ALFALFA survey sample $\alpha.40$, resolving the debate: α = -1.33 $\pm$ 0.02. Our measurement of Ω_{HI}, 4.3 $\pm$ 0.3 $\times$ 10^{-4} h_{70}^{-1}, is 16% higher than

previous determinations based on blind HI catalogs, and we find that the high-mass end of the HI mass function is enhanced relative to previous estimates. This indicates that future 21 cm line surveys, which will be sensitive primarily to galaxies above M_*, will be sensitive to greater numbers of halos than previously thought.

We also showed that the 2DSWML and $1/V_{max}$ methods for estimating the HIMF are in good agreement, which indicates that our understanding of the ALFALFA survey's sensitivity and its selection function is accurate. All estimates are robust against large-scale structures in the survey volume. In Appendix B, we include a full discussion of the 2DSWML method and of the corrections made to the $1/V_{max}$ estimate of the HIMF.

Baryon depletion and photoheating can, in principle, reconcile the difference between the faint-end slopes of dark matter mass functions and their baryonic (stellar and HI) counterparts. Along the same lines, our measurement of the HI mass function showed that the HIMF continues climbing down to $M_{HI} \sim 10^6$ $M_\odot$, and did not observe a turnover. This indicates that the limiting mass scale below which photoheating or other baryonic processes inhibit galaxy formation must be still lower.

Chapter 4 extends this work to consider the impact of survey sensitivity, mass uncertainties, and large-scale structure. A full understanding of these systematic effects will be used to optimize results from the eventual '$\alpha.100$' sample that will contain three times as many galaxies. We find that the inclusion or exclusion of Code 2 sources has only a minor effect on the shape of the HIMF; the same is true for the inclusion of the full redshift extent out to cz $\sim$18,000 km s^{-1}, except that the 2DSWML method is ill-equipped to handle large re-

gions of missing volume to which the $1/V_{max}$ method is easily adaptable. We also showed that poorly informed models of distances to nearby galaxies can strongly bias the faint-end slope of the HI mass function. Finally, we showed that $\alpha.40$ begins to approach homogeneity but is not a perfectly representative sample of the local universe, and suggest that the full ALFALFA survey will lead to improvements.

The dependence of the HI mass function on local environment was discussed in Chapter 5. This analysis used both a PSCz density reconstruction map and a volume-limited spectroscopic sample from SDSS to estimate local densities on a range of scales, from $\sim$2 to $\sim$10 Mpc. Regardless of the scales over which environment is quantified, we find that the characteristic turnover mass for the Schechter function, M_*, increases with increasing density, as expected from the scaling of dark matter halo masses with density. However, the influence of environment on the shape of the HIMF and, in particular, the faint-end slope, tells a more complex story. The results presented in Chapter 5 suggest that, up to some point, higher-density regions are richer in low-mass galaxies, indicated by a steeper faint-end slope. However, at some threshold density, some process (such as ram-pressure stripping) or some combination of processes become dominant over this gas enhancement, effectively flattening the faint-end slope.

Because of limitations in subsampling the dataset and in quantifying environment, these results are highly inconclusive but very suggestive, and the full ALFALFA survey with its larger sample size will be able to more fully explore this possibility.

8.3 The Clustering Characteristics of HI-Selected Galaxies

The correlation function for the $\alpha.40$ sample is presented in Chapter 6. Gas-rich galaxies are known to be among the least clustered populations of objects in the present universe. We find that the superior sensitivity of ALFALFA, and high value of the selection function, allows us to include the full survey redshift range (cz = 0 to 15,000 km s^{-1}) without the introduction of significant noise in this analysis. Assuming a power-law form for the real-space correlation function, $\xi(r) = (r/r_0)^{-\gamma}$, we find $\gamma = 1.51 \pm 0.09$ and $r_0 = 3.3^{+0.3}_{-0.2}$ h^{-1} Mpc.

We also find that HI-selected galaxies are strongly anti-biased relative to the underlying dark matter distribution on small scales (< 1.0 h^{-1} Mpc) and are only slightly anti-biased on larger scales ($\sim$10 h^{-1} Mpc).

Finally, this chapter showed that the $\alpha.40$ sample provides, for the first time, a robust measurement of the clustering of HI-selected galaxies, which can be used to provide observational constraints for theoretical models. While gas-rich galaxies are, currently, poorly modeled in N-body and semi-analytic simulations of the universe at z = 0, this situation is likely to change given the results presented in this dissertation and, especially, the full results when the ALFALFA catalog is complete.

We explored the dependence of clustering on various galaxy properties in Chapter 7, including mass, optical luminosity, and color. We find that the clustering of gas-rich galaxies depends strongly on HI mass and color, and less strongly on luminosity. We are able to split the ALFALFA sample into 3 separate mass subsamples, and show that the clustering increases with mass. This effect is most pronounced on the largest scales probed. There is little difference

on small scales. Taken in combination with the results of Chapter 5, we hypothesize that small-scale processes in dense environments have a weaker influence on the HI masses of constituent galaxies than previously assumed, and that processes occurring on larger scales are more significant. The observed relationships for gas-rich galaxies are the same as those for optically selected samples; that is, more massive, redder, brighter galaxies are more strongly clustered.

Large 21 cm surveys with instruments like the SKA are likely to be less antibiased with respect to dark matter than blind HI surveys have been to date, since they will be most sensitive to the extreme high-mass end of the HIMF.

8.4 Future Work

The greatest extension of the work presented in this dissertation will involve using the full, complete ALFALFA catalog, projected to contain $\sim$30,000 galaxies. Both the increased sample size and the additional volume will improve and refine these results, reducing both statistical and systematic errors (e.g., those due to large-scale structure).

8.4.1 Environmental Processes on Various Scales

More detailed study is necessary to understand the dependence of the HI mass function's dependence on environment, given the results discussed here. What processes are important in changing the shape of the HI mass function, and why do they appear to be functioning on scales larger than clusters? One weakness of the approach presented in Chapter 5 is its dependence on a volume-limited

sample of galaxies (from SDSS or from the PSCz), with smoothing applied to estimate relative density. To better understand the exact relationship between the HIMF and environment, it may be necessary to select voids, groups, and clusters from the ALFALFA volume. This can be achieved first using existing catalogs and, secondly, using iterative structure-finding algorithms. In addition to using this approach to measure a 'bulk' HIMF separately for voids, groups, and clusters, this would allow us to explore the variety of HIMF shapes found within 'similar' environments.

Using known structures to study the dependence of galaxy properties on environment would avoid some of the problems of the δ parameter. In particular, quantifying environment using this parameter can lead to mixing of truly different environments, since, for a given value of N, the distance to the N^{th} nearest neighbor will vary widely for the galaxies in the sample.

Furthermore, by using known structures to explore the HIMF, we could unveil the actual processes that are responsible for the changes in the shape of the HIMF. For instance, we hypothesize that late cold accretion of HI mass may be increasing the faint end of the HIMF in group environments and that the flattening of the faint-end slope in high-density regions measured on large scales must be due to some mechanism other than ram-pressure stripping. These ideas cannot be confirmed with large-scale, global estimates of the HIMF. The search for minihalo satellite candidates in the Local Group suggests a future possibility of a Local Group HI mass function, which would allow an extraordinarily detailed examination of the processes at work. The use of known structures has the added benefit of reducing the systematic uncertainties associated with bulk estimates of the density parameter δ, though the statistical uncertainties (due to

small subsample galaxy counts) are potentially problematic.

Work is also underway (A. Saintonge) to better quantify environment using a modified nearest-neighbors analysis that includes more galaxies to better capture the influence of satellite galaxies in a proper definition of environment. Regardless of method, any way that will more reliably quantify environment will benefit the study of the HI mass function.

8.4.2 Cross-Correlation to Understanding Clustering and Galaxy Properties

The models of Kim et al. [2010], which reproduce the clustering characteristics of the HIPASS sample, can now be exploited to attempt to understand the clustering of ALFALFA galaxies. Conversely, we may find that these models are not able to reliably reproduce the more complex characteristics of $\alpha.40$, particularly the dependence on galaxy characteristics. ALFALFA can therefore contribute to the improvement of these models, working to close the gap between the extremely detailed optical characteristics of simulated galaxies and the poor understanding of where gas fits into the picture. To date, such models can only be loosely compared to HI-selected samples, given the lack of a large survey like ALFALFA and robust measurements of cosmological statistics for these samples. Instead, these models typically focus on fitting the luminosities and stellar characteristics of observed galaxies, which are related to gas reservoirs, since gas fuels star formation, but only loosely.

When a larger sample is available from ALFALFA, we will better under-

stand the relative importance of various galaxy properties in probing the environments of HI-selected populations. The work in this dissertation had a limited ability to isolate the influences of HI mass, galaxy luminosity, and color. Studies of optically-selected samples indicate that the relationships are complex: for red galaxies, clustering strengthens significantly with luminosity, while the trend is weaker for blue galaxies; at any luminosity, red galaxies cluster more strongly than blue galaxies [Zehavi et al., 2005]. Our results show that HI mass is a more significant predictor of clustering, even though we have a significant population of fairly red galaxies in the sample. To add to the confusion, luminosity is correlated with color, as is HI mass. Since red galaxies should be expected to contain less cold gas, on average, than their blue counterparts, it is surprising that both high-mass and red galaxies are our strongest clustering categories. With more galaxies at our disposal, we can cut the sample across HI mass, optical luminosity, and optical color, in order to isolate the strongest effect.

Another approach could compare the clustering dependences of an HI- and an optically-selected sample within the same volume, to determine whether the galaxies with 'missing' low-HI mass reservoirs are explained by an overabundance of gas-poor (or gasless) dwarf galaxies.

In order to avoid statistical issues with small subsamples, another approach can be taken, which involves cross-correlating the ALFALFA catalog with an optical catalog from SDSS. Using subsets of the SDSS catalog, split by various optical properties, would keep sample sizes large while disentangling competing effects. Using this approach would not only determine how the clustering of HI-selected objects is dependent on their optical properties, but also how such objects are biased with respect to various optical samples. Since the biasing rel-

ative to the underlying dark matter distribution is model-dependent, and since we will want to compare the results of future cosmological surveys, some of which will be optically biased, this comparison is important for the cosmological utility of HI-selected galaxies.

Such a cross-correlation investigation is outside the scope of this dissertation and requires a detailed understanding of the systematic effects, survey geometry, and survey sensitivity of the Sloan Digital Sky Survey. A significant complication is the well-known issue of fiber collisions in the SDSS. When objects are selected for spectroscopic followup in the survey (from a complete photometric sample down to r = 17.7), the observing strategy requires that all spectroscopic targets be separated by at least 55 arcseconds. This makes complete spectroscopic catalogs from SDSS slightly biased against clusters, which becomes a significant problem in cross- and auto-correlation analyses. Fiber collisions can be corrected by comparing the complete photometric sample to the spectroscopic sample; any objects not appearing in both would have been left out of the spectroscopic sample due either to fiber collisions or other problems, and this can be corrected by assigning the redshift given to the nearest on-sky neighbor [Krumpe et al., 2010]. At higher redshifts, photometric redshifts can be used, but this does not apply to the ALFALFA volume.

Alternatively, the angular correlation function using the complete photometric catalog can be computed and, via a comparison to the calculated three-dimensional correlation function using the spectroscopic catalog, a correction factor can be obtained that accounts for fiber collisions; each individual SDSS galaxy-galaxy and SDSS-ALFALFA galaxy-galaxy pair would then be corrected by a distinct factor [Li et al., 2006b]. Therefore, the cross-correlation analysis in-

volves calculating both the angular (for fiber-collision correction) and the real-space correlation function for SDSS and for ALFALFA, as inputs to the cross-correlation statistic, even before pair-counting the cross-correlations. This represents a formidable computational challenge, given the large size of the relevant optical samples and the necessity of using random catalogs with a density enhancement of $\sim$20. The results, however, would offer the best chance of determining the relationship between the clustering of HI galaxies and optical properties with statistical significance, and would determine the bias between HI-selected galaxies and galaxies selected by a host of different optical properties.

An additional area of study involves investigating the optical properties (including morphology and star formation history) and the environments of unexpected categories of galaxies in our sample, such as red galaxies with large HI masses. If the findings of West et al. [2009] are any indication, we may uncover a significant population of gas-enriched elliptical galaxies with current ongoing star formation.

8.4.3 Predictions for Future Surveys

Finally, in the near-term, the $\alpha.40$ catalog and our understanding of its HIMF and clustering characteristics can be applied to make more stringent predictions for the next generation of 21 cm experiments. The HI mass function, which we found to be enhanced in extremely high-mass objects relative to previous measurements, indicates the galaxy counts that can be expected for a large galaxy survey in the 21 cm line, such as those proposed for the Square Kilometer Array,

will be higher than in earlier estimates; this, however, is subject to great uncertainties given how little we know about the evolution of the HI mass function. If simulations can be adjusted now that robust benchmarks exist for the z = 0 characteristics of HI-selected galaxies, this may constrain the possible evolutionary tracks that the distribution of gas reservoirs may have followed.

The clustering of HI-selected galaxies, in particular high-mass galaxies, can be applied to make predictions of the strength of the signal that will be obtained with future intensity mapping projects, which will not resolve individual galaxies but will measure the bulk HI on $\sim$10 Mpc scales. The $\alpha.40$ measurement of the HI-selected galaxy bias indicates that, at low redshift, the selection of high-HI mass galaxies ensures a sample that adequately probes the underlying dark matter distribution.

APPENDIX A

CATALOG OF ALFALFA DETECTIONS IN THE ANTI-VIRGO REGION AT δ=+25°

Catalog of 541 detections from the ALFALFA survey covering $22^0 0^m < \alpha(J2000) < 03^h 00^m$ and $24^\circ < \delta(J2000) < 26^\circ$. These data are published in Martin et al. [2009].

Table A.1: HI Candidate Detections

Source	AGC #	HI Coords (J2000) hh mm ss.s+dd mm ss	Opt Coords (J2000) hh mm ss.s+dd mm ss	$cz_{\odot}$ km s^{-1}	$W50$ (ϵ_w) km s^{-1}	F_c Jy km s^{-1}	S/N	rms mJy	Dist Mpc	log M_{HI} $M_{\odot}$	Code
4- 1	321318	22 00 55.4 +24 45 10		-345	29(14)	2.00	11.4	7.01			9 *
4- 2	321319	22 01 27.2 +24 59 47		-371	20(9)	1.08	9.8	5.07			9 *
4- 3	321303	22 02 06.7 +24 15 32	22 02 05.3 +24 15 32	6059	127(18)	1.66	15.4	2.13	88.4	9.49	1
4- 4	321304	22 03 51.1 +25 26 59	22 03 51.1 +25 26 32	2692	142(18)	1.71	15.7	2.04	41.3	8.84	1
4- 5	320941	22 07 50.7 +24 50 01	22 07 51.0 +24 49 57	3412	124(9)	3.71	38.6	1.92	51.4	9.36	1
4- 6	320047	22 08 50.1 +24 42 09	22 08 49.2 +24 41 46	6300	284(40)	1.14	8.0	1.90	91.6	9.35	1
4- 7	320063	22 10 15.2 +25 28 11	22 10 16.7 +25 27 57	7144	124(5)	1.25	11.1	2.26	97.2	9.44	1
4- 8	321305	22 10 21.6 +24 44 11	22 10 22.4 +24 43 39	12185	137(26)	0.85	8.0	2.03	169.2	9.76	1
4- 9	321306	22 11 37.8 +25 43 35	22 11 37.7 +25 43 46	11875	135(16)	0.83	7.2	2.21	164.8	9.73	1
4- 10	321320	22 12 27.6 +24 43 30		-303	25(8)	1.83	14.2	5.46			9 *
4- 11	321204	22 13 05.5 +25 54 37	22 13 06.2 +25 54 56	12652	133(24)	1.10	9.4	2.27	175.9	9.90	1
4- 12	321307	22 14 04.4 +25 41 08	22 14 04.7 +25 40 52	1152	60(5)	1.04	15.6	1.89	18.7	7.94	1
4- 13	321309	22 16 09.2 +25 18 10	22 16 08.7 +25 18 16	5189	157(18)	0.81	8.1	1.78	76.1	9.04	1
4- 14	321368	22 16 56.9 +25 33 49		-459	18(6)	0.43	6.8	3.09			9 *
4- 15	320128	22 17 13.6 +25 12 32	22 17 13.1 +25 12 47	12654	412(15)	2.46	14.0	1.90	175.9	10.25	1
4- 16	321321	22 17 20.3 +25 42 31	22 17 20.2 +25 42 27	12078	39(7)	0.94	15.0	2.19	167.7	9.79	1
4- 17	320139	22 18 05.3 +25 05 53	22 18 07.2 +25 05 32	7049	78(12)	0.70	8.8	1.97	95.8	9.18	1
4- 18	321369	22 18 14.6 +24 42 29		-424	31(8)	1.32	16.8	3.06			9 *
4- 19	320149	22 19 06.5 +24 35 37	22 19 06.2 +24 35 53	8247	178(23)	0.97	8.7	1.86	112.9	9.46	1
4- 20	321350	22 19 38.4 +25 47 36		-427	18(2)	1.60	16.1	4.85			9 *
4- 21	321322	22 19 46.9 +25 41 52	22 19 47.2 +25 41 32	8110	145(24)	0.98	9.8	1.84	111.0	9.45	1
4- 22	321323	22 20 06.3 +25 36 31	22 20 06.6 +25 36 20	12482	357(14)	2.82	14.8	2.26	173.4	10.30	1
4- 23	321389	22 20 23.2 +25 10 33		-428	31(5)	0.77	8.3	3.58			9 *
4- 24	321324	22 20 22.9 +24 23 14	22 20 20.6 +24 23 18	3971	147(22)	2.02	18.2	2.04	58.9	9.22	1
4- 25	321325	22 20 26.8 +25 47 06	22 20 23.9 +25 47 16	11962	396(10)	1.12	6.5	1.95	166.0	9.86	2
4- 26	320162	22 20 44.9 +24 46 48	22 20 41.3 +24 45 52	12349	239(13)	1.01	7.7	1.89	171.5	9.85	1
4- 27	321326	22 20 54.3 +24 45 31	22 20 54.0 +24 45 50	5256	166(35)	1.35	11.0	2.14	76.9	9.27	1
4- 28	321351	22 21 30.5 +24 36 58		-341	26(5)	0.64	8.4	3.22			9 *
4- 29	321352	22 21 45.3 +25 49 29		-440	21(2)	5.79	48.7	5.47			9 *
4- 30	321390	22 22 13.9 +24 33 53		-420	24(4)	0.99	9.6	4.43			9 *
4- 31	321328	22 23 13.8 +24 38 01	22 23 12.5 +24 38 21	12314	307(38)	2.26	14.4	2.00	171.0	10.19	1
4- 32	321391	22 24 17.1 +25 33 50		-403	29(5)	0.60	8.2	2.93			9 *
4- 33	321327	22 24 24.0 +24 15 06	22 24 25.2 +24 14 27	12072	229(19)	1.36	10.2	1.96	167.5	9.95	1
4- 34	321225	22 24 53.4 +25 57 31		-407	22(2)	0.92	12.5	3.33			9 *
4- 35	321329	22 26 26.4 +25 22 32	22 26 25.0 +25 22 27	11954	81(9)	0.81	10.1	1.97	165.9	9.72	1
4- 36	321330	22 26 51.9 +25 00 03	22 26 50.3 +25 00 57	7308	95(8)	0.72	6.7	2.44	99.5	9.22	1
4- 37	321331	22 27 30.5 +25 37 50	22 27 31.5 +25 37 40	12191	226(26)	1.09	8.1	1.98	169.2	9.87	1
4- 38	321332	22 28 14.1 +25 46 54	22 28 13.2 +25 47 11	7254	30(6)	0.56	10.7	2.04	98.7	9.11	1
4- 39	321394	22 29 17.7 +25 00 53		-289	36(9)	0.37	6.5	2.09			9 *
4- 40	321333	22 30 00.5 +25 09 02	22 29 59.2 +25 09 00	11994	224(16)	1.11	8.4	1.97	166.4	9.86	1
4- 41	321334	22 30 12.5 +24 50 57	22 30 12.1 +24 51 03	11893	136(12)	0.80	7.2	2.13	165.0	9.71	1
4- 42	321353	22 30 40.5 +24 38 36		-402	27(3)	3.10	24.7	5.19			9 *
4- 43	321335	22 32 40.0 +25 34 24	22 32 39.5 +25 34 35	12434	225(26)	1.61	11.6	2.06	172.7	10.05	1

TableA.1 – Continued

Source	AGC #	HI Coords (J2000) hh mm ss.s+dd mm ss	Opt Coords (J2000) hh mm ss.s+dd mm ss	$cz_\odot$ km s^{-1}	$W50$ (ϵ_w) km s^{-1}	F_c Jy km s^{-1}	S/N	rms mJy	Dist Mpc	log M_{HI} $M_\odot$	Code
4- 44	321355	22 33 16.2 +24 06 22		-401	31(4)	1.67	11.8	5.52			9 *
4- 45	320690	22 33 43.5 +24 27 03	22 33 50.7 +24 26 37	9968	110(21)	0.89	9.0	2.09	137.4	9.60	1 *
4- 46	12084	22 34 26.7 +25 01 56	22 34 26.6 +25 01 49	12217	357(35)	1.41	8.3	2.02	169.6	9.98	1
4- 47	321356	22 34 52.3 +25 22 57		-425	32(4)	1.35	12.0	4.30			9 *
4- 48	321344	22 35 34.0 +25 11 35	22 35 31.2 +25 10 40	12108	300(6)	1.17	7.6	1.98	168.0	9.89	1
4- 49	321336	22 36 11.8 +24 34 20	22 36 13.4 +24 34 59	12740	158(18)	1.10	10.4	1.87	177.0	9.91	1
4- 50	12106	22 36 33.4 +25 45 51	22 36 32.9 +25 45 47	7380	248(5)	4.14	30.6	1.92	100.5	9.99	1
4- 51	12123	22 37 52.8 +25 11 46	22 37 53.1 +25 11 28	4083	196(19)	3.29	26.5	1.98	60.0	9.45	1
4- 52	320284	22 38 03.6 +24 57 36	22 38 04.5 +24 57 38	12734	342(72)	1.24	7.3	2.06	176.9	9.96	1 *
4- 53	321284	22 38 27.9 +25 55 36	22 38 27.3 +25 55 03	8593	57(22)	0.57	8.4	1.99	117.8	9.27	1 *
4- 54	321337	22 39 03.1 +24 48 35	22 38 59.7 +24 47 58	14967	319(28)	1.39	8.7	1.99	208.8	10.15	1
4- 55	320702	22 40 05.2 +24 41 56	22 40 05.8 +24 41 56	12815	405(168)	4.49	22.2	2.23	178.1	10.53	1 *
4- 56	320306	22 40 41.1 +25 37 01	22 40 39.6 +25 37 26	13649	141(7)	0.74	6.3	2.21	190.0	9.80	2
4- 57	321346	22 41 08.6 +24 35 20	22 41 09.4 +24 34 28	9045	188(7)	1.89	15.3	2.00	124.2	9.84	1
4- 58	321358	22 41 23.6 +25 07 32		-411	29(8)	0.37	7.5	2.00			9 *
4- 59	321392	22 42 43.5 +24 59 13		-373	22(6)	0.44	7.2	2.70			9 *
4- 60	321359	22 42 58.2 +24 05 39		-402	24(5)	0.53	7.6	3.07			9 *
4- 61	321347	22 43 29.4 +24 33 14	22 43 26.8 +24 33 50	7607	161(23)	0.89	8.2	1.89	103.7	9.35	1
4- 62	321360	22 44 16.3 +25 49 25		-442	37(4)	0.90	10.8	2.96			9 *
4- 63	321338	22 44 50.9 +25 25 56	22 44 50.9 +25 26 00	9558	163(19)	1.45	13.5	1.89	131.6	9.77	1
4- 64	321339	22 45 36.2 +25 07 09	22 45 37.5 +25 07 29	7625	156(20)	1.08	9.0	2.15	103.9	9.44	1
4- 65	320379	22 48 33.4 +24 32 10	22 48 33.8 +24 32 05	12319	317(25)	2.48	16.2	1.92	171.0	10.23	1
4- 66	12197	22 48 53.0 +24 50 20	22 48 52.1 +24 50 15	7469	347(17)	3.53	21.3	1.98	101.7	9.93	1
4- 67	321361	22 49 03.9 +24 44 50		-437	25(10)	0.35	6.1	2.44			9
4- 68	321340	22 51 15.3 +24 54 59	22 51 15.8 +24 55 12	13377	289(30)	1.88	13.4	1.85	186.1	10.19	1
4- 69	321341	22 52 14.7 +24 06 03	22 52 15.7 +24 06 09	12133	100(14)	0.65	6.5	2.20	168.3	9.64	1
4- 70	321362	22 52 15.4 +24 38 59		-409	24(2)	1.81	22.6	3.48			9 *
4- 71	321363	22 53 09.7 +24 12 37		-409	25(3)	2.14	22.2	4.10			9 *
4- 72	12233	22 53 44.5 +25 50 26	22 53 44.8 +25 50 40	7571	105(2)	5.62	59.3	2.05	103.2	10.15	1
4- 73	321364	22 54 18.5 +25 12 17		-457	26(2)	2.52	19.0	5.54			9 *
4- 74	321343	22 54 20.7 +25 31 21	22 54 20.4 +25 31 25	9701	49(12)	1.13	17.8	1.98	133.6	9.68	1
4- 75	321365	22 55 46.4 +25 45 06		-398	31(4)	2.82	25.4	4.33			9 *
4- 76	321366	22 55 47.0 +25 21 52		-384	22(5)	2.60	28.1	4.13			9 *
4- 77	321393	22 56 36.0 +24 24 18		-443	24(8)	1.41	14.7	4.18			9 *
4- 78	321387	22 56 40.5 +25 06 50		-372	29(2)	2.12	23.3	3.66			9 *
4- 79	321194	22 57 04.8 +25 43 31	22 57 04.0 +25 43 30	1050	95(2)	4.22	46.1	2.08	16.5	8.43	1
4- 80	321388	22 58 12.3 +24 10 58		-362	21(3)	1.82	25.8	3.21			9 *
4- 81	320876	22 58 33.4 +25 24 31	22 58 35.6 +25 25 30	7428	133(19)	0.78	7.7	1.95	101.1	9.27	1
4- 82	12280	22 58 46.4 +24 38 31	22 58 45.1 +24 38 23	11868	366(20)	1.89	11.0	2.00	164.5	10.08	1
4- 83	321384	22 58 59.5 +24 00 20	22 58 59.6 +24 00 13	7228	192(34)	1.42	11.9	1.92	98.2	9.51	1
4- 84	321385	22 59 15.4 +24 42 53	22 59 15.1 +24 42 34	7246	75(8)	0.87	10.9	2.03	98.5	9.30	1
4- 85	12283	22 59 20.0 +24 06 39	22 59 19.8 +24 06 24	9955	278(5)	1.36	9.1	2.00	137.2	9.78	1
4- 86	320603	22 59 21.0 +24 53 26	22 59 19.9 +24 53 04	7339	235(32)	1.42	9.7	2.15	99.8	9.52	1 *
4- 87	321386	22 59 24.6 +24 48 31	22 59 23.2 +24 48 38	7644	55(7)	0.65	10.0	1.92	104.2	9.22	1

TableA.1 – Continued

Source	AGC #	HI Coords (J2000) hh mm ss.s+dd mm ss	Opt Coords (J2000) hh mm ss.s+dd mm ss	$cz_\odot$ km s^{-1}	$W50$ (ϵ_w) km s^{-1}	F_c Jy km s^{-1}	S/N	rms mJy	Dist Mpc	log M_{HI} $M_\odot$	Code
4- 88	320489	22 59 31.2 +25 07 32	22 59 28.0 +25 07 37	7425	71(8)	0.73	8.6	2.23	101.1	9.24	1
4- 89	12290	22 59 38.2 +24 50 56	22 59 37.4 +24 50 50	7268	455(25)	3.37	16.3	2.03	98.8	9.89	1
4- 90	12289	22 59 42.1 +24 04 26	22 59 41.5 +24 04 29	10165	217(12)	3.82	29.5	1.96	140.2	10.25	1
4- 91	333316	23 00 02.4 +25 33 56	23 00 01.1 +25 32 56	7328	93(8)	0.68	7.8	2.03	99.7	9.20	1 *
4- 92	333259	23 00 05.5 +24 29 18	23 00 05.4 +24 29 04	7330	131(8)	0.91	7.6	2.32	99.7	9.33	1
4- 93	333260	23 00 34.8 +24 47 03	23 00 35.9 +24 46 46	9894	181(18)	0.95	7.9	1.99	136.3	9.62	1 *
4- 94	320891	23 00 42.3 +25 18 38	23 00 43.8 +25 18 33	7613	152(24)	0.97	9.2	1.92	103.8	9.39	1
4- 95	320507	23 00 46.0 +25 42 08	23 00 46.2 +25 41 57	7392	71(22)	1.18	14.3	2.16	100.6	9.45	1
4- 96	332335	23 00 57.5 +25 12 57	23 00 56.2 +25 12 56	11420	291(47)	1.97	12.6	2.06	158.1	10.06	1
4- 97	320516	23 02 05.0 +25 45 48	23 02 05.6 +25 45 40	7513	220(23)	1.50	12.1	1.87	102.3	9.57	1
4- 98	12318	23 02 08.7 +25 40 26	23 02 07.9 +25 40 12	9767	407(32)	1.78	9.8	2.00	134.5	9.88	1
4- 99	332263	23 02 15.5 +24 45 11	23 02 13.7 +24 45 16	7761	221(4)	2.72	18.3	2.23	105.9	9.86	1 *
4-100	332906	23 05 08.2 +25 52 30	23 05 09.2 +25 52 28	9808	122(11)	0.83	8.2	2.02	135.1	9.55	1
4-101	333317	23 05 30.0 +24 50 11		-395	37(10)	0.50	6.2	2.91			9 *
4-102	332907	23 05 41.3 +25 57 29	23 05 42.1 +25 57 05	9870	263(3)	1.52	9.7	2.16	136.0	9.82	1
4-103	333225	23 05 49.3 +25 58 44	23 05 48.2 +25 59 23	-379	21(7)	0.59	9.0	3.00			9 *
4-104	12355	23 06 00.6 +24 33 23	23 05 59.7 +24 33 53	7718	288(15)	2.49	16.3	2.01	105.2	9.81	1
4-105	333265	23 07 28.8 +24 08 27	23 07 29.5 +24 08 40	10637	80(11)	1.00	12.2	2.05	146.9	9.71	1
4-106	12384	23 07 48.5 +24 57 56	23 07 46.4 +24 57 37	10017	257(60)	2.29	16.0	2.00	138.1	10.01	1 *
4-107	12386	23 07 58.5 +24 58 36	23 08 00.9 +24 58 22	10093	302(22)	2.77	16.0	2.23	139.2	10.10	1 *
4-108	333318	23 10 38.5 +24 08 40	23 10 38.9 +24 08 40	12290	216(15)	1.20	8.7	2.10	170.5	9.91	1
4-109	333266	23 12 17.8 +24 09 27	23 12 15.3 +24 09 23	6167	152(25)	0.98	8.6	2.06	88.5	9.26	1
4-110	333267	23 12 50.2 +24 21 21	23 12 50.6 +24 21 18	8117	188(5)	1.16	9.9	1.90	110.9	9.53	1
4-111	333319	23 13 42.5 +24 35 47		-125	21(2)	3.77	35.6	4.85			9 *
4-112	330124	23 13 48.9 +25 28 01	23 13 49.4 +25 28 08	8035	116(18)	1.36	13.1	2.14	109.8	9.59	1
4-113	12432	23 13 56.6 +24 53 21	23 13 58.0 +24 53 42	8339	331(12)	1.58	9.4	2.06	114.1	9.69	1
4-114	333268	23 14 05.0 +24 37 20	23 14 04.2 +24 36 49	16664	121(17)	0.74	6.7	2.24	233.0	9.98	1
4-115	333269	23 15 10.6 +25 45 43	23 15 09.1 +25 45 50	7899	144(13)	1.10	9.6	2.13	107.8	9.48	1
4-116	333270	23 15 12.0 +25 47 53	23 15 11.8 +25 47 29	9674	302(7)	1.04	6.4	2.08	133.2	9.64	2
4-117	12459	23 15 41.5 +24 53 54	23 15 39.3 +24 53 54	7998	285(12)	1.93	11.9	2.15	109.2	9.73	1
4-118	12460	23 15 42.8 +25 05 36	23 15 43.8 +25 05 39	8088	301(7)	3.78	23.7	2.05	110.5	10.04	1 *
4-119	330167	23 16 01.9 +25 33 26	23 16 00.7 +25 33 23	8198	298(23)	2.67	15.5	2.24	112.1	9.90	1
4-120	333271	23 16 23.6 +25 37 45	23 16 22.4 +25 37 46	9674	236(17)	2.25	13.2	2.48	133.2	9.97	1
4-121	12469	23 16 24.0 +24 29 54	23 16 25.0 +24 29 48	8396	358(8)	4.62	23.1	2.36	114.9	10.16	1
4-122	333272	23 17 46.6 +24 34 50	23 17 47.6 +24 34 46	8207	268(13)	1.98	13.5	2.01	112.2	9.77	1
4-123	12490	23 18 39.1 +25 13 57	23 18 38.4 +25 13 57	8085	176(8)	2.77	22.6	2.06	110.5	9.90	1 *
4-124	330230	23 19 19.4 +24 56 08	23 19 19.9 +24 56 02	5811	100(5)	1.37	16.9	1.81	83.4	9.35	1
4-125	333321	23 19 21.7 +24 53 02		-150	30(3)	1.50	13.2	4.49			9 *
4-126	12520	23 20 06.0 +24 13 06	23 20 05.7 +24 13 15	9580	227(12)	7.44	56.7	1.95	131.8	10.48	1
4-127	330249	23 20 12.7 +25 35 53	23 20 13.5 +25 35 39	6057	263(9)	1.75	11.4	2.12	86.8	9.49	1
4-128	333339	23 20 23.6 +24 14 08		-131	49(2)	2.05	15.8	4.10			9 *
4-129	333273	23 20 35.1 +25 21 19	23 20 33.7 +25 20 58	6904	41(16)	0.46	8.0	1.96	93.6	8.98	1
4-130	333274	23 20 37.4 +24 06 32	23 20 37.2 +24 06 30	5996	111(20)	1.01	11.0	1.93	85.9	9.24	1
4-131	333275	23 20 54.2 +24 11 58	23 20 54.0 +24 11 46	9506	174(13)	1.84	15.9	1.95	130.8	9.87	1

TableA.1 – Continued

Source	AGC #	HI Coords (J2000) hh mm ss.s+dd mm ss	Opt Coords (J2000) hh mm ss.s+dd mm ss	$cz_\odot$ km s^{-1}	$W50$ (ϵ_w) km s^{-1}	F_c Jy km s^{-1}	S/N	rms mJy	Dist Mpc	log M_{HI} $M_\odot$	Code
4-132	12541	23 21 17.0 +25 22 53	23 21 18.2 +25 23 07	4450	82(27)	0.85	10.2	2.05	64.0	8.91	1
4-133	333277	23 21 20.5 +25 10 22	23 21 20.3 +25 10 18	5822	56(18)	0.79	12.2	1.90	83.5	9.11	1
4-134	333278	23 21 47.1 +24 17 40	23 21 47.0 +24 17 43	6074	192(15)	1.58	13.0	1.95	87.0	9.45	1
4-135	333279	23 21 50.7 +25 09 42	23 21 50.8 +25 09 30	16715	156(12)	1.16	10.8	1.92	233.8	10.17	1
4-136	333281	23 22 32.3 +24 36 34	23 22 32.3 +24 35 35	9708	37(3)	0.67	10.6	2.26	133.7	9.45	1 *
4-137	332922	23 22 33.9 +25 52 26	23 22 34.6 +25 52 58	8400	295(25)	1.71	11.1	1.99	115.0	9.73	1
4-138	330354	23 23 37.2 +25 12 02	23 23 36.1 +25 11 28	9653	231(25)	1.16	6.9	2.48	132.9	9.68	2
4-139	332071	23 23 43.9 +24 40 39	23 23 40.5 +24 40 10	12434	303(4)	1.57	10.1	1.98	172.6	10.04	1
4-140	333282	23 23 46.6 +24 00 55	23 23 45.0 +24 01 02	5096	126(11)	1.23	11.3	2.17	73.1	9.19	1
4-141	333322	23 23 48.9 +24 38 47		-410	28(7)	0.58	6.1	3.86			9 *
4-142	333283	23 24 10.0 +24 17 05	23 24 09.1 +24 17 01	5915	43(3)	0.86	14.7	1.96	84.7	9.16	1
4-143	333284	23 24 25.3 +24 42 09	23 24 25.9 +24 42 05	9585	111(14)	1.03	11.8	1.86	131.9	9.63	1
4-144	330400	23 24 27.8 +25 22 32	23 24 26.2 +25 23 06	9575	292(24)	2.02	12.0	2.20	131.8	9.92	1
4-145	332925	23 25 08.3 +25 50 37	23 25 09.2 +25 50 39	6608	143(12)	0.86	8.3	1.95	89.4	9.21	1
4-146	330424	23 25 16.9 +25 13 37	23 25 16.8 +25 13 29	5103	155(23)	3.11	26.5	2.11	73.2	9.59	1
4-147	333285	23 25 22.5 +25 08 10	23 25 23.6 +25 08 16	3418	97(15)	0.59	6.0	2.21	49.2	8.53	1
4-148	333286	23 25 51.7 +25 38 19	23 25 51.7 +25 38 19	8582	133(43)	1.63	14.6	2.16	117.6	9.73	1 *
4-149	331874	23 26 00.6 +25 47 01	23 26 00.4 +25 46 46	9613	338(3)	3.53	21.1	2.03	132.3	10.16	1
4-150	333340	23 26 27.5 +25 39 33		-147	16(5)	0.43	6.5	3.36			9 *
4-151	330446	23 26 34.2 +25 38 31	23 26 34.2 +25 38 33	9065	332(3)	2.45	13.3	2.26	124.5	9.95	1
4-152	12598	23 26 39.5 +25 05 03	23 26 39.8 +25 04 49	3474	333(12)	26.47	169.7	1.91	50.0	10.19	1
4-153	12599	23 26 50.4 +25 40 50	23 26 48.0 +25 41 09	6663	430(47)	1.75	8.8	2.07	90.2	9.53	1
4-154	333287	23 28 18.8 +24 20 37	23 28 18.1 +24 20 22	7461	148(23)	0.83	7.4	2.06	101.6	9.30	1
4-155	333288	23 28 40.7 +24 04 41	23 28 40.1 +24 04 55	5659	164(16)	1.00	9.0	1.93	81.0	9.19	1
4-156	331191	23 28 48.9 +24 52 10	23 28 49.1 +24 52 27	4968	119(17)	2.90	30.5	1.94	71.2	9.54	1
4-157	330494	23 29 56.3 +24 58 20	23 29 54.0 +24 58 19	9604	365(21)	1.83	10.1	2.12	132.2	9.88	1
4-158	330500	23 30 10.0 +25 32 01	23 30 09.4 +25 31 59	5745	200(23)	4.61	37.5	1.94	82.2	9.87	1
4-159	333289	23 30 29.8 +24 37 11	23 30 30.5 +24 36 54	9559	127(9)	0.83	7.0	2.32	131.5	9.53	1
4-160	333290	23 31 12.3 +24 52 43	23 31 10.1 +24 52 28	4008	135(19)	1.10	10.1	2.09	57.4	8.93	1
4-161	12643	23 31 28.6 +25 12 10	23 31 28.6 +25 12 12	7004	210(16)	2.68	23.0	1.79	95.1	9.76	1
4-162	12646	23 31 39.2 +25 56 38	23 31 39.1 +25 56 42	8031	166(20)	2.88	23.0	2.17	109.8	9.91	1
4-163	333291	23 31 40.6 +24 02 08	23 31 38.6 +24 02 08	3996	58(8)	0.62	8.4	2.15	57.3	8.68	1
4-164	331194	23 31 43.3 +25 52 03	23 31 42.8 +25 52 18	6768	173(24)	1.07	9.1	2.01	91.7	9.33	1
4-165	332174	23 32 05.1 +25 28 37	23 32 05.7 +25 28 42	3711	179(7)	2.53	17.7	2.38	53.2	9.23	1
4-166	333292	23 32 15.6 +25 08 27	23 32 14.4 +25 08 42	4103	107(10)	1.05	10.8	2.08	58.8	8.93	1
4-167	333293	23 32 47.3 +25 17 50	23 32 47.1 +25 18 15	4082	79(19)	0.87	14.5	1.51	58.5	8.85	1
4-168	330530	23 33 13.3 +25 39 13	23 33 12.0 +25 39 12	7095	147(27)	0.78	6.5	2.20	96.4	9.23	2
4-169	333294	23 33 16.6 +24 22 29	23 33 16.9 +24 22 44	5710	75(17)	0.94	11.3	2.13	81.6	9.17	1
4-170	330537	23 33 26.3 +25 39 11	23 33 26.0 +25 39 00	3718	136(15)	3.73	33.8	2.11	53.3	9.40	1
4-171	333295	23 33 36.9 +24 37 23	23 33 36.2 +24 37 47	7055	141(25)	0.98	9.3	1.98	95.8	9.33	1
4-172	330542	23 33 46.9 +25 46 55	23 33 45.2 +25 47 10	8109	198(10)	2.63	21.0	1.98	110.9	9.88	1
4-173	330544	23 33 47.6 +24 11 50	23 33 49.4 +24 11 27	5607	96(28)	1.20	13.9	1.96	80.1	9.26	1
4-174	333296	23 34 42.5 +24 42 31	23 34 42.4 +24 42 31	10853	107(12)	1.60	18.4	1.88	150.0	9.93	1
4-175	333297	23 35 00.1 +24 44 31	23 34 58.0 +24 44 13	9410	252(8)	1.48	9.6	2.16	129.4	9.77	1

TableA.1 – Continued

Source	AGC #	HI Coords (J2000) hh mm ss.s+dd mm ss	Opt Coords (J2000) hh mm ss.s+dd mm ss	$cz_{\odot}$ km s^{-1}	$W50$ (ϵ_w) km s^{-1}	F_c Jy km s^{-1}	S/N	rms mJy	Dist Mpc	log M_{HI} $M_{\odot}$	Code
4-176	330563	23 35 46.2 +24 35 10	23 35 47.2 +24 35 18	11452	260(20)	2.34	16.3	1.99	158.6	10.14	1
4-177	333298	23 35 56.7 +25 18 18	23 35 56.2 +25 18 34	5059	31(12)	0.43	9.9	1.66	72.3	8.72	1
4-178	333299	23 36 32.2 +24 27 07	23 36 33.8 +24 27 56	11510	102(29)	0.87	8.1	2.35	159.4	9.72	1
4-179	331893	23 36 49.3 +24 02 16	23 36 52.4 +24 03 00	17107	144(16)	0.96	7.3	2.42	239.4	10.11	1 *
4-180	333300	23 37 37.7 +24 34 22	23 37 37.4 +24 34 00	9977	175(25)	0.99	9.0	1.85	137.5	9.64	1
4-181	331413	23 37 44.5 +25 40 31	23 37 44.8 +25 40 26	6742	172(12)	1.91	16.0	2.03	91.3	9.57	1
4-182	331237	23 37 54.5 +25 45 06	23 37 54.1 +25 45 21	11212	154(13)	1.46	13.3	1.98	155.2	9.92	1
4-183	12712	23 38 05.7 +25 43 18	23 38 05.1 +25 43 10	9443	447(46)	1.75	8.3	2.12	127.1	9.82	1 *
4-184	333301	23 39 29.4 +25 41 57	23 39 24.7 +25 42 30	9450	206(25)	1.21	8.9	2.11	130.0	9.68	1
4-185	332950	23 40 04.4 +25 57 08	23 40 04.6 +25 57 10	11317	187(0)	0.83	6.6	2.17	156.7	9.68	1
4-186	330766	23 41 06.1 +24 56 24	23 41 06.4 +24 57 05	11496	281(26)	1.18	8.0	1.97	159.2	9.85	1
4-187	330765	23 41 06.5 +25 10 02	23 41 06.0 +25 10 18	11878	471(55)	2.66	13.4	1.88	164.7	10.23	1
4-188	330768	23 41 16.4 +25 33 31	23 41 15.9 +25 33 08	9393	479(60)	4.36	20.0	2.03	127.1	10.22	1 *
4-189	332110	23 41 47.8 +25 43 05	23 41 46.4 +25 43 09	11678	241(10)	0.85	5.8	2.12	161.9	9.72	2
4-190	333302	23 41 51.6 +24 13 14	23 41 51.3 +24 13 27	10590	339(54)	1.39	8.7	1.93	146.3	9.85	1
4-191	331281	23 42 01.9 +25 59 56	23 42 00.9 +25 59 53	9276	236(18)	1.34	9.0	2.17	127.1	9.71	1
4-192	12766	23 45 07.4 +25 30 53	23 45 06.5 +25 31 05	11754	339(29)	1.92	12.0	1.93	163.0	10.08	1
4-193	333303	23 45 29.5 +24 11 58	23 45 28.5 +24 12 19	11500	239(38)	1.26	9.5	1.92	159.3	9.88	1 *
4-194	333304	23 47 04.6 +24 44 26	23 47 04.5 +24 44 31	11064	276(38)	1.49	9.0	2.23	153.1	9.92	1
4-195	333342	23 47 12.8 +25 43 30		-128	33(4)	2.03	16.4	4.65			9 *
4-196	333305	23 48 03.7 +24 29 04	23 48 04.9 +24 28 57	14788	80(7)	0.68	7.9	2.13	206.3	9.83	1
4-197	333306	23 48 14.0 +24 23 15	23 48 14.5 +24 23 30	6968	313(80)	1.33	7.3	2.29	94.6	9.45	1
4-198	330927	23 48 19.0 +24 04 52	23 48 19.6 +24 05 27	10079	176(41)	0.93	7.7	2.02	139.0	9.63	1
4-199	333307	23 48 18.7 +24 35 06	23 48 19.4 +24 35 30	5349	159(14)	1.07	12.3	1.54	76.1	9.16	1
4-200	333210	23 48 41.2 +25 55 08	23 48 41.6 +25 54 40	1608	62(3)	0.60	7.7	2.20	23.3	7.88	1
4-201	333308	23 48 44.3 +24 18 24	23 48 46.7 +24 18 45	5344	73(11)	0.56	7.0	2.08	76.0	8.88	1
4-202	333309	23 49 13.7 +25 27 10	23 49 12.3 +25 27 42	7174	155(6)	0.98	8.4	2.09	97.5	9.34	1
4-203	333310	23 49 47.3 +24 07 50	23 49 46.7 +24 08 00	1649	52(12)	1.29	19.7	2.00	23.8	8.24	1
4-204	333324	23 50 30.3 +24 31 01		-420	21(10)	0.86	10.5	3.74			9 *
4-205	333215	23 51 44.3 +25 50 31	23 51 43.8 +25 50 25	7307	192(11)	1.08	8.8	1.98	99.5	9.40	1
4-206	331380	23 53 48.9 +25 35 34	23 53 45.2 +25 35 17	11510	64(19)	0.78	8.5	2.54	159.5	9.67	1
4-207	333311	23 54 53.8 +24 52 41	23 54 53.8 +24 52 52	10244	168(15)	1.13	9.5	2.05	141.4	9.73	1 *
4-208	331012	23 55 48.6 +25 30 43	23 55 50.1 +25 30 21	17327	220(55)	1.59	10.5	2.28	242.6	10.34	1 *
4-209	333326	23 56 14.3 +25 17 00		-345	27(2)	1.30	18.9	2.85			9 *
4-210	332903	23 57 48.6 +24 00 12	23 57 47.9 +24 00 27	10924	105(14)	3.29	34.4	2.07	151.1	10.25	1
4-211	333312	23 58 34.7 +24 53 56	23 58 34.3 +24 54 20	10809	277(42)	1.41	9.3	2.03	149.5	9.87	1
4-212	331047	23 59 12.4 +25 56 30	23 59 13.7 +25 56 29	7010	147(28)	1.14	9.3	2.25	95.3	9.39	1
4-213	333313	24 00 00.1 +24 54 35	23 59 59.3 +24 54 26	11181	313(20)	1.65	11.3	1.85	154.8	9.97	1
4-214	102622	00 06 52.0 +24 13 38	00 06 51.5 +24 13 30	10252	156(23)	3.03	28.2	1.92	141.6	10.16	1
4-215	102623	00 07 18.7 +25 05 18	00 07 17.4 +25 05 16	10896	182(20)	1.44	12.6	1.89	150.8	9.89	1
4-216	102680	00 07 58.7 +24 01 37		-410	18(7)	0.40	7.2	2.64			9 *
4-217	76	00 08 51.1 +24 32 34	00 08 49.3 +24 32 24	4583	195(19)	3.26	25.6	2.03	64.7	9.51	1
4-218	79	00 09 04.4 +25 37 07	00 09 04.4 +25 37 05	4336	189(2)	5.22	41.2	2.06	64.6	9.71	1
4-219	102625	00 09 36.1 +25 33 24	00 09 35.5 +25 33 11	4642	129(15)	1.07	11.0	1.90	65.5	9.03	1

TableA.1 – Continued

Source	AGC #	HI Coords (J2000) hh mm ss.s+dd mm ss	Opt Coords (J2000) hh mm ss.s+dd mm ss	$cz_\odot$ km s^{-1}	$W50$ (ϵ_w) km s^{-1}	F_c Jy km s^{-1}	S/N	rms mJy	Dist Mpc	log M_{HI} $M_\odot$	Code
4-220	89	00 09 54.1 +25 55 43	00 09 53.3 +25 55 26	4569	383(29)	6.40	35.5	2.06	64.6	9.80	1 *
4-221	94	00 10 26.5 +25 50 01	00 10 26.0 +25 49 55	4592	305(12)	9.24	54.8	2.16	64.6	9.96	1 *
4-222	102681	00 10 29.0 +24 31 57		-229	31(16)	0.48	7.3	2.56			9 *
4-223	102626	00 10 35.9 +24 22 19	00 10 35.9 +24 21 51	14676	150(36)	1.08	10.6	1.85	204.8	10.03	1
4-224	102627	00 10 36.2 +24 30 03	00 10 34.8 +24 30 05	11554	177(43)	1.10	9.1	2.04	160.2	9.82	1
4-225	102628	00 10 38.7 +24 31 55	00 10 38.4 +24 31 22	10214	116(10)	1.71	16.7	2.12	141.0	9.90	1 *
4-226	101	00 10 55.7 +25 33 46	00 10 57.0 +25 33 36	10230	471(25)	3.54	19.4	1.73	141.3	10.22	1
4-227	102629	00 11 14.6 +24 42 13	00 11 11.7 +24 42 17	14827	441(74)	1.96	6.9	2.89	206.9	10.30	1 *
4-228	102682	00 11 35.2 +25 34 06	00 11 37.3 +25 33 34	10474	175(7)	0.77	7.3	1.79	144.8	9.58	1 *
4-229	100079	00 12 06.2 +24 59 09	00 12 05.6 +24 59 13	10753	335(37)	3.24	19.8	2.00	148.8	10.23	1
4-230	102630	00 13 12.8 +25 36 35	00 13 12.8 +25 36 14	6238	96(16)	0.76	10.4	1.66	88.4	9.15	1
4-231	102631	00 13 14.8 +24 49 10	00 13 13.6 +24 49 14	11517	279(37)	1.36	7.2	2.54	159.7	9.91	2
4-232	102632	00 13 16.9 +24 15 51	00 13 15.9 +24 15 15	6684	187(12)	1.19	9.2	2.10	90.6	9.36	1
4-233	102634	00 16 03.0 +25 04 02	00 16 00.9 +25 03 44	10742	208(17)	0.82	7.1	1.79	148.6	9.63	1
4-234	102635	00 16 11.4 +24 51 04	00 16 12.1 +24 50 57	9491	94(12)	1.03	12.0	1.96	130.8	9.62	1 *
4-235	102636	00 16 58.8 +24 04 10	00 16 56.6 +24 04 48	6617	163(20)	0.76	6.9	1.91	89.7	9.16	1
4-236	102637	00 16 59.7 +24 25 59	00 16 58.6 +24 26 12	6702	48(8)	2.33	34.0	2.17	90.9	9.66	1
4-237	165	00 17 42.4 +24 40 19	00 17 41.8 +24 40 03	6042	332(22)	3.52	23.1	1.87	85.5	9.78	1 *
4-238	100146	00 17 58.4 +24 33 53	00 17 59.7 +24 33 45	5798	255(25)	2.98	22.6	1.84	82.0	9.67	1 *
4-239	102638	00 18 35.8 +24 27 00	00 18 36.1 +24 26 55	5761	90(25)	2.31	21.7	2.50	81.4	9.56	1
4-240	102639	00 18 56.7 +24 09 47	00 18 55.4 +24 09 47	5984	149(18)	1.11	10.0	2.03	84.6	9.27	1
4-241	102683	00 19 27.7 +25 35 15		-128	25(2)	2.72	19.0	6.08			9 *
4-242	102640	00 20 28.4 +25 06 26	00 20 27.8 +25 06 07	11443	76(12)	1.17	13.1	2.26	158.7	9.84	1
4-243	102641	00 20 39.5 +25 13 27	00 20 41.6 +25 13 40	10677	174(17)	0.95	6.9	2.33	147.7	9.69	1
4-244	102642	00 21 23.0 +25 43 31	00 21 21.2 +25 43 49	10191	233(15)	1.27	8.9	2.08	140.8	9.77	1
4-245	102684	00 22 06.7 +25 29 01	00 22 06.8 +25 29 09	7447	33(4)	0.73	13.7	2.01	101.6	9.25	1
4-246	102250	00 22 13.1 +25 56 34	00 22 13.7 +25 56 37	9106	305(20)	1.16	6.7	2.20	125.3	9.63	1
4-247	102251	00 22 27.1 +25 57 42	00 22 25.5 +25 57 50	10261	197(32)	1.45	11.1	2.08	141.8	9.84	1
4-248	102644	00 22 50.6 +25 48 31	00 22 51.5 +25 47 21	7018	103(21)	1.09	12.6	1.90	95.5	9.37	1 *
4-249	100202	00 22 51.5 +24 08 54	00 22 50.0 +24 09 08	12528	239(69)	1.07	7.8	1.99	174.2	9.88	1 *
4-250	100204	00 22 58.1 +25 13 18	00 22 57.3 +25 12 44	12207	419(27)	1.65	8.4	2.09	169.6	10.05	1
4-251	228	00 23 56.8 +24 18 19	00 23 56.6 +24 18 21	5682	284(19)	6.81	39.9	2.26	80.2	10.01	1
4-252	102645	00 23 57.4 +25 10 13	00 23 58.5 +25 10 39	4410	51(18)	0.63	9.5	2.04	61.8	8.75	1
4-253	102646	00 24 01.4 +24 05 36	00 24 00.9 +24 05 02	12588	286(36)	1.16	8.0	1.90	175.0	9.92	1
4-254	100215	00 24 02.5 +25 39 54	00 24 02.2 +25 39 45	7371	255(10)	1.56	12.8	1.69	100.5	9.57	1
4-255	102685	00 24 43.4 +24 31 38		-122	28(3)	1.46	13.9	4.26			9 *
4-256	102647	00 24 59.4 +25 41 47	00 25 02.8 +25 41 58	9743	265(38)	0.94	6.7	1.94	134.4	9.60	1
4-257	102686	00 25 05.7 +25 17 29		-364	27(3)	1.00	15.8	2.59			9 *
4-258	244	00 25 30.8 +24 48 27	00 25 30.0 +24 48 39	4553	230(25)	2.66	22.4	1.74	63.9	9.41	1
4-259	102254	00 25 31.9 +25 56 53	00 25 32.4 +25 55 55	14586	176(34)	1.49	11.7	2.14	203.6	10.16	1
4-260	102687	00 25 49.7 +25 18 05		-357	25(2)	2.33	45.9	2.15			9 *
4-261	248	00 26 07.1 +25 43 08	00 26 07.0 +25 43 30	10170	469(82)	6.37	27.6	2.20	140.5	10.47	1 *
4-262	102311	00 26 22.1 +24 38 31	00 26 21.0 +24 38 35	3576	104(10)	1.17	13.1	1.96	49.9	8.84	1
4-263	102648	00 26 26.0 +25 31 09	00 26 25.8 +25 30 53	4464	128(18)	1.04	12.9	1.59	62.6	8.98	1

TableA.1 – Continued

Source	AGC #	HI Coords (J2000) hh mm ss.s+dd mm ss	Opt Coords (J2000) hh mm ss.s+dd mm ss	$cz_\odot$ km s^{-1}	$W50$ (ϵ_w) km s^{-1}	F_c Jy km s^{-1}	S/N	rms mJy	Dist Mpc	log M_{HI} $M_\odot$	Code
4-264	102649	00 26 53.0 +25 49 01	00 26 55.4 +25 48 55	10579	226(38)	1.27	9.1	2.08	146.4	9.81	1
4-265	102650	00 27 06.4 +24 25 24	00 27 08.0 +24 25 43	11030	82(6)	0.57	6.8	2.09	152.8	9.50	1
4-266	102651	00 27 07.4 +24 59 14	00 27 07.6 +24 59 07	6378	196(24)	1.63	13.4	1.94	86.4	9.46	1
4-267	261	00 27 17.7 +24 10 11	00 27 17.2 +24 10 09	7409	338(71)	1.73	9.9	2.13	101.1	9.62	1
4-268	102688	00 27 32.6 +25 42 27		-160	22(2)	5.04	35.1	6.44			9 *
4-269	102689	00 27 42.7 +25 07 55		-340	24(2)	2.21	22.7	4.21			9 *
4-270	102652	00 28 05.6 +24 21 05	00 28 06.4 +24 21 10	10879	131(24)	0.65	6.4	1.97	150.6	9.54	1
4-271	273	00 28 06.4 +25 59 53	00 28 06.7 +25 59 47	5600	157(11)	2.42	20.4	2.11	78.9	9.55	1
4-272	102690	00 28 47.9 +25 22 33		-356	20(2)	1.61	25.6	2.92			9 *
4-273	102653	00 29 15.5 +25 24 43	00 29 15.1 +25 24 47	3749	114(11)	1.65	18.6	1.85	52.3	9.03	1
4-274	102692	00 29 53.4 +24 43 58	00 29 51.9 +24 44 35	5439	111(10)	0.91	9.8	1.96	76.6	9.10	1
4-275	102654	00 30 13.1 +25 50 24	00 30 08.8 +25 49 48	9914	245(38)	1.53	9.4	2.32	136.9	9.83	1
4-276	102655	00 30 16.7 +24 18 00	00 30 13.6 +24 17 59	1651	38(11)	0.50	8.5	2.09	23.1	7.80	1
4-277	102693	00 30 20.9 +25 30 18		-352	24(2)	1.64	24.1	2.98			9 *
4-278	302	00 30 26.3 +25 08 19	00 30 26.1 +25 08 32	5413	120(10)	1.42	13.5	2.15	76.2	9.29	1
4-279	102656	00 30 55.0 +24 37 26	00 30 54.4 +24 37 42	13679	332(70)	1.46	9.2	1.95	190.7	10.10	1
4-280	102657	00 32 29.0 +24 49 31	00 32 27.6 +24 49 54	5372	165(7)	0.77	7.5	1.78	75.5	9.01	1 *
4-281	102694	00 32 46.4 +24 45 57		-357	29(2)	4.28	40.4	4.23			9 *
4-282	102658	00 34 14.7 +25 01 57	00 34 15.4 +25 01 44	5537	234(15)	0.94	7.1	1.95	77.9	9.13	1
4-283	337	00 34 24.3 +24 36 13	00 34 24.8 +24 36 13	5334	144(13)	5.01	48.7	1.91	74.9	9.82	1
4-284	102659	00 34 25.6 +24 21 49	00 34 24.6 +24 21 44	12874	620(85)	2.70	9.5	2.05	179.2	10.31	1
4-285	102581	00 35 14.2 +25 58 56		-359	23(2)	0.74	13.0	2.49			9 *
4-286	102695	00 35 32.1 +24 09 18		-363	25(2)	10.01	61.8	6.94			9 *
4-287	100323	00 35 46.6 +25 40 18	00 35 39.1 +25 40 43	9844	131(14)	0.64	7.0	1.76	135.9	9.45	1 *
4-288	352	00 35 47.2 +24 14 28	00 35 46.6 +24 14 14	5149	171(14)	1.89	15.3	2.11	72.2	9.37	1
4-289	354	00 35 57.4 +24 01 42	00 35 57.9 +24 02 15	5612	225(8)	1.26	8.4	2.23	78.9	9.27	1
4-290	102696	00 36 41.9 +25 20 01		-350	25(2)	5.75	47.4	5.20			9 *
4-291	102660	00 37 08.5 +24 54 51	00 37 08.6 +24 54 56	5145	188(17)	0.95	7.2	2.14	72.2	9.07	1
4-292	100347	00 37 13.7 +25 49 42	00 37 11.6 +25 50 21	9025	128(26)	0.90	7.4	2.40	124.3	9.52	2 *
4-293	100350	00 37 44.2 +24 12 22	00 37 44.1 +24 12 29	4653	107(13)	0.78	8.5	1.98	64.2	8.88	1
4-294	102697	00 37 52.6 +25 42 37		-351	38(11)	2.07	22.3	3.28			9 *
4-295	102661	00 38 00.0 +25 33 54	00 38 00.1 +25 34 06	7772	171(18)	0.87	9.5	1.57	106.4	9.37	1
4-296	101771	00 38 11.9 +25 23 52	00 38 11.0 +25 23 45	5195	99(13)	1.77	20.9	1.90	72.9	9.35	1
4-297	102229	00 38 23.6 +25 26 11	00 38 24.0 +25 26 10	3314	82(8)	1.60	22.5	1.73	45.9	8.90	1
4-298	102662	00 38 26.6 +25 37 31	00 38 21.4 +25 36 48	9915	275(29)	1.15	7.1	2.19	137.0	9.71	1 *
4-299	100377	00 38 49.3 +25 43 09	00 38 47.9 +25 43 04	7243	243(24)	1.56	10.7	2.09	98.8	9.55	1
4-300	398	00 38 57.9 +25 38 31	00 38 57.3 +25 38 21	4613	199(2)	1.84	18.3	1.59	63.8	9.25	1
4-301	101658	00 39 06.5 +25 36 37	00 39 00.8 +25 35 53	10328	101(11)	0.57	6.9	1.83	142.9	9.44	1
4-302	102663	00 39 18.4 +24 43 54	00 39 18.6 +24 43 43	5157	122(7)	1.60	15.9	2.03	72.3	9.29	1
4-303	100395	00 39 26.7 +25 15 02	00 39 25.7 +25 15 21	4648	171(21)	2.62	22.6	1.98	63.8	9.40	1
4-304	100762	00 41 07.0 +25 35 57	00 41 10.1 +25 35 07	10001	378(46)	1.45	10.3	1.60	138.2	9.81	1
4-305	438	00 41 27.8 +25 30 01	00 41 27.8 +25 30 00	4537	351(4)	5.40	37.4	1.72	63.8	9.71	1
4-306	100470	00 42 00.4 +25 34 09	00 42 00.8 +25 33 56	10139	214(27)	1.76	18.2	1.48	140.2	9.91	1
4-307	102664	00 42 04.3 +25 20 14	00 42 06.4 +25 19 55	10211	187(15)	0.89	6.9	2.09	141.2	9.62	1

TableA.1 – Continued

Source	AGC #	HI Coords (J2000) hh mm ss.s+dd mm ss	Opt Coords (J2000) hh mm ss.s+dd mm ss	$cz_\odot$ km s^{-1}	$W50$ (ϵ_w) km s^{-1}	F_c Jy km s^{-1}	S/N	rms mJy	Dist Mpc	log M_{HI} $M_\odot$	Code
4-308	102665	00 42 26.0 +25 10 54	00 42 26.5 +25 11 01	3285	92(7)	2.54	25.8	2.29	45.4	9.09	1 *
4-309	102268	00 42 58.8 +25 54 24	00 42 59.2 +25 54 09	4490	167(12)	0.90	8.1	1.92	62.6	8.92	1
4-310	102666	00 43 11.0 +24 29 11	00 43 10.1 +24 29 21	13282	330(85)	1.82	9.7	2.31	185.1	10.17	1
4-311	102301	00 43 43.3 +25 52 03	00 43 42.8 +25 51 48	5180	166(8)	0.83	6.8	2.11	72.5	9.01	1
4-312	102667	00 44 16.5 +25 45 51	00 44 16.8 +25 45 57	13547	174(6)	1.03	7.8	2.24	188.9	9.94	1
4-313	102668	00 44 35.6 +25 41 57	00 44 35.6 +25 41 14	14157	136(19)	0.81	7.5	2.05	197.6	9.87	1 *
4-314	102669	00 44 49.8 +25 27 01	00 44 46.7 +25 26 44	12004	347(31)	3.21	21.2	1.81	166.9	10.32	1
4-315	100502	00 45 04.7 +25 47 45	00 45 06.5 +25 46 56	13653	190(40)	1.02	6.4	2.58	190.4	9.94	1
4-316	102670	00 45 18.3 +24 08 38	00 45 18.3 +24 08 41	2025	63(11)	0.72	9.7	2.06	27.9	8.12	1
4-317	100509	00 45 57.6 +25 14 28	00 45 54.0 +25 15 53	13596	687(148)	4.33	12.2	2.30	189.6	10.56	1 *
4-318	102272	00 46 14.7 +25 55 08	00 46 14.5 +25 55 05	13899	278(64)	1.64	9.9	2.21	194.0	10.16	1
4-319	102671	00 46 23.1 +25 26 31	00 46 22.9 +25 26 25	13458	209(28)	1.19	8.6	2.11	187.6	9.99	1
4-320	102672	00 46 24.3 +25 04 56	00 46 24.0 +25 04 14	5275	170(40)	1.13	6.8	2.85	73.9	9.16	1
4-321	100553	00 46 56.1 +24 44 53	00 46 56.2 +24 45 26	12674	85(22)	1.13	11.4	2.38	176.4	9.92	1
4-322	102673	00 48 41.7 +24 35 30	00 48 41.3 +24 35 03	7144	149(31)	1.04	9.6	1.99	97.4	9.37	1 *
4-323	102674	00 49 15.8 +25 17 02	00 49 14.4 +25 17 35	13890	131(18)	1.11	11.4	1.89	193.8	9.99	1
4-324	102699	00 49 37.9 +24 02 00		-332	23(2)	7.51	59.9	5.53			9 *
4-325	102276	00 49 52.2 +25 56 57	00 49 52.1 +25 56 39	4984	149(4)	1.03	8.5	2.23	69.6	9.07	1
4-326	102277	00 49 54.1 +25 49 24	00 49 54.0 +25 50 01	14122	452(38)	1.64	6.7	2.41	197.2	10.18	1
4-327	100594	00 50 22.2 +24 31 08	00 50 12.9 +24 29 51	10096	452(50)	4.72	20.0	2.34	139.6	10.34	1 *
4-328	102700	00 51 05.0 +25 33 01	00 51 07.0 +25 32 46	13769	63(10)	0.77	11.1	1.94	192.1	9.83	1
4-329	102675	00 51 27.7 +25 53 28	00 51 29.5 +25 54 06	14838	513(54)	1.91	6.9	2.40	207.4	10.29	1
4-330	102717	00 52 27.4 +25 03 44		-130	24(6)	0.89	7.2	5.45			9 *
4-331	534	00 52 30.9 +24 21 01	00 52 30.2 +24 21 04	10165	568(33)	7.53	30.2	1.96	140.6	10.55	1
4-332	102282	00 53 18.7 +25 55 00	00 53 18.6 +25 54 43	6980	237(16)	1.62	11.8	1.99	95.2	9.54	1
4-333	101042	00 53 59.4 +25 47 22	00 54 02.3 +25 48 07	13691	379(41)	1.31	6.9	2.19	191.1	10.05	2
4-334	102676	00 54 18.6 +25 26 08	00 54 17.6 +25 26 15	9974	213(26)	0.87	6.7	1.97	137.9	9.59	2
4-335	100626	00 54 41.2 +24 52 01	00 54 41.5 +24 52 11	13521	114(20)	1.78	18.1	2.06	188.6	10.17	1
4-336	101857	00 54 55.8 +25 53 42	00 54 56.0 +25 53 23	13774	247(28)	1.01	5.9	2.42	192.2	9.94	2 *
4-337	102702	00 55 33.1 +25 13 59		-367	28(8)	0.82	14.4	2.32			9 *
4-338	100635	00 55 36.5 +24 11 55	00 55 35.4 +24 11 43	4975	78(7)	0.90	12.5	1.81	70.2	9.02	1
4-339	102703	00 56 03.0 +25 35 53		-350	23(4)	0.44	8.9	2.19			9 *
4-340	102714	00 56 18.5 +24 02 21	00 56 16.4 +24 03 12	4963	27(16)	0.32	6.7	1.99	69.2	8.56	2
4-341	102677	00 56 30.8 +25 32 39	00 56 34.2 +25 33 03	14812	57(7)	0.57	7.3	2.29	207.1	9.76	1
4-342	102678	00 56 34.1 +24 18 55	00 56 37.7 +24 18 54	6501	171(28)	1.10	9.8	1.92	88.3	9.31	1 *
4-343	102704	00 57 18.7 +24 26 29		-344	24(2)	3.47	39.7	3.82			9 *
4-344	102705	00 58 55.9 +25 29 28		-378	26(4)	0.72	14.8	2.02			9 *
4-345	102679	00 59 46.6 +24 25 43	00 59 46.0 +24 26 03	5549	148(20)	0.68	6.3	1.99	77.6	8.98	1
4-346	113870	01 00 05.1 +25 49 51		-140	23(2)	3.68	34.3	4.72			9 *
4-347	113746	01 00 27.2 +25 53 23	01 00 27.9 +25 53 41	12457	184(19)	1.09	8.7	2.05	173.5	9.89	1
4-348	113808	01 02 12.6 +25 31 51	01 02 11.5 +25 32 16	9615	47(6)	0.58	9.7	1.92	132.9	9.38	1
4-349	643	01 03 06.1 +24 58 09	01 03 05.6 +24 58 19	9041	279(3)	4.27	28.3	2.02	124.7	10.19	1
4-350	113809	01 03 38.8 +24 10 12	01 03 36.0 +24 10 08	13687	213(18)	1.09	7.5	2.24	191.0	9.97	1
4-351	113810	01 04 07.3 +24 17 24	01 04 05.3 +24 17 53	13789	209(9)	1.20	8.7	2.14	192.5	10.02	1

TableA.1 – Continued

Source	AGC #	HI Coords (J2000) hh mm ss.s+dd mm ss	Opt Coords (J2000) hh mm ss.s+dd mm ss	$cz_\odot$ km s^{-1}	$W50$ (ϵ_w) km s^{-1}	F_c Jy km s^{-1}	S/N	rms mJy	Dist Mpc	log M_{HI} $M_\odot$	Code
4-352	113837	01 04 55.6 +25 25 38		-131	20(3)	9.05	59.1	7.20			9 *
4-353	113811	01 04 59.2 +24 29 35	01 04 59.0 +24 29 42	13885	228(13)	1.58	10.1	2.30	193.9	10.15	1
4-354	113812	01 05 56.1 +24 34 08	01 05 55.2 +24 34 01	13524	202(18)	1.20	8.5	2.23	188.7	10.00	1 *
4-355	110025	01 06 18.3 +25 32 41	01 06 11.8 +25 33 06	6710	154(6)	1.52	13.9	1.97	91.4	9.48	1 *
4-356	110030	01 06 38.8 +24 12 46	01 06 39.3 +24 12 57	11744	316(11)	1.03	6.0	2.13	163.3	9.81	2
4-357	113813	01 07 00.1 +25 35 26	01 07 01.9 +25 34 59	5026	69(12)	0.81	10.1	2.14	69.9	8.97	1
4-358	113764	01 07 34.4 +25 59 38		-427	16(2)	73.23	593.7	6.24			9 *
4-359	113839	01 08 02.5 +25 42 00		-424	14(3)	42.43	423.3	5.32			9 *
4-360	113840	01 08 02.8 +25 54 31		-426	15(2)	207.01	989.8	10.88			9 *
4-361	113842	01 08 45.6 +25 52 00		-425	12(3)	31.13	385.2	4.45			9 *
4-362	113843	01 10 33.9 +25 06 00		84	16(3)	0.80	11.2	3.58			9 *
4-363	113814	01 15 50.2 +24 22 07	01 15 50.4 +24 21 51	8170	274(19)	3.44	22.9	2.03	112.4	10.01	1
4-364	110178	01 16 13.8 +25 11 08	01 16 12.5 +25 11 30	8949	226(19)	1.22	8.6	2.11	123.5	9.64	1 *
4-365	113871	01 17 12.4 +24 49 54		-134	22(4)	0.67	9.5	3.12			9 *
4-366	113845	01 17 18.0 +24 08 12	01 17 21.7 +24 08 16	8192	82(17)	0.58	6.7	2.11	112.7	9.24	1
4-367	113815	01 17 38.5 +24 08 04	01 17 37.9 +24 08 31	11645	287(29)	1.88	12.4	2.00	162.0	10.07	1
4-368	113846	01 18 32.7 +24 30 05		-131	23(3)	1.23	11.3	4.77			9 *
4-369	113816	01 18 46.7 +24 27 08	01 18 47.5 +24 27 18	4956	144(6)	0.97	9.2	1.97	68.7	9.03	1
4-370	113817	01 18 47.7 +24 11 17	01 18 46.6 +24 11 23	8265	120(20)	1.54	15.9	1.96	113.7	9.67	1
4-371	113818	01 20 14.5 +24 44 53	01 20 15.5 +24 44 58	11610	213(13)	1.35	9.3	2.21	161.6	9.92	1
4-372	884	01 20 58.4 +25 31 37	01 20 58.6 +25 32 54	9039	117(21)	0.66	7.0	1.95	124.9	9.38	1 *
4-373	113754	01 23 05.2 +25 54 42	01 23 07.6 +25 53 57	16430	130(9)	0.71	6.4	2.18	230.5	9.95	2
4-374	113820	01 23 49.3 +24 32 27	01 23 47.5 +24 32 47	16235	200(16)	1.02	7.3	2.23	227.7	10.10	1
4-375	113821	01 24 18.0 +25 36 01	01 24 17.8 +25 36 30	8155	247(6)	2.22	15.3	2.06	112.3	9.82	1
4-376	113822	01 24 21.2 +25 29 01	01 24 21.4 +25 29 01	8108	252(3)	3.03	20.8	2.05	111.6	9.95	1
4-377	113848	01 24 30.7 +25 36 01		-342	21(2)	9.89	82.9	5.51			9 *
4-378	113849	01 26 59.7 +24 58 56		-255	35(7)	0.80	8.3	3.57			9 *
4-379	113850	01 27 41.5 +25 26 12		-248	17(5)	0.80	12.1	3.32			9 *
4-380	113823	01 28 47.0 +24 54 11	01 28 47.8 +24 53 23	8664	124(34)	1.07	10.6	2.01	119.6	9.56	1 *
4-381	113851	01 30 03.6 +24 22 42		-264	23(2)	2.94	26.4	4.94			9 *
4-382	1073	01 30 07.6 +25 52 00	01 30 07.5 +25 51 48	3674	153(2)	9.80	82.1	2.15	50.1	9.76	1
4-383	110389	01 30 52.8 +25 55 00	01 30 50.2 +25 54 54	11211	268(26)	1.17	7.6	2.11	156.0	9.83	1
4-384	113852	01 33 37.2 +24 48 21		-263	23(3)	1.73	15.2	5.05			9 *
4-385	112465	01 35 11.4 +24 00 29	01 35 10.1 +24 00 25	3315	111(2)	6.04	62.8	2.03	45.0	9.46	1
4-386	113853	01 35 12.9 +25 05 02		-240	23(2)	1.12	14.2	3.48			9 *
4-387	113824	01 35 43.9 +25 47 46	01 35 45.1 +25 47 40	16335	250(5)	1.34	9.3	2.03	229.3	10.22	1
4-388	113854	01 37 47.6 +24 31 40	01 37 47.7 +24 32 12	10243	48(12)	0.56	8.1	2.19	142.2	9.43	1 *
4-389	113855	01 38 07.5 +24 55 29		-108	34(2)	8.43	58.5	5.39			9 *
4-390	113856	01 38 58.7 +24 10 38		-108	28(3)	2.36	21.3	4.49			9 *
4-391	113825	01 43 27.9 +24 46 49	01 43 27.2 +24 46 47	3848	107(9)	1.38	15.6	1.91	52.4	8.95	1
4-392	110497	01 43 48.6 +24 14 51	01 43 48.4 +24 14 27	10421	221(41)	1.15	8.2	2.12	144.8	9.75	1
4-393	113826	01 44 04.8 +24 51 58	01 44 03.5 +24 51 59	12884	124(8)	0.78	6.9	2.24	180.0	9.78	1
4-394	113827	01 44 23.2 +25 45 42	01 44 22.7 +25 45 36	10181	143(18)	1.10	9.6	2.13	141.5	9.71	1
4-395	113307	01 44 30.1 +25 51 07	01 44 29.4 +25 50 47	4905	193(4)	2.31	17.6	2.10	67.6	9.40	1 *

TableA.1 – Continued

Source	AGC #	HI Coords (J2000) hh mm ss.s+dd mm ss	Opt Coords (J2000) hh mm ss.s+dd mm ss	$cz_{\odot}$ km s^{-1}	$W50$ (ϵ_w) km s^{-1}	F_c Jy km s^{-1}	S/N	rms mJy	Dist Mpc	log M_{HI} $M_{\odot}$	Code
4-396	113731	01 44 49.3 +25 27 43	01 44 49.6 +25 27 43	3323	107(2)	2.52	26.7	2.02	45.0	9.08	1
4-397	111329	01 45 12.9 +24 55 07	01 45 11.8 +24 56 18	12706	326(43)	1.07	6.0	2.21	177.5	9.90	2 *
4-398	110509	01 45 13.0 +25 47 18	01 45 12.9 +25 47 32	3378	67(14)	1.06	15.6	1.83	45.7	8.72	1
4-399	1230	01 45 32.7 +25 31 17	01 45 32.5 +25 31 15	3837	88(2)	11.82	145.4	1.92	52.2	9.88	1
4-400	113244	01 46 00.0 +25 37 01	01 45 59.3 +25 36 20	3293	123(2)	1.39	14.0	2.00	44.5	8.81	1
4-401	113857	01 46 22.6 +25 21 14	01 46 23.0 +25 21 32	13096	236(33)	1.37	8.9	2.24	183.1	10.03	1 *
4-402	113828	01 46 37.9 +25 11 32	01 46 37.7 +25 11 40	3675	91(16)	0.92	11.0	1.95	49.9	8.73	1
4-403	113829	01 46 48.2 +25 31 35	01 46 46.5 +25 31 59	12809	225(5)	1.28	8.4	2.27	179.0	9.99	1
4-404	1244	01 46 50.4 +24 28 07	01 46 49.5 +24 27 53	3122	189(2)	4.59	36.5	2.04	42.1	9.28	1
4-405	110531	01 47 48.7 +25 35 04	01 47 47.6 +25 34 25	12241	345(5)	2.57	14.0	2.21	170.9	10.25	1
4-406	113297	01 47 54.2 +25 51 59	01 47 54.1 +25 52 24	12188	408(5)	3.09	15.4	2.20	170.2	10.32	1
4-407	113830	01 49 55.0 +25 40 36	01 49 56.2 +25 41 06	14341	34(8)	0.53	9.1	2.20	201.0	9.70	1
4-408	113831	01 50 04.9 +25 25 31	01 50 04.0 +25 25 31	10216	110(3)	2.52	27.9	1.92	142.0	10.08	1
4-409	113832	01 53 49.6 +24 35 17	01 53 49.8 +24 34 22	7397	56(7)	0.50	7.7	1.92	101.8	9.09	1
4-410	113833	01 54 14.7 +25 39 36	01 54 14.0 +25 39 37	12589	476(46)	2.40	10.6	2.13	176.0	10.24	1
4-411	113834	01 54 49.8 +25 15 26	01 54 47.1 +25 15 32	14020	149(38)	0.79	7.0	2.08	196.4	9.86	1
4-412	113858	01 55 22.8 +24 19 07		-104	29(2)	11.03	61.9	7.08			9 *
4-413	113835	01 56 25.6 +24 58 32	01 56 28.7 +24 58 45	12552	126(22)	0.85	7.8	2.17	175.5	9.79	1
4-414	113300	01 58 01.3 +25 53 34	01 58 00.5 +25 53 44	12593	281(36)	1.20	7.5	2.13	176.1	9.94	1
4-415	1451	01 58 30.0 +25 21 32	01 58 30.1 +25 21 36	4918	325(15)	3.94	23.6	2.07	67.8	9.63	1
4-416	1453	01 58 46.5 +24 38 53	01 58 45.8 +24 38 34	4897	85(3)	2.92	35.6	1.97	67.8	9.50	1
4-417	1455	01 58 48.5 +24 53 55	01 58 48.1 +24 53 31	5120	138(3)	28.37	236.5	2.27	67.8	10.49	1 *
4-418	1462	01 59 09.1 +25 23 02	01 59 09.6 +25 23 08	5062	214(6)	3.96	30.0	2.01	67.8	9.63	1
4-419	110712	01 59 16.9 +24 24 48	01 59 15.6 +24 24 59	3886	176(29)	1.27	10.7	1.99	52.8	8.92	1
4-420	110720	01 59 55.6 +24 18 59	01 59 55.2 +24 18 44	5096	262(18)	1.48	10.7	1.91	67.8	9.21	1
4-421	113836	01 59 55.5 +24 58 32	01 59 56.5 +24 58 33	4885	79(12)	1.65	20.7	2.00	67.2	9.24	1
4-422	1478	02 00 15.5 +24 15 08	02 00 14.9 +24 15 08	4846	95(2)	3.52	38.8	2.07	67.8	9.58	1
4-423	1479	02 00 20.6 +24 29 18	02 00 19.1 +24 28 26	4929	497(109)	1.40	6.6	1.90	67.8	9.18	2
4-424	1486	02 00 23.0 +24 34 49	02 00 23.4 +24 34 50	5067	234(3)	4.62	35.0	1.93	67.8	9.70	1
4-425	122860	02 01 16.8 +24 04 26	02 01 16.3 +24 04 12	9827	159(46)	1.03	9.0	2.01	136.6	9.66	1
4-426	1551	02 03 39.2 +24 04 26	02 03 37.6 +24 04 31	2670	119(2)	13.80	137.9	2.04	35.8	9.62	1
4-427	122137	02 03 45.3 +24 01 26	02 03 43.0 +24 01 41	4814	172(14)	1.34	10.2	2.23	66.1	9.14	1
4-428	1561	02 04 05.8 +24 12 24	02 04 05.0 +24 12 28	606	39(2)	3.13	54.7	2.00	8.6	7.73	1
4-429	122861	02 05 07.5 +24 58 14	02 05 07.9 +24 58 15	4842	169(4)	1.89	15.9	2.04	66.5	9.29	1
4-430	1575	02 05 11.5 +24 40 08	02 05 10.2 +24 39 58	4836	196(3)	4.67	38.3	1.94	66.5	9.69	1
4-431	120018	02 05 19.3 +25 06 21	02 05 20.3 +25 06 19	4815	202(6)	4.10	31.4	2.05	66.1	9.63	1
4-432	122909	02 06 06.9 +24 03 06		-140	23(3)	9.89	62.9	6.93			9 *
4-433	122862	02 07 30.3 +24 15 28	02 07 29.0 +24 16 04	9606	204(6)	0.98	7.2	2.13	133.5	9.61	1
4-434	122863	02 07 58.1 +25 36 54	02 07 56.1 +25 36 51	2771	51(8)	0.66	10.3	1.96	37.1	8.33	1
4-435	122864	02 08 21.2 +24 56 08	02 08 21.1 +24 56 08	8093	34(2)	2.04	39.0	1.93	111.9	9.78	1
4-436	1648	02 09 13.5 +25 34 23	02 09 14.1 +25 34 14	4878	344(8)	5.08	31.1	1.97	67.0	9.73	1
4-437	1675	02 11 13.4 +25 49 03	02 11 13.8 +25 48 38	5109	221(49)	1.98	14.3	2.08	70.0	9.36	1
4-438	122865	02 11 28.2 +25 20 32	02 11 30.2 +25 20 52	5106	145(22)	1.02	9.4	2.00	70.4	9.08	1
4-439	122866	02 11 31.4 +24 12 51	02 11 31.6 +24 12 54	2790	37(4)	0.59	10.4	2.04	37.3	8.29	1

TableA.1 – Continued

Source	AGC #	HI Coords (J2000) hh mm ss.s+dd mm ss	Opt Coords (J2000) hh mm ss.s+dd mm ss	$cz_\odot$ km s^{-1}	$W50$ (ϵ_w) km s^{-1}	F_c Jy km s^{-1}	S/N	rms mJy	Dist Mpc	log M_{HI} $M_\odot$	Code
4-440	122910	02 12 25.6 +24 35 13		-125	21(3)	13.38	98.5	6.20			9 *
4-441	122911	02 13 06.1 +25 22 04	02 13 05.2 +25 22 39	11844	207(20)	1.90	13.5	2.19	165.6	10.09	1 *
4-442	1706	02 13 33.8 +25 51 19	02 13 33.9 +25 51 19	4794	255(7)	2.37	16.0	2.08	70.0	9.44	1
4-443	1711	02 13 45.4 +24 53 10	02 13 45.7 +24 53 26	2640	149(4)	1.71	16.9	1.85	35.3	8.70	1
4-444	122441	02 15 21.6 +24 06 47	02 15 21.6 +24 06 57	2543	99(2)	2.08	22.6	2.05	34.0	8.75	1
4-445	122867	02 15 32.2 +25 02 41	02 15 31.3 +25 02 36	5100	121(5)	2.09	22.6	1.87	70.2	9.39	1
4-446	122912	02 15 41.4 +24 45 15		-127	24(2)	5.56	42.0	5.77			9 *
4-447	1739	02 15 45.1 +25 12 22	02 15 44.2 +25 12 23	5094	251(5)	5.37	38.8	1.95	70.0	9.79	1 *
4-448	121503	02 16 10.1 +25 13 56	02 16 10.7 +25 14 13	4987	236(7)	9.79	71.0	2.01	68.6	10.04	1 *
4-449	122913	02 16 18.5 +25 26 18		9767	73(9)	0.83	10.8	1.99	136.0	9.56	1 *
4-450	122914	02 18 03.8 +25 22 08		-125	23(3)	3.10	29.4	4.68			9 *
4-451	122915	02 18 14.8 +24 09 27		-123	26(2)	18.42	62.1	12.39			9 *
4-452	122451	02 18 21.4 +24 32 20	02 18 22.7 +24 32 07	12863	87(8)	2.48	26.7	2.22	180.2	10.28	1 *
4-453	122916	02 18 31.8 +25 05 15		-121	24(3)	3.67	30.2	5.32			9 *
4-454	122868	02 18 52.6 +24 07 27	02 18 50.2 +24 07 34	13167	86(24)	1.51	16.4	2.20	184.6	10.08	1
4-455	122797	02 20 49.3 +25 58 54	02 20 49.0 +25 59 02	11197	62(9)	0.99	13.3	2.10	156.5	9.76	1
4-456	122917	02 21 10.2 +25 13 11		-114	26(3)	4.22	22.5	7.80			9 *
4-457	1812	02 21 24.0 +25 25 22	02 21 25.3 +25 25 22	4527	236(7)	2.73	19.3	2.05	61.8	9.39	1
4-458	122798	02 22 15.8 +25 52 37	02 22 18.7 +25 52 39	9756	166(21)	1.52	12.2	2.15	135.9	9.82	1 *
4-459	120189	02 22 49.5 +25 56 43	02 22 51.6 +25 56 39	14693	152(53)	1.47	12.1	2.19	206.4	10.17	1
4-460	120193	02 22 55.0 +25 18 53	02 22 55.3 +25 18 35	4584	174(21)	4.24	38.0	1.89	62.7	9.59	1 *
4-461	122918	02 23 16.9 +24 07 53		-134	32(2)	9.79	70.5	5.33			9 *
4-462	120202	02 23 33.6 +25 27 55	02 23 31.8 +25 27 54	11289	410(33)	1.16	6.4	1.97	157.8	9.83	2
4-463	120207	02 23 52.0 +25 32 31	02 23 51.9 +25 32 30	5111	251(23)	1.56	11.7	1.87	70.4	9.26	1
4-464	122919	02 23 52.6 +24 19 59		-129	27(3)	22.46	122.1	7.65			9 *
4-465	122869	02 24 35.4 +24 16 23	02 24 35.1 +24 16 56	9367	157(10)	0.97	8.2	2.11	130.4	9.59	1
4-466	1860	02 24 40.5 +25 33 56	02 24 42.2 +25 33 39	9665	447(9)	2.38	12.6	1.89	134.6	10.01	1
4-467	120216	02 24 42.5 +25 32 01	02 24 43.4 +25 31 56	10477	271(24)	1.13	7.4	2.09	146.2	9.76	1
4-468	122870	02 24 45.1 +25 13 21	02 24 45.4 +25 13 27	10441	146(7)	1.34	11.3	2.18	145.7	9.83	1
4-469	122871	02 25 08.7 +24 50 08	02 25 10.4 +24 49 42	5431	64(10)	0.89	13.3	1.84	75.1	9.07	1
4-470	122872	02 25 16.6 +24 27 12	02 25 19.3 +24 27 14	12068	194(7)	0.91	7.3	1.98	169.0	9.79	1
4-471	122873	02 25 41.0 +25 47 31	02 25 42.9 +25 47 04	10283	162(41)	0.89	8.1	1.93	143.5	9.64	1
4-472	120240	02 25 54.1 +24 51 25	02 25 55.7 +24 51 25	10799	79(16)	0.95	11.0	2.15	150.9	9.71	1
4-473	120241	02 25 55.2 +24 49 23	02 25 57.4 +24 49 17	12209	402(95)	1.47	8.1	2.03	171.0	10.01	2 *
4-474	122811	02 26 07.4 +25 58 30	02 26 05.6 +25 58 01	5900	135(6)	0.78	7.3	2.07	82.0	9.09	1 *
4-475	122874	02 26 14.8 +24 26 34	02 26 14.7 +24 26 02	6386	135(10)	0.69	6.6	1.99	87.8	9.10	1
4-476	122875	02 26 46.9 +24 40 08	02 26 45.7 +24 40 18	9688	223(4)	1.56	11.4	2.04	135.0	9.83	1 *
4-477	121820	02 26 54.1 +25 02 10	02 26 54.1 +25 01 57	9635	399(7)	1.86	10.6	1.97	134.2	9.90	1
4-478	122876	02 27 15.4 +24 05 49	02 27 14.8 +24 06 30	10113	139(23)	0.75	6.8	2.09	141.1	9.55	1
4-479	122877	02 27 33.7 +24 51 58	02 27 31.9 +24 52 12	6100	81(5)	0.83	10.0	2.03	85.0	9.15	1
4-480	122878	02 27 47.2 +24 58 36	02 27 46.6 +24 58 45	10548	258(12)	1.64	11.2	2.03	147.3	9.92	1
4-481	122879	02 27 49.1 +24 15 23	02 27 51.6 +24 15 36	6226	125(3)	1.36	12.2	2.21	86.8	9.38	1
4-482	122880	02 28 03.8 +25 31 39	02 28 01.7 +25 31 31	1831	30(4)	0.46	8.6	2.09	24.3	7.81	1
4-483	122920	02 28 25.0 +24 43 10		-124	27(3)	22.68	147.3	6.36			9 *

TableA.1 – Continued

Source	AGC #	HI Coords (J2000) hh mm ss.s+dd mm ss	Opt Coords (J2000) hh mm ss.s+dd mm ss	$cz_{\odot}$ km s^{-1}	$W50$ (ϵ_w) km s^{-1}	F_c Jy km s^{-1}	S/N	rms mJy	Dist Mpc	log M_{HI} $M_{\odot}$	Code
4-484	122881	02 28 29.6 +25 30 53	02 28 28.7 +25 30 36	9056	124(7)	1.33	13.4	1.98	126.0	9.70	1
4-485	1955	02 28 54.4 +25 20 22	02 28 54.3 +25 20 38	5207	227(4)	4.54	32.7	2.06	71.8	9.74	1
4-486	122921	02 29 09.3 +25 18 53	02 29 08.7 +25 19 28	11567	253(42)	2.20	15.3	2.02	161.9	10.13	1 *
4-487	122922	02 29 10.4 +25 37 48		-113	32(3)	25.95	99.9	9.97			9 *
4-488	1970	02 29 54.5 +25 15 15	02 29 54.1 +25 15 22	1913	226(3)	6.09	42.8	2.11	25.4	8.97	1
4-489	122442	02 30 15.7 +24 06 38	02 30 16.3 +24 06 32	6064	214(6)	2.41	18.6	1.97	84.4	9.61	1
4-490	122882	02 30 52.2 +24 58 38	02 30 54.9 +24 58 51	5929	205(52)	0.99	6.6	2.33	82.4	9.20	1
4-491	120300	02 30 54.8 +25 24 42	02 30 52.9 +25 24 27	5346	166(47)	0.96	8.3	2.00	73.8	9.09	2 *
4-492	122400	02 31 21.3 +25 42 31	02 31 22.1 +25 42 45	938	31(2)	0.84	15.4	2.13	12.7	7.51	1
4-493	122883	02 31 42.7 +24 18 03	02 31 42.1 +24 17 19	5409	194(17)	1.60	13.2	1.95	74.7	9.32	1
4-494	122884	02 32 53.7 +25 09 32	02 32 53.4 +25 09 16	2433	119(2)	5.28	49.8	2.17	32.4	9.12	1
4-495	120325	02 32 54.0 +25 05 43	02 32 54.3 +25 05 36	5416	143(5)	2.66	24.4	2.02	74.8	9.55	1
4-496	2025	02 33 14.1 +25 30 08	02 33 14.4 +25 30 21	11091	489(30)	3.63	15.0	2.22	155.1	10.31	1
4-497	2052	02 34 24.2 +25 16 02	02 34 25.8 +25 16 10	5700	173(3)	1.19	10.2	1.97	79.0	9.24	1
4-498	122886	02 35 01.1 +24 39 31	02 35 00.2 +24 39 29	5574	162(6)	1.51	13.0	2.04	77.2	9.33	1
4-499	122887	02 35 23.5 +25 33 21	02 35 22.0 +25 32 41	11026	370(2)	2.85	16.9	1.96	154.3	10.20	1 *
4-500	122888	02 35 49.1 +24 45 51	02 35 49.9 +24 46 06	11119	171(32)	0.81	6.5	2.14	155.6	9.66	1
4-501	2082	02 36 16.5 +25 25 25	02 36 16.2 +25 25 23	707	192(2)	41.20	277.6	2.39	9.8	8.97	1
4-502	120995	02 37 00.9 +25 37 31	02 37 00.2 +25 37 30	10837	509(55)	2.40	9.6	2.18	151.6	10.11	1
4-503	122889	02 39 43.7 +24 58 52	02 39 44.3 +24 59 03	12825	250(7)	1.60	10.5	2.16	180.0	10.09	1 *
4-504	122890	02 41 20.6 +24 16 28	02 41 20.1 +24 16 13	13975	204(25)	1.19	8.3	2.25	196.5	10.03	1 *
4-505	122891	02 41 55.6 +24 46 01	02 41 57.3 +24 45 54	14013	87(30)	0.70	6.9	2.40	197.0	9.81	1
4-506	122892	02 41 55.5 +24 55 50	02 41 55.7 +24 56 10	9870	124(3)	1.48	14.2	2.09	137.8	9.82	1 *
4-507	122893	02 42 05.8 +24 41 47	02 42 06.4 +24 41 12	9827	180(22)	1.06	7.4	2.38	137.2	9.67	1
4-508	122894	02 42 27.0 +24 49 24	02 42 30.6 +24 49 18	7696	100(12)	0.68	7.0	2.16	106.8	9.26	1
4-509	122424	02 45 07.8 +25 56 30	02 45 07.1 +25 56 10	1536	46(7)	0.41	6.3	2.12	20.4	7.61	2
4-510	122895	02 46 13.0 +24 50 25	02 46 14.1 +24 50 12	7431	70(5)	0.67	8.1	2.19	103.1	9.22	1
4-511	2251	02 47 02.3 +24 51 31	02 47 02.6 +24 51 27	7470	309(3)	2.90	17.9	2.05	103.6	9.87	1
4-512	122896	02 47 02.8 +25 44 27	02 47 03.1 +25 44 20	5924	160(8)	0.86	7.7	1.97	82.3	9.14	1
4-513	122897	02 47 10.5 +25 48 43	02 47 09.6 +25 48 25	10479	278(7)	2.85	18.7	2.04	146.6	10.16	1
4-514	122898	02 48 33.0 +24 27 10	02 48 33.9 +24 27 13	5988	132(5)	1.35	13.9	1.88	83.3	9.34	1
4-515	122899	02 48 59.0 +24 22 09	02 49 00.3 +24 22 23	9947	138(4)	1.18	11.4	1.96	139.0	9.73	1
4-516	122900	02 50 29.0 +24 18 46	02 50 27.3 +24 18 34	1391	30(3)	1.00	19.9	1.98	18.5	7.91	1
4-517	2333	02 51 14.1 +25 24 02	02 51 12.5 +25 24 25	7171	311(4)	1.74	9.9	2.23	99.4	9.61	1
4-518	122901	02 51 32.7 +24 09 47	02 51 31.4 +24 09 47	10068	283(14)	2.42	15.2	2.11	140.8	10.05	1
4-519	122902	02 52 03.9 +24 55 10	02 52 02.9 +24 55 19	7161	158(22)	0.99	6.5	2.69	99.3	9.36	1
4-520	122903	02 53 03.7 +25 19 37	02 53 03.2 +25 19 27	6707	70(3)	1.62	22.0	1.95	92.8	9.52	1
4-521	2357	02 53 08.0 +25 29 08	02 53 07.1 +25 29 24	6692	411(8)	2.24	11.2	2.17	92.6	9.66	1
4-522	122135	02 53 36.0 +25 56 36	02 53 35.7 +25 56 44	11157	113(28)	0.93	8.8	2.23	156.4	9.73	1
4-523	122925	02 53 43.6 +24 22 09		-270	22(2)	2.57	26.6	4.38			9 *
4-524	122430	02 54 18.6 +25 56 38	02 54 17.9 +25 56 26	14907	149(11)	1.04	8.3	2.29	210.0	10.03	1
4-525	120593	02 55 13.4 +24 38 01	02 55 12.5 +24 38 07	6983	114(8)	0.66	6.4	2.13	96.8	9.16	1
4-526	122904	02 57 38.7 +24 43 54	02 57 38.5 +24 43 20	10450	310(11)	1.40	7.1	2.50	146.4	9.85	1
4-527	122808	02 57 52.5 +25 57 54	02 57 52.7 +25 58 11	10411	251(32)	1.31	8.8	2.09	145.8	9.82	1

TableA.1 – Continued

Source	AGC #	HI Coords (J2000) hh mm ss.s+dd mm ss	Opt Coords (J2000) hh mm ss.s+dd mm ss	$cz_\odot$ km s^{-1}	$W50$ (ϵ_w) km s^{-1}	F_c Jy km s^{-1}	S/N	rms mJy	Dist Mpc	log M_{HI} $M_\odot$	Code
4-528	120678	02 57 58.9 +25 25 18	02 57 59.1 +25 25 26	6837	127(10)	1.34	13.5	1.97	94.8	9.45	1 *
4-529	120695	02 58 02.0 +25 26 50	02 58 04.1 +25 26 56	6995	125(17)	2.16	21.3	2.02	97.0	9.68	1 *
4-530	120707	02 58 08.1 +24 02 15	02 58 09.8 +24 02 01	10484	362(85)	1.36	6.6	2.43	146.9	9.84	2
4-531	120727	02 58 18.2 +25 27 03	02 58 18.1 +25 26 57	10507	202(24)	2.68	19.8	2.12	147.2	10.14	1
4-532	122905	02 58 35.8 +24 18 29	02 58 37.4 +24 18 34	10133	325(7)	1.65	9.7	2.12	141.9	9.89	1
4-533	2442	02 58 35.9 +25 17 04	02 58 35.6 +25 16 47	10452	173(4)	4.67	40.2	1.97	146.4	10.37	1
4-534	120763	02 58 43.4 +25 23 22	02 58 44.4 +25 23 41	10330	287(15)	2.05	11.7	2.30	144.7	10.00	1 *
4-535	2445	02 58 44.1 +25 45 43	02 58 43.9 +25 45 29	7216	211(16)	2.45	17.8	2.11	100.2	9.76	1
4-536	122906	02 59 03.6 +24 04 56	02 59 05.2 +24 05 01	10233	112(6)	0.86	9.6	1.88	143.3	9.62	1
4-537	122907	02 59 29.9 +24 00 54	02 59 26.1 +24 00 32	10044	221(8)	1.27	8.4	2.27	140.6	9.77	1
4-538	122810	02 59 30.7 +25 53 52	02 59 31.6 +25 54 12	10402	202(10)	1.18	8.8	2.09	145.7	9.77	1
4-539	2455	02 59 43.3 +25 14 20	02 59 42.3 +25 14 14	373	68(2)	54.30	626.6	2.33	7.8	8.89	1
4-540	122908	02 59 55.3 +24 05 09	02 59 55.4 +24 04 53	10133	187(5)	1.50	12.3	1.98	141.9	9.85	1 *
4-541	2457	02 59 55.1 +24 13 28	02 59 54.9 +24 13 31	10216	305(4)	3.82	24.3	2.02	143.1	10.27	1

The comments associated with the sources marked with an asterisk in column 12 of Table A.1 are given here:

4- 1: HVC: small cloud
4- 2: HVC: small cloud
4- 10: HVC: small, isolated cloud
4- 14: HVC: smaller, relatively isolated cloud in a densely populated region.
4- 18: HVC: bright knot of a larger cloud in a densely populated region.
4- 20: HVC: bright knot in an elongated filament in a densely populated region.
4- 23: HVC: bright knot in filament in densely populated region.
4- 28: HVC: more isolated cloud in velocity, but near a densely populated region.
4- 29: HVC: bright peak in a large cloud in a densely populated region.
4- 30: HVC: bright knot in HVC field/filament.
4- 32: HVC: small knot in relatively dense field.
4- 34: HVC: elongated cloud in somewhat dense field.
4- 39: HVC: small, isolated cloud
4- 42: HVC: large, elongated cloud
4- 44: HVC: cloud in dense region
4- 45: cz mismatch with previous detection (Lawrence et al. 1999, MNRAS, 308, 897), but within quoted measurement error
4- 47: HVC: small, more isolated cloud near somewhat dense region
4- 52: poor width precision due to shape of HI profile and poor spectral definition
4- 53: signal merges into strong rfi, severely affecting parameters
4- 55: blend of emission from AGC320702 (224005.8+244156) and AGC320701 (223957.0+244139); not separable spatially or kinematically. Parameters uncertain.
4- 58: HVC: small cloud
4- 59: HVC: small, relatively isolated cloud not far from some others
4- 60: HVC: small, relatively isolated cloud in a field with several others
4- 62: HVC: clumpy/patchy cloud in a region near others
4- 70: HVC: bright knot in elongated filament in densely populated region
4- 71: HVC: bright knot in elongated filament in densely populated region
4- 73: HVC: bright peak in elongated cloud, in densely populated region
4- 75: HVC: bright peak in cloud complex
4- 76: HVC: bright peak in cloud complex
4- 77: HVC: bright knot in cloud complex
4- 78: HVC: bright knot in cloud complex
4- 80: HVC: bright knot in elongated filament, part of a larger cloud complex
4- 86: blend with emission from UGC12290; deblending good, but centroiding and parameters for AGC320603 moderately uncertain.
4- 91: alternative opt.id with 230003.9+253359 (extremely LSB); position poorly determined. Associated with group SRGb016 (Mahdavi et al. 1999).

4- 93: alternative opt.id with 230038.4+244704
4- 99: blend of emission from two nearby galaxies; deblending not possible. Lower-velocity emission is likely from the small nearby galaxy at 230212.7+244549; most of the emission is from the larger source.
4-101: HVC: small cloud
4-103: HVC: small cloud
4-106: double system with partial deblending: detection includes some flux from neighbor, UGC12386. Parameters and centroiding affected.
4-107: double system with partial deblending: detection includes some flux from neighbor, UGC12384. Parameters and centroiding affected.
4-111: HVC: large cloud
4-118: blend with emission from nearby AGC330159 (231529.1+250714); not separable spatially or kinematically. Parameters uncertain.
4-123: blend with emission from AGC330217 (231840.2+251601) to the north and signs of interaction. Parameters and centroiding affected.
4-125: HVC: elongated cloud
4-128: HVC: small clump; somewhat difficult to distinguish from bulk of Galactic HI emission.
4-136: note narrow profile width for object at cz = 9700 km/s
4-141: HVC: small isolated cloud
4-148: blend with emission from AGC331873 (232601.0+254137); signal merges into strong rfi, affecting parameters on high cz end of profile; deblending poor due to spatial and kinematic overlap
4-150: HVC: small clump; somewhat difficult to distinguish from bulk of Galactic HI emission
4-179: near region of poor spatial and spectral coverage; affects centroiding and distance between HI center and opt.id
4-183: poor spectral definition on low-cz end of profile
4-188: opt.id's appear to be an interacting system
4-193: poor spectral definition, with ambiguous opt.id: possible alternative id with 234528.5+241303 and 234526.8+241307; emission may be a blend from all three
4-195: HVC: large cloud; somewhat difficult to distinguish from bulk of Galactic HI emission
4-204: HVC: small cloud
4-207: possible alternative opt.id with 235449.2+245058, but less likely
4-208: blend of AGC331012 with AGC331014 (235548.1+253031) and 331015 (235551.8+252933); no deblending. Parameters uncertain: signal is intrinsically broad, but also merges into rfi and approaches end of ALFALFA bandwidth.
4-209: HVC: isolated large cloud
4-216: HVC: small, isolated knot
4-220: UGC94 nearby at same cz; deblending good, parameters mostly unaffected
4-221: UGC89 nearby at same cz; deblending good, parameters mostly unaf-

fected
4-222: HVC: small, isolated cloud
4-225: possible alternative opt.id with 001034.8+243005, farther from HI center.
4-227: poor spectral definition (high noise)
4-228: significant polarization mismatch
4-234: possible alternative opt.id with 001611.8+245211
4-237: parameters affected by proximity to AGC100146 (001759.7+243345)
4-238: parameters affected by proximity to UGC165
4-241: HVC: bright knot of a larger cloud
4-248: opt.id ambiguous; other possible counterpart at 002258.5+254723 but farther from HI center
4-249: near border of region; detection and parameters will be improved once data is available for +23 degree strip to the South
4-255: HVC: bright knot of a larger cloud extending to the east
4-257: HVC: bright knot
4-260: HVC: bright knot of a much larger complex of clouds
4-261: unresolved broad blend of emission from HCG 001. Centered on UGC248 pair (disturbed optical morphology); blend of flux from group, with uncertain parameters. Not separable spatially or kinematically.
4-268: HVC: bright knot in an extended filament/cloud structure
4-269: HVC: bright knot in a much larger complex of clouds
4-272: HVC: bright knot in a much larger complex of clouds
4-277: HVC: bright knot in a much larger complex of clouds
4-280: possible alternative opt.id with 003237.2+244854, but much less likely
4-281: HVC: one of the many knots in this densely populated region
4-285: HVC: one of the many knots in this densely populated region.
4-286: HVC: one of the many knots in this densely populated region.
4-287: opt.id is distant ($\sim$ 1.6 arcmin), but has previous optical redshift measurement of cz=9838 km/s
4-290: HVC: one of the many knots in this densely populated region.
4-292: signal merges with strong rfi; parameters uncertain
4-294: HVC: one of the many knots in this densely populated region.
4-298: opt.id ambiguous; other possible counterpart at 003828.2+253616
4-308: parameters affected by poor coverage in this region
4-313: possible alternative opt.id with 004433.2+254143
4-317: blend with AGC100511 (004557.3+251308); significantly broadens the profile
4-322: alternative opt.id with 004842.0+243554; flux may be a blend.
4-324: HVC: one of two measured peaks in a large cloud in this densely populated region
4-327: blend of emission from AGC100594 and AGC100596 (005025.6+243114); not separable spatially or kinematically
4-330: HVC: bright peak near a filament that blends into Galactic emission
4-336: blend of emission from AGC101857 and AGC101858 (005456.4+255308);

not separable spatially or kinematically
4-337: HVC: bright peak in a densely populated region
4-339: HVC: isolated, small cloud in a densely populated region
4-342: alternative opt.id with 005629.1+241911
4-343: HVC: larger cloud in a densely populated region.
4-344: HVC: compact cloud in a densely populated region
4-346: HVC: bright peak in a large cloud, distinct from Galactic emission.
4-352: HVC: large cloud of emission, not entirely spectrally distinct from bulk of Galactic HI emission.
4-354: alternative opt.id with 010555.1+243435; emission may be a blend of both
4-355: poor centroiding; galaxy emission plus a low-level emission extended in redshift space. Identified as AGC110025, but some emission is likely to be associated with nearby galaxies at 010619.0+253212.3 and 010617.4+253342.9. Unmeasured broader emission covers 6330 km/s to 6600 km/s, and the measurement of AGC110025 likely contains some contamination from this additional gas.
4-358: HVC: subclump of Wright's Cloud complex
4-359: HVC: subclump of Wright's Cloud complex
4-360: HVC: entirety of the Wright's Cloud clump in +25deg region. Contains many sub-clumps which were measured separately; bulk of the cloud extends into +27deg region (see Saintonge et al. 2008)
4-361: HVC: small clump, part of Wright's Cloud complex
4-362: HVC: possibly an unusual positive velocity cloud. Significant polarization squint; caution as signal may be spurious.
4-364: signal merges with strong rfi; parameters affected and some flux lost
4-365: HVC: clump in an area with several more clouds
4-368: HVC: clump in an area with several more clouds
4-372: interacting triplet: UGC884, AGC113819 (012059.7+253207) and AGC110845 (012052.9+253247). Likely to be at a lower velocity; rfi strongly affects profile and parameters. Actual velocity width of emission closer to 400 km/s.
4-377: HVC: large isolated cloud
4-378: HVC: one of two clumps in a small faint cloud
4-379: HVC: one of two clumps in a small faint cloud
4-380: signal merges with strong rfi; parameters uncertain
4-381: HVC: the brightest of several small distinct clouds in this region
4-384: HVC: one of the small clouds in this region
4-386: HVC: one of the small clouds in this region
4-388: opt.id ambiguous: second possible counterpart at 013744.3+243153
4-389: HVC: cloud not completely spectrally distinct from bulk of Galactic HI emission
4-390: HVC: companion to cloud HI013806.2+245528, not completely spectrally distinct from bulk of Galactic HI emission
4-395: possible alternative opt.id with AGC113308 (014429.3+255114, no previ-

ous cz), but associated with AGC113307 because of match with previous measurement
4-397: opt.id ambiguous; other possible counterpart at 014509.6+245535
4-401: opt.id ambiguous; other possible counterpart at 014619.6+252140
4-412: HVC: cloud not completely spectrally distinct from bulk of Galactic HI emission
4-417: flux much larger than previous measurements (available through our public digital archive site)
4-432: HVC: clumpy, irregular shape, well separated spectrally from galactic HI. Extends about half a degree; may extend into unavailable region to the south at +23deg
4-440: HVC: central and brightest region of a filament that extends for over 2 degrees
4-441: alternative opt.id with 021303.9+252126
4-446: HVC: group of knots
4-447: parameters uncertain; signal may be contaminated by AGC121503 (021610.7+251413). Deviation from 2-horn profile shape and possible signs of interaction
4-448: parameters uncertain; signal may be contaminated by UGC1739. Deviation from 2-horn profile shape and possible signs of interaction
4-449: no discernible optical counterpart
4-450: HVC: one of the many knots in this densely populated region
4-451: HVC: patchy cloud that may extend farther south into unavailable region at +23deg
4-452: galaxy shows clear signs of interaction: optical morphology very irregular, and presence of an extension of the HI, wider than the main emission from the galaxy
4-453: HVC: one of the many knots in this densely populated region
4-456: HVC: one of the many knots in this densely populated region
4-458: HI disk seems elongated toward the NW, low surface brightness extension to one side of the galaxy
4-460: asymmetric shape (both HI and optical image)
4-461: HVC: one knot in one of the filaments that extends to the east of HI022823.7+244309
4-464: HVC: elongated filament
4-473: parameters uncertain because of low S/N
4-474: opt.id ambiguous; other possible counterpart AGC122812 (022606.5+255855)
4-476: possible alternative opt.id with AGC122133 (022646.8+244236), but more than 2 arcmin from HI centroid
4-483: HVC: bright central knot of system of filaments that extend throughout the grid
4-486: opt.id ambiguous; other possible counterpart (or blend of emission) at 022908.9+251808. Parameters uncertain.
4-487: HVC: one knot in one of the filaments that extends to the north of

HI022823.7+244309
4-491: parameters affected by high noise; width determination uncertain
4-499: alternative opt.id twice the distance away at 023528.2+253343
4-503: outer isophotes asymmetrical, distribution appears to stretch to the east.
4-504: possible alternative opt. id with 024124.5+241627, similar galaxy but more than 1 arcmin away
4-506: possible alternative opt. id with 024154.6+245657, but about 1 arcmin farther away from HI center
4-523: HVC: small cloud
4-528: parameters uncertain; signal partly blended with AGC120695 (025804.1+252656). Width could be larger that what is measured here.
4-529: parameters uncertain; signal partly blended with AGC120678 (025759.1+252526)
4-534: HI disk is elongated; appears affected by two nearby neighbors AGC120727 (025818.1+252657) and UGC2442.
4-540: HI emission links this galaxy with UGC2457, 10 arcmin north

APPENDIX B

DETAILED HIMF ESTIMATION METHODS

These details are published as an Appendix to Martin et al. [2010].

B.1 Details of Corrections to the 1/V$_{max}$ Method

B.1.1 Width-Dependent Sensitivity Correction

Giovanelli et al. [2005a] predicted, from the precursor survey observations, that ALFALFA in full two-drift mode could expect an approximate integrated flux detection threshold, S$_{int,th}$ in Jy km s^{-1}, dependent upon profile width as follows:

$$S_{int,th} = \begin{cases} 0.15\, S/N\, (W_{50}/200)^{1/2},\ W_{50} < 200 \\ 0.15\, S/N\, (W_{50}/200),\ \ W_{50} \geq 200 \end{cases} \tag{B.1}$$

In practice, however, ALFALFA outperforms this detection threshold, and we therefore use the data itself to fit a detection limit as described in Section 3.3.3.

The width-dependent sensitivity correction is based on the distribution of observed profile widths. We also assume that the distribution of observed galaxies gives an indication of the true underlying distribution. We are therefore interested in working with as many sample galaxies as possible, and thus we consider a detection threshold S$_{int,th}$ as a function of W$_{50}$ that indicates the limits of ALFALFA's detection ability, rather than a strict completeness limit as in

the 2DSWML case (Section 3.3.4).

The completeness correction is based on the relationship of galaxy mass to the distribution of profile widths W_{50}. It is known that HI profile widths and masses are correlated, and we observe a mass-dependent spread in the distribution of profile width. We determine the profile width distribution as a function of mass by binning $\alpha.40$ galaxies by $\log(M_{HI}/M_{\odot})$ and fitting to each histogram a Gumbel (or Extreme Value Type 1) distribution:

$$f(x) = \frac{1}{\beta}\, e^{\frac{x-\mu}{\beta}}\, e^{-e^{\frac{x-\mu}{\beta}}} \tag{B.2}$$

where μ parametrizes the center of the distribution and β its breadth. The profile width distributions feature narrow central peaks and extended skewed tails, which the Gumbel distribution is designed specifically to model.

We find that the center of the profile width distribution increases linearly with $\log(M_{HI}/M_{\odot})$, and the breadth decreases linearly with $\log(M_{HI}/M_{\odot})$. We derive a relationship between $\log(M_{HI}/M_{\odot})$ and the parameters μ and β, in order to extrapolate to any mass and infer the underlying distribution of W_{50} to which a given galaxy belongs, $P(W_{50}, M_{HI})$. The probability of detecting a galaxy in a given mass bin depends on the profile width distribution for that bin, as well as the limiting profile width $W_{50,lim}$ beyond which that galaxy would not be detectable by ALFALFA. We are seeking a correction factor C that will account for the profile width-integrated flux bias and that satisfies the relationship

$$N_{galaxies}(M_{HI}) = C\, N_{obs}(M_{HI}) \tag{B.3}$$

where $N_{galaxies}$ is the corrected galaxy count to be input for the calculation of the HIMF, and N_{obs} is the observed galaxy count. In terms of the derived distribution $P(W_{50}, M_{HI})$, we have

$$C = \frac{\int_{-\infty}^{+\inf} P(W_{50}, M_{HI})\, dW_{50}}{\int_{-\infty}^{W_{50,lim}} P(W_{50}, M_{HI})\, dW_{50}} \tag{B.4}$$

Since a bin is made up of galaxies with varying $W_{50,lim}$, we apply this correction to each individual galaxy, rather than on a mass bin-by-bin basis. The sum over effective search volume, $\Sigma 1/V_{max}$, therefore becomes $\Sigma C/V_{max}$.

To be conservative, we have included the errors on our derived linear relationships between $\log(M_{HI}/M_{\odot})$ and the Gumbel distribution parameters μ and β in our final error analysis for the HI mass function.

B.1.2 Large Scale Structure Correction

The $1/V_{max}$ method would be biased by large scale structure if we counted galaxies in overdense regions with the same weight as their counterparts in voids. Instead, we want to consider the effective search volume $V_{max,eff}$ in such a way that overdense regions are counted as contributing more effective volume to the overall survey.

We modify $\Sigma 1/V_{max}$ to include weighting by the average density $n(V_{max})$ interior to D_{max}, normalized to the average density of the Universe. The expression for measuring the HIMF then becomes $\Sigma 1/n(V_{max})V_{max}$ [Springob et al., 2005]. We obtain $n(V_{max})$ from the PSCz density reconstruction of Branchini et al. [1999], using their Cartesian map of evenly-spaced grid points out to 240

Mpc h^{-1} smoothed to 3.2 Mpc h^{-1} and using our assumed value h = 0.7. For values $D_{max} \lesssim 85$ Mpc, the average density interior to D_{max} becomes equal to the average density in the PSCz map, so no correction is needed. The large scale structure correction is therefore small compared to the Poisson counting error for galaxies with $\log(M_{HI}/M_{\odot}) > 9.0$, which are found at large distances.

This weighting scheme for galaxy counts in over- and under-abundant regions corrects the relative counts between different environments, so that clusters and superclusters don't dominate the shape of the measured HIMF.

B.2 Details of the 2DSWML Method

In the case of a sample such as α.40, which is not flux-limited and instead depends on additional observables, we must consider a bivariate or two-dimensional stepwise maximum likelihood (2DSWML) approach. In this bivariate case, the likelihood of finding a galaxy with HI mass $M_{HI,i}$ and velocity width $W_{50,i}$ at distance D_i is given by

$$\ell_i = \frac{\phi(M_{HI,i}, W_{50,i})}{\int_{W_{50}=0}^{\infty} \int_{M_{HI}=M_{HI,lim}(D_i,W_{50})}^{\infty} \phi(M_{HI}, W_{50})\, dM_{HI} dW_{50}} \tag{B.5}$$

where $M_{HI,lim}(D_i, W_{50})$ is the minimum detectable mass at distance D_i for a galaxy with velocity width W_{50}, calculated using the completeness relationship in integrated flux-velocity width space as described above.

We proceed by splitting the distribution in bins of $m = \log(M_{HI}/M_{\odot})$ and $w = \log W_{50}$, and assume a constant value within each bin. This leads to

the Two-Dimensional Step Wise Maximum Likelihood (2DSWML) technique, where the parameters of the two-dimensional distribution can now be written as ϕ_{jk} ($j = 1, 2, ..., N_m$ and $k = 1, 2, ..., N_w$). The individual likelihood for each galaxy (Equation B.5) becomes

$$\ell_i = \frac{\sum_j \sum_k V_{ijk}\phi_{jk}}{\sum_j \sum_k H_{ijk}\phi_{jk}\Delta m \Delta w}, \tag{B.6}$$

where the set of coefficients V_{ijk} are used to ensure that only the value for the bin to which galaxy i belongs appears in the numerator and the coefficients H_{ijk} are used to enforce the summation in the denominator to go only over the area in the (m, w) plane where galaxies could be detectable at distance D_i. More precisely,

$$V_{ijk} = \begin{cases} 1 & \text{if galaxy } i \text{ belongs to mass bin } j \text{ and width bin } k \\ 0 & \text{otherwise} \end{cases} \tag{B.7}$$

and, if we denote the completeness function in the (m, w) plane for galaxies at distance D_i by $C_i(m, w)$,

$$H_{ijk} = \frac{1}{\Delta m \Delta w} \int_{w_k^-}^{w_k^+} \int_{m_j^-}^{m_j^+} C_i(m, w)\, dmdw \tag{B.8}$$

where m_j^- and m_j^+ are the HI mass at the lower and upper boundary of mass bin j correspondingly and similarly w_k^- and w_k^+ are the upper and lower boundaries of width bin k. The completeness function in the mass-width plane, $C_i(m, w)$, is directly derived from the $\alpha.40$ sample data, as in Fig. 3.1. For the 2DSWML

method we restrict ourselves to galaxies above a strict completeness cut as a function of W_{50}, where the completeness is 1, excluding 321 galaxies ($\sim 3\%$ of α.40) from the calculation of the mass function.

The goal of the 2DSWML approach is to find the values of the parameters ϕ_{jk} that maximize the joint likelihood of finding all the galaxies in the sample simultaneously, $\mathcal{L} = \prod_i \ell_i$. In practice it is more convenient to maximize the log-likelihood, which using Equation B.6, can be written as

$$\begin{aligned} \ln \mathcal{L} = \sum_i \ln \ell_i = \sum_i \sum_j \sum_k V_{ijk} \ln(\phi_{jk} \Delta m \Delta w) \\ - \sum_i \ln \left(\sum_j \sum_k H_{ijk} \phi_{jk} \Delta m \Delta w \right) + \text{const.} \end{aligned} \tag{B.9}$$

$\ln \mathcal{L}$ is maximized by setting the partial derivatives with respect to each of the parameters equal to zero, giving

$$\phi_{jk} = \frac{\sum_i V_{ijk}}{\sum_i \frac{H_{ijk}}{\sum_m \sum_n H_{imn} \phi_{mn}}} = \frac{n_{jk}}{\sum_i \frac{H_{ijk}}{\sum_m \sum_n H_{imn} \phi_{mn}}} \tag{B.10}$$

where n_{jk} is the galaxy count in bin j, k. The Maximum Likelihood values for each parameter can be found by iterating Equation B.10 until a stable solution is obtained. Finally, the HI mass distribution can be derived by the bivariate HI mass-velocity width distribution by marginalizing over velocity width, or

$$\phi_j = \sum_k \phi_{jk} \, \Delta w. \tag{B.11}$$

Marginalizing the bivariate distribution over HI mass leads, instead, to the projected velocity width function for HI bearing galaxies, which will be the focus of a forthcoming publication.

As Equations B.5 and B.6 imply, the overall normalization is lost in the process, and only the relative values of the parameters ϕ_{jk} are meaningful. Fixing the amplitude gives the HI mass function.

B.2.1 HIMF Amplitude

To transform the calculated probability density function into an HI mass function (e.g. transform the unitless $\{\phi_k \Delta m\}$ into space densities) we evaluate the amplitude of the HIMF by matching the integral of the distribution to the inferred average density of galaxies in the survey volume $\bar{n}$, as in Zwaan et al. [2003]. Davis and Huchra [1982] discuss various estimators for $\bar{n}$ that strike different balances between stability against poor knowledge of the selection function of the survey and immunity to large-scale structure. Since we believe we have a good understanding of the selection function out to $cz = 15,000$ km s^{-1}, we choose to adopt the estimator that is least prone to bias, denoted by n_1, defined as

$$n_1 = V_{survey}^{-1} \int \frac{n(D)\, dD}{S(D)} \tag{B.12}$$

where $n(D)\, dD$ is the number of galaxies in a spherical shell of thickness dD and radius D, and V_{survey} is the total survey volume. The selection function $S(D)$ is the fraction of galaxies detectable at distance D and is given by

$$S(D) = \frac{\int_{w_{min}}^{w_{max}} \int_{m_{lim}(w,D)}^{m_{max}} \phi(m,w)\, dm\, dw}{\int\int \phi(m,w)\, dm\, dw}. \tag{B.13}$$

In the case of the 2DSWML method we evaluate n_1 by the expression

$$n_1 = V_{survey}^{-1} \sum_i \frac{1}{\sum_j \sum_k H_{ijk} \phi_{jk} \Delta m \Delta w}. \tag{B.14}$$

Equation B.14 corresponds to weighing each detected galaxy in the survey by the inverse of the selection function at the galaxy's distance, effectively correcting each detection by the fraction of galaxies that cannot be detected at distance D_i.

APPENDIX C

VATT IMAGE REDUCTION FOR THE ALFALFA METALLICITIES PROJECT

C.1 Introduction

A primary strength of the ALFALFA survey is its ability to detect dwarf galaxies in the local volume, including those with extremely low surface brightnesses which are missed by optical surveys like the SDSS. The ALFALFA metallicities project has identified a sample of low-mass (log $M_{HI}/M_{\odot} < 7.5$) galaxies with narrow velocity widths ($W_{50} \leq 80$ km s^{-1}) that are typically very low surface brightness but nevertheless gas-rich. This unique sample may be used to probe the star formation properties of dwarf galaxies which retain their baryons through cosmic time. Star formation in dwarf galaxies is poorly understood; the Grebel [2005] survey of Local Group dwarfs found a variety of different star formation histories for dwarfs that were otherwise morphologically similar. Of particular interest is the metallicity-luminosity relationship for the lowest-mass gas-rich galaxies, and whether this relationship is a function of a galaxy's environment.

As the ALFALFA catalog expands, so too will our ability to identify samples of gas-rich, low-luminosity galaxies for optical followup. The ALFALFA survey can contribute to this body of knowledge by providing neutral hydrogen information in addition to color and morphology from the SDSS, while the ALFALFA metallicities project adds nebular abundance measurements of metallicities.

In the Spring of 2009, the ALFALFA metallicities project was granted time on

the Palomar 5m telescope to observe nebular abundances for a set of ALFALFA-identified galaxies with low HI masses and narrow HI line widths within 11 Mpc. These galaxies are also being targeted for GALEX observations, giving a fuller picture of the star formation history of these dwarfs and the origin of the metallicity-luminosity relation. Spectroscopy of nebular abundances requires accurate astrometry of Hα knots in the galaxy of interest, so members of the ALFALFA collaboration provide R-band and Hα filtered images of galaxies well in advance of Palomar observations. Before arriving at Palomar, these images are reduced, Hα knots are identified, and finding charts based on nearby bright stars are prepared. For the Spring 2009 Palomar run, Hα images were provided by collaborators with access to WISE (Noah Brosch and Oded Spector) and SMARTS (Rebecca Koopmann), telescopes with which we are accustomed to working. For the first time, however, 5 nights of telescope time at the Vatican Advanced Technology Telescope (VATT) were allotted to Aileen O'Donoghue (St. Lawrence University) and Christopher Corbally (Vice Director of the Vatican Observatory), who successfully observed 25 galaxies identified from ALFALFA. This Appendix details the reduction methods applied to the images from the VATT.

C.2 VATT Observations and Palomar Abundance Measurements

The Vatican Advanced Technology Telescope (VATT) is located at the Mount Graham International Observatory in Arizona[1]. The imaging camera is the

[1]http://vaticanobservatory.net/VATT.html

VATT 4k, with a pixel scale of 0.375 arcseconds per pixel, a gain of 1.8 e^-/DN, and a read noise of 3.5 e^-. A set of Hα filters are available with FWHM $\sim$ 70 Å, spaced 50 Å apart. For the extremely nearby galaxies selected here, we needed only to use the Hα 658 filter, capable of detecting galaxies out to a redshift $\sim$ 4,000 km s^{-1}($\sim$ 55 Mpc), well beyond the 11 Mpc limit that defines the sample proposed for observation at Palomar.

The observing run took place over five nights, from March 18 to March 22, 2009, resulting in images (in R-band and Hα) of 25 galaxies for use in the ALFALFA metallicities project. The images are not of photometric quality, since calibration images were not available, but the weather was good for all but the last night and the best seeing achieved was $\sim$ 2 arcsec. Without photometry, the VATT images cannot be used to derive star formation rates, and are therefore only used to identify slit positions for nebular abundance observations at Palomar.

The observing run at the Palomar 5m telescope took place over four nights, from March 26 to March 29, 2009 during moon-dark time. Targets were selected to fill the observing time from about 0800 LMST to 1700 LMST while keeping the airmass for each target low. This limitation led to the addition of some 'late-night' ALFALFA sources, which didn't fall into the original proposal; time at the VATT was used to obtain Hα images of several of these objects. From the galaxies in the original proposal to Palomar, those which were actually observed were selected to have a good chance of a successful nebular abundance measurement. Those galaxies which had been targeted for spectroscopy in the SDSS were checked; if the spectrum did not contain Hα, these galaxies were rejected. Other criteria included visible patchiness and brighter blue regions where active

star formation was suspected. The final criterion applied was our ability to obtain Hα maps of the galaxy before the start of the observations at Palomar. The Palomar single-slit (2 arcsec) observations used the D55 dichroic and the same observational setup used for the PhD theses of Stierwalt [2010] and Saintonge [2007].

Of the 25 galaxies observed at the VATT, only two (AGC252519, AGC252211) were used in the Spring 2009 Palomar observing run, since the last two nights at Palomar were cloudy. The others have all been reduced and are available for the ALFALFA metallicities project in the future. Of the remaining 23 galaxies, Hα knots were found in 17. These images, and the corresponding identification of star forming regions, can be used for future ALFALFA followup.

C.3 Reduction of VATT R-band and Hα Images

The VATT 4k chip is a mosaicing imager, which effectively records two images when an object is observed. Data is recorded as a multi-extension FITS file, a format which allows the storage of data from each of many mosaic elements. In the case of the VATT 4k, there are two extensions for each FITS file, and each extension (called, in this case, 'im1' and 'im2') contains the image and header from one half of the chip.

Since these images must be processed for use in the ALFALFA Hα projects, there are two options for dealing with the multi-extension fits files. First, we can proceed by editing and processing each image in two halves, treating each 'im1' extension as a separate image from the corresponding 'im2' extension. In this case, we would apply the corrections of trim, overscan, zero biasing, and flat-

fielding separately to each of the extensions, and merge the two images in the final stages. In principle, this approach is possible, but in practice it is messy and is not preferred. Instead, we choose to trim and overscan correct each extension, and then merge them together into a single FITS file using the CCDPROC module in the MSCRED package. The MSCRED package was originally designed for use with the NOAO Mosaic Camera, which consists of 8 CCDs, and the software in this package is designed for the processing and reduction of mosaiced images.

C.3.1 Trim, Overscan and Merge

The first stage of processing is specific to the CCD used. VATT 4k trim and overscan parameters were provided by Richard Boyle (S.J.), VATT Telescope Scientist. Setting these parameters in the CCDPROC package will apply an overscan bias correction, then trim and merge the two extensions into a single FITS file. The bias or overscan region of the chip is around the edge of the CCD, and these rows and/or columns are not actually exposed when science or calibration images are taken. Overscan regions are used to correct images for dark current, with the disadvantage that they don't cover the entire CCD to provide full two-dimensional information. A median value can be subtracted from the entire chip, but using overscan regions as opposed to full dark images will miss dead or bright pixels. Some observers therefore take separate dark images for the same exposure time as a science image. Other conditions should be similar between dark and science images, including temperature which can affect the behavior of the CCD. On the other hand, corrections from overscan regions automatically take place at the same time as the science images they are used to

correct, so that they reflect any time variability or change in condition, a key advantage over dark images. In this case, we are not concerned with absolute photometric calibration, so dark frames are not strictly necessary and the overscan regions will do to remove the dark current. Therefore, here we set the CCDPROC parameter to apply an overscan correction using the bias section. For the VATT4k, the bias section (set in parameter 'biassec') is [1020:1035, 9:2024]. The overscan bias correction can be applied in more than one way. In the first, a different bias value is found for each line from the bias pixels in that line. In the second, which we use, overscan columns are averaged together, and a smooth Legendre polynomial function is applied to correct the image at each line.

A trim is also applied to each of the image extensions separately; the trim section (set in parameter 'trimsec') is [9:1016,9:2024]. Finally, the parameter 'merge' is set to 'yes,' so that CCDPROC will merge the amplifiers from the same CCD and all of the FITS extensions along with their headers will be merged into a single-extension FITS file.

Once CCDPROC is run, the resulting VATT images will be 2016 by 2016 pixels and will be merged into a single amplifier. However, the detector section (DETSEC) header parameter will still read [9:1016, 9:2024] as if the image were still a half-square mosaic. Any packages which use this header keyword will therefore read the image as being much smaller than it truly is. This must be changed so that the MSCRED package can be used later on to create finding charts. Using HEDIT and setting the DETSEC to [9:2024, 9:2024] will correct this potential problem.

A subsets file should also be created and placed in the IRAF $home directory, to translate the FILTER keyword in the header from the VATT system to the

IRAF filters (i.e. R and H_ALPHA). The filter 'TOP 4 BOT 1' corresponds to H_ALPHA, and 'TOP 3 BOT 1' corresponds to R for this run (however, this could change in the future, and these parameters should be checked from the image headers and the observing logs). The subsets file should be formatted as follows:

'TOP 2 BOT 1' B

'TOP 4 BOT 1' H_ALPHA

'TOP 3 BOT 1' R

For the MSCRED package to read and translate the file, the parameter ssfile should be set via

cl > mscred.ssfile='home$subsets'

C.3.2 Zero Biasing and Flatfielding

Now that the images have been trimmed and merged, the MSCRED package does not need to be used anymore, and the modules in the NOAO IMRED package should be used instead. Because similarly-named modules are in both the MSCRED and IMRED packages (such as CCDPROC), it is important to close out the MSCRED package before continuing.

In the next stage, all of the zero images should be combined using the ZEROCOMBINE package, and then applied to all of the images using CCDPROC. On the first night of VATT observations, no zero fields were taken, so those from the second night were used without any apparent problems.

Typically, the process of flatfielding is relatively straightforward. Dome flats

from each filter are combined with FLATCOMBINE and then science images in each filter are separately corrected via CCDPROC. For the entire VATT run, however, there was a problem with the R-band dome flats. These were only exposed for 0.7 second, which was enough time for high counts (about 35,000, where the chip saturates at 60,000) but not enough time to account for the shutter effect. Because the CCD shutter takes time to open and to close, a short exposure time will have a 'ghost' image of the shutter. 0.7 seconds is not enough time to integrate this effect out. Thus, the R-band 'flat' images are not truly flat, and have a good deal of structure due to the shutter itself. Figure C.1 shows an example frame of an R-band domeflat from this run, with the 'pinwheel' shape of the shutter's signature clearly visible. For contrast, see Figure C.2, another sample frame but this time in Hα; these domeflats lack the structure seen in the R-band. The Hα filtered flats were exposed for 30 seconds, so they do have high counts as well as uniform, flat illumination. Figures C.1 and C.2 also demonstrate the different structure on the left and right sides of the VATT 4k chip, with the left-hand side of each image appearing brighter than the right-hand side.

Several approaches were considered for dealing with the shutter problem in the R-band domeflat frames. First, we could have used the Hα dome flats on images in both R band and Hα. These flats are well-illuminated, but the inclusion of the filter changes the optical path and could introduce structure that does not exist in the R-band science images. Secondly, we could have used the few twilight flats for an illumination correction, but the twilight flats have very low counts (10 - 12 thousand) and there are only two of them (too few for a median subtraction). The final option, and that which was ultimately used, was to use the bad R-band dome flats and work around the structure introduced by the shutter. This approach was used quite successfully. The bad flats have little

structure in the center of the image, where the galaxies were typically located on the chip. Since no photometry was to be used, this approach served well enough for the identification of Hα knots. Furthermore, the Hα images themselves are very well-corrected by the flat fields in that band, and these are the images that contain the most important information, so extraordinary efforts to correct the R-band images would result in very little gain.

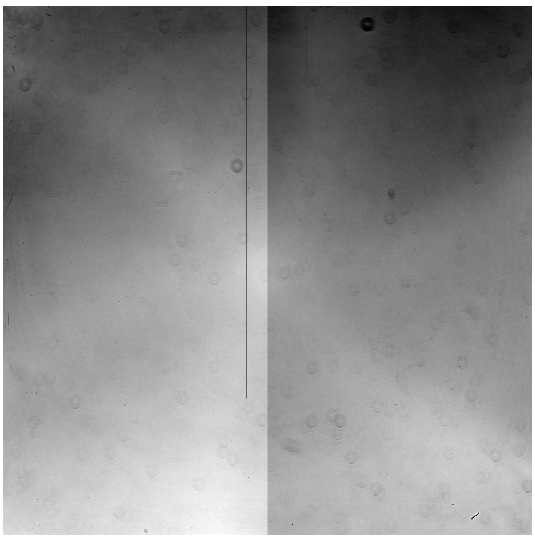

Figure C.1 R-band dome flat image from the VATT.

C.3.3 Smoothing and Scaling

Cosmic rays were automatically subtracted using COSMICRAYS (in the CRUTIL package) and residuals were cleaned up using IMEDIT. The VATT images were very large compared to the galaxies, so IMEDIT was used primarily to rid the area immediately surrounding the galaxy of cosmic rays. Particularly large or egregious problems in the rest of the image were also corrected.

Next, science images must be separated by galaxy so that all images can be

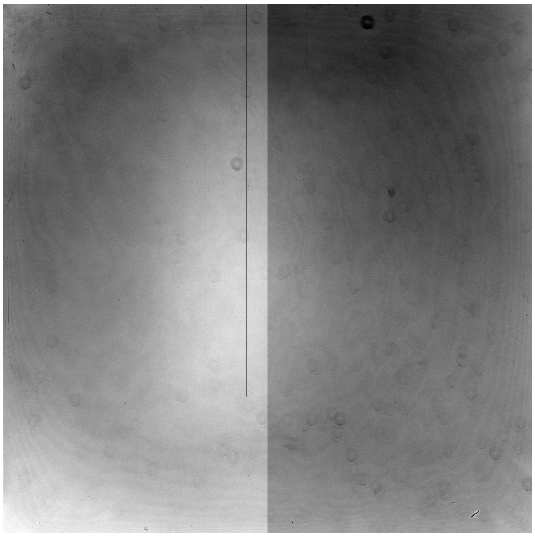

Figure C.2 Hα-band dome flat image from the VATT.

smoothed and trimmed together. All galaxies on the VATT run had a single Hα image (with exposure times of either 1500 or 1800 seconds) and a single R band image (with an exposure time of 600 seconds). Because there can be substantial shifts between the R and Hα image, GETSHFTS.cl and DOALIGN.cl (provided by collaborator J. Salzer) line the two images up based on a selection of reference stars.

If there is a substantial difference between the full width at half maximum (FWHM) of a point object (i.e. star) in the two galaxy images, they must be smoothed to match the poorest-quality image. J. Salzer's program FINDFWHM determines the FWHM of stars in each (R and Hα) image, in pixels. In cases where the FWHM of the two images are not the same, the higher-quality image should be smoothed using GAUSS (in the IMFILTER package), and the resulting image should be checked again.

Finally, the counts of each R-band image were scaled to the counts of their respective Hα image using another module provided by J. Salzer, called

GETSCALE. As inputs, this procedure requires the image scale in arcseconds per pixel (0.375 for the VATT), the FWHM of the PSF obtained from the images and FINDFWHM, and the exposure time of the Hα image. GETSCALE uses bright but unsaturated stars to correct the scaling between the images. Because of the structure in the R-band images due to the shutter effect in the dome flats, it is wise to use stars as close to the galaxy of interest as possible while scaling. This region is the most important and the shutter effect makes it impossible to obtain a simple scaling relationship between the Hα and R-band image.

Finally, any Hα knots were identified by subtracting the 'off' (R-band) image from the 'on' (Hα) image, to obtain a continuum-subtracted Hα image. The VATT images are flipped North-South, which was corrected as the final step.

Figure C.3 shows the results of the previous steps for a particular galaxy observed at the VATT, AGC252211. The top and middle panels show the R-band and Hα images, respectively, after overscan, trim, zero-biasing, flatfielding, scaling, and flipping have been completed. The bottom panel shows the final result of image reduction: a continuum-subtracted Hα image. In this case, there are several Hα knots clearly visible which were targeted for a nebular abundance measurement at Palomar.

C.3.4 Finding Charts

The procedure for making finding charts from the continuum-subtracted Hα image was determined by Amelie Saintonge, and the same procedure can be used for any galaxy image, whether from the VATT, WISE, WIYN, or SMARTS. Many of the galaxies observed at Palomar for the ALFALFA metallicities project

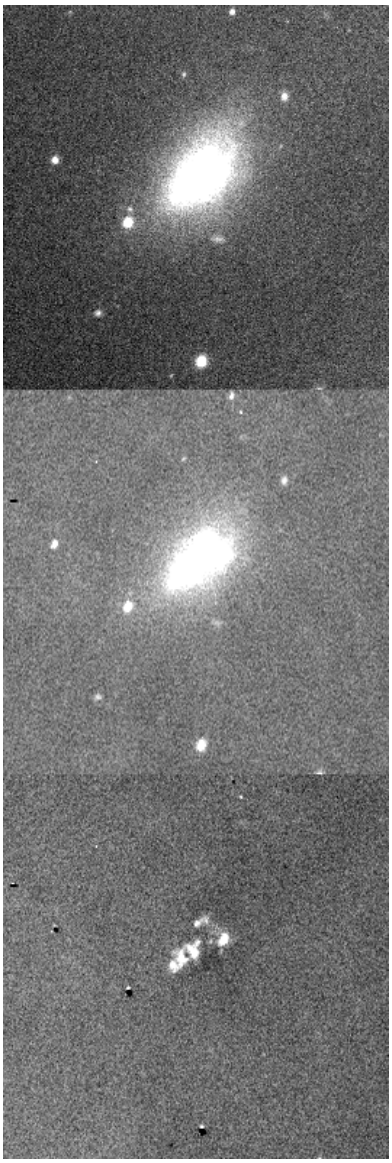

Figure C.3 Top panel: R-band (continuum) image of galaxy AGC252211, taken at the VATT on 19 March 2009. Image has been trimmed and zoomed to show only the galaxy of interest. Middle panel: Hα image of same. Bottom panel: continuum-subtracted Hα image of AGC252211, with bright HII regions clearly identifiable. All panels: image is shown after zero-biasing, flatfielding, scaling, and flipping.

are extremely faint and have low surface brightnesses. It is not a simple task, therefore, to point the telescope to the proper coordinates and assess whether or not the galaxy has fallen in the slit. Instead, we prepare a finding chart before observations at Palomar. During the observations, we simply point the telescope to a bright, easy-to-find star within roughly 2 arcmin of the galaxy, and then off-set the telescope by the predetermined amount.

To prepare a finding chart, procedures in the MSCRED package are used to overlay a stellar catalog on the image (MSCGETCATALOG and MSCTVMARK), match stars in the image with those in the catalog to properly set the world coordinate system in the image header (MSCCMATCH), and obtain pixel coordinates of both the HII regions and the reference stars which will be used for pointing.

C.4 Conclusions and Palomar Observations

While the images recorded at the VATT are not of photometric quality, we were able to obtain positions of HII regions from the majority of them (19 of 25) for use in the ALFALFA metallicities project. Two of the galaxies, AGC252211 and AGC252519, were successfully observed at Palomar for the run immediately following the VATT observations. The VATT galaxies were specifically selected to fill in the later part of the Palomar observing night. Since the final two nights of the Palomar run were clouded out, most of the galaxies presented here were not observed. The results of the nebular abundance measurements made at Palomar in Spring 2009 are detailed in Stierwalt [2010]. As an example of the utility of the VATT images to the ALFALFA metallicities project, Figure C.4 displays

the reduced spectra derived by Sabrina Stierwalt for the three individual knots within AGC252211.

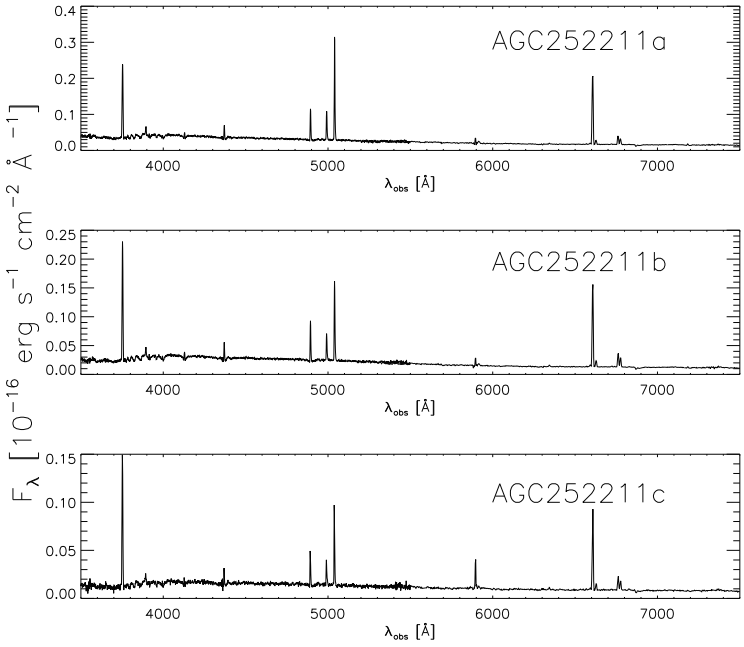

Figure C.4 Spectrum obtained for AGC252211 during Spring 2009 nebular abundances observing run on the Palomar 5m telescope.

Table C.1 contains relevant information about each of the galaxies targeted for at the VATT in Spring 2009, including whether Hα knots were successfully detected in the galaxy. Comments in the fourth column, "Galaxy Notes", give a sense of our expectation of success. In some cases, poor image quality affected the results. Generally, those galaxies in which Hα emission was not detected were very faint or smooth, so that the nondetection is not surprising.

Table C.1. VATT Spring 2009 ALFALFA Galaxy Targets

AGC Nr	Other name	J2000 Coords	Galaxy Notes	Target [a]	Hα Detection?[b]
3672		070627.5+301919	Irregular with large, apparent clumps		Yes; strong, several large clumps
3775	057-006	071552.5+120654	dSp with apparent clumps	†	Yes; several knots
181471	KK 67	080324.6+150828	Irregular, faint and patchy LSBG	†	Yes; faint, single knot
191791	D634-03	090853.8+143502	Irregular, very faint and patchy LSBG	* ‡	No; clouded out
204464		100425.1+023331	Irregular, faint and patchy disk		Yes; strong, single knot
201051	KDG068	101432.0+032159	Irregular, very faint LSBG		No
204288		101451.5+033853	Smooth stellar disk without apparent clumps		No
5539	036-075	101555.1+024109	Bright Irregular with apparent clumps		Yes; strong, several clumps
205165		103704.8+152015	Irregular, bright disk with large clumps	*	Yes; strong, single knot
200603	066-029	104917.1+122520	Smooth stellar disk with apparent clumps	†	Yes; strong, several knots
219303		112313.5+134254	Extreme LSBG [c]		No; poor im. quality, not detected in R-band
215296		112655.2+145003	Irregular, faint with smooth distribution		Yes; strong, single knot
219309		114219.5+262726	Irregular, patchy LSBG		No
8638	131-018	133919.7+244630	Irregular, bright and patchy with large clumps	*	Yes; very strong, large clump
238847		134509.7+272011	Irregular, faint with apparent clumps		Yes; extremely faint and small, single knot
233681		134716.0+131038	Irregular, very faint and patchy LSBG		Yes; faint, single knot
231980	KKH86	135433.5+041440	Irregular, very faint and patchy LSBG	* ‡	No
9540	KKR13	144852.6+344232	V. Irregular faint triple system		Yes; very strong, several knots
252211	MTT-113	150350.2+005841	Irregular, bright disk with large clumps		Yes; very strong, large clump
252519	KKR15	150355.9+002544	Irregular, faint with extremely bright clumps		Yes; very strong, large clump
251419	MTT-124	150408.3+013128	Bright Spiral with apparent clumps		Yes; strong, single knot
252398	MTT-162	150528.7+011733	Irregular, faint and patchy with apparent clumps		Yes; faint, single knot
252399	MTT-192	150611.3+020546	Bright edge-on spiral		Yes; strong, several clumps

The observing team (O'Donoghue and Corbally) also targeted three special objects, with limited success; information on these objects is contained in Table C.2. The first, NGC 3169 Tail, may contain a detection of star formation in a tidal dwarf galaxy within the tidal tail of the NGC 3169/3166/3165 group. This tidal dwarf candidate was observed at Palomar by Martin and Stierwalt during an observing run in Spring 2011 (March 28, 29 and 30). Analysis of star formation and of the metallicity in this tidal dwarf will be the focus of a forthcoming work.

The second special object, HCG 059, was added by the VATT observers, but is at a higher redshift ($\sim$4,000 km s^{-1}), so any Hα emission would have been lost to the filter selected for the ALFALFA galaxies at redshifts less than 3,000 km s^{-1}. The third object, NGC 5953 Tail, was another attempt to detect a tidal dwarf galaxy in a known HI bridge. This observation failed due to poor seeing and image quality, leading to poor continuum subtraction, and would therefore not be worth later followup at a large telescope. However, there is a hint of a serendipitous detection of Hα in AGC/NGC 9902 to the south of the NGC 5953/54 system, and a better quality image may result in a tidal dwarf galaxy in the targeted system itself, so this observation may be worth repeating at a later date.

Table C.1 (cont'd)

AGC Nr	Other name	J2000 Coords	Galaxy Notes	Target [a]	Hα Detection?[b]
250103	021-013	150805.8+013905	Bright disk with apparent clumps		Yes; strong, single knot
252250		151043.1+033018	Irregular with evident clump		Maybe; marginal quality image

[a] Objects denoted with a * in this column are targets of the ALFALFA Metallicities Project. Where noted with ‡, they will also be targeted by 11HUGS. There is some overlap between the ALFALFA project and 11HUGS for calibration purposes. Objects denoted with a †in this column are ALFALFA GALEX Targets.

[b] Notes expand on the quality of the detection.

[c] May be associated with the Leo Triplet HI plume (Stierwalt 2009, ALFALFA 9)

Table C.2. VATT Spring 2009 Special Objects

Object Name	J2000 Coords	Object Notes	Hα Detection?
NGC 3169 Tail	101516.0+033304	HI tail in NGC3169/3166/3165 interacting group [a]	Possible; 101445.6+033859[b]
HCG 059	114826.6+124240	Hickson Compact Group; several large, bright galaxies	No; wrong Hα filter (higher redshift than ALFALFA targets)
NGC 5953 Tail	153430.0+151300	HI tail in NGC5953/54; searching for star formation indicating TDG	No; extremely poor image quality

[a] Targeted in order to search for star formation indicating a Tidal Dwarf Galaxy Candidate

[b] There is a possible Hα knot in the image at the listed coordinates. There is only a faint corresponding signal in the R-band image.

BIBLIOGRAPHY

K. N. Abazajian, J. K. Adelman-McCarthy, M. A. Agüeros, S. S. Allam, C. Allende Prieto, D. An, K. S. J. Anderson, S. F. Anderson, J. Annis, N. A. Bahcall, and et al. The Seventh Data Release of the Sloan Digital Sky Survey. *ApJS*, 182:543–558, June 2009. doi: 10.1088/0067-0049/182/2/543.

F. B. Abdalla, C. Blake, and S. Rawlings. Forecasts for dark energy measurements with future HI surveys. *MNRAS*, 401:743–758, January 2010. doi: 10.1111/j.1365-2966.2009.15704.x.

I. K. Baldry, K. Glazebrook, J. Brinkmann, Ž. Ivezić, R. H. Lupton, R. C. Nichol, and A. S. Szalay. Quantifying the Bimodal Color-Magnitude Distribution of Galaxies. *ApJ*, 600:681–694, January 2004. doi: 10.1086/380092.

I. K. Baldry, K. Glazebrook, and S. P. Driver. On the galaxy stellar mass function, the mass-metallicity relation and the implied baryonic mass function. *MNRAS*, 388:945–959, August 2008. doi: 10.1111/j.1365-2966.2008.13348.x.

R. Barkana and A. Loeb. Detecting the Earliest Galaxies through Two New Sources of 21 Centimeter Fluctuations. *ApJ*, 626:1–11, June 2005a. doi: 10.1086/429954.

R. Barkana and A. Loeb. Probing the epoch of early baryonic infall through 21-cm fluctuations. *MNRAS*, 363:L36–L40, October 2005b. doi: 10.1111/j.1745-3933.2005.00079.x.

D. G. Barnes, L. Staveley-Smith, W. J. G. de Blok, T. Oosterloo, I. M. Stewart, A. E. Wright, G. D. Banks, R. Bhathal, P. J. Boyce, M. R. Calabretta, M. J. Disney, M. J. Drinkwater, R. D. Ekers, K. C. Freeman, B. K. Gibson, A. J. Green, R. F. Haynes, P. te Lintel Hekkert, P. A. Henning, H. Jerjen, S. Juraszek, M. J. Kesteven, V. A. Kilborn, P. M. Knezek, B. Koribalski, R. C. Kraan-Korteweg, D. F. Malin, M. Marquarding, R. F. Minchin, J. R. Mould, R. M. Price, M. E. Putman, S. D. Ryder, E. M. Sadler, A. Schröder, F. Stootman, R. L. Webster, W. E. Wilson, and T. Ye. The Hi Parkes All Sky Survey: southern observations, calibration and robust imaging. *MNRAS*, 322:486–498, April 2001. doi: 10.1046/j.1365-8711.2001.04102.x.

S. Basilakos, M. Plionis, K. Kovač, and N. Voglis. Large-scale structure in the HI Parkes All-Sky Survey: filling the voids with HI galaxies? *MNRAS*, 378: 301–308, June 2007. doi: 10.1111/j.1365-2966.2007.11781.x.

C. M. Baugh, C. G. Lacey, C. S. Frenk, A. J. Benson, S. Cole, G. L. Granato, L. Silva, and A. Bressan. Predictions for the SKA from hierarchical galaxy formation models. *NewAR*, 48:1239–1246, December 2004. doi: 10.1016/j.newar.2004.09.008.

M. R. Blanton. Galaxies in SDSS and DEEP2: A Quiet Life on the Blue Sequence? *ApJ*, 648:268–280, September 2006. doi: 10.1086/505628.

M. R. Blanton and J. Moustakas. Physical Properties and Environments of Nearby Galaxies. *ARA&A*, 47:159–210, September 2009. doi: 10.1146/annurev-astro-082708-101734.

M. R. Blanton, D. W. Hogg, N. A. Bahcall, I. K. Baldry, J. Brinkmann, I. Csabai, D. Eisenstein, M. Fukugita, J. E. Gunn, Ž. Ivezić, D. Q. Lamb, R. H. Lupton, J. Loveday, J. A. Munn, R. C. Nichol, S. Okamura, D. J. Schlegel, K. Shimasaku, M. A. Strauss, M. S. Vogeley, and D. H. Weinberg. The Broadband Optical Properties of Galaxies with Redshifts $0.02 < z < 0.22$. *ApJ*, 594:186–207, September 2003. doi: 10.1086/375528.

M. R. Blanton, D. Eisenstein, D. W. Hogg, D. J. Schlegel, and J. Brinkmann. Relationship between Environment and the Broadband Optical Properties of Galaxies in the Sloan Digital Sky Survey. *ApJ*, 629:143–157, August 2005. doi: 10.1086/422897.

R. G. Bower, A. J. Benson, R. Malbon, J. C. Helly, C. S. Frenk, C. M. Baugh, S. Cole, and C. G. Lacey. Breaking the hierarchy of galaxy formation. *MNRAS*, 370:645–655, August 2006. doi: 10.1111/j.1365-2966.2006.10519.x.

M. Boylan-Kolchin, V. Springel, S. D. M. White, A. Jenkins, and G. Lemson. Resolving cosmic structure formation with the Millennium-II Simulation. *MNRAS*, 398:1150–1164, September 2009. doi: 10.1111/j.1365-2966.2009.15191.x.

E. Branchini, L. Teodoro, C. S. Frenk, I. Schmoldt, G. Efstathiou, S. D. M. White, W. Saunders, W. Sutherland, M. Rowan-Robinson, O. Keeble, H. Tadros, S. Maddox, and S. Oliver. A non-parametric model for the cosmic velocity field. *MNRAS*, 308:1–28, September 1999. doi: 10.1046/j.1365-8711.1999.02514.x.

R. Braun and D. A. Thilker. The WSRT wide-field H I survey. II. Local Group features. *A&A*, 417:421–435, April 2004. doi: 10.1051/0004-6361:20034423.

F. H. Briggs and S. Rao. The Shape of the H i Mass Function for Late-Type Galaxies over the Range M_{HI} 10^7 to 10^{10} M_{Sun}. *ApJ*, 417:494–+, November 1993. doi: 10.1086/173328.

J. S. Bullock, K. R. Stewart, M. Kaplinghat, E. J. Tollerud, and J. Wolf. Stealth Galaxies in the Halo of the Milky Way. *ApJ*, 717:1043–1053, July 2010. doi: 10.1088/0004-637X/717/2/1043.

D. Burstein and C. Heiles. Reddenings derived from H I and galaxy counts - Accuracy and maps. *AJ*, 87:1165–1189, August 1982. doi: 10.1086/113199.

B. Catinella, M. P. Haynes, R. Giovanelli, J. P. Gardner, and A. J. Connolly. A Pilot Survey of H I in Field Galaxies at Redshift z ˜ 0.2. *ApJL*, 685:L13–L17, September 2008. doi: 10.1086/592328.

D. Ceverino and A. Klypin. The Role of Stellar Feedback in the Formation of Galaxies. *ApJ*, 695:292–309, April 2009. doi: 10.1088/0004-637X/695/1/292.

T.-C. Chang, U.-L. Pen, K. Bandura, and J. B. Peterson. An intensity map of hydrogen 21-cm emission at redshift z$\sim$0.8. *Nature*, 466:463–465, July 2010. doi: 10.1038/nature09187.

J. Cho and C. Park. Internal Extinction in the Sloan Digital Sky Survey Late-Type Galaxies. *ApJ*, 693:1045–1055, March 2009. doi: 10.1088/0004-637X/693/2/1045.

A. Chung, J. H. van Gorkom, J. Kenney, H. Crowl, B. Vollmer, and D. Schiminovich. The Evolution of H I Disks in the Virgo Cluster. In J. G. Funes & E. M. Corsini, editor, *Formation and Evolution of Galaxy Disks*, volume 396 of *Astronomical Society of the Pacific Conference Series*, pages 127–+, October 2008.

A. Chung, J. H. van Gorkom, J. D. P. Kenney, H. Crowl, and B. Vollmer. VLA Imaging of Virgo Spirals in Atomic Gas (VIVA). I. The Atlas and the H I Properties. *AJ*, 138:1741–1816, December 2009. doi: 10.1088/0004-6256/138/6/1741.

J. J. Condon, W. D. Cotton, E. W. Greisen, Q. F. Yin, R. A. Perley, G. B. Taylor, and J. J. Broderick. The NRAO VLA Sky Survey. *AJ*, 115:1693–1716, May 1998. doi: 10.1086/300337.

H. H. Crowl and J. D. P. Kenney. The Stellar Populations of Stripped Spiral Galaxies in the Virgo Cluster. *AJ*, 136:1623–1644, October 2008. doi: 10.1088/0004-6256/136/4/1623.

J. Davies, R. Minchin, S. Sabatini, W. van Driel, M. Baes, P. Boyce, W. J. G. de Blok, M. Disney, R. Evans, V. Kilborn, R. Lang, S. Linder, S. Roberts, and R. Smith. A multibeam HI survey of the Virgo cluster - two isolated HI clouds? *MNRAS*, 349:922–932, April 2004. doi: 10.1111/j.1365-2966.2004.07568.x.

M. Davis and J. Huchra. A survey of galaxy redshifts. III - The density field and the induced gravity field. *ApJ*, 254:437–450, March 1982. doi: 10.1086/159751.

G. De Lucia and J. Blaizot. The hierarchical formation of the brightest cluster galaxies. *MNRAS*, 375:2–14, February 2007. doi: 10.1111/j.1365-2966.2006.11287.x.

A. Dressler, A. Oemler, Jr., W. J. Couch, I. Smail, R. S. Ellis, A. Barger, H. Butcher, B. M. Poggianti, and R. M. Sharples. Evolution since z = 0.5 of the Morphology-Density Relation for Clusters of Galaxies. *ApJ*, 490:577–+, December 1997. doi: 10.1086/304890.

G. Efstathiou. A model of supernova feedback in galaxy formation. *MNRAS*, 317:697–719, September 2000. doi: 10.1046/j.1365-8711.2000.03665.x.

G. Efstathiou, R. S. Ellis, and B. A. Peterson. Analysis of a complete galaxy redshift survey. II - The field-galaxy luminosity function. *MNRAS*, 232:431–461, May 1988.

P. Erdoğdu, O. Lahav, J. P. Huchra, M. Colless, R. M. Cutri, E. Falco, T. George, T. Jarrett, D. H. Jones, L. M. Macri, J. Mader, N. Martimbeau, M. A. Pahre, Q. A. Parker, A. Rassat, and W. Saunders. Reconstructed density and velocity fields from the 2MASS Redshift Survey. *MNRAS*, 373:45–64, November 2006. doi: 10.1111/j.1365-2966.2006.11049.x.

K. B. Fisher, M. Davis, M. A. Strauss, A. Yahil, and J. Huchra. Clustering in the 1.2-Jy IRAS Galaxy Redshift Survey. I - The redshift and real space correlation functions. *MNRAS*, 266:50–+, January 1994a.

K. B. Fisher, M. Davis, M. A. Strauss, A. Yahil, and J. P. Huchra. Clustering in the 1.2-JY IRAS Galaxy Redshift Survey - Part Two - Redshift Distortions and Xi/r/p, PI. *MNRAS*, 267:927–+, April 1994b.

A. S. Font, R. G. Bower, I. G. McCarthy, A. J. Benson, C. S. Frenk, J. C. Helly, C. G. Lacey, C. M. Baugh, and S. Cole. The colours of satellite galaxies in groups and clusters. *MNRAS*, 389:1619–1629, October 2008. doi: 10.1111/j.1365-2966.2008.13698.x.

E. Freeland, A. Stilp, and E. Wilcots. H I Observations of Five Groups of Galaxies. *AJ*, 138:295–304, July 2009. doi: 10.1088/0004-6256/138/1/295.

W. Freudling, B. Catinella, M. Calabretta, E. Momjian, M. Zwaan, S. Linder, R. Minchin, K. O'Neil, and L. Staveley-Smith. The ALFA Ultra Deep Survey (AUDS). In R. Minchin & E. Momjian, editor, *The Evolution of Galaxies Through the Neutral Hydrogen Window*, volume 1035 of *American Institute of Physics Conference Series*, pages 242–245, August 2008. doi: 10.1063/1.2973592.

W. Freudling, L. Staveley-Smith, B. Catinella, R. Minchin, M. Calabretta, E. Momjian, M. Zwaan, M. Meyer, and K. O'Neil. Deep 21 cm H I Observations at z $\sim$ 0.1: The Precursor to the Arecibo Ultra Deep Survey. *ApJ*, 727: 40–+, January 2011. doi: 10.1088/0004-637X/727/1/40.

M. Fukugita and P. J. E. Peebles. The Cosmic Energy Inventory. *ApJ*, 616:643–668, December 2004. doi: 10.1086/425155.

M. Fukugita, C. J. Hogan, and P. J. E. Peebles. The Cosmic Baryon Budget. *ApJ*, 503:518–+, August 1998. doi: 10.1086/306025.

M. Fumagalli, M. R. Krumholz, J. X. Prochaska, G. Gavazzi, and A. Boselli. Molecular Hydrogen Deficiency in H I-poor Galaxies and its Implications for Star Formation. *ApJ*, 697:1811–1821, June 2009. doi: 10.1088/0004-637X/697/2/1811.

G. Gavazzi, A. Boselli, W. van Driel, and K. O'Neil. Completing H I observations of galaxies in the Virgo cluster. *A&A*, 429:439–447, January 2005. doi: 10.1051/0004-6361:20041678.

M. Geha, M. R. Blanton, M. Masjedi, and A. A. West. The Baryon Content of Extremely Low Mass Dwarf Galaxies. *ApJ*, 653:240–254, December 2006. doi: 10.1086/508604.

R. Giovanelli, M. P. Haynes, B. R. Kent, P. Perillat, B. Catinella, G. L. Hoffman, E. Momjian, J. L. Rosenberg, A. Saintonge, K. Spekkens, S. Stierwalt, N. Brosch, K. L. Masters, C. M. Springob, I. D. Karachentsev, V. E. Karachentseva, R. A. Koopmann, E. Muller, W. van Driel, and L. van Zee. The Arecibo Legacy Fast ALFA Survey. II. Results of Precursor Observations. *AJ*, 130:2613–2624, December 2005a. doi: 10.1086/497432.

R. Giovanelli, M. P. Haynes, B. R. Kent, P. Perillat, A. Saintonge, N. Brosch, B. Catinella, G. L. Hoffman, S. Stierwalt, K. Spekkens, M. S. Lerner, K. L. Masters, E. Momjian, J. L. Rosenberg, C. M. Springob, A. Boselli, V. Charmandaris, J. K. Darling, J. Davies, D. Garcia Lambas, G. Gavazzi, C. Giovanardi, E. Hardy, L. K. Hunt, A. Iovino, I. D. Karachentsev, V. E. Karachentseva, R. A. Koopmann, C. Marinoni, R. Minchin, E. Muller, M. Putman, C. Pantoja, J. J. Salzer, M. Scodeggio, E. Skillman, J. M. Solanes, C. Valotto, W. van Driel, and L. van Zee. The Arecibo Legacy Fast ALFA Survey. I. Science Goals, Survey Design, and Strategy. *AJ*, 130:2598–2612, December 2005b. doi: 10.1086/497431.

R. Giovanelli, M. P. Haynes, B. R. Kent, A. Saintonge, S. Stierwalt, A. Altaf, T. Balonek, N. Brosch, S. Brown, B. Catinella, A. Furniss, J. Goldstein, G. L. Hoffman, R. A. Koopmann, D. A. Kornreich, B. Mahmood, A. M. Martin, K. L. Masters, A. Mitschang, E. Momjian, P. H. Nair, J. L. Rosenberg, and B. Walsh. The Arecibo Legacy Fast ALFA Survey. III. H I Source Catalog of the Northern Virgo Cluster Region. *AJ*, 133:2569–2583, June 2007. doi: 10.1086/516635.

R. Giovanelli, M. P. Haynes, B. R. Kent, and E. A. K. Adams. Are Newly Discovered H I High-Velocity Clouds Minihalos in the Local Group? *ApJL*, 708: L22–L25, January 2010. doi: 10.1088/2041-8205/708/1/L22.

N. Y. Gnedin. Effect of Reionization on Structure Formation in the Universe. *ApJ*, 542:535–541, October 2000. doi: 10.1086/317042.

N. Y. Gnedin, K. Tassis, and A. V. Kravtsov. Modeling Molecular Hydrogen and Star Formation in Cosmological Simulations. *ApJ*, 697:55–67, May 2009. doi: 10.1088/0004-637X/697/1/55.

F. Governato, B. Willman, L. Mayer, A. Brooks, G. Stinson, O. Valenzuela, J. Wadsley, and T. Quinn. Forming disc galaxies in ΛCDM simulations. *MNRAS*, 374:1479–1494, February 2007. doi: 10.1111/j.1365-2966.2006.11266.x.

F. Governato, C. Brook, L. Mayer, A. Brooks, G. Rhee, J. Wadsley, P. Jonsson, B. Willman, G. Stinson, T. Quinn, and P. Madau. Bulgeless dwarf galaxies and dark matter cores from supernova-driven outflows. *Nature*, 463:203–206, January 2010. doi: 10.1038/nature08640.

E. K. Grebel. Stellar Populations in the Local Group of Galaxies. In J. Mikolajewska & A. Olech, editor, *Stellar Astrophysics with the World's Largest Telescopes*, volume 752 of *American Institute of Physics Conference Series*, pages 161–174, March 2005. doi: 10.1063/1.1893349.

E. K. Grebel and J. S. Gallagher, III. The Impact of Reionization on the Stellar Populations of Nearby Dwarf Galaxies. *ApJL*, 610:L89–L92, August 2004. doi: 10.1086/423339.

L. Guzzo, M. A. Strauss, K. B. Fisher, R. Giovanelli, and M. P. Haynes. Redshift-Space Distortions and the Real-Space Clustering of Different Galaxy Types. *ApJ*, 489:37–+, November 1997. doi: 10.1086/304788.

M. R. Haas, J. Schaye, and A. Jeeson-Daniel. Disentangling galaxy environment and host halo mass. *ArXiv e-prints*, March 2011.

E. Hawkins, S. Maddox, S. Cole, O. Lahav, D. S. Madgwick, P. Norberg, J. A. Peacock, I. K. Baldry, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, R. Cannon, M. Colless, C. Collins, W. Couch, G. Dalton, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, C. S. Frenk, K. Glazebrook, C. Jackson, B. Jones, I. Lewis, S. Lumsden, W. Percival, B. A. Peterson, W. Sutherland, and K. Taylor. The 2dF Galaxy Redshift Survey: correlation functions, peculiar velocities and the matter density of the Universe. *MNRAS*, 346:78–96, November 2003. doi: 10.1046/j.1365-2966.2003.07063.x.

M. P. Haynes and R. Giovanelli. The pattern of H I deficiency in the Virgo cluster. *ApJ*, 306:466–482, July 1986a. doi: 10.1086/164357.

M. P. Haynes and R. Giovanelli. The connection between Pisces-Perseus and the Local Supercluster. *ApJL*, 306:L55–L59, July 1986b. doi: 10.1086/184705.

M. P. Haynes, R. Giovanelli, and G. L. Chincarini. The Influence of Enviri-onment on the H I Content of Galaxies. *ARA&A*, 22:445–470, 1984. doi: 10.1146/annurev.aa.22.090184.002305.

P. A. Henning, L. Staveley-Smith, R. D. Ekers, A. J. Green, R. F. Haynes, S. Juraszek, M. J. Kesteven, B. Koribalski, R. C. Kraan-Korteweg, R. M. Price, E. M. Sadler, and A. Schröder. H I-bright Galaxies in the Southern Zone of Avoidance. *AJ*, 119:2686–2698, June 2000. doi: 10.1086/301374.

M. Hoeft and S. Gottloeber. Dwarf Galaxies in Voids: Dark Matter Halos and Gas Cooling. *ArXiv e-prints*, January 2010.

M. Hoeft, G. Yepes, S. Gottlöber, and V. Springel. Dwarf galaxies in voids: suppressing star formation with photoheating. *MNRAS*, 371:401–414, September 2006. doi: 10.1111/j.1365-2966.2006.10678.x.

M. Hoeft, G. Yepes, and S. Gottlöber. Too Small to Form a Galaxy: How the UV Background Determines the Baryon Fraction. In J. Davies & M. Disney, editor, *IAU Symposium*, volume 244 of *IAU Symposium*, pages 279–283, May 2008. doi: 10.1017/S1743921307014093.

D. W. Hogg, J. Bovy, and D. Lang. Data analysis recipes: Fitting a model to data. *ArXiv e-prints*, August 2010.

A. M. Hopkins. On the Evolution of Star-forming Galaxies. *ApJ*, 615:209–221, November 2004. doi: 10.1086/424032.

M. J. Irwin, V. Belokurov, N. W. Evans, E. V. Ryan-Weber, J. T. A. de Jong, S. Koposov, D. B. Zucker, S. T. Hodgkin, G. Gilmore, P. Prema, L. Hebb, A. Begum, M. Fellhauer, P. C. Hewett, R. C. Kennicutt, Jr., M. I. Wilkinson, D. M. Bramich, S. Vidrih, H.-W. Rix, T. C. Beers, J. C. Barentine, H. Brewington, M. Harvanek, J. Krzesinski, D. Long, A. Nitta, and S. A. Snedden. Discovery of an Unusual Dwarf Galaxy in the Outskirts of the Milky Way. *ApJL*, 656:L13–L16, February 2007. doi: 10.1086/512183.

A. Jenkins, C. S. Frenk, S. D. M. White, J. M. Colberg, S. Cole, A. E. Evrard, H. M. P. Couchman, and N. Yoshida. The mass function of dark matter haloes. *MNRAS*, 321:372–384, February 2001. doi: 10.1046/j.1365-8711.2001.04029.x.

G. Kauffmann, S. D. M. White, T. M. Heckman, B. Ménard, J. Brinchmann, S. Charlot, C. Tremonti, and J. Brinkmann. The environmental dependence of the relations between stellar mass, structure, star formation and nuclear activity in galaxies. *MNRAS*, 353:713–731, September 2004. doi: 10.1111/j.1365-2966.2004.08117.x.

R. C. Kennicutt, Jr. The Global Schmidt Law in Star-forming Galaxies. *ApJ*, 498: 541–+, May 1998. doi: 10.1086/305588.

B. R. Kent, R. Giovanelli, M. P. Haynes, A. M. Martin, A. Saintonge, S. Stierwalt, T. J. Balonek, N. Brosch, and R. A. Koopmann. The Arecibo Legacy Fast Alfa Survey. VI. Second HI Source Catalog of the Virgo Cluster Region. *AJ*, 136: 713–724, August 2008. doi: 10.1088/0004-6256/136/2/713.

H.-S. Kim, C. M. Baugh, A. J. Benson, S. Cole, C. S. Frenk, C. G. Lacey, C. Power, and M. Schneider. The spatial distribution of cold gas in hierarchical galaxy formation models. *ArXiv e-prints*, February 2010.

M. D. Kistler, H. Yüksel, J. F. Beacom, A. M. Hopkins, and J. S. B. Wyithe. The Star Formation Rate in the Reionization Era as Indicated by Gamma-Ray Bursts. *ApJL*, 705:L104–L108, November 2009. doi: 10.1088/0004-637X/705/2/L104.

A. Klypin, A. V. Kravtsov, O. Valenzuela, and F. Prada. Where Are the Missing Galactic Satellites? *ApJ*, 522:82–92, September 1999. doi: 10.1086/307643.

B. S. Koribalski, L. Staveley-Smith, V. A. Kilborn, S. D. Ryder, R. C. Kraan-Korteweg, E. V. Ryan-Weber, R. D. Ekers, H. Jerjen, P. A. Henning, M. E. Putman, M. A. Zwaan, W. J. G. de Blok, M. R. Calabretta, M. J. Disney, R. F. Minchin, R. Bhathal, P. J. Boyce, M. J. Drinkwater, K. C. Freeman, B. K. Gibson, A. J. Green, R. F. Haynes, S. Juraszek, M. J. Kesteven, P. M. Knezek, S. Mader, M. Marquarding, M. Meyer, J. R. Mould, T. Oosterloo, J. O'Brien, R. M. Price, E. M. Sadler, A. Schröder, I. M. Stewart, F. Stootman, M. Waugh, B. E. Warren, R. L. Webster, and A. E. Wright. The 1000 Brightest HIPASS Galaxies: H I Properties. *AJ*, 128:16–46, July 2004. doi: 10.1086/421744.

M. Krumpe, T. Miyaji, and A. L. Coil. The Spatial Clustering of ROSAT All-Sky Survey AGNs. I. The Cross-correlation Function with SDSS Luminous Red Galaxies. *ApJ*, 713:558–572, April 2010. doi: 10.1088/0004-637X/713/1/558.

S. D. Landy and A. S. Szalay. Bias and variance of angular correlation functions. *ApJ*, 412:64–71, July 1993. doi: 10.1086/172900.

R. H. Lang, P. J. Boyce, V. A. Kilborn, R. F. Minchin, M. J. Disney, C. A. Jordan, M. Grossi, D. A. Garcia, K. C. Freeman, S. Phillipps, and A. E. Wright. First results from the HI Jodrell All Sky Survey: inclination-dependent selection effects in a 21-cm blind survey. *MNRAS*, 342:738–758, July 2003. doi: 10.1046/j.1365-8711.2003.06535.x.

I. Lewis, M. Balogh, R. De Propris, W. Couch, R. Bower, A. Offer, J. Bland-Hawthorn, I. K. Baldry, C. Baugh, T. Bridges, R. Cannon, S. Cole, M. Colless, C. Collins, N. Cross, G. Dalton, S. P. Driver, G. Efstathiou, R. S. Ellis, C. S. Frenk, K. Glazebrook, E. Hawkins, C. Jackson, O. Lahav, S. Lumsden, S. Maddox, D. Madgwick, P. Norberg, J. A. Peacock, W. Percival, B. A. Peterson, W. Sutherland, and K. Taylor. The 2dF Galaxy Redshift Survey: the environmental dependence of galaxy star formation rates near clusters. *MNRAS*, 334: 673–683, August 2002. doi: 10.1046/j.1365-8711.2002.05558.x.

C. Li and S. D. M. White. The distribution of stellar mass in the low-redshift Universe. *MNRAS*, 398:2177–2187, October 2009. doi: 10.1111/j.1365-2966.2009.15268.x.

C. Li, G. Kauffmann, Y. P. Jing, S. D. M. White, G. Börner, and F. Z. Cheng. The dependence of clustering on galaxy properties. *MNRAS*, 368:21–36, May 2006a. doi: 10.1111/j.1365-2966.2006.10066.x.

C. Li, G. Kauffmann, L. Wang, S. D. M. White, T. M. Heckman, and Y. P. Jing. The clustering of narrow-line AGN in the local Universe. *MNRAS*, 373:457–468, December 2006b. doi: 10.1111/j.1365-2966.2006.11079.x.

C. H. Lineweaver, L. Tenorio, G. F. Smoot, P. Keegstra, A. J. Banday, and P. Lubin. The Dipole Observed in the COBE DMR 4 Year Data. *ApJ*, 470:38–+, October 1996. doi: 10.1086/177846.

J. Loveday. The K-band luminosity function of nearby field galaxies. *MNRAS*, 312:557–566, March 2000. doi: 10.1046/j.1365-8711.2000.03179.x.

D. S. Madgwick, O. Lahav, I. K. Baldry, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, R. Cannon, S. Cole, M. Colless, C. Collins, W. Couch, G. Dalton, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, C. S. Frenk, K. Glazebrook, C. Jackson, I. Lewis, S. Lumsden, S. Maddox, P. Norberg, J. A. Peacock, B. A. Peterson, W. Sutherland, and K. Taylor. The 2dF Galaxy Redshift Survey: galaxy luminosity functions per spectral type. *MNRAS*, 333:133–144, June 2002. doi: 10.1046/j.1365-8711.2002.05393.x.

F. A. Marín, N. Y. Gnedin, H.-J. Seo, and A. Vallinotto. Modeling the Large-scale Bias of Neutral Hydrogen. *ApJ*, 718:972–980, August 2010. doi: 10.1088/0004-637X/718/2/972.

A. M. Martin, R. Giovanelli, M. P. Haynes, A. Saintonge, G. L. Hoffman, B. R. Kent, and S. Stierwalt. The Arecibo Legacy Fast Alfa Survey. VIII. H I Source Catalog of the Anti-Virgo Region at $\delta = +25^\circ$. *ApJS*, 183:214–224, August 2009. doi: 10.1088/0067-0049/183/2/214.

A. M. Martin, E. Papastergis, R. Giovanelli, M. P. Haynes, C. M. Springob, and S. Stierwalt. The Arecibo Legacy Fast ALFA Survey: X. The HI Mass Function and Omega_HI From the 40% ALFALFA Survey. *ArXiv e-prints*, August 2010.

K. L. Masters. *Galaxy flows in and around the Local Supercluster*. PhD thesis, Cornell University, United States – New York, 2005.

K. L. Masters, M. P. Haynes, and R. Giovanelli. The Impact of Distance Uncertainties on Local Luminosity and Mass Functions. *ApJL*, 607:L115–L118, June 2004. doi: 10.1086/422100.

L. Mayer, F. Governato, and T. Kaufmann. The formation of disk galaxies in computer simulations. *Advanced Science Letters*, 1:7–27, June 2008.

M. J. Meyer, M. A. Zwaan, R. L. Webster, L. Staveley-Smith, E. Ryan-Weber, M. J. Drinkwater, D. G. Barnes, M. Howlett, V. A. Kilborn, J. Stevens, M. Waugh, M. J. Pierce, R. Bhathal, W. J. G. de Blok, M. J. Disney, R. D. Ekers, K. C. Freeman, D. A. Garcia, B. K. Gibson, J. Harnett, P. A. Henning, H. Jerjen, M. J. Kesteven, P. M. Knezek, B. S. Koribalski, S. Mader, M. Marquarding, R. F. Minchin, J. O'Brien, T. Oosterloo, R. M. Price, M. E. Putman, S. D. Ryder, E. M. Sadler, I. M. Stewart, F. Stootman, and A. E. Wright. The HIPASS catalogue - I. Data presentation. *MNRAS*, 350:1195–1209, June 2004. doi: 10.1111/j.1365-2966.2004.07710.x.

M. J. Meyer, M. A. Zwaan, R. L. Webster, M. J. I. Brown, and L. Staveley-Smith. The Weak Clustering of Gas-rich Galaxies. *ApJ*, 654:702–713, January 2007. doi: 10.1086/508799.

R. F. Minchin. *Properties of galaxies found in a deep blind neutral hydrogen survey*. PhD thesis, Department of Physics and Astronomy, Cardiff University 5 The Parade, Cardiff, Glamorgan, CF24 3YB, Wales, UK, 2001.

S. T. Myers, F. B. Abdalla, C. Blake, L. Koopmans, J. Lazio, and S. Rawling. The Billion Galaxy Cosmological HI Large Deep Survey (BiG-CHILDS). In *astro2010: The Astronomy and Astrophysics Decadal Survey*, volume 2010 of *Astronomy*, pages 219–+, 2009.

P. Norberg, C. M. Baugh, E. Hawkins, S. Maddox, D. Madgwick, O. Lahav, S. Cole, C. S. Frenk, I. Baldry, J. Bland-Hawthorn, T. Bridges, R. Cannon, M. Colless, C. Collins, W. Couch, G. Dalton, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, K. Glazebrook, C. Jackson, I. Lewis, S. Lumsden, J. A. Peacock, B. A. Peterson, W. Sutherland, and K. Taylor. The 2dF Galaxy Redshift Survey: the dependence of galaxy clustering on luminosity and spectral type. *MNRAS*, 332:827–838, June 2002. doi: 10.1046/j.1365-8711.2002.05348.x.

P. Noterdaeme, P. Petitjean, C. Ledoux, and R. Srianand. Evolution of the cosmological mass density of neutral gas from Sloan Digital Sky Survey II - Data Release 7. *A&A*, 505:1087–1098, October 2009. doi: 10.1051/0004-6361/200912768.

D. Obreschkow, D. Croton, G. De Lucia, S. Khochfar, and S. Rawlings. Simulation of the Cosmic Evolution of Atomic and Molecular Hydrogen in Galaxies. *ApJ*, 698:1467–1484, June 2009. doi: 10.1088/0004-637X/698/2/1467.

T. Oosterloo and A. Iovino. HI observations of compact groups of galaxies. *PASA*, 14:48–51, April 1997.

C. Park, Y.-Y. Choi, M. S. Vogeley, J. R. Gott, III, and M. R. Blanton. Environmental Dependence of Properties of Galaxies in the Sloan Digital Sky Survey. *ApJ*, 658:898–916, April 2007. doi: 10.1086/511059.

S. S. Passmoor, C. M. Cress, and A. Faltenbacher. Clustering of HI galaxies in HIPASS and ALFALFA. *ArXiv e-prints*, January 2011.

U.-L. Pen, L. Staveley-Smith, J. B. Peterson, and T.-C. Chang. First detection of cosmic structure in the 21-cm intensity field. *MNRAS*, 394:L6–L10, March 2009. doi: 10.1111/j.1745-3933.2008.00581.x.

J. B. Peterson, R. Aleksan, R. Ansari, K. Bandura, D. Bond, J. Bunton, K. Carlson, T.-C. Chang, F. DeJongh, M. Dobbs, S. Dodelson, H. Darhmaoui, N. Gnedin, M. Halpern, C. Hogan, J.-M. Le Goff, T. T. Liu, A. Legrouri, A. Loeb, K. Loudiyi, C. Magneville, J. Marriner, D. P. McGinnis, B. McWilliams, M. Moniez, N. Palanque-Delabruille, R. J. Pasquinelli, U.-L. Pen, J. Rich, V. Scarpine, H.-J. Seo, K. Sigurdson, U. Seljak, A. Stebbins, J. H. Steffen, C. Stoughton, P. T. Timbie, A. Vallinotto, and C. Teche. 21-cm Intensity Mapping. In *astro2010: The Astronomy and Astrophysics Decadal Survey*, volume 2010 of *Astronomy*, pages 234–+, 2009.

D. J. Pisano, D. G. Barnes, B. K. Gibson, L. Staveley-Smith, K. C. Freeman, and V. A. Kilborn. HI in Local Group Analogs: What does it Tell Us about Galaxy Formation? In I. Saviane, V. D. Ivanov, & J. Borissova, editor, *Groups of Galaxies in the Nearby Universe*, pages 33–+, 2007.

C. Power, C. M. Baugh, and C. G. Lacey. The redshift evolution of the mass function of cold gas in hierarchical galaxy formation models. *MNRAS*, 406: 43–59, July 2010. doi: 10.1111/j.1365-2966.2010.16481.x.

W. H. Press and P. Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *ApJ*, 187:425–438, February 1974. doi: 10.1086/152650.

W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in FORTRAN. The art of scientific computing*. 1992.

J. X. Prochaska and J. Tumlinson. *Baryons: What, When and Where?*, pages 419–+. 2009.

J. X. Prochaska and A. M. Wolfe. On the (Non)Evolution of H I Gas in Galaxies Over Cosmic Time. *ApJ*, 696:1543–1547, May 2009. doi: 10.1088/0004-637X/696/2/1543.

J. X. Prochaska, S. Herbert-Fort, and A. M. Wolfe. The SDSS Damped Lyα Survey: Data Release 3. *ApJ*, 635:123–142, December 2005. doi: 10.1086/497287.

S. M. Rao, D. A. Turnshek, and D. B. Nestor. Damped Lyα Systems at z < 1.65: The Expanded Sloan Digital Sky Survey Hubble Space Telescope Sample. *ApJ*, 636:610–630, January 2006. doi: 10.1086/498132.

S. Rawlings, F. B. Abdalla, S. L. Bridle, C. A. Blake, C. M. Baugh, L. J. Greenhill, and J. M. van der Hulst. Galaxy evolution, cosmology and dark energy with the Square Kilometer Array. *NewAR*, 48:1013–1027, December 2004. doi: 10.1016/j.newar.2004.09.024.

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. *MNRAS*, 374:2–15, January 2007. doi: 10.1111/j.1365-2966.2006.11204.x.

M. Ricotti. Late gas accretion on to primordial minihaloes: a model for Leo T, dark galaxies and extragalactic high-velocity clouds. *MNRAS*, 392:L45–L49, January 2009a. doi: 10.1111/j.1745-3933.2008.00586.x.

M. Ricotti. The First Galaxies and the Likely Discovery of their Fossils in the Local Group. *ArXiv e-prints*, November 2009b.

M. S. Roberts and M. P. Haynes. Physical Parameters along the Hubble Sequence. *ARA&A*, 32:115–152, 1994. doi: 10.1146/annurev.aa.32.090194.000555.

J. L. Rosenberg and S. E. Schneider. The Arecibo Dual-Beam Survey: Arecibo and VLA Observations. *ApJS*, 130:177–199, September 2000. doi: 10.1086/317347.

J. L. Rosenberg and S. E. Schneider. The Arecibo Dual-Beam Survey: The H I Mass Function of Galaxies. *ApJ*, 567:247–257, March 2002. doi: 10.1086/338377.

E. V. Ryan-Weber, A. Begum, T. Oosterloo, S. Pal, M. J. Irwin, V. Belokurov, N. W. Evans, and D. B. Zucker. The Local Group dwarf Leo T: HI on the brink of star formation. *MNRAS*, 384:535–540, February 2008. doi: 10.1111/j.1365-2966.2007.12734.x.

A. Saintonge. The Arecibo Legacy Fast ALFA Survey. IV. Strategies for Signal Identification and Survey Catalog Reliability. *AJ*, 133:2087–2096, May 2007a. doi: 10.1086/513515.

A. Saintonge. *Properties of low mass dwarf galaxies in the ALFALFA survey*. PhD thesis, Cornell University, 2007b.

A. Saintonge, R. Giovanelli, M. P. Haynes, G. L. Hoffman, B. R. Kent, A. M. Martin, S. Stierwalt, and N. Brosch. The Arecibo Legacy Fast Alfa Survey. V. The H I Source Catalog of the Anti-Virgo Region at $\delta = +27^\circ$. *AJ*, 135:588–604, February 2008. doi: 10.1088/0004-6256/135/2/588.

T. Sanchis, G. A. Mamon, E. Salvador-Solé, and J. M. Solanes. The origin of H I-deficiency in galaxies on the outskirts of the Virgo cluster. II. Companions

and uncertainties in distances and deficiencies. *A&A*, 418:393–411, May 2004. doi: 10.1051/0004-6361:20034158.

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. *ApJ*, 232:352–364, September 1979. doi: 10.1086/157295.

W. Saunders, M. Rowan-Robinson, and A. Lawrence. The spatial correlation function of IRAS galaxies on small and intermediate scales. *MNRAS*, 258: 134–146, September 1992.

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. *ApJ*, 500:525–+, June 1998. doi: 10.1086/305772.

M. Schmidt. Space Distribution and Luminosity Functions of Quasi-Stellar Radio Sources. *ApJ*, 151:393–+, February 1968. doi: 10.1086/149446.

J. M. Schombert, S. S. McGaugh, and J. A. Eder. Gas Mass Fractions and the Evolution of Low Surface Brightness Dwarf Galaxies. *AJ*, 121:2420–2430, May 2001. doi: 10.1086/320398.

U. Seljak. Analytic model for galaxy and dark matter clustering. *MNRAS*, 318: 203–213, October 2000. doi: 10.1046/j.1365-8711.2000.03715.x.

Z. Shao, Q. Xiao, S. Shen, H. J. Mo, X. Xia, and Z. Deng. Inclination-dependent Luminosity Function of Spiral Galaxies in the Sloan Digital Sky Survey: Implications for Dust Extinction. *ApJ*, 659:1159–1171, April 2007. doi: 10.1086/511131.

G. S. Shostak and R. J. Allen. Accurate H I profiles of several nearby galaxies. *A&A*, 81:167–171, January 1980.

J. D. Simon and M. Geha. The Kinematics of the Ultra-faint Milky Way Satellites: Solving the Missing Satellite Problem. *ApJ*, 670:313–331, November 2007. doi: 10.1086/521816.

J. M. Solanes. Looking for clues to the nature of HI deficiency in cluster spirals. *ArXiv Astrophysics e-prints*, June 2001.

J. M. Solanes, A. Manrique, C. García-Gómez, G. González-Casado, R. Giovanelli, and M. P. Haynes. The H I Content of Spirals. II. Gas Deficiency in Cluster Galaxies. *ApJ*, 548:97–113, February 2001. doi: 10.1086/318672.

V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, J. A. Peacock, S. Cole, P. Thomas,

H. Couchman, A. Evrard, J. Colberg, and F. Pearce. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435:629–636, June 2005. doi: 10.1038/nature03597.

C. M. Springob, M. P. Haynes, and R. Giovanelli. Morphology, Environment, and the H I Mass Function. *ApJ*, 621:215–226, March 2005. doi: 10.1086/427432.

C. M. Springob, K. L. Masters, M. P. Haynes, R. Giovanelli, and C. Marinoni. SFI++. II. A New I-Band Tully-Fisher Catalog, Derivation of Peculiar Velocities, and Data Set Properties. *ApJS*, 172:599–614, October 2007. doi: 10.1086/519527.

S. Stierwalt, M. P. Haynes, R. Giovanelli, B. R. Kent, A. M. Martin, A. Saintonge, I. D. Karachentsev, and V. E. Karachentseva. The Arecibo Legacy Fast Alfa Survey. IX. The Leo Region H I Catalog, Group Membership, and the H I Mass Function for the Leo I Group. *AJ*, 138:338–361, August 2009. doi: 10.1088/0004-6256/138/2/338.

M. E. C. Swanson, M. Tegmark, M. Blanton, and I. Zehavi. SDSS galaxy clustering: luminosity and colour dependence and stochasticity. *MNRAS*, 385: 1635–1655, April 2008. doi: 10.1111/j.1365-2966.2008.12948.x.

A. V. Tikhonov and A. Klypin. The emptiness of voids: yet another overabundance problem for the Λ cold dark matter model. *MNRAS*, 395:1915–1924, June 2009. doi: 10.1111/j.1365-2966.2009.14686.x.

E. Tolstoy, V. Hill, and M. Tosi. Star-Formation Histories, Abundances, and Kinematics of Dwarf Galaxies in the Local Group. *ARA&A*, 47:371–425, September 2009. doi: 10.1146/annurev-astro-082708-101650.

M. C. Toribio, J. M. Solanes, R. Giovanelli, M. P. Haynes, and A. Martin. HI Content and Optical Properties of Field Galaxies from the ALFALFA Survey. II. Multivariate Analysis of a Galaxy Sample in Low Density Environments. *ArXiv e-prints*, March 2011a.

M. C. Toribio, J. M. Solanes, R. Giovanelli, M. P. Haynes, and K. L. Masters. HI Content and Optical Properties of Field Galaxies from the ALFALFA Survey. I.Selection of a Control Sample. *ArXiv e-prints*, March 2011b.

N. Trentham. The galaxy luminosity function in clusters and the field. *MNRAS*, 294:193–+, February 1998. doi: 10.1046/j.1365-8711.1998.01084.x.

N. Trentham and S. Hodgkin. The luminosity function of the Virgo Cluster from M_B=-22 to -11. *MNRAS*, 333:423–442, June 2002. doi: 10.1046/j.1365-8711.2002.05440.x.

R. van de Weygaert, K. Kreckel, E. Platen, B. Beygu, J. H. van Gorkom, J. M. van der Hulst, M. A. Aragon-Calvo, P. J. E. Peebles, T. Jarrett, G. Rhee, K. Kovac, and C. -. Yip. The Void Galaxy Survey. *ArXiv e-prints*, January 2011.

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. *AJ*, 113:1618–1637, May 1997. doi: 10.1086/118379.

M. A. W. Verheijen, N. Trentham, B. Tully, and M. Zwaan. The HI Mass Function in Ursa Major. In J. E. Hibbard, M. Rupen, & J. H. van Gorkom, editor, *Gas and Galaxy Evolution*, volume 240 of *Astronomical Society of the Pacific Conference Series*, pages 507–+, 2001.

Y. Wang, X. Yang, H. J. Mo, and F. C. van den Bosch. The Cross-Correlation between Galaxies of Different Luminosities and Colors. *ApJ*, 664:608–632, August 2007. doi: 10.1086/519245.

D. F. Watson, A. A. Berlind, and A. R. Zentner. A Cosmic Coincidence: The Power-Law Galaxy Correlation Function. *ArXiv e-prints*, January 2011.

D. H. Weinberg, A. Szomoru, P. Guhathakurta, and J. H. van Gorkom. On the population of H I dwarf galaxies. *ApJL*, 372:L13–L16, May 1991. doi: 10.1086/186012.

A. A. West, D. A. Garcia-Appadoo, J. J. Dalcanton, M. J. Disney, C. M. Rockosi, and Ž. Ivezić. H I-Selected Galaxies in the Sloan Digital Sky Survey. II. The Colors of Gas-Rich Galaxies. *AJ*, 138:796–807, September 2009. doi: 10.1088/0004-6256/138/3/796.

A. A. West, D. A. Garcia-Appadoo, J. J. Dalcanton, M. J. Disney, C. M. Rockosi, Ž. Ivezić, M. C. Bentz, and J. Brinkmann. H I-Selected Galaxies in the Sloan Digital Sky Survey. I. Optical Data. *AJ*, 139:315–328, February 2010. doi: 10.1088/0004-6256/139/2/315.

O. I. Wong, E. V. Ryan-Weber, D. A. Garcia-Appadoo, R. L. Webster, L. Staveley-Smith, M. A. Zwaan, M. J. Meyer, D. G. Barnes, V. A. Kilborn, R. Bhathal, W. J. G. de Blok, M. J. Disney, M. T. Doyle, M. J. Drinkwater, R. D. Ekers, K. C. Freeman, B. K. Gibson, S. Gurovich, J. Harnett, P. A. Henning, H. Jerjen, M. J. Kesteven, P. M. Knezek, B. S. Koribalski, S. Mader, M. Marquarding, R. F. Minchin, J. O'Brien, M. E. Putman, S. D. Ryder, E. M. Sadler, J. Stevens, I. M. Stewart, F. Stootman, and M. Waugh. The Northern HIPASS catalogue - data presentation, completeness and reliability measures. *MNRAS*, 371:1855–1864, October 2006. doi: 10.1111/j.1365-2966.2006.10846.x.

M. C. H. Wright. The tail of M33 and the adjacent hydrogen cloud. *ApJ*, 233: 35–38, October 1979. doi: 10.1086/157363.

I. Zehavi, Z. Zheng, D. H. Weinberg, J. A. Frieman, A. A. Berlind, M. R. Blanton, R. Scoccimarro, R. K. Sheth, M. A. Strauss, I. Kayo, Y. Suto, M. Fukugita, O. Nakamura, N. A. Bahcall, J. Brinkmann, J. E. Gunn, G. S. Hennessy, Ž. Ivezić, G. R. Knapp, J. Loveday, A. Meiksin, D. J. Schlegel, D. P. Schneider, I. Szapudi, M. Tegmark, M. S. Vogeley, and D. G. York. The Luminosity and Color Dependence of the Galaxy Correlation Function. *ApJ*, 630:1–27, September 2005. doi: 10.1086/431891.

M. Zwaan, M. Meyer, and L. Staveley-Smith. The velocity function of gas-rich galaxies. *ArXiv e-prints*, December 2009.

M. A. Zwaan, F. H. Briggs, D. Sprayberry, and E. Sorar. The H i Mass Function of Galaxies from a Deep Survey in the 21 Centimeter Line. *ApJ*, 490:173–+, November 1997. doi: 10.1086/304872.

M. A. Zwaan, L. Staveley-Smith, B. S. Koribalski, P. A. Henning, V. A. Kilborn, S. D. Ryder, D. G. Barnes, R. Bhathal, P. J. Boyce, W. J. G. de Blok, M. J. Disney, M. J. Drinkwater, R. D. Ekers, K. C. Freeman, B. K. Gibson, A. J. Green, R. F. Haynes, H. Jerjen, S. Juraszek, M. J. Kesteven, P. M. Knezek, R. C. Kraan-Korteweg, S. Mader, M. Marquarding, M. Meyer, R. F. Minchin, J. R. Mould, J. O'Brien, T. Oosterloo, R. M. Price, M. E. Putman, E. Ryan-Weber, E. M. Sadler, A. Schröder, I. M. Stewart, F. Stootman, B. Warren, M. Waugh, R. L. Webster, and A. E. Wright. The 1000 Brightest HIPASS Galaxies: The H I Mass Function and Ω_{HI}. *AJ*, 125:2842–2858, June 2003. doi: 10.1086/374944.

M. A. Zwaan, M. J. Meyer, L. Staveley-Smith, and R. L. Webster. The HIPASS catalogue: Ω_{HI} and environmental effects on the HI mass function of galaxies. *MNRAS*, 359:L30–L34, May 2005. doi: 10.1111/j.1745-3933.2005.00029.x.