The Large–Scale Distribution of Neutral Hydrogen in the Local Universe

Virginia Alexandra Kilborn

Faculty of Science
The University of Melbourne

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy.

March 12, 2001

Abstract

A blind H I sample of 533 galaxies has been produced by visually cataloguing $\sim 11\%$ of the H I Parkes All Sky Survey (HIPASS). HIPASS data with $\delta < -62$ was catalogued, including the South Celestial Cap (SCC). Of the galaxies catalogued, 140 had not been previously optically identified. An automatic galaxy finder was developed to search through HIPASS data in three dimensions, and produce a catalogue of galaxies with determinable limits. The galaxy finder, MultiFind, was tested with simulations. A test run on 65% of the SCC region, found no additional galaxies compared to the visual catalogue, to a flux limit of ~ 60 mJy. The HIMF was calculated from the SCC catalogue, and no cosmologically significant population of dwarf H I galaxies was found in the sample, although the faint–end slope is steeper than previous determinations. The HIMFs of spiral and late–type galaxies was also investigated. A small population of H I clouds without optical counterparts was discovered, and one of these clouds, HIPASS J1712-64, had no optical counterpart to very faint limits. The nature, distance and possible origin of these clouds is discussed. Eighteen new galaxies found in the SCC region were observed at the Australia Telescope Compact Array. These detections included low surface brightness galaxies, and compact high surface brightness galaxies. The H I properties of the different optical classes of galaxies is discussed.

This is to certify that

(i) the thesis comprises only my original work, except where indicated in the preface

(ii) due acknowledgement has been made in the text to all other material used,

(iii) the thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Preface

The data used in the thesis was the product of the HIPASS collaboration. As a result, there are many contributors to the data collection and reduction. As part of the HIPASS collaboration, I contributed to the project as a whole doing ~ 10 weeks observing at the Parkes Telescope, 2 weeks at the Australia Telescope Compact Array and 1 week of optical observing at the MSSSO 40 inch telescope. I have been heavily involved in data cubing of the survey, both in the region of sky used for this thesis and elsewhere.

•The galaxy finder, MultiFind, was developed fully by myself, on the suggestion of Multibeam working group members. I acknowledge Lister Staveley-Smith for the suggestion of using MIRIAD as the basis. I also acknowledge helpful feedback from users of MultiFind, in particular Baerbel Koribalski and Erwin de Blok. The simulation cubes used for testing were constructed by Niall Ryan. The MultiFind lists from the SCC region were sorted for real detections by Ivy Wong.

•Parts of Chapter 6 were produced as a collaborative effort, with the Multibeam working group.

·The initial discovery of the H I cloud, HIPASS J1712-64, was made by myself, and follow–up observations were made by myself at the ATCA, and at Parkes. All H I and optical data reduction was done by myself, except where indicated in the text.

·I acknowledge Lister Staveley–Smith for making narrow–band observations at Parkes, and Patricia Knezek for making optical B and R-band observations at the CTIO 1-m telescope. I acknowledge Malte Marquarding for help with the optical data reduction, and Michael Brown for assistance using Sextractor.

·This research has benefited from many discussions with Multibeam group members, in particular Lister Staveley-Smith and Rachel Webster who read a number of drafts. Mary Putman, Brad Gibson, and the paper's referee, enlightened me about the issues with HVCs, and helped with the discussion on the nature of HIPASS J1712-64.

·The ESO/UKST plate work was done wholly by David Malin.

·The Hα spectrum of HIPASS J1712-64 was observed by J. Bland-Hawthorn, and I thank him, and Mary Putman for supplying the spectrum and for help with interpretation.

·The new population of HVCs was discovered by myself, with the exception of J718-59 and J1616-55 who were found by Baerbel Koribalski and Lister Staveley-Smith.

•Chapter 4 & Chapter 5: I acknowledge the use of two programs written by Lister Staveley-Smith: SCHECFIT, which fits a Schechter function to a $\sum 1/V_{max}$ method HIMF, and SIMUL which produced simulations of galaxy population. The code used for the density independent HIMF calculation was made available to me by Will Saunders.

Parts of this thesis have been published:

• Chapter 4 has been published in abridged form in Publications of the Astronomical Society of Australia (Kilborn, V., Staveley-Smith, L. & Webster, R. L. 1999). This Chapter now contains a larger data set than published, and a more robust calculation of the HIMF.

• Chapter 6 has been published in part as a HIPASS paper in the Astronomical Journal (Kilborn et al., 2000). This chapter now contains new data and analysis not contained in the published paper.

Acknowledgements

Firstly I would like to thank my supervisors, Rachel Webster and Lister Staveley–Smith. Your help and encouragement has been exceptional. In particular, thanks to Rachel for the support she has given by providing the opportunity to travel to three international conferences. And thanks to Lister for patiently going through all aspects of my thesis work, and for always having the time to help.

During the past three years, I have worked closely with the Multibeam working group. I have been very fortunate to have such a pool of people to probe for answers to scientific and technical questions. In particular, I would like to thank Baerbel Koribalski, David Malin, Ken Freeman, Erwin de Blok, Mary Putman, Brad Gibson, Ian Stewart, Trish Henning and Pat Knezek for discussions and help in all aspects of this thesis. In addition, I would like to thank the Australia Telescope National Facility for awarding me a studentship, and for the opportunity to work at the Facility throughout my candidature.

I would like to acknowledge those who have read the many drafts of my thesis: Rachel, Lister, Mum, Liz and Richard. Thankyou for tirelessly reading drafts and correcting spelling mistakes when I forgot to run a spell check.

Thanks to those who helped with computing: Lister, Brett Holman, Michael Drinkwater, Michael Brown, Matt (LaTex man) Whiting, David Barnes, Erwin de Blok, Malte Marquarding, and the astro group in general.

To members of the Melbourne astro group, past and present - its been great studying with you, and many thanks for the support and kitchen cricket that has kept me going. Also to my friends, thanks for your understanding and friendship.

I'd like to thank my family for supporting me every step of the way – financially and personally. In particular, thanks to my parents, Marion and Phil, for their unending encouragement. To Liz and Rod, thanks for your help and friendship.

Finally, to my husband, Richard. Thanks for never going to bed while I was working late at uni, for doing the washing and cleaning while I was working, and for putting up with me going observing all the time. Your friendship, love and support has made all the difference.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The study of the structure of the universe has largely been based on optical catalogues of galaxies. The study of 'nebulae' before the era of extragalactic astronomy was pioneered by Charles Messier, who catalogued 103 'nebulae' in 1781. Sir William Herschel and his son John catalogued 'galactic nebulae' that lay along the plane of the Milky Way, and 'extragalactic nebulae' that lay away from the plane, although it was thought that both type of nebulae were associated with the Milky Way. By the early 1900s, about ten thousand nebulae had been catalogued, including the "General Catalogue of Galaxies" by John Herschel and the "New General Catalogue (NGC)", by J. Dreyer. In 1924, Edwin Hubble determined that extragalactic nebulae were separate and distinct to our own galaxy by observing cepheid variable stars in M31 and then applying the period–luminosity relationship to determine the distance to M31. Since then, many thousands of galaxies have been catalogued optically, including such catalogues as the Uppsala General Catalogue, UGC (Nilson, 1973), the European Southern Observatory catalogue, ESO (Lauberts, 1982), and the Third Reference Catalogue of Bright Galaxies, RC3 (de Vaucouleurs et al., 1991). See Sandage et al. (1975), Snow (1991) and Friedlander (1985) for more information on the history of optical galaxy catalogues.

The emission observed in optical catalogues comes primarily from stars. However, there is strong evidence that optical light only accounts for a small amount of the total mass in the universe (eg: Trimble (1987)). In addition, optical observations are biased toward high surface brightness objects, and a substantial part of the optical population may be missed in typical optical surveys (Disney, 1976). Instead of looking at the visible star light, the dynamics and distribution of matter can be traced using the most abundant element in the universe, hydrogen. Hydrogen, in its neutral state (H I), emits a photon with a wavelength of 21cm. The emission occurs when an electron in the ground state goes through a spin-flip transition, and moves from the triplet to singlet state (see Figure 1.1). The probability of the transition is low ($\sim 2.868 \times 10^{-15} s^{-1}$), and the spin temperature of the gas is governed by atom collision, and the absorption and emission of photons from either an existing radiation field or from Lyman α photons (Verschur and Kellerman, 1988). This means that a bright external radiation field is not required for H I emission, and the emission is independent from stellar light.

H I emission was first detected by Ewen and Purcell (1951), who found H I in the Milky Way. Kerr and Hindman (1953) were the first to detect extragalactic H I, in the Magellanic Clouds, using a 36 foot antenna in Australia. Since then, many thousands of extragalactic objects have been observed in the 21cm line, either from 'pointed' observations, with the pointing centre based on positions from optical catalogues (eg: Fisher and Tully (1981a)), or from 'blind' H I surveys, where an area of sky was 'scanned' for 21cm emission (eg: Sorar (1994); Henning (1995); Schneider et al. (1998)). In the past, H I observations in the literature were dominated by pointed observations, as the telescope time required for blind observations was prohibitive. However, in recent years, improved instrumentation has meant that scanning methods have been practical for a number of blind H I surveys. Even so, the number of galaxies catalogued from blind H I surveys is only $\sim$ 1500 galaxies at present. Table 1.1 shows the largest blind H I surveys with the area surveyed, and observing parameters. The surveys have increased in sensitivity and size as technology has improved.

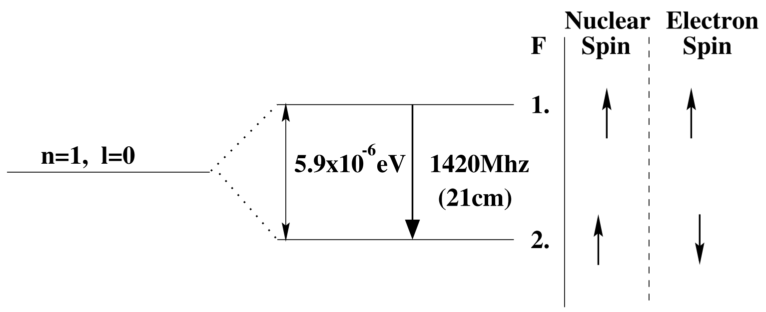

Figure 1.1: Energy level diagram for the hyperfine splitting of the ground state the hydrogen atom (Harwitt, 1988). A 21cm photon is emitted in this transition.

Observers	**Volume**[a] (Mpc^3)	**Vel. Range** (km s^{-1})	$\mathbf{M_{lim}}$[b] ($M_\odot$)	#[c]	**Region**
Shostak (1977)	500	–485 to –232 63 to 1330	$\sim 10^8$	0	Void
Fisher and Tully (1981b)	220	< 3000	$\sim 10^8$	0	M81 region
Krumm and Brosch (1984)	1×10^4	6500 to 9500 5500 to 8500	8×10^9		Hercules Void Perseus Void
Weinberg et al. (1991)	150	4500 to 5700 2650 to 3850	$\sim 10^8$	16	Perseus-Pisces Supercluster & Void
Sorar (1994)	4000	< 7400	5×10^7	66	General Field
Henning (1995)	1.1×10^4	–400 to 6800	$\sim 10^7$	37	ZOA/General
Spitzak and Schneider (1998)	6000	< 8340	$\sim 10^7$	75	General Field
Rosenberg and Schneider (2000)	$\sim 10^5$	–654 to 7977	$\lesssim 10^7$	265	General Field
ZOA[d]	1.2×10^5	-1200 to 12700	$\sim 10^6$	$\gtrsim 1000$	ZOA
HIPASS[e]	8×10^6	-1200 to 12700	$\sim 10^6$	$\gtrsim 5000$	$\delta < +2$

[a] see Schneider (1997) for more detail
[b] For a galaxy of H I velocity width 100 km s^{-1}
[c] Number of galaxies found in survey
[d] Estimated total number from shallow ZOA results (Henning et al., 2000)
[e] $\sim$ 1000 galaxies catalogued to date (see Koribalski (2000), Kilborn et al. (1999))

Table 1.1: Blind H I surveys.

1.1 The Space Density of Neutral Hydrogen

The distribution of neutral hydrogen is a key ingredient in theories of galaxy formation and evolution. Galaxies are formed when concentrations of hydrogen collapse under gravity. Galaxies that contain a large reservoir of hydrogen are likely to continue star formation, while those with little hydrogen will live with an aging star population. The evolutionary sequence of galaxies might be obtained by studying their hydrogen content, to determine their "stage of life".

The density of H I at higher redshifts has been measured using the Lyman–α, Lyα, forest seen as absorption in the spectra of quasars (see Rauch (1998) for a recent review). It is found that the number of absorbers is constant until a redshift of $z = 2$, when that number steeply rises. Determining the nature of Lyα absorbers at high redshift, and comparing this to the H I mass distribution at the present epoch is necessary to test models of galaxy formation.

The space density of H I at the present epoch can be measured using either pointed or blind H I surveys. However, H I surveys based on optical catalogues risk being biased towards galaxies of high optical surface brightness, which is problematic unless the H I mass distribution follows the optical brightness distribution. Optical surveys also preclude the possibility of detecting galaxies with no visible optical star emission at all, or protogalaxies which are yet to form stars.

1.1.1 H I Mass Functions

The distribution of H I mass has been studied both from blind H I surveys, and from pointed observations from optical catalogues. Similar results have been obtained from both types of surveys. The H I Mass Function (HIMF) describes the number of galaxies per log H I mass interval, per Mpc^{-3}, and is the H I equivalent of the optical luminosity function. Current observations have shown that most of the H I mass in the local universe exists in galaxies of 10^9 to $10^{10} M_\odot$ in H I and that there is not a significant population of gas-rich dwarf galaxies that add to the H I mass density. The HIMF has been generally well fit using a Schechter function (Schechter, 1976):

$$\Phi(M_{\rm HI})dM_{\rm HI} = \theta^*(\frac{M_{\rm HI}}{{M_{\rm HI}}^*})^{-\alpha} exp(\frac{-M_{\rm HI}}{{M_{\rm HI}}^*})d(\frac{M_{\rm HI}}{{M_{\rm HI}}^*}) \tag{1.1}$$

The Schechter parameters are the slope of the low mass end (α), the characteristic 'knee' in the function (${M_{\rm HI}}^*$), and the normalisation (θ^*).

The HIMF can be based on the H I masses derived from a complete sample of optical galaxies, or on a sample of galaxies derived from a blind H I survey. This thesis will determine whether there is a difference in the shape of the HIMF from the two methods, and hence whether the two observational methods detect similar samples.

Figure 1.2 plots previously determined HIMFs. Mass functions derived from blind H I surveys are shown. There is some agreement between all functions in the 'mid' mass range (ie from 10^9 to 10^{10} $M_\odot$). However, the extremities of the mass function have been dominated by low number statistics in the samples.

Zwaan et al. (1997) derive a faint slope of α=1.2, from a driftscan survey using the Arecibo telescope, while Schneider et al. (1998) derive a faint slope that rises steeply at low masses ($< 10^8$ $M_\odot$), in a manner similar to the optical luminosity function derived by Loveday (1997). Both of these H I surveys used the same instrument and conducted a similar survey with a similar sample size. The difference in these results highlights the difficulties in determining the sensitivity and detectability of an H I survey. Zwaan et al. (1997) include all sources detected in the sidelobes of the telescope, and used a detectability function for corrections, while Schneider

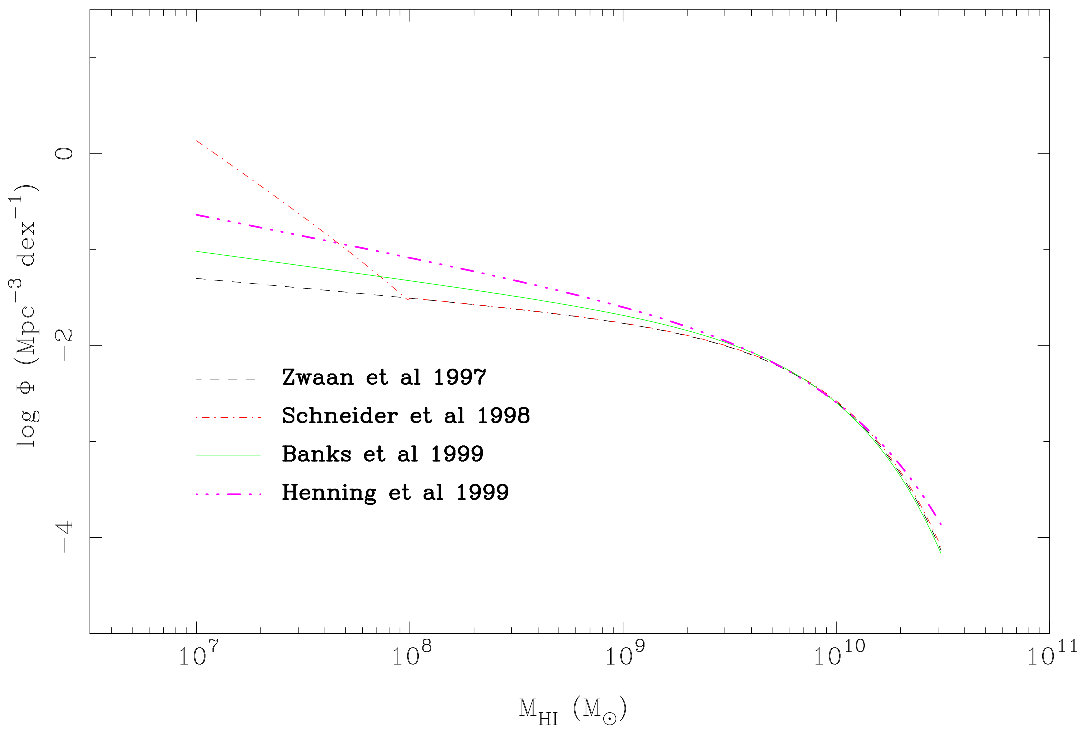

Figure 1.2: Previously determined H I Mass Functions from blind H I surveys

et al. (1998) only use detections detected in the main beam of the telescope. The detectability of galaxies in the noise is also modelled differently in the two samples, and it is likely that a combination of the different analysis techniques accounts for the discrepant low-mass slopes.

Banks et al. (1999) find an HIMF with a faint slope of $\alpha = 1.3$ from their survey of the Centaurus A Group, which is comparable to the Zwaan et al. (1997). The Zone of Avoidance survey (ZOA) which also uses the Multibeam receiver on the Parkes Radio Telescope (eg: Juraszek et al. (2000)), derived an H I mass function with a slope of 1.51 (Henning et al., 2000), which is much steeper than slopes determined from other blind H I surveys. The mass function derived from an optical catalogue by Briggs (1990) is similar to that of Zwaan et al. (1997) and Banks et al. (1999).

1.1.2 H I Content and Optical Morphology

The H I content of galaxies of different optical morphology has been widely studied using 21-cm observations of optically catalogued galaxies (eg: Kamphuis et al. (1996); Haynes and Giovanelli (1991); Fisher and Tully (1981a); Krumm and Salpeter (1980); Shostak (1978)).

H I is predominately found in 'normal' spiral galaxies, with the later types making up the difference (Rao and Briggs, 1993; Hoffman et al., 1989). The ratio of H I mass to Blue luminosity (M_{HI}/L_B) seems to increase with later galaxy type (Shostak, 1978), and Solanes et al. (1996), found that the Sb galaxies generally had a higher H I mass compared to the other spirals. The H I surface density of galaxies of different types does not seem to be predictable, with some studies finding a constant H I surface density for galaxies of any type (Broeils and Rhee, 1997; van der Hulst et al., 1987), and others finding an increase for later Hubble types (Roberts and Haynes, 1994).

The only blind H I study of the H I content of galaxies of different morphological type used a small region of the South Cap from HIPASS (Marquarding, 2000). Once again the majority of the H I mass was found in spiral galaxies, and the H I mass distribution of the spirals showed a Gaussian distribution centred on an H I mass of 10^9 $M_\odot$. The late–type galaxies in the sample were better fit with a Schechter function of faint end slope α=2.25. This is a similar mass distribution as seen by Hoffman et al. (1989) who looked at the H I content of optically identified galaxies in the Virgo cluster.

The H I content of dwarf galaxies varies with their morphology. Low luminosity irregular dwarfs are typically H I–rich, while dwarf spheroidal galaxies are rarely detected in neutral hydrogen (Carignan, 1999). Blue compact dwarf galaxies (BCDs) seem to be compact in both their H I and optical emission (Staveley-Smith et al., 1992; van Zee et al., 1995). Smoker et al. (2000) looked at UV excess and BCDs and found that the most H I rich galaxies were those with the lowest metallicities. The H I content of these galaxies increased with lower luminosity.

Irregular spatial and velocity distribution of neutral hydrogen can give a good indication of the history of interactions of a galaxy (Haynes et al., 1984). Such observations have been used in determining the interaction history of galaxies such as starburst or active galaxies, to determine if an interaction was the trigger for the active star formation or energy output seen today (eg: Mundell et al. (1995)). However, Jackson et al. (1987) looked at a sample of star-burst galaxies and found no evidence for interactions from the regular profiles seen.

The H I content of E and S0 galaxies is low, with only 10–15% of these galaxies typically detected in neutral hydrogen. It is commonly thought that the majority of the H I seen in these galaxies is from recent or past accretion from a gas–rich companion, as the observed H I distribution is typically irregular (Duprie and Schneider, 1996). Early–types galaxies that are optically classified as peculiar are more likely to contain H I (Bregman et al., 1992), as are those Ellipticals that lie in compact groups (Huchtmeier et al., 1995). There is some evidence for a small population of low–luminosity early–type galaxies and dwarf Ellipticals that have an intrinsic content of neutral hydrogen in them (Oosterloo and Morganti, 2000; Sadler, 1997), suggesting these galaxies had a different evolutionary path to more luminous, H I poor Ellipticals.

Low Surface Brightness (LSB) galaxies are defined as having a central surface brightness fainter than 23 B mag arcsec^{-2} (Impey and Bothun, 1997). They display the same range in size as is seen for high surface brightness galaxies, ie from dwarf to large disk galaxies (McGaugh, 1994; de Blok et al., 1995). LSB galaxies are typically gas-rich (Schombert et al., 1992) excepting the class of LSB dwarf elliptical galaxies, where few of these galaxies are detected in H I (Kraan-Korteweg and Tammann, 1979). For other LSB dwarf galaxies here is an indication that the gas richness increases with decreasing luminosity (Staveley-Smith et al., 1992). This means that blind H I surveys may be ideal way of finding these LSB galaxies. LSB galaxies might be missed in typical optical surveys (eg: Bothun et al. (1997)), so any HIMF based on optical catalogues may be biased against these gas-rich objects.

1.1.3 Dependence on Environment

The H I content of galaxies depends to a degree on the environment in which the galaxy resides (Haynes et al., 1984). Galaxies near the centre of clusters are likely to be H I deficient, (Giovanelli and Haynes, 1985; Solanes et al., 2000) and preliminary results from HIPASS support this idea, with a recent survey of the Fornax cluster finding an H I deficiency in the cluster galaxies (Waugh, 2000). Galaxies with medium to large optical sizes in clusters tend to be more H I deficient than galaxies of a smaller optical size (Valluri and Jog, 1991). Gas removal effects do not seem to be as acute in smaller galaxy groups. Giuricin et al. (1985) studied 230 galaxies in groups and found their H I content did not correlate with the distance from the group center. There is some indication that galaxies residing in regions of low density are more H I rich than higher density regions (Huchtmeier et al., 1997). This conclusion seems reasonable, as the probability of interactions is much reduced in a low galaxy density environment.

To determine whether a population of H I clouds exists that are not optically visible, blind H I surveys have been conducted in many of environments: voids, clusters, groups and the general field. No H I clouds have been found to date that are not gravitationally bound to systems with optical light. From the surveys in voids, it can be concluded that H I galaxies of 'normal' mass ($\sim 10^9 M_\odot$) do not fill the voids. The first blind H I survey (Shostak, 1977) searched several hundred degrees of optically blank sky, and one source was detected. This was later found to be a High Velocity Cloud that was gravitationally bound to the Galaxy. A search for massive H I clouds ($> 10^{10} M_{\odot}$) in parts of the Perseus-Pisces and the Hercules voids (Krumm and Brosch, 1984) did not find any clouds, while Hoffman et al. (1992) surveyed a nearby void for gas-rich dwarf galaxies and also made no detections. Henning and Kerr (1989) did find one small LSB spiral in an H I search of the void near (R.A., Dec.) = (24h,31°), but they did not find a large population of these galaxies. The only H I mass range not covered is small H I masses, of $M_{\rm HI} \lesssim 10^7 M_{\odot}$, where the previous surveys have not been sensitive.

H I surveys in groups of galaxies have placed fairly good limits on the H I content of these environments and the existence of H I clouds. Fisher and Tully (1981b) did not find any clouds of neutral hydrogen in the direction of the M81 group, to a limit of $\sim 10^8 M_{\odot}$. Materne et al. (1979) determine similar limits on the existence of H I clouds in the groups NGC 1023 and CVnII, with no detection of clouds in their survey. Similarly, Lo and Sargent (1979) detected no H I clouds in the M81, NGC 1023 and CVnI groups, with similar limiting masses. Clusters have also been searched using blind H I surveys, even though, so this environment may not be the ideal place to search for intergalactic clouds of neutral hydrogen. Barnes et al. (1997) searched the Fornax and Centaurus clusters, and did not find any H I detections without optical counterparts, to the limit of $\sim 10^8 M_{\odot}$.

1.1.4 High Velocity Clouds

One class of H I objects still has a contentious origin: the High Velocity Clouds (HVCs). High velocity clouds typically have heliocentric velocities of -300 km s^{-1} < V < 400 km s^{-1}. A large number of HVCs can be directly linked with the Magellanic Clouds and Magellanic Stream (Putman et al., 1998), but there is a population of comparatively isolated and compact clouds that are unlikely to be of Magellanic origin. Blitz et al. (1999) suggest these clouds may have been formed during the formation of our Local Group, and Braun and Burton (1999) suggest the clouds may have distances up to 1 Mpc. However, from a search for molecular gas in HVCs, Combes and Charmandaris (2000) find that the clouds they studied were more likely to be at a distance of 20 kpc than 1 Mpc. Methods for determining the distance to HVCs include looking for absorption in stellar spectra, or observing their H-α emission, and applying a model for the ionising field of the Galaxy (eg: Weiner and Williams (1996), Bland-Hawthorn and Maloney (1999)). The former method can provide an upper limit on the distance to the cloud, although a non-detection does not necessarily provide a lower limit. The latter method is sensitive to errors in the ionisation model and it is yet to be proven whether this is an effective way of determining the distance to HVCs.

1.2 The H I Parkes All Sky Survey

The H I Parkes All Sky Survey (HIPASS) has been conducted at the Parkes Radio Telescope in NSW, Australia, from 1997. Table 1.2 gives the observing parameters for HIPASS. The survey uses a new, 13-beam 'Multibeam' receiver, mounted on the 64-m telescope. HIPASS has surveyed the whole Southern sky south of declination $+2^{o}$, and is now continuing into the north to $+25^{o}$. The telescope scans the sky in declination strips 8^{o} long, and the effective integration time at each point on the sky at the completion of observations is 460 seconds.

Parameter	HIPASS Value
Sky Coverage	$\delta < 2^{o}$
Southern Sky Extension	$\delta \leqslant 25^{o}$
Integration time per beam	460 s
FWHP beamwidth	$14.'0$
Gridded FWHP beamwidth	$15.'5$
Velocity Range	-1200 to 12,700 km s^{-1}
Velocity Resolution	18 km s^{-1}
Channel Separation	13.19 km s^{-1}
Theoretical RMS noise	13 mJy $beam^{-1}$
3σ H I Mass Limit[a]	$10^6 D^2_{Mpc} M_{\odot}$

[a] For $\Delta V = 100\,\mathrm{km\ s^{-1}}$

Table 1.2: HIPASS Observing Parameters

The bandpass for the survey is 64 MHz wide, with 1024 channels. This provides velocity coverage from $V_{start} = -1200\mathrm{km\ s^{-1}}$, to $V_{end} = 12,700\mathrm{km\ s^{-1}}$, with a channel width of 13.19 km s^{-1} and a velocity resolution after gridding of 18 km s^{-1}. The data is bandpass corrected, then gridded into $8^{o} \times 8^{o}$ cubes. There are 388 data cubes in the full HIPASS survey, and a further 150 for Northern HIPASS. Figure 1.3 shows a 3-dimensional view of part of a HIPASS datacube. Visible are several galaxies, Galactic H I emission, and an H I recombination line.

The theoretical RMS noise for the final datacubes is 13 mJy $beam^{-1}$, and the H I mass limit for the survey is $\sim 10^6 D^2_{Mpc}\ M_{\odot}$, where D is the distance to the source[1]. The beam size

[1] Assuming an H I linewidth of 100 km s^{-1}

for the gridded data is approximately 15.5 arcminutes, however due to the scanning motion of the telescope across the sky, the positional errors of detections in the data are typically less than 3 arcminutes, depending on the H I flux of the source (Barnes, 1998b).

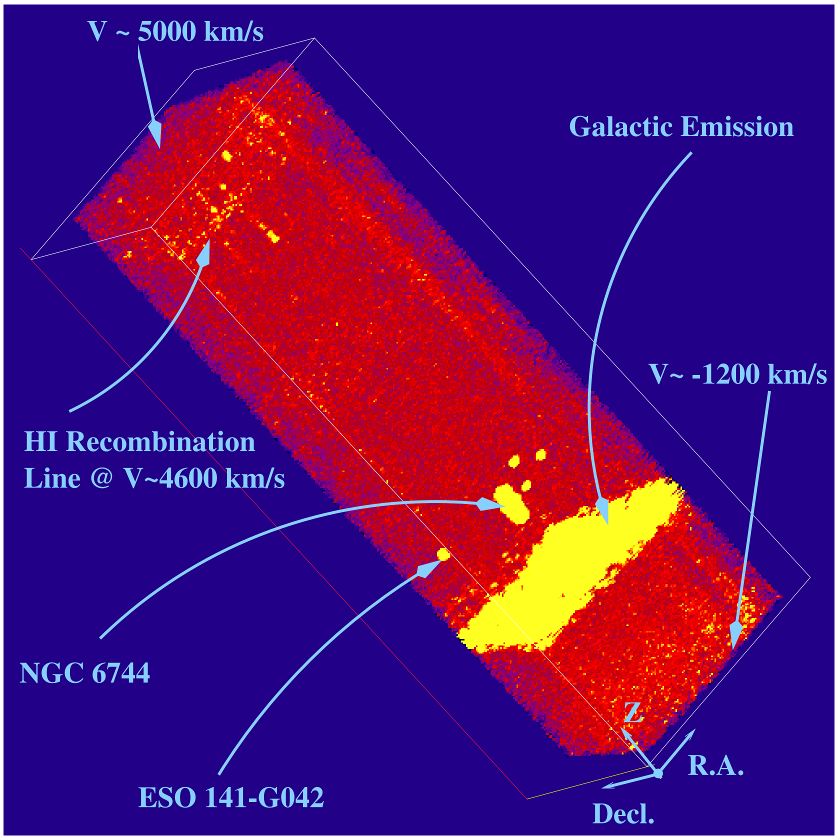

Figure 1.3: HIPASS cube H041. The velocity range shown is from -1200–5000 km s^{-1}. Galaxies and Galactic emission can be seen.

1.2.1 HIPASS Data Reduction

The bandpass removal is conducted in real time at the Parkes Telescope (Barnes et al., 1998), using the data reduction package LIVEDATA. The usual method of bandpass removal is to remove a reference spectrum from a target spectrum, width the spectra generated by observing on and off a source. However, due to the scanning motion of the telescope across the sky in the HIPASS blind observations, the method of bandpass removal was modified. Each spectrum observed was divided by a reference spectrum. The reference spectrum was derived from a combination of earlier and later spectra. These reference spectra were acquired nearby in time to the original spectrum to avoid temporal variations of the bandpass. Sidelobes are generated during this stage of processing, which affects large extended sources such as HVCs and nearby galaxies.

Once spectra for a region of sky are obtained, they are gridded into datacubes, using the package GRIDZILLA (Barnes, 1998a). Each cube is $8^o \times 8^o$ in size, and the pixel size is $4' \times 4'$. Each pixel contains weighted data from 90 possible spectra, within a radius of 6 arcmin. To avoid interference in the data, the median of the spectra is used to determine the flux for each pixel in the cube. The reader is referred to Barnes et al. (2000) for explicit details of the HIPASS data reduction techniques.

1.2.2 HIPASS Data Improvement

The online bandpass removal of the data from the survey has been inefficient in the vicinity of bright continuum sources. The bandpass removal is over-estimated near bright continuum sources, which results in the baseline being negative at the low velocity planes of a cube, and positive in the high velocity planes of a cube (eg. see Figure 1.4(a)). This means that if an H I source is situated near a continuum source, it is likely to be missed in a search of the datacube. Wright and Stewart (2000) developed an algorithm to correct the bandpass for this effect. The algorithm, LUTHER, makes a polynomial fit to the continuum sources in a data cube, then subtracts this template from each of the spectra in the gridded cube. Figure 1.4(b) shows the previous spectrum which has been corrected using LUTHER. An H I galaxy is now readily visible in the spectrum due to the improved bandpass removal.

1.2.3 Definitions of H I Parameters

A number of measured and derived values for the H I detections of each galaxy are used throughout this thesis.

(1) *Measured Values*:

The coordinates, Right Ascension (R.A.) and Declination (Dec.) (J2000), of a source are produced from IMFIT, by fitting a Gaussian to a velocity integrated flux map. Galactic Longitude (l) and Galactic Latitude (b) are determined using the coordinate transform program, COCO. The recessional velocities (V) are heliocentric velocities in the optical convention unless stated otherwise (km s^{-1}). The 20% and 50% velocity widths (ΔV_{20} and ΔV_{50}) are the width of the H I profile at 20% and 50% of the peak flux (km s^{-1}). The peak flux (S_p) is defined as the

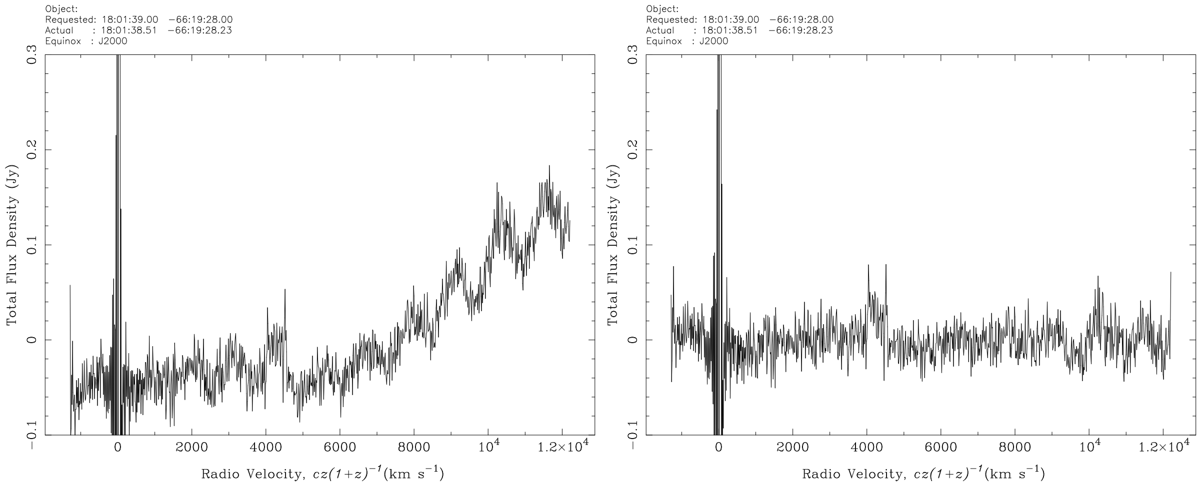

Figure 1.4: The LUTHER algorithm: (a) The first panel shows the spectrum at the position of a continuum source in the original HIPASS data cubes, and (b) the second panel shows the spectrum once the LUTHER algorithm has been run on it. A galaxy can easily be seen in the second panel at ~ 4100 km s^{-1}.

maximum flux point in the galaxy profile (Jy)[2]. The integrated flux (S_{int}) is the sum of all flux over the entire velocity width of a profile (Jy km s^{-1}), and is also referred to as the total flux of a galaxy.

(2) *Derived Values*:

The derived values use simple cosmological models, especially when calculating distances and masses. The value of Hubble's constant used in this thesis is $H_0 = 75$ km s^{-1} Mpc^{-1}. For transparency in equations which rely on H_0, h_{75} is $H_0/75$.

• The distance to a galaxy is generally derived using the Hubble flow, and a corrected value for the velocity of a galaxy.

$$D = \frac{V_{LG}}{75\, h_{75}\ \text{Mpc}} \tag{1.2}$$

where V_{LG} is the Local Group velocity, corrected for the motion of the Sun in the Galaxy, and the Galaxy in the Local Group.

• The H I mass of a galaxy is derived assuming that the galaxy is transparent, and has a typical

[2]The peak flux is dependent on the spatial binning of the data, and the implications of this are discussed in a later chapter

spin temperature. The mass is calculated using:

$$M_{\rm HI} = 2.356 \times 10^5 D^2 \int S\,dv\,M_\odot \tag{1.3}$$

where S is the spatially integrated flux in Jy, D is the distance to the galaxy in Mpc, and the integral is over the velocity width of the galaxy in km s^{-1}.

• The column density of a galaxy is derived when higher resolution synthesis data has been obtained, using:

$$N_{HI} = 1.81 \times 10^{18} \int T(K)\,dv \;\; \text{atoms}\,\text{cm}^{-2} \tag{1.4}$$

where

$$T(K) = \frac{1.36\,S\,\lambda^2}{\theta_\alpha \theta_\beta} \tag{1.5}$$

θ_α and θ_β are the major and minor axes of the beam in arcseconds, λ is the wavelength of the observations in cm, and S is the spatially integrated flux in units of mJy beam^{-1}.

1.3 Outline of this Thesis

This thesis presents the analysis of a blind H I survey from HIPASS. The main data sample and catalogue are described in Chapter 2, including survey techniques, follow-up observations, and the characteristics of the blind H I survey. An automated galaxy finder is described in Chapter 3, including simulations, and tests of the finder on real data. Chapter 4 describes methods for determining the distribution of H I mass, and HIMFs, including simulations of the effectiveness of the different HIMF estimates. Chapter 5 presents the H I Mass Function for the sample. The completeness of the catalogue is also determined in this chapter, and the variation in the HIMF as a function of galaxy type is investigated. One particularly interesting galaxy, with no visible optical counterpart is discussed in detail in Chapter 6, and a population of such clouds is discussed as a whole. New galaxies from the survey region that have been observed in at the Australia Telescope Compact Array are discussed in Chapter 7. The conclusions are presented in the final chapter.

Chapter 2

South Celestial Cap Sample

A catalogue of candidate galaxies was determined in the South Celestial Cap (SCC) region of the sky, $\delta <$ -62°. This was the first region of the sky to be completed to the full sensitivity of HiPASS. The SCC region represents $\sim$ 11% of the full HiPASS survey. The solid angle covered in the region is 0.74 steradians, and the total volume covered over the full velocity range is $\sim 1 \times 10^6$ Mpc^{-3}.

The candidate galaxy list was formed by visually searching the data cubes. Each of the candidates was then checked in the NASA/IPAC Extragalactic Database[1](NED). The final catalogue contains 533 galaxies, and 140 of these (26%) have no previous optical identification. In addition, new redshifts were determined for 124 previously catalogued optical galaxies. All galaxies found in this region had a positive radial velocity.

Optical images for all of the galaxies were obtained from the First Generation Digital Sky Survey[2](DSS I); for those with no optical counterpart in the First Generation Survey, the image from the Second Generation Digital Sky Survey (DSS II) was obtained. A small number of the new detections lying outside the Zone of Avoidance and Magellanic Clouds do not have obvious optical counterparts in either of the DSS images. One of these detections, HIPASS J1712-64, is discussed in Chapter 7. Optical observations and higher resolution H I observations were made for a number of detections. These observations as well as the general properties of the sample will be outlined in this chapter. Optical images and spectra for the full SCC Sample can be found in Appendix B.

[1]The NASA/IPAC Extragalactic Database (NED) is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

[2]The Digitised Sky Survey, provided by the Space Science Institute, based on photographic data from the UK Schmidt Telescope.

2.1 Detecting Galaxies in HIPASS cubes

Galaxies were detected by searching through the cubes visually. The region searched consists of 45 $8^o \times 8^o$ cubes, with 1024 velocity planes in each cube. Each cube was searched at least three times, over different searching axes. The visualisation program KVIEW (Gooch, 1995) was used to display a cube in two dimensions, and allowed the user to step through the third dimension of the cube. Firstly, a cube was displayed in R.A. and Dec., and the cube was stepped through the velocity axis searching for flux peaks. The cube was then displayed as R.A. and Velocity, and was stepped through in Dec., and finally the cube was displayed as Dec. and Velocity and stepped through R.A. The first display method was the best at finding bright, narrow-width galaxies, while the other two display methods aided in finding fainter galaxies with larger velocity widths. Figure 2.1 shows the three display modes of KVIEW.

The number of galaxies in any particular cube was not constant: Large-scale structure in the galaxy distribution was observed, with some cubes containing up to 30 galaxies while others only had one galaxy.

The LUTHER bandpass removal algorithm was developed after the cubes had been searched (Wright and Stewart, 2000). Once LUTHER, was run on all of the cubes, the cubes were re-searched for galaxies near continuum sources. Approximately 40 new galaxies were found in these cubes, especially at the high velocity end of the cubes where the bandpass removal had previously been inadequate.

2.2 Galaxy Selection Criteria

To generate a consistent list of H I detections, it was necessary to impose some selection criteria when cataloguing the galaxies. The selection criteria for this visual sample were: (1) a detection must be easily visible above the noise ($\sim 3\sigma$ in peak flux), (2) the detection must have spatial extent either greater or equal to the beamsize (excludes interference), and (3) it must be visible over two or more velocity planes. Negative velocity regions were searched, but the Galactic region from -100 km s^{-1} $\lesssim$ V $\lesssim$ 200 km s^{-1} was not searched for galaxies. Other regions excluded from the search were the Magellanic Stream, and the Large and Small Magellanic Clouds. The Large Magellanic Cloud (LMC) lies at R.A. = 05h23m34.5s, Dec. = –69d45m22s (J2000), at a velocity of 278 km s^{-1}. The Small Magellanic Cloud (SMC) lies at R.A. = 00h52m44.8s, Dec. = –72d49m43s (J2000), at a velocity of 158 km s^{-1}. The Magellanic Stream and Leading Arm can be seen in Figure 2.2, which shows a peak flux map of the SCC region,

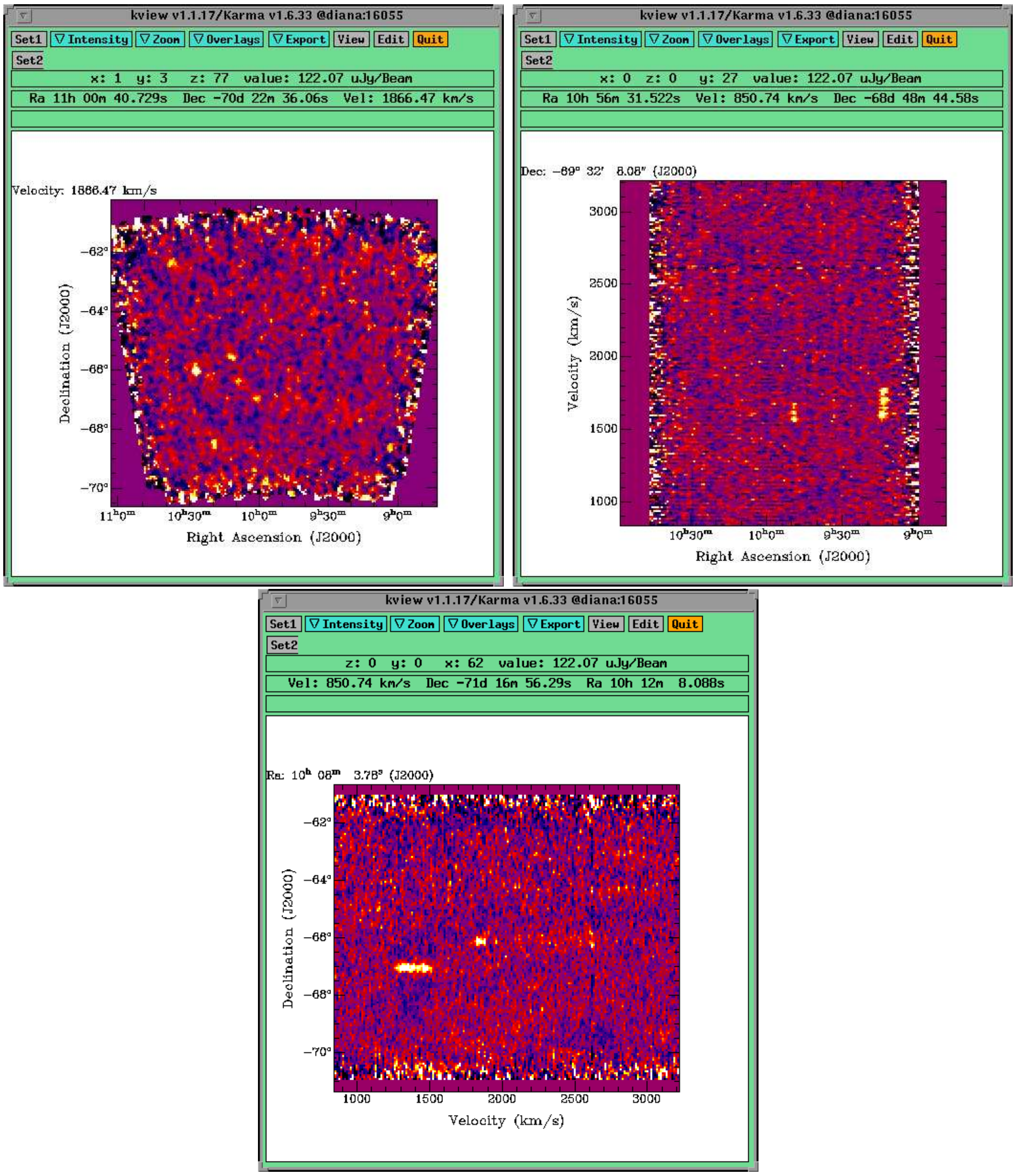

Figure 2.1: KVIEW display modes.

between a velocity of 200 km s^{-1} and 800 km s^{-1}. Some regions obviously contained High Velocity Clouds, with a continuous distribution of resolved, narrow width objects, with V $\lesssim$ 450 km s^{-1}; these regions were not searched. Any detection that was found in the HVC velocity range but which had no other HVCs in the same cube was catalogued as a galaxy. Therefore a small number of the detections included in the sample have ambiguous identities[3]. These particular detections and their implications are discussed further in Chapter 6. In addition it is likely that some nearby galaxies are not included in the catalogue if their H I emission lies near that of the HVCs. Higher spatial and velocity resolution observations will be needed to distinguish these galaxies from the general HVC population.

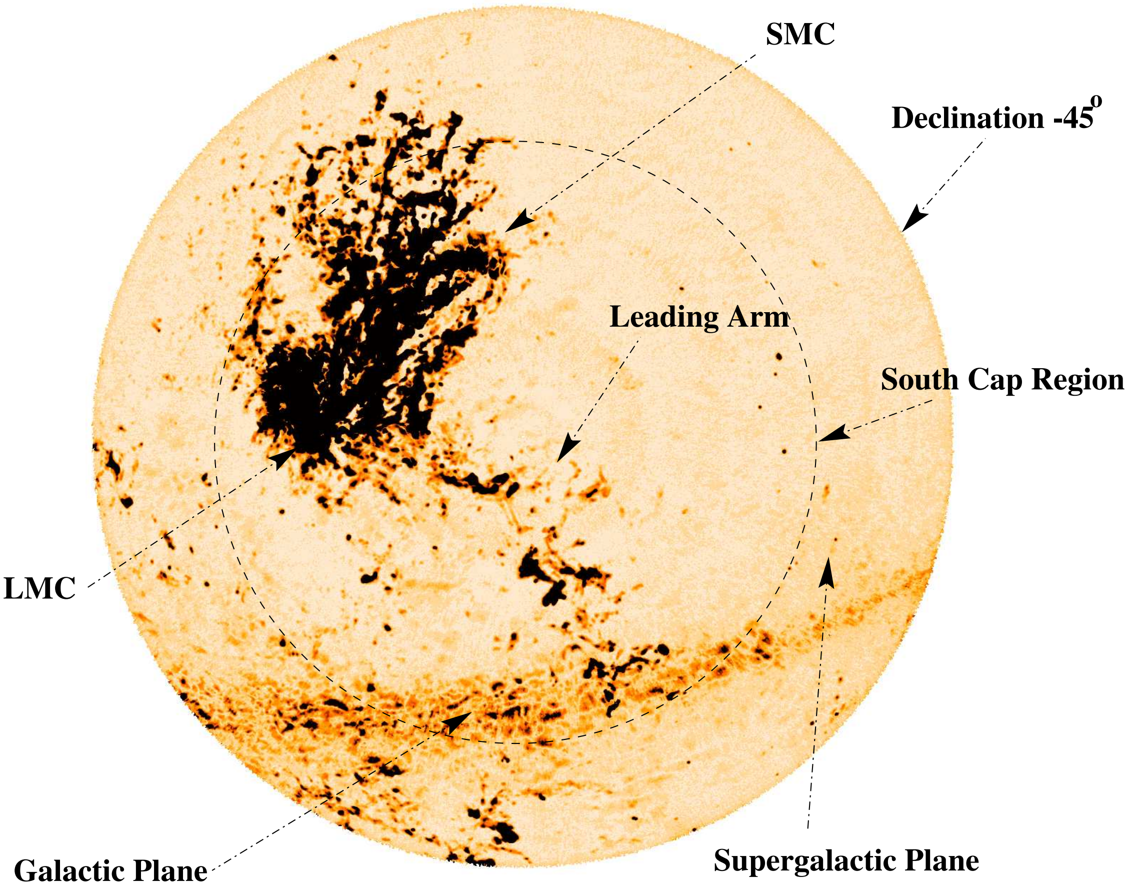

Figure 2.2: Peak flux map of the velocity range 200 km $s^{-1} \lesssim$ V $\lesssim$ 800 km s^{-1}, showing the Galaxy and Magellanic Clouds. The South Celestial Pole is in the centre of the image.

2.3 Galaxy Detection Consistency

Since visual cataloguing is subjective, the SCC sample has been independently checked by two methods. Marquarding (2000) catalogued the SCC region up to a velocity of $\sim$ 2500 km s^{-1}.

[3]7 detections are ambiguous as to their identity: They are noted as such in the catalogue.

Comparison with the catalogue generated for Marquarding's thesis found 95% agreement between the two lists. Each catalogue contained ~ 10 galaxies that were not in common, so the original lists were modified to include these. The second check was made using an automatic galaxy finder on the region (see Chapter 3). Ivy Wong[4] used automatic galaxy finder lists[5] as a basis for cataloguing all galaxies in approximately 65% of the SCC region, through the entire HIPASS velocity range. Wong catalogued 6 faint galaxies (peak flux ~ 40 mJy) that were not included in the original SCC catalogue. So there is 99% agreement between the galaxies catalogued by Wong and the SCC sample in the corresponding regions, and the extra 6 galaxies were added to the SCC sample.

2.4 Parameterization of the Galaxies

Parameters for each of the detections were measured in a semi-automated manner. A script was developed to step through the fitting process, accepting user input at each stage. The first stage was to fit a Gaussian to velocity integrated map of each detection to determine both the central position of the galaxy as well as the spatial extent of the H I. This fitting was done using the MIRIAD task IMFIT. The central position was then used to generate an integrated spectrum of the detection, using a box size based on the extent of the H I. The spectrum was generated using the MIRIAD task MBSPECT, which also gave a measurement of the peak and integrated flux of each detection, as well as the 50% and 20% velocity width, RMS noise and heliocentric velocity. The heliocentric velocity was measured as the central velocity at the 20% velocity width. Spectra of each galaxy in the SCC sample can be found in Appendix B.

The consistency of these fluxes was checked with galaxies previously observed in H I. Two catalogues were used to test the fluxes. Huchtmeier and Richter (1989) compiled a catalogue of 9500 galaxies, including H I data from nearly 900 other sources. Forty-one galaxies in the SCC sample had entries for total integrated flux in this catalogue. The second comparison catalogue was compiled by Mathewson et al. (1992), who published H I data from pointed observations of 551 galaxies observed with the Parkes Telescope. Thirty-three galaxies from this catalogue were present in the SCC sample. Figure 2.3(a) shows a comparison of integrated flux for the overlapping 74 galaxies from both catalogues. The fluxes from the homogeneous Mathewson et al. (1992) catalogue agree very well with the HIPASS derived fluxes, while the heterogeneous Huchtmeier and Richter (1989) fluxes agree to a lesser extent. There is no relation between the

[4] *private communication*
[5] MultiFind, see Chapter 3.

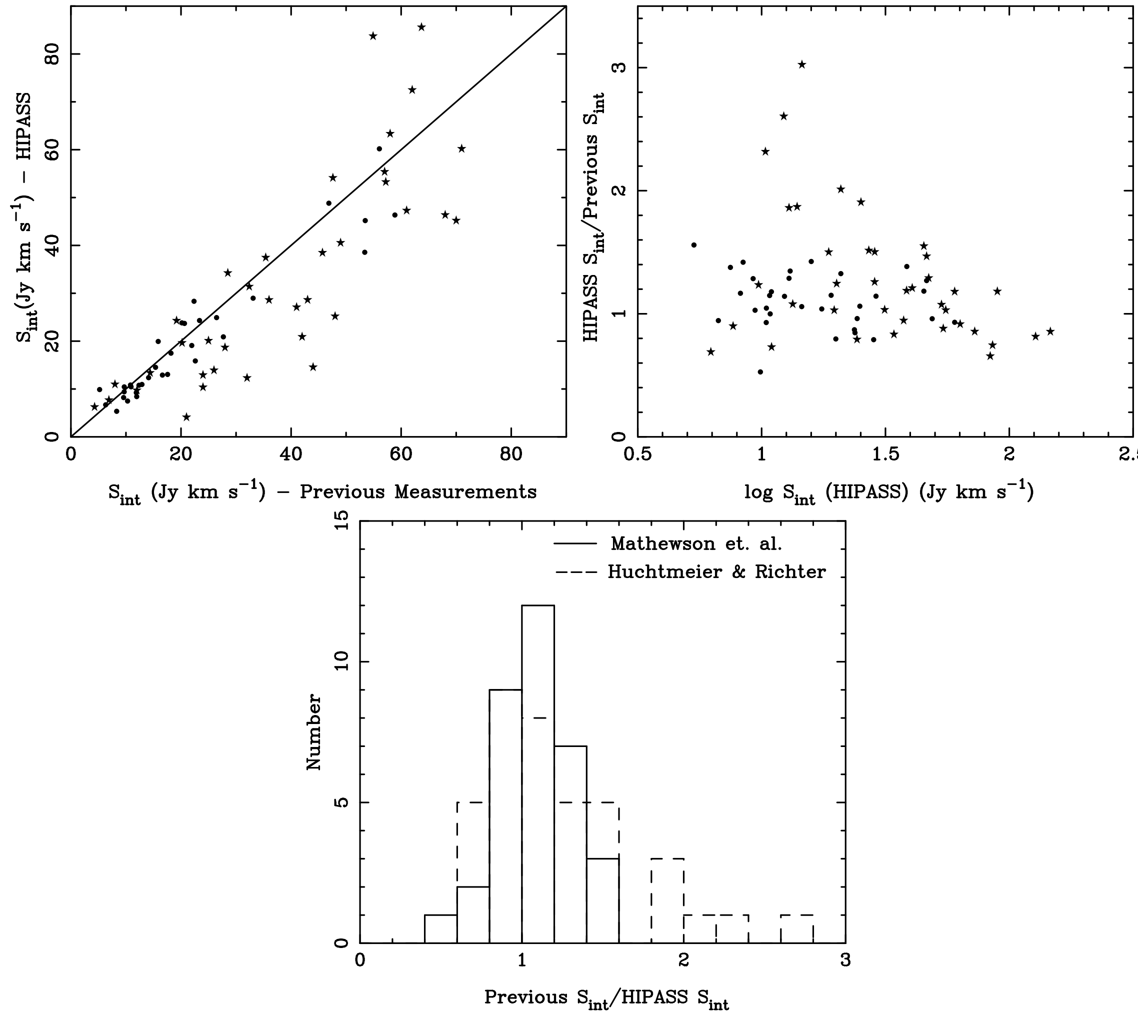

Figure 2.3: Integrated flux measurements from the Huchtmeier and Richter (1989) catalogue (stars) and the Mathewson et al. (1992) catalogue (circles), compared to HIPASS measurements. Panel (a) [Top left]: The Mathewson et al. (1992) measurements compare quite well with HIPASS fluxes, while the Huchtmeier and Richter (1989) compare to a lesser extent. The HIPASS fluxes tend to be lower than the previous measurements. Panel (b) [Top right]: No relationship is seen between the agreement of HIPASS fluxes with previous measurements and the total amount of flux. Panel (c) [Bottom]: Most of the Mathewson et al. (1992) fluxes agree with the HIPASS fluxes within 20%. The scatter is larger for the Huchtmeier and Richter (1989) fluxes.

flux determined and the correlation of previous fluxes to the HIPASS fluxes (see Figure 2.3(b)). The fluxes agree to $\sim$ 20% for the Mathewson et al. (1992) comparison and significantly worse for the Huchtmeier and Richter (1989) (Figure 2.3(c)).

2.5 Velocity Corrections and Distance Estimates

The recessional velocity determined for galaxies from the HIPASS data are heliocentric values. The galaxy velocities in this thesis are defined from the optical convention, $V = c\,\frac{\Delta\lambda}{\lambda_0}$, rather than the radio convention $V = c\,\frac{\Delta f}{f_0}$. The galaxy recessional velocities presented in this Chapter have been corrected for the motion of the Sun around the Galaxy, and also the motion of the Galaxy in the Local Group. The correction used is the IAU convention,

$$V_{LG} = V + 300 \sin l \cos b \tag{2.1}$$

where V_{LG} is the Local Group velocity, V is the heliocentric velocity in optical convention, and l and b are the Galactic Longitude and Latitude. This correction does not take account of peculiar motions due to gravitational interactions with other galaxies, so is an approximation. At low velocities, especially where V $<$ 500 km s^{-1}, the distance that can be derived from these corrected velocities is only an estimate. Different distance corrections and implications will be discussed further in subsequent chapters.

2.6 Synthesis follow-up Observations

Follow-up H I synthesis observations were made at the Australia Telescope Compact Array[6] (ATCA), in NSW, Australia. This telescope comprises of six antennas lying on a 6 km East-West array. Earth rotation is required to obtain full U-V coverage. These observations are necessary in some cases due to the large beam size of the Parkes Telescope. Often there will be two or more optical galaxies within the Parkes beam of an H I detection, so higher resolution follow-up observations are important to determine which of the galaxies in the beam contribute to the H I flux (eg. Figure 2.4(a)). A second situation which requires higher resolution follow-up, is when no obvious optical counterpart is seen in the field. In this case, a good H I position is needed for follow-up optical searches (eg. Figure 2.4(b)). 44 candidates were observed with the ATCA, using either the 750m or 375m arrays, of which 27 galaxies were detected. The observations were made using an 8MHz bandpass with 512 channels, which gave a channel separation of 3.3 km s^{-1}, and a FWHM resolution of 4.0 km s^{-1} before smoothing. The data were reduced using the MIRIAD data reduction package. A secondary calibrator was observed

[6]The Australia Telescope Compact Array is funded by the Commonwealth of Australia for operation as a National Facility managed by CSIRO.

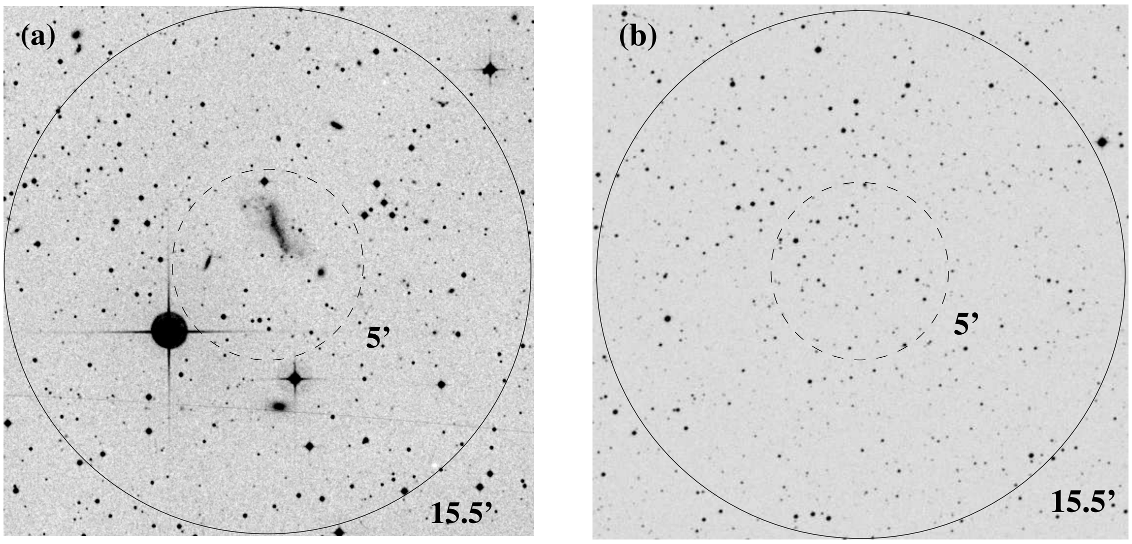

Figure 2.4: Galaxies within the Parkes beam: (a) Several galaxies of unknown redshift within the Parkes beam. (b) A field without any bright optical counterpart. High resolution synthesis imaging is required to determine (a) which galaxy/ies are producing the H I emission, and (b) the position of an optical counterpart to the H I emission. The outer circle shows the approximate HIPASS gridded beam size, and the inner circle indicates the approximate accuracy for HIPASS positions.

with each galaxy to calibrate the phase, and the primary calibrator, PKS 1934-638 was used to set the primary flux scale. The data were edited and calibrated, and the bandpass was removed. Bright continuum sources were then removed by fitting and subtracting straight lines to the line-free channels in the visibility domain using the MIRIAD task UVLIN. The data were then cubed using natural weighting to give the source the highest signal-to-noise possible. The cubes were then CLEANed until the absolute maximum residual fell below 3 times the theoretical noise for the cube (typical RMS $\sim$ 7 mJy/beam). The data were then RESTORed, and HANNING smoothed.

Table 2.1 gives the H I derived parameters for the galaxies from the ATCA follow-up observations. The columns are:

(1) HIPASS name.

(2) and (3) Central Right Ascension and Declination (J2000) from the ATCA data. This position represents the highest column density point in each galaxy.

(4) Heliocentric velocity - optical correction has been applied.

(5) and (6) 50 % and 20% velocity width, derived from the MIRIAD task, MBSPECT (in km s^{-1}).

(7) Peak Flux - derived from the spatially integrated Hanning smoothed spectrum, using MBSPECT (in mJy).
(8) Integrated Flux - also derived from MBSPECT (in Jy km s^{-1}).
(9) Observation month and Year.
(10) ATCA array used for the observations.

17 candidates were not detected in the ATCA observations. 5 candidates had a low peak flux ($\sim$ 40 mJy), which may have been too faint to detect. 5 candidates were the ambiguous HVC/galaxy detections with low systemic velocities ($\sim$ 400 km s^{-1}). These are easily detected in the HIPASS data, but were resolved out in the ATCA data (all were observed using the 750m array). The remaining 7 candidates were determined to be interference in the original HIPASS data, and were subsequently taken out of the SCC catalogue. These candidates were included in the original sample from cubes which had a gridding error when they were first produced. Once this error was realised and corrected, the cubes were re-searched, and these 7 candidates along with several others were excluded from the final catalogue.

Optical images of all the newly detected galaxies with H I column density contours overlaid are shown in Chapter 7. The beams for these observations can be seen in the bottom left hand corner of the images. The newly discovered galaxies are mostly faint dwarf galaxies, or galaxies near bright stars, as expected. Several interesting galaxies, including some Low Surface Brightness galaxies were detected, and they are described in detail in Chapter 7.

2.7 Comparison with HIPASS Parameters

The positional accuracy of the original HIPASS detections can be compared with the follow-up ATCA observations. Figure 2.5(a) shows the difference in position between the ATCA and HIPASS observations. No correlated offset can be seen between the two sets of H I positions, and the RMS of the scatter for the offset is 1.9 arcmins including both R.A. and Dec. The RMS scatter for the R.A. measurements is 2.1 arcmin, and 1.7 for Dec. This compares to the predicted range of positional errors from the gridding process of $\sim$ 3 arcmin for a 3σ detection to $\sim$ 1 arcmin for a 10σ detection (Barnes, 1998b). The detection with the largest ATCA to HIPASS offset is lying on an interference spike in the original HIPASS cube, which hampered the calculation of its central position.

Figure 2.5(b) shows the ATCA and HIPASS measurements for the 50% velocity width. The two measurements agree well for most of the sample. There is one galaxy for which the velocity

Source	**RA** J2000	**Dec** J2000	**Vel** km s^{-1}	$\mathbf{\Delta V_{50}}$ km s^{-1}	$\mathbf{\Delta V_{20}}$ km s^{-1}	$\mathbf{S_p}$ mJy	$\mathbf{S_{int}}$ Jy.km s^{-1}	**Obs**	**Arr**
HIPASS J0039-76	00 39 08	-76 20 12	1739	56	92	76	3.1	Jun99	750
HIPASS J0116-63	01 16 48	-63 30 01	2301	119	136	17	1.3	Jun99	750
HIPASS J0159-83	02 00 11	-83 59 16	3498	93	109	45	3.0	Dec98	750
HIPASS J0209-75	02 09 01	-75 56 11	1653	175	188	48	5.14	Oct97	375
HIPASS J0305-69	03 05 24	-68 56 31	1364	117	131	33	2.4	May00	750
HIPASS J0411-70	04 11 38	-70 14 13	7155	65	79	30	1.1	May00	750
HIPASS J0511-72	05 11 34	-72 18 21	1525	72	92	64	3.8	Oct97	375
HIPASS J0554-71	05 54 42	-71 55 50	1473	58	64	44	2.2	Oct97	375
HIPASS J0729-75	07 28 43	-75 03 21	1118	80	136	92	6.3	Jun99	750
HIPASS J0751-62†	07 51 33	-62 19 59	6106	56	77	24	1.3	Dec99	375/750
HIPASS J0908-64	09 08 21	-64 42 10	1839	39	73	73	3.4	Jun99	750
HIPASS J1005-73	10 05 03	-73 51 15	1243	42	58	152	5.7	Jun98	750
HIPASS J1150-69	11 51 35	-69 48 26	6810	26	73	35	1.2	May00	750
HIPASS J1242-77	12 42 06	-77 55 19	2654	64	76	57	2.9	Jun98	750
HIPASS J1247-77	12 47 35	-77 34 53	413	29	50	122	4.1	Jun98	750
HIPASS J1339-72	13 39 23	-72 54 54	3220	136	149	32	2.4	Jun98	750
HIPASS J1407-75	14 07 28	-75 53 24	2795	108	145	89	8.0	Jun98	750
HIPASS J1456-72	14 56 42	-72 33 10	3110	121	154	41	3.2	May00	750
HIPASS J1512-72	15 11 56	-72 51 28	3000	408	424	211	61.1	May00	750
HIPASS J1648-69	16 48 13	-69 08 14	4683	313	352	105	16.1	Jun99	750
HIPASS J1650-62	16 49 49	-62 33 05	4226	89	122	42	2.7	May00	750
HIPASS J1705-68	17 05 55	-68 57 33	7826	288	348	21	3.5	Dec99	750
HIPASS J1816-67	18 15 45	-67 01 36	3312	183	190	58	7.0	Jun99	750
HIPASS J2039-63	20 39 01	-63 46 15	1644	52	65	91	4.1	Dec99	750
HIPASS J2214-67	22 13 46	-67 24 19	3178	41	56	36	1.2	Jun99	750
HIPASS J2247-73	22 48 14	-73 48 49	3378	154	167	13	1.3	Oct97	375
HIPASS J2309-74	23 08 40	-74 49 52	2073	98	147	19	1.4	Jun98	750

†Table values for HIPASS J0751-62 derived from combined image from both 375m and 750m observations

Table 2.1: Derived parameters for the 27 galaxies detected with the ATCA

width determined from HIPASS is much higher than the ATCA derived velocity widths. This galaxy, HIPASS J1650-62, has a low signal-to-noise, and the difference in measurements is due to noise in the spectrum. The integrated flux detected in the ATCA observations is just 22% of the flux detected in HIPASS.

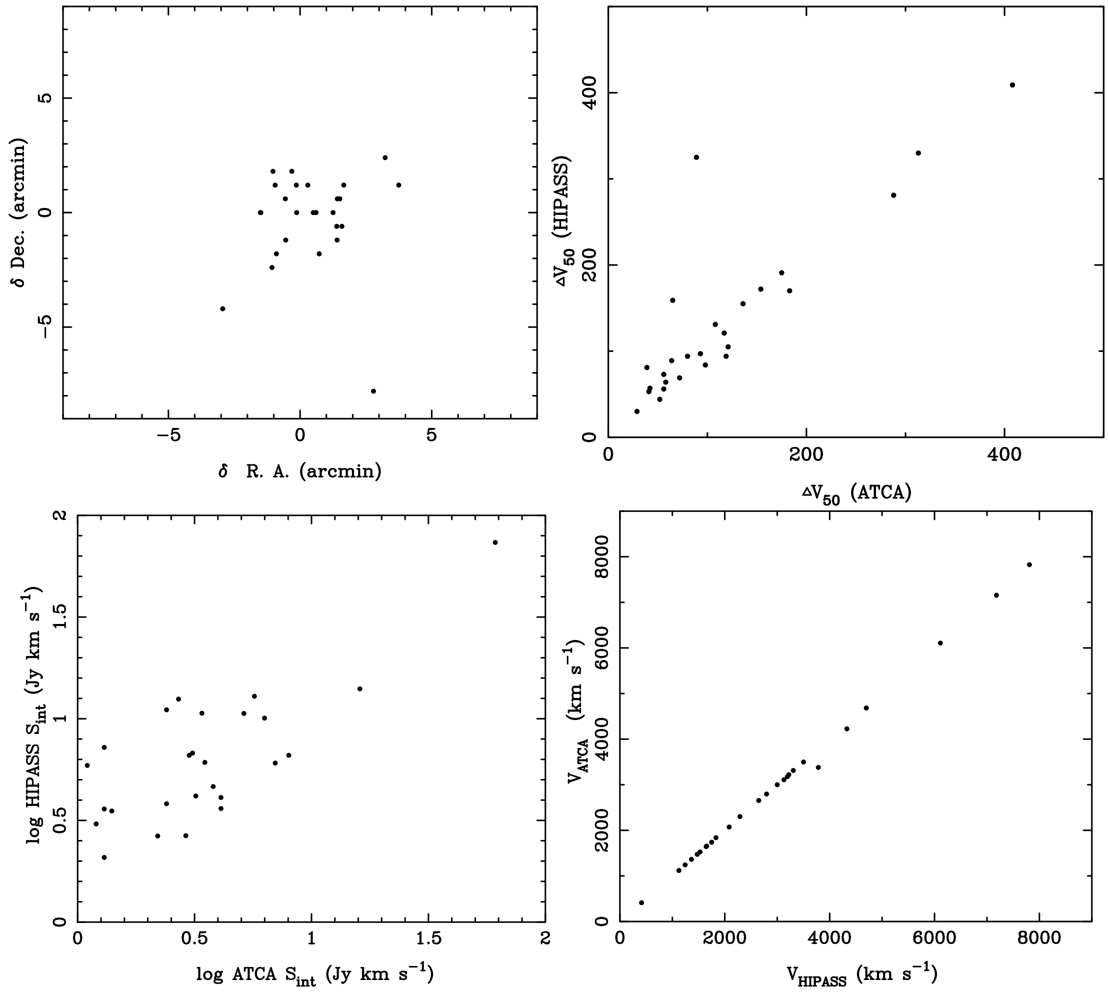

Figure 2.5: Panel (a) [Top Left]: Difference in R.A. and Dec. for the ATCA Measurements and HIPASS measurements. Panel (b) [Top Right]: ΔV_{50} from HIPASS versus ΔV_{50} from the ATCA. Panel (c) [Bottom Left]: Total flux from the ATCA and HIPASS data: Several galaxies have been fully detected in the ATCA observations, while most are missing some flux. Panel (d) [Bottom Left]: Velocities determined from the ATCA and HIPASS. Very good agreement is seen, with one exception, HIPASS J 1650-62 (see text).

One other source, HIPASSJ0908-64 also has a substantial difference between the HIPASS derived ΔV_{50} and the ATCA ΔV_{50}. This galaxy is a strong source, but the ATCA map indicates some extended emission, and the lower flux emission may have been missed in the ATCA spectrum, and indeed, 70% of the HiPASS flux is missed in the ATCA observations.

A comparison of total flux from the ATCA and HiPASS observations is shown in Figure 2.5(c). Most of the galaxies had a much lower total flux detection from the ATCA observations, while the ATCA detected the full flux for 5 of the 27 galaxies. The velocities determined from the ATCA and HiPASS observations matched very well (see Figure 2.5(d)). There was one discrepant value, HiPASS J1650-62, where the majority of the H I flux was missed in the ATCA observations, leading to the conclusion that the higher velocity channels of this detection contain extended emission.

2.8 Optical Follow-up Observations

An optical follow-up program was undertaken for the HiPASS survey, using the MSSSO 40-inch telescope at Siding Spring, the 1-metre Curtis Schmidt Telescope at CTIO, and the SWOPE telescope at Las Campanas. Optical follow-up is desirable for the newly catalogued HiPASS galaxies to determine colours and characteristics. 223 galaxies were observed during this program, in R and B-bands. About half of the objects observed lie in the SCC region, and at this time, 22 are reduced. Table 2.2 gives the names of the galaxies that have been observed and reduced, as well as their optical values. Optical data was reduced by Marquarding (2000), using the IRAF data reduction package. The columns for this Table are:

(1) HIPASS name.
(2) Previous optical Identity (if any).
(3) and (4) Right Ascension and Declination (J2000) of the optical galaxy in the field.
(5) Heliocentric Velocity from HiPASS observations (km s^{-1}).
(6) R total magnitude.
(7) B total magnitude.
(8) Extinction, E(B-V) (Schlegel et al., 1998).

HIPASS Name	Optical ID	RA J2000	Dec J2000	Vel km s^{-1}	R mag	B mag	E(B-V)
HIPASS J0125-76	-	01 25 07	-76 58 19	1688	17.03	-	0.04
HIPASS J0343-71	-	03 43 14	-71 48 59	1348	16.67	16.15	0.04
HIPASS J0635-70	-	06 35 36	-70 52 55	1583	16.30	15.53	0.06
HIPASS J0653-73	-	06 53 50	-73 40 35	1202	16.57	15.71	0.13
HIPASS J0717-79	-	07 17 40	-79 27 04	1561	17.35	15.85	0.35
HIPASS J0759-71	-	07 59 55	-71 50 49	1362	16.90	15.27	0.26
HIPASS J0903-64	-	09 03 19	-64 56 59	1669	14.05	-	0.21
HIPASS J1039-71	-	10 39 40	-71 55 23	1537	17.79	16.45	0.15
HIPASS J2039-63	-	20 39 08	-63 46 52	1645	15.89	14.75	0.06
HIPASS J0118-68	AM0117-681	01 18 34	-68 01 09	2039	15.04	14.42	0.02
HIPASS J0128-65	AM0126-653	01 28 23	-65 17 09	1639	14.54	13.69	0.02
HIPASS J0720-72	AM0720-723	07 20 02	-72 43 20	1501	15.88	14.60	0.26
HIPASS J0732-77	ESO017-G002	07 32 25	-77 55 03	1624	14.30	-	0.25
HIPASS J1049-80	ESO019-G004	10 49 00	-80 13 15	2009	15.56	14.51	0.13
HIPASS J1104-83	ESO007-G003	11 04 33	-83 14 46	2050	17.48	16.54	0.28
HIPASS J1121-75	ESO038-G011	11 21 35	-75 52 51	1879	16.03	14.70	0.30
HIPASS J1332-80	ESO021-G003	13 32 29	-80 26 29	2274	14.21	12.86	0.24
HIPASS J0838-75	ESO090-G004	08 38 44	-75 09 38	1138	14.78	-	0.13
HIPASS J0901-64	ESO090-G015	09 01 57	-64 52 46	1669	14.00	-	0.18
HIPASS J1209-75	ESO039-G006	12 09 06	-75 29 32	1899	15.51	-	0.35

Table 2.2: Optical follow-up observations of HIPASS galaxies (from Marquarding (2000)).

2.9 Optical Classification

All of the galaxies in the SCC sample were classified according to their optical morphology. The classification followed the scheme described in "The Carnegie Atlas of Galaxies" (Sandage and Bedke, 1994). Classifications of previously catalogued optical galaxies were checked with the NED database. Incorrect or misleading classifications may have been made with galaxies of small optical angular extent due to the poor resolution of the DSS I images. Due to the poor optical resolution, the optical types have been divided into four classes for analysis - Early Types (E/S0), Early Spirals (Sa-b), Late-Spirals (Sc-d), and Late-types (Sm, Im and d), ignoring possible bars. In addition, a number of galaxies were unable to be classified due to optical obscuration from the Galactic plane. The final problem for optical classification of galaxies was possible confusion in the Parkes beam. Appendix A.1 lists the classifications of the galaxies in the SCC sample. Galaxies were flagged if they were seen to be confused, interacting, or if an optical identification was not possible. For the purpose of this table, a galaxy was labelled as confused if there was more than one galaxy of unknown redshift within a 5 arcmin radius of the HIPASS position.

2.10 Characteristics of the Sample

Appendix A.1 gives the values calculated from the HIPASS data, as well as previous optical identifications and optical colours from either the ESO-LV catalogue (Lauberts and Valentijn, 1989) or from follow-up optical observations. Previous optical identification is also listed.

Figure 2.6 shows the velocity distribution for HIPASS galaxies in the South Celestial Cap compared with that of all NED[7] galaxies of known velocity in the same region. Galaxies from HIPASS with a radial velocity of less than 2000 km s^{-1} outnumber those with known optical redshifts. It is also evident that the velocity distribution of HIPASS galaxies does not follow the same structure as the optical galaxies. This is expected as the sensitivity limits for optical surveys and HIPASS are different. At larger velocities, HIPASS is only sensitive to the highest mass H I galaxies, and optical surveys are sensitive to the highest surface brightness galaxies, and the two are not equivalent. In addition, HIPASS may not detect nearby clusters of galaxies, due to their deficiency in neutral gas. Solanes et al. (2000) find that two thirds of the clusters they surveyed showed a deficiency of H I near the core which would suggest that HIPASS would miss such clusters. So while many optical surveys concentrate on optical

[7]It is noted that the NED optical catalogue is heterogeneous, thus comparison is difficult

clusters, HIPASS may not detect them at all.

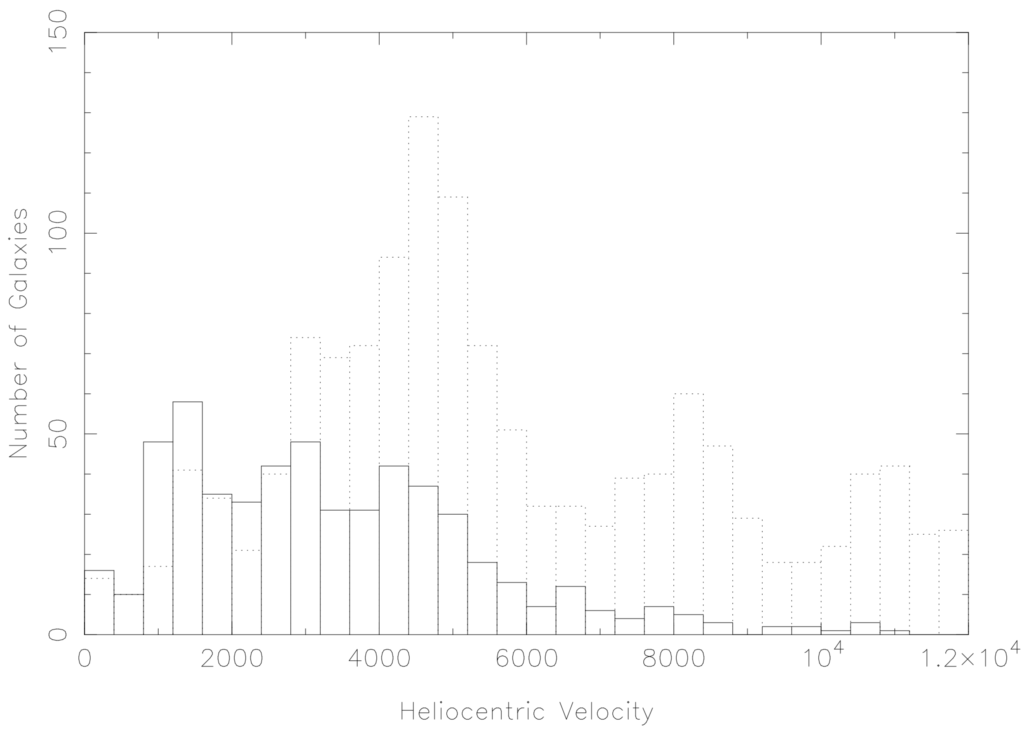

Figure 2.6: Velocity Distribution of the SCC sample (solid line), and all galaxies in the NED database with known redshift (dotted line).

The H I mass distribution is shown in Figure 2.7(a). The dashed line represents the distribution for the whole sample, and the solid line shows the H I mass distribution of newly detected galaxies. The mass range for the sample is larger than for previous surveys, with the lowest mass galaxy having $\sim 1 \times 10^6 M_\odot$ of H I and the highest mass galaxy with $\sim 8.8 \times 10^{10} M_\odot$ of H I. For the distance determination, a Hubble constant of $H_0 = 75$ km s^{-1} Mpc^{-1} is assumed. Most of the H I mass in the sample is contained in galaxies with $M_{\rm HI} \sim 10^9$ - 10^{10} $M_\odot$. The highest mass galaxy detected in the sample is HIPASS J2123-69, with a velocity of 10247 km s^{-1}. There are several optical galaxies visible near the position of this H I detection, so higher resolution H I observations are needed to determine the optical counterpart. Of the ten most massive galaxies in the sample, only one lies closer than 6500km s^{-1}, and that galaxy is NGC 6744, which lies at a velocity of $V_{LG} = 715$km s^{-1}. The H I mass distribution of the new galaxies follows that of the optical galaxies, although only new galaxies are present in the extreme H I mass bins.

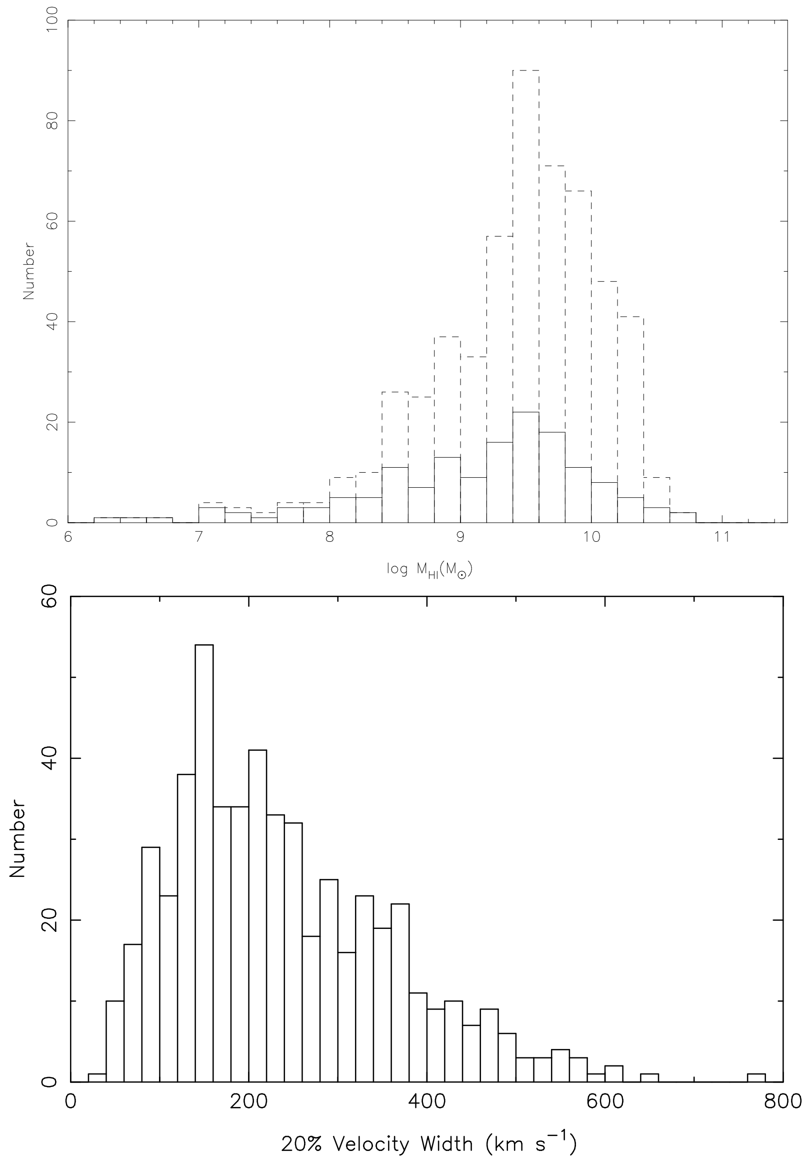

Figure 2.7: Panel (a) [Top]: H I Mass Distribution for the SCC Sample. The dashed line shows the H I mass distribution for the full sample, and the solid line shows the H I mass distribution of the newly detected galaxies. Panel (b) [Bottom]: Distribution of 20% velocity widths for the SCC Sample.

Figure 2.7(b) shows the distribution of 20% velocity widths (ΔV_{20}) for the sample. The largest velocity width is ΔV_{20} = 770 km s^{-1}, but this is a confused detection in the Parkes beam comprising the galaxies NGC 6872 and IC 4970. The smallest velocity width galaxy in the sample has ΔV_{20} = 35 km s^{-1}, which is just twice the velocity resolution of HIPASS. The mean ΔV_{20} for the sample is ΔV_{20} = 240 km s^{-1}, while the mean 50% velocity width for the sample is ΔV_{50} = 182 km s^{-1}.

2.11 Types of Galaxies found in the Region

The majority of galaxies in the SCC sample are classified as Spirals (366 galaxies, or 68%). There are equal numbers of Sa and Sb galaxy types, with a smaller number of galaxies classified as Sc. The number of S0s in the sample is low, 34 or $\sim$ 6%, and there were just 2 galaxies classified as Elliptical. The majority of the newly detected galaxies were Irregular galaxies or Dwarfs. 40 galaxies could not be optically identified due to either their Galactic obscuration, or the presence of 2 or more possible candidates. 11 galaxies were visibly interacting, and 35 galaxies were confused.

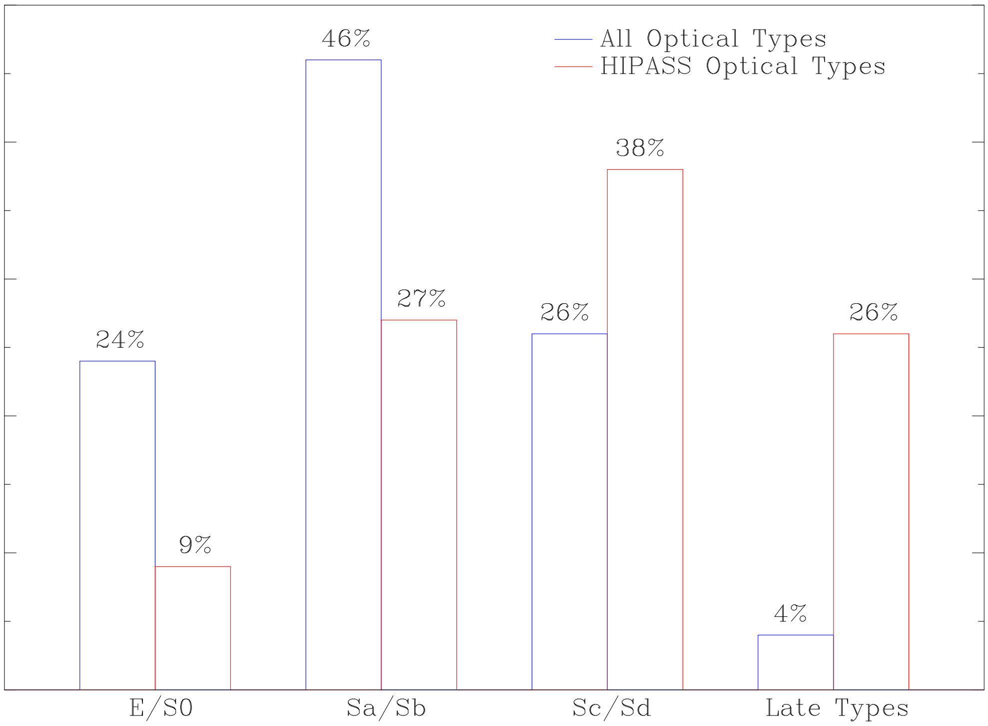

Figure 2.8: Classifications for all optically identified galaxies in the SCC region (blue line), and all HIPASS galaxies in the SCC region (red line).

2.12 Optical Properties of the H I Selected Galaxies

1396 galaxies with known redshifts have been catalogued optically in the SCC region, according to the NED database. Of these, 793 have morphological classifications from NED. The percentage break-up of type, of all optical galaxies in the SCC region with classifications compared to the break-up of optical types of the HIPASS galaxies is shown in Figure 2.8.

The distribution of galaxy type in the optical catalogues is quite different to that in the HIPASS catalogue. The optical catalogues contain a much higher percentage of early-type galaxies, while the HIPASS catalogue has a much greater number of late-type galaxies. The percentage of spirals in each sample is similar, but the early spirals dominate in the optical sample while the late spirals dominate the HIPASS catalogue.

2.13 Mass-to-Light Ratios

Figure 2.9 shows the distribution of mass-to-light ratios for the SCC sample. Most of the galaxies detected with HIPASS have an $\mathcal{M}_{\rm HI}/L_B < 5$, but 28 galaxies (or 5%) have a higher $\mathcal{M}_{\rm HI}/L_B$. The galaxy with the highest $\mathcal{M}_{\rm HI}/L_B$ is HIPASS J0019-77, with $\mathcal{M}_{\rm HI}/L_B = 24$. This is a previously known ESO galaxy, but the HIPASS observation represents the first H I detection. This particular detection may be confused however, as there are two optical galaxies close by, and only one has a measured redshift.

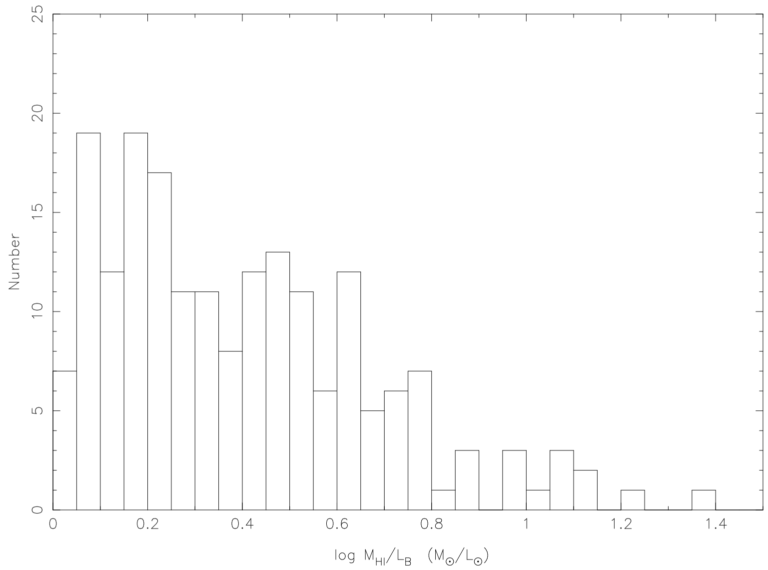

Figure 2.9: Distribution of log H I mass-to-light ratios for galaxies with published ESO-LV B-magnitudes.

2.14 Comparison with previous Optical Detections

The distribution of the HIPASS galaxies on the sky is similar to the distribution of already known galaxies in the region. Figure 2.10 shows the HIPASS galaxies plotted, with the previously catalogued galaxies with a redshift in the HIPASS detection range. The newly discovered HIPASS galaxies do not fill the voids, and follow the general distribution of the optically catalogued galaxies. The exception is near the Galactic plane, where the catalogue of optical galaxies is incomplete. Several optical clusters can be seen in Figure 2.10, with the most obvious being the PAVO II cluster at (l, b) = (332.3,-23.5), V=4167 km s^{-1}. This cluster is classed as a 'rich' galaxy cluster by Abell et al. (1989), meaning it contains more than 30 galaxy members, yet just one HIPASS galaxy is detected near the cluster at the cluster velocity. Once the full HIPASS catalogue is complete, the clustering of the optical galaxies compared to the HIPASS galaxies can be measured to quantify the differences between the two samples.

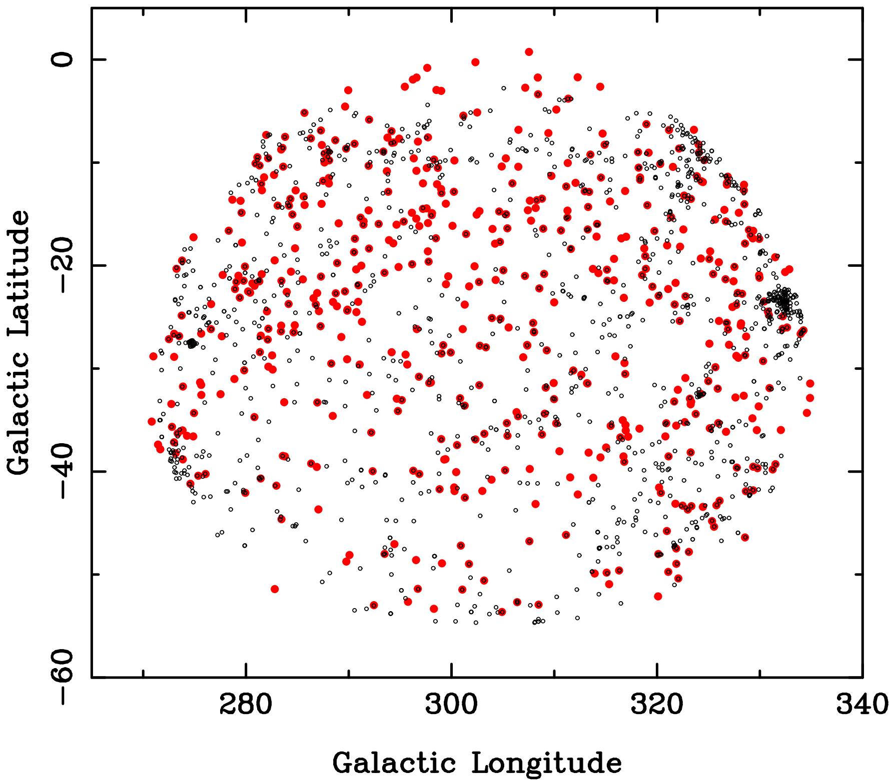

Figure 2.10: Distribution of SCC galaxies on the sky. The red circles represent HIPASS galaxies, and the black circles show previously catalogued optical galaxies in the same velocity range as the SCC sample.

2.15 Distribution of the SCC Sample

Figure 2.11 shows the distribution of galaxies, colour coded for their velocities. Clustering of the galaxies is evident, as the distribution of galaxies seems to move from left to right accross the plot with increasing velocity, until a velocity of 4500 km s^{-1}, where the galaxy distribution evens out somewhat.

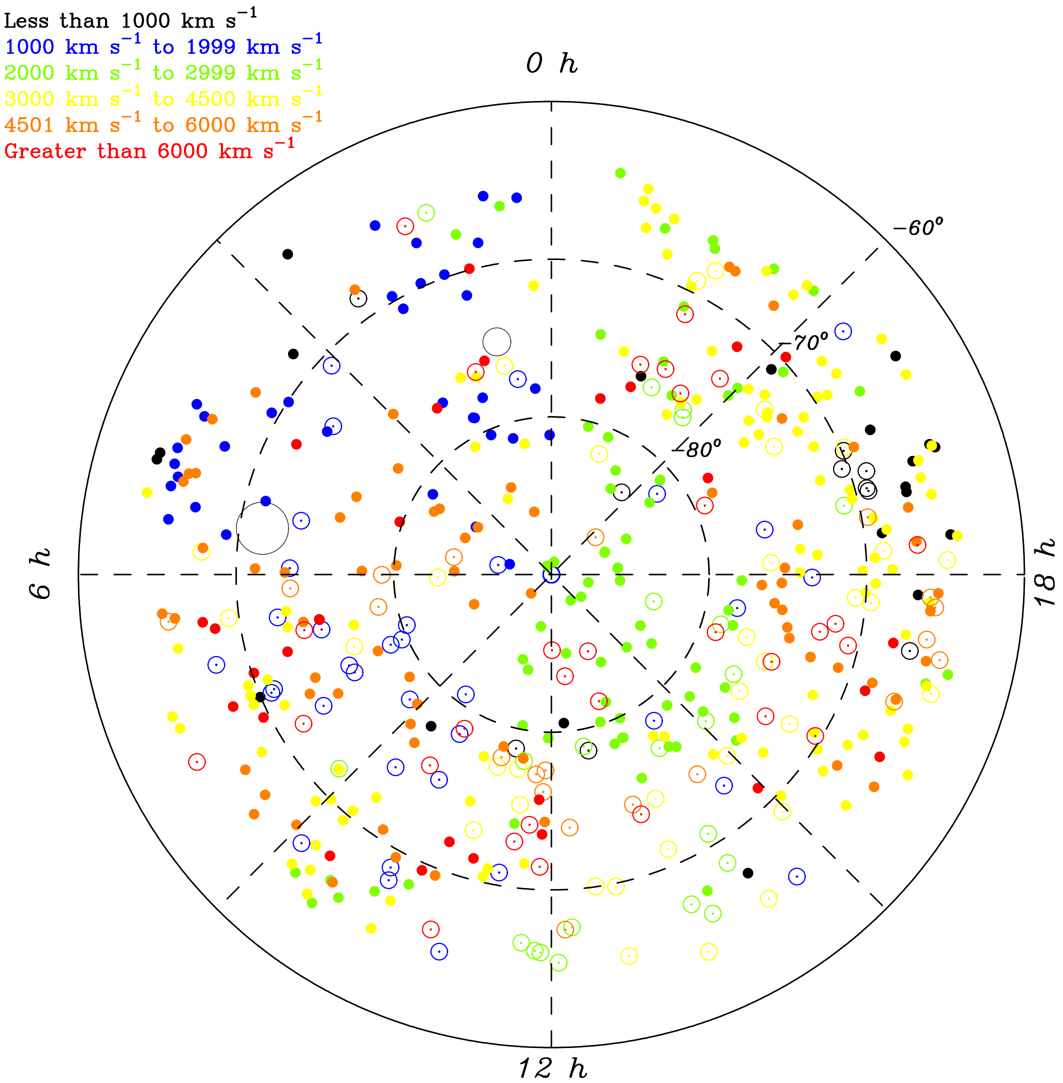

Figure 2.11: Galaxies in the SCC Region. The Magellanic Clouds are plotted as large black circles between the R.A. of 0h and 6h. New galaxies are indicated by ⊙ and previously known galaxies are filled circles.

2.16 Comparison with Large–Scale Optical Structures

The distribution of HIPASS galaxies for different velocity ranges is now compared to known optical structures based on maps by Fairall (1998). Figure 2.12 and Figure 2.13 show the distribution of HIPASS galaxies: those with previous optical counterparts are indicated by filled circles and galaxies with no previous optical identifications are shown with circles with a central dot. The Magellanic Clouds are shown in blue, and the plane of the Milky Way is towards the bottom of each plot.

The edge of the Centaurus wall is present in the low velocity region (V < 1000km s^{-1}). A number of new detections were discovered in this region. Mostly they are ambiguous H I detections without optical counterparts. This velocity range represents only $\sim$ 0.05% of the total volume of the SCC region, but this is the velocity range for finding low–mass galaxies ($M_{HI} \lesssim 10^8\, M_{\odot}$).

Between velocities 1000 km s^{-1} and 1999 km s^{-1}, the Fornax wall is the main structure, and the direction to the Fornax cluster indicated. The Fornax wall is seen face–on, while the Centaurus wall is an edge–on feature. In the next velocity region (2000 km s^{-1} to 2999 km s^{-1}) a fragmentation of the Fornax wall is present, and the Eridanus void can be seen to the top-right of the plot. The structure of galaxies continues moving to the right as we look deeper in velocity, and between velocities of 3000 km s^{-1} and 3999 km s^{-1} the Pavo spur is quite obvious, and the Eridanus void is still present. There are a number of new galaxies towards the plane of the Galaxy at this velocity, and it is possible we are seeing a new structure that joins the Fornax and Centaurus clusters via the Fornax wall.

The edge of the Centaurus wall is present again between the velocities of 4000 km s^{-1} and 4999 km s^{-1}, and there are still no galaxies present in the Eridanus void. The next velocity range, 5000 km $s^{-1}$to 5999 km s^{-1}, contains the start of the Sculpter void, along with a several new galaxies towards the plane of the Milky Way. The number of galaxies found at this velocity range has decreased from the numbers being detected in the lower velocity ranges.

Due to the sparsity of HIPASS detections above a velocity of 6000 km s^{-1}, no optical structures have been indicated. In the velocity range 6000 km s^{-1} to 7000 km s^{-1}, the number of new HIPASS galaxies outnumbers the number of galaxies HIPASS detected with known optical counterparts. Once again the new galaxies tend to be found towards the plane of the

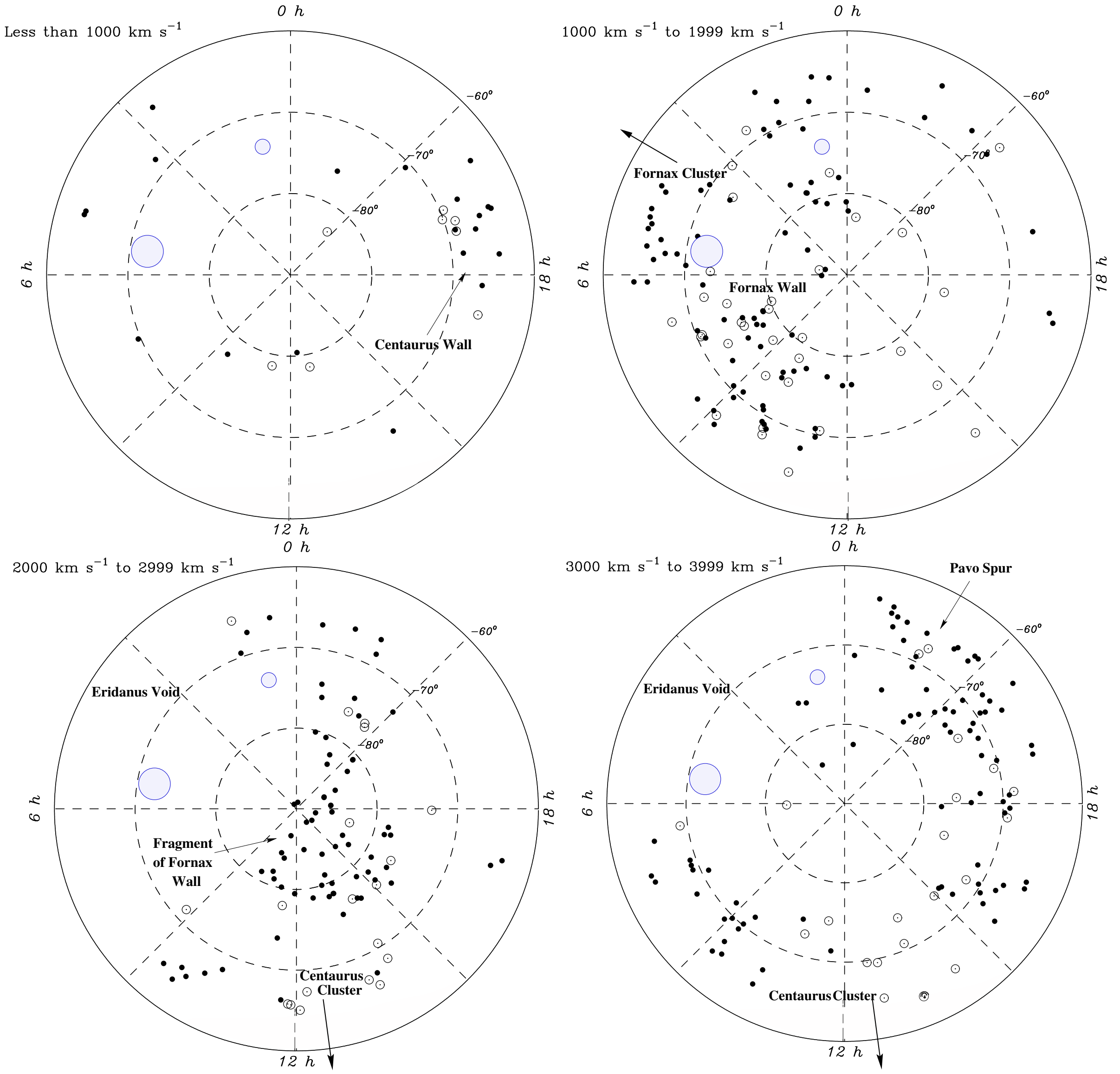

Figure 2.12: Distribution of South Cap galaxies I. Filled circles are HIPASS detections with a known optical counterpart, and the circles with a dot are new HIPASS detections. The plots are: (a) V $<$ 1000km s^{-1}. (b) 1000km s^{-1} $<$ V $<$ 2000km s^{-1}. (c) 2000km s^{-1} $<$ V $<$ 3000km s^{-1}. (d) 3000km s^{-1} $<$ V $<$ 4000km s^{-1}.

Galaxy. There are still a significant number of new galaxies being detected above a velocity of 7000 km s^{-1}, and in this velocity range only large H I mass galaxies have flux limits above the HIPASS sensitivity limit.

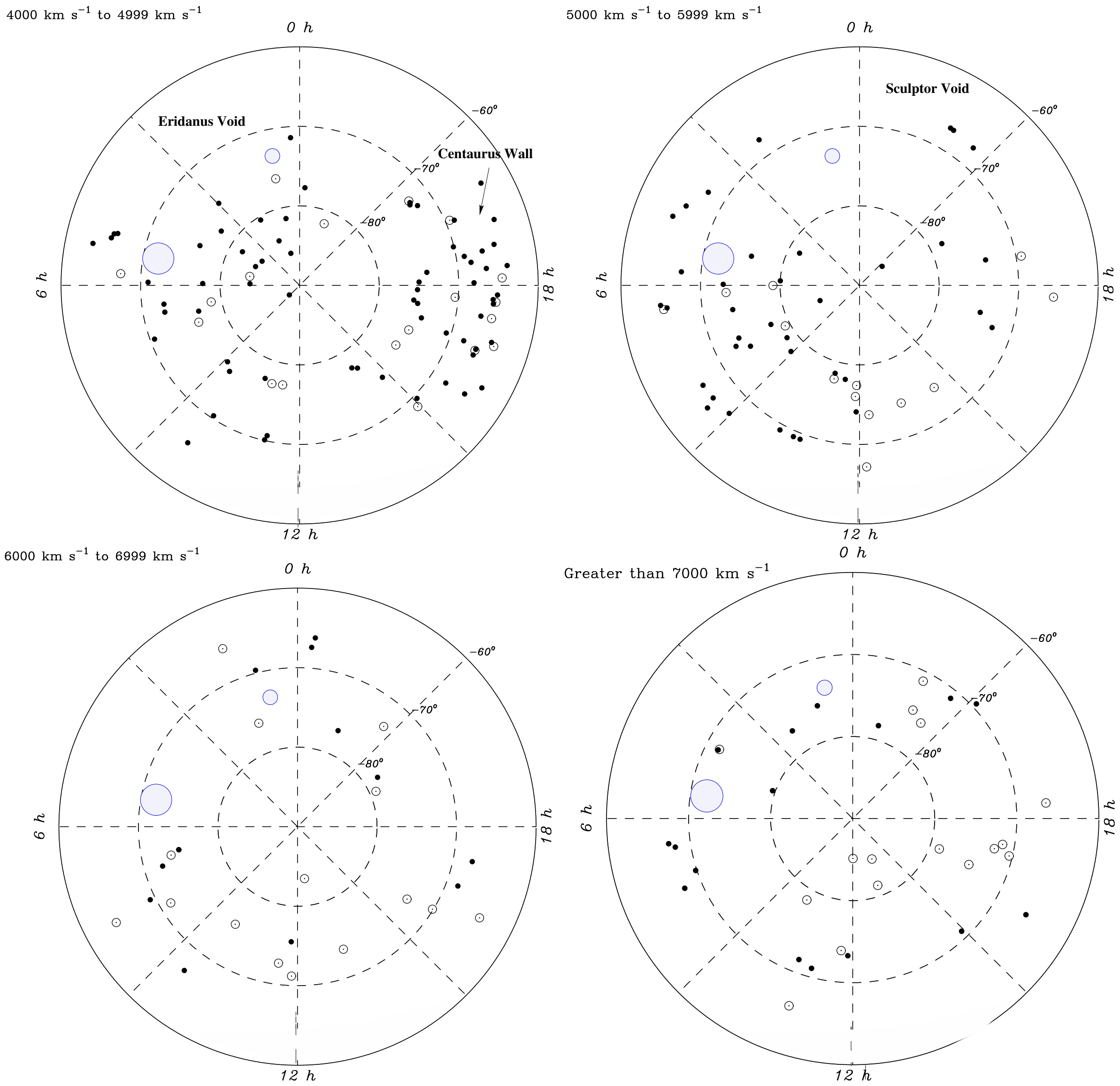

Figure 2.13: Distribution of South Cap galaxies II. Filled circles are HIPASS detections with a known optical counterpart, and the circles with a dot are new HIPASS detections. The plots are: (e) 4000km s^{-1} < V < 5000km s^{-1}. (f) 5000km s^{-1} < V < 6000km s^{-1}. (g) 6000km s^{-1} < V < 7000km s^{-1}. (h) V > 7000km s^{-1}.

Figure 2.14 shows all galaxies collapsed along the Galactic longitude and latitude axes, versus velocity. The plot of longitude versus velocity shows a large continuous structure starting at a velocity of $\sim$ 1000 km s^{-1}, and ending around $\sim$ 5000 km s^{-1}. The structure starts with

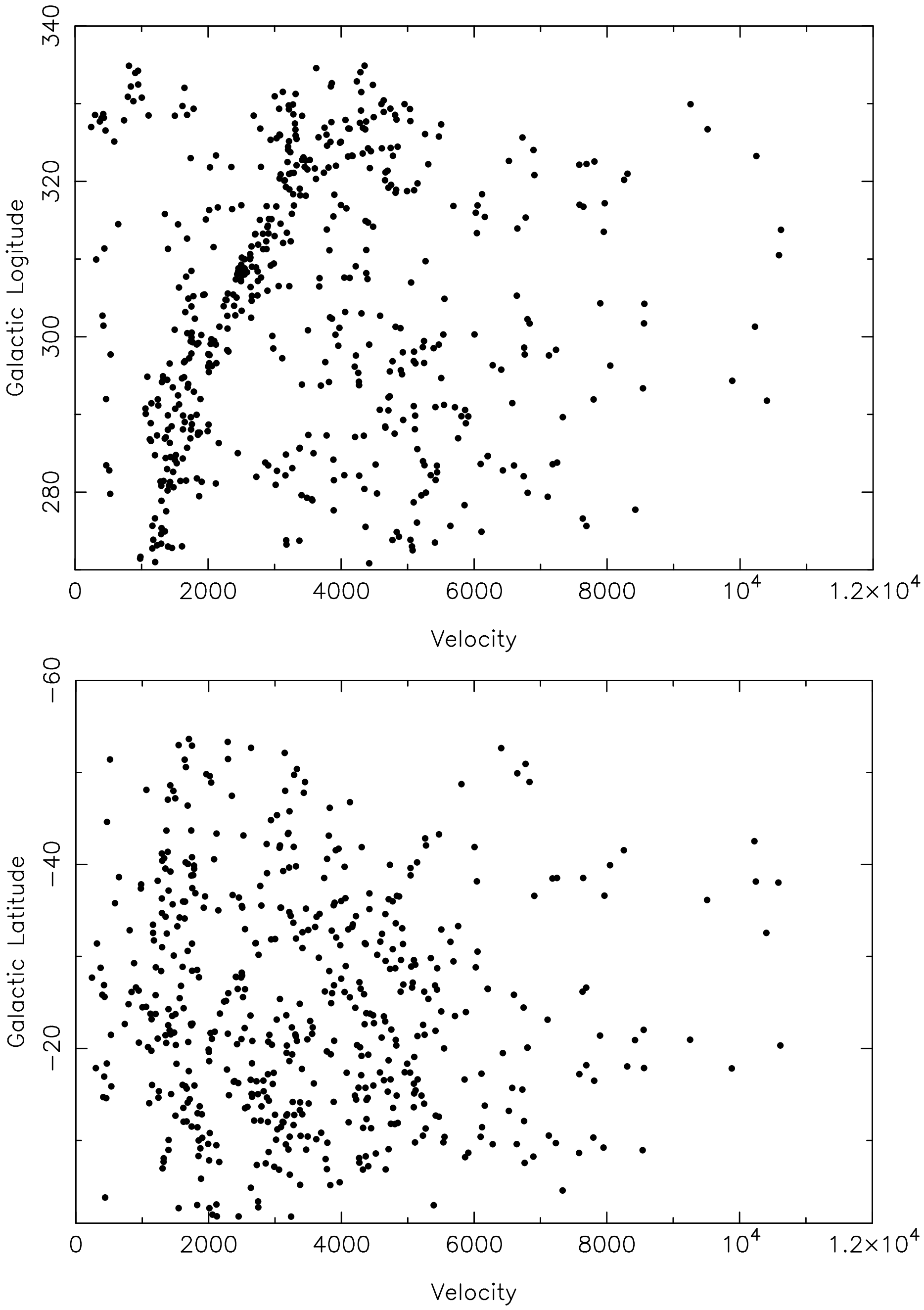

Figure 2.14: Galactic longitude vs velocity [Top] and Lltitude vs. velocity [Bottom].

the Fornax wall at low velocities, and with increasing velocity moves to the Fornax fragment and Pavo spur. This continuous structure is not seen in the Latitude versus Velocity plot.

Chapter 3

Galaxy Finding

3.1 Introduction

Galaxies are sparsely located in the HIPASS datacubes. While the eye is good at pattern matching, rejecting interference and finding irregular objects, visually searching the large HIPASS data set is not practical and it is likely that several different people (with different biases) would be needed to produce the catalogue, leading to limits that are hard to quantify. Computer-generated lists of candidates are a better way to produce a complete catalogue from the whole dataset. The aim of constructing this type of automated galaxy finder is to produce a catalogue with well characterised detection limits. Such a catalogue is necessary when determining H I Mass Functions, densities, and limits from H I samples.

An algorithm has been developed which automatically scans through the data cubes, and detects galaxies in 3 dimensions. There were several ways that the algorithm could be tackled, including using wavelet transformations, or fitting Gaussians to spectra in a cube. The method described in this chapter is the simplest conceptually, and aims for a robust calculation of the detection limits of the output catalogue.

The galaxy finding routine was written in PERL. PERL is a scripting language that allows external programs to be called. It is not compiled, and it is very transportable. The finder does not require a great number of mathematical calculations, so the speed of the finder should not be greatly affected by using PERL as the scripting language. The script uses the MIRIAD library extensively, and each operation that is performed on the datacubes uses a particular MIRIAD command.

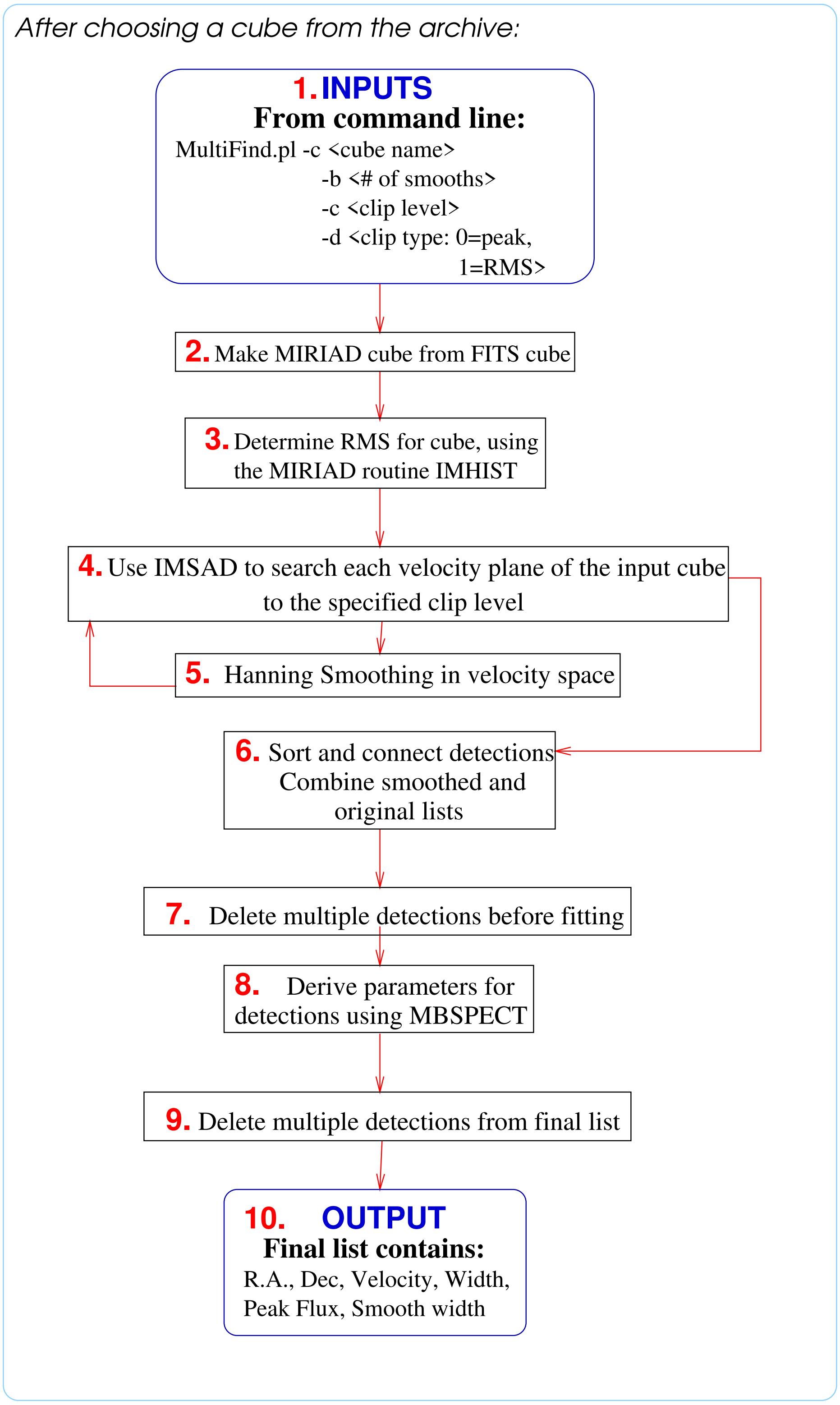

Figure 3.1: Flowchart of the MultiFind algorithm

3.2 MultiFind - A Simple Algorithm

Figure 3.1 shows the steps taken in producing the candidate galaxy list, using the MultiFind algorithm.

MultiFind searches for real sources with flux peaks above a user-defined level. The search level can be defined in one of two ways:
(a) A peak flux clip level, or
(b) A signal–to–noise (S/N) ratio.

Connected pixels above the clip level on each velocity plane are recorded. The plane-by-plane pixel search program is the MIRIAD routine, IMSAD. Figure 3.2 shows an example of the IMSAD routine detecting several galaxies in the HiPASS data. IMSAD fits a Gaussian to each 'clump' of pixels detected above the cut-off value, to produce a central position.

Detections found within 5 arcminutes of each other on consecutive velocity planes are recorded as possible detections. Stipulating that a real galaxy must be detected on two or more planes is an effective means of rejecting noise spikes, which are rarely correlated on two or more planes.

The detectability of galaxies depends on the degree of smoothing along the velocity axis. Lowering the noise in the cube can increase the signal-to-noise of a detection, making the galaxy more detectable. Two algorithms were tested to try and achieve this. The first method HANNING smoothed the data cube, and the second method smoothed the cubes with a top-hat filter (using IMBIN). The HANNING smoothing method was chosen for MultiFind for several reasons. As the optimal detectability of a galaxy occurs when the smoothing width equals the velocity width of the galaxy, it is necessary to smooth with as many widths as possible. In practice, it was found that the time taken for the MIRIAD routine IMBIN to complete each binning was prohibitive, and binning only a few times is not effective in finding large-width low-flux galaxies as it is likely they will lie in between the smoothing widths. This effect is partially overcome when using a HANNING smooth. In the Hanning smoothing algorithm, the smoothed cube consists of 50% of the original velocity plane, added to 25% of the velocity planes directly before and after. In this way, the smoothed cube has not 'missed' any channels, as could happen in poorly binned data.

New detection levels were determined for the smoothed cubes due to their reduced noise levels. For the S/N cut-off option, the new flux level was determined by calculating the RMS in the smoothed cube, and then applying the same S/N cut-off. For the peak flux searches,

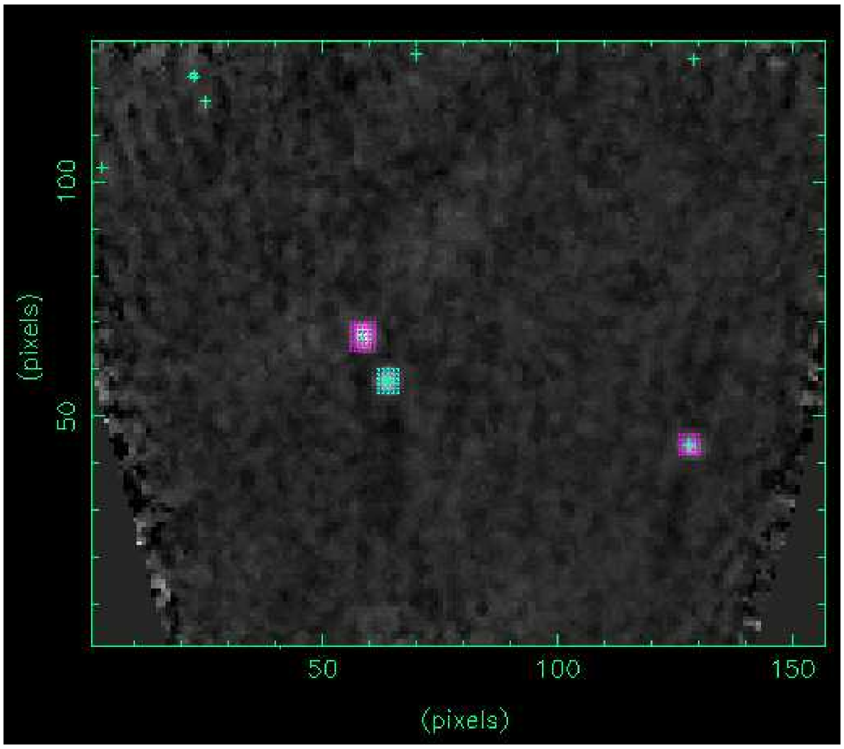

Figure 3.2: Example display from the MIRIAD task IMSAD. The coloured boxes show galaxies that lie above the specified cut-off. The green crosses without a coloured box around them are noise peaks above the cut-off level.

the flux level is reduced by 1.2 for each subsequent binning. It was expected that the noise would be reduced in the smoothed cube by $\sqrt{2}$. However in the HIPASS data, the noise is not Gaussian, and reduces by closer to 1.2, reflecting baseline noise.

The smoothed cubes are searched in the same way as the original cubes, and a list of detections produced. The number of smoothed cubes made is defined by the user. The detections from the smoothed cubes are then checked against each other and against the detections from the original cube, and any duplicates are deleted. The preliminary list contains the R.A. and Dec. of each detection, the number of planes the detection appeared over, and the central velocity plane for the detection.

Each detection is now parameterised to determine the heliocentric velocity, velocity width, and peak flux. The sequence of steps is:

1. A velocity integrated moment map is produced over the velocity planes that the detection was above the cut-off level. The MIRIAD routine MOMENT is used to make this map.
2. IMFIT is used to determine the central position of the detection by fitting a Gaussian to the

moment map. This fitting routine is occasionally unsuccessful when the signal-to-noise of the detection is low. In this case, the fitting is abandoned, and the original detection position from IMSAD is used.

3. A spectrum of each candidate galaxy is generated using the task MBSPECT, which also parameterises the detection, determining the heliocentric velocity, velocity width and peak flux for the detection.

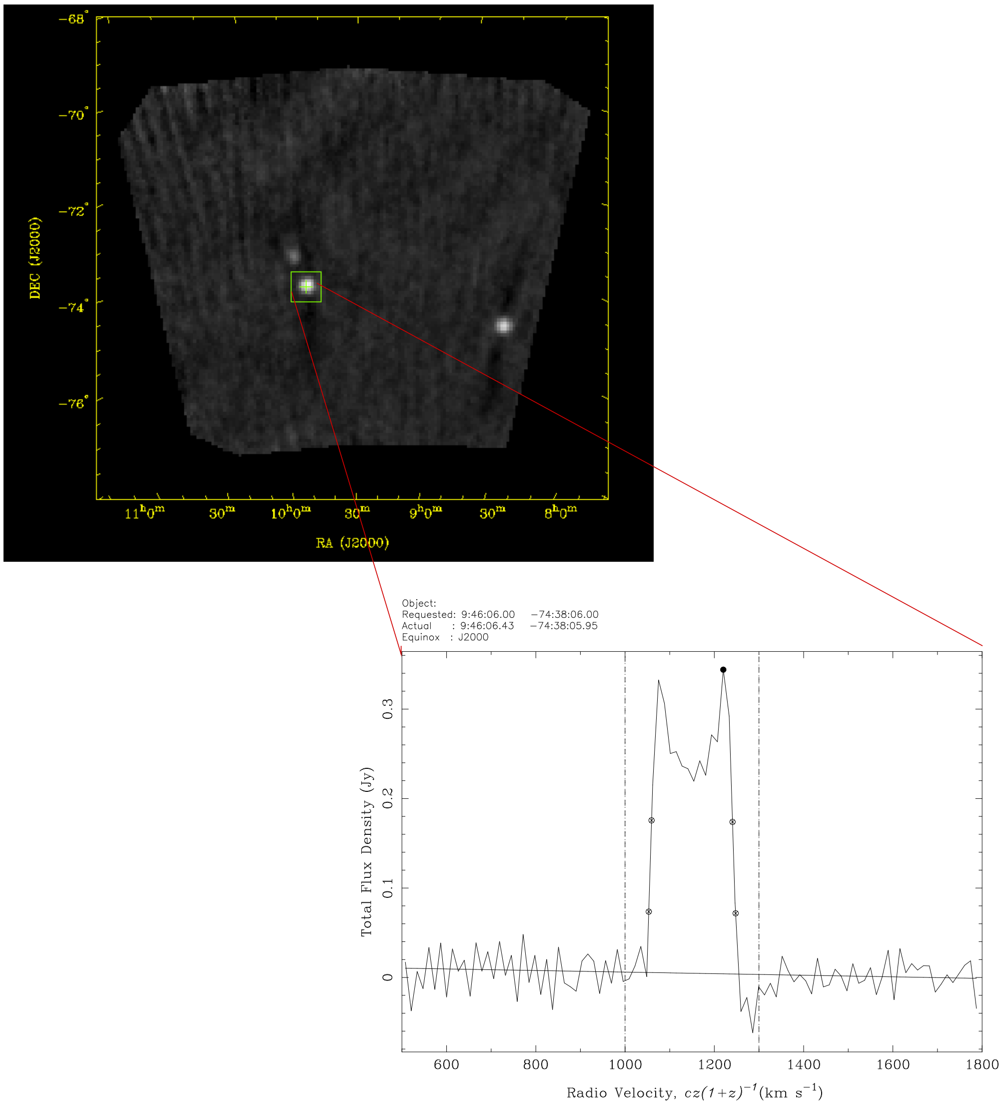

Figure 3.3: MultiFind galaxy display. Top image: A velocity integrated moment map of a galaxy candidate is displayed within the green box. Bottom image: A spectrum of the candidate galaxy produced by MBSPECT, with the 50% and 20% widths and peak flux marked with circles.

The display of the galaxy finder at this stage shows the position of the detection in the cube, followed by the spectrum with the fitted parameters plotted over the top. An example of the images can be seen in Figure 3.3.

The fitted values are then checked for multiple detections. A number of continuum sources are detected above a velocity of about 9000 km s^{-1}, due to inadequate bandpass removal. The LUTHER algorithm has done much to rectify this situation, however rather than have several detections of the one continuum source above this velocity, the detections are simply correlated for position, and recorded once in the list. The final list contains all detections above the given cut-off, some continuum sources, and interference spikes.

Before using the galaxy finder it was tested to check that the results matched expectations. Two testing methods were used: firstly on simulated data, and then on real data from the South Cap region.

3.3 Simulated Cubes

The simulated cubes used to test the galaxy finder were based on real HIPASS cubes, with added sources of known velocity width and peak flux. To ensure the simulated cubes reflect the real data, the sources that were added were actual HIPASS galaxies that were 'cut-out' from particular cubes and then 'pasted' into the simulation cube, with a scaled peak flux and velocity width. The noise in the spectrum was scaled to keep it similar to the noise in real galaxies. Three different profile types were chosen for the simulations - a 'square' profile, a 'double-horn' profile, and a 'triangular' profile. The profiles for the galaxies used are shown in Figure 3.4. Of course the range of different spectra found in the SCC sample spans a greater range than the above 3 profiles, but they are always some combination of the above.

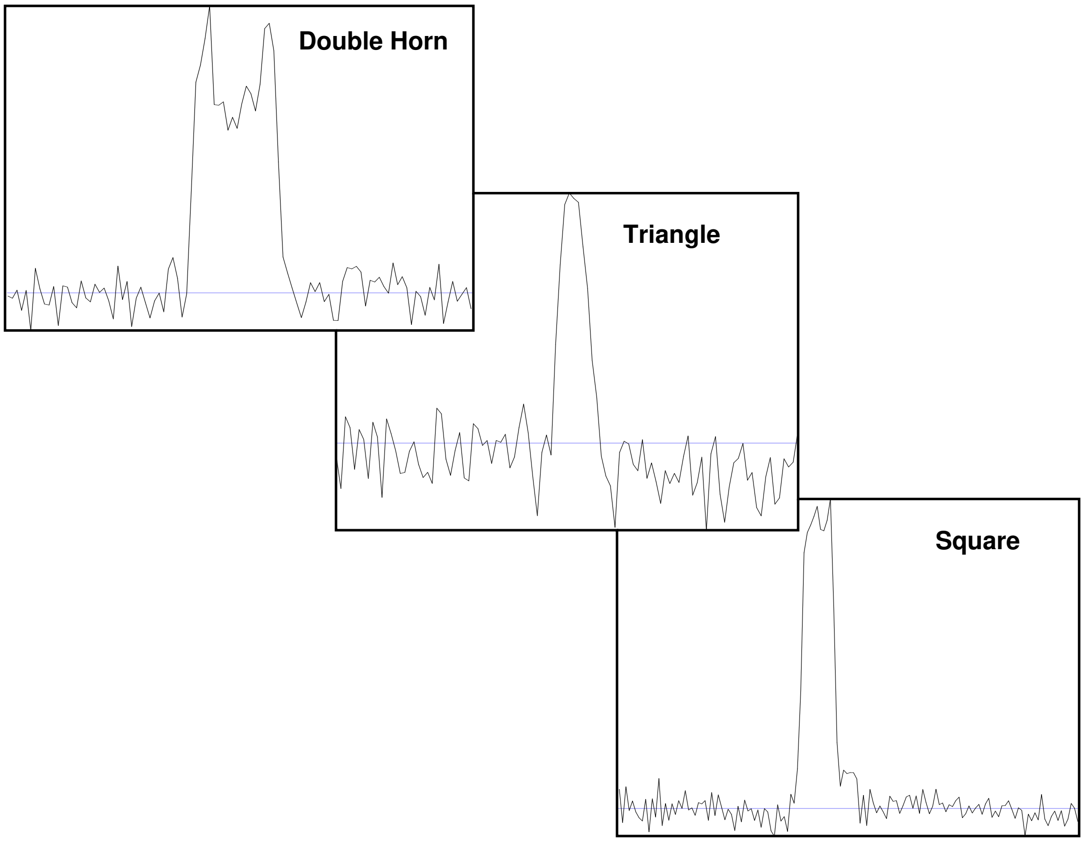

Figure 3.4: Profile Types Used in Simulations

The simulated cubes were produced by Niall Ryan, who wrote an IDL script to construct the cubes. 18 simulation cubes were produced. Three different HIPASS cubes were chosen as the basis of the simulations, with varying RMS noise levels. Thus the detectability of galaxies in data of varying quality was determined. Table 3.3 lists the test cubes, and the type of the sources pasted into them. Each cube contained 150 simulated galaxies, distributed throughout the cube. The galaxies were scaled in a $5 \times 5 \times 6$ grid, giving 6 galaxies of each type in each simulation cube. The galaxies were separated spatially by $\sim 1^{o}$, and by ~ 1200 km s^{-1} in velocity. The simulated galaxies were inserted into the cube on top of the existing noise, which means that some sources with a peak flux below the MultiFind cut-off limit will be pushed above the limit by noise, and vice-versa. Due to the possibility of systematic noise fluctuations, the flux and width properties of galaxies at the different grid positions was rotated at different velocity intervals in the simulated cubes. This ensured that the noise characteristics of the cube would not overly influence one galaxy type.

COARSE SEARCH

Cube	Profile Type	RMS (mJy beam^{-1})	Widths (km s^{-1})	Peak Flux (mJy)	# of Sources
H011	Square	13	360,270,190,90,50	200,100,60,40,20	150
H011	Triangle	13	360,270,190,90,50	200,100,60,40,20	150
H011	Double Horn	13	360,270,190,90,50	200,100,60,40,20	150
H013	Square	11	360,270,190,90,50	200,100,60,40,20	150
H013	Triangle	11	360,270,190,90,50	200,100,60,40,20	150
H013	Double Horn	11	360,270,190,90,50	200,100,60,40,20	150
H024	Square	15	360,270,190,90,50	200,100,60,40,20	150
H024	Triangle	15	360,270,190,90,50	200,100,60,40,20	150
H024	Double Horn	15	360,270,190,90,50	200,100,60,40,20	150

FINE SEARCH

Cube	Profile Type	RMS (mJy beam^{-1})	Widths (km s^{-1})	Peak Flux (mJy)	# of Sources
H011	Square	13	360,270,190,90,50	80,70,60,50,40	150
H011	Triangle	13	360,270,190,90,50	80,70,60,50,40	150
H011	Double Horn	13	360,270,190,90,50	80,70,60,50,40	150
H013	Square	11	360,270,190,90,50	80,70,60,50,40	150
H013	Triangle	11	360,270,190,90,50	80,70,60,50,40	150
H013	Double Horn	11	360,270,190,90,50	80,70,60,50,40	150
H024	Square	15	360,270,190,90,50	80,70,60,50,40	150
H024	Triangle	15	360,270,190,90,50	80,70,60,50,40	150
H024	Double Horn	15	360,270,190,90,50	80,70,60,50,40	150

Table 3.1: Details of simulated cubes. A coarse set of cubes was made first to determine the approximate detection limits on peak flux, then a fine set of cubes was produced to better determine the limits for MultiFind.

3.3.1 Results I

The first cubes that MultiFind was tested on had a typical range of velocity widths, and a large range of peak fluxes, to determine the approximate performance of the finder. The MultiFind input parameters used were a peak flux cut-off of 45 mJy and due to time constraints, only 2 x binning were performed. The results from these first tests can be seen in Figure 3.5. It can be seen that in all cases the galaxy finder successfully found all of the 200 and the 100 mJy peak flux sources, and the majority of the 60 mJy peak flux sources at all velocity widths. However, only a couple of 40 mJy sources were picked up in any of the cubes, consistent with the peak cut-off of 45 mJy. The different sized circles in Figure 3.5 represent detections at different velocities. For each simulation, the galaxies missing are from a variety of different velocities or 'circles'. Thus velocity that the galaxy was placed at in the simulation cube does not affect its detectability.

MultiFind is most successful in finding galaxies with square profiles, which is not surprising as the criteria for galaxy detection is that the peak flux of the galaxy must be above the specified cut-off for two or more velocity channels.

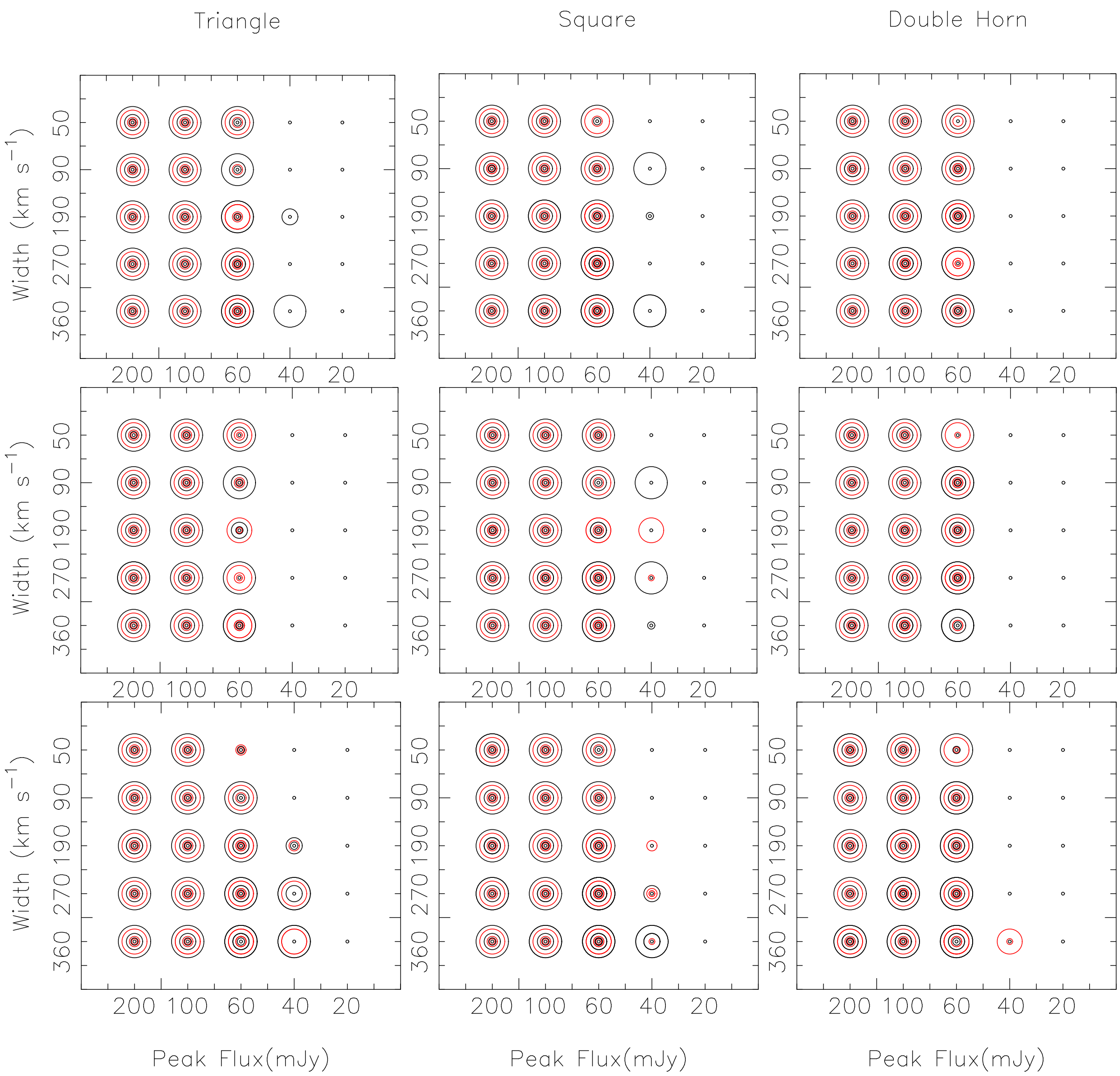

Figure 3.5: Galaxy Finder Tests I: Larger circles indicate simulated galaxies that were detected by MultiFind. Six galaxies of each velocity width and peak flux were placed in each cube. Each line represents a different simulated cube: From top to bottom, they are cubes H011, H013 and H024.

MultiFind performs most poorly on triangular profiles of narrow velocity width. This is due to just one velocity channel containing flux above the cut-off limit. Therefore the galaxy does not pass the criteria that it must be above the cut-off limit on two consecutive velocity channels. The efficiency of detecting double horned profiles was slightly worse than the square profiles, with the low velocity width sources occasionally being missed.

There was some variation between the detectability of the galaxies between data cubes of different quality. However, this was not directly correlated with the RMS noise of the cube. The highest detection rate for 40 mJy sources was in the cube with the highest RMS noise. One reason for this is that a higher percentage of 40 mJy sources may have been sitting on noise peaks that brought them above the cut-off level.

The percentage of galaxies detected in the first set of simulated cubes versus peak flux (over all velocity widths) can be seen in Table 3.2. It is once again apparent from this table that MultiFind is very efficient at finding square profile galaxies, and less so for triangle and double horn profiles, especially at small velocity widths. At a peak flux of 60 mJy, MultiFind detected 90% of the galaxies with triangle profiles, compared with 93% of the input double horn galaxies, and 96% for galaxies with square profiles. Figure 3.6 shows the number of galaxies detected as a function of velocity width. There is a slight trend towards detecting larger width galaxies of a given peak flux more easily, although this is not apparent for the double horned profiles.

The detection limits are still unclear from these course cubes, so MultiFind was run on simulated cubes of finer peak flux resolution to determine the limits more accurately.

PROFILE TYPE	PERCENTAGE DETECTED **Peak Flux**				
	200 mJy	**100 mJy**	**60 mJy**	**40 mJy**	**20 mJy**
Double horn	100	100	93	2	0
Triangle	100	100	90	14	0
Square	100	100	96	11	0
Overall	100	100	93	9	0

Table 3.2: Percentage of Galaxies Detected using MultiFind, for the first set of simulation cubes (all widths combined). All of the 200 and 100 mJy sources were detected, while the detection rate for galaxies with a peak flux of 40 mJy was low.

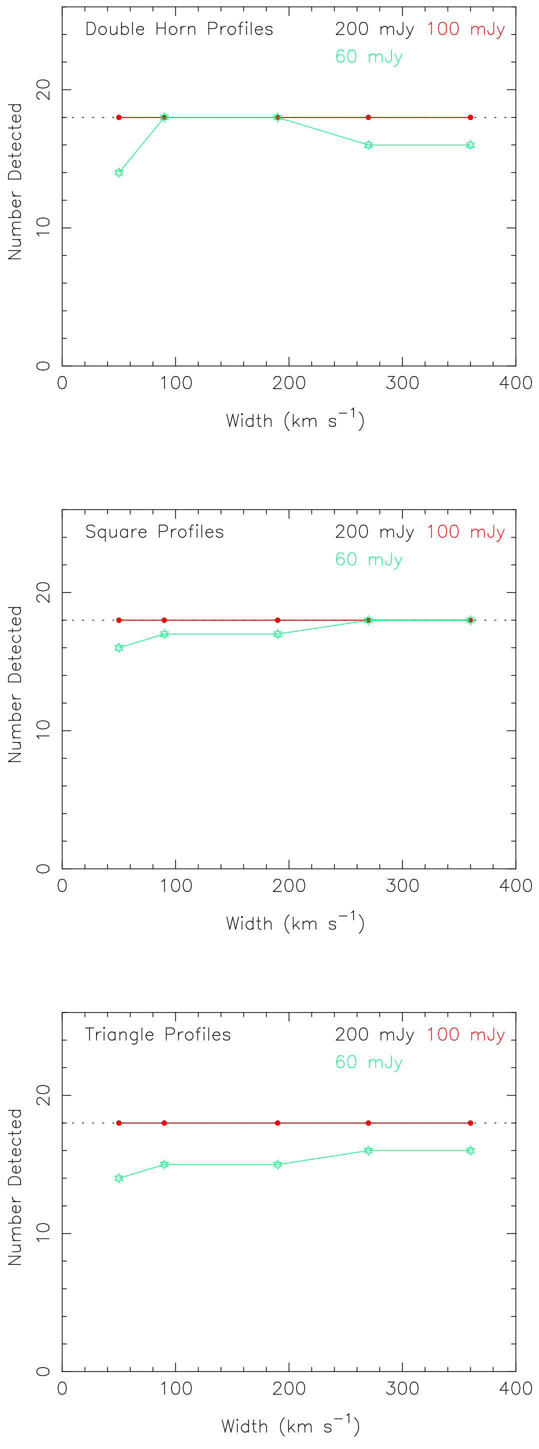

Figure 3.6: Simulated galaxies detected as a function of peak flux and velocity width. There is a slight trend towards detecting more low–flux large–width galaxies than narrow width galaxies for the square and triangle profiles.

3.3.2 Results II

A second set of simulated data cubes was made to more closely determine the cut-off limits for MultiFind. This dataset contained simulated sources with a finer grid of peak fluxes, but the same velocity widths as previously. Figure 3.7 shows the galaxies detected in these cubes. The detection of galaxies with triangular profiles with small velocity widths is less efficient than the detection of galaxies with the other two profile types. The detection of the square profiles is remarkably good, with a large number of sources with a peak flux of 50 mJy being detected.

Table 3.3 shows the percentage of galaxies detected for each peak flux over all velocity widths.

There is an obvious drop-out of detection numbers around 50 mJy, with only one third of the simulated galaxies of all profile types being detected, compared to the detection rate of 60 mJy galaxies, where 93% were detected. Simulated galaxies with triangular profiles had a lower detection rate than the double horn and square profiles, especially at low velocity widths. This was expected, as with the lowest simulated velocity width (50 km s^{-1}), the galaxy is present on only 3-4 channels, and unless the profile is steeply rising (as with the double horn and square profiles), it is less likely that the will be two consecutive velocity channels with a peak flux above the cut-off limit.

Figure 3.8 shows the number of galaxies detected with different velocity widths. Once again, apart from the double horn profiles there is a trend towards a better detectability of low peak flux galaxies with larger velocity widths.

PROFILE TYPE	PERCENTAGE DETECTED **Peak Flux**				
	80 mJy	**70 mJy**	**60mJy**	**50 mJy**	**40mJy**
Double horn	100	98	94	27	3
Triangle	98	94	87	29	4
Square	100	99	97	55	8
Overall	99	97	93	37	5

Table 3.3: Percentage of galaxies detected using MultiFind, for the second set of simulation cubes (all widths combined).

Figure 3.7: Galaxy Finder Tests II: Larger circles indicate simulated galaxies that were detected by MultiFind. Six galaxies of each velocity width and peak flux were placed in each cube. Each line represents a different simulated cube: From top to bottom, they are cubes H011, H013 and H024.

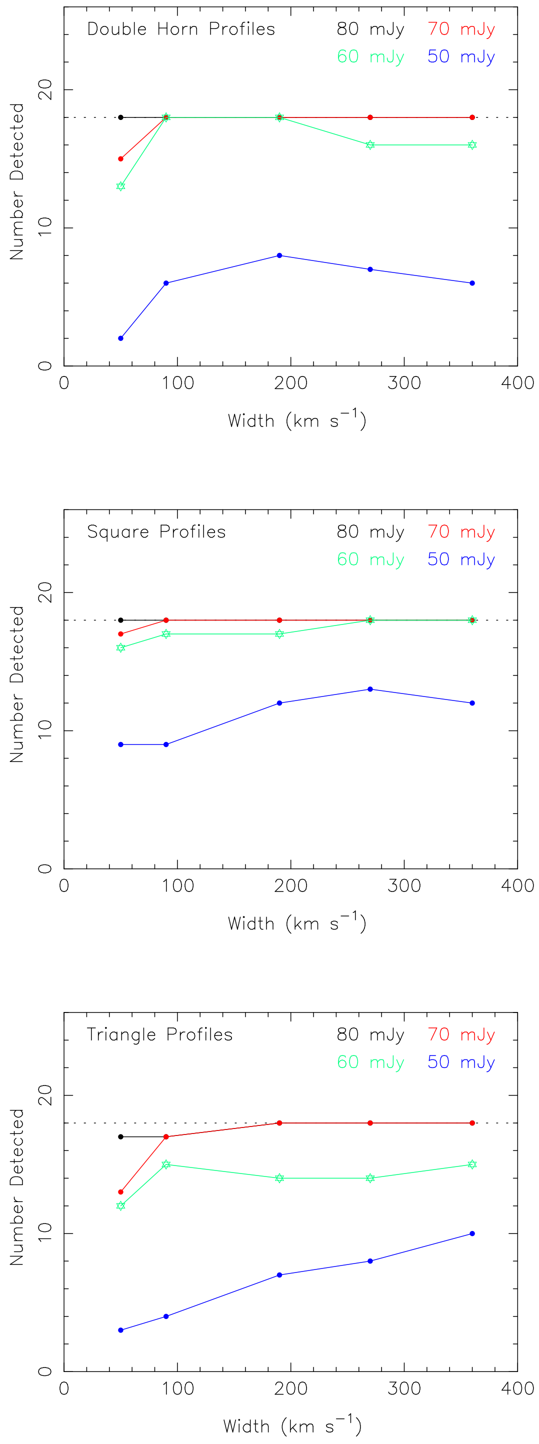

Figure 3.8: Simulated galaxies that were detected as a function of width and peak flux. A general trend toward detecting more galaxies at lower flux with a larger velocity width is seen.

3.3.3 Accuracy of the Output Values

MultiFind produces a list of coordinates, peak fluxes, velocity and velocity widths in the output list, which can now be tested against the simulation input values. As the simulation galaxies are actual HiPASS galaxies, their widths were measured using the MIRIAD routine MBSPECT. These widths were then compared with the output from MultiFind.

Figures 3.9 and Figure 3.10 show the input and output values of Velocity Width and Peak Flux for the three different profile shapes. In all cases, there is a good relation between the input and output values, and there is some spread in the measured values. On each plot, the dashed line shows where the input equals the output. Some data points are substantially offset from the mean. This is more noticeable in the velocity width determination, and is due to noise peaks perturbing the fitting.

The closest matched output of velocity width and peak flux was with the double horn profile galaxies. There is relatively little scatter around the velocity width determinations, and the same is seen in the peak flux plot. MultiFind derived a lower output velocity width for the triangle profile galaxies compared to the other profile types.

In the case of the peak fluxes, the mean of the lowest peak flux data is systematically perturbed. This is due to the imposed cut-off limit of 45 mJy. There is a spread of MultiFind derived peak fluxes at all peak fluxes. This spread is due to noise, position fitting and systematic errors, and the typical RMS of the scatter is about 4 mJy for each of the peak fluxes. The spread seems higher than this upon inspection of Figure 3.9. However this is caused by the large scatter of a small number of fluxes, rather than the majority.

The comparison between input and output central velocities has not been made as the galaxy finder lists were automatically matched to the input lists, within a velocity of 200 km s^{-1}. Therefore any detections that had a determined velocity which deviated from this range will not have been matched, so it is not possible to test the effectiveness of MultiFind in determining central velocities in this manner.

Th calculation of output values for the detected galaxies is an indication of the actual value, but in practice will need to be redone using a semi–automated process to determine more accurate values. The incidence of galaxies lying close together, noise spikes and other irregularities in the data results in errors that are easily removed with visual inspection, but would take sophisticated programming to remove automatically.

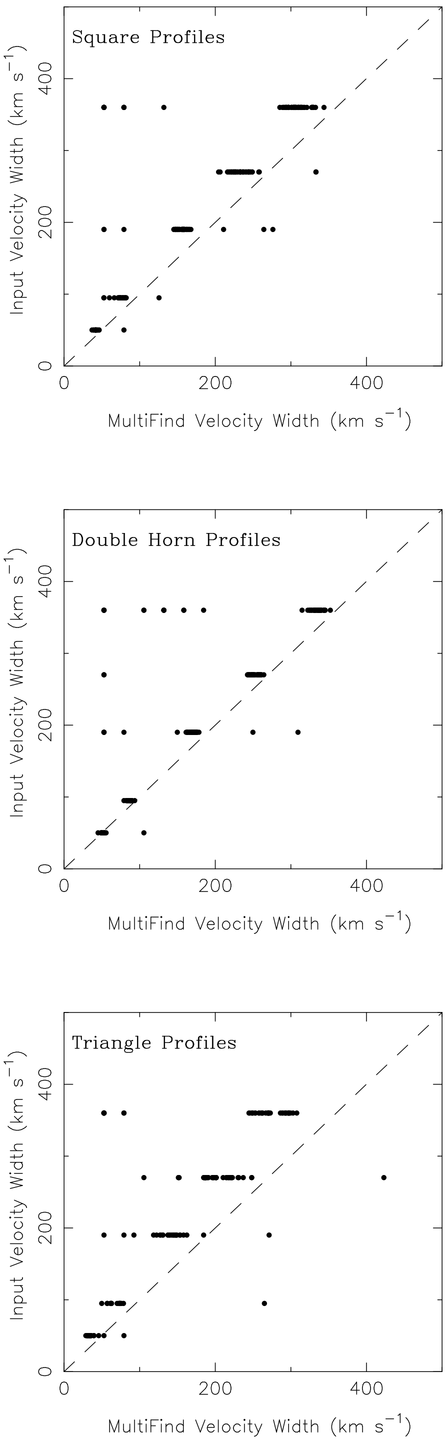

Figure 3.9: Comparison of Input/Output Velocity Widths. The dashed line shows the case where the input velocity width equals the output velocity width.

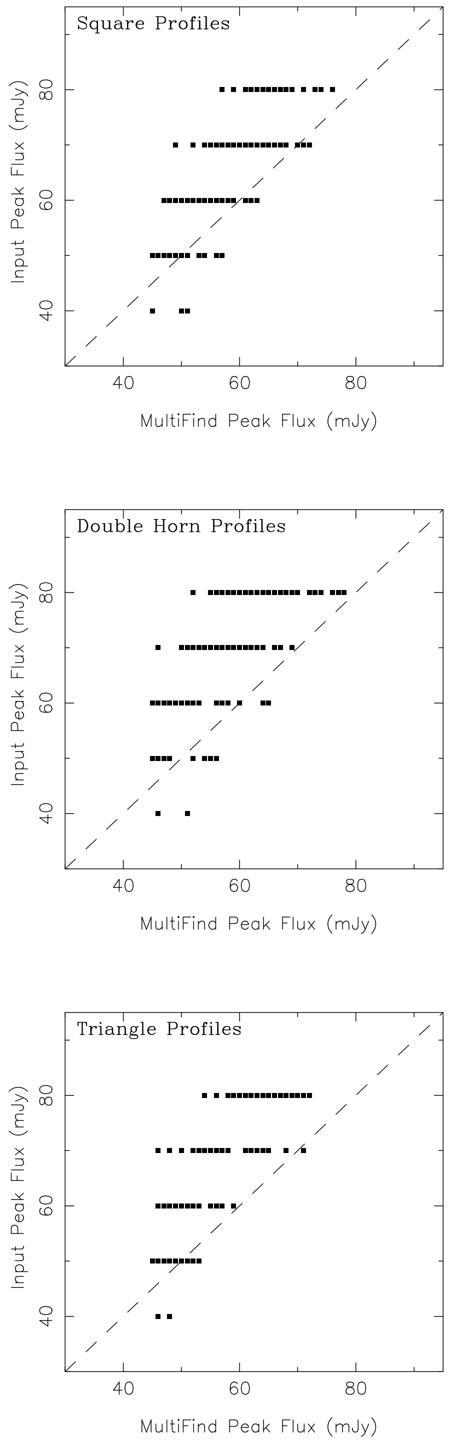

Figure 3.10: Comparison of Input/Output Peak Fluxes. The dashed line shows the case where the input flux equals the output flux.

3.3.4 Summary

From the simulated cubes, it can be seen that the galaxy finder is 93% accurate to detecting sources with a peak flux of greater than 60 mJy, with all profile types combined. The detection rate jumps to 95.5% when galaxies with only square or double-horn profiles are considered. The fraction of total galaxies with each profile type is unknown for real data, and the performance of MultiFind on galaxies with 'combination' profiles is also unknown. Thus MultiFind was also tested on real galaxies in the SCC sample, to determine whether the limits found from the simulations are representative of the limits in a real galaxy sample.

3.4 Real Data: The SCC Region

MultiFind was run on 23 cubes in the SCC region in order to test the algorithm on real data, and to see if any additional galaxies were picked up by the galaxy finder that were not detected by eye. By using the real data, it is also possible to set more stringent limits on the reliability of the galaxy finder, and test the finder with differing profile shapes and noise levels. Only those regions away from the Galactic Plane were examined, as MultiFind does not perform well in the presence of residual continuum emission. This test on real data will test how well MultiFind works compared to a visual catalogue produced using the human eye. Thus galaxies might be missed in both catalogues. It would have be useful to test MultiFind on a HIPASS region that has subsequently had more telescope integration time. Thus it could be determined whether noisy peaks detected MultiFind in the less sensitive data are in fact real detections. As this type of data set is not available at the moment, and MultiFind is not probing well down into the noise (ie: the minimum detection is $\sim 4.5\sigma$), the normal HIPASS cubes are used for the test.

MultiFind was run with the same parameters as in the simulated cubes section. The data cubes used had increased RMS noise around the edges due to the number of scans being lower on the edges, so the edges were masked generously to avoid these high RMS regions.

Figure 3.11(a) shows a histogram of the peak fluxes of galaxies in the test region (dotted histogram), with a histogram of those galaxies missed by MultiFind (solid histogram). Most of the galaxies in the region have a peak flux around 50 mJy, and but with a cut–off flux limit of 45 mJy, MultiFind misses most of these galaxies. There were 226 galaxies visually detected in the test region, of which MultiFind detected 105. No new galaxies were found by MultiFind in the region.

To better replicate the primary measure used to detect galaxies using the MultiFind algorithm, the average flux of a galaxy, which is defined as

$$S_{av} = \frac{S_{int}}{\Delta V_{50}}. \tag{3.1}$$

was investigated. This flux measurement is not as sensitive to noise spikes which may increase the peak flux of a galaxy without increasing its detectability. Figure 3.11(b) shows the average flux of galaxies in the HIPASS test region, and the galaxies not-detected by MultiFind. MultiFind is 94% complete to an average flux of 60 mJy.

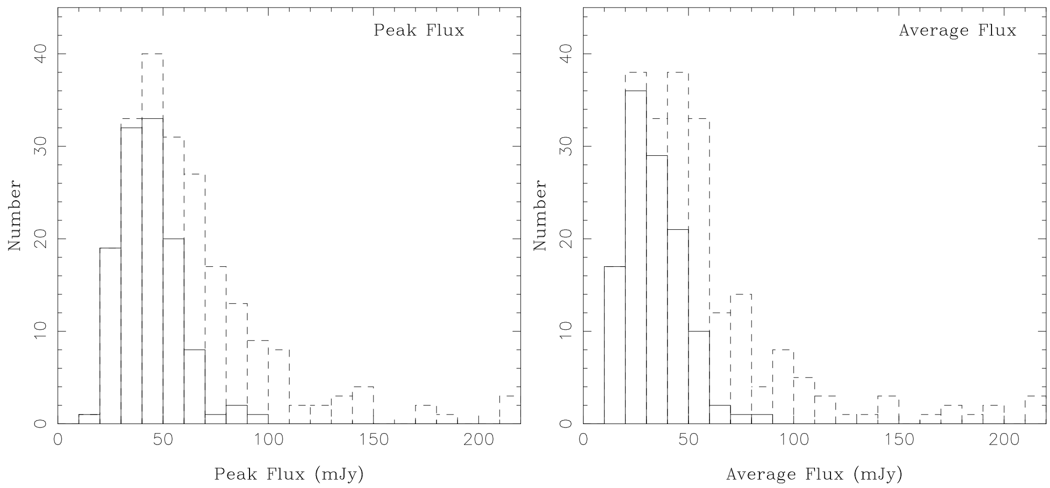

Figure 3.11: Fluxes of Galaxies in the MultiFind Test Region. The dotted histograms show the fluxes of galaxies detected by eye in the HIPASS test region, and the solid histogram shows the galaxies not detected by MultiFind. Panel(a) [Left]: Peak flux distribution of galaxies. Panel(b) [Right]: Average flux distribution of galaxies.

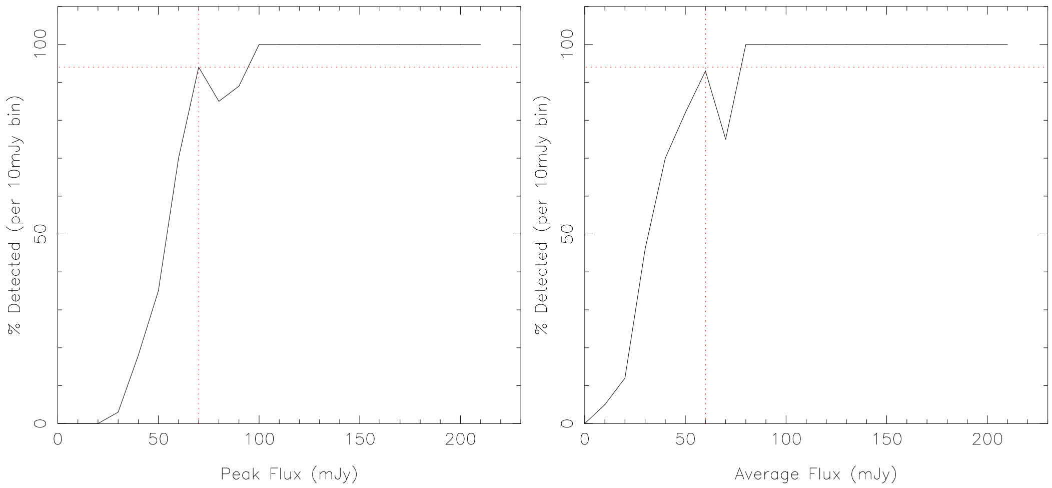

Figure 3.12: Percentage Completeness in the HIPASS Test Region, using two flux definitions - peak and average flux. The red lines show the flux at which the sample is 94% complete.

The percentage completeness versus flux for both the peak and average flux of the galaxies in the HIPASS test region is shown in Figure 3.12. The 94% completeness limit is shown on each plot. There is some variation in the completeness lines where MultiFind begins to miss galaxies (around 90mJy), and this is likely to be due to low number statistics in the flux bins. In the average flux bin of 45 mJy, which was the limit set for the detection algorithm, MultiFind detects $\sim$ 65% of the galaxies detected by eye.

Figure 3.13 shows the derived MultiFind peak flux and velocity width versus those measured for the same galaxies using MIRIAD measuring routines manually. There is good agreement with the automatically determined velocity width, with a small amount of scatter produced by noise peaks in the spectra. There is much greater scatter with the MultiFind derived peak fluxes, and generally the MultiFind flux tends to be lower than the measured peak flux. This could be caused by an offset in the central position of the generated spectrum in the automated routine.

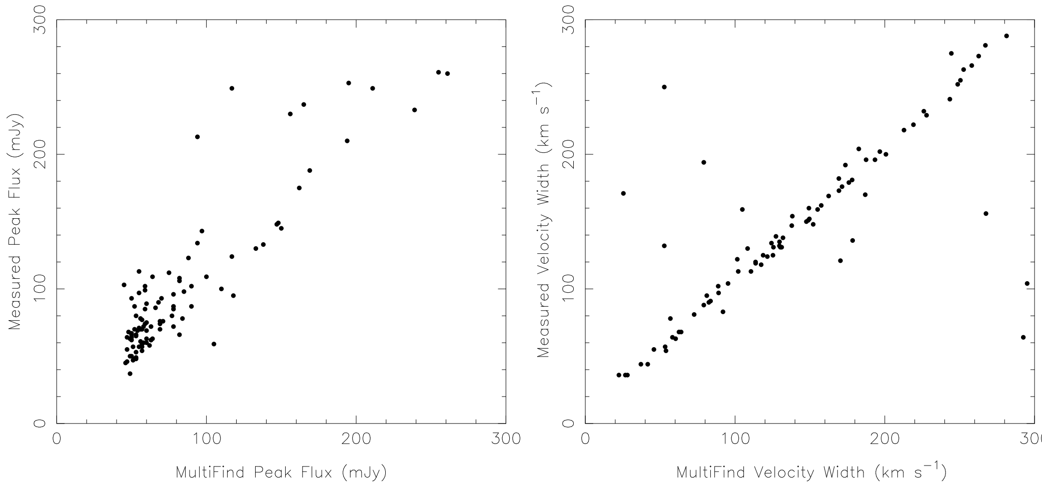

Figure 3.13: Comparison of Input/Output values of of peak flux and velocity width for real galaxies in the SCC region.

3.4.1 Missed Galaxies

The four galaxies that MultiFind did not detect above a peak or average flux of 60 mJy all had triangle profiles, and therefore they were not above the detection limit for two consecutive planes. To better detect galaxies with this profile type, either the search level would have to be lowered, or the criterion that a detection must be above the detection level for two consecutive planes not applied.

3.4.2 False Detections

In addition to the the 'real' galaxies detected by MultiFind, the output list contains a number of 'bad' or 'false' detections. These detections are either noise spikes in the data, the GPS satellite (which wreaks havoc at a velocity of $\sim$ 8300 km s^{-1}), or recombination lines from the Galactic H I. The easiest noise rejection method is to cut the sample at a sensible peak flux limit. This means that a number of real galaxies may be cut from the sample, but it also means that the number of detections the user receives more truly reflects the number of real galaxies detected. A PERL script was written to try and reject the recombination lines, which occur at specific frequencies. The typical number of false detections to real detections in the tested data was $\sim$ 3 to 1. As this is quite a high rate, another PERL script was written to help the user check through the output of MultiFind, and determine which sources are real and eliminate those which are not. This script, PARAMETER.PL, also parameterises those galaxies that are determined to be real, and interrogates the NED database automatically to find any previous identifications for the detection.

3.4.3 Running Time

The time taken for Multifind to run depends on how clean the cube is, and how much computer memory is available. In general, MultiFind was run on a Sparc Ultra 10 machine with 1024MB of RAM. On this machine with no other processes running, it would take between 45 minutes and 2 hours for MultiFind to analyse a cube. The running time is greatly decreased if all visual displays are turned off. However it may be desirable to inspect the galaxies as they are found, so there is provision to turn the displays on and off as necessary.

3.5 Discussion and Possible Improvements

MultiFind has been shown to be effective in detecting galaxies to defined limits in HIPASS data. The efficiency of detecting galaxies at a peak flux of 70 mJy or an average flux of 60 mJy is $\sim 94\%$, however, as can be seen by Figure 3.12, the efficiency of Multifind at the detection limit (in this case 45 mJy), is above 50% for the average flux definition, therefore lowering the flux limit for the search may be possible to detect more galaxies in the data. However, there is a trade–off in lowering the detection level, as the number of false detections will also increase as the detection level is lowered.

The hanning smoothing width used in these tests due to time constraints was just 2, however a better width would be 5 or 6, which would smooth the data by over 500 km s^{-1}. This may enable the signal–to–noise of low–flux, large–width galaxies to be boosted above the detection limit. As galaxies generally do not have a top–hat profile, the smoothing or binning function could be varied in the future to concentrate on detecting certain types of profile shapes. However this raises the problem of pre–empting which galaxies the finder will detect.

There are several improvements to the current finder which could make it faster. Firstly, the algorithm that performs the binning and hanning smoothing on the data sets is not very efficient. Using a different algorithm would speed up the process. An example of a better algorithm is the AIPS++ hanning smoothing or binning function. In addition, eliminating more false detections before the detections are parameterised would increase the speed of MultiFind. The correlation of the galaxies over velocity channels actually takes a significant amount of time, and it is possible this routine would be faster if it was performed in a program written in a compilable language such as FORTRAN or C.

3.6 Endnote

MultiFind has now been run on the whole of the HIPASS dataset, excluding the Galactic plane, by Baerbel Koribalski (Koribalski, 2000). It took several months to run MultiFind (or gfinder as it was so named then) on the $\sim$ 300 cubes, and a sample of the brightest HIPASS galaxies was produced, which contains $\gtrsim$ 600 galaxies above a peak flux of 120 mJy. The work is ongoing, and presently, this detection limit is being lowered to 90 mJy with the aim of producing a catalogue with the brightest $\sim$ 1000 H I galaxies in the southern sky.

Chapter 4

Constructing H I Mass Functions

4.1 The H I Mass Function (HIMF)

The H I mass function, $\phi(M_{\mathrm{HI}})$ describes the fraction of galaxies per H I mass interval per Mpc^{-3}. It is determined from a flux-limited sample such that

$$dN(M_{\mathrm{HI}}) = \phi(M_{\mathrm{HI}})\, V(M_{\mathrm{HI}})\, dM_{\mathrm{HI}} \tag{4.1}$$

where $dN(M_{\mathrm{HI}})$ is the observed number of galaxies within the mass range $(M_{\mathrm{HI}} - dM_{\mathrm{HI}}/2, M_{\mathrm{HI}} + dM_{\mathrm{HI}}/2)$, and $V(M_{\mathrm{HI}})$ is the volume in which galaxies of mass M_{HI} are observable to a limiting mass of $\mathcal{M}_{\mathrm{HI,lim}}$. The HIMF is the H I analogue to the optical Luminosity Function (eg: Efstathiou et al. (1988)). This function can be represented either non-parametrically via a binned histogram, or parametrically by an analytic function, typically a Schechter function (Equation 1.1) or a Gaussian.

The HIMF is necessary to derive quantities such as the H I mass density in the local universe, the distribution of H I masses, or even the expected number of Ly-α absorbers in a volume in the foreground of a redshifted quasar. It is an important ingredient in models of cosmology and galaxy evolution. The integral over the HIMF gives us the total H I gas density in the local universe, which can then be compared to high redshift Ly-α systems, and used to determine models of structure formation and star formation (eg: Storrie-Lombardi et al. (1996)). The HIMF can measure the density of gas-rich, low-mass, low-luminosity dwarf galaxies and LSB galaxies that might have been missed in optical surveys, but may contain a substantial fraction of the local H I mass.

The effect of galaxy clustering may influence the HIMF calculated from a flux limited sample of limited breadth and depth. Two methods of calculating the HIMFare now explored, the $\sum 1/\mathrm{V}_{max}$ method, which assumes a uniform density, and the maximum likelihood method is unaffected by clustering in a sample.

4.2 $\sum 1/\mathrm{V}_{max}$ Method

The $\sum 1/\mathrm{V}_{max}$ method (Schmidt, 1968) requires a flux limited sample, from which the volume within which each detection would just be detectable, to the limit of the survey is calculated. The calculation assumes that the space density of galaxies is uniform (Binggeli et al., 1988).

From Equation 4.1, it can be seen how one might construct an H I mass function using this method. It is as follows:

$$\phi(M_{\mathrm{HI}})dM_{\mathrm{HI}} = \sum_{M-dM/2}^{M+dM/2} \frac{1}{V_{max,i}} \tag{4.2}$$

where dM_{HI} is a discrete mass interval, and $V_{max,i}$ is the maximum volume in which each galaxy, i, of mass M_{HI} could just be detected.

It is usual to determine the $\sum 1/\mathrm{V}_{max}$ in each mass interval, and plot this as a density histogram. Often, a Schechter function (Schechter, 1976) is subsequently fit to the data to provide an analytic form for the HIMF.

The error in each point is governed by the number statistics of the sample, and can be determined using

$$\sigma = \left(\sum \frac{1}{V_{max}^2} \right)^{1/2} \tag{4.3}$$

Thus each galaxy is weighted by its contribution to the sum (Marshall, 1985). At low H I masses, the effective volume surveyed is much smaller than for galaxies of higher H I mass. This means that the error in the low and high mass points can be substantial due to poor number statistics, but the calculated errors reflect this.

A test for the uniformity of a sample requires that the value $\langle V/V_{max} \rangle = 0.5$, where V is the observed volume associated with each galaxy. $\langle V/V_{max} \rangle$ can also be calculated for each mass bin to determine the uniformity of galaxies in each mass interval.

4.2.1 Application to SCC Sample

The main problem faced when applying the $\sum 1/\mathrm{V}_{max}$ method to the SCC sample is quantifying the detection limits for the survey so that the maximum observable volume for each galaxy can be calculated. In a survey such as HIPASS, the detectability of galaxies changes as the quality of the data decreases or increases. Currently we have no measure of the detectability of galaxies throughout the data, so the average noise level in the survey is used to calculate the maximum observable volumes. In addition, no arbitrary signal–to–noise is specified, but a sensitivity function is fit to the data to determine the detectability limits.

The two main observable parameters of galaxies in HIPASS are peak flux and velocity width. The extent that the velocity width of a galaxy affects its detectability is unclear. Galaxies of different profile types will depend on velocity width to differing extents for their detectability. This effect has been shown in the tests on the reliability of the MultiFind galaxy finding routine (Chapter 3).

If the detectability of a galaxy depends only on the peak flux density of that galaxy, then the maximum detectable distance can be calculated simply by using the inverse square law

$$D_{max} = D\sqrt{\frac{S_p}{S_{lim}}} \tag{4.4}$$

where S_p is the peak flux density of an individual galaxy, S_{lim} is the limiting peak flux density for the galaxy, and D is its distance.

If the detectability of a galaxy depends on its total flux and velocity width, then we use the H I mass of a galaxy:

$$M_{\rm HI} = 2.356 \times 10^5 D^2 \int S\,dv\, M_\odot \tag{4.5}$$

where D is the distance to a galaxy in Mpc, and S is the spatially integrated flux density in Jy, and dv is the velocity extent of the galaxy in km s^{-1}. For a top–hat profile galaxy, $\int S\,dv \cong S_{ave} \times \Delta V$, where S_{ave} is the average flux density for the galaxy in Jy, and ΔV is its velocity width in km s^{-1}. The average flux density for a galaxy is given by

$$S_{ave} = \frac{M_{\rm HI}}{2.356 \times 10^5 D^2 \times \Delta V} \tag{4.6}$$

The maximum distance, D_{max}, that a particular galaxy of mass $M_{\rm HI}$, could be detected depends on the detectability function of the survey. The flux from a galaxy must be detected above the noise in the data. The noise is reduced by smoothing the data, and the maximum signal–to–noise of a galaxy occurs when the smoothing width, ΔV_s, equals the velocity width of the galaxy. The noise in this case depends on the number of smoothed channels, N, and the RMS in of the unsmoothed data, σ_0. With smoothing, the noise is reduced by a factor N^{β}, where β = 0.5 for uncorrelated noise. Therefore the noise for a smoothed data–cube is

$$\sigma = \sigma_0/N^{\beta} = \sigma_0(\Delta V_s/\Delta V_0)^{-\beta} \tag{4.7}$$

where ΔV_0 is the velocity channel separation in the unsmoothed cube. So the signal–to–noise ($\mathcal{SN} = S_{ave}/\sigma$) for a galaxy in an optimally smoothed cube is

$$\mathcal{SN} = \frac{M_{\mathrm{HI}}}{2.356 \times 10^5 D^2 \times \Delta V \times \sigma_0 \times (\Delta V_s/\Delta V_0)^{-\beta}} \tag{4.8}$$

The derivation of the completeness of a catalogue will determine the signal–to–noise limit for the data–set, or a limiting average flux density, $S_{ave,lim} = \mathcal{SN} \times \sigma_0$.

Thus the maximum detectable distance can be derived from Equation 4.8, so that

$$D_{max} = \sqrt{\frac{M_{\mathrm{HI}}}{2.356 \times 10^5 S_{ave,lim} \Delta V (\Delta V_s/\Delta V_0)^{-\beta}}} \tag{4.9}$$

where $S_{ave,lim}$ is in Jy, and D_{max} is in Mpc. For the case of unsmoothed data, $\Delta V_s = \Delta V_0$, and in the case of optimally smoothed data, $\Delta V_s = \Delta V$.

Equation 4.9 can be modified according to the determined detectability function, ie for a sample where the detectability is determined as a function of integrated flux, the corresponding D_{max} is

$$D_{max} = \sqrt{\frac{M_{\mathrm{HI}}}{2.356 \times 10^5 S_{int,lim} (\Delta V_s/\Delta V_0)^{-\beta}}} \tag{4.10}$$

where $S_{int,lim}$ is the limiting integrated flux, in Jy km s^{-1}, that a galaxy could contain and still be detectable in the sample.

As HIPASS is a bandwidth-limited survey, if the calculated D_{max} is greater than the bandwidth limit, D_{BW}, then $D_{max} = D_{BW}$. For HIPASS, $D_{BW} = 169$ Mpc.

Once the maximum detectable distance has been determined for each galaxy in the sample, this is converted into a maximum detectable volume

$$V_{max} = \frac{1}{3} D^3_{max}\, \Omega \tag{4.11}$$

where V_{max} is the maximum volume a galaxy could have been detected in, and Ω is the solid angle of the survey.

Finally, Equation 4.2 is invoked to produce the HIMF in the form of a binned histogram which can then be fit with an analytic function.

4.2.2 Previous Examples using $\sum 1/\mathrm{V}_{max}$

A number of blind H I surveys have been undertaken recently, and the HIMFs calculated typically use the $\sum 1/\mathrm{V}_{max}$. Two of these HIMFs were constructed using blind H I samples from the Arecibo telescope, but with different techniques for calculating galaxy detectability. This lead to different results when determining the HIMF.

Zwaan et al. (1997) (hereafter ZBSS) constructed an HIMF from a sample of 66 galaxies, in a region of sky $\sim$ 65 square degrees in area. Their observations were made in drift-scan mode, and produced two data cubes which consisted of strips of constant Declination and varying Right Ascension. The detectability of galaxies in this sample was determined by the ratio of signal–to–noise, and also by the sensitivity of Arecibo to detect galaxies offset to the centre of the beam. The calculation of D_{max} in was made using an Equation similar to Equation 4.10, where $\beta = 0.5$. The survey also needed to take into account sources that did not lie in the primary beam of the telescope, so their detectability was modified by another function, $\mathrm{I}(\theta)$, which depended on the angle, θ, that the source was detected from the primary beam. The HIMF that ZBSS calculate has Schechter parameters of $\alpha = 1.2$, and $\log M_{\mathrm{HI}}{}^* = 9.8$. No up–turn in galaxy numbers is seen at the low-mass end of the H I mass function.

Schneider et al. (1998) (hereafter SSR) made a blind survey using pointed observations with the Arecibo telescope, rather than using the telescope in drift-scan mode. This ensured that all of the sources detected were within the primary beam of the telescope, and a beam correction was not necessary. In addition, the sample from ZBSS was added, using only galaxies detected in the main beam, giving a final sample size of 120 sources. The detectability function that SSR define uses a new parameter, V_{tot}, which describes the "completeness volume" in which a source should "on average" have been detected over the entire survey area, taking into account offsets from the beam centre and other effects. From an analysis of V/V_{max} for different velocities, SSR find that $\beta = 0.25$ compared to the $\beta = 0.5$ used by ZBSS. This gives an increased weighting to low-mass galaxies. In addition, SSR correct for the density of galaxies in the region from optical catalogues, although these corrections make little difference to the final HIMF. The HIMF calculated for this sample shows a steep upturn at an H I mass of $\sim 10^7\ M_\odot$, similar to that seen in the optical by Loveday (1997).

The ZOA multibeam shallow–survey (Henning et al., 2000) (hereafter ZOA) find that the detectability for galaxies in their data involves both total flux and velocity width such that $\beta = 0.3$. This is similar to the form found by SSR. The Schechter parameters found from this survey are similar to SSR, with $\alpha = 1.51$, log ${M_{\rm HI}}^* = 9.7$ and $\theta^* = 0.006$. This is a much steeper faint–end slope than found by ZBSS.

4.3 Maximum Likelihood Analysis

The maximum likelihood method for determining the characteristics of a dataset was developed by Sandage et al. (1979). The method assumes that the HIMF has a universal form (ie that the relative densities of galaxies of different H I masses are the same everywhere). Then the probability that a galaxy at a distance r_i has the observed mass $M_{{\rm HI},i}$, in a complete, H I flux-limited catalogue, which is:

$$p(M_{{\rm HI},i}|r_i) = \begin{cases} \frac{\phi(M_{{\rm HI},i})}{\Phi(M_{{\rm HI},lim}(r_i))} & \text{if} M_{{\rm HI},i} \geq M_{{\rm HI},lim}(r_i), \\ 0 & \text{otherwise.} \end{cases} \tag{4.12}$$

where $\phi(M_{{\rm HI},i})$ and $\Phi(M_{{\rm HI},lim})$ are the assumed forms of the HIMF and its integral, and $M_{{\rm HI},lim}(r_i)$ is the minimum H I mass detectable at a distance r due to the flux limitation.

The probability, $p(M_{{\rm HI},i}|r_i)$ is calculated for each galaxy in the sample, and the the product of all the probability densities is derived,

$$\mathcal{P} = \prod_i p(M_{{\rm HI},i}|r_i) \tag{4.13}$$

Solanes et al. (1996) determined a computationally convenient likelihood function:

$$\Lambda = -2 \sum_i ln\,\mathcal{P} \tag{4.14}$$

By miminising this function for a range of parameters, the maximum likelihood estimate for the sample can be calculated.

The maximum likelihood method is useful in determining the HIMF without any knowledge of the density or clustering of galaxies. However, the method assumes a certain functional fit, and any fine detail is lost (eg turn-up at the faint end of the HIMF). In addition, the density of galaxies of different H I masses may not be a constant, for example, in clusters compared to the general field. This effect might be averaged out in a sample of sufficient size.

4.3.1 Application SCC sample

To calculate the maximum likelihood function for a sample a functional form is needed. We will use the Schechter function, although any functional form such as a Gaussian or a power-law is acceptable.

$\phi(M_{\mathrm{HI},i})$ is the value of the function at mass $M_{\mathrm{HI}}=$ for galaxy i. Therefore using the Schechter function,

$$\phi(M_{\mathrm{HI},i}) = t^{\alpha}\, e^{-t} \tag{4.15}$$

where $t = M_{\mathrm{HI}}/M_{\mathrm{HI}}{}^{*}$.

$\Phi(M_{\mathrm{HI},lim}(r_i))$ is the integral of $\phi(M_{\mathrm{HI},i})$

$$\Phi(M_{\mathrm{HI},lim}(r_i)) = \int_t^{\infty} t^{\alpha}\, e^{-t}\, dt \tag{4.16}$$

We now integrate so that it is computable.

$$\Phi(M_{\mathrm{HI},lim}(r_i)) = \left[e^{-t}\, \frac{1}{\alpha+1}\, t^{\alpha+1} \right]_t^{\infty} + \int_t^{\infty} \frac{1}{\alpha+1}\, t^{\alpha+1}\, e^{-t}\, dt \tag{4.17}$$

The integral of the second term of Equation 4.17 is the gamma function: $\Gamma(a,b) = \int_b^{\infty} t^{a-1}\, e^{-t}\, dt$.

If we let $a-1 = \alpha+1$, then

$$\int_{M_{\mathrm{HI},lim}}^{\infty} t^{\alpha}\, e^{-t}\, dt = -e^{-t} \frac{1}{\alpha+1}\, t^{\alpha+1} + \frac{1}{\alpha+1} \Gamma(a,t) \tag{4.18}$$

and substituting back for $a = \alpha+2$,

$$\Gamma(\alpha+1,t) = \frac{1}{1+\alpha}\left[\Gamma(\alpha+2) - e^{-t}\,t^{\alpha+1}\right] \tag{4.19}$$

which is now a computable function.

The probability density, $\mathcal{P}$ is calculated for a range of values of α and $M_{\rm HI}{}^*$, by evaluating $p(M_{{\rm HI},i}|r_i)$ for each galaxy in the sample and taking the product of these individual probability densities. The minimum value of Λ gives the maximum likelihood for the sample, and error contours are χ^2 errors with 2 degrees of freedom.

4.3.2 Previous Examples using Maximum Likelihood

Due to small number statistics, the maximum likelihood method has not been used on blind H I surveys previously. ZBSS found upon experimenting with this method that the faint end slope derived was not robust when applied to samples with small number statistics at edges of the H I mass distribution.

Solanes et al. (1996) used maximum likelihood methods to determine the H I mass content of spiral galaxies from a range of optical catalogues of galaxies. They find that their samples are equally well fit by Gaussians as Schechter functions, and that the Sb types of galaxies typically have a larger H I mass than other spiral types.

The robustness of the the maximum likelihood method and the $\sum 1/{\rm V}_{max}$ method will now be tested with simulations.

4.4 Simulations

Simulations were performed to check the derived mass functions produced by the two methods. The simulations involved creating datasets with known properties, and analysing them using the HIMF routines, and determining whether the resulting HIMFs reflected the properties of the input catalogue.

4.4.1 Details of the Simulator

A simulator program was created by Lister Staveley-Smith. This program, SIMUL, generates a random list of galaxies, using a Schechter function as input. Galaxies of particular masses were created, then randomly placed throughout the chosen volume. A velocity width was assigned to each galaxy using the Mass-Velocity width relation of $\Delta V = 0.16 M_{\rm HI}{}^{1/3}$ (Fisher and Tully, 1981b). There is some scatter in this relationship, so each velocity width was dispersed around a Gaussian of mean 0, variance 1. The total flux of each detection was calculated by dividing the mass of the galaxy by its distance squared. The peak flux density was given by dividing the integrated flux by the velocity width of the simulated galaxy. The final list of galaxies produced was clipped above a user defined cut-off function, usually defined as a peak flux limit. The typical number of galaxies produced in a simulated data set ranged from 100 to 300 galaxies, with more galaxies being produced when the input low–mass slope (α) was steep.

4.4.2 $1/\mathrm{V}_{max}$ Tests

The simulations tested whether the galaxy sample input into the $1/\mathrm{V}_{max}$ method was represented by the HIMF that was produced. Several simulation data–sets were produced by varying the input low–mass slope (α) and characteristic mass ($M_{\rm HI}{}^*$). These data–sets were then used as an input into the $\sum 1/V_{max}$ mass function, and finally the Schechter parameters were derived for the HIMF. Figure 4.1 shows the input versus output Schechter parameters. These match very well for all simulation input values of α and $M_{\rm HI}{}^*$.

4.4.3 Maximum Likelihood Tests

The maximum likelihood method was tested in the same manner as the $\sum 1/V_{max}$ method, varying α and $M_{\rm HI}{}^*$. The output values can be seen in Figure 4.2. These values are not as accurate as the $\sum 1/V_{max}$ method, although a good agreement is still seen between the input and output of the method. It is likely that low number statistics are biasing the output in from the simulations, as a high integrated flux cut-off was used and the typical simulation sample size was ~ 150. The maximum likelihood method tended to give an α that was flatter than the input, and a $M_{\rm HI}{}^*$ that was slightly larger than the input.

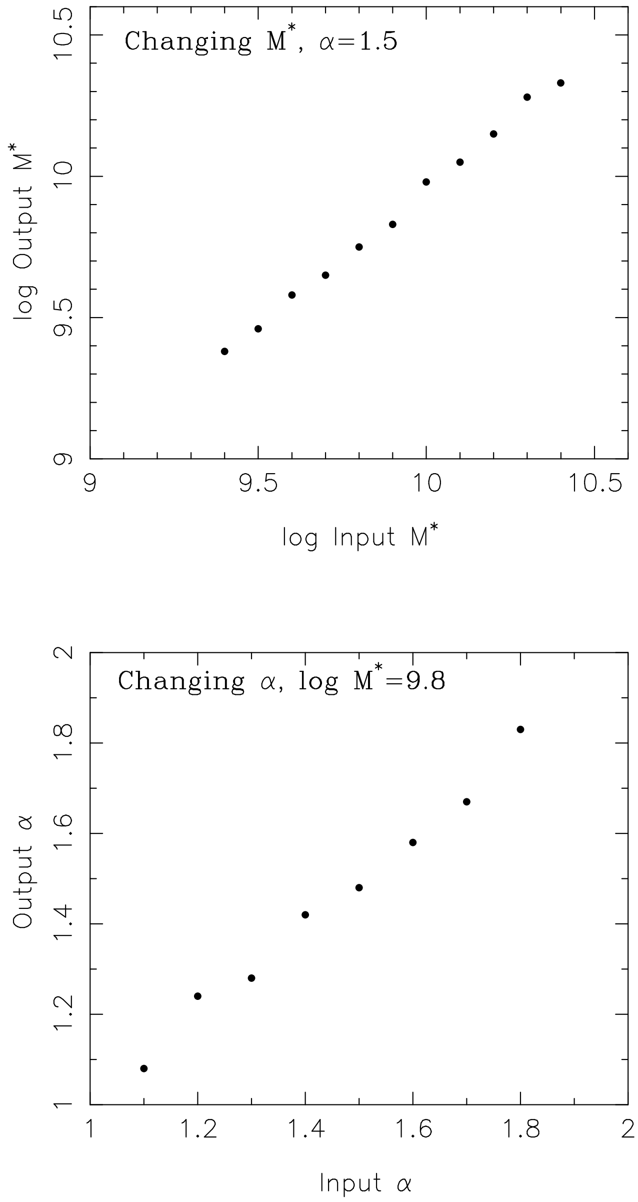

Figure 4.1: Testing the $\sum 1/V_{max}$ method, with varying $M_{\rm HI}{}^*$ [Top] and α [Bottom]. Good agreement is seen between the input and output values.

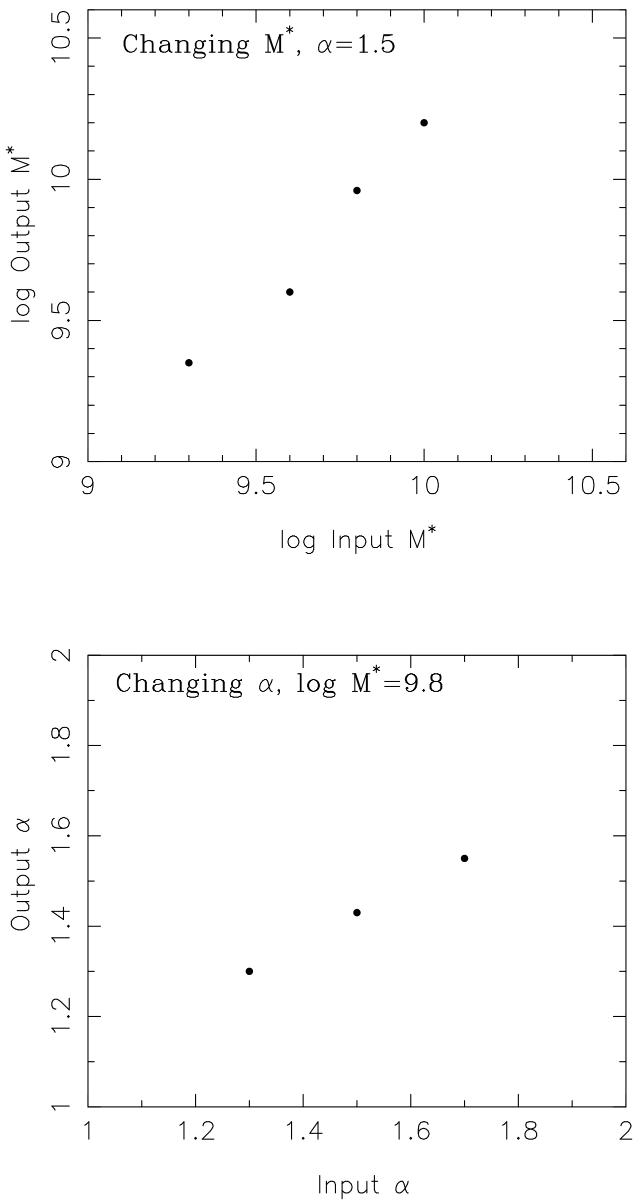

Figure 4.2: Testing the maximum likelihood method, with varying $M_{\rm HI}{}^{*}$ [Top] and α [Bottom]. Good agreement is seen between the input and output values, however α tends to be measured more shallower than the input value.

Chapter 5

H I Mass Distribution of the SCC

5.1 Introduction

The H I mass function (HIMF) for the SCC sample is determined using the methods described in the previous chapter. Before any meaningful HIMF could be derived, the completeness limit for the sample needed to be characterised. This was done using three different flux estimates, and the resulting HIMFs from each completeness limit are shown for comparison. The HIMF was then tested for its robustness. The effect of binning, and consistency in the detection limit and V_{max} determination was investigated. As the $\sum 1/V_{max}$ method may be sensitive to large–scale structure, the maximum likelihood method was also used to estimate the shape of the H I mass distribution. In addition, a density independent method of calculating the $\sum 1/V_{max}$ method was investigated.

As the SCC sample has been classified into optical types, the H I mass distribution is investigated for galaxies of different Hubble type. Finally the total H I density is calculated and implications are discussed.

5.2 Completion Limits for the SCC Sample

The SCC catalogue contains galaxies with a range of peak and integrated fluxes, and it is likely that not all galaxies at the lowest fluxes have been detected. It is therefore necessary to determine the completeness level of the catalogue before any statistical measures can be calculated. Once the completeness level for the survey is determined, either the distribution of 'missing' galaxies can be modelled and applied to the whole sample, or we can determine the level at which the sample is complete and ignore any data below this cut-off level.

To determine the completeness of the sample, the H I parameters that govern the detectability of a galaxy need to be characterised. The peak or average flux might be the primary measure of the detectability of a galaxy, but the dependence on velocity width must also be investigated. The orientation of a galaxy determines its peak flux and velocity width, so ideally the detectability would be related to the intrinsic properties of a galaxy, such as its integrated flux or H I mass.

5.2.1 Peak Flux versus Average Flux

The peak flux of galaxies in the SCC sample is defined as the maximum flux pixel in the galaxy's spectrum. Thus the peak flux of a galaxy may be influenced by the presence of a noise spike in the H I spectrum. As an example, Figure 5.1 shows the spectrum of HIPASS J0341-77. The detectability of the galaxy is unlikely to be governed by the spike in the spectrum. A more robust measurement of the flux might be the average flux. For this reason both the peak flux and average flux limits will be investigated for the sample.

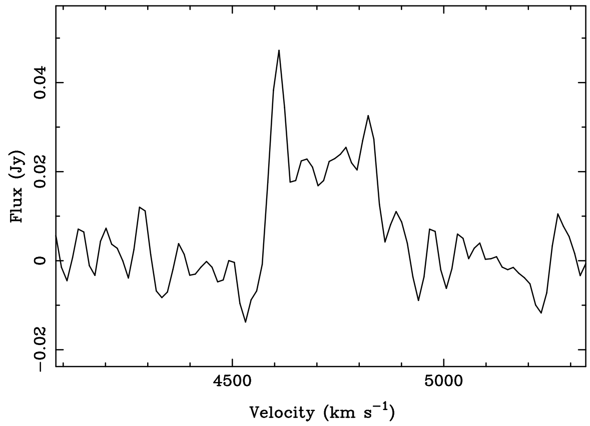

Figure 5.1: HIPASS J0341-77. One spike in the spectrum which defines the peak flux of the spectrum may not be a good representation of the detectability of the galaxy in the data.

First of all, the distribution of fluxes was compared to that predicted by theory. The distribution of peak and average flux is shown in Figure 5.2. Theoretically, the number of galaxies of flux S, increases as[1] $N(S) \sim S^{-5/2}$. This function was fit against the flux histograms, and is shown as a black line in the plots. The incompleteness at the low-flux end of the histogram was modelled by the green curve[2] which was manually adjusted until the combined blue curve gave a good fit of the data.

[1] Assuming Euclidean geometry, and a homogeneous distribution

[2] The functional form of the green curve is $10(1.0{+}\cos(\pi/2.0((S - S_{cut})/S_{wid} - 1.0)))$, where S_{cut} marks the end of the cosine curve, and S_{wid} is the width of the cosine curve

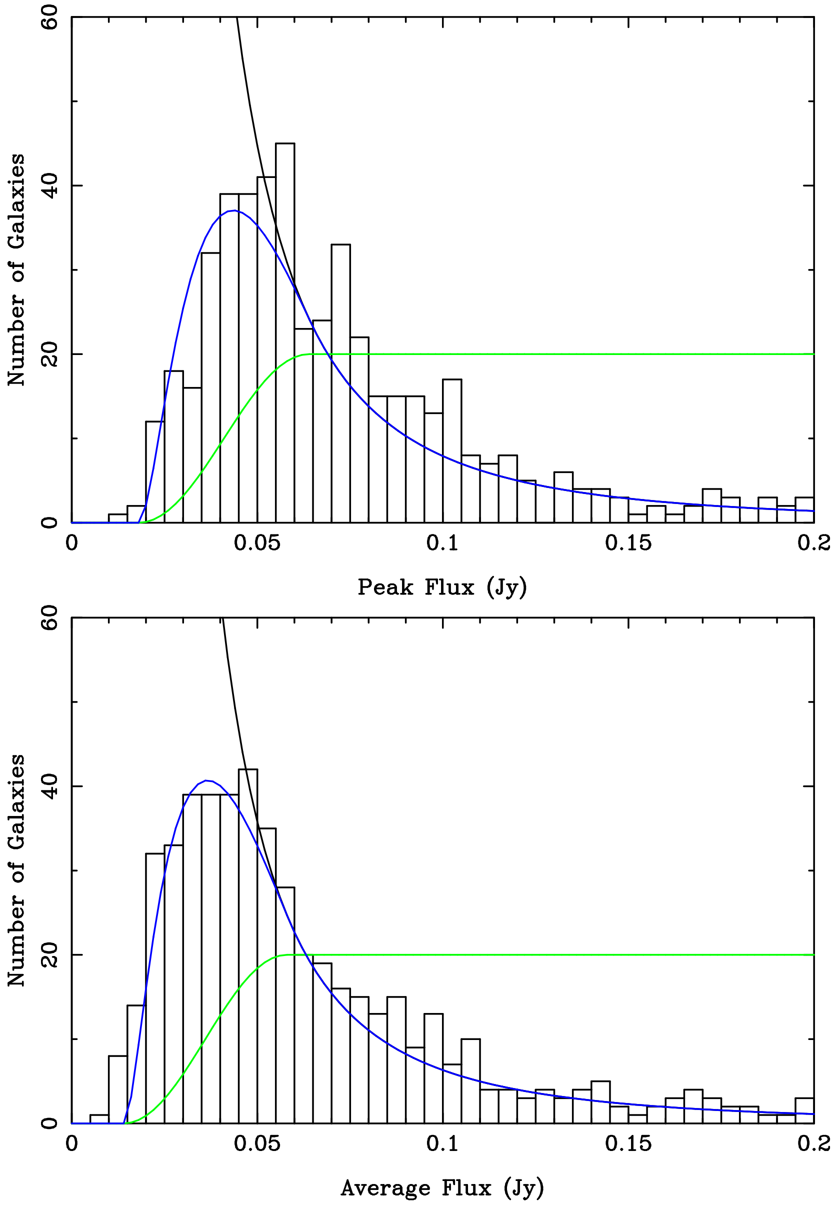

Figure 5.2: Distribution of peak flux [Top] and average flux [Bottom] for the SCC sample. The curves show modelling of the completeness of the data. The black curve shows $N(S) \sim S^{-5/2}$ (see text), the green curve was manually fit to the data, and the blue curve is the combined completeness function.

The flux at which the black and blue curves meet represents the theoretical 100% completeness limit for the sample. This occurs at a peak flux of 64 mJy for the peak flux histogram, and at an average flux of 58 mJy for the average flux histogram. The sample is approximately 50% complete at a peak flux of 41 mJy and average flux of 36 mJy. In theory, it would now be possible to apply this function to the data to correct for galaxies missed in the survey. However, the relationship for expected number of galaxies only holds in a uniformly distributed sample, and as we have already seen, the SCC sample shows strong evidence of clustering. Also, the functionality of the expected flux drop-off due to sensitivity and bandwidth limitations is not well known. Due to these uncertainties the functions will not be used to correct the data, but the determined 100% completeness limits will be tested for robustness, then applied to the data.

5.2.2 Dependence on Velocity Width

In Chapter 3 it was seen that a loose correlation exists between the detectability of low flux sources and velocity width. The automatic finder was more efficient in detecting galaxies with low peak fluxes that had large velocity widths rather than those with small velocity widths. If there is a dependence on the velocity width of a galaxy for its detectability, then we would expect that more low-flux, large velocity width galaxies to be detected than low-flux, small velocity width galaxies.

Figure 5.3 shows peak and average flux versus ΔV_{50} for the SCC sample. No correlation is seen between the flux of the detected galaxies and their velocity width. Therefore the galaxies in this sample appear to just be selected according to their peak or average flux. The completeness limits determined from the flux histograms are shown on Figure 5.3.

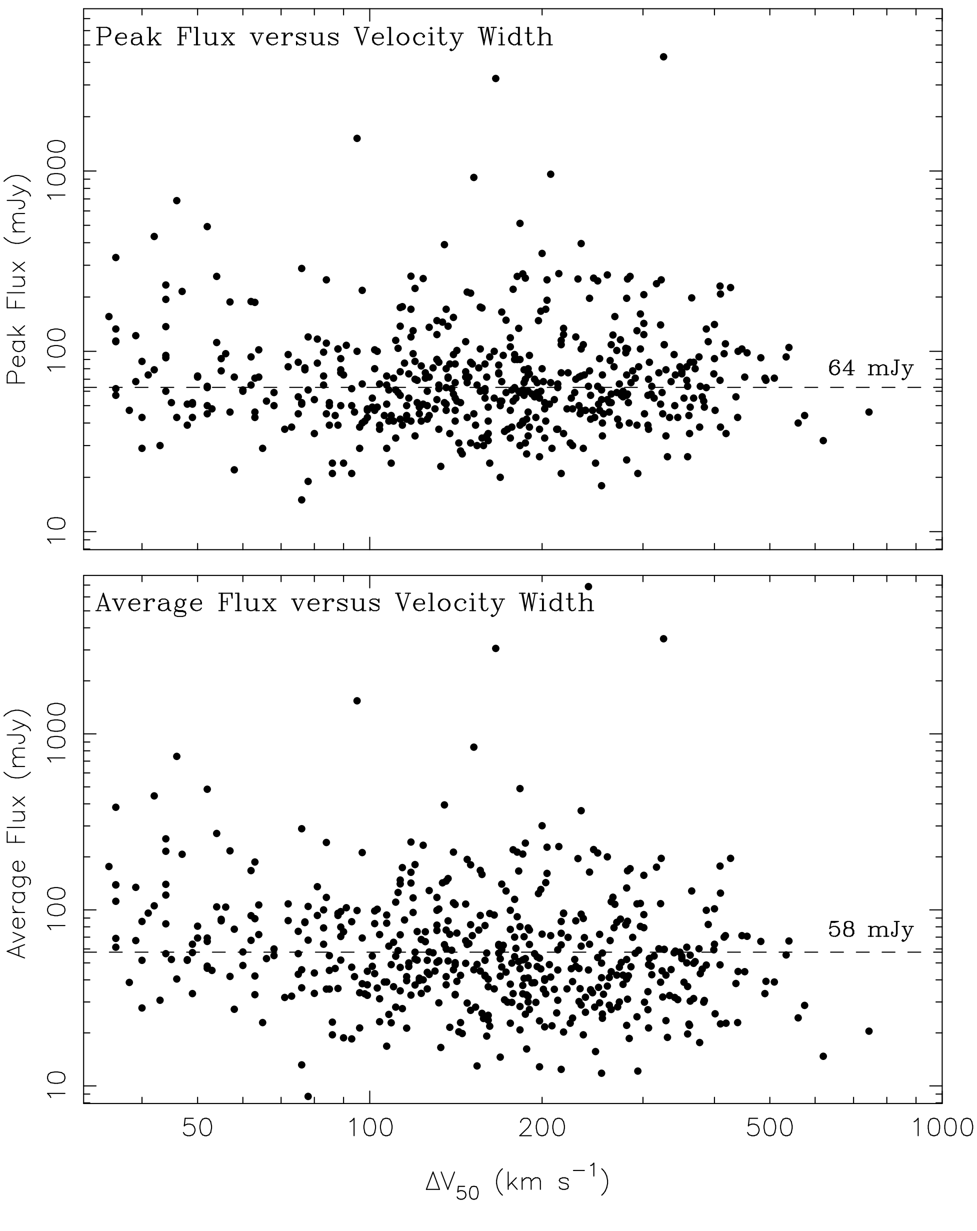

Figure 5.3: Average and Peak Flux versus 50% Velocity Width for the SCC Sample. No correlation of peak flux with velocity width is seen in either of the panels. The dashed lines show the 100% completeness limits for peak flux (64 mJy) and average flux (58 mJy).

5.2.3 Integrated Flux

A better way of determining completeness limits for a sample would be to define the sample in terms of the integrated flux of a galaxy. This is the only parameter that will not change with differing inclination of a galaxy, or velocity resolution of the observations. However, the total flux of a galaxy depends on a combination of velocity width and peak flux, meaning that a galaxy with a peak flux of 100 mJy and a velocity width of 50 km s^{-1} will have the same integrated flux as a galaxy with a peak flux of 20 mJy and a velocity width of 250 km s^{-1}. The first galaxy will be much easier to detect in our data than the second galaxy. For a high limit of integrated flux, this is not problematic, but as the integrated flux level decreases, the detectability of large velocity width galaxies also decreases, meaning that the detectability function cannot be modelled as a function of integrated flux alone. Figure 5.4 shows the same functions as previously scaled to a histogram of the integrated fluxes of galaxies in the sample. The completeness limit given by these functions is $S_{int} \sim 9.2$ Jy km s^{-1}. The fit for the high–flux galaxies was slightly below optimal, and a more sophisticated model would be needed to fit the data better here.

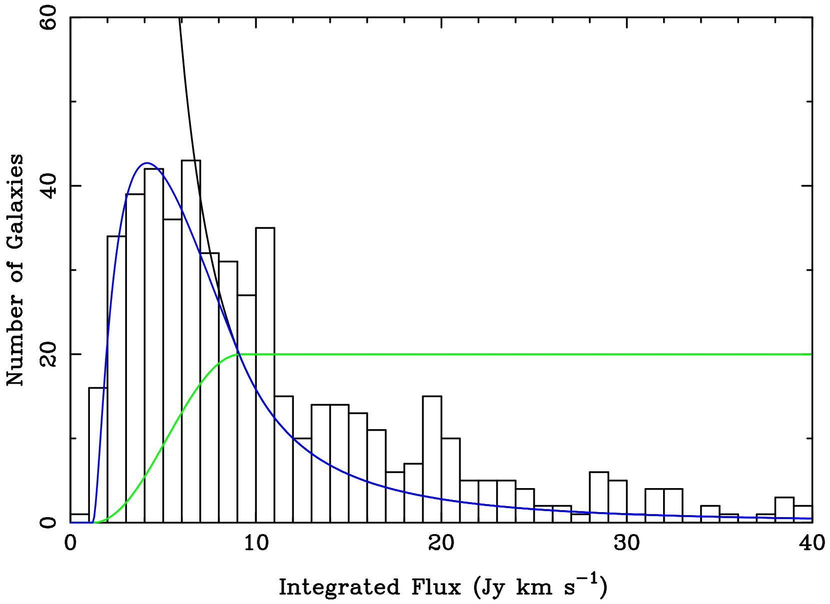

Figure 5.4: Distribution of integrated flux, with the expected distribution of flux shown. The completeness limit from these functions is $S_{int} \sim 9.2$ Jy km s^{-1}.

Figure 5.5 shows the integrated flux versus ΔV_{50} for the SCC sample. The completeness limit of 9.2 Jy km s^{-1} is shown, as well as the average flux limit of 58 mJy. The integrated flux limit at the shown completeness limit seems to be adequate, however, the average flux limit follows the trend of the data better.

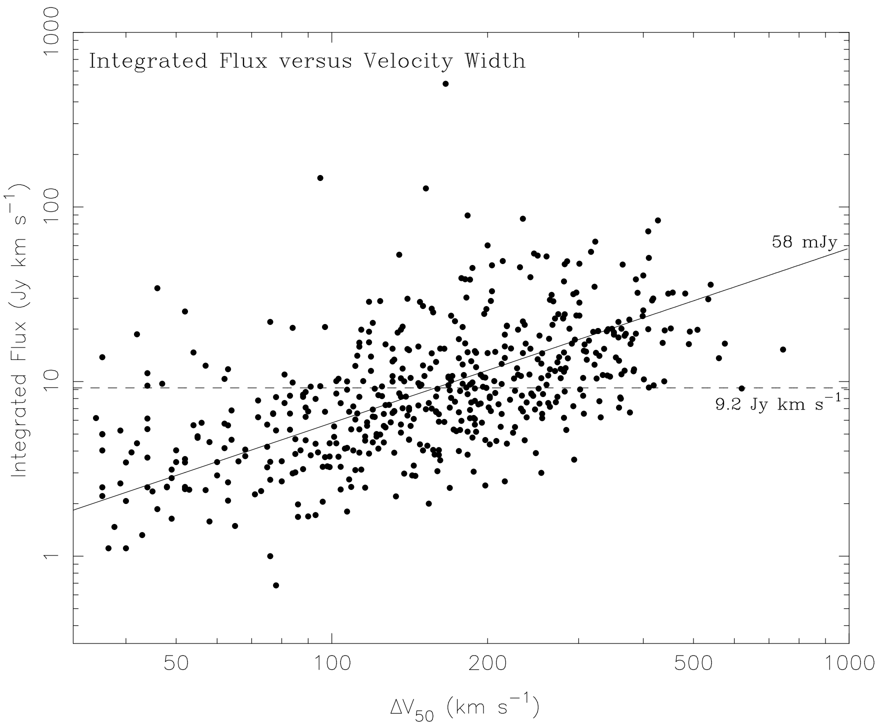

Figure 5.5: Integrated Flux versus 50% Velocity Width for the SCC Sample. The dashed line shows an integrated flux limit of 9.2 Jy km s^{-1}, and the solid line shows an average flux limit of 58 mJy.

5.3 Velocity and Distance Estimates

The radial velocities determined from HIPASS need to be corrected for the motion of the Sun in the Galaxy, and the Galaxy in the Local Group. There are several possible ways of making this correction, and up this point in this thesis, the IAU correction has been used (Equation 2.1). One alternative correction is to use the POTENT program (Bertschinger et al., 1990), which makes corrections for the local galaxy motion from the variation in the Cosmic Microwave Background (CMB). Another correction that could be applied was derived by Yahil et al. (1977) (YTS), and is given by:

$$V_{LG} = V_{sys} - 79 \cos l \cos b + 296 \sin l \cos b - 36 \sin b \tag{5.1}$$

which is similar to the IAU convention method.

Figure 5.6 shows the differences in velocity estimates for the sample using the different methods. The uncorrected velocities are shown for comparison: these are higher than the corrected velocities below a velocity of 1000 km s^{-1}, due to the direction of the SCC with respect to the motion of the Galaxy in the Local Group.

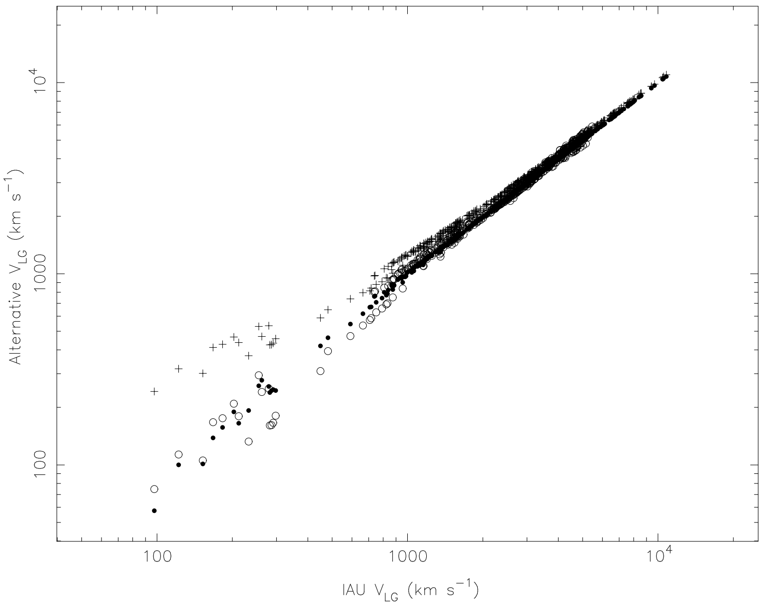

Figure 5.6: Comparing different velocity estimators. Uncorrected velocities are shown with the crosses, the filled circles show the YTS correction and the open circles show the POTENT velocity correction.

Above a velocity of $\sim$ 1000 km s^{-1}, all corrections produce essentially the same output velocities. The difference the velocity corrections make on the mass determined for a galaxy are shown in Figure 5.7. The difference in the calculated H I masses for the different velocity calculations are small, but some of the POTENT corrections at low velocities produce smaller mass galaxies than the other methods.

The velocity correction chosen for the calculation of the HIMF is the IAU correction. This is the most conservative correction and was chosen above the POTENT correction which seems to produce a divergence in the output velocities, that is especially visible in Figure 5.7.

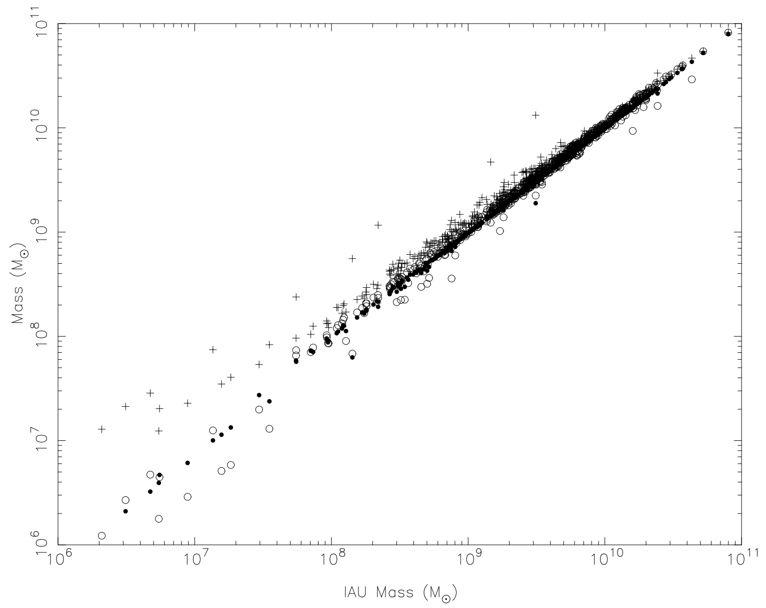

Figure 5.7: Comparison of the derived mass from the different Local Group velocity estimators. The open circles show the mass calculated using the POTENT velocity, the filled circles show the mass using the YTS method, and the crosses show the mass calculated with uncorrected velocity.

5.4 The HIMF for the South Cap region

The HIMF for the SCC sample has been determined using the $\sum 1/V_{max}$ method (Schmidt, 1968). Figure 5.8 shows the HIMF for the SCC sample, after applying an average flux limit of 58 mJy. Equation 4.9 was used to determined the maximum detectable distance and hence volume for each galaxy in the sample. There are 231 galaxies used to make the function, which is less than half of the SCC sample.

A traditional fit to the data is made with a Schechter function (Schechter, 1976):

$$\Phi(M_{\rm HI})dM_{\rm HI} = \theta^*(\frac{M_{\rm HI}}{{M_{\rm HI}}^*})^{-\alpha}exp(\frac{-M_{\rm HI}}{{M_{\rm HI}}^*})d(\frac{M_{\rm HI}}{{M_{\rm HI}}^*}) \quad (5.2)$$

where α is the faint–end slope, ${M_{\rm HI}}^*$ is the characteristic H I mass and θ^* is the normalisation.

A least squares fit of a Schechter function to the SCC data gave the following Schechter parameters: $\alpha = 1.52 \pm 0.07$, $\log {M_{\rm HI}}^* = 10.1 \pm 0.08$ $M_\odot$, and normalisation, $\theta^* = 3.2 \pm 0.5 \times 10^{-3}$. The solid line on Figure 5.8 shows the Schechter function with these parameters.

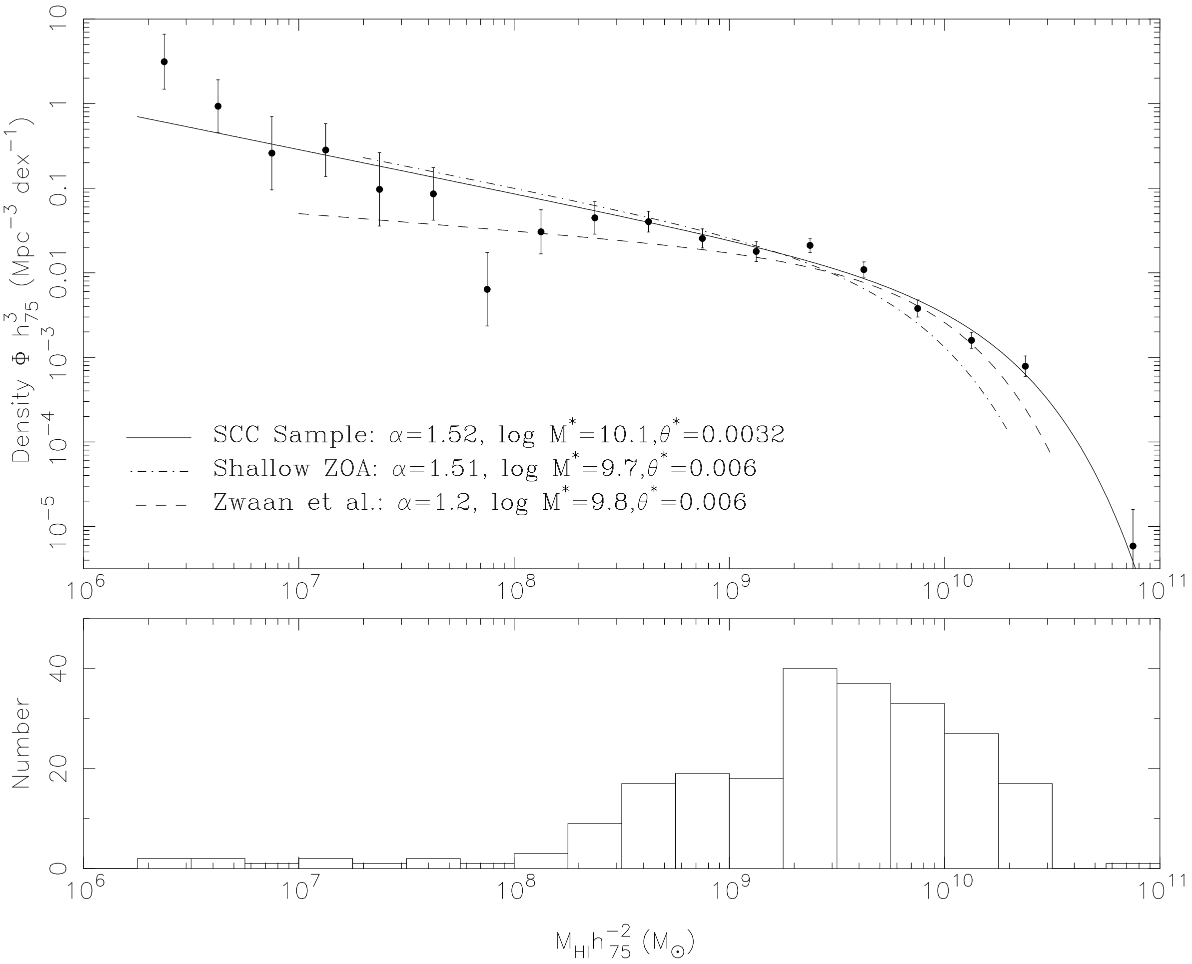

Figure 5.8: HIMF for SCC sample. The dotted curve is from Zwaan et al. (1997) the dot-dash curve is from the shallow ZOA survey (Henning et al., 2000), and the solid line shows the Schechter curve from a least squares fit to the SCC data.

For comparison, the Zwaan et al. (1997) Schechter fit (dashed line) and the recent calculation of the HIMF from the Zone of Avoidance (dot-dashed line) are shown on the same plot. The ZOA mass function was derived from data also observed using the Multibeam receiver at Parkes (Henning et al., 2000). The HIMF derived from the SCC sample has a steeper slope to that of Zwaan et al. (1997), and is similar to the ZOA calculation.

The fitted Schechter curve seems to match the data quite well, apart from a 'dip' in the H I density at an H I mass of $\sim 10^8 M_\odot$. No matter how the HIMF is constructed, this dip is always present, which suggests that a better fit to the data may comprise of two distinct functional slopes. This shape is similar to that found with optical luminosity functions (eg:

Trentham (1998); Loveday (1997)), and may be fit better by a power–law and flat Schechter function combination.

The SCC sample has a higher $M_{\rm HI}{}^*$ than other previous surveys. This may be due to this survey including a significant number of galaxies with $M_{\rm HI} > 10^{10}\ M_\odot$. The number statistics of the low mass bins are poor, with just 10 galaxies with $M_{\rm HI} < 10^8 M_\odot$.

5.4.1 The effect of Binning

When using the $\sum 1/V_{max}$ method, the data has to be binned, and this may have an effect on the final mass function. The data was thus binned with different widths and offsets, and the Schechter parameters calculated. Table 5.1 shows these Schechter parameters derived from a number of different binnings of the data.

Start Bin log $M_{\rm HI}$ ($M_\odot$)	**Bin Width** log $M_{\rm HI}(M_\odot)$	$\boldsymbol{\alpha}$	$\boldsymbol{M_{\rm HI}{}^*}$ log $M_{\rm HI}(M_\odot)$	$\boldsymbol{\theta^*}$
5.9	0.25	1.48	10.005	0.0041
6.0	0.25	1.507	10.037	0.0035
6.1	0.25	1.533	10.135	0.0027
5.9	0.5	1.571	10.264	0.0020
6.0	0.5	1.434	9.939	0.0050
6.2	0.5	1.490	10.061	0.0035

Table 5.1: Effect of binning on the Schechter Parameters

There is a small variation seen in the Schechter parameters with varying binning, but these values are generally within the original quoted errors.

5.4.2 Testing the cut-off limits

The derived Schechter parameters for the HIMF can be tested for robustness by varying the flux cut–off limit. In particular, it is expected that the $\langle V/V_{max}\rangle$ value for the survey will be 0.5 when the chosen sample is complete. Figure 5.9 shows different values for $\langle V/V_{max}\rangle$, and α and $M_{\rm HI}{}^*$ with different cut–offs.

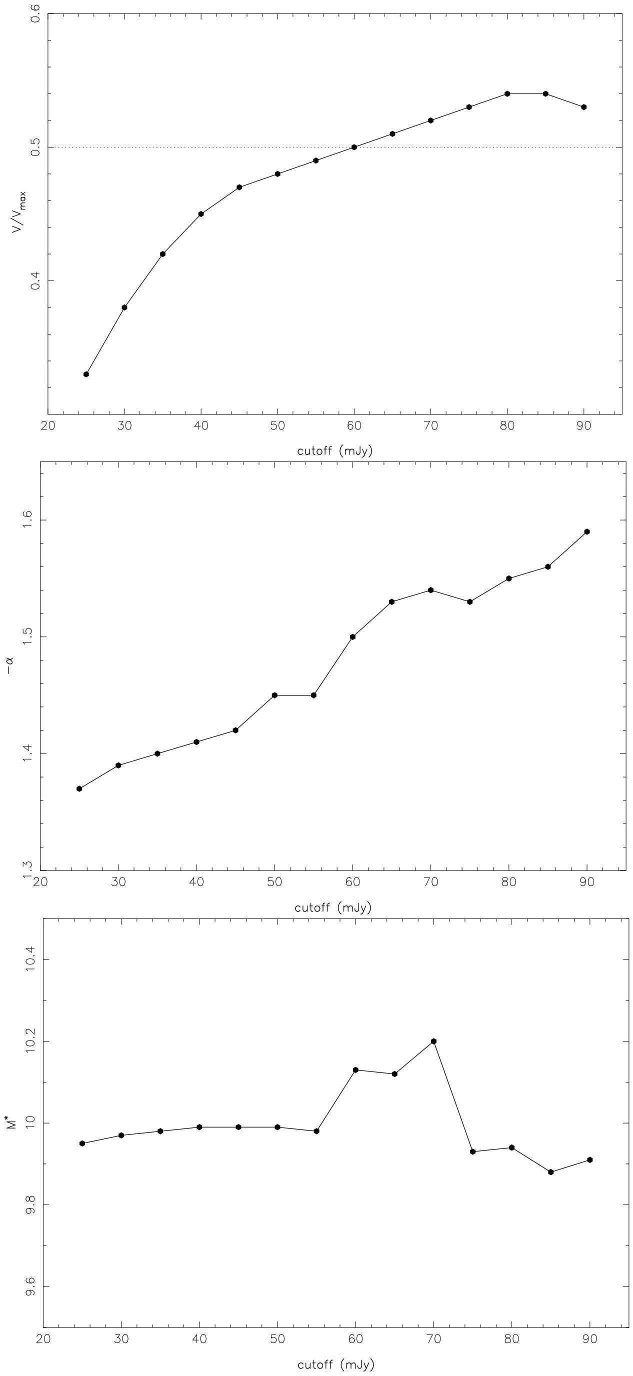

Figure 5.9: [Top]: Change of$\langle V/V_{max}\rangle$ with sample cut-off. [Centre]: Change of faint end slope (α) with cut-off. [Bottom]: Change of $M_{\rm HI}{}^*$ with cut-off.

These plots are consistent with the chosen cut-off of an average flux of 58 mJy, and $\langle V/V_{max}\rangle$

is 0.5 at a cut–off of 60 mJy. The faint end slope, α continually steepens with the raising of the cut-off level. The value of log $M_{\mathrm{HI}}{}^*$ is fairly constant apart from a jump in its value at a cut-off limit of 60-70 mJy.

We can test the cut-off value that has been chosen for the HIMF by adding up the H I mass density from HIMFs derived with different cut–off values. The total mass density will be underestimated for an incomplete sample, and once the completeness level is reached, the total mass density should remain constant, until low number statistics begin to dominate.

Figure 5.10 shows a plot of total H I mass density versus flux cut-off level. It can be seen that the total mass density starts to even out at a flux cut-off of $\gtrsim$ 50 mJy. Thus the chosen flux limit of 58 mJy is consistent as a completeness limit.

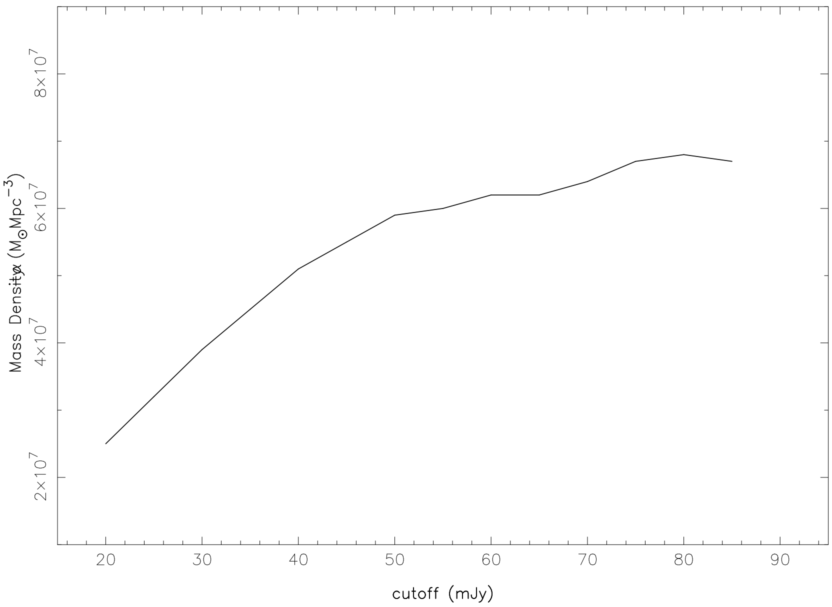

Figure 5.10: Total H I mass density from HIMFs with varying flux cut–off limits. The total H I mass density seems to even out near a flux of 50 mJy, indicating a completeness limit.

5.4.3 Other selection parameters

Completion limits were also derived for peak and integrated flux, so the HIMF can now be calculated using these limits to determine whether all analyses give similar results. Figure 5.11 shows the calculated mass functions. The maximum observable distance of galaxies in the peak flux sample was calculated using Equation 4.4, and using Equation 4.10 for the integrated flux cut–off sample. There were 263 galaxies in the peak flux cut–off sample, and 249 galaxies in the integrated flux sample. The HIMFs derived from these two samples are consistent with the

HIMF derived earlier. The integrated flux limit sample has a slightly different mass distribution to the peak flux limit sample, with less low–mass galaxies, and more galaxies at higher masses. Even so, the low–mass slope of both of the HIMFs is consistent with a slope of α=1.52. The high–mass end of both the samples is well fit by a Schechter function with a characteristic H I mass of log $M_{\rm HI}{}^{*}$=10.1 $M_{\odot}$.

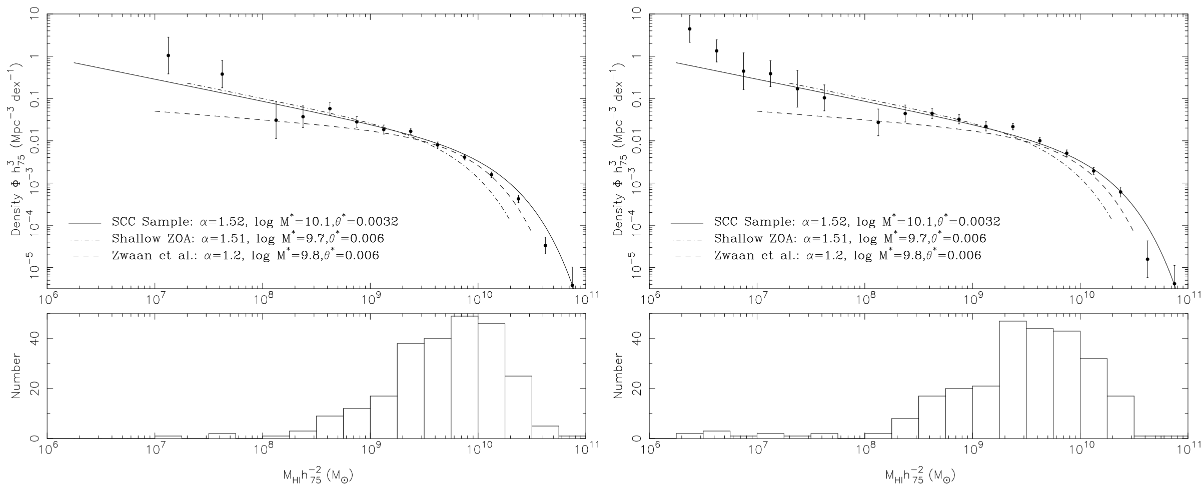

Figure 5.11: The HIMF using an integrated flux limit of 9.2 Jy km s^{-1}, which includes 249 galaxies [Left], and a peak flux limit of 64 mJy, which includes 263 galaxies [Right].

5.4.4 Consistency of Sample Selection and V_{max}

So far the HIMF has been calculated using three different flux limits for the sample. Each of the HIMFs has given similar results, even though the selected samples are slightly different in each case. When determining the value of V_{max} for each galaxy in the selected sample, it was necessary to be consistent between the selection limit and the determined maximum distance for each galaxy in the sample. This would not be a problem if the average and peak flux are equivalent, but as is shown in Figure 5.12, this is not the case.

So for each different flux limit, the limiting observable flux was calculated to reflect the original flux limit imposed on the sample. Table 5.2 shows the equations used to calculate the maximum observable distance in each of the different flux limit scenarios. See Chapter 4 for full details.

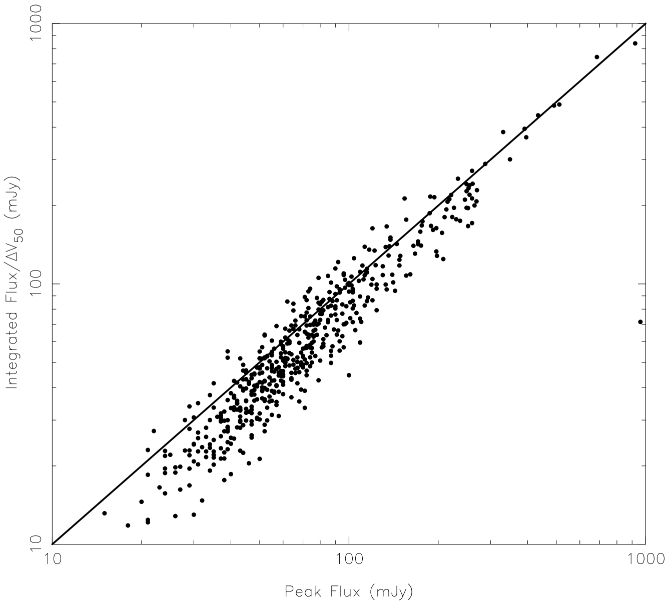

Figure 5.12: Peak flux versus average flux. There is good agreement between the two flux measures for large fluxes, but at small fluxes, the average flux is lower than the peak flux.

Flux cut–off	Maximum Distance
Peak Flux, S_p	$D_{max} = D\sqrt{\frac{S_p}{S_{p,lim}}}$
Average flux, $S_{av}{}^{\dagger}$	$D_{max} = \sqrt{\frac{M_{\mathrm{HI}}}{2.356\times 10^5\, S_{av,lim}\, \Delta V}}$
Integrated Flux, S_{int}	$D_{max} = \sqrt{\frac{M_{\mathrm{HI}}}{2.356\times 10^5\, S_{int,lim}}}$

† For the SCC sample, $\beta = 0$

Table 5.2: Maximum distance determination for galaxies, with differing sample selection paramaters.

5.5 Dependence on Large Scale Structure of the SCC Sample

The shape of the HIMF can be influenced by the presence of clustering within the sample. This is especially important for the lowest mass galaxies in the sample which can only be observed to small distances. If the local universe is overdense, then the numbers of these objects will be overestimated.

A density independent method of calculating the HIMF has been applied to the SCC sample, to determine if the rising low-mass slope is due to an overdensity in the local universe. The method, developed by Saunders et al. (1990), calculates the number of galaxies in distance rings, and then normalises the sample by determining the expected number of galaxies if there was no clustering. Figure 5.13 shows the resulting HIMF from this method. Most of the points agree quite well, but the slope has been lowered slightly in the low-mass end of the function. The existing fit is still consistent with the data, which gives a good indication that the galaxy density in the SCC region is not a major influence on the determination of the HIMF.

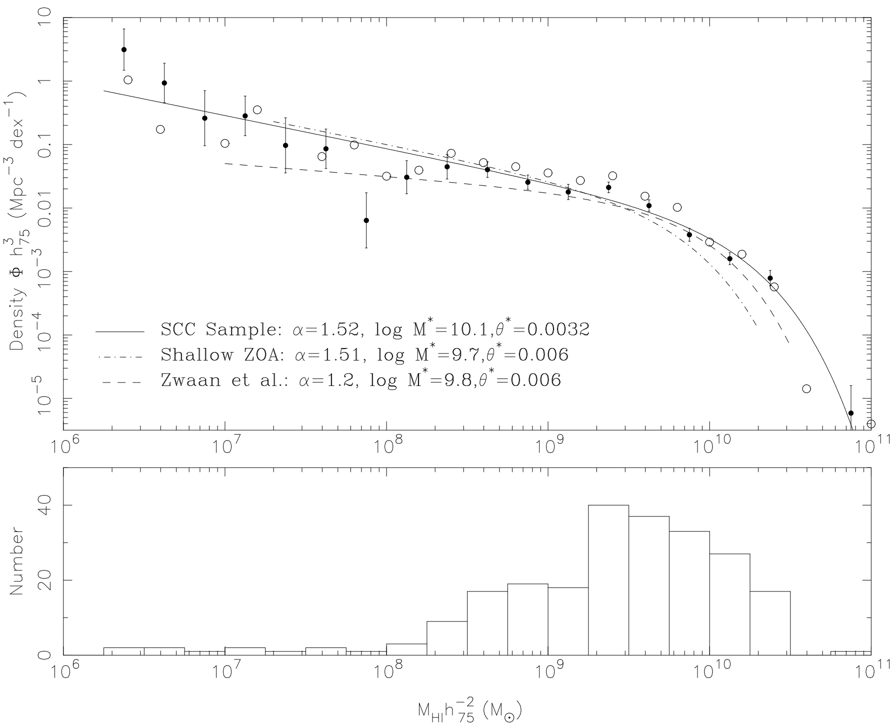

Figure 5.13: The SCC sample HIMF (filled circles), and with a density independent $\sum 1/V_{max}$ method (open circles) which accounts for the clustering of galaxies. The faint–end slope is flattened slightly, but no significant difference between the two methods is seen.

The maximum likelihood determination (Yahil et al., 1977) of the HIMF should be less sensitive to the effects of large scale structure that the $\sum 1/V_{max}$ method. Figure 5.14 shows the maximum likelihood estimation of the HIMF using a parametric fit of a Schechter function. The flux limit used to produce this estimate was $S_{int} > 9.2$ Jy km s^{-1}, consistent with the limits determined previously. The Schechter parameters that are derived from this curve are $\alpha = 1.38$, and log $M_{\rm HI}{}^{*} = 10.03\ M_{\odot}$. The characteristic mass, $M_{\rm HI}{}^{*}$, determined from the maximum likelihood method matches well to the $\sum 1/V_{max}$ result. However, the faint–end slope is shallower than that determined using the $\sum 1/V_{max}$ method. This is consistent with the simulation results, which found a tendancy for an output slope that was less than the input. In the case of the SCC data, this shallower slope may be due to low number statistics. There were only three galaxies with $M_{HI} < 10^{8}\,M_{\odot}$, that also had $S_{int} > 9.2$ Jy km s^{-1} (see the histogram in Figure 5.11). Due to the large error coutours and taking into account the simulation results it is unclear whether the shallower faint end slope estimated using the maximum likelihood method is evidence of clustering in the SCC sample.

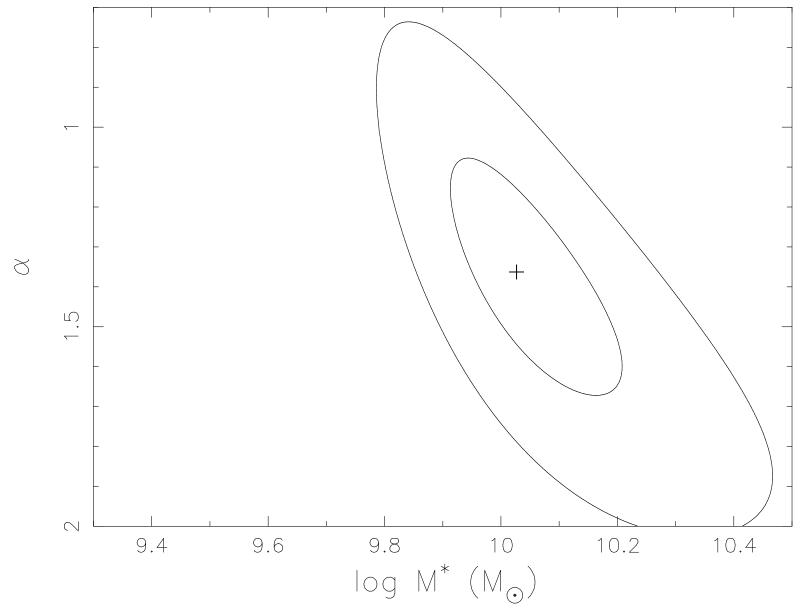

Figure 5.14: Maximum Likelihood HIMF for SCC sample. The cross shows the best fit Schechter parameters, which are α=1.38 and log $M_{\rm HI}{}^{*}$=10.03 $M_{\odot}$. The contours show the 99% and 95% probability curves.

5.6 Exclusion of Possible High Velocity Clouds

There are a number of possible HVCs in the SCC sample, as mentioned in Chapter 2. The distance to these detections is uncertain, and it is possible they are associated with the Magellanic Stream. The general properties of these clouds are discussed further in Chapter 6. The HIMF is now calculated excluding these clouds. Figure 5.15 shows the $\sum 1/V_{max}$ method HIMF excluding the clouds. The number of low-mass detections is almost halved without including these objects. However, the slope of the HIMF is unchanged at the low mass end, although the error bars have increased in size.

If this population of low-mass, optically unidentified H I clouds are in fact extragalactic, then their contribution to the H I mass density in the local universe is minimal. Chapter 6 discusses one of these H I clouds in more detail, and the population as a whole is also investigated.

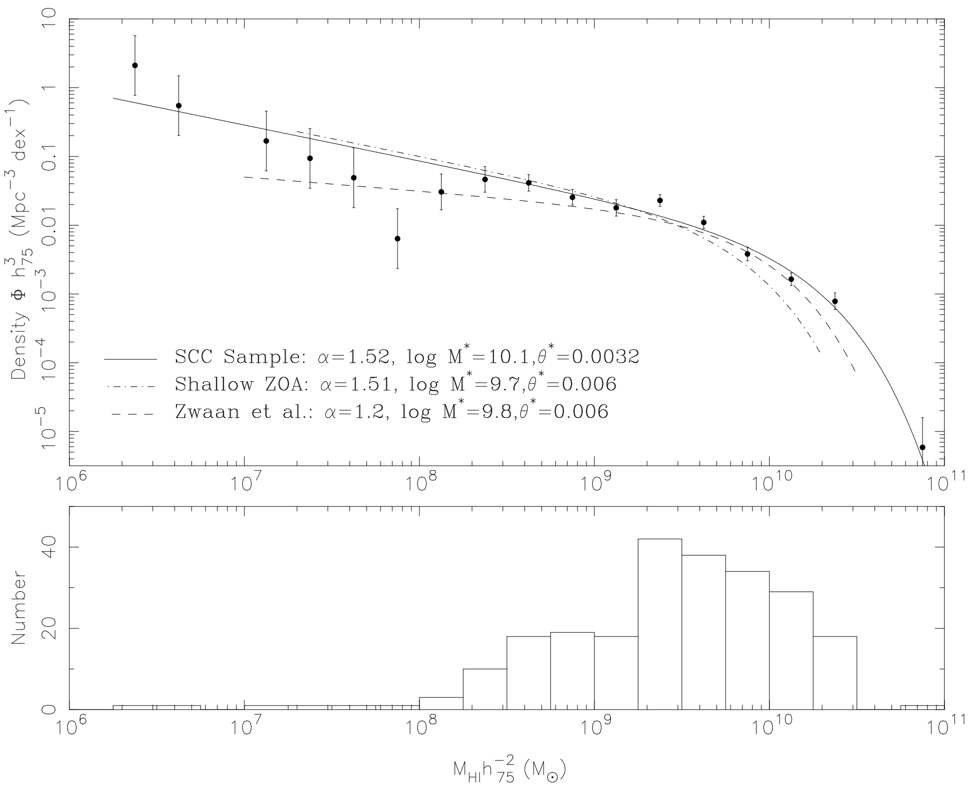

Figure 5.15: HIMF without possible HVCs. The same Schechter parameters as previously determined still fit this data well, although the number of low–mass detections has halved.

5.7 The HIMF over Hubble Type

The SCC sample was classified into optical type as detailed in Chapter 2. To study the HIMF for galaxies of different optical type, the galaxies were divided into 3 groups: spirals, late–types and early–types. HIMFs for galaxies of different optical type are shown in Figure 5.16. The H I mass distribution of spiral galaxies could be fit by a Gaussian distribution, with the parameters log ${M_{\rm HI}}^* = 8.95\ M_\odot$, log σ=0.74. The late-types follow more closely to a power-law distribution, of slope 2.15. There is some deviation from this power–law for late–type galaxies of $M_{\rm HI} \sim 10^9\ M_\odot$. This may be due to incorrect classification of galaxies due to the poor resolution of the DSS I images used. The early–type galaxies also show a power-law distribution, although due to small number statistics, no attempt was made to fit a functional form.

The functional forms fit to the galaxies of different Hubble type are similar to those found by Marquarding (2000). Table 5.3 shows the values of the functional forms that could be fit to spirals and late-type galaxies.

The dip that was seen in all the $\sum 1/V_{max}$ mass functions at an H I mass of $M_{\rm HI} \sim 10^8$ may be caused by the H I distributions of the different optical types.

Author	Optical Type	Fit Type	Parameters
This work	Spiral	Gaussian	log ${M_{\rm HI}}^* = 8.95$ log $\sigma = 0.74$
	Late–Type	Power–Law	$\alpha = 2.15$
Marquarding (2000)	Spiral	Gaussian	log ${M_{\rm HI}}^* = 9.05$ log $\sigma = 0.45$
	Late–Type	Power–Law	$\alpha = 2.25$

Table 5.3: Functional fits of galaxies with different Hubble type.

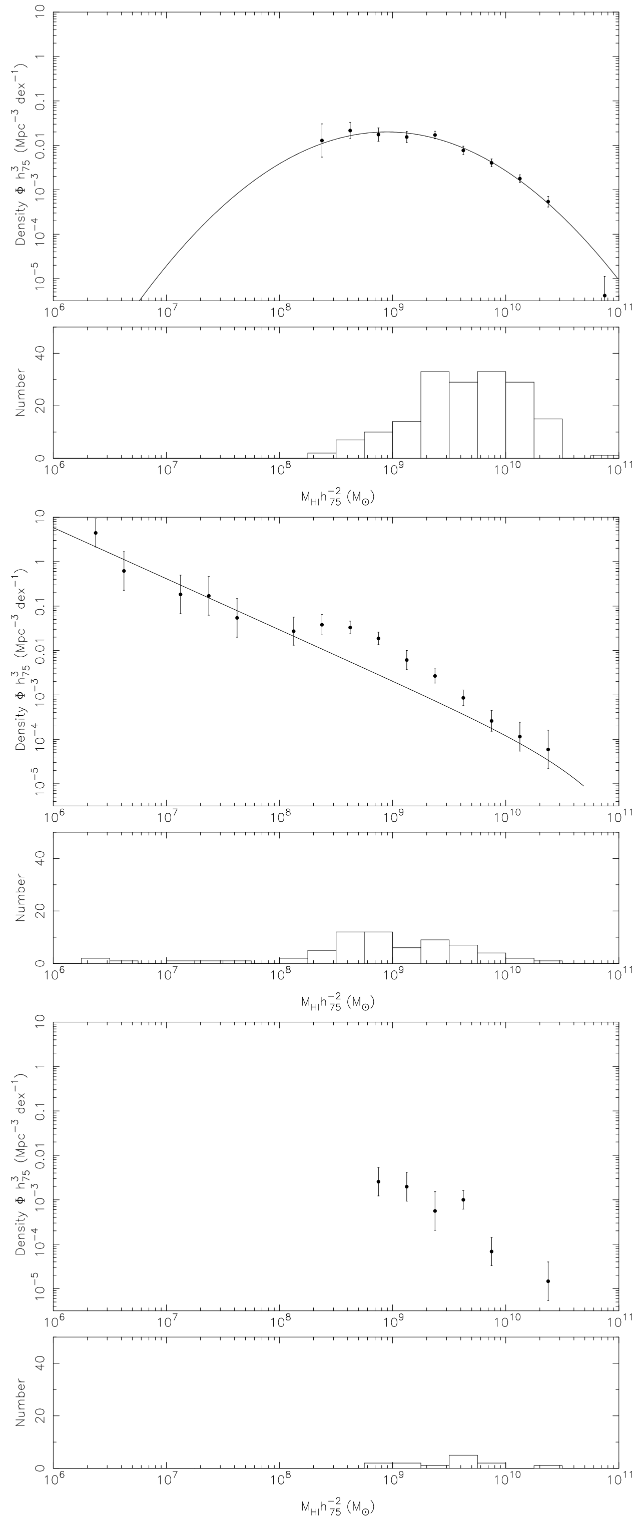

Figure 5.16: HIMF for different Hubble types. Panel(a) [Top]: HIMF for spirals. A Gaussian of log $M_{\rm HI}{}^{*} = 8.95$ and log $\sigma = 0.74$ was fit bye-eye to the data. Panel(b)[Centre]: Late–Type galaxies. A power–law of slope 2.15 is plotted. Panel(c)[Bottom]: Early–Type galaxies. No functional fit is made due to low number statistics.

5.8 The H I content of the local universe

The total H I mass density of the local universe can be determined by integrating under the best fit Schechter function. The integral under the Schechter function from Figure 4 is given by $\rho_{HI} = \Gamma(2-\alpha)\theta^* M_{\rm HI}{}^* = 6.7 \times 10^7 h_{75} M_\odot$ Mpc^{-3}. The H I mass density from adding the data points in the H I density histogram gives the observed total H I mass density, $\sum M_{HI}/V_{max}$. This is a slightly lower value for the total H I mass density of $\rho_{HI} = 6.5 \times 10^7 h_{75} M_\odot$ Mpc^{-3}, as the contribution of galaxies with $M_{HI} < 10^6$ or $M_{\rm HI} > 10^{10.5}$ $M_\odot$ is not included. Table 5.9 shows a comparison of the values derived in the SCC region and from previous surveys. The H I mass density derived in the SCC region is slightly higher than previous calculations, including that of the ZOA shallow survey (Henning et al., 2000). Rao and Briggs (1993) determine ρ_{HI} = 3.6 $\times 10^7 h_{75} M_\odot$ Mpc^{-3}, from H I observations based on optical catalogues. This is about half the H I mass density derived from the SCC sample.

Figure 5.17 shows the H I mass density of galaxies of different mass. Galaxies with an $M_{\rm HI}$ $\sim 10^6 - 10^8$ $M_\odot$ contribute only 6% of the total H I mass density from the calculated HIMF. Most of the H I mass is found in galaxies with $M_{\rm HI}$ from $10^9 - 10^{10.25}$ $M_\odot$.

The critical density of the universe is the density at which the expansion of the universe would eventually be halted. It is defined as (Padmanabhan, 1999):

$$\rho_c = \frac{3H_0^2}{8\pi G} \cong 1.57 \times 10^{-29}\, h_{75}^2\, \mathrm{g\,cm^{-3}} \tag{5.3}$$

The cosmological H I mass density, Ω_{HI} is defined as ρ_{HI}/ρ_c. For this sample, $\Omega_{HI} = 4.1 \pm 0.4 \times 10^{-4} h_{75}^{-1}$. Assuming that the percentage of He is 25% of the total gas density, $\Omega_{gas} = 5.1 \pm 0.5 \times 10^{-4} h_{75}^{-1}$. These values are slightly higher than previously determined values from other blind H I surveys (see Table 5.9).

5.9 Discussion

The HIMFs derived from the SCC sample are similar to previous results, although a slightly steeper value for the Schechter parameter α is determined. The faint–end slope of the HIMF does not indicate the presence of a large, previously missed population, of gas–rich dwarf galaxies, and the mass contribution of galaxies with $M_{\rm HI}\sim 10^6 - 10^8$ $M_\odot$ is just 6% of the total H I mass observed in the sample. The dependence on clustering in the SCC sample is not large, and two density–independent methods gave similar results for the HIMF as the general $\sum 1/V_{max}$ method did.

The value for the characteristic mass, $M_{\rm HI}{}^*$, found for the SCC sample is much higher than any previous blind H I survey. The SCC region contains more galaxies than other H I surveys in the highest mass bins. It was found that the contribution of galaxies with $M_{\rm HI} \sim 10^{10.25}$ is 15% of the total H I mass density. Previous surveys have not had sufficient volume to get good number statistics for galaxies of this H I mass, and it is possible that the Schechter function fits have dropped off too steeply at the high–mass end in previous surveys.

The dependence on velocity width, ΔV for the detectability of galaxies in the SCC sample was found to be different to previous determinations. The reason for this is unclear, especially as the ZOA survey used the same Multibeam receiver as for HIPASS. The detectability of galaxies in HIPASS will need to be characterised further for the full survey, using automated galaxy finding lists as a basis.

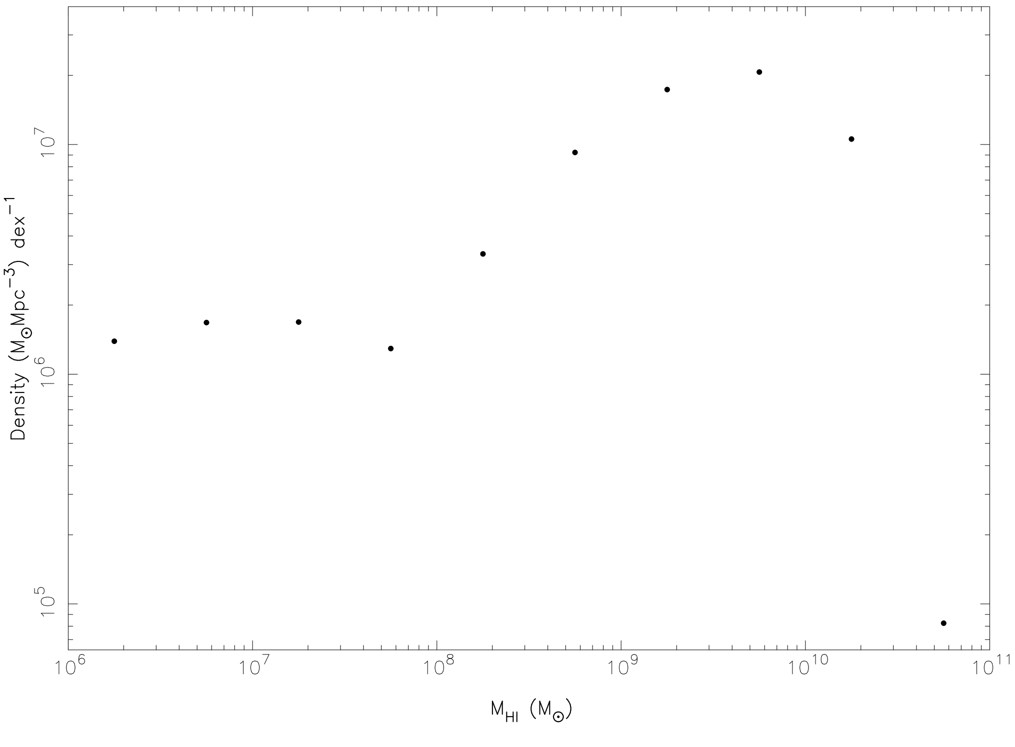

Figure 5.17: Space density of galaxies of different H I masses. Galaxies with $M_{\rm HI} \sim 10^6 - 10^8$ $M_\odot$ make up only 6% of the total H I mass density. Most of the H I mass is contained in galaxies with $M_{\rm HI} \sim 10^9 - 10^{10.25}$ $M_\odot$.

There are a few classes of H I–rich objects that may not be included in this sample, and hence will not be represented in the HIMF. Low column density galaxies ($N_{HI} \sim 10^{18}$ cm^{-2}) would be missed due to the sensitivity of HIPASS. There is evidence of decreasing H I column density with decreasing optical luminosity (de Blok et al., 1996), so it is possible that the galaxies hardest to detect in the optical due to their low surface brightness, may also be hard to detect in H I due to their low H I surface density. In addition, galaxies with a spatial extent much larger than the Parkes beam may be missed due to data reduction techniques.

'Normal' high velocity clouds have not been included in this sample. However, if it is determined that a subset of these clouds are actually placed throughout the Local Group (eg: Blitz et al. (1999), Braun and Burton (1999)), then the H I mass of these clouds will need to be included in any HIMF. Individually, the H I mass of these clouds is most likely to be small, however as a group they will make a larger contribution to the total H I mass in the local universe.

Author	Schechter Parameters			Hi Mass Density		Cosmological Density	
	α	$M_{\mathrm{HI}}{}^{*}$ ($M_{\odot}$)	θ^{*}	$\rho_{HI,tot}$ ($h_{75} M_{\odot}$ Mpc^{-3})	$\rho_{HI,obs}$ ($h_{75} M_{\odot}$ Mpc^{-3})	Ω_{HI} (h_{75}^{-1})	Ω_{gas} (h_{75}^{-1})
This Work	1.52	1.14×10^{10}	3.2×10^{-3}	6.7×10^{7}	6.0×10^{7}	$4.1\pm0.4 \times 10^{-4}$	$5.1\pm0.4 \times 10^{-4}$
Henning et al. (2000)	1.51	5.0×10^{9}	6×10^{-3}	5.4×10^{7}	$\cdots$	3.5×10^{-4}	4.4×10^{-4}
Zwaan et al. (1997)	1.2	6.3×10^{9}	6×10^{-3}	4.4×10^{7}	4.05×10^{7}	2.8×10^{-4}	3.6×10^{-4}

Table 5.4: Comparison of H I Parameters from this work and previous blind H I surveys.

Chapter 6

HIPASS J1712-64

6.1 Intergalactic H I Clouds

Primordial concentrations of hydrogen are thought to have collapsed from density fluctuations in the early universe to form galaxies. It is possible that some of these concentrations have not yet started to form stars, and exist in the local universe as clouds of mostly neutral hydrogen. Studies of LSB galaxies have demonstrated the complexity of star formation in different environments, with these galaxies in particular having a large reservoir of gas, yet little star formation (eg: Bell et al. (1999); van der Hulst et al. (1993)). If there is some threshold gas density above which star formation can occur, it is conceivable there exists a population of hydrogen gas clouds that have not yet reached this minimum density threshold.

Limits to the presence of such a population of H I-rich clouds or 'protogalaxies' have been investigated by a number of blind H I surveys conducted in a range of environments, eg: clusters (Barnes et al., 1997; Dickey, 1997; Weinberg et al., 1991; Materne et al., 1979), groups (Kraan-Korteweg et al., 1999; Banks et al., 1999; Fisher and Tully, 1981b; Lo and Sargent, 1979), voids (Krumm and Brosch, 1984; Hoffman et al., 1992), and in the general field (Spitzak and Schneider, 1998; Henning, 1995; Sorar, 1994; Shostak, 1977). In addition, there have been numerous observations in off-source calibration regions for galaxies which have been optically-selected. For example, Fisher and Tully (1981a) serendipitously detected a couple of dozen galaxies in the calibration scans from their survey, all with optical counterparts.

No H I detections were confirmed to be intergalactic H I clouds without optical emission from these surveys, although Spitzak and Schneider (1998) did find one H I detection without an optical counterpart lying nearby a bright star which might obscure a low surface brightness counterpart. However, the existence of a population of high mass H I clouds has not been ruled