

# Cosmic Variance in the HI Mass Function

Stephen E. Schneider

Department of Astronomy, University of Massachusetts, Amherst, MA 01003, USA  
email: schneider@astro.umass.edu

**Abstract.** I review the history of blind HI surveys and the evidence for cosmic variance in the HI mass function. Even the most sensitive current surveys are challenged by noise levels that make determining the shape of the mass function difficult, and this can lead to discrepant results. However, some differences in the mass function appear to be real environmental effects, and this suggests that there are locales where dark (or dim) HI-rich galaxies are more likely to be found. The requirements for significantly deeper surveys beyond the local supercluster to sample cosmic variance in a wider variety of environments are modeled and discussed.

**Keywords.** galaxies; mass function; stellar content; statistics

---

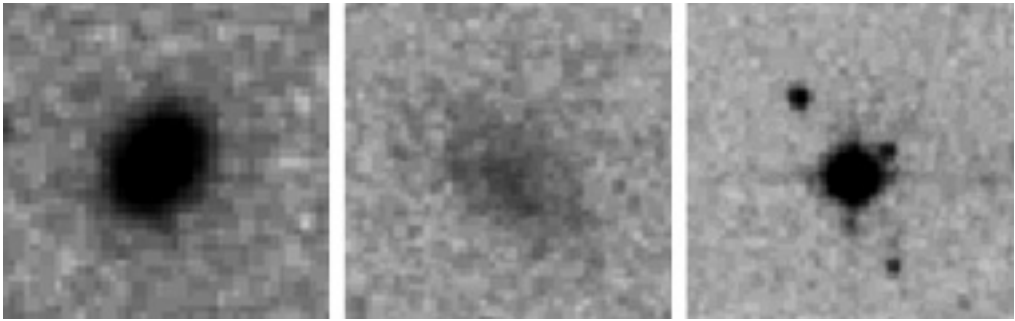
## 1. Introduction

Dark galaxies—at least those that contain baryonic matter—are expected to have hydrogen and helium and little other detectable matter. Searches for 21 cm emission from neutral hydrogen therefore appear to be a likely way to uncover such objects. “Blind” surveys for HI-emitting objects have demonstrated that optical surveys do in fact systematically miss large numbers of dim and nearby objects whose baryonic matter content is primarily gaseous. Some very dark objects have been found, although all, so far, appear to have at least a small amount of star formation present.

The HI mass function is a convenient shorthand way of describing the population of objects containing neutral hydrogen in the extragalactic universe. It can be characterized as a Schechter function, with an exponential turnover above masses of about  $10^{10} M_{\odot}$  and a power-law with a slope  $\alpha$  at lower masses. Early attempts to characterize the HI mass function based on extrapolations from low surface brightness (LSB) galaxies suggested that the mass function might be very “steep,” that is, rising to high counts for low mass objects. Some of these studies suggested that the slope might not be much less steep than the value  $\alpha = -2$  at which the integrated luminosity function becomes divergent (Impey *et al.* 1988).

However, the first attempts to quantitatively estimate the HI mass function implied a much shallower function. These studies were based largely upon extrapolations from the luminosity function of optically-discovered galaxies or targeted surveys of known groups. The power-law slope found in these studies was  $\alpha \approx -1.25$ , or perhaps even shallower ( $\alpha \approx -1$ ) in high-density environments (Briggs 1990; Briggs & Rao 1993). Such shallow slopes are similar to the estimated slope of the visible-wavelength luminosity function, and this would tend to suggest that there few or no dark galaxies.

In the subsequent decade and a half, a number of blind HI surveys of increasing size have been carried out to attempt to better characterize the HI content of extragalactic objects. However, these surveys have also shown significant disagreement in the HI mass functions derived from them. Nevertheless, it appears that there may be greater agreement between all of these surveys than is at first apparent. A number of seeming inconsistencies can be resolved if there is variance in the shape of the mass function



**Figure 1.** Three galaxies detected in the ADBS, each with approximately  $10^9$  solar masses of HI. Negative images of the galaxies from the digital sky survey are shown to the same scale.

dependent upon the local environment. Environments where the mass function is particularly steep then might represent the most likely places to uncover dark galaxies.

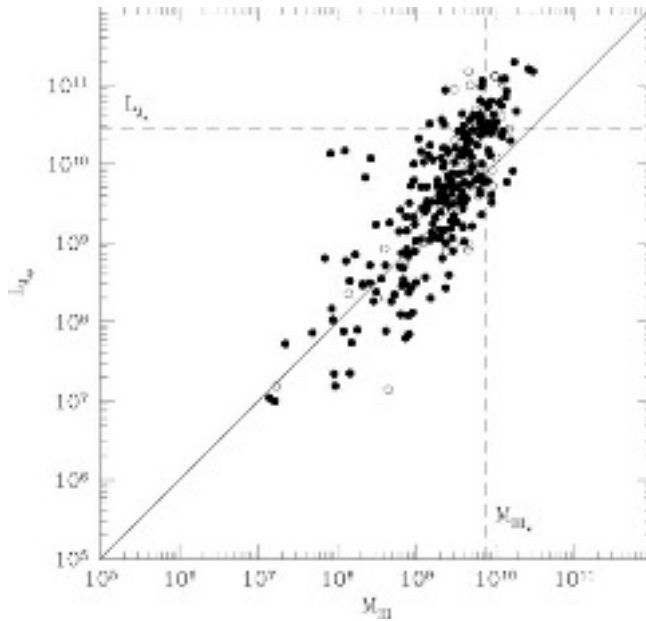
## 2. Gas versus Stellar Content

Before examining the details of the HI mass function, it is informative to look back at some of the properties of the population of extragalactic objects revealed by past blind HI surveys. I will focus on the results from the Arecibo Dual Beam Survey (ADBS, Rosenberg & Schneider 2000), because we collected nearly complete information about the stellar content of the sources detected for the galaxies detected in this HI survey.

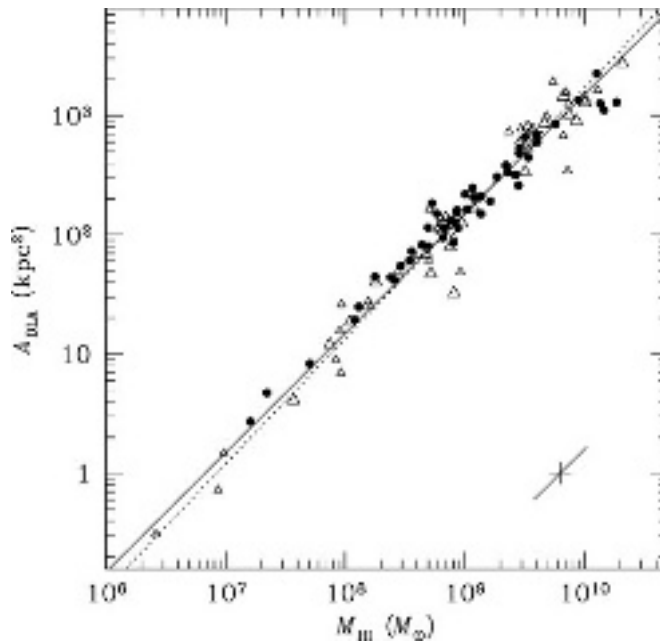
First, it is interesting to examine just how different are the HI and optical properties of these galaxies. In fact, sources of the same fairly high HI mass content look surprisingly dissimilar when viewed at visible wavelengths. In Figure 1, images of three galaxies from the ADBS are shown side-by-side with the same linear and brightness scales. The galaxies all have approximately  $10^9 M_{\odot}$  of neutral hydrogen, but clearly they have widely different stellar distributions. The first galaxy is a “normal” high surface brightness galaxy, while the second is a typical LSB that had not been cataloged previously. The third image shows that some overlooked HI sources are nearly point-like when viewed optically; such sources can be bright but omitted from optical catalogs because of their stellar appearance.

The differences between the stellar and gas content of galaxies can be shown quantitatively by plotting the J-band luminosity versus the HI-mass of HI-selected galaxies (Figure 2). Although there is a correlation, what is more significant is the enormous spread exhibited—exceeding two orders of magnitude at a given HI mass. With such a large variance, any flux-limited selection at visible wavelengths versus 21 cm will yield very different samples.

Finally, I would like to point out one further feature of the HI-selected sources that was commented upon by Michael Disney (these proceedings). We measured the cross-sectional area of the galaxies above a column density of  $2 \times 10^{20} \text{ cm}^{-2}$  and plotted them against the total HI masses of the galaxies. There is a surprisingly tight correlation across an enormous range of masses (Figure 3). This correlation is seen when HI measurements of optically-selected galaxies are included as well (Rosenberg & Schneider 2003). This is a puzzling correlation that seems to suggest a nearly uniform surface brightness of HI emission, at least for the highest density portions of galaxies’ interstellar medium.



**Figure 2.** Comparison of the J-band luminosity versus the HI mass of the ADBS and Arecibo Slice galaxies. The dashed lines indicate the “L-star” values (at the knee of the respective luminosity functions), while the 45-degree line indicates where the gas and stellar mass is roughly equal. (See Rosenberg *et al.* 2005, Figure 5, for further details)



**Figure 3.** The cross sectional area within which galaxies have a column density of neutral hydrogen greater than  $2 \times 10^{20} \text{ cm}^{-2}$  is plotted against the integrated HI mass of the galaxies. The plot is based on high-resolution synthesis observations of both HI- and optically-selected galaxies. (See Rosenberg & Schneider 2003 for further details.)

### 3. The Challenges of Blind Surveys

A major difficulty of carrying out blind HI surveys is that modern instruments remain relatively insensitive to the 21 cm line. Constructing large samples of HI-selected sources requires hundreds or thousands of hours with the largest radio telescopes. As a result, to avoid the problems of small-number statistics, we are driven to “scrape the bottom of the barrel” in order to extract as many sources as possible. Unfortunately, when we do this, it is much more challenging to determine the completeness of the samples.

Surveys to date have established that the completeness of HI surveys depends on the line width of the HI profiles, although there is not universal agreement about the best way to represent this functionally. It almost certainly depends on the shape of the profiles, although this is even more difficult to quantify. Baseline procedures may continue to hide wide-line sources even at higher signal-to-noise levels, and radio-frequency interference is a substantial, worsening problem.

The best method of establishing completeness levels is to introduce realistic fake sources throughout a survey, and processing those signals in the same way as real signals to determine the fraction that are recovered. This was carried out with the ADBS, which showed a roll-off in completeness below a nominal signal-to-noise level of about 10 (Rosenberg & Schneider 2002). I would argue that surveys that collect sources below this level should be assumed to be partially incomplete, or else they need to demonstrate their completeness through careful tests. The more conservative the estimate of completeness, the smaller the effective volume can be claimed to have been searched. Therefore surveys with more conservative claims for completeness have tended to derive steeper HI mass functions.

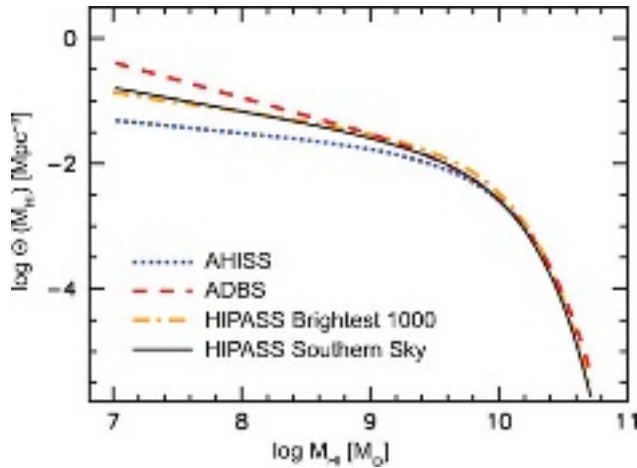
Finally, it is important to note that in attempting to identify low-mass sources, *all* surveys are biased toward the nearby volume of space. And because we live in a relatively high-density portion of the universe, this inevitably biases our estimates of the population of the lowest-mass sources to a volume of space that is atypically dense. This is an important point that I will return to in the final section.

### 4. Is There One HI Mass Function?

The first Arecibo blind HI surveys, the “Arecibo Slice” (Spitzak & Schneider 1998) and the “AHISS” (Zwaan *et al.* 1997) both strove for depth rather than large areal coverage. This strategy was partly based on expectations at the time that if there *were* sources missing from optical surveys, they likely had low column densities or had very low HI masses. These first surveys demonstrated that many more sources were detected than proposal referees and time allocation committees expected! And instead of being very low mass objects or extended clouds, the new HI sources had masses much like other galaxies at 21 cm, but were LSB or otherwise peculiar when observed optically.

For the same observing time on a given telescope, a larger number of sources can be detected by minimizing the time spent per point. The ADBS pushed for a larger sample by following this strategy, using drift scans with an effective integration time of about 12 seconds. As noted earlier, the ADBS inserted fake sources to understand its completeness better. In addition, each strip was observed twice in order to improve reliability. The survey sampled a large fraction of the sky visible to Arecibo and detected 265 objects, about 5 times more sources as the earlier Arecibo surveys, but in a similar amount of time.

The Arecibo surveys differed in observing strategy and analysis techniques, and they set the stage for the subsequent debate about the HI mass function. The AHISS found a shallow slope for the mass function ( $\alpha \approx -1.2$ ) shown as a dotted line in Figure 4, the ADBS found a relatively steep slope ( $\alpha \approx -1.5$  when excluding the region around



**Figure 4.** A few of the current estimates of the HI mass function are displayed. The ADBS has a Schechter-function fit with  $\alpha \approx -1.5$ , while the AHISS has  $\alpha \approx -1.2$ . Fits to the brightest 1000 galaxies and the full southern sky sample found by HIPASS yield the two intermediate curves.

the Virgo Cluster) shown as a dashed line in Figure 4. The Arecibo Slice showed a fairly shallow slope at high HI masses, but the (small) numbers of sources detected at low masses were in better agreement with the faint end of the mass function derived from the ADBS. While some of these differences were caused by small number statistics, I will argue below that much of the difference was caused by cosmic variance.

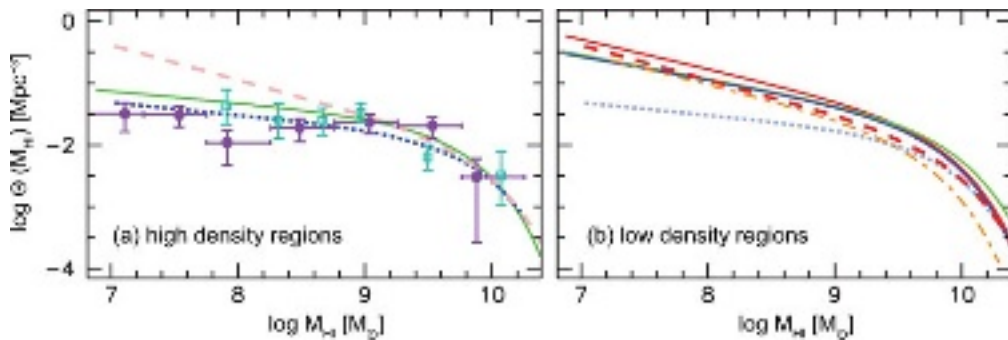
At about the same time as these Arecibo blind surveys were reporting their initial findings, a new 13-element multibeam feed array was being installed on the Parkes radiotelescope. This offered an opportunity to carry out a much larger survey called “HIPASS” (Barnes *et al.* 2001). HIPASS covered the entire sky visible from Parkes and ultimately detected about 20 times more sources than the ADBS. It is, however, more strongly biased toward the local universe because of its lower sensitivity.

Early results from Parkes multibeam receiver initially agreed with the higher Arecibo estimates of the slope of the mass function, yielding a value of  $\alpha \approx -1.5$  for galaxies detected in the zone of avoidance portion of the survey (e.g., Henning *et al.* 2000). However, analysis of the brightest galaxies in the southern sky yielded a lower value of  $\alpha \approx -1.3$  (Zwaan *et al.* 2003), although more recent reanalysis, including all galaxies detected in the southern sky, suggests a somewhat steeper slope of  $\alpha \approx -1.37$  (Zwaan *et al.* 2005). These curves are shown as a dot-dash and solid line in Figure 4, which overlap heavily.

The history of these estimates of the HI mass function might appear to be a process of homing in on the final value of the HI mass function, working from extremes toward a final value that lies somewhere in the middle. However, there are some problems with this description. Could the differences in these determinations be caused not by number statistics or analysis differences, but instead by intrinsic variations in the HI mass function?

## 5. Cosmic Variance in the HI Mass Function

It has been noted for many years that galaxies in high density regions, such as clusters of galaxies, have a deficit of HI. This does not however necessarily imply that the shape of the HI mass function is flatter. Removal of HI might cause the mass function to be



**Figure 5.** The difference in the HI mass function in high and low density regions is compared. The mass functions based on the ADBS and AHSS are shown in both panels for reference. (a) Open squares come from the ADBS in the vicinity of the Virgo cluster, while filled circles from the Ursa Major cluster. The solid line is the HIPASS fit to the HI mass function in the quadrant of the southern sky in which the Fornax cluster is found. (b) Fits from HIPASS for the three other quadrants of the southern sky are shown as solid curves. The fit from the southern ZoA HIPASS survey is shown as a dash-dot line.

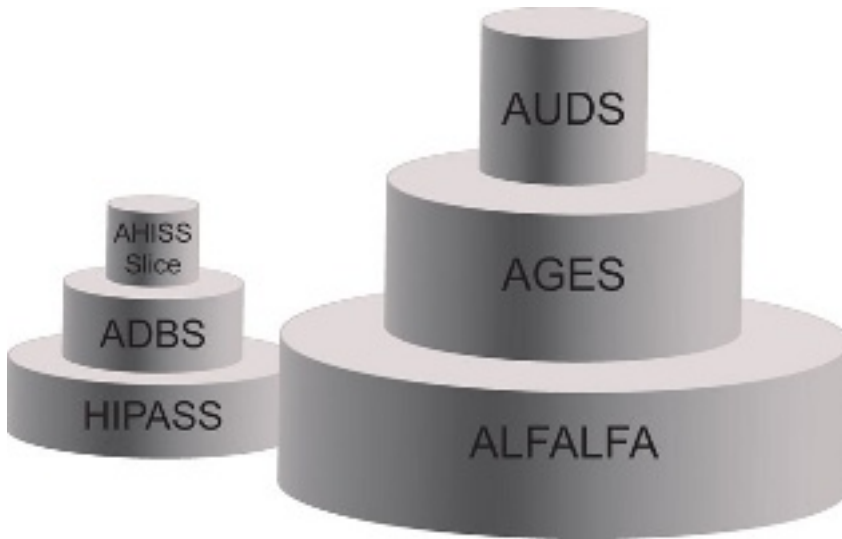
depressed at all masses, or indeed, the mass function might be even steeper in high-density regions, as was suggested recently by Zwaan *et al.* (2005). I will argue, though, that the results from blind HI surveys all point toward an HI mass function that grows steeper outside of dense regions.

The ADBS survey covered the Virgo cluster, but Virgo was excluded from the main analysis of the mass function so as to avoid biasing the results by such a prominent local structure. In fact, when the region around Virgo was analyzed separately, it showed a shallow mass function, more similar to the shallow slope of the AHSS (Figure 5). Other regions of high density, such as the Ursa Major cluster (Verheijen *et al.* 2001) also show a flat mass distribution. (The Ursa Major cluster points were shifted downward to adjust for the high overall density of galaxies in that region.)

Furthermore, HIPASS also made mass function fits to separate quadrants of the sky. The quadrant containing the nearby Fornax cluster matches the flat mass function seen for the other high-density regions seen in Figure 5. By contrast, the other three quadrants of the sky are in much better agreement with the steep mass function found in the ADBS, and the fit reported for the southern ZoA galaxies (Henning *et al.* 2000) also appears to be in better agreement with the steeper mass function.

These differences have several important consequences when considering future determinations of the HI mass function: (1) Since the shape of the mass function is variable, it is improper to use estimation techniques (such as stepwise maximum likelihood methods) that assume a uniform mass function shape everywhere; (2) Since the faint end slope is largely determined by nearby galaxies, it is important to recognize any local structures that may affect the results; and (3) In order to examine the faint end of the mass function in environments outside of the local supercluster, we need to carry out substantially deeper surveys than have been completed before now.

Point (2) is probably the explanation for the peculiar mass function derived by Schneider *et al.* (1998). The Arecibo Slice was deeper than most blind surveys that have been carried out to date, but it examined a small area in the direction of the Pisces-Perseus supercluster, with an intervening void. As a result, the higher masses were drawn from a population of galaxies in the cluster environment, while the lower masses were drawn from a small nearby volume. As a result the mass function for different mass ranges was



**Figure 6.** Schematic representation of the overall properties of various blind HI surveys, indicating their sensitivity by height and their areal coverage by the width of each “layer of the cake.” The past surveys discussed in this article are shown on the left, while the next generation of surveys, currently being carried out at Arecibo are shown at right. (See text.)

drawn from disjoint regions of space with significantly different HI properties, leading to an apparent jump in number counts at low masses.

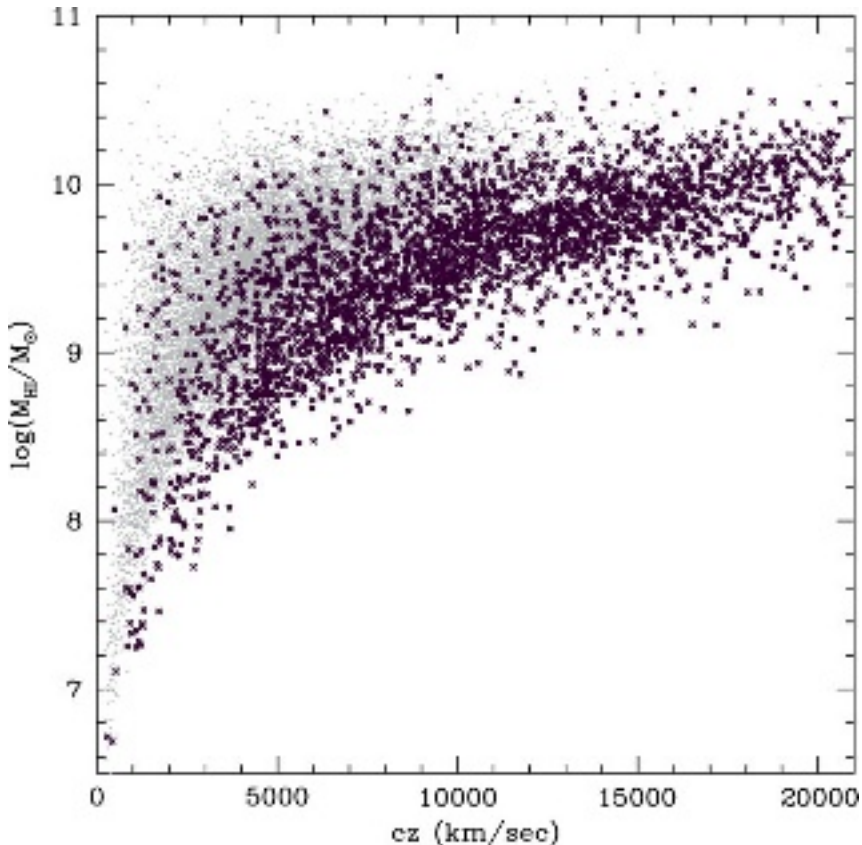
If the evidence points toward a steeper mass function in low-density regions, why then did Zwaan *et al.* (2005) reach the opposite conclusion? The answer appears to be in the method they used for determining the local density: they used nearest neighbors among the HI-selected galaxies themselves. As is shown very clearly by Luca Cortese in these proceedings, the HI-selected sources do not exhibit a higher density around a cluster. Even though they follow the general large-scale structure, HI-selected sources may even show a slight deficit at the cluster position. This suggests that the steepest mass function may be found in the boundaries of the high-density regions of the large-scale structure—where the density of HI sources has grown higher, but at densities still low enough that mergers and other destructive events have not reduced the population of low mass sources.

## 6. The Next Generation of Blind Surveys

The HI mass function exhibits variations that have begun to be explored by past surveys. A new generation of surveys is underway using the Arecibo L-band Feed Array (ALFA). There are several extragalactic surveys being carried out with this instrument, following the “wedding cake” strategy (Figure 6) of a shallow wide survey (Arecibo Legacy Fast ALFA survey, or ALFALFA), a more-sensitive, smaller-area survey (ALFA Galaxy Environments Survey, or AGES), and an even higher sensitivity survey covering a small area (Arecibo Ultra-Deep Survey, or AUDS).

The ALFALFA survey will recover the largest number of sources, but its sensitivity is not much deeper than the ADBS. Therefore, the deeper surveys will be critical for exploring the low-mass end of the HI mass function in environments beyond the local supercluster. This was pointed out by Zwaan *et al.* (2003), where they showed that despite the size and areal coverage of HIPASS or the ADBS, for low-mass galaxies, neither had





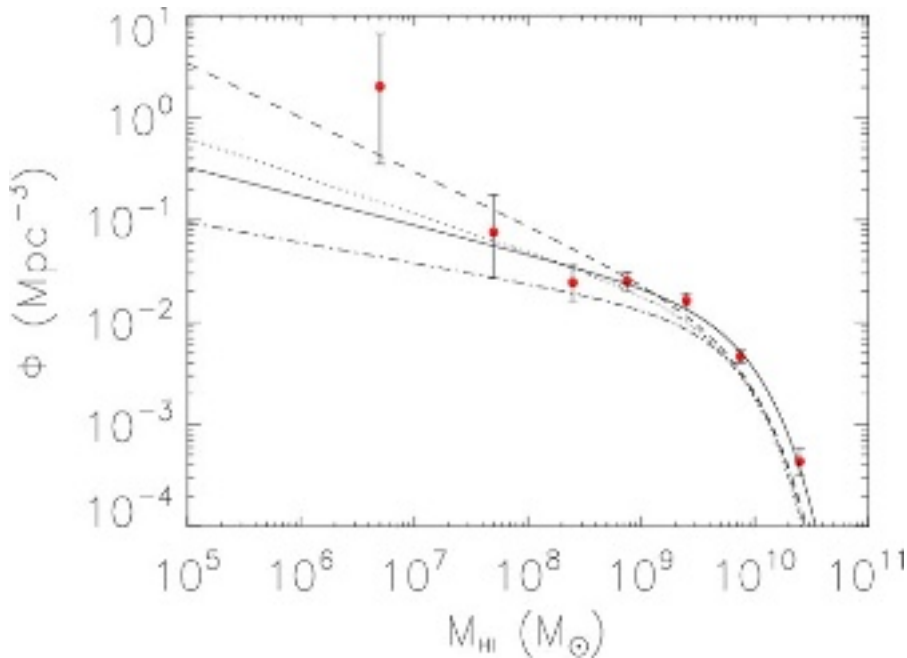
**Figure 7.** Simulation showing the masses and distances of sources detected by two surveys of different depth, but equal total observing time. The shallower (gray dots) survey detects more sources, but only detects low mass sources to a small distance. The deeper (dark x's) survey collects fewer sources, but detects lower-mass sources at all distances, permitting the exploration of environments not accessible to the shallower survey.

sampled as large a volume of space as the original AHISS and Arecibo Slice surveys. The AGES survey will in fact be deeper than either of those surveys, cover a much larger area, and ultimately sample a wide variety of directions.

Another way of illustrating the importance of the wedding cake strategy is shown in Figure 7. Here a simulation was used to estimate the masses and distances of sources that might be detected by surveys of roughly the same relative integration time as the AGES and ALFALFA surveys. The simulation assumed the HIPASS mass function, then distributed the galaxies with random orientations, and attempted to estimate which galaxies would be recovered given plausible completeness limits. Each of the simulated surveys used the same total integration time, so the shallower survey detected many more sources, but the complementarity of the deeper survey is that at all distances it can detect lower mass sources. AGES will in fact survey deeper than any of the past blind HI surveys, exploring a wider range of environments than was possible previously.

Initial results from the first area covered by AGES are shown in Figure 8, reproduced from a first draft of Robbie Auld's thesis (2007). As can be seen, the data points lie relatively low for high mass sources, but there is a hint of a higher number count at the lowest masses. This is very similar to what was seen in the Arecibo Slice, another deep,





**Figure 8.** Initial results for the HI mass function from the first fields observed by AGES (Auld 2007). The values found are reminiscent of the Arecibo Slice.

small area survey. It will be very interesting to see as AGES progresses whether this continues to be found. If it does, it will suggest that the HI mass functions determined from shallow surveys to date have been biased by the density of low mass sources in our local universe.

## References

- Auld, R. 2007, PhD thesis, University of Cardiff, Wales.  
 Barnes, D. G., *et al.* 2001, *MNRAS* 322, 486  
 Briggs, F. H. 1990, *AJ* 100, 999  
 Briggs, F. H. & Rao, S. 1993, *ApJ* 417, 494  
 Henning, P. A., *et al.* 2000, *AJ* 119, 2686  
 Impey, C. Bothun, G. & Malin, D. 1988, *ApJ* 330, 634  
 Rosenberg, J. L. & Schneider, S. E. 2000, *ApJS* 130, 177  
 Rosenberg, J. L. & Schneider, S. E. 2002, *ApJ* 567, 247  
 Rosenberg, J. L. & Schneider, S. E. 2003, *ApJ* 585, 256  
 Rosenberg, J. L., Schneider, S. E. & Posson-Brown, J. 2005, *AJ* 129, 1311  
 Schneider, S. E., Spitzak, J. G. & Rosenberg, J. L. 1998, *ApJL* 507, L9  
 Spitzak, J. G. & Schneider, S. E. 1998, *ApJS* 119, 159  
 Zwaan, M. A., Briggs, F. H., Sprayberry, D. & Sorar, E. 1997, *ApJ* 490, 173  
 Verheijen, M. A. W., Trentham, N., Tully, B. & Zwaan, M. 2001, *ASPC* 240, 507  
 Zwaan, M. A., *et al.* 2003, *AJ* 125, 2842  
 Zwaan, M. A., Meyer, M. J., Staveley-Smith, L. & Webster, R. L. 2005, *MNRAS* 359, 30