View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of Queensland eSpace

The Astronomical Journal, 125:2842–2858, 2003 June

© 2003. The American Astronomical Society. All rights reserved. Printed in U.S.A.

THE 1000 BRIGHTEST HIPASS GALAXIES: THE H I MASS FUNCTION AND $\Omega_{\rm H\,I}$

M. A. Zwaan,[1] L. Staveley-Smith,[2] B. S. Koribalski,[2] P. A. Henning,[3] V. A. Kilborn,[4] S. D. Ryder,[5] D. G. Barnes,[1] R. Bhathal,[6] P. J. Boyce,[7] W. J. G. de Blok,[8] M. J. Disney,[8] M. J. Drinkwater,[9] R. D. Ekers,[2] K. C. Freeman,[10] B. K. Gibson,[11] A. J. Green,[12] R. F. Haynes,[2] H. Jerjen,[10] S. Juraszek,[12] M. J. Kesteven,[2] P. M. Knezek,[13] R. C. Kraan-Korteweg,[14] S. Mader,[2] M. Marquarding,[2] M. Meyer,[1] R. F. Minchin,[8] J. R. Mould,[15] J. O'Brien,[10] T. Oosterloo,[16] R. M. Price,[3] M. E. Putman,[17] E. Ryan-Weber,[1,2] E. M. Sadler,[12] A. Schröder,[18] I. M. Stewart,[18] F. Stootman,[6] B. Warren,[10] M. Waugh,[1] R. L. Webster,[1] and A. E. Wright[2]

Received 2002 December 3; accepted 2003 February 19

ABSTRACT

We present a new, accurate measurement of the H I mass function of galaxies from the HIPASS Bright Galaxy Catalog, a sample of 1000 galaxies with the highest H I peak flux densities in the southern ($\delta < 0°$) hemisphere. This sample spans nearly 4 orders of magnitude in H I mass [$\log (M_{\rm H\,I}/M_\odot) + 2 \log h_{75} = 6.8$–$10.6$] and is the largest sample of H I–selected galaxies to date. We develop a bivariate maximum likelihood technique to measure the space density of galaxies and show that this is a robust method, insensitive to the effects of large-scale structure. The resulting H I mass function can be fitted satisfactorily with a Schechter function with faint-end slope $\alpha = -1.30$. This slope is found to be dependent on morphological type, with late-type galaxies giving steeper slopes. We extensively test various effects that potentially bias the determination of the H I mass function, including peculiar motions of galaxies, large-scale structure, selection bias, and inclination effects, and we quantify these biases. The large sample of galaxies enables an accurate measurement of the cosmological mass density of neutral gas: $\Omega_{\rm H\,I} = (3.8 \pm 0.6) \times 10^{-4}\, h_{75}^{-1}$. Low surface brightness galaxies contribute only ~15% to this value, consistent with previous findings.

Key words: galaxies: ISM — galaxies: luminosity function, mass function — ISM: general — radio emission lines — surveys

1. INTRODUCTION

Estimates of the H I mass function (HIMF), the distribution function of galaxies as a function of H I mass, have been based on very small samples of galaxies. The most extensive blind H I surveys to date, the Arecibo H I Strip Survey (AHISS; Zwaan et al. 1997) and the Arecibo Dual-Beam Survey (ADBS; Rosenberg & Schneider 2002), resulted in 66 and 265 galaxy detections, respectively. On the contrary, state-of-the-art measurements of the optical equivalent of the HIMF, the galaxy luminosity function (LF), are based on samples of typically a few tens of thousands of galaxies (the Two-Degree Field: Folkes et al. 1999, Norberg et al. 2002; the Sloan Digital Sky Survey: Blanton et al. 2001). These large-area, low-redshift surveys have resulted in a very accurate census of the local galaxy population and have shown that the power-law faint-end slope of the LF is "flat" ($\alpha = -1.2$) down to the lowest luminosities ($M_B - 5 \log h_{75} < -15$). The shape of the LF and the derived integral cosmological luminosity density ρ_L provide important constraints to galaxy evolution models. However, for an accurate assessment of the distribution and content of baryons in the universe, it is also important to measure the HIMF with high accuracy.

The H I Parkes All-Sky Survey (HIPASS) is a blind survey of the whole southern sky south of $\delta = 25°$ in the velocity range -1200 km s^{-1} $< cz <$ 12,700 km s^{-1}. Analysis of the extragalactic component of the survey is ongoing, and the final galaxy sample is estimated to have ~7000 entries. This deep catalog is expected to make important contributions to the mapping and understanding of large-scale structure, the interpretation of QSO absorption lines, analysis of galaxy groups and clusters, etc. Here we use the first product of the HIPASS database, the HIPASS Bright Galaxy Catalog (BGC; Koribalski et al. 2002), which consists of the 1000 galaxies with the highest peak flux densities. This sample is 4 times as large as the next-largest H I–selected galaxy sample (ADBS) and therefore enables a

[1] School of Physics, University of Melbourne, Melbourne, VIC 3010, Australia; mzwaan@ph.unimelb.edu.au.
[2] Australia Telescope National Facility, CSIRO, P.O. Box 76, Epping, NSW 1710, Australia.
[3] Institute for Astrophysics, University of New Mexico, 800 Yale Boulevard, NE, Albuquerque, NM 87131.
[4] Jodrell Bank Observatory, University of Manchester, Macclesfield, Cheshire SK11 9DL, UK.
[5] Anglo-Australian Observatory, P.O. Box 296, Epping, NSW 1710, Australia.
[6] School of Engineering and Industrial Design, University of Western Sydney, P.O. Box 555, Campbelltown, NSW 2560, Australia.
[7] Department of Physics, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, UK.
[8] Department of Physics and Astronomy, University of Wales, Cardiff, P.O. Box 913, Cardiff CF2 3YB, UK.
[9] Department of Physics, University of Queensland, Brisbane, QLD 4072, Australia.
[10] Research School of Astronomy and Astrophysics, Mount Stromlo Observatory, Cotter Road, Weston, ACT 2611, Australia.
[11] Centre for Astrophysics and Supercomputing, Swinburne University of Technology, P.O. Box 218, Hawthorn, VIC 3122, Australia.
[12] School of Physics, University of Sydney, NSW 2006, Australia.
[13] WIYN, Inc., 950 North Cherry Avenue, Tucson, AZ 85726.
[14] Departamento de Astronomía, Universidad de Guanajuato, Apdo. Postal 144, 36000 Guanajuato, GTO, Mexico.
[15] National Optical Astronomy Observatory, P.O. Box 26732, 950 North Cherry Avenue, Tucson, AZ 85726.
[16] ASTRON, Postbus 2, NL-7990 AA Dwingeloo, Netherlands.
[17] Center for Astrophysics and Space Astronomy, University of Colorado at Boulder, 389 UCB, Boulder, CO 80309.
[18] Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK.

more careful analysis of the shape of the HIMF, as well as its dependence on various galaxy parameters.

An accurate measurement of the HIMF is highly relevant to several modern astrophysical problems. First, the HIMF provides a measurement of the cosmological mass density of H I, $\Omega_{\rm H\,I}$, in the local universe. This is an important benchmark in the mapping out of the evolution of the neutral gas mass density from before the epoch of reionization, when the vast majority of baryons were in the form of H I, to the present epoch, where the mass in stars outweighs that in neutral gas (Lanzetta, Wolfe, & Turnshek 1995; Storrie-Lombardi & Wolfe 2000). At intermediate redshifts, damped Lyα systems are used to trace the evolution of $\Omega_{\rm H\,I}$, but because of their low number density dN/dz at low z, damped Lyα cannot be used to accurately measure $\Omega_{\rm H\,I}$ at $z = 0$.

Second, the shape of the HIMF faint-end slope provides useful input to galaxy formation models. Both the LF and the HIMF are required in order to link the local galaxy population to the model prediction, since both functions measure different baryonic components of galaxies. The HIMF measures the distribution of the *mass* of the innate cool gas in galaxies, whereas the LF describes the distribution of the light emission of processed material (i.e., stars), which is nontrivially linked to its mass. Many dwarf galaxies have large gas fractions (e.g., Roberts & Haynes 1994), which means that the HIMF is closely related to the low-mass end of the total mass function of galaxies. Of course, this relation breaks down in regions of high galaxy density, where gas-poor dwarf elliptical galaxies dominate the number counts of low-mass galaxies. Measuring the low-mass end of the mass function is specifically interesting with regard to cold dark matter (CDM) theory, which predicts an abundance of low-mass objects, which might be detectable at 21 cm. However, the detectability could be decreased if gas is ejected by early supernovae (Dekel & Silk 1986; Babul & Rees 1992; Babul & Ferguson 1996) or photoevaporated by the cosmic UV background during reionization (Barkana & Loeb 1999). Also, at lower redshifts can ionization by the UV background of lower column density regions of H I disks have an effect on the slope of the HIMF (Corbelli, Salpeter, & Bandiera 2001).

Recently, there has been considerable controversy over the faint-end slope of the HIMF. Zwaan et al. (1997) found a slope of $\alpha = -1.2$, consistent with previous estimates based on optically selected galaxies (Briggs & Rao 1993). Similar values for α were found by Kraan-Korteweg et al. (1999). Schneider, Spitzak, & Rosenberg (1999) report a flat HIMF with an extremely steep low-mass slope below $M_{\rm H\,I} = 10^8$. In a more recent analysis Rosenberg & Schneider (2002) advocated a faint-end slope of $\alpha = -1.53$.

The aim of this paper is to calculate the HIMF with greater accuracy using the HIPASS BGC, the largest sample of H I–selected galaxies available to date. Particular emphasis is directed toward understanding possible biases in the calculation of the HIMF. In § 2 we briefly summarize the HIPASS specifics and describe the BGC. In § 3 various estimators of the HIMF are discussed and a new method, the two-dimensional stepwise maximum likelihood (SWML) method, is described. The H I mass function is presented in § 4, which includes a detailed discussion of possible biases that may influence the HIMF calculation. In § 5 the results are compared with previous measurements of the HIMF. The contribution of different galaxy types is discussed in § 6. In § 7 we present a discussion of the H I mass density $\rho_{\rm H\,I}$, and in § 8 the selection function of the survey is discussed. Finally, in § 9 the conclusions are presented. We use $H_0 = 75$ km s^{-1} Mpc^{-1} throughout this paper.

2. THE HIPASS BRIGHT GALAXY CATALOG

The BGC (Koribalski et al. 2002) is a catalog of the 1000 galaxies with the largest H I peak flux densities,[19] S_p, in the southern sky ($\delta < 0°$). The sample is based on HIPASS, for which the observing strategy and reduction details are described in Barnes et al. (2001). Here we briefly summarize the HIPASS survey strategy.

The observations were conducted in the period from 1997 to 2000 with the Parkes 64 m radio telescope,[20] using the 21 cm multibeam receiver (Staveley-Smith et al. 1996). The telescope scanned along strips of 8° in declination, and data were recorded for 13 independent beams, each with two polarizations. A total of 1024 channels over a total bandwidth of 64 MHz were recorded, resulting in a channel separation of $\Delta v = 13.2$ km s^{-1} and a velocity resolution of $\delta v = 18$ km s^{-1} after Tukey smoothing. The total velocity coverage is -1200 to 12,700 km s^{-1}. After bandpass calibration, continuum subtraction, and gridding into $8° \times 8°$ cubes, the typical rms noise is 13 mJy beam^{-1}. This leads to a 3 σ column density limit of $\approx 6 \times 10^{18}$ cm^{-2} for gas filling the beam. The spatial resolution of the gridded data is 15′.5.

An automatic galaxy-finding algorithm (see Kilborn et al. 2002) was applied to the HIPASS data set to identify all sources with $S_p > 60$ mJy. The list of potential detections was inspected by eye to separate radio-frequency interference and bandpass ripples from real H I sources. Since the noise in the HIPASS data is considerably higher at low Galactic latitude, the list of detections was complemented with $|b| < 3°$ detections from the H I zone of avoidance survey (Henning et al. 2000). Furthermore, to avoid confusion with the Milky Way and high-velocity clouds, the range $|v| < 350$ km s^{-1} was excluded from the list and substituted with known nearby galaxies. From the resulting list, the 1000 galaxies with the highest H I peak fluxes were selected, resulting in a selection limit of $S_p > 116$ mJy. Further details of the selection, as well as properties of the BGC galaxies, are given in Koribalski et al. (2002). An important point to note here is that only four BGC sources outside the zone of avoidance ($b > 5°$) have no optical counterpart on the Digitized Sky Survey. One is obscured by the LMC, one is a tidal cloud of NGC 2442 (Ryder et al. 2001), and the other two have heliocentric velocities of less than 500 km s^{-1} and are probably Magellanic debris (Koribalski et al. 2002).

HIPASS achieves 100% coverage over the southern sky and has a very uniform noise level. However, there are regions of the sky near strong sources of radio continuum in which the noise level is elevated. Near such sources, it is possible that we are missing bright galaxies. As above, the typical rms noise level is 13 mJy beam^{-1}. However, above an rms noise level of 23 mJy beam^{-1}, the faintest BGC objects are detected with a signal-to-noise ratio (S/N) of no more

[19] In the remainder of this paper we refer to "peak flux density" as "peak flux."

[20] The Parkes telescope is part of the Australia Telescope, which is funded by the Commonwealth of Australia for operation as a National Facility managed by the CSIRO.

than 5 in each velocity channel. Within the BGC velocity range, 2.5% of all southern HIPASS spectra have an rms noise greater than 23 mJy beam^{-1}. This gives a useful upper limit to the incompleteness of the BGC.

3. CALCULATING THE H I MASS FUNCTION

3.1. *Methods*

We define the H I mass function $\theta(M_{\rm H\,I})$ as the space density of objects in units of h_{75}^3 Mpc^{-3}. The HIMF is normally calculated per decade of H I mass and plotted on a logarithmic scale. For fitting purposes, we use the Schechter function (Schechter 1976) defined by

$$\theta(M_{\rm H\,I})dM_{\rm H\,I} = \theta^* \left(\frac{M_{\rm H\,I}}{M^*_{\rm H\,I}}\right)^{\alpha} \exp\left(-\frac{M_{\rm H\,I}}{M^*_{\rm H\,I}}\right) dM_{\rm H\,I} \ , \quad (1)$$

characterized by the three parameters α, $M^*_{\rm H\,I}$, and θ^*, which define the slope of the power law, the H I mass corresponding to the "knee," and the normalization, respectively.

The $1/V_{\rm max}$ method, originally developed by Schmidt (1968) to study the evolution of QSOs, simply consists of summing in H I mass bins the reciprocals of $V_{\rm max}$, the volume corresponding to the maximum distance $D_{{\rm lim},i}$ at which an object can be detected. This is a popular method for determining H I mass functions because the values of $V_{\rm max}$ can be readily evaluated for every survey, if the survey sensitivity is well understood (see Zwaan 1997; Rosenberg & Schneider 2002; Kilborn 2000). The method also works well for surveys that are not purely flux limited but where the survey selection is a complicated function of various galaxy parameters and telescope properties. For the BGC, $D_{{\rm lim},i}$ can simply be found by multiplying the distance D_i at which the object is detected by $(S_i/S_{\rm lim})^{1/2}$, where S_i and $S_{\rm lim}$ are, respectively, the peak flux of object i and the limiting peak flux of the sample, which is 116 mJy.

The main concern about the $V_{\rm max}$ method is that it is potentially sensitive to the influence of large-scale structure. The $V_{\rm max}$ method intrinsically assumes that the galaxy population used to evaluate luminosity functions or mass functions is homogeneously distributed in space. In sensitivity-limited samples (as opposed to volume-limited samples), a correlation exists between the mass or luminosity of objects and the distance at which they are preferentially found. Underdense or overdense regions may therefore yield under- or overrepresentations of objects of corresponding mass or luminosity. Maximum likelihood techniques (Sandage, Tamman, & Yahil 1979, and the SWML method, Efstathiou, Ellis, & Peterson 1988) are designed to be insensitive to density fluctuations, and it is therefore important to test whether these methods can be employed for our H I–selected galaxy sample.

The key to the maximum likelihood techniques is to find the parent distribution θ that yields the maximum joint probability of detecting all objects in the sample. The probability that a galaxy with H I mass $M_{{\rm H\,I},i}$ will be detected can be expressed as

$$p(M_{{\rm H\,I},i}|D_i) = \frac{\theta(M_{{\rm H\,I},i})}{\int_{M_{{\rm H\,I,lim}(D_i)}}^{\infty} \theta(M_{\rm H\,I})dM_{\rm H\,I}} \ , \quad (2)$$

where $M_{{\rm H\,I,lim}(D_i)}$ is the minimal detectable H I mass at distance D_i in megaparsecs. Put differently, $p(M_{{\rm H\,I},i}|D_i)$ is the fraction of galaxies in the survey volume with H I mass $M_{{\rm H\,I},i}$ that are near enough to be brighter than the survey detection limit. Finding for which $\theta(M_{\rm H\,I})$ the product of the probabilities,

$$\mathscr{L} = \prod_{i=1}^{N_{\rm g}} p_i \ , \quad (3)$$

is maximal then gives the maximum likelihood solution for the mass function. The disadvantage of this procedure is that it is parametric; that is, it requires an analytical expression for θ, which is usually taken to be the Schechter function, as in equation (1). The SWML is a modification of this procedure and measures θ at fixed intervals of $\log M_{\rm H\,I}$ by iteration. This procedure does not depend on a functional form for θ.

As described above, the maximum likelihood estimators require the calculation of $M_{{\rm H\,I,lim}}(D_i)$, the minimal detectable H I mass at distance D_i. For a hypothetical sample limited by integrated H I flux, this parameter could be simply calculated by $M_{{\rm H\,I,lim}}(D) = 2.36 \times 10^5 D^2 \int S\,dV_{\rm lim}$, where $\int S\,dV_{\rm lim}$ is the integrated flux limit in Jy km s^{-1} and D is the distance in megaparsecs. For optical redshift surveys, the analogous value of minimal detectable luminosity, $L_{\rm lim}(D)$, can be determined because optical redshift surveys are generally flux limited, and maximum likelihood methods can be readily applied. H I–selected samples are seldom integrated-flux limited, but instead the detectability of signals depends on both the velocity width ΔV and peak flux S_p (Zwaan et al. 1997; Rosenberg & Schneider 2000) or, in the case of the BGC, solely on S_p. Therefore, a unique relation between distance and the minimal detectable H I mass does not exist, and the standard maximum likelihood techniques cannot be employed.

In principle, for every galaxy in the sample a parameter $M_{{\rm H\,I,lim}}(D_i,\ P_i)$ can be calculated, which is the minimal detectable H I mass at distance D_i, with profile shape P_i equal to that of galaxy i. Rosenberg & Schneider (2002) applied this method to their ADBS sample and found an HIMF with faint-end slope $\alpha = -1.53$. However, this method can lead to incorrect measurements of the HIMF. The SWML essentially consists of determining for each bin k in $\log M_{\rm H\,I}$ the volume accessible to that H I mass by summing the reciprocals of space densities of galaxies that could be detected in the volume, out to the maximum distance at which a galaxy in bin k could be detected. A summation over all galaxies with $M_{\rm H\,I} < M_{{\rm H\,I,lim}}(D_i, P_i)$ is not the same.

In the following we briefly demonstrate that this implementation of the maximum likelihood technique for the BGC can result in a severe overestimation of the faint-end slope α. We fill volumes equal to that of the BGC with synthetic galaxy samples using HIMFs with slopes $\alpha = -1.0$ and $\alpha = -1.4$. The galaxies follow a general trend of $\Delta V \propto M_{\rm H\,I}^{1/3}$, similar to what is observed in the BGC, and we introduce scatter on all galaxy properties equal to that seen in the BGC. Next, we select from these volumes all galaxies with $S_p > 116$ mJy and calculate the HIMF following the SWML method as described above. From these synthetic peak-flux–selected samples, we also select integrated-flux–limited subsamples, for which $S_{\rm int} > 25$ Jy km s^{-1}, and also apply the SWML method to these samples. We test by means of a $V/V_{\rm max}$ test that the cutoff at $S_{\rm int} > 25$ Jy km s^{-1}

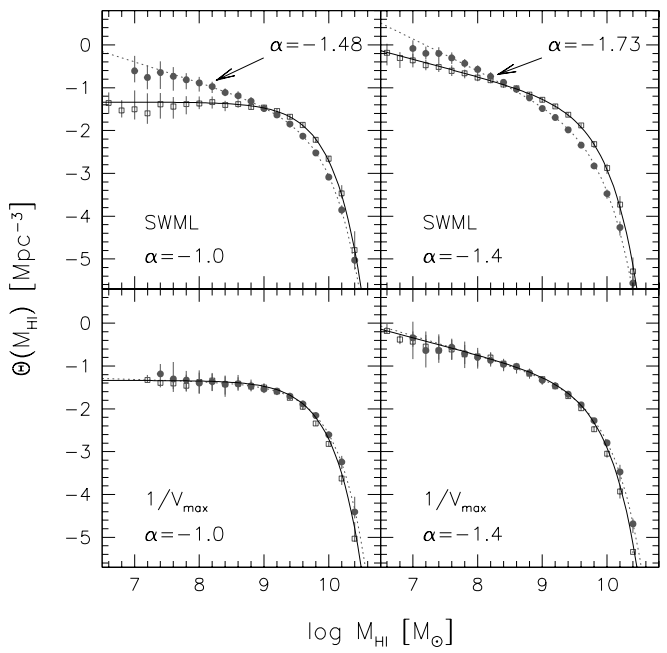

FIG. 1.—Testing the conventional SWML method on peak-flux–limited samples. Each panel shows the input Schechter function that is used to construct synthetic galaxy samples (*solid line*), the output H I mass function of the total sample (*circles*), and the integrated-flux–limited subsample (*squares*). In the top two panels the SWML method is applied, in the bottom two the $1/V_{\max}$ method is used. The SWML application on a peak-flux–limited sample clearly gives incorrect results.

provides a statistically complete subset. Finally, the $1/V_{\max}$ method is applied to both the peak-flux–limited samples and the integrated-flux–limited subsamples. All simulations are performed 100 times.

The results are presented in Figure 1. The top two panels represent the results from the SWML method, and the bottom two panels those from the $1/V_{\max}$ method. In each panel the solid line shows the input Schechter function, the circles show the measured HIMFs for the peak-flux–limited samples, and the squares show the measured HIMF for the integrated-flux–limited subsamples. The dotted lines are a Schechter function fitted to the points.

Both methods recover the input HIMF very well if they are applied to the integrated-flux–limited subsamples. However, this figure illustrates clearly that the SWML method results in a severe overestimation of the faint-end slope if it is applied to the peak-flux–limited samples.

We note that the examples presented here are specific to the BGC. To what degree the output HIMF is in error depends very strongly on the selection function of the H I survey, and there are situations possible in which the SWML coincidentally gives the correct answer.

3.2. *The Two-dimensional SWML Method (2DSWML)*

On galaxy samples that are other than flux limited, other parameters have to be included in the maximum likelihood calculation of the space density of objects. For a given H I source at distance D, the peak flux is directly proportional to its H I mass $M_{\rm H\,I}$ and inversely proportional to the velocity width over which the flux is distributed. Therefore, the velocity width is the second parameter we have to include in the maximum likelihood analysis. A two-dimensional maximum likelihood algorithm can then be employed to find the true galaxy distribution. In the following we use w_{20}, the profile width measured at 20% of peak intensity, as a measurement of the velocity width.

The probability of detecting galaxy i with H I mass $M^i_{\rm H\,I}$ and velocity width w^i_{20} at distance D_i is given by

$$p(M^i_{\rm H\,I}, w^i_{20}|D_i) = \frac{\theta(M^i_{\rm H\,I}, w^i_{20})}{\int_{w_{20}=0}^{\infty}\int_{M_{\rm H\,I}=M^i_{\rm H\,I,lim}}^{\infty}\theta(M_{\rm H\,I}, w_{20})\,dM_{\rm H\,I}\,dw_{20}}\ , \quad (4)$$

where

$$M^i_{\rm H\,I,lim}(w_{20}) = M^i_{\rm H\,I}\frac{S_{\rm lim}}{S_i}\frac{w_{20}}{w^i_{20}} \quad (5)$$

is the minimal H I mass the galaxy could have to be detectable at distance D_i if it had a velocity width w_{20}, and $\theta(M_{\rm H\,I}, w_{20})$ is the two-dimensional density distribution function. However, we have no a priori knowledge of the functional form of $\theta(M_{\rm H\,I}, w_{20})$, which implies that a parametric maximum likelihood method (Sandage et al. 1979) cannot be applied.

Alternatively, we can define the logarithmically binned two-dimensional density distribution function

$$\theta_{jk} = \theta(M, W)\ , \quad (6)$$

where

$$j = 1,\ \ldots, N_M\ , \quad k = 1,\ \ldots, N_W\ , \quad (7)$$

and N_M and N_W are the number of bins in M and W, respectively, and we define

$$M = \log M_{\rm H\,I}\ , \quad W = \log w_{20}\ . \quad (8)$$

The logarithm of the likelihood of detecting all galaxies in the sample can now be expressed as

$$\ln\mathscr{L} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_M}\sum_{k=1}^{N_W} V(M_i - M_j, W_i - W_k)\ln\theta_{jk} - \sum_{i=1}^{N_g}\ln\left(\sum_{j=1}^{N_M}\sum_{k=1}^{N_W} H_{ijk}\theta_{jk}\right) + c\,, \quad (9)$$

where c is a constant and V is a function defined by

$$V(x,y) = \begin{cases} 1, & \text{if } |x| \le \Delta M/2 \text{ and } |y| \le \Delta W/2\ , \\ 0, & \text{otherwise}\ , \end{cases} \quad (10)$$

which makes the first sum only go over galaxies in bin $\Delta M\Delta W$. The function H_{ijk} is defined as the fraction of the bin accessible to source i:

$$H_{ijk} = \frac{1}{\Delta M\Delta W}\int_{W=W_k-\Delta W/2}^{W^+}\int_{M=M^-}^{M_j+\Delta M/2} dM\,dW\ , \quad (11)$$

where

$$M^- = \max\left[M_j - \Delta M/2,\quad \min(M_j + \Delta M/2, M_{{\rm lim},i})\right]\ ,$$

$$M_{{\rm lim},i} = M_i + \log(S_{\rm lim}/S_i) - W_i + W_k\ ,$$

$$W^+ = \min\left[W_k + \Delta W/2,\quad \max(W_k - \Delta W/2, W_{{\rm lim},i})\right]\ ,$$

$$W_{{\rm lim},i} = W_i - \log(S_{\rm lim}/S_i) - M_i + M_j\ .$$

This implies that

$$H_{ijk} = \begin{cases} 1, & \text{if } M_i - M_j - W_i + W_k \\ & \quad + \log(S_{\rm lim}/S_i) + \Delta W/2 + \Delta M/2 < 0 \ , \\ 0, & \text{if } M_i - M_j - W_i + W_k \\ & \quad + \log(S_{\rm lim}/S_i) - \Delta W/2 - \Delta M/2 > 0 \ . \end{cases}$$

Maximum likelihood solutions for θ_{jk} are now found by differentiating $\mathscr{L}$ and applying the usual additional constraint to fix the normalization (see, e.g., Efstathiou et al. 1988). We then arrive at

$$\theta_{jk} = \sum_{i=1}^{N_g} V(M_i - M_j, W_i - W_k) \Bigg/ \left[\sum_{i=1}^{N_g} H_{ijk}\, \Delta M\, \Delta W \left(\sum_{l=1}^{N_M} \sum_{m=1}^{N_W} \theta_{lm} H_{ilm}\, \Delta M\, \Delta W \right)^{-1} \right] \quad (12)$$

and most-likely values for θ_{jk} are found by iterating equation (12). Stable solutions for θ_{jk} are usually found after $\sim$25 iterations.

Finally, the HIMF can be calculated from

$$\theta(M_j) = \sum_{k=1}^{N_W} \theta(M_j, W_k) \ . \quad (13)$$

In the remainder of this paper, we refer to this method as the two-dimensional stepwise maximum likelihood or 2DSWML method. A very similar technique was used by Loveday (2000) to calculate the *K*-band luminosity function from a *B*-band selected galaxy sample. The implementation of the 2DSWML method described in this paper is designed to work on the peak-flux–limited BGC, but the method could be adjusted to work on samples other than peak-flux limited. This will be of particular interest for the full sensitivity HIPASS galaxy catalog (Meyer et al. 2003a) for which the selection criteria are a combination of peak flux and velocity width.

3.3. *The Selection Function*

Whereas for the $V_{\rm max}$ method the normalization of the mass function is automatically determined, for the maximum likelihood methods the normalization has to be determined afterward. First we have to evaluate the selection function $S(D)$, which describes the probability that an object at distance D is detected by the survey. For a flux-limited sample the selection function would normally be calculated with

$$S(D) = \frac{\int_{\max[M_{\rm lim}(D), M_{\rm low}]}^{M_{\rm high}} \theta(M) dM}{\int_{M_{\rm low}}^{M_{\rm high}} \theta(M) dM} \ , \quad (14)$$

where $M_{\rm high}$ and $M_{\rm low}$ are the highest and lowest values of $\log M_{\rm H\,I}$ in the sample. For our two-dimensional distribution function this would translate to

$$S(D) = \frac{\int_{W_{\rm low}}^{W_{\rm high}} \int_{\max[M_{\rm lim}(D,W), M_{\rm low}]}^{M_{\rm high}} \theta(M, W) dM\, dW}{\int_{W_{\rm low}}^{W_{\rm high}} \int_{M_{\rm low}}^{M_{\rm high}} \theta(M, W) dM\, dW} \ , \quad (15)$$

where $W_{\rm high}$ and $W_{\rm low}$ are the highest and lowest values of $\log w_{20}$ in the sample and $M_{\rm lim}(D, W)$ is the minimal detectable H I mass of a galaxy with logarithmic profile width W and at distance D megaparsecs. The problem is that we cannot derive an analytical expression for $M_{\rm lim}(D, W)$. It would be possible to derive an empirical relation for $M_{\rm lim}(D, W)$, but it is far better to simply make use of the actual data. Therefore, we evaluate the selection function in equation (15) for each galaxy individually and then average the values of $S(D_i)$ in bins of distance.

The mean galaxy density $\bar{n}$ is then determined by correcting the measured distance distribution of objects by the selection function. There are various methods described in the literature for determining $\bar{n}$ (Davis & Huchra 1982; Willmer 1997). Here we choose to use the minimum-variance estimator (Davis & Huchra 1982), but tests with different estimators gave very similar results. In the calculation of $\bar{n}$, the selection function is weighted with the inverse of the second moment of the two-point correlation function, which we set to $J_3 = 8000\ h_{75}^{-3}$ Mpc^3 (Meyer et al. 2003b). The value of $\bar{n}$ is found to be very insensitive to the exact value of J_3. Since at very low and high distances the number of galaxies is small, the selection function is not accurately known in those regions. We therefore choose to limit the calculation of $\bar{n}$ to $0.001 < S(D) < 0.1$, which roughly corresponds to the distance range 10 Mpc $< D <$ 50 Mpc. Limiting the calculation to this distance range also ensures that the overdense volume within 2 correlation lengths around the Milky Way is not taken into account in the calculation of $\bar{n}$. Finally, we normalize the H I mass function by setting

$$\int_{W_{\rm low}}^{W_{\rm high}} \int_{M_{\rm low}}^{M_{\rm high}} \theta(M, W) dM\, dW = \bar{n} \ . \quad (16)$$

3.4. *Testing the 2DSWML Method*

Before applying the two-dimensional SWML method to the BGC, we run a number of simulations to test whether the method produces reliable results. Using a method similar to the simulations in § 3.1, we fill volumes with synthetic galaxy samples and select sources from the volume in the same manner as the BGC is selected. The galaxy samples are constructed such that the distribution and the correlation statistics of peak flux, velocity width, and H I mass are similar to those of the BGC, and the samples typically contain $\sim$1000 galaxies. Each test is based on 100 simulations.

Figure 2 shows the results of three different tests. In the left panels the input HIMFs are represented by solid lines, whereas the measured HIMFs are shown by circles and error bars. In the right panels the histograms show the redshift distribution of the simulated data, and the solid curves are the reconstructed selection functions multiplied by $\Omega D^2\, \Delta D \bar{n}$. The dotted lines represent 1 σ uncertainties on the selection functions. From top to bottom, we show simulated data based on (1) a faint-end slope $\alpha = -1.1$ and an overdensity at $\sim$30 Mpc, (2) $\alpha = -1.5$ and an underdensity at $\sim$20 Mpc, and (3) $\alpha = -1.3$ and an underdensity at $\sim$30 Mpc and an overdensity at $\sim$10 Mpc.

It is clear that the 2DSWML method is capable of recovering the input HIMF with high accuracy, independent of strong over- and underdensities in the redshift distribution of galaxies. Also, our implementation of the calculation of the selection function gives satisfactory results. The selection function corresponds well to the overall redshift distribution and is insensitive to strong density variations. Motivated by the success of the 2DSWML method on

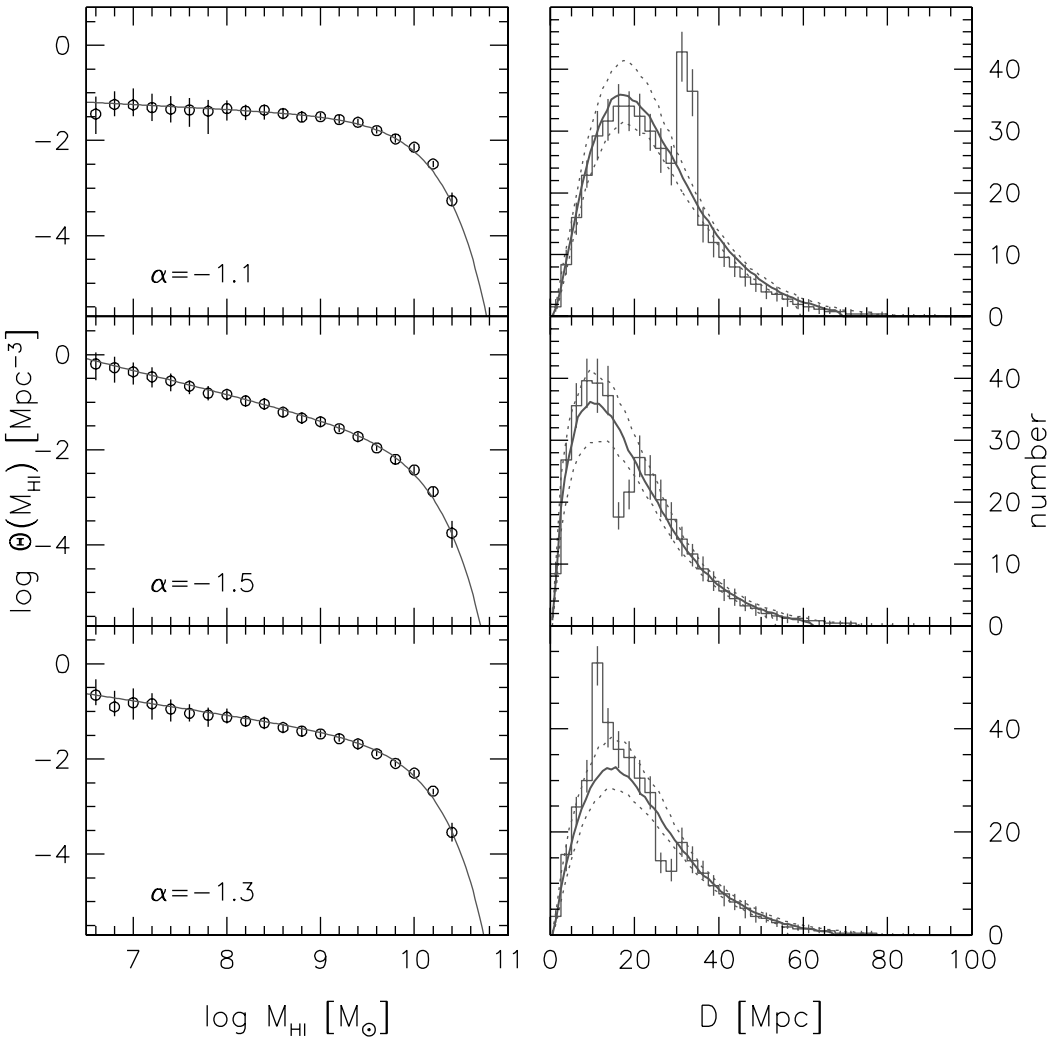

FIG. 2.—Testing the 2DSWML method. *Left*, input HIMFs (*solid lines*) and recovered HIMFs (*circles and error bars*); *right*, redshift distribution of the simulated data (*histograms)* and reconstructed selection functions multiplied by $\Omega D^2 \Delta D \bar{n}$ (*solid lines*). The dotted lines represent 1 σ uncertainties on the selection functions.

synthetic data, we chose to apply it to BGC. The results are described in the next section.

4. RESULTS

A gray-scale representation of the density distribution function θ_{jk} is shown in Figure 3. This function is calculated with bin sizes $\Delta M = 0.2$ and $\Delta W = 0.1$. This choice of bin sizes results in a 24×15 grid containing 360 bins in $dM\,dW$, of which 142 contain galaxies. The mean occupancy of the bins is therefore seven galaxies per bin. In Figure 3 we also plot the data points on which the analysis is based. It is clear that the low-mass bins contain very few galaxies, which explains the noisy character of the lower left part of the figure. Not surprisingly, there is a strong correlation between the H I mass of a galaxy and its observed velocity width. Note that the plotted velocity width is the observed width of the H I signals, uncorrected for galaxy inclination.

A comparison between the gray scales and the data points illustrates that a peak-flux–limited sample is, at a particular H I mass, slightly biased toward galaxies with low velocity widths, or alternatively, at a particular velocity width, slightly biased toward galaxies with high H I masses. Although the survey contains a large fraction of high H I mass, low velocity width galaxies, the true space density of these objects is very low.

The H I mass function can now be found by applying equation (13) to the two-dimensional density distribution θ_{jk}. The result is shown in Figure 4, where the circles show the measured space density of objects per decade of H I mass and the solid line is the best-fit Schechter function, with parameters $\alpha = -1.30 \pm 0.08$, $\log(M^*_{\rm H\,I}/M_\odot) = 9.79 \pm 0.06$, and $\theta^* = (8.6 \pm 2.1) \times 10^{-3}$ Mpc^{-3}. The Schechter function provides an excellent fit to the data. In fact, thanks to the large sample size and hence small Poisson errors, we convincingly show here for the first time that the H I mass function of galaxies can be satisfactorily described by a Schechter function. Note, however, that the shape and the normalization of the HIMF are determined without using a Schechter function as an assumption about the intrinsic shape of the HIMF.

The uncertainties on the Schechter parameters are determined with the jackknife error estimator (Lupton 1993). This is done by simply dividing the sample into 24 equal

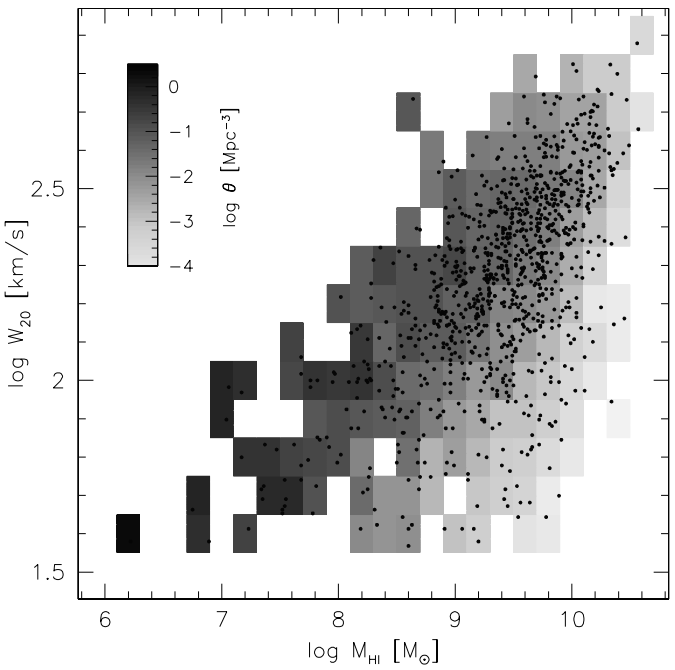

FIG. 3.—Bivariate galaxy density distribution in the ($M_{\rm H\,I}$, w_{20})-plane, derived via the 2DSWML method. The gray scales are logarithmic and represent the space density per decade of $M_{\rm H\,I}$ and decade of w_{20}. The distribution of data points is shown by circles.

regions of sky and calculating the H I mass function 24 times, leaving out a different region each time. A Schechter function is fitted to each mass function, and the errors on the parameters are found by

$$\sigma_x^2 = \frac{N-1}{N}\sum (x-\bar{x})^2 \,, \qquad (17)$$

where $N = 24$. The quoted uncertainties therefore include statistical errors and incorporate the effects of large-scale structure, but they do not include systematic errors.

As is normally the case in these fits, the errors in α and $\log M_{\rm H\,I}$ are correlated in the sense that steeper faint-end slopes imply higher values of $\log M_{\rm H\,I}$. The correlation coefficient between α and $\log M_{\rm H\,I}$ is $r = -0.28$, but because of the small number of jackknife samples ($N = 24$), the error on r is large.

The two-dimensional distribution function θ_{jk} can also be integrated along lines of constant velocity width to obtain the "velocity width function," the space density of objects as a function of velocity width or, after inclination correction, as a function of rotational velocity. This analysis will be the topic of a forthcoming paper.

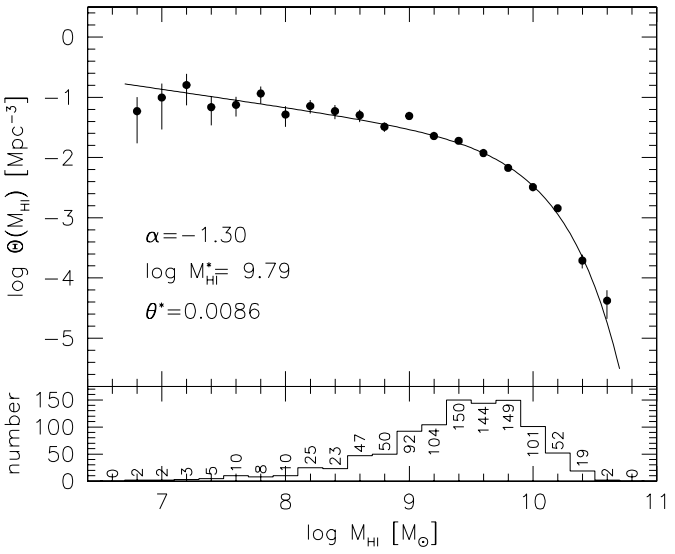

FIG. 4.—*Top*, BGC H I mass function; *bottom*, distribution of H I masses in the BGC per bin of 0.2 dex.

4.1. *Biases in the Calculation of the HIMF*

In the following we investigate the influence of various selection effects on the shape of the HIMF and on the measurement of the integral H I mass density. The effects of the biases are summarized in Table 1.

4.1.1. *Noise on* H I *Detection Spectra*

The BGC consists of the brightest galaxies in the total HIPASS galaxy sample, and therefore the S/N of the detections is high. The average rms noise in the HIPASS data is 13 mJy beam^{-1} per 18.0 km s^{-1}, which means that with the BGC detection limit of 116 mJy the lowest possible S/N level would be 9. In reality, because the detections are spread out over more than one resolution element, the S/Ns are generally higher. However, since most detections are much wider than one resolution element, the measured peak flux is often an overestimation of the galaxy's noise-free peak flux. This becomes increasingly important for flat-topped profiles, in which a large fraction of the channels are close to the peak flux.

We test this selection bias by adding Gaussian noise with an rms dispersion of 13 mJy to the H I profiles of synthetic galaxy samples that resemble the BGC and selecting galaxies by their measured peak flux. Next we calculate HIMFs from these samples using the 2DSWML method. Two extreme situations are tested, one in which all profiles are Gaussian and one in which all profiles are double-horned. Figure 5 shows the true HIMF of the simulated galaxies as a solid line, the recovered HIMF for double-horned profiles as circles, and the recovered HIMF for Gaussian profiles as crosses.

For both cases the selection bias causes a small global overestimation of the HIMF. For the Gaussian profiles, α is overestimated by $\Delta\alpha \approx 0.03$, whereas for the double-horned profiles the change in α is negligible. For the BGC, which is dominated by galaxies showing double-horned profiles, the integral H I mass density $\rho_{\rm H\,I}$ is probably overestimated by less than 15%.

4.1.2. *The Eddington Effect*

Another possible source of error in the H I mass function is the so-called Eddington effect, which is a steepening of the low-mass slope due to distance uncertainties. For the vast majority of the BGC galaxies, distances are calculated from their recession velocities, and peculiar velocities can introduce errors in these distance estimates and hence in the H I masses. In a relatively nearby sample such as the BGC the Eddington effect is potentially much more important than for deeper surveys such as AHISS, for which the peak of the galaxy distance distribution is much higher and relative distance uncertainties are smaller.

To test the severity of the Eddington effect, we add Gaussian noise to the recessional velocities in our synthetic

TABLE 1
BIASES IN THE HIMF DETERMINATION

Bias	Effect on HIMF	Effect on $\Omega_{\rm H\,I}$
Selection bias	Overall increase ($\approx$10%)	$\lesssim$15% over
Eddington effect	Steepening ($\Delta\alpha \approx -0.05$)	8% over
H I self-absorption	Overall decrease ($\approx -10\%$)	$\lesssim$15% under
Minimal velocity width	Flattening ($\Delta\alpha \approx 0.03$)	1% under
Confusion	Increase in $M^*_{\rm H\,I}$ ($\approx$15%) and decrease in θ^* ($\approx$15%)	None
Cosmic variance	Small error in shape	Uncertainty $\sim$10%

galaxy samples and select galaxies from the samples similar to the selection of the BGC. The results are summarized in Figure 6. The solid line shows the input HIMF and the symbols show the recovered HIMFs for samples for which dispersions of 50 km s^{-1} (*squares*), 100 km s^{-1} (*triangles*), 200 km s^{-1} (*crosses*), and 400 km s^{-1} (*circles*) are added. Each simulation was performed 100 times. The 2DSWML method has been used to calculate the HIMFs.

Figure 6 shows that the Eddington effect is a potentially important bias in our analysis. If a velocity dispersion of 200 km s^{-1} is added to our synthetic galaxy samples, the measured HIMF slope increases from $\alpha = -1.30$ to $\alpha = -1.43$. For a velocity dispersion of 100 km s^{-1}, α rises to -1.36; for 50 km s^{-1} the change in α is no longer measurable. As expected, the Eddington effect has almost no influence on the measured space density of sources around $M_{\rm H\,I} = M^*_{\rm H\,I}$; therefore the effect on the integral H I mass density $\rho_{\rm H\,I}$ is small. The inset in Figure 6 shows the recovered selection function $S(D)$ for each of the simulations, compared with the selection function of a galaxy sample without noise added to the distances (*dashed line*). The selection functions have been multiplied by D^2 and are normalized to a peak value equal to unity. The x-axis is logarithmic to better show the effects on the selection function at small distances. This clearly illustrates that the Eddington effect artificially steepens the selection function at small distances.

Our model of adding Gaussian noise to the recession velocities of galaxies oversimplifies the true peculiar-velocity distribution. The real dispersion is a function of local density (Strauss, Ostriker, & Cen 1998) and should be directional to some degree, instead of random. However, Sheth & Diaferio (2001) recently showed that the expected peculiar-velocity distribution in CDM simulations is nearly Gaussian. The one-dimensional pairwise peculiar-velocity dispersion σ_{12} is found to be $\approx$300 km s^{-1} on scales of 1 h^{-1}_{100} Mpc (e.g., Jing, Börner, & Suto 2002) and lower on smaller scales. This value compares with a velocity noise of $\sigma = 2^{-1/2}\sigma_{12} = 212$ km s^{-1}. Clusters of galaxies contribute significantly to this high value and the velocity field outside of clusters, where late-type galaxies dominate the statistics, is known to be much colder (e.g., Strauss et al. 1998). For example, Willick et al. (1997) find $\sigma = 125$ km s^{-1} for local ($cz < 3000$ km s^{-1}) spiral galaxies. More locally, the velocity field is even colder than this: the peculiar-velocity dispersion of galaxies within 5 Mpc of the Milky Way is only 60 km s^{-1} (Schlegel et al. 1994). Considering the latter two values, our simulations with dispersions of 50 and 100 km s^{-1} probably define the boundaries of a realistic test of the Eddington effect. From Figure 6 we conclude that our estimated HIMF probably overestimates the steepness of the faint-end slope α with $\Delta\alpha < 0.05$. Over the range of H I masses where we can reliably measure the HIMF, the H I mass density $\rho_{\rm H\,I}$ is maximally overestimated by 8%.

Our results are different from those of Rosenberg & Schneider (2002) who find that for their sample the Eddington effect is unimportant even for a dispersion of 600 km s^{-1}. This difference can be explained by the difference in survey depth between the BGC and ADBS. In the analysis

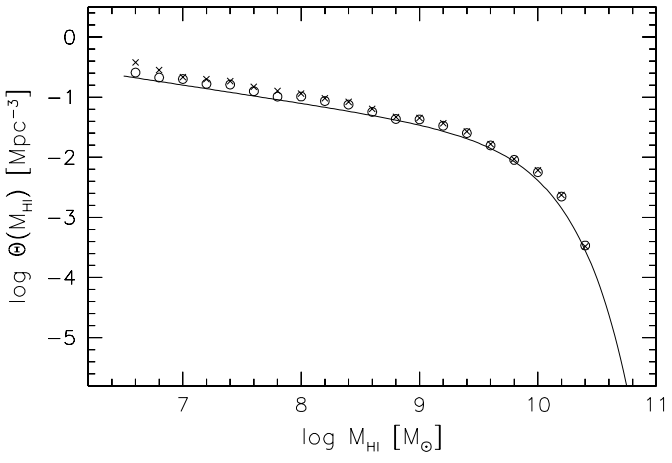

FIG. 5.—Effect of noise on the detection spectra of the HIMF determination. The solid line shows the input HIMF with a low-mass slope of $\alpha = -1.3$ that has been used to create synthetic galaxy samples. Noise with an rms fluctuation of 13 mJy has been added to the synthetic galaxy spectra. The points show the recovered 2DSWML HIMF assuming Gaussian profiles (*crosses*) and double-horned profiles (*circles*).

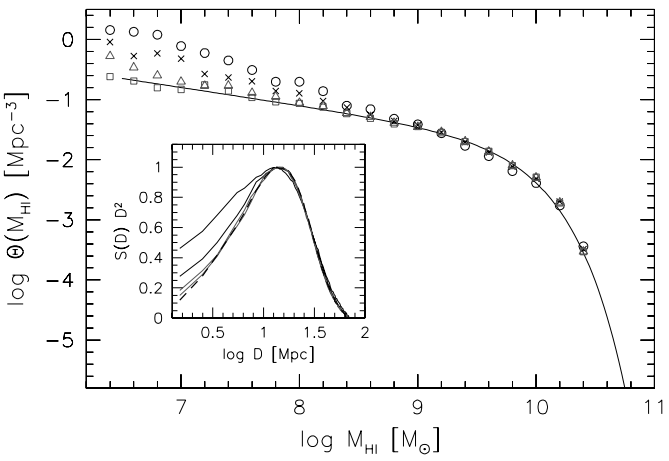

FIG. 6.—The Eddington effect. The solid line shows the HIMF of simulated data sets with a low-mass slope of $\alpha = -1.3$. The symbols show the recovered 2DSWML HIMFs after Gaussian noise is added to the recession velocities of the galaxies: dispersion of 50 km s^{-1} (*squares*), 100 km s^{-1} (*triangles*), 200 km s^{-1} (*crosses*), and 400 km s^{-1} (*circles*).

of the deep HIPASS catalog, the Eddington effect will be less important.

Of course, the H I mass estimates of the nearest galaxies suffer the most from the distance uncertainties. A solution would be to impose a lower distance limit on the galaxy sample above which the distance uncertainties are believed to be small. However, for samples other than integrated flux limited, imposing a lower distance limit causes a slight drop of the HIMF at low H I masses. This happens because in the low H I mass bins only those galaxies will be selected that have a high peak flux compared to other galaxies of the same H I mass at the same distance.

4.1.3. *Inclination Effects*

The inclination of galaxies might lead to two independent biases in the HIMF determination. One effect, H I self-absorption, is discussed in the next subsection (§ 4.1.4). Here we investigate the effect of applying in the galaxy selection procedure a minimal velocity width w_m below which galaxy signals cannot be reliably distinguished from radio-frequency interference. Since galaxies with low inclinations have smaller velocity widths, this selection effect might lead to an underrepresentation of face-on galaxies. Via the Tully-Fisher relation (Tully & Fisher 1977) and the relation between H I mass and optical luminosity (e.g., Roberts & Haynes 1994), it is known that galaxies with low H I masses have lower velocity widths. The selection bias therefore becomes progressively more important for dwarf galaxies, possibly causing a flattening of the H I mass function. Lang et al. (2002) were the first to calculate the severity of this effect and concluded that for their sample of galaxies the HIMF could be underestimated by 18.7% at $M_{\rm H\,I} = 2 \times 10^7\ M_\odot$. Here we make specific calculations for the BGC.

The velocity width measured at 50% of the peak flux can be written as

$$w_{50} = (w_0^2 \sin^2 i + w_{\rm tur}^2)^{1/2} + w_{\rm inst}\ , \qquad (18)$$

where w_0 is the intrinsic velocity spread due to rotation, $w_{\rm tur}$ is the velocity width due to turbulence in the gas layer, and $w_{\rm inst}$ is the contribution of instrumental broadening to the velocity width. For $w_{\rm tur}$ we adopt the conservative value of 6 km s^{-1} from Verheijen & Sancisi (2001) and for $w_{\rm inst}$ we use the Bottinelli et al. (1990) estimate of $w_{\rm inst} = 0.13 \times \delta v$, where δv is the velocity resolution. For our case, where $\delta v = 18$ km s^{-1}, we find that $w_{\rm inst} = 2.3$ km s^{-1}.

For detections to be included in the BGC, signal must be found in at least two consecutive channels. The BGC is limited to sources with $S_p > 116$ mJy, but the original search limit was approximately 50% lower, at 60 mJy (see Koribalski et al. 2002). We therefore assume that the minimal velocity width for inclusion in the BGC, $w_{\rm min}$, applies to the 50% level and hence $w_{50} > w_{\rm min} = 26.4$ km s^{-1}. Following Lang et al. (2002), we can calculate from equation (18) the minimal inclination angle $i_{\rm min}$ that a galaxy needs to have for it to be included in the sample

$$i_{\rm min} = \arcsin \left[\frac{(w_{\rm min} - w_{\rm inst})^2 - w_{\rm tur}^2}{w_0^2} \right]^{1/2} . \qquad (19)$$

Using the values given above we find that $i_{\rm min} = 13\overset{\circ}{.}5$ for $w_0 = 100$ km s^{-1} and $i_{\rm min} = 6\overset{\circ}{.}7$ for $w_0 = 200$ km s^{-1}. Recall that w_0 is the total velocity width of a galaxy, which is twice the rotational velocity.

From this we can estimate ζ, the fraction of galaxies potentially missed at every w_0, by integrating from $i = 0$ to $i = i_{\rm min}$ over a randomly oriented sample:

$$\zeta = \int_0^{i_{\rm min}} \sin i\, di = 1 - \cos i_{\rm min} = 1 - \left[1 - \frac{(w_{\rm m} - w_i)^2 - w_t^2}{w_0^2} \right]^{1/2} . \qquad (20)$$

We find that the values for ζ are very small: $\zeta = 2.7\%$ (0.7%) for $w_0 = 100$ km s^{-1} (200 km s^{-1}). We note that our adopted values for $w_{\rm inst}$ and $w_{\rm tur}$ are very conservative, and much higher values can be found in the literature. With higher values of $w_{\rm inst}$ and $w_{\rm tur}$, the resulting ζ would become even smaller.

Finally, in order to express ζ as a function of H I mass, we need to adopt a relation between $M_{\rm H\,I}$ and w_0. Lang et al. (2002) adopt the relation $w_0 = 0.42 M_{\rm H\,I}^{0.3}$, where in their case w_0 is derived from w_{20}. Here we use w_{50} and find that $w_0 = 0.35 M_{\rm H\,I}^{0.3}$ is a better fit to the BGC data. We note, however, that the scatter in the correlation between $M_{\rm H\,I}$ and w_0 is very large. An expression of ζ as a function of $M_{\rm H\,I}$ should therefore be regarded as illustrative. If we boldly apply the relation, we find that $\zeta = 10\%$ (5%) for $M_{\rm H\,I} = 2 \times 10^7\ M_\odot$ ($5 \times 10^7\ M_\odot$).

In Figure 7 we show the effect of the inclination bias. The solid line is an HIMF with $\alpha = -1.30$ and $\log M_{\rm H\,I}^* = 9.80$, and the dotted line is what would have been measured if galaxies with low inclination angles were missed in the survey. The dashed line is the same, but here we test the extreme case that $w_{\rm tur}$ and $w_{\rm inst}$ are negligible. The inset shows $\Theta(M_{\rm H\,I})$ on a linear scale to better show the differences between the lines. The inclination bias only becomes apparent at very low H I masses, implying that only the last two bins of our measured HIMF are affected by the bias, and then only slightly. Over the H I mass range where the HIMF is measured, the mass density $\rho_{\rm H\,I}$ is underestimated only by approximately 1%.

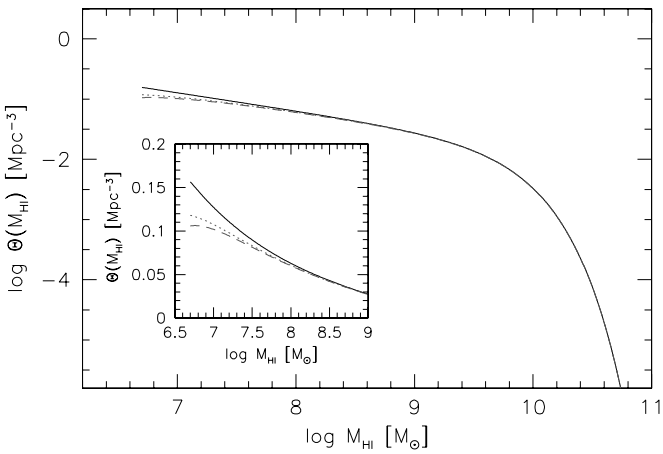

FIG. 7.—Effect of minimal velocity width on the HIMF measurement. Shown are the input HIMF (*solid line*), the 2DSWML HIMF that is measured if face-on galaxies are missed because of the minimal velocity width requirement (*dotted line*) and the same, but for $w_{\rm tur} = w_{\rm inst} = 0$ (*dashed line*). The inset shows $\Theta(M_{\rm H\,I})$ on a linear scale to better show the differences between the lines.

4.1.4. H I *Self-Absorption*

In the Milky Way the cold neutral phase of the interstellar medium can be seen as H I self-absorption (HISA) against the H I background (e.g., Knapp 1974; Gibson et al. 2000). It is not well known to what extent this HISA reduces the 21 cm emission-line flux of external galaxies. Deep, high-resolution 21 cm maps of nearby spirals (Braun 1997) show that between 60% and 90% of the H I flux comes from a network of cool gas ($T < 300$ K), which is expected to have a high opacity. However, not many instances of H I absorption against background continuum sources are observed in these regions. Dickey et al. (2000) find that in the Small Magellanic Cloud the abundance of cool-phase gas is only 15%, a factor of 2 lower than what is found in the Milky Way. The temperature of this gas is very low, typically 40 K or less. From detailed absorption studies they calculated the HISA correction as a function of column density, finding that the correction is negligible for $N_{\rm H\,I} < 3 \times 10^{21}$ cm^{-2} and rises to 1.4 at $N_{\rm H\,I} = 10^{22}$ cm^{-2}. These low correction values agree with the analysis of Zwaan, Verheijen, & Briggs (1999), who find that the column density distribution function of galaxies follows $N_{\rm H\,I}^{-3}$, expected for optically thin gas, up to $N_{\rm H\,I} = 8 \times 10^{21}$ cm^{-2}.

Haynes & Giovanelli (1984) plotted the H I mass of galaxies as a function of inclination and measured $f_{\rm H\,I}$, the ratio of HISA corrected to measured flux, for different morphological types. Zwaan et al. (1997) considered the effect of HISA on the HIMF, and based on Haynes & Giovanelli's (1984) result, they found that the average value of $f_{\rm H\,I}$ for a randomly oriented galaxy sample is approximately 1.10. They concluded that $M^*_{\rm H\,I}$ and θ^* would increase by no more than 10% if HISA effects were taken into account. The integral density $\rho_{\rm H\,I}$ could be underestimated by $\sim$15%, maximally. Lang et al. (2002) address the issue of HISA by plotting the distribution of inclination angles of their HIPASS galaxies and find that the number of highly inclined galaxies is lower than what is expected for a randomly oriented sample. From this they derive that $f_{\rm H\,I}$ is 1.25 and that $\rho_{\rm H\,I}$ could be underestimated by as much as 28.5%.

In Figure 8 the distribution of the cosine of the inclination angle i of cataloged BGC galaxies is drawn as a solid line (see Jerjen et al. 2003 for optical properties of the BGC sample). For a randomly oriented sample, the distribution should be flat. The BGC distribution is clearly underabundant at low and high inclinations, suggesting the effects of HISA at high i and the minimal velocity width effect (see § 4.1.3) at low i. However, also shown in Figure 8 is the cos i distribution for the 1000 brightest optically selected southern galaxies from the Lyon-Meudon Extragalactic Database (LEDA, *dashed line*), and the 1000 brightest optically selected southern galaxies with available 21 cm data (*dotted line*). The distributions are indistinguishable. This implies that the distribution of cos i is not determined by HISA and minimal velocity width, but by the inclination measurements available from the literature. As was discussed by Huizinga & van Albada (1992), purely circular isophotes are seldom observed in galaxies, leading to an apparent underrepresentation of face-on disks. On the other end of the distribution, three different effects cause a deficit of galaxies. First, because of dust obscuration, highly inclined galaxies drop out of the optical sample. These galaxies are also likely to be missed when our H I–selected sample is cross-correlated with optical catalogs. Second, the intrinsic thickness of galaxies causes uncertainties in i, especially for highly inclined galaxies. This effect causes the spike at cos $i = 0$ and may cause a deficit at low i. Third, HISA might cause an underrepresentation of high-i galaxies. However, since the optical information on the BGC is incomplete, it is not possible to disentangle the effects of HISA and optical dust extinction. We therefore conclude that based on the information at hand, no meaningful measurement of the effects of HISA can be made. In the absence of reliable intrinsic measurements, we adopt the externally measured value of 15% underestimation of $\Omega_{\rm H\,I}$, derived from the results of Zwaan et al. (1997).

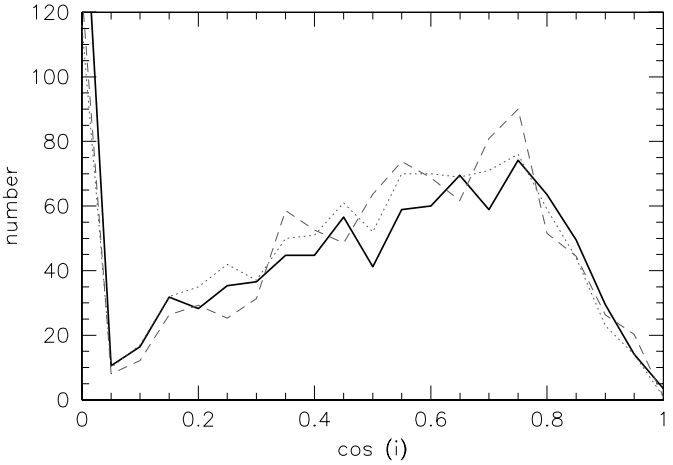

FIG. 8.—Distribution of the cosine of the inclination angle: BGC galaxies with optical information (*solid line*), the 1000 brightest southern optically selected galaxies from LEDA (*dashed line*), and the 1000 brightest southern optically selected galaxies with available 21 cm data (*dotted line*).

4.1.5. *Confusion*

The relatively coarse spatial resolution of HIPASS (15.′5) causes confusion of some signals with neighboring galaxies. Via Australia Telescope Compact Array follow-up observations and comparison with literature values, Koribalski et al. (2002) identified 11 compact groups and 44 pairs of galaxies in the BGC whose signals might chance to coincide in frequency. Furthermore, there are 67 sources that are flagged as "confused."

It is expected that some of the confused sources would not make it into the BGC if they were resolved, because their peak flux would drop below the selection limit. In that case we could overestimate the normalization of the HIMF, as artificial sources would enter the sample. On the other hand, if the unconfused sources had peak fluxes above the selection limit, galaxies would shift to higher mass bins, leading to an overestimation of $M^*_{\rm H\,I}$.

We test the two extremes by splitting all sources marked as "confused" or "pair" into two sources of equal H I mass and distinguish two cases: (1) the peak flux of the new sources is equal to that of their parent source, but their velocity width is halved, (2) the velocity width of the new sources is equal to that of their parent source, but the peak flux is halved. In the latter case only the new sources that have $S_p > 116$ mJy are counted. For case 1 we find that $M^*_{\rm H\,I}$ drops by $\sim$15%, θ^* increases by $\sim$15%, and α does not change significantly. Since the changes in θ^* and $M^*_{\rm H\,I}$ are

balanced, there is no net effect on the H I mass density $\rho_{\rm H\,I}$. In case 2 the changes in θ^* and $M^*_{\rm H\,I}$ are comparable, but in addition there is a marginal decrease in α of $\approx$0.02. Since in this case some sources drop below the detection limit, $\rho_{\rm H\,I}$ drops slightly, by $\approx$2%.

4.1.6. *Cosmic Variance*

In this section we investigate to what extent the shape of the HIMF depends on the region of the sky that is investigated. Previous blind H I surveys were based on relatively small regions of sky (66 deg^2 for AHISS and 430 deg^2 for ADBS), whereas the present analysis is based on a survey covering 2×10^5 deg^2. This large area would be expected to guarantee a fair sampling of the local volume, but on the other hand, the BGC is a shallow survey compared with AHISS and ADBS.

Figure 9 shows 2DSWML HIMFs for the four different quadrants of the southern sky. Around $M_{\rm H\,I} = M^*_{\rm H\,I}$, the variation in the HIMFs is only mild, but at $10^8\ M_\odot$ the estimated space density varies by a factor of 5, with the first quadrant being the most deviant. Not only is the estimated average density lowest there, but the low-mass end is also flatter. The cosmic variance is also reflected in the number of sources in each quadrant. The third quadrant contains 313 sources, whereas the fourth quadrant contains only 159 sources.

This exercise shows that selecting small regions of sky for galaxy surveys can introduce substantial uncertainties in the estimated space density of galaxies. We stress, however, that several properties of a survey define its immunity against large-scale structure: the sensitivity, the covered area of sky, and the shape of the covered area. The former two properties together define the search volume of the survey. Surveys such as AHISS and ADBS cover small declination ranges but 24$^{\rm h}$ in right ascension, whereas the HIPASS SCC survey covers more square degrees but is concentrated only on the south celestial cap. In addition, the immunity against large-scale structure depends on whether the search volume extends well beyond the local overdensity surrounding the Milky Way.

To quantify these statements and test the sensitivity to large-scale structure of blind H I surveys, we compute for each survey the number of 10^3 and 20^3 Mpc^3 cells that are probed for galaxies with different H I masses. The sizes of 10 and 20 Mpc were chosen because they span the typical sizes of voids in the local universe (Hoyle & Vogeley 2002; Plionis & Basilakos 2002). For a survey to be insensitive to large-scale structure it seems reasonable to require that it probe several such cells. Note that in the calculation of the number of cells we do not require that all cells be fully probed, but instead we calculate the number of cells from which each survey selects its detections. The results are given in Table 2, which also shows the search volumes for different H I masses for all surveys.

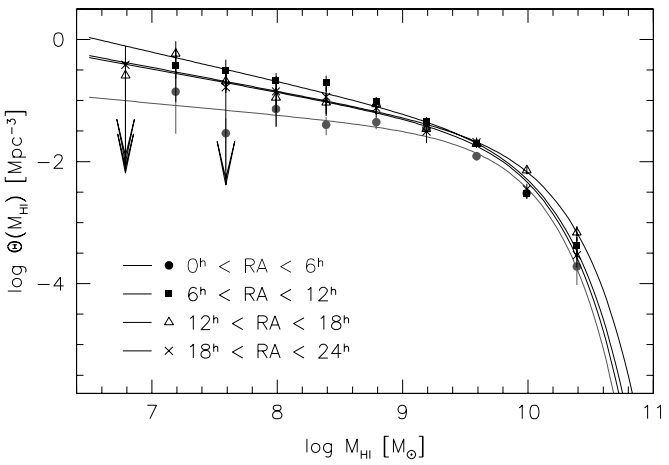

FIG. 9.—Effect of cosmic variance on the HIMF shape. The different symbols correspond to HIMF determinations in four different quadrants of the southern sky.

An important conclusion is that all surveys, except for the strip surveys AHISS and ADBS, select their $M_{\rm H\,I} = 10^8\ M_\odot$ sources from only one 20^3 Mpc^3 cell and at most a few 10^3 Mpc^3 cells. At $M_{\rm H\,I} = 10^9\ M_\odot$ most surveys cover a large number of cells, although the shallow, large-scale samples selected from HIPASS still probe fewer cells than the other surveys. At $M_{\rm H\,I} = 10^{10}\ M_\odot$ all surveys are probably insensitive to large-scale structure effects. Since these are the H I masses that dominate the total H I mass density (see § 7), measurements thereof based on HIPASS data are very secure.

Of course, this analysis may be too complementary to the strip surveys. Although these surveys go through many more cells than the large-scale surveys, their sensitivity to large-scale structure is not reduced as much as suggested by Table 2, because their cells are not fully sampled. This can

TABLE 2

NUMBER OF CELLS AND VOLUMES PROBED BY BLIND H I SURVEYS

	$M_{\rm H\,I}$ ($\sim 10^8\ M_\odot$)			$M_{\rm H\,I}$ ($\sim 10^9\ M_\odot$)			$M_{\rm H\,I}$ ($\sim 10^{10}\ M_\odot$)		
SAMPLE	10^3	20^3	Volume	10^3	20^3	Volume	10^3	20^3	Volume
HIPASS BGC	2	1	5	30	4	90	1500	190	1600
AHISS	60	15	0.35[a]	200	50	1.6[a]	200	50	1.6[a]
Arecibo Slice	3	1	0.25	30	8	4.2	50	12	6.5
HIPASS SCC	2	1	0.6	10	2	10	200	20	180
HIPASS HIZSS	2	1	0.25	14	4	4.7	100	24	84
ADBS	20	5	0.5	120	20	8.6	700	100	51

NOTE.—For each H I mass, the first column gives the number of probed 10^3 Mpc^3 cells and the second column gives the number of probed 20^3 Mpc^3 cells. The third column gives the approximate search volumes in 1000 Mpc^3. All search volumes are calculated using the quoted rms noise levels in the respective papers and assume optimal smoothing and a relation between velocity width and H I mass as given in Zwaan et al. 1997.

[a] Not counting sidelobes.

also be seen from the number of galaxies that the different surveys detect at each H I mass. ADBS detected seven galaxies with $M_{\rm H\,I} < 10^8\ M_\odot$, whereas the BGC has 38 galaxies in that mass range. Surely, the larger search volume and higher number of galaxies in the BGC must reduce its susceptibility to the effects of large-scale structure.

4.2. *Comparison with Other Methods*

Here we compare the results of the 2DSWML analysis of the BGC with those from other, more conventional estimators. Figure 10 shows the results of the standard $1/V_{\rm max}$ method (*open circles*), and the standard SWML application to an integrated-flux–limited subsample ($S_{\rm int} > 25$ Jy km s^{-1}; *triangles*). For comparison, the 2DSWML solution is also reproduced in this figure (*solid line*). The lower portion of Figure 10 shows the averaged $V/V_{\rm max}$ values for each bin in H I mass. The circles and triangles represent the total BGC and the integrated-flux–limited subsample, respectively, and values of $V_{\rm max}$ are calculated on the basis of their respective selection criteria.

The HIMFs from the 2DSWML method and the standard $1/V_{\rm max}$ method are in very good agreement. This is particularly striking because the latter method makes no corrections for the effects of large-scale structure, which are obviously present in our data (see Koribalski et al. 2002). The mean value of $V/V_{\rm max}$ for the $1/V_{\rm max}$ method is 0.503 ± 0.009, supporting the assertion that the sample is statistically complete and that the effects of large-scale structure average out over the whole sample. Zwaan et al. (1997) tested the effects of density variations on their $1/V_{\rm max}$ determination of the HIMF and also concluded that their calculation was insensitive to large-scale structure. For H I masses less than $10^8\ M_\odot$, the $1/V_{\rm max}$ method finds higher space densities than the 2DSWML method. This is most likely the result of the fact that these galaxies are all drawn from the very local universe (<10 Mpc; see Fig. 11), which we know to be overdense. Also, the Poisson errors on these points are very high, since they are based on small numbers of galaxies.

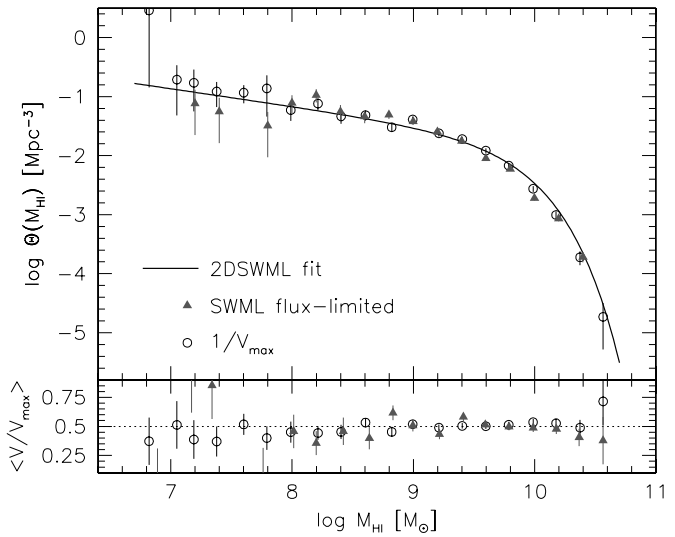

FIG. 10.—Comparison of different HIMF estimators. *Top*: HIMF of BGC galaxies derived via the $1/V_{\rm max}$ method (*open circles*) and the standard SWML HIMF of an integrated-flux–limited subsample (*triangles*). The 2DSWML HIMF from Fig. 4 is reproduced as a solid line. *Bottom*: Median values of $V/V_{\rm max}$ in each bin, with 1 σ uncertainties, via the $1/V_{\rm max}$ method (*circles*) and the standard SWML method (*triangles*).

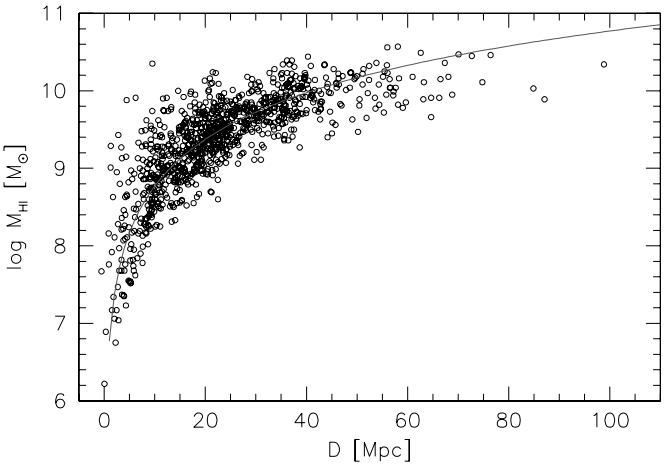

FIG. 11.—H I masses of BGC galaxies as a function of their distance. The solid line indicates an integrated flux limit of 25 Jy km s^{-1}, above which the sample is "complete."

The SWML HIMF is also in good agreement with both other methods, but the uncertainties on the space densities for $M_{\rm H\,I} < 10^8\ M_\odot$ are large because the integrated-flux–limited subsample contains only a small number of low-mass galaxies. The mean $V/V_{\rm max}$ value is 0.500 ± 0.013, which suggests that the integrated-flux–limited subsample is complete.

5. COMPARISON WITH OTHER SURVEYS

In the previous section we showed that the results on the HIMF are internally consistent. We now compare the results with those of previous H I surveys. In Figure 12 we reproduce the best-fit Schechter function to the BGC HIMF as a solid line. In addition, we plot the HIMFs from AHISS (Zwaan et al. 1997), the Arecibo Slices (AS; Schneider et al.

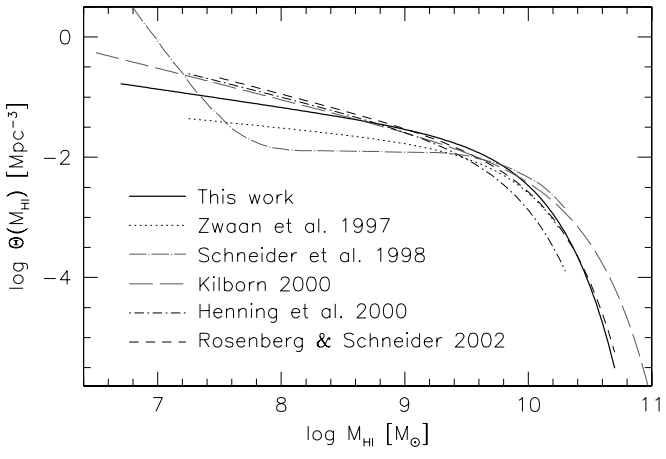

FIG. 12.—Comparison of the HIPASS BGC HIMF with previous calculations from blind H I surveys. Shown are HIMFs from AHISS (Zwaan et al. 1997), AS (Schneider et al. 1998), the HIPASS SCC (Kilborn 2000), the HIPASS HIZSS (Henning et al. 2000), and ADBS (Rosenberg & Schneider 2002). All mass functions have been converted to the same value of $H_0 = 75$ km s^{-1} Mpc^{-1} and are only plotted over the H I mass range in which they have been measured.

TABLE 3
SCHECHTER PARAMETERS OF H I MASS FUNCTIONS

Sample	α	$M^*_{\rm H\,I}$ ($h_{75}^{-2}\,M_\odot$)	θ^* ($10^{-4}\,h_{75}^3$ Mpc^{-3})	Ref.
HIPASS BGC	-1.30	9.79	86	1
AHISS	-1.20	9.80	59	2
HIPASS SCC	-1.52	10.1	32	3
HIPASS HIZSS	-1.51	9.70	60	4
ADBS	-1.53	9.88	50	5

REFERENCES.—(1) This paper; (2) Zwaan et al. 1997; (3) Kilborn 2000; (4) Henning et al. 2000; (5) Rosenberg & Schneider 2002.

1998), ADBS (Rosenberg & Schneider 2002) and the HIPASS SCC (Kilborn 2000). The ranges over which the Schechter function have been plotted indicate the H I mass bracket within which the HIMFs have been reliably determined from the different surveys. All curves have been converted to $H_0 = 75$ km s^{-1} Mpc^{-1}. The Schechter parameters of the various HIMFs are given in Table 3.

All curves agree very well at the high-mass end, which can also be seen from the value of $\log M^*_{\rm H\,I}$, which is $\approx$9.8 for all surveys. Discrepancies arise at the low-mass end, where Zwaan et al. (1997) found $\alpha = -1.2$ and Rosenberg & Schneider found $\alpha = -1.53$. Undoubtedly, part of these differences can be explained by small number statistics, but different treatments of the survey completeness and different HIMF estimators may also cause discrepancies.

AHISS (Zwaan et al. 1997) and ADBS (Rosenberg & Schneider 2002) are both drift-scan surveys performed with the Arecibo telescope. AHISS is more sensitive, with a 1 σ rms noise of $\sim$0.75 mJy beam^{-1} compared with $\sim$3.5 mJy beam^{-1} per 32 km s^{-1} for ADBS, but the sky coverage of ADBS was much larger, which is reflected in the number of detections: 66 in AHISS and 265 in ADBS. Both galaxy samples are based on by-eye examinations of the H I spectra, which means that the completeness limits have to be calculated a posteriori. Assessing the completeness limits of the drift-scan surveys is complicated because the sensitivity is a function of declination offset, which is particularly important if sidelobe detections have to be taken into account (AHISS). Zwaan et al. (1997) derived an analytical expression for the "detectability" of sources; Rosenberg & Schneider (2002) used a large number of synthetic sources to test their survey's completeness. One of the differences in the analysis of both surveys is the treatment of the variation of sensitivity S with velocity width, which can be expressed as $S \propto \Delta V^{\beta}$. Assuming optimal smoothing and uncorrelated noise, the expected value of β is 0.5, which was adopted by Zwaan et al. (1997). Rosenberg & Schneider (2002) argue that $\beta = 0.75$, which leads to a reduced sensitivity to large line width sources and hence, via the Tully-Fisher relation, to a reduced sensitivity to high H I mass sources.

With a peak flux limit of 116 mJy, the BGC is a relatively shallow survey compared with ADBS and AHISS. However, thanks to the BGC's large sky coverage, the number counts at the low-mass end are better than those from AHISS and ADBS, and the BGC HIMF extends to lower H I masses than either of the Arecibo surveys. This low sensitivity causes all low-mass galaxy detections to be at small distances (see Fig. 11), which implies that the Eddington effect might have a significant influence on the HIMF shape.

Davies et al. (2001) recently advocated an HIMF slope of $\alpha \approx -2$, which they derived by extracting H I spectra from the HIPASS public release database, and compared the distance distribution of detections with what would be expected for different HIMFs. This method intrinsically assumes that a relationship exists between distance and the minimal detectable H I mass at that distance. We have argued in § 3.1 that this assumption does not hold for H I–selected galaxy samples. The validity of the Davies et al. (2001) result is therefore unclear.

6. DEPENDENCE OF THE HIMF ON GALAXY TYPE

The BGC was cross-correlated with LEDA to find matches with cataloged galaxies (see Jerjen et al. 2003). In addition to this, Ryan-Weber et al. (2002) searched for uncataloged BGC galaxies on the Digitized Sky Survey and determined morphological types for these new galaxies. In total, morphological type information is now available for 892 galaxies out of our total sample of 1000. The 108 unclassified galaxies consist mainly of galaxies at low Galactic latitude, groups or pairs for which a unique match between H I signal and optical galaxy could not be made, and detections for which a cross-correlation with LEDA galaxies is otherwise ambiguous.

In order to determine the morphological type dependence on the HIMF, we divide the BGC into five subsets of galaxies: E–S0, Sa–Sb, Sbc–Sc, Scd–Sd, and Sm–Irr. Not surprisingly for an H I–selected sample, the subset of early-type galaxies contains only 43 galaxies, which is marginally sufficient to calculate a meaningful HIMF. In addition, we calculate the HIMFs for galaxy samples divided into "late" and "early," where galaxy types later than Sb are regarded as late.

Figure 13 shows the H I mass functions for different morphological types. We have chosen to apply the $1/V_{\rm max}$ method here, because the subsamples are uncomfortably small for a two-dimensional analysis. For reference, the best-fit HIMF for the total BGC is indicated by a dotted line in each panel. Best-fit Schechter functions to the individual HIMFs are also shown, but note that for the E–S0 subsample the normalization and $M^*_{\rm H\,I}$ are very poorly defined. All HIMFs are multiplied by 1000/892 to correct for the incompleteness in morphological classification. Table 4 includes the best-fit Schechter parameters for all morphological types.

There is no clear variation in the HIMF slope between types Sa and Sd; the HIMF is flat with $\alpha \sim -1.0$. Only for galaxies in the Sm–Irr bin do we see a steepening of the HIMF to $\alpha = -1.4$. This effect is very similar to what is

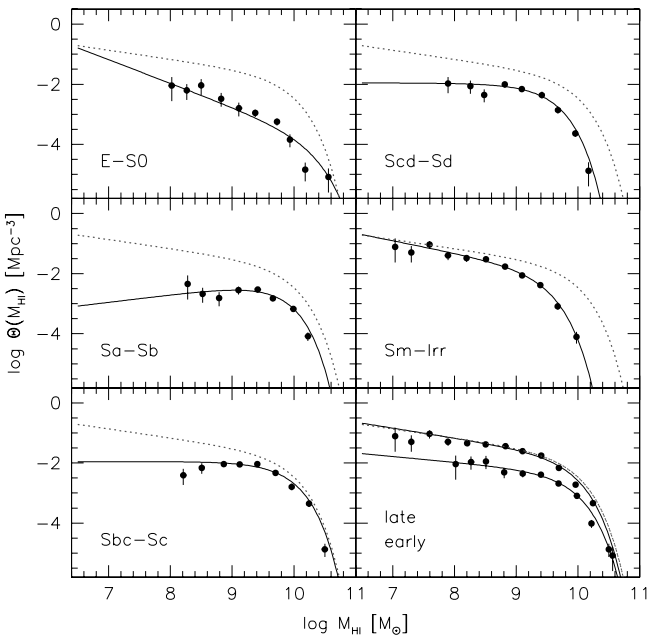

FIG. 13.—H I mass function for different morphological types. Shown are Schechter fits to the type-specific HIMFs (*solid lines*) and the HIMF for the total sample (*dashed line*). In the bottom right panel, late-type galaxies are all types later than Sb.

found for optical luminosity functions (Marzke et al. 1998). Zwaan, Briggs, & Sprayberry (2001) used the Marzke et al. (1998) luminosity functions and fitted relations between H I mass and optical luminosity to derive type-specific H I mass functions. These are in good agreement with what is found here and also show a steepening for the very late type galaxies. Apparently, both the number density of low-luminosity and low H I mass galaxies are dominated by late-type galaxies.

Another conclusion from Figure 13 is that the characteristic mass $M^*_{\rm H\,I}$ is much higher for types Sbc–Sc than for the other types. This reflects that most galaxies with high H I masses are of types Sbc–Sc and that these types are the most dominant contributors to the total H I mass density.

Similarly, the HIMF can be divided into galaxies of different optical surface brightness. The question of how much low surface brightness (LSB) galaxies contribute to the local H I density and the total local baryon budget has been addressed before by several authors. Based on local galaxy samples with targeted 21 cm spectroscopic follow-up, Briggs (1997a, 1997b) concluded that LSB galaxies make a $\sim$10% addition to the total H I mass density. Zwaan et al. (2001) used AHISS to find that galaxies with central surface brightness greater than 23.0 mag arcsec^{-2} contribute 18% to the H I density. The luminosity density contained in LSB galaxies has also been found to be low (Sprayberry et al. 1997; Driver 1999; de Jong & Lacey 2000; Zwaan et al. 2001).

Measurements of optical surface brightness for the BGC are drawn from LEDA. We choose to use the mean effective surface brightness, $\mu_{\rm eff}$, which is defined as the mean surface brightness inside an aperture enclosing one-half the total light. Unfortunately, the measurements are not complete: $\mu_{\rm eff}$ is available for only for 600 galaxies in the BGC sample. The uncertainties in the HIMFs for LSB and high surface brightness (HSB) galaxies are therefore large. We divide the sample into two subsamples, where we use $\mu_{\rm eff} = 24.0$ mag arcsec^{-2} as the boundary between an LSB and an HSB galaxy. This separation results in 96 LSB and 504 HSB galaxies. For an exponential disk, the value of $\mu_{\rm eff} = 24.0$ mag arcsec^{-2} compares to a central surface brightness of $\mu_{\rm eff} = 22.2$ mag arcsec^{-2}. The values of $\mu_{\rm eff}$ are uncorrected for inclination and dust extinction.

Figure 14 shows HIMFs for the LSB and HSB subsamples. Both are corrected by a factor 1000/600 to account for the incompleteness in surface brightness measurements. Similarly to what was found by Briggs (1997a), we see that the HIMF for LSB galaxies is steeper ($\alpha = -1.35$) than that for HSB galaxies ($\alpha = -1.2$). The obvious reason for this is that there is a relation between surface brightness and H I mass, in the sense that LSB galaxies have lower total H I masses. LSB galaxies therefore increasingly populate the lower H I mass bins, resulting in a steeper H I mass function compared with HSB galaxies. Nonetheless, LSB galaxies contribute only very little to the total H I mass density. Based on the Schechter fits, we find that this contribution is $\sim$15%.

TABLE 4
TYPE-SPECIFIC H I MASS FUNCTIONS

Type	α	$M^*_{\rm H\,I}$ ($h_{75}^{-2}\,M_\odot$)	θ^* ($10^{-4}\,h_{75}^3$ Mpc^{-3})
E–S0	−1.78	10.2[a]	0.92[a]
Sa–Sb	−0.75	9.65	22
Sbc–Sc	−1.00	9.77	49
Scd–Sd	−1.01	9.42	46
Sm–Irr	−1.41	9.32	61
Late	−1.33	9.75	78
Early	−1.19	9.76	21

[a] Values uncertain because a Schechter function is a poor fit to the data points.

7. THE H I MASS DENSITY

Figure 15 shows the H I mass density $\rho_{\rm H\,I}$ contained in galaxies of different H I mass. The solid line shows the converted best-fit Schechter function, and the dotted and dashed lines

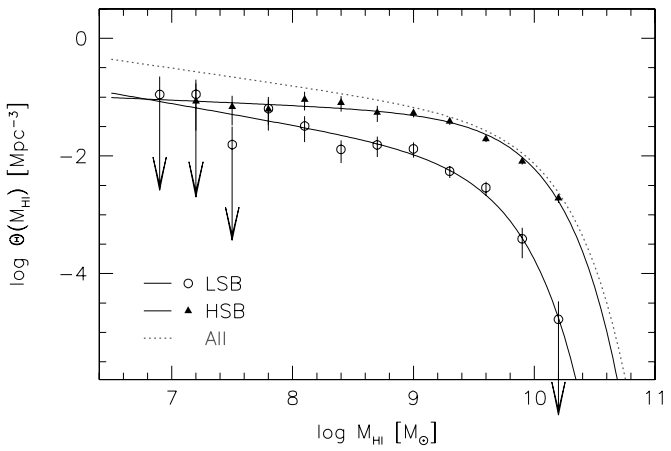

FIG. 14.—H I mass function for LSB and HSB galaxies. The HIMF for the total sample is drawn as a dotted curve.

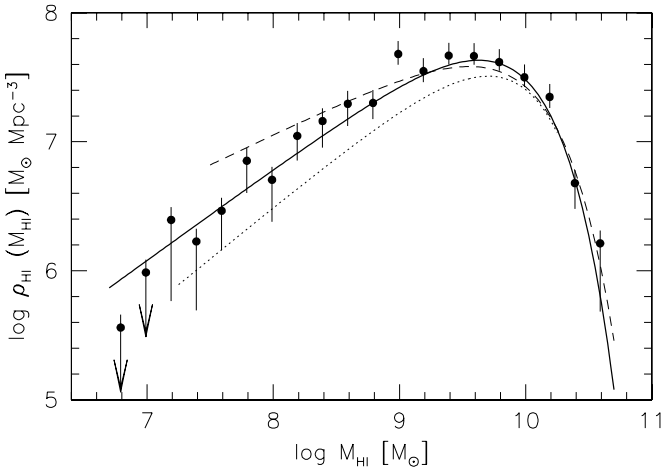

FIG. 15.—Distribution of H I mass density over different H I masses. The points show the measured mass density derived by multiplying the 2DSWML HIMF by $M_{\rm H\,I}$. The lines show the converted best-fit Schechter function (*solid line*), the mass density distribution function from Zwaan et al. (1997; *dotted line*) and that derived by Rosenberg & Schneider (2002; *dashed line*).

show the H I mass densities derived by Zwaan et al. (1997) and Rosenberg & Schneider (2002), respectively. The bivariate H I mass density distribution in the ($M_{\rm H\,I}$, W_{20})-plane is shown in Figure 16. This figure clearly shows that the gas mass density is dominated by galaxies with H I masses around $M^*_{\rm H\,I}$ and velocity widths of $\sim$250 km s^{-1}.

The total H I mass density contained by galaxies in the local universe is calculated with $\rho_{\rm H\,I} = \theta^*\Gamma(2+\alpha)M^*_{\rm H\,I}$, and is found to be $\rho_{\rm H\,I} = (6.9 \pm 1.1) \times 10^7\ h_{75}\ M_\odot$ Mpc^{-3}. A straight summation of $M_{\rm H\,I}\,\theta(M_{\rm H\,I})$ using the points in Figure 4 gives a slightly lower value of $\rho_{\rm H\,I} = (6.6 \pm 1.1) \times 10^7$ $h_{75}\ M_\odot$ Mpc^{-3} because this calculation does not include the extensions of the Schechter fit beyond the region where we can measure it reliably. Taking into account all biases as summarized in Table 1, we find $\rho_{\rm H\,I} = (6.1 \pm 1.0) \times 10^7\ h_{75}$ $M_\odot$ Mpc^{-3}, or $(4.1 \pm 0.7) \times 10^{-33}$ g cm^{-3}, where the error is the 1 σ uncertainty derived with the jackknife method, which includes random errors and the effects of large-scale structure. For comparison, Zwaan et al. (1997) found $\rho_{\rm H\,I} = 4.3 \times 10^7\ h_{75}\ M_\odot$ Mpc^{-3} and using Rosenberg & Schneider's (2002) Schechter parameters we find $\rho_{\rm H\,I} = 7.1 \times 10^7\ h_{75}\ M_\odot$ Mpc^{-3}. Briggs (1990) used luminosity functions and a conversion factor from luminosity to H I mass to find $\rho_{\rm H\,I} = 7.0 \times 10^7\ h_{75}\ M_\odot$ Mpc^{-3}, a value very close to our measurement.

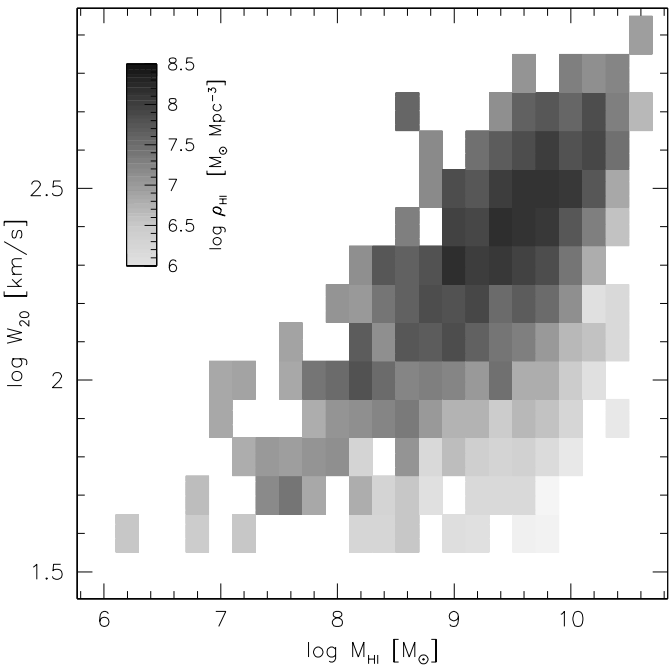

FIG. 16.—Bivariate H I mass density distribution in the ($M_{\rm H\,I}$, w_{20}) plane. The gray scales are logarithmic and represent the mass density in $M_\odot$ Mpc^{-3} per decade of $M_{\rm H\,I}$ and decade of w_{20}. The H I mass density is dominated by high-mass, high velocity width galaxies.

Converting $\rho_{\rm H\,I}$ to the more convenient $\Omega_{\rm H\,I}$, the H I mass density as a fraction of the critical density of the universe, we find $\Omega_{\rm H\,I} = (3.8 \pm 0.6) \times 10^{-4}\ h_{75}^{-1}$, and after making a correction for the 24% helium mass fraction, we find $\Omega_{\rm atomic} = (4.8 \pm 0.8) \times 10^{-4}\ h_{75}^{-1}$. This measurement of $\Omega_{\rm H\,I}$ is very robust because the galaxies that contribute most to the gas density are also the ones that dominate the counting statistics ($M_{\rm H\,I} = 10^9$–$10^{10}\ M_\odot$). Uncertainties in the faint-end slope α therefore contribute little to the total error in $\Omega_{\rm H\,I}$. Furthermore, since these galaxies are found at large distances, and therefore over a large region of the sky, the effects of large-scale structure are unimportant.

Fukugita, Hogan, & Peebles (1998) estimated that the mass fraction in molecular gas is approximately 80% of that of atomic gas. Keres, Yun, & Young (2002) measure the CO luminosity function and derive a mass fraction in molecular gas of approximately 60%. Averaging these two values, we find that the total mass density in cool gas in the local universe is $\Omega_{\rm cool\,gas} = 7.6 \times 10^{-4}\ h_{75}^{-1}$. To put these numbers into perspective, the most recent determination of the baryon density via the primeval deuterium abundance is $\Omega_{\rm baryon} h_{75}^2 = 0.035 \pm 0.004$ (Burles, Nollett, & Turner 2001). Microwave background anisotropy measurements give slightly higher values (e.g., de Bernardis et al. 2000). We estimate from this that cool gas in galaxies makes up approximately 2% of the total baryon density in the local universe. The total mass in diffuse ionized intergalactic gas is much higher than this. From low-redshift *Hubble Space Telescope* spectra, Penton, Shull, & Stocke (2000) derived that the mass density in the Lα forest is approximately 20% of $\Omega_{\rm baryon}$.

Our derived value of $\Omega_{\rm H\,I}$ is approximately 5 times lower than that at redshifts 2–4 (Ellison et al. 2001; Péroux et al. 2001, Storrie-Lombardi, McMahon, & Irwin 1996). A gradual conversion from neutral gas to stars in the disks of galaxies is generally believed to cause this decline in $\Omega_{\rm H\,I}$ (Lanzetta et al. 1995; Pei, Fall, & Hauser 1999), but recent results on high column density QSO absorption-line systems at intermediate redshifts have confused this picture. Lane (2000) searched for intermediate-z 21 cm absorption in Mg II–selected systems and used Mg II statistics to bootstrap the $\Omega_{\rm H\,I}$ values. The same statistics was used by Rao & Turnshek (2000) for a Lyα survey of $z < 1.65$ Mg II–selected systems. Both authors find values of $\Omega_{\rm H\,I}$ ($z < 1.65$) consistent with those at higher redshifts, indicating that the neutral gas density does not evolve strongly from high z to the present. Churchill (2001) recently performed an unbiased survey for low-z Mg II systems and used the same statistics for Mg II systems as Rao & Turnshek (2000) did, and derived $\Omega_{\rm H\,I}$ ($z = 0.05$), which is a factor 5 larger than our results. The uncertainties on all these measurements are very

large, since they suffer from small number statistics. The evolution of $\Omega_{\rm H\,I}$ from $z = 2$ to the present time therefore remains very uncertain, and more intermediate-redshift optical and 21 cm surveys are required to constrain the evolution of H I.

8. THE BGC SELECTION FUNCTION

As described in § 3.3, the 2DSWML method allows us to calculate the selection function of the BGC. The selection function is an important input into calculations of clustering in the BGC, such as the two-point correlation function (Meyer et al. 2003b). It is therefore interesting to test whether the calculated selection function $S(D)$ is a valid approximation of the true selection of the survey.

In Figure 17 we compare the measured redshift distribution of BGC galaxies with the distribution implied by the derived selection function. The overall redshift distribution agrees well with the calculated curve. This is also reflected in the predicted number counts of galaxies from the selection function, which is 0.0456 galaxies deg^{-2}. This is only 6% lower than the measured number count in the BGC of 0.0485 galaxies deg^{-2}. This implies that if we had used the number counts to normalize the HIMF, in a manner similar to what is occasionally done for optical LFs (e.g., Norberg et al. 2002), we would have found essentially the same normalization.

By comparing the measured with the derived distribution, two overdensities become apparent: one at ~22 Mpc and one at ~36 Mpc. See Koribalski et al. (2002) for a detailed discussion of the large-scale structure in the BGC. Also shown in Figure 17 is the redshift distribution of the integrated-flux–limited subsample and the predicted distribution calculated with equation (14). The same density structure can be seen here, although less pronounced than for the full galaxy sample.

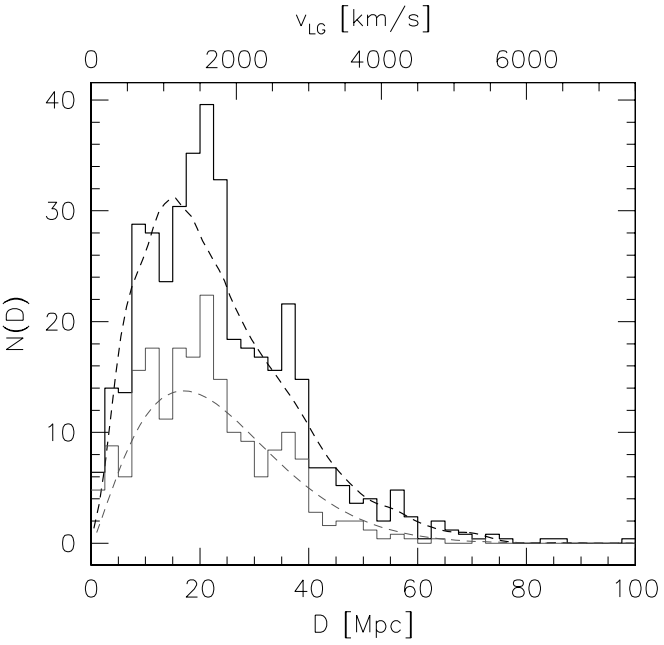

FIG. 17.—Distance distribution of the BGC. Measured redshift distribution of galaxies in the BGC (*histogram*) and predicted distribution based on the selection function $S(D)$, which is calculated with the 2DSWML method (*dark dashed curve*). Both distributions are normalized to $\Delta D = 1$ Mpc. The light lines show the same for a subsample with $S_{\rm int} > 25$ Jy km s^{-1}.

9. CONCLUSIONS

We have used the HIPASS Bright Galaxy Catalog (Koribalski et al. 2002), consisting of the 1000 southern galaxies with highest peak flux densities, to measure the H I mass function of galaxies in the local universe. This is the largest sample of galaxies ever used to measure the HIMF and contains galaxies over the H I mass range $\log (M_{\rm H\,I}/M_\odot) + 2 \log h_{75} = 6.8$–$10.6$. We have developed a bivariate stepwise maximum likelihood method to measure the HIMF. This method solves for H I mass and velocity width and then integrates over velocity width to find the HIMF. We show that this method is a reliable estimator and insensitive to the effects of large-scale structure. The resulting HIMF can be well fitted with a Schechter function with parameters $\alpha = -1.30 \pm 0.08$, $\log (M^*_{\rm H\,I}/M_\odot) = 9.79 \pm 0.06\ h_{75}^{-2}$, and $\theta^* = (8.6 \pm 2.1) \times 10^{-3}\ h_{75}^3$ Mpc^{-3}. We find that the faint-end slope of the HIMF is dependent on morphological type. Late-type galaxies show the steepest faint-end slopes, and these galaxies dominate the statistics at low H I masses. We extensively test the influence of possible biases in the HIMF determination, including peculiar motions of galaxies, inclination effects, selection biases, and large-scale structure, and we quantify these biases. The integral H I mass density in the local universe is found to be $\rho_{\rm H\,I} = (6.1 \pm 1.0) \times 10^7\ h_{75}\ M_\odot$ Mpc^{-3}, contributing a fraction $\Omega_{\rm H\,I} = (3.8 \pm 0.6) \times 10^{-4}\ h_{75}^{-1}$ of the critical density of the universe.

REFERENCES

Babul, A., & Ferguson, H. C. 1996, ApJ, 458, 100
Babul, A., & Rees, M. J. 1992, MNRAS, 255, 346
Barkana, R., & Loeb, A. 1999, ApJ, 523, 54
Barnes, D. G., et al. 2001, MNRAS, 322, 486
Blanton, M. R., et al. 2001, AJ, 121, 2358
Bottinelli, L., Gouguenheim, L., Fouqué, P., & Paturel, G. 1990, A&AS, 82, 391
Braun, R. 1997, ApJ, 484, 637
Briggs, F. H. 1990, AJ, 100, 999
———. 1997a, ApJ, 484, 618
———. 1997b, ApJ, 484, L29
Briggs, F. H., & Rao, S. 1993, ApJ, 417, 494
Burles, S., Nollett, K. M., & Turner, M. S. 2001, ApJ, 552, L1
Churchill, C. W. 2001, ApJ, 560, 92
Corbelli, E., Salpeter, E. E., & Bandiera, R. 2001, ApJ, 550, 26
Davies, J. I., de Blok, W. J. G., Smith, R. M., Kambas, A., Sabatini, S., Linder, S. M., & Salehi-Reyhani, S. A. 2001, MNRAS, 328, 1151
Davis, M., & Huchra, J. 1982, ApJ, 254, 437
de Bernardis, P., et al. 2000, Nature, 404, 955
de Jong, R. S., & Lacey, C. 2000, ApJ, 545, 781
Dekel, A., & Silk, J. 1986, ApJ, 303, 39
Dickey, J. M., Mebold, U., Stanimirović, S., & Staveley-Smith, L. 2000, ApJ, 536, 756
Driver, S. P. 1999, ApJ, 526, L69
Efstathiou, G., Ellis, R. S., & Peterson, B. A. 1988, MNRAS, 231, 431
Ellison, S. L., Yan, L., Hook, I. M., Pettini, M., Wall, J. V., & Shaver, P. 2001, A&A, 379, 393
Folkes, S., et al. 1999, MNRAS, 308, 459
Fukugita, M., Hogan, C. J., & Peebles, P. J. E. 1998, ApJ, 503, 518
Gibson, S. J., Taylor, A. R., Higgs, L. A., & Dewdney, P. E. 2000, ApJ, 540, 851
Haynes, M. P., & Giovanelli, R. 1984, AJ, 89, 758
Henning, P. A., et al. 2000, AJ, 119, 2686
Hoyle, F., & Vogeley, M. S. 2002, ApJ, 566, 641
Huizinga, J. E., & van Albada, T. S. 1992, MNRAS, 254, 677
Jerjen, H., et al. 2003, in preparation
Jing, Y. P., Börner, G., & Suto, Y. 2002, ApJ, 564, 15
Keres, D., Yun, M., S., & Young, J. S. 2003, ApJ, 582, 659
Kilborn, V. A. 2000, Ph.D. thesis, Univ. Melbourne
Kilborn, V. A., et al. 2002, AJ, 124, 690

Knapp, G. R. 1974, AJ, 79, 527
Koribalski, B. S., et al. 2002, AJ, submitted
Kraan-Korteweg, R. C., van Driel, W., Briggs, F., Binggeli, B., & Mostefaoui, T. I. 1999, A&AS, 135, 255
Lane, W. M. 2000, Ph.D. thesis, Univ. Groningen
Lang, R. H., et al. 2002, MNRAS, in press
Lanzetta, K. M., Wolfe, A. M., & Turnshek, D. A. 1995, ApJ, 440, 435
Loveday, J. 2000, MNRAS, 312, 557
Lupton, R. 1993, Statistics in Theory and Practise (Princeton: Princeton Univ. Press)
Marzke, R. O., da Costa, L. N., Pellegrini, P. S., Willmer, C. N. A., & Geller, M. J. 1998, ApJ, 503, 617
Meyer, M. J., et al. 2003a, in preparation
———. 2003b, in preparation
Norberg, P., et al. 2002, MNRAS, 336, 907
Pei, Y. C., Fall, S. M., & Hauser, M. G. 1999, ApJ, 522, 604
Penton, S. V., Shull, J. M., & Stocke, J. T. 2000, ApJ, 544, 150
Péroux, C., Irwin, M. J., McMahon, R. G., & Storrie-Lombardi, L. J. 2001, Ap&SS, 277 (Suppl.), 551
Plionis, M., & Basilakos, S. 2002, MNRAS, 330, 399
Rao, S. M., & Turnshek, D. A. 2000, ApJS, 130, 1
Roberts, M. S., & Haynes, M. P. 1994, ARA&A, 32, 115
Rosenberg, J. L., & Schneider, S. E. 2000, ApJS, 130, 177
———. 2002, ApJ, 567, 247
Ryan-Weber, E., et al. 2002, AJ, 124, 1954
Ryder, S. D., et al. 2001, ApJ, 555, 232
Sandage, A., Tammann, G. A., & Yahil, A. 1979, ApJ, 232, 352 (STY79)
Schechter, P. 1976, ApJ, 203, 297
Schlegel, D., Davis, M., Summers, F., & Holtzman, J. A. 1994, ApJ, 427, 527
Schmidt, M. 1968, ApJ, 151, 393
Schneider, S. E., Spitzak, J. G., & Rosenberg, J. L. 1998, ApJ, 507, L9
Sheth, R. K., & Diaferio, A. 2001, MNRAS, 322, 901
Staveley-Smith, L., et al. 1996, Publ. Astron. Soc. Australia, 13, 243
Sprayberry, D., Impey, C. D., Irwin, M. J., & Bothun, G. D. 1997, ApJ, 482, 104
Storrie-Lombardi, L. J., McMahon, R. G., & Irwin, M. J. 1996, MNRAS, 283, L79
Storrie-Lombardi, L. J., & Wolfe, A. M. 2000, ApJ, 543, 552
Strauss, M. A., Ostriker, J. P., & Cen, R. 1998, ApJ, 494, 20
Tully, R. B., & Fisher, J. R. 1977, A&A, 54, 661
Verheijen, M. A. W., & Sancisi, R. 2001, A&A, 370, 765
Willick, J. A., Strauss, M. A., Dekel, A., & Kolatt, T. 1997, ApJ, 486, 629
Willmer, C. N. A. 1997, AJ, 114, 898
Zwaan, M. A., Briggs, F. H., & Sprayberry, D. 2001, MNRAS, 327, 1249
Zwaan, M. A., Briggs, F. H., Sprayberry, D., & Sorar, E. 1997, ApJ, 490, 173
Zwaan, M. A., Verheijen, M. A. W., & Briggs, F. H. 1999, Publ. Astron. Soc. Australia, 16, 100