THE VIRGO CLUSTER THROUGH THE AGES

by

Rhys Taylor

A THESIS SUBMITTED TO CARDIFF UNIVERSITY
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

August 2010

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

You just keep on trying till you run out of cake
And the science gets done...

Glados - Portal

Acknowledgements

I would like to thank my parents, although I'm not entirely sure that I should. Their initial response of, "Looking for dark matter ? You can't even find your shoes" was not particularly encouraging, and rather more prescient than I suspect they intended. Despite this dubious beginning they nevertheless provided essential moral support throughout the whole process.

I must also thank my supervisor Jon Davies whose guidance was crucial in so many areas. Without his advice and encouragement much of this thesis would never have been written, especially the search for intracluster hydrogen (as he pointed out, the search for gravitational waves has generated even greater numbers of non-detections). His feedback on the early drafts led to many improvements both in the scientific content and writing style. Numerous other colleagues in Cardiff University deserve credit; special mentions must go to Luca Cortese, whose seemingly limitless knowledge puts Google to shame, and Robbie Auld, whose detailed knowledge of radio astronomy and patience to explain it to me were equally valuable.

I also thank the staff at Arecibo for their very welcoming approach and the much useful information I learned from them. Particular credit must go to Robert Minchin who taught me more than anyone else about observing and data reduction procedures, and whose automatic scripts developed for AGES have made the whole process so much easier. The other co-observers of AGES, namely Wim van Driel, Emmanuel Momjian Karen O'Neil, Stephen Schneider and the telescope operators also deserve credit.

There are innumerable other friends at Cardiff who also provided support in their own ways. Rory Smith deserves a special mention, as his frequent smoking breaks provided innumerable helpful discussions. The sacrifice of his lungs to the cause of astronomy is much appreciated. Similarly Tom Hughes and Matt Smith also helped with much interesting discourse, while sensibly managing to avoid chronic lung damage. Together with Gwenifer Raymond (whose alternative suggestion for my thesis title I shall refraim from printing), Sara Carver, Sarah Stickler/Harry, Ian Harry, Vicky Roloff and too many others to list, they all made the department an enjoyable and truly unique place, which I will greatly miss.

Finally I cannot fail to acknowledge the examination team of Rod Davies, Steve Eales and Peter Coles, who did their best to make me feel as relaxed as possible on the day. Their forgiveness for my lapse in basic numeracy is extremely appreciated, as are their helpful suggestions for future work.

Abstract

The results of the deep HI survey AGES are presented for two regions in the Virgo Cluster. Covering a total of 15 square degrees to a sensitivity of 0.6 mJy/beam, 73 objects are detected within the cluster, with 109 detections in the background. A new automatic extractor is described which gives greater completeness and reliability than previous methods. The majority of the cluster detections are associated with galaxies previously identified in the optical Virgo Cluster Catalogue, but nearly 30% of the cluster detections are new objects below the VCC completeness limit. Their optical and HI characteristics are described.

No definite optically dark galaxies are identified, however 4 intriguing candidates are reported. All have possible optical counterparts but these are extremely faint, and their HI velocity widths appear inconsistent for such objects when compared to the more certain associations. The likelyhood that these are really dark galaxies is discussed, but further observations are required for a more definitive analysis.

Cluster galaxies are found to be significantly HI deficient but it is not clear where their missing gas has gone. An automated algorithm is described to try to recover faint extended HI features, but no detections are made. The HI may be ionised by the intracluster medium or the sensitivity of the survey insufficient - the relative merits of these views are assessed.

A small fraction ($\sim$10%) of the early-type galaxies identified in the VCC are detected in HI. Evidence that some of these are morphologically evolving via gas loss, while others are recent additions to the cluster as yet unaffected by their new environment, is discussed. I also compare and contrast the two separate areas studied, describe the cluster's HI mass-function from AGES and other surveys, and discuss the overall influence of the cluster environment on galaxy properties.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

During the last century there have been many fundamental discoveries in astronomy. The Milky Way was shown to be just a single "island universe", and the "spiral nebulae" actually other galaxies, as large as our own and they themselves as numerous as the stars our galaxy contains. Moreover, the entire Universe was shown to be far from the static, eternal place it was once thought - instead it is seen as dynamic and ever-expanding, indeed with an expansion that appears to be accelerating. Many of these discoveries have relied on the small fraction of the electromagnetic spectrum visible to the human eye, optical light. But the previous century has also seen this restricted view of the Universe broaden immensely.

Radio observations of the sky first began in the 1930's with Karl Jansky's radio observations of the Milky Way. Since then, virtually the entirety of the EM spectrum is now observable, albeit to different levels of sensitivity and resolution. Such a vastly increased wavelength range has allowed discoveries that would once have been inconceivable - from the direct detection of pulsars to the inference of huge amounts of "dark matter", of a form still unknown. Some discoveries, such as radiation at the 21 cm wavelength, were predicted beforehand, but the consequences of such discoveries - dark matter a prime example - have been truly unexpected, and radically altered our view of the Universe.

It is the 21cm radiation that is the basis of this thesis. First detected in 1951 by Ewen and Purcell, in the same year Muller and Oort were able to use it to estimate galactic rotation. It was predicted by van der Hulst in 1945, who realised that such radiation could arise from electron energy level transitions within the neutral hydrogen atom. This transition relates to the alignment of the spin of the proton and electron within the atom. An atom with the proton and electron in parallel spin is in a higher energy state than if they were anti-parallel.

The transition from parallel alignment to anti-parallel alignment releases energy as a 21cm wavelength photon. This can occur spontaneously, but with very low probability, occurring only every few million years or so for a given atom. Collisions between hydrogen atoms allow the exchange of electrons. In this way, the atoms may change from the lower to higher energy state, or the reverse, or there may be no change of energy state. Collisional excitation or de-excitation is much more frequent, occurring on a timescale of a few hundred years rather than millions. The sheer number of neutral hydrogen atoms ensures that sufficient photons are emitted as to be relatively easily detectable to an observer on Earth.

Neutral atomic hydrogen, or HI, is generally much easier to detect than molecular hydrogen, H_2. Since only collisions between HI atoms are necessary to cause 21cm emission, it does not require very special conditions to be detected. It can be found in ordinary, quiescent galaxies, ongoing mergers, and, notably, entirely independently of any optical radiation. It can of course cool to become H_2, or become ionised into its constituent protons and electrons, but this requires more precise conditions, and both of these states are more technically challenging to detect.

Optical imaging typically employs broad-band filters, losing most of the frequency information of a detected source. The HI in a galaxy, however, emits over a much smaller range of frequencies, so a comparable broad-band filter would collect more noise than signal. Instead, heterodyne receivers are necessary. These collect photons over a wide range of frequencies but record the photons' frequency in many different bins or channels. An HI detection

is thus automatically also a redshift measurement, since the frequency measurement, and so recessional velocity, is inherent in the detection process.

This makes HI surveys a relatively easy way to detect the main gas component of a galaxy and measure its recessional velocity. The width of the detected line - the number of channels it spans - provides kinematic information on how fast the gas is rotating. However, while optical surveys have been performed for centuries, large-area HI surveys are a much more recent development. The problems are technical - it is now easy to build optical cameras with a wide field of few (of order a square degree) with CCD arrays of millions of pixels, but nothing comparable to this exists for radio wavelengths.

However, significant progress has been made. Multibeam receivers may only have of order 10 pixels, rather than 10 million, but this is still 10 times better than a single pixel. Thanks to the instruments installed on the Parkes 64m and the Lovell 76m dishes, almost the entire sky has now been observed at the 21cm wavelength, an unthinkable prospect with single pixel receivers. While optically-targeted - or pointed - HI observations are still often required, HI mapping (often called "blind" surveys, since they map entire areas and not just specific optical targets) is becoming more prevalent.

This thesis concerns the results of AGES, the Arecibo Galaxy Environment Survey, using observations of the Virgo Cluster. This is one of the most sensitive blind HI surveys to date, here exploring the nearest rich cluster. Details of the survey can be found in chapter 2. In this chapter, I discuss why studying the gas is interesting and the benefits of blind surveys as opposed to pointed observations. In particular, I discuss how studying the gas can be key to understanding the effects of environment on galaxy evolution. I also outline the theory behind "dark galaxies" and the missing satellite problem. The characteristics of the Virgo Cluster are summarised, and finally I outline the specific goals of this study.

1.2 Gas and Galaxy Formation

Star formation is not well understood. The very basic model is of gas gravitationally collapsing until nuclear fusion releases enough energy to halt further contraction. This is a view universally accepted, but within this there is little consensus on any of the details. For instance, does the gas have to be purely molecular or can it instead be atomic, or a mixture of the two ? If it has to be molecular, how does H_2 form from HI - is the presence of dust necessary ? What are the exact conditions necessary to trigger star formation, and are there circumstances in which the collapse of the gas can be prevented ?

The situation is infinitely worse on the scale of whole galaxy formation and evolution, for which even a "zeroth-order" model is far from universally accepted. There are fundamental disagreements on what galaxies are even made of. Dark matter is accepted not only as a reality, but the dominant constituent of galaxies by much of the astronomical community (e.g. Silk 2006). A few go further, purporting that a dark matter halo without stars can comprise a galaxy (e.g. Davies et al. 2006, Minchin et al. 2007). Yet what for many is fundamental is utterly rejected by others, who hold that dark matter does not exist - instead, it is our theories of gravitation which are wrong (Milgrom 1983).

In the most conventional view, a galaxy is a system of stars (and sometimes gas) bound together in a dark matter halo, with the components supported against gravitational collapse by their rotation and/or random velocity dispersion. Within this are a veritable plethora of controversies over fundamental points of galaxy formation and evolution. The standard model is so-called ΛCDM cosmology (Λ being a term to explain the acceleration of the expansion), in which cold dark matter (CDM) fragments into a very wide mass range of gravitationally-bound clumps. Baryons, initially gas, fall into these halos and form stars, and these proto-galaxies merge over time to form the galaxies seen in today's Universe.

Theories of star formation and galaxy evolution are thus intimately related, with understanding the behaviour of the gas playing a crucial role in both. The major success of the standard model has been to accurately reproduce the type of large-scale structures observable - filaments and voids consisting of galaxies, clusters and superclusters of galaxies. Its downfall lies in its predictions at the much smaller scale of individual galaxies, where the baryonic (i.e. gas) physics becomes far more important.

Before outlining some of the problems that occur at these scales, it is important to note that numerical simulations of the formation of large-scale structure, in ΛCDM cosmology, deal with the dark matter and baryons quite separately, if the baryons are included at all. It is believed that dark matter particles can be modelled as collisionless particles whose only interaction is gravitational. This makes them computationally very cheap, with the number of particles as high as 10^{10} in the case of the Millennium Simulation (Springel et al. 2005).

Baryonic particles, which are known to experience complex physical processes, are far more computationally expensive and cannot be simulated in such numbers. As mentioned above, much of the physics of star formation is uncertain. A common approach is to consider a particular (much smaller) section of the dark matter simulation and re-simulate this with baryonic particles included. Another is the "semi-analytic" method, where the baryons are not simulated, but are modelled analytically using observationally-derived relations to the dark matter. The point is that given the computational constraints and physical uncertainties, disagreements from observations should perhaps be expected.

1.2.1 The Missing Satellites Problem

Nonetheless, some of the discrepancies between the predictions of the standard model and observational reality are fundamental. One of the key issues is the "missing satellite" problem. The standard model predicts about 10 times more low-mass dark matter halos (i.e. dwarf galaxies) than are actually observed. The exact difference depends on the current state of the art : observations have discovered more faint dwarf galaxies, but increasing the resolution of simulations also predicts yet more satellites.

Key to understanding the extent of the missing satellite problem were numerical studies by Klypin et al. 1999 and Moore et al. 1999. At that time simulations were predicting 300-500 dwarf galaxies in a Local Group environment, but only about 40 were known. The model of Moore et al., when adapted to simulate a Virgo Cluster environment, actually reproduced the observed mass distribution very well. Some process appeared to be suppressing substructure in galactic halos, but not in clusters. The authors went so far as to state that "Either the hierarchical model is fundamentally wrong, or the substructure lumps are present in the galactic halo and contain too few baryons to be observed." More advanced, higher-resolution n-body simulations have not helped, with the *Via Lactea* simulation producing about 50,000 subhalos in a halo the mass of the Milky Way (Kuhlen et al. 2008).

The most important observational development in this area was the discovery of 12 new satellites of the Milky Way by the ongoing Sloan Digital Sky Survey, with another 8 or so objects in the Local Group as a whole (Simon & Geha 2007, for the SDSS see chapter 3). This still leaves the total far short of the predicted hundreds, but other factors must be accounted for : the SDSS does not survey the whole sky (part of which is obscured by the Milky Way), and sensitivity effects prevent a 100% satellite detection efficiency. Moreover, the distribution of the galaxies is not uniform across the sky, being better described by a 3-dimensional disc surrounding the Milky Way (Kroupa et al. 2005). Accounting for these factors to estimate the true number of satellites is not straightforward.

Consequently there is not yet a clear consensus on whether the new discoveries are sufficient to solve the problem or not. Models by Tollerud et al. 2008 suggest that they are, with most of the 500-1,000 undetected subhalos simply being too faint to detect (but not beyond the capabilities of planned future surveys). In contrast Koposov et al. 2009 find that it is necessary to invoke "squelching" - the quenching of star formation in low-mass halos by heating in the epoch of reionisation - for agreement with observational data. Similarly Simon and Geha, 2007, have to invoke the same process, noting that otherwise there is still a factor 4 difference between the number of predicted and observed dwarf galaxies.

As is so many areas of astronomy, the field is highly contentious. As well as external heating, internal "feedback" from stars, supernovae and active galactic nuclei (AGN) has been proposed as a solution - such processes could quench star formation by removing the gas. Like heating during reionisation, the physics of none of these processes (on galactic scales) is at all well understood. Reliance on uncertain theoretical models to explain observations is clearly unsatisfactory.

1.2.2 Dark Galaxies

Understanding of the gas physics is by no means perfect either, but 21cm astronomy does at least offer the prospect of an observationally testable solution to the missing satellites problem. Theory and observations suggest that HI can exist without forming stars. Some of the dark halos might be optically dark, but detectable by their 21cm emission. These are so-called "dark galaxies", which hereafter are defined as gravitationally bound dark matter halos containing some baryonic matter, but not in the form of stars ("dark halos" will be defined as gravitationally bound dark matter halos devoid of any baryonic matter).

The theoretical basis for dark galaxies comes from models of star formation, which suggest there is a gas density threshold that must be exceeded for star formation to occur. Essentially, this is because the gas must be of sufficient density in order to cool sufficiently for collapse. If the density is too low, the gas may become ionised due to background radiation. The exact details are non-trivial, relating to gas dynamics, intensity of UV background radiation, and metallicity and dust content (Schaye 2004). As an approximate figure, Schaye gives a star formation column density threshold of $\approx 3 \times 10^{20}$ atoms cm^{-2} for the Milky Way. He notes, however, that in low surface brightness galaxies star formation is seen to occur where the *average* column density is below this, most likely due to local density peaks (see also Minchin 2001).

Observational evidence for a star formation threshold is implied from observations of neutral gas that has no associated optical emission. Examples of such features include High Velocity Clouds (HVCs) of the Milky Way and other galaxies (Westmeier et al. 2008) and streams of gas removed tidally or otherwise from other galaxies (Koopmann et al. 2008). Even within galaxies HI is not always associated with the optical component, the HI usually extending beyond the stellar disc by a factor $\sim$1.7 (Broelis & Rhee 1997). In some cases the difference can be much greater - Cannon, Salzer & Rosenburg 2009 describe a dwarf galaxy with the HI extending 44 optical scale lengths beyond the stellar component.

If only to demonstrate the complexity of the issue, it is useful to note that the very opposite case can also be found. Truncated HI discs, where the HI radius is less than the optical, are discussed in section 1.3.3 and are probably an effect of environment. More problematic may be extended UV discs, which apparently indicate that young stars can be found well beyond the optical disc, sometimes in regions below the predicted star formation threshold (Thilker et al. 2007). One possible explanation is provided by Roškar et al. 2008, who present numerical models in which stars migrate significant distances across the galactic disc radially. This could have important implications for estimating the star formation threshold.

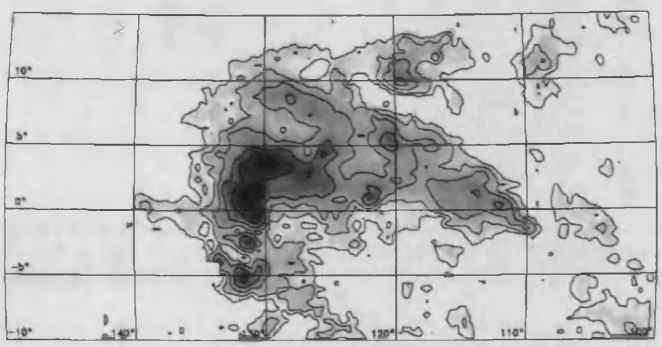
This issue aside, it is known and undisputed that HI can be found on occasion without significant associated star formation. However, claims for *bona fide* dark galaxies detected by 21cm emission are infinitely more controversial. It is important to note that such objects will by their nature be difficult to identify. Firstly, resolution in HI surveys is usually much less than at optical wavelengths, simply because the wavelength is so much longer. This means that associating the HI to a unique optical counterpart is very difficult. Optical spectroscopy is also required to determine if an optical counterpart is at a similar redshift to the HI detection. Higher spatial resolution HI measurements may still then be necessary to determine if a true association exists.

Secondly, supposing that a dark cloud of hydrogen is identified, proving that is a genuine dark galaxy is even harder. Again this necessitates high-resolution HI observations to accurately determine the scale of the disc, in order to estimate its dark matter content. This assumption would rest on assuming the HI to be gravitationally bound, which is not necessarily the case for tidal debris. Unless the object is very isolated any possible tidal (or other interaction) formation scenarios must be ruled out. A tidally formed dark galaxy would certainly be interesting, but would not have the same consequences for the missing satellites problem as a primordial object.

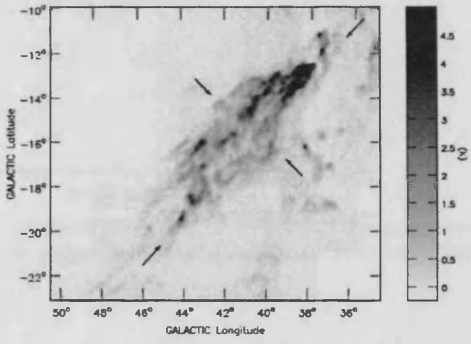
Nonetheless, several interesting potential candidate dark galaxies have been reported. Most of the dwarf galaxies of the Local Group have not been detected in HI, with some constrained by observation to upper limits of just a few solar masses of HI, e.g. Grcevich & Putman 2009. These authors find that the chance of a detection relates to distance from the nearest large galaxy (i.e. the Milky Way or Andromeda), with those closer than $\sim$270 Kpc likely to be undetected. Yet HVCs are now known to exist within this radius of the Milky Way and there have been suggestions that at least some of these may be dark galaxies (e.g. Blitz et al. 1999, Giovanelli et al. 2010).

Giovanelli et al. 2009 note two problems with this idea. Firstly, if the clouds have the properties needed to explain the missing satellites problem, then they should be detectable in more distant galactic systems, but they are not. Secondly their structural properties are inconsistent with the simulated missing satellites (i.e. underconcentrated

and too large). However, they identify a population of HVCs from the ALFALFA HI survey (see chapter 2) data that would both be undetectable outside the Local Group and fit the requirements of simulations, so the possibility remains open.



(a) HI map of complex H in Galactic coordinates from the Leiden/Dwingeloo Survey of Galactic neutral hydrogen. The image shows the integrated intensity for velocities between -230 and -150 km s^{-1}. The contours are spaced logarithmically from column densities of 4.6×10^{18} to 1.5×10^{20}. The telescope has a beam size of 36' (figure and caption from Simon et al. 2006).



(b) GBT HI image of the Smith Cloud at 100 km/s showing the cometary morphology suggesting the Cloud is moving toward the plane and interacting with the Galactic ISM (figure and caption from Nichols & Bland-Hawthorn 2009.

Figure 1.1: HI images of HVC dark galaxy candidates.

A couple of individual HVCs are worth mentioning as possible candidates, shown in figure 1.1. Complex H is a massive HVC ($\sim2\times10^7$ $M_\odot$ of HI) which, if gravitationally bound, would be dark matter dominated, and is apparently devoid of stars (Simon et al. 2006). Against this, its velocity width is much less than would be expected for a galaxy this size, so alternative interpretations such as cold gas accretion onto the Milky Way cannot yet be ruled out (see figure 1.1 left panel). Another candidate, the Smith Cloud, appears to have survived passage through the Galactic disc, a process which would require dark matter to avoid destruction (Nichols & Bland-Hawthorn 2009). Again alternative explanations, such as tidal stripping from an infalling dwarf galaxy, cannot yet be eliminated (see figure 1.1 right panel).

Perhaps the most celebrated and most controversial candidate is not in the Local Group but the Virgo Cluster (see section 1.4) - the HI feature known as VIRGOHI21. This was discovered by Davies et al. 2004 using Jodrell Bank, and subsequently observed at Arecibo (Minchin et al. 2005, Haynes, Giovanelli & Kent 2007) and with much higher resolution mapping at Westerbork (Minchin et al. 2007). The feature is a 25' long HI tail extending north from the one-armed spiral NGC 4254, ending in the VIRGOHI21 feature itself (see figure 1.2, reproduced from Minchin et al. 2007). Minchin et al. used HST observations to constrain optical emission in the I-band to an upper limit of about 31 mag $arcsec^{-2}$ at the position of VIRGOHI21 - it would be quite impossible to detect optically.

The extremely sharp change in velocity of the stream led Minchin et al. to suggest that VIRGOHI21 could be the HI disc of a dark galaxy. If gravitationally bound its total mass would be $\sim10^{10}$ $M_\odot$ - massive enough that its interaction with NGC 4254 could explain the prominent spiral arm. Though quite HI massive at $\sim10^7$ $M_\odot$, its column density is sufficiently low (3×10^{19}) to explain why it remains dark. Of course, since an interaction is occurring with NGC 4254, it is necessary to prove that VIRGOHI21 is not the result of this interaction, rather than the cause of it.

Haynes, Giovanelli & Kent discovered that far from being at the end of the HI stream, VIRGOHI21 is actually embedded in the middle of a much larger feature, with a projected length of about 250 Kpc (figure 1.3). These authors propose that the stream results from galaxy harassment - a high-velocity interaction(s) with one or more unidentified companions. Duc & Bournaud 2008 attempted to numerically model such an event, but it is unclear if such a process can reproduce the sharp break in velocity of the gas.

There are other aspects to the VIRGOHI21 complex which mean it can only be described as weird. UV emission has been detected in GALEX data extending southwest from NGC 4254 (Thilker et al. 2007), and, perhaps merely

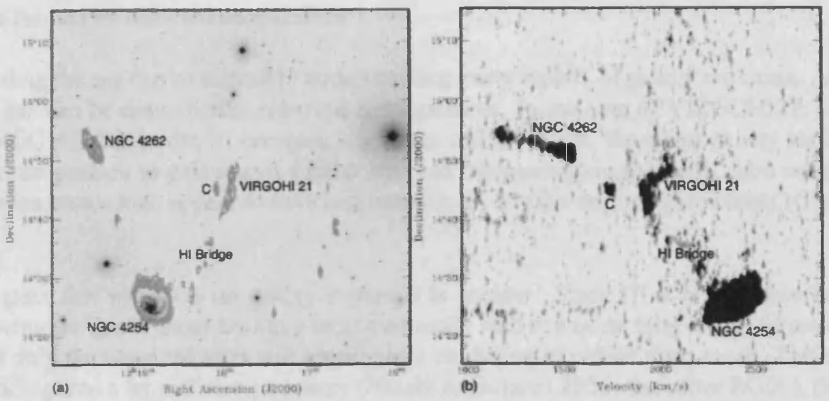


Figure 1.2: (a) HI contour map of the 21 cm observations, superimposed on a 1 deg^2 negative Digitised Sky Survey image. Contours are from 2.5×10^{19} to 2×10^{20} cm^{-2} at intervals of 2.5×10^{19} cm^{-2}. (b) Declination-velocity projection of the data cube (figure and caption from Minchin et al. 2007).

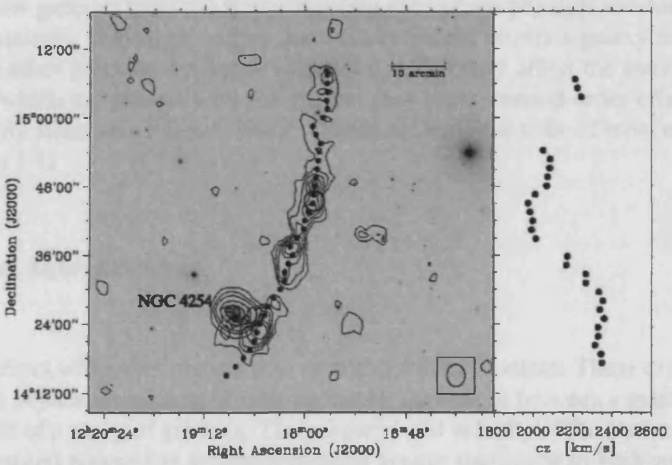


Figure 1.3: Left: HI flux contours extracted from the ALFALFA survey dataset, which mapped the full field represented in the image, superposed on a DSS2 Blue image. The 36 filled dots indicate the locations of beam centers for the successive LBW observations. The 3' circle in midstream indicates the position of Virgo HI21 reported by Minchin et al. (2005). The ellipse on the bottom right indicates the size of the Arecibo beam. Right: The velocity of the HI emission peak as seen in LBW pointings. Some of the LBW spectra yielded poor baselines or poor peak definition (figure and caption from Haynes, Giovanelli & Kent 2007).

coincidence, NGC 4262 is an early-type galaxy detected in HI. Additionally, Kantharia, Pramesh Rao & Sirothia 2007 find evidence that ram pressure stripping is occurring in NGC 4254. It seems we may still be far from a complete understanding of this very complex system. What is clear is that HI surveys can uncover features that are wholly unexpected, challenging the accepted paradigms of galaxy evolution.

1.3 Gas and Galaxy Evolution

In the hierarchical merging framework, galaxy formation and evolution are difficult to distinguish, with merging an integral part of the ongoing evolution even of contemporary quiescent galaxies. Since the standard model still has fundamental problems reproducing the correct number of observed contemporary galaxies, it is unsurprising that its description of the behaviour of galaxies (with regards to environment) is no less mired in controversy and uncertainty.

For example, is galaxy morphology evolution a significant process occurring in most galaxies, or a phenomenon peculiar to only a few specific cases ? Do certain environments elicit greater changes than others ? Can individual aspects - such as colour - be modified without much altering other properties ? What processes alter the gas content - are these entirely different depending on environment ? Most fundamentally, is it nature (where a galaxy is born) or nurture (the subsequent processes it experiences) that determines galaxy morphology ? Can galaxies of similar

morphologies be formed by different mechanisms ?

Again understanding the gas can be critical to understanding many aspects of galaxy evolution. As described above it is known that gas can be dramatically removed from galaxies. In the case of VIRGOHI21, only a few percent of the mass of NGC 4254 is in the HI complex - in terms of HI content, the spiral galaxy itself has been barely affected. Yet, in comparison to galaxies of similar size and morphology in the field, there are galaxies in cluster (and other) environments which appear to have lost virtually all of their original gas content (Giovanelli & Haynes 1985).

Exactly what impact this will have on galaxy evolution is unclear. Since HI is accepted to be the fuel for star formation, removing the gas without resupply must eventually lead to a quenching of star formation. As the young blue stars die off only the older red stars will remain, so a reddening of colour must result. This may well result in a spiral galaxy fading into a lenticular-type galaxy (Boselli & Gavazzi 2006 - hereafter BG06), though the gas-loss mechanism itself is debated (Dressler 2004). Both of these papers also question how significant such a transformation can be - structural differences between lenticular (or S0) and spiral galaxies suggest that, at the very least, not all S0s can have formed by simple gas loss.

To try and understand how galaxies evolve it is worth noting the various physical mechanisms that may act upon them in different environments. Broadly speaking there are two main effects a galaxy may experience : gravitational interactions (with other galaxies and larger structures) which may affect the stars, gas, or both, and "gastrophysical" processes, which act primarily on the gas but may cause second-order effects on the stars as well. These processes are briefly summarised below. Most of these are believed to be of most relevance within a cluster environment (see section 1.4).

1.3.1 Gravitational Interactions

Tidal encounters

A galaxy may tidally interact with other objects with or without direct contact. These can cause a wide variety of effects, depending on the objects interacting. A tidal encounter may occur between a galaxy and another galaxy or with the gravitational field of a group of galaxies. The geometry and velocity of the encounter can greatly influence the end result, with low-speed encounters generally causing greater damage than high-speed encounters (Toomre & Toomre 1972). This is because during a low-speed encounter, the galaxy is subjected to the same gravitational forces for a longer period of time.

Toomre & Toomre, and many others since, have explored how tidal encounters can disrupt the stellar disc, causing bridges and tails. The VIRGOHI21 complex is perhaps an example of disruption to the gas disc without creating any corresponding stellar tail (Duc & Bournaud 2007). Tidal encounters may even remove enough material that it can recondense into a new galaxy (e.g. Bournaud 2010). Besides removing material, they may also be essential in forming "grand-design" spiral structure (Bottema 2003, Smith 2009) and inducing star formation (e.g. Woods, Geller & Barton 2006).

Harassment

Harassment is really only a term to describe multiple, high-speed close tidal encounters, usually within a cluster environment (Moore et al. 1996). Although high-speed encounters are by nature less disturbing than low-speed encounters, they occur frequently (about 1 per Gyr in a cluster environment - Moore et al.), being the most common type of tidal interaction within a cluster owing to the high velocity dispersion. It was proposed as a mechanism to explain disturbed spirals without any obvious interacting companions, and may cause morphological evolution from spirals to ellipticals/lenticulars.

Mergers

While one galaxy may pass through another during a tidal encounter, only if its relative speed is less than the escape velocity will it be captured and the two actually merge. Mergers are divided into two types : major mergers, in which the merging galaxies are of comparable mass, and minor mergers, in which the masses are very different.

Each plays different roles in galaxy evolution. Major mergers have been proposed to explain the formation of elliptical galaxies (e.g. Van der Wel et al. 2009), while minor mergers may cause bulge formation in lower-mass galaxies as well as disc thickening (Hopkins et al. 2009, Moster et al. 2010).

Mergers can of course have important consequences for the gas content of the interactors, but so-called "dry mergers" can occur in which only gravitational forces are at work. These are mergers in which none of the galaxies have a significant quantity of gas, and may be very significant for the formation of elliptical galaxies (Khochfar & Burkert 2003). Both wet and dry mergers are intrinsic to the standard model, and establishing the typical merger rate is still a difficult process (Stewart et al. 2009).

1.3.2 Interactions Affecting the Gas

Ram Pressure Stripping

The intragalactic medium is never entirely empty. The motion of the galaxy causes the intragalactic gas to exert a force on the intergalactic medium. If this force is greater than the gravitational restoring force of the galaxy, its gas will be removed. Various factors dictate how significant this effect will be, including the mass of the galaxy, the density and temperature of the intragalactic medium, and the velocity of the galaxy through the medium. There is little doubt that this process does occur (BG06) but like all environmental effects, its significance is debated.

Vollmer et al. 2001 find that the gas is stripped into long streams and the star formation rate decreases, and that ram pressure cannot account for morphological change. In complete contrast Kapferer et al. 2009 find that there is a huge (a factor 10) increase in overall star formation, with most of this being in the streams rather than the galaxy. The newly-formed stars can be recaptured by the galaxy and form a bulge-like structure. Smith 2009 does not find any such "star-forming wakes", but does suggest that strong morphological transformation may result, turning dwarf irregulars into dwarf spheroidals. In contrast to all these, Dressler (2004) believes that ram pressure stripping is only one of many processes acting to remove gas, describing its role as "possibly redundant."

Accretion

In a cluster environment with hot, X-ray emitting gas, ram pressure stripping and other gas-depleting processes are generally agreed to be significant. However in more tranquil environments it is possible that gas infall (from cooling of the intragalactic medium) may still be occurring. Gas infall into dark halos is an intrinsic feature of the standard model, and HVCs of the Milky Way are perhaps examples of gas accretion within our own galaxy (Grcevich & Putman 2009). Models suggest that the evolution of the gas is highly dependent upon environment. Tonnesen, Bryan & Van Gorkum 2007 depict cluster galaxies as experiencing net accretion beyond $\sim$2.4 Mpc (radially from the cluster center), with those within experiencing a net gas loss.

The infalling gas itself may originate from a combination of sources - some of it primordial but some of it also stripped material re-accreting into the disc (Larson, Tinsely & Caldwell 1980). Dwarf galaxies may also supply larger spirals with HI, though by themselves the Local Group dwarfs would be insufficient to explain the current star formation rate of the Milky Way (Grcevich & Putman 2009). Accretion might also be able to rejuvenate early-type galaxies, with fresh gas infall leading to new star formation (Cortese & Hughes 2009, Kannappan et al. 2009).

Strangulation

Whatever gas reservoir supplies the accreting gas is likely to be more extended and diffuse than the galactic HI discs themselves. Gas further out in the gravitational potential of the galaxy is obviously more weakly bound to it, and therefore more easily removed. Strangulation - or starvation - is the removal of this gas (Bekki, Couch & Shioya 2002). Once star formation has consumed the HI remaining in the disc it then ceases altogether, with no gas resupply available. In clusters the diffuse gas could be removed by ram pressure stripping. Since this gas is further out in the galaxy's potential well than gas in the disc, it could be stripped at greater distances from the cluster center than the disc gas, and could not be re-accreted (again Bekki, Couch & Shioya 2002).

The strangulation mechanism was proposed to explain the existence of S0s by Larson, Tinsely & Caldwell 1980. This would be a rather more gradual change than direct stripping of the disc but with the same final result. It would

explain why S0s are found in clusters where ram pressure stripping is thought to be less effective (Dressler 2004, cites Virgo). The difficulty is proving that the necessary gas reservoirs exist (Dressler 2004, Benson et al. 2000).

Viscous Stripping and Thermal Evaporation

Both of these processes relate to the interface between the cold, dense galactic gas (HI) and the hot, tenuous intragalactic medium. Viscous stripping is a fluid effect that occurs due to the motion of the galaxy through the medium, with the difference in viscosities of the gasses causing drag (Nulsen 1982). Thermal evaporation (Cowie & Songaila 1977) occurs by the heating of the cold galactic gas by the hot medium; it is not directly dependent on the motion of the galaxy. The exact effects of these processes are difficult to quantify, though viscous stripping is expected to produce similar results to ram pressure stripping (BG06).

1.3.3 Observations

The manner in which many of the above environmental effects operate can only be explored in detail through simulations. Even when a galaxy is clearly observed to be experiencing a particular effect it may be difficult to predict the end result of the process (simulations are not so different, with different models of the same process predicting different results - for example see Agertz et al. 2007). However, this in no way negates the importance of observations. All simulations must ultimately concord with observations or they explain nothing. Observations act as the dominant constraint on models of galaxy evolution.

One of the key problems in theories of galaxy evolution, already mentioned, is whether or not morphology evolution is significant. That it can occur at all is well established with polar ring galaxies being a prime example. These are disc galaxies with stellar/gas rings orbiting (usually, but not necessarily - Brosch et al. 2010) over the poles of the disc. These are very rare and unusual systems - Dressler 2004 describes one example, NGC 4650A : "when a spheroidal system tries to recapture a disk, the result looks odd and forced - like a dog wearing shoes". In this instance it seems that an early-type and late-type galaxy are in the act of merging, producing an object that is a combination of both.

Their are further examples of galaxies that resemble a combination of early and late-type objects. Cortese & Hughes 2009 describe a sample based on colour and stellar mass. It is well known that in colour-magnitude space galaxies fall into two basic populations : a red sequence of mainly early-type galaxies, and a blue cloud of late-type galaxies (for an example see figure 1.5 but note that the blue sequence morphologies here are unusual - see section 1.3.4). Cortese & Hughes examined galaxies in the transition region between these two groups. Some of their sample objects (NGC 5701 in particular) show early-type central components with distinct, detached, late-type rings or discs in orbit, and some of these are also gas rich.

To understand the significance of individual examples for galaxy evolution as a whole, it is necessary to understand their context. The morphology-density relation (Dressler 1980) describes the relative abundance of early-type galaxies in clusters, and the relative prevalence of spirals in the field. Dressler found that while S0s and ellipticals constitute $\sim$90% of the population by number in dense clusters, in the field this decreases to only $\sim$10%, the remainder being spirals and irregulars. The question as to whether this reflects nature (early-types being predominantly formed in clusters) or nurture (early-types being more readily evolved from late-types in clusters) remains open.

Dressler (1980, 2004) concluded that nature is the dominant factor. In his view, a galaxy formed in a dense environment will tend to remain there, and it is the environment at and soon after formation that dominates : "Galaxy type is destiny." He acknowledges that environmental effects later in a galaxy's life may have some influence over it, but views this as of relatively little importance. This is because the morphology density relation is extremely weak - with a change in volume density of a factor $\sim$1,000, the fraction of spirals merely doubles. If current conditions were important, one would expect a much stronger relation. Instead, it seems that earlier conditions, when density variations were much less, that is the main factor in determining morphology.

Dressler (and others such as BG06) also notes that the bulge-to-disc ratio is different in S0s compared to spirals, with S0s having larger bulges. While the loss of spiral structure can apparently be explained by the cessation of star formation (e.g. Elmegreen et al. 2002), it is less obvious how the tightly bound inner bulges could be altered by gas loss. As described above, certain numerical models (such as Kapferer et al. 2009) may offer solutions to this.

But there are also purely observational reasons why, perhaps, the effect of environment on morphology evolution should not be neglected, and certainly instances that illustrate many of the environmental processes outlined above.

As mentioned earlier it has been known for many years that some galaxies have very little HI gas in comparison to galaxies of similar morphological type in the field. This relation was quantified by Giovanelli & Haynes 1985 as HI *deficiency*. This is defined as :

$$HI_{def} = log(MHI_{ref}) - log(MHI_{obs}) \tag{1.1}$$

Where HI_{def} is the HI deficiency, MHI_{ref} is the HI mass of the reference galaxy, and MHI_{obs} is the HI mass of the observed galaxy. A galaxy with a deficiency of 1.0 therefore has 10% of the gas of an equivalent field object, with a deficiency of 2.0 it would have 1%, etc. The expected HI mass can be calculated by a linear equation (Gavazzi et. al. 2008, using the parameters of Solanes et. al. 1996) :

$$log(MHI_{ref}) = a + b \times log(d) \tag{1.2}$$

Where a and b depend upon the morphological type and d is the optical diameter in Kpc. Late-type galaxies within clusters are known to be HI deficient, with deficiency perhaps the best indication of interaction with the cluster environment (BG06). They also show truncated HI discs in comparison to field galaxies, as much as 5 times smaller than the optical radius (Cayatte et al. 1990). There are also known to be more normal spirals without significant HI deficiency or truncated gas discs, which presumably are still on their initial descent into the cluster (Cayatte et al. 1994, BG06).

Cayette et al. 1990 suggested that these features are indicative of ram pressure stripping (which will more easily remove gas at the edge of the galactic disc where it is less tightly bound), particularly given that in some cases the HI is seen to have a sharp edge in the direction of M87 (which in Virgo corresponds to the cluster X-ray center, see section 1.4) and extended in the opposite direction. In Cayette et al. 1994 the authors expanded this interpretation. Some of the sample show HI deficiency but not a truncation of the gas disc, so some other process must also be acting to cause more even stripping across the whole disc, perhaps viscous stripping and/or thermal evaporation.

Cayette et al. 1994 also concluded that tidal interactions are unlikely to be the dominant method of HI removal, given that deficient galaxies have characteristics of ram pressure stripping. Tidal stripping can remove material from anywhere in the disc and so should not cause a radially truncated HI disc. Any plumes or streams of stripped material may be aligned in any direction relative to the galaxy, depending on the geometry of the interaction. The importance of ram pressure stripping is also suggested by the truncation of the Hα disc. This emission line traces hydrogen ionised to HII, usually by young ($\lesssim 4\times10^6$ years old), bright OB stars. This makes it a good tracer of recent star formation. Since ram pressure stripping causes a truncation of the HI disc, a quenching of star formation in the outskirts naturally follows (BG06). It is difficult to explain this by the more random nature of tidal stripping.

Ram pressure stripping has been shown by many different models to draw out long HI tails (Kapferer et al. 2009, Smith 2009, Vollmer et al. 2001). By the nature of ram pressure, these will have to be aligned only along the direction of motion of the galaxy (streams from tidal interactions could be aligned in any direction, depending on the geometry of the encounter). There is good evidence that such tails do exist, for example Chung et al. 2007 describes 7 galaxies with HI tails consistent with ram pressure or viscous stripping, up to 10' in length ($\sim$47 Kpc projected length).

Despite Cayatte et al.'s findings, tidal encounters can have dramatic effects on both the gas and stars. Kent et al. 2007, as part of the ALFALFA survey (see chapter 2) found an HI complex with a projected span of 170 Kpc and an HI mass of 5×10^8 $M_\odot$, which the authors attributed to tidal stripping from NGC 4445. An even more spectacular tail was discovered by Koopmann et al. 2008, who describe an HI stream 500 Kpc in projected length, associated with NGC 4532. This is the longest known HI stream within a cluster, and also the most distant from the cluster center, making it difficult to explain this feature by ram pressure stripping.

Whatever the eventual outcome of gas removal, tidal interactions can affect the stellar component directly. Kenney et al. 2008 discovered Hα filaments spanning 120 Kpc projected length, linking the very disturbed spiral NGC 4438 and the giant elliptical M86. The authors contend that a collision between the two would explain the unusual morphology of NGC 4438. Although Hα often traces young star formation, in this case it is more likely to be

due to heating of the cold gas stripped from NGC 4438. Unlike the previous examples, this indicates that neutral hydrogen removed from galaxies may become ionised.

The most impressive example of tidal interactions on the stellar component must surely be the discovery of ~100 Kpc long stellar streams around M87, discovered by Mihos et al. 2005. Their figure 1 is reproduced below as figure 1.4. The authors attribute the complex, very low surface brightness features to the hierarchical assembly of the cluster. While some features - such as the diffuse halo around M87 - are undoubtedly extragalactic, galactic cirrus must be a major caveat. The authors note that this central cluster region lies behind the center of a ring of cirrus, with the region itself apparently uncontaminated (but for the complexities of distinguishing cirrus from extragalactic features, see Cortese et al. 2010).

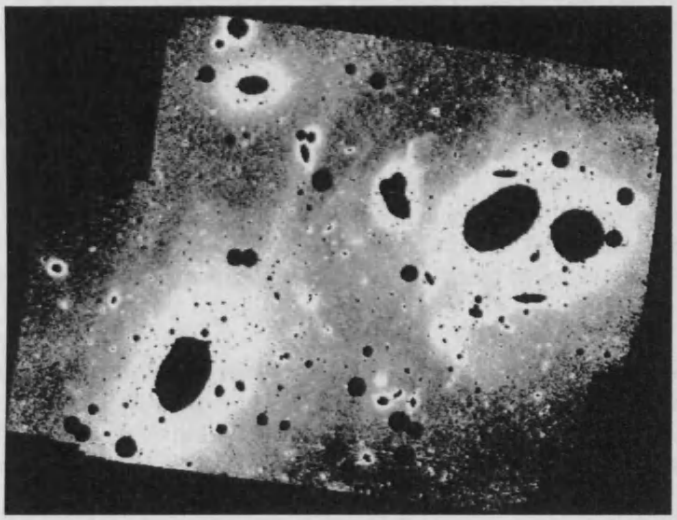


Figure 1.4: Diffuse light in the Virgo Cluster core. North is up; east is to the left. The white levels saturate at $\mu_V \sim 26.5$, while the faintest features visible have a surface brightness of $\mu_V \sim 28.5$. Figure and caption from Mihos et al. 2005. The image spans about $1.5 \times 1.5°$.

1.3.4 Morphological Transformation

It is clear that galaxies *can* be affected even at $z = 0$ by their environment, and in some cases very dramatically so. Gas removal is certain, but what is much harder to infer is what happens *after* the gas is removed. Galaxy-galaxy interactions are certainly capable of directly altering morphology (e.g. Kenney et al. 2008), but it is not yet clear if they can change a galaxy from one Hubble type to another, or if this effect is a mainstay of evolution. Nor is it certain if merely removing the gas, without directly influencing the stars, can cause any sort of significant morphological change.

It has been mentioned already that some numerical models of environmental effects, particularly ram pressure stripping, do predict morphological change. Observations suggest that these mechanisms are really at work - they are not an artificial construct - but do not agree on the relative significance of each type. It is at least clear that the effects of environment can not only create impressive structures but also remove very significant levels of gas. Many of the streams described above constitute only about 10% of the HI mass of the galaxies from which they seem to have been extracted (Kent et al. 2007, Koopman et al. 2008), but other galaxies without any (detected) extended HI features have far higher HI deficiencies, of order 2.0 (BG06).

The blue cloud and red sequence have already been mentioned. At the simplest level, galaxies in the blue cloud are late-type, star-forming and gas rich. Those in the red sequence are "red and dead" - they are early-type, show little or no evidence of star formation, and have very low gas content. Transition galaxies[1], which lie between the two populations, seem to consist mainly of gas-deficient spirals (Cortese & Hughes 2009, Hughes & Cortese 2009). The obvious inference is that late-type galaxies lose gas and evolve in colour and morphology onto the red

[1]This is far from an ideal term, since it implies that these objects are in a state of change, but the alternative of "green valley" is also unsatisfactory since galaxies do not appear green !

sequence.

Naturally there is a wealth of subtlety behind this apparently simple picture. The transition region is not populated exclusively by deficient spirals, as described in Cortese & Hughes 2009. Nor are the red sequence or blue cloud populated entirely by early or late type galaxies respectively. Fundamentally of course, a galaxy changing colour does not necessitate a change on morphology, or vice versa (for example Lisker, Grebel & Binggeli 2006 discovered dwarf elliptical galaxies with spiral structure). While individual galaxies may show good evidence for morphological change, suggestions that this is a widespread process perhaps inevitably rely on inference.

A key line of evidence is that relative populations of the blue cloud and red sequence are not constant with redshift. While the mass in the red sequence appears to have approximately doubled, that in the blue cloud has remained roughly constant since $z = 1.0$ (Bell et al. 2007). Given that most environmental effects tend to cause gas depletion and hence a quenching of star formation, growth in the red sequence is not unexpected.

More quantitatively, however, Bell et al. showed that the observed growth in the red sequence is in agreement with the observed star formation rate in the blue cloud. There is not enough star formation in the red sequence to account for its growth, so it appears that the blue galaxies must experience global star formation quenching and migrate to the red sequence. Bell et al. showed that only quenching is needed to account for this. They also state that most of the growth comes from massive ($\gtrsim 3\times10^{10}$ $M_\odot$) blue galaxies - if they did not have their star formation quenched then we would observe far more local, massive blue discs. To maintain the constant stellar mass in the blue sequence, the quenching of the massive discs must be balanced by intense star formation in smaller blue galaxies.

Strictly speaking of course these conclusions relate only to colour, not morphology. Hughes & Cortese 2009 found that most of the galaxies in the transition region are HI deficient, but this too does not necessarily imply morphological evolution, since different models do not agree on whether morphological change can result from gas removal. Other, quite different lines of evidence, however, more directly imply morphological transformation occurs.

Conselice et al. 2003 performed very sensitive HI observations on a sample of cluster dwarf ellipticals and detected about 10% in HI. In combination with other early-type detections from the literature, the authors argue that a significant fraction of the dE's in the Virgo Cluster are gas-depleted dwarf irregulars. Their detections have roughly the same gas content as dwarf irregulars, far higher than that of Local Group dwarf ellipticals. However, different interpretations of the same data are possible. Whereas Conselice et al. interpret clumpy substructure in one detection to indicate an early state of transition, it could be argued that this is morphological misclassification (see also chapter 5 for other examples of different interpretations).

Evidence for transformation of spirals into lenticulars is also controversial. There is a subtle but crucial difference between Dressler's position that "galaxy type is destiny" and simply saying that environment does not matter. Rather, Dressler holds that environment *does* create S0 galaxies from spirals (though the environmental mechanism(s) remains uncertain) but these will tend to remain in the cluster for their entire lives, hence the existence of a weak morphology density relation. Mechanisms aside, the population of S0s within clusters should grow with decreasing redshift if this is true. This is supported by Dressler et al. 1999, but Andreon 1998 blames morphological misclassification, claiming that there is no evidence for such a change.

While the formation of S0s from spirals may seem relatively likely, given that there are so many ways to remove gas, the opposite process may also occur. Cortese & Hughes 2009 describe examples where accretion may be at work in early-type galaxies. Kannappan, Guie & Baker 2009 cite examples where merging of spirals may create red sequence objects, and also examples of S0s building up a spiral disc through gas accretion. Their examples run completely counter to Dressler's "dog wearing shoes", simply appearing as blue S0s, with little to describe them as "odd and forced". Their figure 3, showing some of their examples, is reproduced in figure 1.5.

The effects of environment are clearly very controversial. There is at least consensus that environment is important - the morphology-density relation makes that difficult to dismiss - but what exactly environment actually does is unclear. Many of the effects of environment described above are believed to be most significant within galaxy clusters - some processes, such as thermal evaporation or harassment, cannot happen at all in the field. It is also apparent that the gas may have an important role in galaxy evolution, being essential for star formation and readily affected by environment.

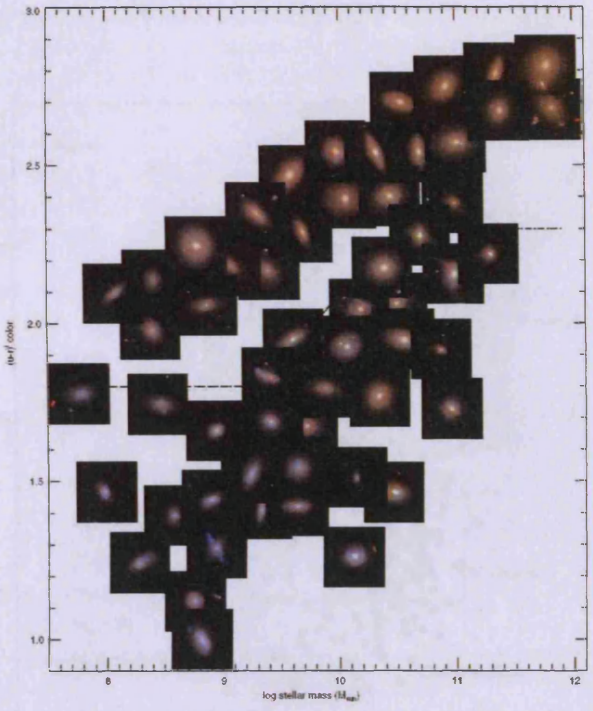


Figure 1.5: SDSS $ugriz$ color-composite cutouts of E/S0s from both sequences, arranged by u-r color and stellar mass.We include examples from both the HyperLeda+ and NFGS samples, using the u-r to U-R conversion given in Section 2.4. Images are slightly shifted to minimise overlap and sized arbitrarily. Figure and caption reproduced from Kannappan, Guie & Baker 2009.

As has been alluded to previously, the Virgo Cluster is of particular significance. Most of the individual examples described so far have been deliberately chosen as Virgo Cluster objects, as this is the primary topic of study of this thesis. To understand the importance of deep HI observations within this environment, the cluster as a whole is described below.

1.4 The Virgo Cluster

The historical importance of this particular environment stems from its proximity. At approximately 17 Mpc distance (Gavazzi et al. 1999) it is near enough to detect dwarf galaxies, far more common than giants but more difficult to detect owing to their lower luminosities. It is also significant due to its richness, with well in excess of 1,000 members in the infamous Virgo Cluster Catalogue (VCC) of Binggeli, Sandage & Tammann 1985 (BST85). Compare this to the Local Group : Virgo spans about 5 Mpc projected length in declination, while the Local Group is about 1 Mpc along its longest axis, yet there is roughly a factor 50 difference in the number of members.

The VCC has become the bible of Virgo Cluster studies, due to its large area ($\sim$140 deg^2) and completeness to faint objects (to an apparent photographic magnitude of about +18). Details on how the catalogue was compiled are outlined in chapter 3; here the results are described. The other essential source of information, also described more fully in chapter 3, is the Galaxy On-line Database Milano Network, GOLDMine. This multi-wavelength database contains information on every VCC galaxy, sometimes only the original VCC data but usually much more detailed, including optical and UV photometry and HI data.

The VCC galaxies using GOLDMine data are shown in figure 1.6 (courtesy Luca Cortese, private communication). The cluster does not have a simple structure but is composed of several different sub-clusters or clouds. For clarity only the main clouds listed in BST85 are shown, though further studies subdivide this structure further (Gavazzi et al. 1999). The main cluster, or subcluster A, is centered on the giant elliptical M87, and is at a distance of about 17 Mpc. The M and W clouds are believed to be at distances of about 32 Mpc. Of course the real boundaries of the clouds are not so well-defined as they are shown here - individual distance determinations are only available for a few objects (see chapter 3 and Gavazzi et al. 1999).

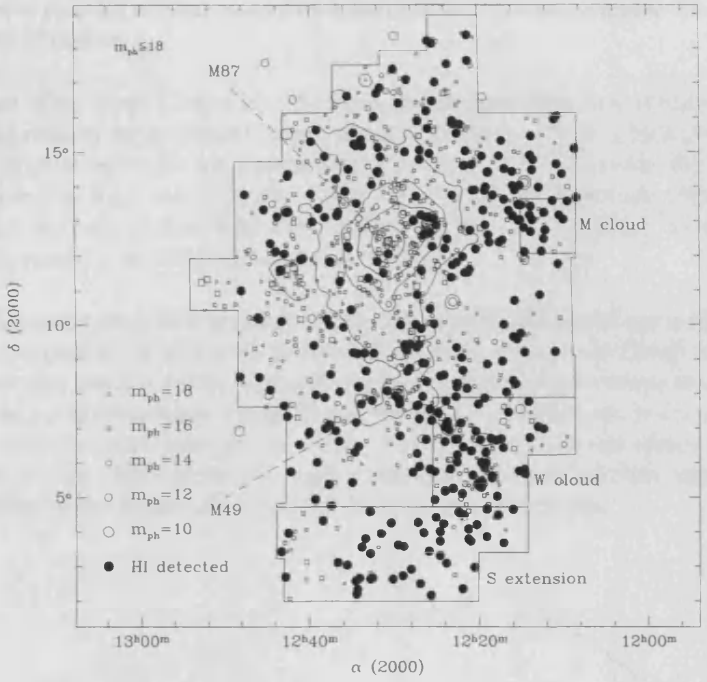


Figure 1.6: The major features of the Virgo Cluster. Circles are late-type galaxies (filled circles are HI detected) and squares early-types. Contours show the X-ray emitting gas from observations of the ROSAT satellite.

Fortunately redshift data is far easier to obtain than actual distance measurements, and this can give some clues to the large-scale structure of the cluster. It cannot be used as an accurate proxy for distance, since in the massive, dense and complex environment of the cluster velocity dispersion is very high. This gives rise to the classic "finger of God" effect seen in redshift surveys. In the field, where velocity dispersion is much lower, velocity traces distance, and a complex structure of filaments and voids is seen. At the intersection of these filaments are galaxy clusters with much higher dispersions, producing elongated "fingers". An example is shown in figure 1.7.

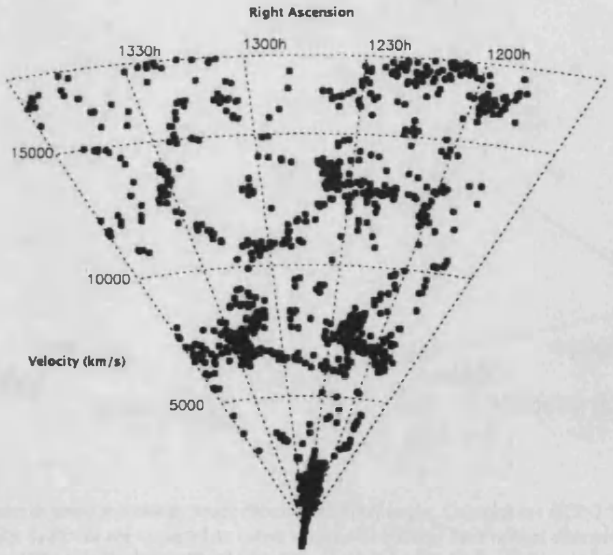


Figure 1.7: Wedge diagram to show the finger-of-god effect, where clusters are seen as elongated structures in redshift space. Reproduced from Giovanelli et al. 2007, showing the Virgo Cluster at $z \approx 0$.

BST85 listed a total of 2096 galaxies, of which 854 are spectroscopically confirmed as members in GOLDMine. An additional 904 have been classed as members or possible members but still lack redshift measurements, giving a total of 1758 cluster galaxies (the remainder have been confirmed by redshift observations as background objects). Throughout, those galaxies without redshifts are assumed to be members unless otherwise stated. Distances are those assigned in GOLDMine.

The velocity dispersion of the Virgo Cluster is ~767 km/s, the standard deviation of velocity measurements from GOLDMine. The total velocity range is much larger, about 4,000 km/s. There is little evidence for a foreground population (BST85), so galaxies which are blueshifted (by as much as 1,078 km/s in the case of VCC 594) have probably been accelerated by the cluster's potential well. While galaxies appear abundant up until ~3,000 km/s there is a dearth of galaxies beyond this (until about 3,500 km/s from GOLDMine), so this is often taken as the edge of the cluster (Giovanelli et al. 2007, Gavazzi et al. 1999).

Despite the velocity dispersion the infalling clouds can be seen in position-redshift space (figure 1.8). The complex structure of the cluster suggests it is still in the process of assembly, with whole clouds of galaxies still infalling. This makes it an interesting place to study, with galaxies having experienced various levels of environmental influence. Differences in the distribution of morphologies are readily evident both in the distribution of velocities (figure 1.9) and in position-velocity space (figure 1.10). The various subclusters appear more visible in the distribution of early-type objects. The distribution pretty much agrees with BG06, who suggested that some of the late-types near the center of the cluster (on the sky) are actually infalling clouds.

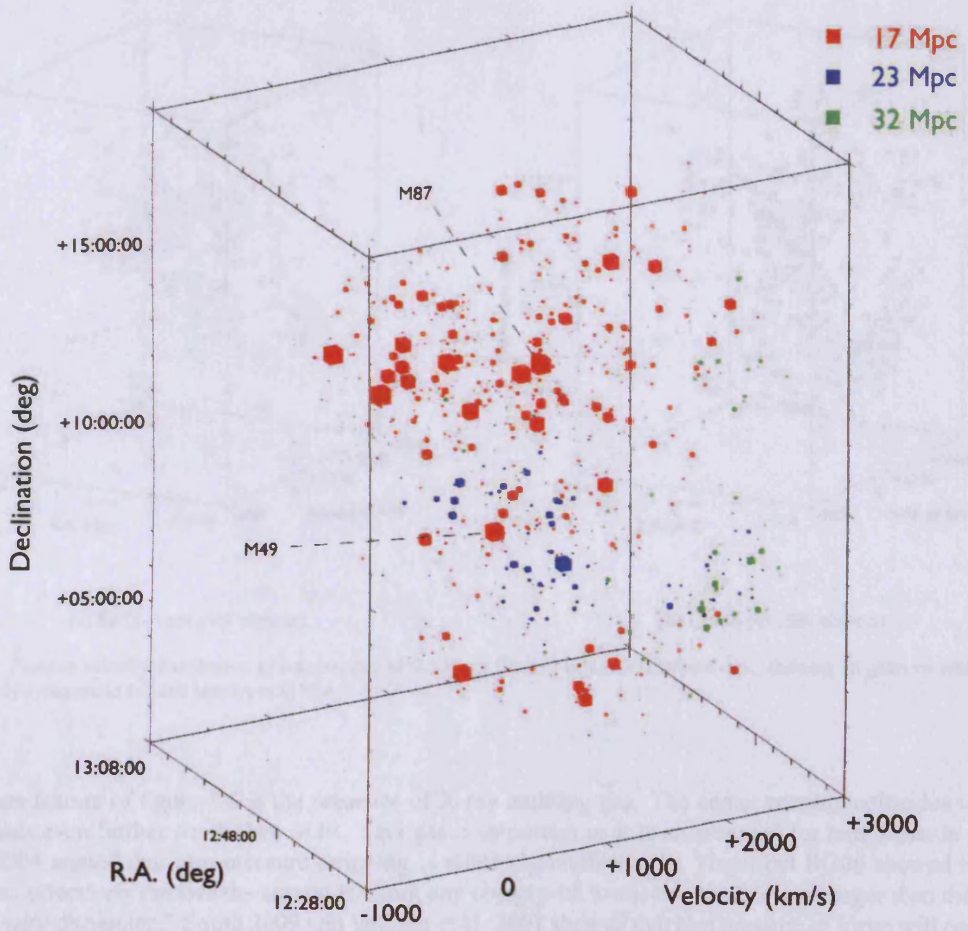
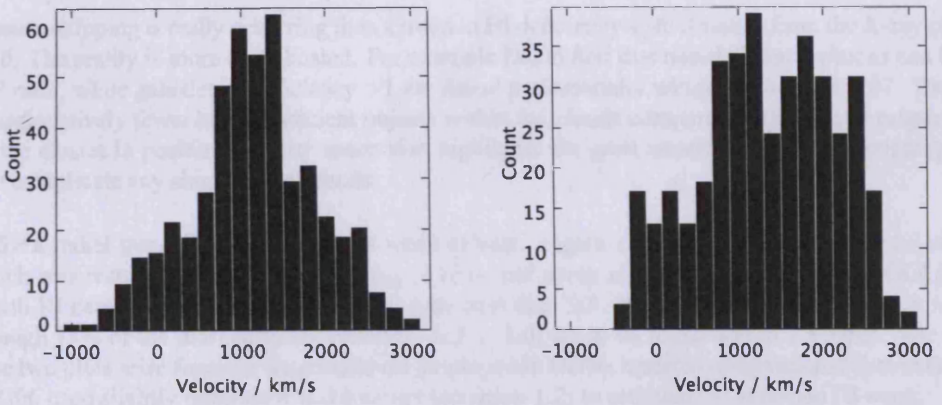
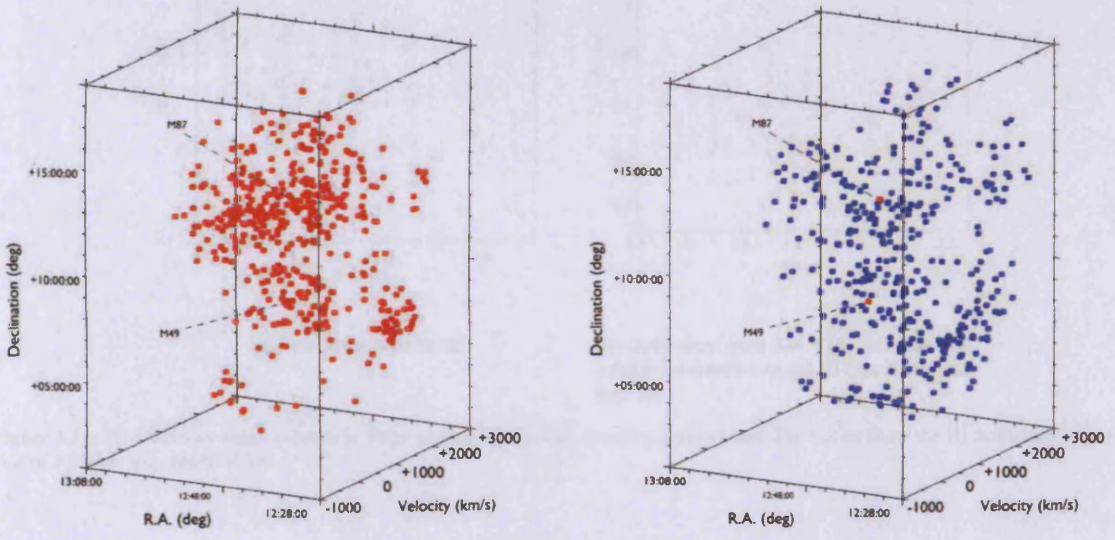


Figure 1.8: View of the Virgo Cluster in position-velocity space from an arbitrary angle. Galaxies are all 852 VCC objects listed in GOLDMine with a measured recessional velocity. Galaxies are rendered as cubes whose size reflects their optical diameter from the GOLDMine database, exaggerated by a factor 2 for clarity. Colour indicates assigned distance.



(a) Early-types (467 objects) **(b)** Late-types (387 objects)

Figure 1.9: Velocity distribution of components of the Virgo Cluster from GOLDMine data, showing all galaxies with measured redshifts.



(a) Early-types (467 objects) **(b)** Late-types (387 objects)

Figure 1.10: Position-velocity distribution of components of the Virgo Cluster, using GOLDMine data, showing all galaxies with measured redshifts. Early-types are in red and late-types in blue.

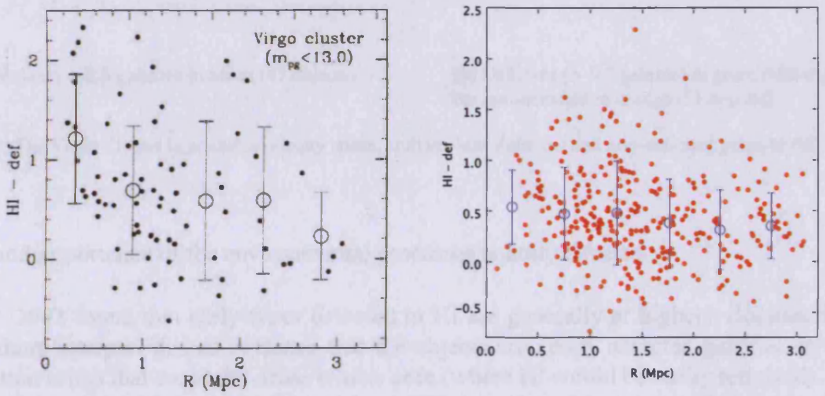
One obvious feature of figure 1.6 is the presence of X-ray emitting gas. The center roughly coincides with M87 but it extends even further south than M49. This gas is important as it is an essential for ram pressure stripping to occur. D04 argued that ram pressure stripping is relatively ineffective in Virgo, but BG06 showed that "ram pressure can effectively remove the atomic HI from any object with a velocity similar to or larger than the average cluster velocity dispersion." Smith 2009 and Vollmer et al. 2001 showed that ram pressure in Virgo will completely remove the HI from dwarf galaxies during their first pass through the cluster core.

The gas may be partly self-replenishing : HI removed by ram pressure stripping may fall into the cluster core where it will be evaporated by the X-ray component (Smith 2009). It may also be partly primordial, falling into clusters whilst their constituent galaxies were also forming, compressional heating having been shown to able to produce the observed temperatures (Sarazin 1988). Sarazin notes that observations of heavy elements mean that

not all of the gas can be primordial.

If ram pressure stripping is really occurring then a trend in HI deficiency with distance from the X-ray center might be expected. The reality is more complicated. For example BG06 find that non-deficient galaxies can be found in Virgo at all radii, while galaxies of deficiency >1 are found preferentially within 2.2 Mpc of M87. They also find that there are relatively fewer highly deficient objects within the clouds compared to the cluster proper. The complexity of the cluster in position-velocity space also highlights the great uncertainties when assigning distances, which may complicate any simple radial trends.

Evidence for a radial trend in HI deficiency is weak at best. Figure 1.11 shows the deficiency as estimated by BG06, which was restricted to galaxies with m_{pg} <13.0, and using all possible galaxies from GOLDMine (i.e. all those with HI detections and of morphological type later than S0). If there is a trend at all then it is extremely weak, although 75% of the most deficient galaxies ($def > 1.0$) are to be found within 1.5 Mpc[2]. The differences between the two plots arise from the magnitude cut employed in BG06, updated observational data in GOLDMine, and that BG06 used slightly different a and b values (equation 1.2) to estimate the expected HI mass.



(a) Deficiency from BG06

(b) Deficiency from 344 VCC galaxies with redshift measurements and HI detections, later than S0.

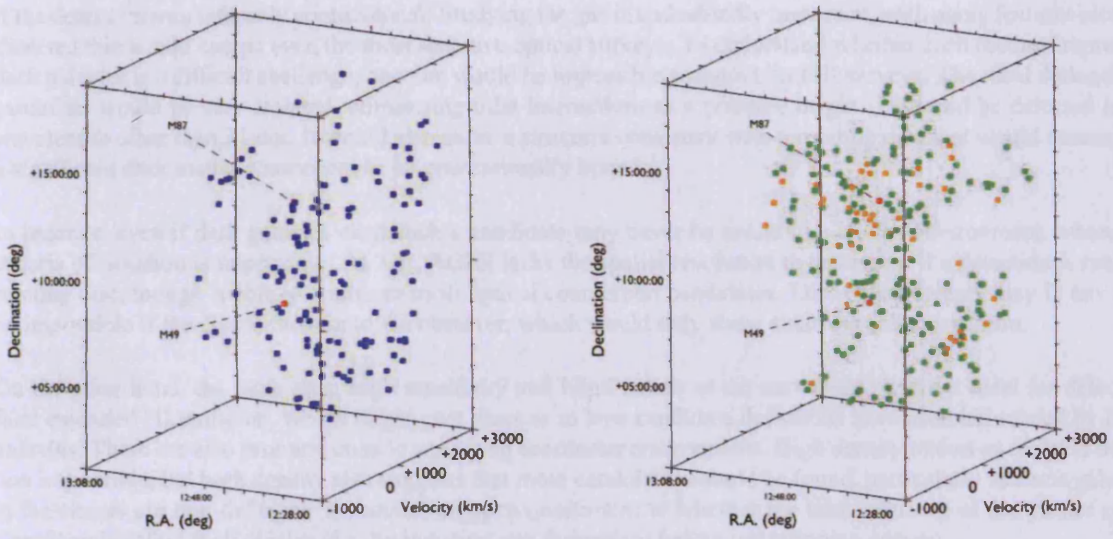
Figure 1.11: HI deficiency measurements in Virgo varying with radius from the X-ray center. The circles show the HI deficiency averaged in bins of 0.5 Mpc with errors of 1σ.

A trend is perhaps more visible in position-velocity space. Figure 1.12 shows that whereas the non-deficient galaxies are more or less uniformly distributed, the deficient galaxies better trace the structure of the clouds (intriguingly similar to the distribution of galaxies when split by morphology). This appears to be especially true of the most deficient galaxies of all, those that have been observed in HI but not detected (this is of course subject to varying sensitivity effects since they were not all observed to the same depth - additionally, 1,186 galaxies remain without any HI observations).

This may also be apparent from figure 1.6 via the general lack of HI detections around M87. This may partly be a real effect - BG06 show that there is a (very broad) correlation with HI deficiency and projected distance from the X-ray center - but a major caveat is that it is also likely to be partly a selection effect. M87 hosts an AGN and is a strong continuum source, which makes reliable observations within about 1 degree difficult (see chapter 2).

It is worth summarising the complex effects of the cluster on the HI. Deficiency varies very broadly with radius, but non-deficient and highly deficient galaxies are mixed. Galaxies are seen with extended HI features, and there are also some HI features whose association with particular galaxies is unclear. The most deficient galaxies generally show strongly truncated HI discs, but some are deficient but with relatively normal gas discs. Small wonder then,

[2]BG06 show there is a much more obvious trend for the Coma cluster, but here observations have been made out to 14 Mpc projected distance, rather than the 3 Mpc for Virgo



(a) Non-deficient (< 0.3 galaxies in blue (141 objects)

(b) Deficient (> 0.3 galaxies in green (185 objects), HI observed but non-detected in orange (51 objects)

Figure 1.12: The Virgo Cluster in position-velocity space, split to show deficient and non-deficient galaxies (all are late-types).

that the nature and importance of the environmental processes is hotly debated.

Conselice et al. 2003 found that early-types detected in HI are generally at higher velocities than those not detected. The authors interpret this as evidence that the objects are newly accreted galaxies, or possibly on high angular momentum orbits that avoid the dense cluster core (where HI would be easily removed). Similarly, Davies et al. 2004 found that, in a blind survey, the number of HI detections increased with velocity (the result broadly resembles that of figure 1.9, right panel, since late-types are usually HI detected and early-types are not). They interpret this to mean that the HI detections are largely of an infalling and unvirialised population.

There is a fundamental bias introduced by HI observations of optically selected targets. The optical survey may be complete to some luminosity, but this in no way means it will include every object to a given HI mass. As discussed earlier, there are some objects which may have extremely high HI mass-to-light ratios that would go entirely undetected in optical surveys. Thus, observations of a given sensitivity of an (optically targeted) HI survey will not necessarily detect every object actually present above the HI mass sensitivity.

The only way to remove this bias is through a fully sampled HI survey. This literally means fully sampling an area in HI, rather than integrating over individually targeted galaxies. The remainder of this thesis concerns the results of the Arecibo Galaxy Environment Survey in two areas in the Virgo Cluster. Technical details and goals relating to the specific capabilities of this survey are given in chapter 2. In the final section of this chapter, I summarise the current discussion and outline the goals of a deep, blind HI survey in the Virgo Cluster.

1.5 Summary, Goals and Thesis Outline

It is difficult to overstate the controversy that surrounds the study of galaxy evolution. At times it can seem that there is little or no consensus on even the most fundamental issues, with debate still raging (albeit amongst a minority) as to whether dark matter is essential to galaxy formation or even exists at all. The situation is little improved even working within the standard model. Although there are many proposed environmental effects, most of which have at least been demonstrated convincingly to really occur, there remains much discussion as to their relative importance, and, more fundamentally, their effects on galaxy evolution.

Yet there is some measure of agreement. Environment is accepted as an important factor in galaxy evolution, even if the details remain intensely controversial. Studying the gas is undoubtedly important, with many features already detected that would escape even the most sensitive optical surveys. To understand whether such features represent dark galaxies is a difficult challenge, one that would be impossible without blind HI surveys. The ideal dark galaxy candidate would be very isolated, eliminating tidal interactions as a possible origin. It should be detected in no wavelength other than 21 cm. It would also show a structure consistent with a rotating disc that would necessitate a significant dark matter component to be gravitationally bound.

In practice, even if dark galaxies exist such a candidate may never be found in a cluster environment, where the criteria of isolation is impossible. At 3.5', AGES lacks the spatial resolution to determine if a detection is really a rotating disc, though is able to eliminate most optical counterpart candidates. Disc determination may in any case be impossible if the disc is face-on to the observer, which would only show a narrow velocity profile.

On the other hand, the large area, high sensitivity and blind nature of the survey also make it ideal for detecting faint extended HI emission, which might give clues as to how candidate detections have been influenced by other galaxies. There are also pros and cons to surveying the cluster environment. High density makes an isolated detection impossible, but high density also suggests that more candidates should be found, particularly as some galaxies in the cluster are non-deficient. It remains an open question as to whether the tidal influence of the cluster could significantly affect such objects (i.e. by inducing star formation) before gas stripping occurs.

Extended HI features are interesting in their own right, giving direct evidence of the environmental mechanisms at work. Ram pressure and viscous stripping, as well as tidal encounters, are believed to be able to strip gas into long streams, with properties dependent on the particular mechanism. Thermal evaporation, in contrast, causes ionisation, so would not create such features. The fate of the stripped gas is also an open question - possibly becoming ionised, possibly forming stellar streams, but at least in some cases in Virgo it remains as HI.

The cluster environment at least offers the assurance that gas is being lost - it is difficult to see how gas could ever be accreted (except in the occasional merger) given all of the processes at work to remove it. This effectively removes the prospect of rejuvenating early-types through HI, leaving the focus on whether or not such objects can be formed by gas removal from late-type galaxies. Study of those early-types which do possess HI is obviously crucial to answering this question. Do they, for instance, really all lie on the cluster outskirts, as Conselice et al. 2003 claim, or is this a selection effect ? Do they possess relics of a previous morphology, or are they as red and dead as ordinary early-types ? Finally, are they really very much richer in gas than other early-types, or is this a sensitivity effect ?

As has been discussed, different components of the cluster are distributed differently. Objects detected by an HI survey but not in the VCC are likely to be optically faint but gas rich, so such a survey may detect new infalling components. This will give an optically unbiased view of how the cluster is assembling. It will also by its nature observe every galaxy within the target areas, in many cases observing early-type galaxies for the first time in HI. This should give a much better idea of how HI relates to star formation - in GOLDMine, while 61% of late-types have been observed and detected, just 8% of early-types have even been observed, and less than 2% are reported as detected (many of which are flagged as unreliable).

The remainder of the thesis is organised as follows. In chapter 2, the technical details of AGES such as the observing strategy and data reduction and analysis are described. The targeted fields and goals of the survey as a whole are outlined and some of the previous discoveries are summarised.

Chapter 3 outlines the other sources of data used, including other sources of HI data and observations from other wavelengths. This includes the VCC, GOLDMine, the Galaxy Evolution Explorer Satellite, the Sloan Digital Sky Survey and the Isaac Newton Telescope Wide Field Survey. The respective survey areas and sensitivity limits are described, as are methods of data analysis such as integrated aperture photometry.

Chapter 4 is concerned with methods of source extraction from HI surveys, which pose the same problems as optical surveys but with the added complication of a third dimension. Existing source extraction methods such as a visual inspection and the automatic programs Polyfind and Duchamp are described, and estimates made of their completeness and reliability. Taking cues from the merits and detriments of these methods, a new automatic program is described.

Chapter 2

The Arecibo Galaxy Environment Survey

2.1 Introduction

The Arecibo Galaxy Environment Survey is a blind, neutral hydrogen survey using the ALFA (Arecibo L-band Feed Array) instrument mounted on the Arecibo telescope. The combination of the 7-beam receiver with the world's largest single dish radio telescope (305m diameter) enables surveys of unprecedented sensitivity. Currently 4 HI surveys with ALFA are operating from Arecibo, including the Zone of Avoidance Survey (Henning et al. 2008). As HI is optically thin, this survey can see through the galactic plane to reveal galaxies that, to optical surveys, would be totally hidden by intervening dust.

The other 3 surveys form a "wedding cake" in terms of sensitivity and area. The first tier represents the shallowest but largest area survey. This is ALFALFA, the Arecibo Legacy Fast ALFA survey (Giovanelli et al. 2005). When complete, this will cover 7000 square degrees of the Arecibo sky to a depth of $\approx$2.2 mJy/beam rms noise - a huge increase over such predecessors as HIPASS, which had a sensitivity of $\approx$13 mJy/beam. AGES represents the second tier, being much smaller in area with a target of 200 square degrees, but significantly deeper at as low as 0.5 mJy/beam. The final tier is the smallest and deepest of all : AUDS, the ALFA Ultra Deep Survey, aims to cover less than 0.4 square degrees but to an extraordinary target depth of $\approx$0.05 mJy/beam (Freudling et al. 2008).

The surveys each have very different goals and motivations. There is little reason to suppose that distant galaxies in the galactic plane are any different to distant galaxies at higher latitudes. However, detection of nearby objects could significantly increase knowledge of the Local Group, a primary goal of the ZOA survey. The large area of ALFALFA will enable a statistically significant number of detections ($\sim$20,000) over a wide range of environments to a very high sensitivity compared to previous surveys. The extreme depth (but small area) of AUDS is designed to enable HI detections at a redshift $\sim$0.1, which is extremely challenging for HI observations and has never before been attempted with a blind survey.

This thesis is focused on the results of AGES, the middle tier of the wedding cake. While ALFALFA will provide excellent statistics across a large continuous area, by being more focused on selected, separate areas AGES can hope to detect fainter objects to greater distances, but still in a wide variety of environments. This chapter presents an overview of the survey as a whole, from its goals and motivations, the technical details of the survey strategy, data reduction and analysis techniques, to a review of some of the discoveries AGES has already made. It draws heavily from the AGES observing proposal, private communications with the AGES consortium, Auld et al. 2006 (hereafter A06) and Auld 2007.

2.2 Survey goals

Previous blind HI surveys have enormously increased our understanding of the gas content of the Universe. The principle advantage of a blind survey is that it is not biased by existing optical observations, increasing its discovery potential - pointed HI observations of galaxies could never hope to uncover an object such as VIRGOHI21 (Davies et al. 2004, Minchin et al. 2005 - see chapter 1). Until relatively recently, the limitations of single pixel receivers made HI surveys to any reasonable depth hugely impractical.

The development of the 13 pixel receiver mounted on the Parkes 64m, and the 4 beam instrument on the Lovell 76m dish, enabled for the first time HI observations of the whole sky through the HIPASS (Staveley-Smith et al. 2000)

and HIJASS (Kilborn 20002) surveys. Both of these have, however, suffered from poor sensitivity (13 mJy/beam is equivalent to 2×10^8 $M_\odot$ at the distance of Virgo). This limits their ability to detect dwarf galaxies except in the very local volume. Since dwarfs may be the most sensitive tracers of environmental influences, detecting such objects may be of vital importance in understanding galaxy evolution.

The dish size of course dictates angular resolution as well as sensitivity, and HIPASS and HIJASS have 15 and 12' beams respectively, Such large beams negate the possibility of resolving all but the very largest of HI sources, and makes the identification of unique optical counterparts very difficult. At 3.5', the far better angular resolution of Arecibo makes it much easier to pinpoint the optical counterpart. This becomes especially useful in a dense cluster environment when attempting to associate a weak HI detection to an object without an optical redshift measurement. An example is shown in figure 2.1, which compares the HIPASS and AGES beam sizes.

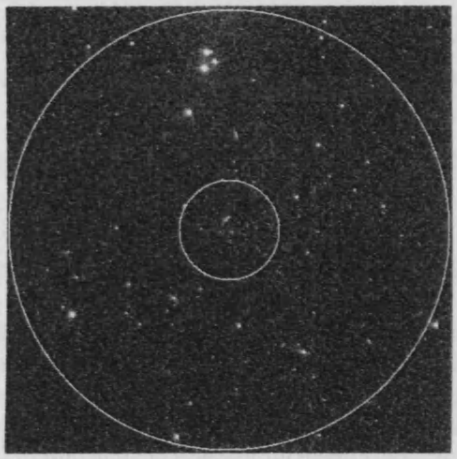


Figure 2.1: Object 66 from the VC1 area (see chapter 5) as seen in the SDSS g-band. The inner circle has diameter 3.5' (Arecibo) and includes only 1 possible optical counterpart. The outer circle has diameter 15.5' (Parkes) and encloses numerous optical candidates.

While the much greater dish size offers improved sensitivity and angular resolution, the ALFA receiver itself offers better velocity resolution than previous instruments. Whereas both HIPASS and HIJASS had a velocity resolution of 18 km/s, ALFA has 5 km/s. Standard practise is to Hanning smooth the data to 10 km/s - this gives greater sensitivity, ensures consistency with ALFALFA for comparison purposes, and also greatly reduces Gibbs ringing from the Milky Way (see section 2.5). This is still superior to the previous surveys (which with Hanning smooth have a resolution of 27 km/s), making AGES better suited to search for dwarf galaxies.

The bandwidth of AGES observations is significantly greater than that of HIPASS. Whereas HIPASS could reach 12,700 km/s (Barnes et al. 2001) AGES can reach 20,000 km/s. Although at this distance AGES is only sensitive to very massive galaxies (see next section), which are much rarer than dwarfs, it also observes a much greater volume - a total of $\sim$10,000 Mpc^3 for a 5 square degree field. This means that AGES is also well suited to detecting HI giants as well as dwarfs.

The larger telescope and increased instrumental performance enables AGES to potentially tackle many different issues. Many of these are extensions of previous studies for which AGES may give new insights. For example, previous determinations of the HI mass function by different methods (optically targeted and blind HI surveys) have given quite different results. This is illustrated by Kovac et al. 2005 figure 2, reproduced here in figure 2.2 - all of the examples here are from blind HI surveys of different areas of the sky. It is clear that there is considerable variation with environment. When complete AGES will have the results of a large area, deep survey spanning many different environments from which to construct HI mass functions, with galaxies detected over a very wide range of HI masses.

One of the important reasons in determining the faint-end slope of the HI mass functions is to test its agreement with cosmological models, which predict a steep slope for dark matter halos, a manifestation of the so-called "missing satellite" problem. As discussed further in chapter 1, previous HI surveys have discovered only a scant handful of candidate dark galaxies that might solve this problem. Modelling by Davies et al. 2006 shows that in terms of searching for dark galaxies, AGES is by far the most efficient survey to date, with 23% of its detections expected to be dark. Although the figure is highly dependent upon model parameters, the great sensitivity and

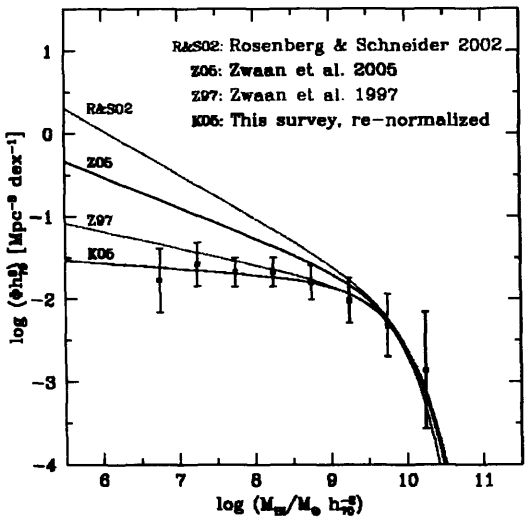


Figure 2.2: The HI mass function from different surveys, reproduced from Kovac et al. 2005.

velocity resolution undoubtedly place AGES at the forefront of the search for dark galaxies.

The identification of dark galaxies relies on the relatively high spatial resolution of AGES, as illustrated above. This is also important in mapping diffuse HI structures, such as tidal tails or HI streams. These are direct signatures of galaxy interactions, mergers, and the effects of ram-pressure stripping, potentially providing substantial information on how galaxy evolution proceeds. The 3.5' resolution of Arecibo means it has the capacity to directly resolve features such as VIRGOHI21 (although high resolution interferometric mapping remains important), a capacity which HIPASS almost totally lacked due to its much lower resolution.

Another consequence of Arecibo's greater resolution is the potential ability to map the low-column density extent of HI in massive galaxies. This will only be possible for the largest of galaxies, but again, is a capability lacked by HIPASS and most other previous HI surveys. It will not, of course, have the exquisite detail of a targeted interferometric survey such as THINGS (e.g. Walter 2008), but will cover a much larger total survey area. By studying how sharp the edge of the galactic HI discs are, AGES may have a role in determining how background radiation ionises HI discs (though surveys of higher spatial resolution are more suitable).

While its better resolution offers the prospect of resolving more sources than HIPASS, the majority of detections by AGES will still be unresolved. In these instances the resolution still enables a more reliable identification of optical counterparts (or lack thereof), but its primary advantages are its greater sensitivity and velocity resolution - AGES will be better suited to detecting dwarf galaxies at larger distances. Not only will this aid in constraining the HIMF but it will also help with understanding the nature of different environments. For example, how well do HI dwarfs survive in a cluster environment ? Are they preferentially found in certain areas (i.e. the outskirts) or at particular velocities ? Are there more gas streams in groups or clusters ? How well do optically and HI selected catalogues compare at low masses in different environments ?

By addressing these and other issues, and any serendipitous findings which may occur, AGES hopes to increase our understanding of galaxy evolution in the nearby universe. One of its key features is the range of environments which it will examine, from the very low density voids to the high density Virgo Cluster. By studying the HI gas, the fuel for star formation, with such great sensitivity and resolution, and in such widely differing density regimes, AGES hopes to give a new insight into the importance of environmental mechanisms on galaxy evolution.

2.3 Survey fields

The 14 fields selected for observation by AGES consist of the local void, 3 isolated galaxies, 2 galaxy pairs, 4 galaxy groups, and 4 galaxy clusters (including two different areas in Virgo, and also M33 as this field incorporates the Pisces-Perseus cluster). A brief outline of each of these is given below, including their observational status at the time of writing.

2.3.1 The Local Void

The Local Void is thought to be at least 45 Mpc in diameter and to be very empty (Tully 2008). First identified in the Nearby Galaxies Atlas of Tully and Fisher (1987), it describes a large part of the sky that is under-abundant in (but not totally devoid of) low redshift galaxies. A NED search of this field reveals no galaxies with optical redshift measurements out to a velocity of ≈4,700 km/s, with no known well-defined structure out to at least 10,000 km/s. The AGES field is a 5×4 degree field centred on an R.A. of 18:38:00, declination +18:00:00.

Although this region has of course been studied by HIPASS, the greater sensitivity of AGES may give a better prospect of detecting HI-rich, low optical surface brightness galaxies that may have been missed in optical surveys. If these form a substantial population it could significantly alter the view of the void - at the least, AGES will place new constraints on how empty the void really is. Any galaxies that are detected will make for an interesting comparison with those found in other environments. Observations have not yet started.

2.3.2 Isolated galaxies

3 isolated galaxies have been selected for study. As with the Local Void, AGES may discover new objects that have been missed by optical studies, but will also examine how the properties of the gas discs have evolved in isolation. These objects have large apparent diameters for which AGES may be able to resolve the HI discs, and so provide kinematic information on isolated objects. The galaxies chosen are as follows :
1) NGC 1156 (irregular galaxy). The field is 2.5×2 degrees centred on an R.A. of 03:00:00, declination +25:12:00. The observations are complete and this object is detailed in Auld 2007 and Minchin et al. 2010 (in preparation).
2) UGC 2082 (edge-on Sc galaxy). The field is 2.5×2 degrees centred on an R.A. of 02:36:00, declination +25:48:00. Observations are complete.
3) NGC 5523 (highly inclined SA galaxy). The field is 2.5×2 degrees centred on an R.A. of 14:15:00, declination +29:20:00. Observations have not yet started.

2.3.3 Galaxy pairs

2 galaxy pairs have been chosen for observation. The first is NGC 7332/9, a field 2.5×2 degrees centred on R.A. of 22:36:00, declination +23:48:00. NGC 7332 is classed in NED as a peculiar S0 while NGC 7339 is an SAB. The two are separated by 5.2', a projected distance of 35 kpc, and have 140 km/s difference in velocities. The pair are thus interesting as a possible interacting system. Previous claims for a detection in NGC 7332 have been controversial, with a detection by Knapp, Kerr & Williams (1978) using a circular feed at Arecibo being contested as a sidelobe detection of NGC 7339 by Biermann, Clarke & Fricke (1979) and Haynes (1981). With its greatly reduced sidelobes (see section 2.5) AGES may give a more definitive answer. Observations for this field are complete (Minchin et al. 2010 in preparation).

The second targeted pair is NGC 2577 / UGC 4375, an S0 and an SAB respectively. These are separated by 9.2' and just 4 km/s in velocity, so again may be an interacting system. The field for survey is 5×4 degrees centred on R.A. of 08:24:00, declination +22:30:00. The area has never been blindly mapped in HI with observations of this quality (HIPASS would not even have the resolution to distinguish between the two, let alone search for new companions). AGES observations have not yet started.

2.3.4 Galaxy groups

Galaxy groups may represent an important area of galaxy evolution. If clusters are composed of infalling groups, then their constituent galaxies may have undergone "preprocessing" - evolution within the group environment prior to the formation of the cluster proper. While they may be directly relevant to cluster formation, groups are also interesting in their own right. With a generally much lower velocity dispersion and galaxy density than within clusters, galaxies experience a quite different tidal environment as well as a different intergalactic medium.

4 groups are targeted with AGES, which are as follows :
1) NGC 628, precursor observations. A 5×1 degree field centred on R.A. 01:39:00, declination +15:46:00. Observations are complete (A06).
2) The Leo I group. A 5×4 degree field centred on R.A. 10:45:00, declination +11:48:00. This group is particularly rich in early-type galaxies and it will be interesting to compare the percentage detected in HI compared with other environments (HI is known to be common in early-types in the field (Grossi et al. 2009) but rare in clusters (Alighieri et al. 2007). Observations have not yet started.
3) NGC 7448. Another 5×4 degree field centred on R.A. 23:00:00, +15:59:00. In contrast to Leo I this group has a number of both early and late-type members. AGES observations are ongoing.
4) NGC 3193. Another 5×4 degree field centred on R.A. 10:03:00, declination +21:53:00. 9 other group members are known and the group is part of a well-defined galaxy filament. Observations are ongoing.

2.3.5 Galaxy clusters

Galaxy clusters contain by far the highest density of galaxies of any environment. Whereas the Local Group is approximately 1 Mpc across and contains 41 'classical' members (Mateo 1998), the Virgo Cluster spans perhaps 5 Mpc but contains ≈1,800 members (using data for VCC galaxies from the GOLDMine database). Similarly their velocity dispersions are very different, with a range of velocities of Local Group objects of ≈700 km/s (Mateo 1998) but as much as 4,000 km/s for the Virgo Cluster (see chapter 1). Processes such as galaxy harassment and ram pressure stripping, while perhaps possible in a group environment, are naturally expected to be very much stronger in clusters. Clusters, then, are interesting as the most extreme environment galaxies inhabit.

4 fields are targeted by AGES within galaxy clusters. This includes the Perseus-Pisces filament, a field 5×4 degrees centred on R.A. 01:34:00, declination +30:40:00. The primary target of this field is, however, the local galaxy M33, which is known to have many associated tidal HI features (Grossi et al. 2008). AGES observations have commenced but only a few scans have been completed.

As the nearest rich cluster (∼17 Mpc), Virgo is a natural target for AGES, but too large to observe its entirety to full AGES depth in this survey. Instead, 2 separate areas are targeted. The first, VC1, is a 10×2 degree field centred on R.A. 12:30:00, declination +08:00:00. Observations for this field are complete for a 5×1 degree area in the southern half, with the remainder ongoing. The second field is a 5×1 degree region centred on R.A. 12:48:00, declination +11:36:00. AGES observations for this field are complete. These fields are the major subject of this thesis and are discussed in chapters 5 and 6.

The final cluster targeted is Abell 1367. The field spans 5×4 degrees, centered on R.A. 11:44:00, declination +19:50:00. This is sufficient to cover the entirety of the cluster given its ≈92 Mpc distance. Observations are complete for the central 5×1 degree region and are discussed in Cortese et al. 2008, for the rest of the field they are ongoing.

2.3.6 The AGES volume

In each field the background region is also surveyed. The AGES volume is the collective of the background regions of each targeted area, to the redshift sensitivity limit of up to 20,000 km/s (the survey also includes part of the foreground to as much as -2,000 km/s). This is only a redshift of 0.07, a cosmologically insignificant distance of 282 Mpc, so AGES cannot hope to study galaxy evolution with redshift. Rather the goal is to increase sensitivity to HI giants by surveying a larger volume. At this distance the mass sensitivity is $\sim 2\times10^9$ $M_\odot$, so any detections here will by definition be HI massive.

Constraining the HIMF at low masses may have important consequences for cosmology, e.g. dark galaxies and the missing satellite problem. But it is equally important to understand how the HIMF varies at the high mass end, as previous surveys have found differences even at 10^9 $M_\odot$ in different parts of the sky (Rosenburg & Schneider 2002). A large volume, blind HI survey not constrained by optical biases such as AGES may give a much clearer view of how and perhaps why the HIMF varies with environment.

AGES sensitivity, velocity and spectral resolution, are unprecedented for a blind HI survey over such a wide range of environments. This is achieved not only by the telescope and instrument performance, but also the observing strategy. This is discussed in the next section.

2.4 Observing strategy and telescope details

Like other Arecibo HI surveys such as ALFALFA and AUDS, AGES is a drift-scan survey. The extraordinary sensitivity of Arecibo is due to its 305m spherical primary reflector. The dish itself is fixed, but the platform containing the secondary and tertiary reflectors (which correct for spherical aberration) can be rotated, to give $\approx$40 degrees of declination coverage. This procedure is of course simplified by the spherical nature of the dish, which avoids large differences in gain with changes in the instrument pointing. To avoid the beam spilling over the edge of the dish only a part of it can be covered by the instrument beams at any one time. This still enables a huge area of illumination of 213×237 m at zenith. With this footprint, the zenith angle of the instrument is kept below 15 degrees to avoid overspill.

The basic observing strategy is to drive the telescope to the starting position of the scan and let the sky drift overhead for the required duration. Each scan of the sky is performed 3 times. The advantage of keeping the instrument angular position relative to the dish constant is that it eliminate any angle-dependent effects such as telescope gain. The system temperature varies only slowly over the course of each scan, and baseline ripple caused by standing waves between the receiver and dish should remain constant. This allows for an accurate estimate of this ripple so that it can easily be removed.

The system temperature is the equivalent temperature of a blackbody if it were to generate the same power as is observed to be coming from the receiver. This includes all sources : the actual astronomical source, the background signal from the sky, noise within the receiver itself, and loss in transmission from the antenna to receiver. For ALFA the system temperature is typically $\sim$30 K, for HIPASS $\sim$20 K, and for the VLA $\sim$40 K.

While using the full 305 m diameter of Arecibo would give even better sensitivity, there are many advantages to only using part of the dish, beside the obvious freedom from only being able to observe the sky directly overhead. Instead of using a single beam with the footprint of the entire dish, it is possible to use multiple beams that illuminate only part of it. Each observes a different part of the sky simultaneously, analogous to the individual pixels in an optical CCD. In a scanning mode of observing, the same part of the sky may be observed by multiple beams, allowing very high sensitivity over a relatively large area. This is why the advent of multibeam receivers has been so revolutionary to HI surveys.

6 of the 7 beams of ALFA are arranged in a hexagonal pattern with the remaining beam in the center. At the start of each scan ALFA is rotated to maintain the same parallactic angle with every beam equidistant at 1.8' in declination. The beams are slightly elliptical (enabling a larger area of the dish to be illuminated), as is the overall beam pattern. Ideally on a single scan the whole area mapped by ALFA would be Nyquist sampled, but due to the beam separation and size this is not quite the case. To ensure Nyquist sampling, individual scans are staggered by 1.5' in declination. With partially overlapping beams, each point on the sky, for the full 3 passes, is in effect scanned by 25 beams. Since a source takes 12s to cross the beam this gives a 300s integration time.

The beam pattern of ALFA is shown in figure 2.3. Except at the edges (where coverage is not fully sampled and the sensitivity is consequently lowered), every beam will scan the same point in the sky. This helps reduce the effect of varying gain and sidelobes of each of the beams. The effects of radio frequency interference (RFI) can also be reduced by the observing strategy. Each scan is numbered - in each of the 3 passes, the odd numbered scans are all observed first, and only then are the even numbered scans started. This staggers the scans in time. Any transient or time-varying RFI present in a scan is less likely to be present (or as strong) in corresponding, overlapping scans. This at least mitigates the effects of time-varying RFI, though it will not eliminate it completely.

For the smaller of the fields discussed in this thesis, VC2, sufficient time was allocated to observe its entirety in a single observing campaign (January-February 2007). Observations consisted of 120 scans each 20 minutes in duration covering 5 degrees of R.A (40 scans at unique declinations, each done 3 times). At the start of each scan, as is standard, a signal from a high-temperature noise diode is injected for 5s for calibration (5s is also recorded with the diode off). To minimise overheads, an IDL script is used to prepare an observing command file that automatically controls the pointing and rotation of the instrument, and begins and ends the data recording appropriately.

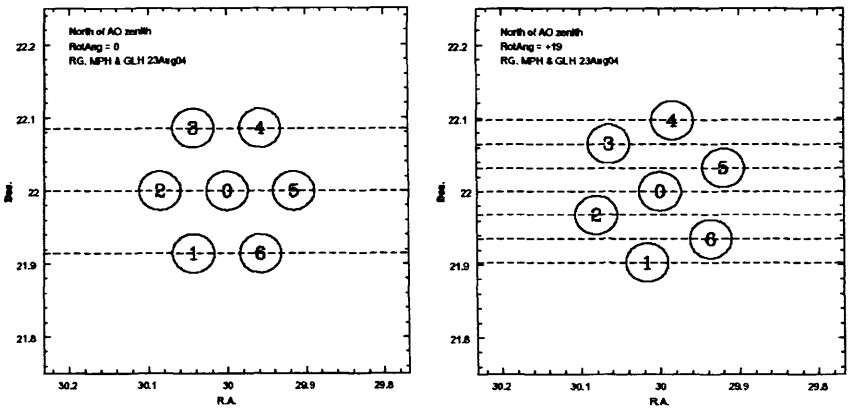


Figure 2.3: ALFA's beam pattern. Left : unrotated. Right : rotated to an angle of 19 degrees. Beams as shown as circular for simplicity. Dashed lines represent the beam tracks across the sky. Image from a memo of Giovanelli, Haynes and Hoffman, 2004.

For the larger VC1 field it was not possible to observe the whole field in a single campaign as Virgo is not visible to Arecibo for that long. The telescope scheduling was only known for a month ahead at the time the observations commenced. To try and ensure useful data was obtained during that run, the field was split into quarters, the goal being to at least obtain another 5 degree field in Virgo as a good comparison with VC2. Thus scans were again 20 minutes in duration with the same calibration length beforehand. In the event, the full 3 passes were obtained for the target quarter, and a second quarter completed to 2 out of the 3 passes necessary (observations began in January 2008 and are ongoing).

From VC1 onwards, during the calibration the telescope tracks the sky, since both the diode and the sky are observed simultaneously. If this were not done, the telescope would be drifting for about 10s, (in actuality slightly more, allowing for the time it takes to settle into position). This is comparable to the time it takes for a source to cross the beam. This would create a risk that a continuum source could move in or out of the beam during the calibration, altering the recorded power levels. Tracking the sky ensures that the same continuum sources are present throughout, so essentially only the power from the diode, which is controlled, is allowed to vary.

AGES observations to date have used the Wideband Arecibo Pulsar Processors (WAPPs)[1]. These backends were originally designed for pulsar observations but have been adapted for spectral line observations. The bandwidth is 100 Mhz with ~24 Khz resolution, for a total of 4,096 channels. The AGES frequency range is 1,331 to 1,431 MHz, equivalent to a redshift range from approximately -2,000 to +20,000 km/s. The velocity resolution is equivalent to 5.3 km/s at the rest frequency of HI.

Each of the 7 beams measures two linear polarisations separately and simultaneously. Data is recorded every second for every beam, in both polarisations and in every channel, as 4-bit floating point numbers. Since the 21cm line is not intrinsically polarised (see chapter 1), the use of the two polarisations can help in establishing whether a detection is real or false (see next section and chapter 4).

For both on-site and remote observing the recorded data is viewed in realtime, allowing for monitoring of RFI as well as any instrumental problems. Often, the receiver of a single polarisation in a single beam was not operating correctly, recording a maximum power level a factor 10 lower than the values of the others. The data reduction tools, which were adapted from HIPASS, do not allow this data to be rejected. This has probably resulted in a slight increase in noise levels, but from the results of chapters 5 and 6 it does not appear to have had a significant impact. The methods of data reduction are discussed in the next section.

After a source has been detected by the methods described in chapter 4, if there is any reason to doubt the validity of the detection follow-up observations are obtained. Since these are normally faint sources, Arecibo is used for this. 10% of AGES scheduled observing time is alloted for this purpose. Although the central pixel of ALFA has been used occasionally for this, ordinarily the L-band wide receiver is used. This is slightly more sensitive with a system temperature ~5 K lower than ALFA. These observations use the position-switching technique, with 5

[1]Recent observations are using the Mock spectrometer, which offers higher bandwidth, but since AGES has only just begun using this instrument and it was not available for any of the observations in this thesis, it is not discussed here.

minutes on source and 5 minutes off (see below). The data can be quickly reduced and another observation taken immediately if needed.

2.5 Data reduction and analysis

A program by Robert Minchin initially scales the velocity (frequency) of the data from the recorded geocentric into solar system barycentric coordinates. Thereafter data reduction is performed using the AIPS++ programs LIVEDATA and GRIDZILLA. These have been adapted from their original HIPASS configuration to process AGES data (by O'Neil, de Block, Minchin, Auld and Kilborn of the AGES consortium, in collaboration with Mark Calabrettra who was involved with the design of the original software). The LIVEDATA program performs flux and bandpass calibrations while GRIDZILLA assembles the processed data into the final cube.

In a pointed survey, one of the most common strategies to calibrate the data is the position-switching method. Once the source has been observed (the on scan), the telescope returns (ideally) to its initial orientation and integrates for the same duration (the off-scan), in both cases tracking the sky. The purpose of the off-scan is to record data to the same depth and with the same telescope pointing on a blank patch of sky. This is designed to give an accurate measure of the noise and bandpass. The off-scan can then be subtracted from the on-scan to remove the baseline while leaving the source. Problems can occur if the source is very extended (such that when the telescope moves to what should be the off-scan, it is still recording the source), or if another source is now present at the original orientation, but these are very rare circumstances and for the most part the technique is very successful.

In a drift scan the situation is more complicated. Since each scan takes 20 minutes and even a small field may have 120 individual scans, it is hugely impractical to double the survey time with the method described above. However, the majority of the sources are point sources. This means that a source detected in one scan will almost certainly not be detected in the overlapping scans (in the same beam). The off-scans are thus the appropriate scans from the survey itself, which are at very similar pointings (if not actually identical) so should have a very similar baseline.

This strategy has been shown to work well for point sources, but bright extended sources do cause problems. The normal procedure in LIVEDATA is to calculate the median of each scan and subtract this as the bandpass (this is done for each channel). While the median is generally robust to a small fraction of the scan at much higher flux values than the rest (i.e. point sources), if a source fills a sufficient fraction of the scan it can raise the median. This causes over-subtraction of the bandpass, which is manifested in the data as apparent negative "lobes" that fill the length of the scan (the whole swathe of R.A. - for HIPASS, which scanned in declination, such lobes cover the whole swathe of declination (Barnes et al. 2001)). An example is shown in the left panel of figure 2.4.

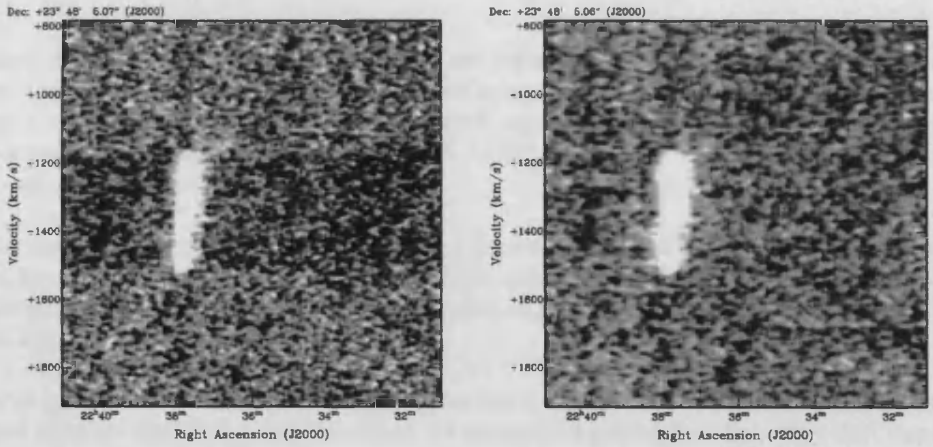


Figure 2.4: Example of over-subtraction of the bandpass from an extended source, using the standard median estimation (left). Using the "MinMed" estimator effectively removes the negative lobes (right). The image is of NGC 7339, courtesy Robert Minchin.

It is possible to remove these negative lobes through alternative estimates of the bandpass, such as MinMed or the similar procedure MedMed. These were originally developed for high velocity clouds detected in HIPASS (Putman

et al. 2002), which can be much larger than even the beam size of Parkes. For MinMed, the scan is divided into an arbitrary number of "boxes", such that each is long enough to contain the entirety of the extended emission. The median of the scan in each box is calculated and the minimum of the medians is used as the bandpass. Similarly, MedMed takes the median of the medians and uses this value as the bandpass.

In this way the flux from the extended source is effectively ignored when calculating the median of the scan (or at least greatly reduced by using the median of medians). This removes the negative sidelobes, as shown in the right panel of figure 2.4. The trade-off for this is that since a smaller fraction of the bandpass is analysed, the bandpass estimate is less accurate, which can raise the noise level. Additionally, it can be shown (Putman et al. 2002, and also see chapter 7) that an extended source must fill a significant fraction of the scan, approximately 30% or more, before the standard procedure will start to actually suppress it (rather than just causing the negative lobes seen above). For these reasons MinMed and MedMed have not been used in reducing the data in this thesis.

One problem traditionally associated with Arecibo is the problem of sidelobe contamination (for example see section 2.3.3). This is much reduced using ALFA, whose sidelobes are illustrated in figure 2.5 (this is reproduced from A06). These are actually well removed thanks to the procedures adopted here, with each beam making an equal contribution to each pixel in the final cube. Using the standard median estimate of the bandpass, a source would have to appear in 4 or more beams simultaneously for sidelobe contamination to become significant (A06) - otherwise, most beams will be unaffected by sidelobe contamination and so it will be averaged out. Point sources should thus not suffer from sidelobe contamination, a fact which is borne out in the data analysis. Extended sources may still suffer slightly.

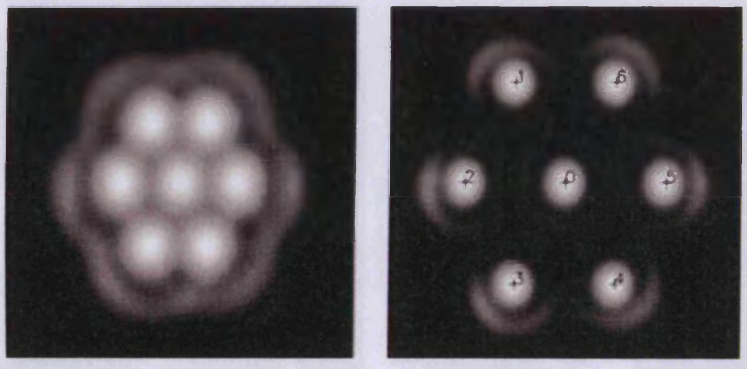


Figure 2.5: An illustration of the ALFA beam footprint to emphasise the strength and symmetry of the sidelobes. (a) A highly stretched image of the combined response of all seven ALFA beams, the image size is 32×32 arcmin. (b) The beam separations have been stretched to 2.5 times the true separations to highlight the sidelobes of the beams. The angular scales of each beam is correct and unchanged. The image size in this case is 48 48 arcmin (Courtesy: Carl Heiles). This figure and caption reproduced from A06.

Since the receivers in each polarisation of each beam are different, individual calibration is necessary. The correction factors have been determined by Robert Minchin using the ALFACAL program[2], based on the individual temperatures of the receivers and assuming a typical zenith angle of 5 degrees. The calculated fluxes from AGES data have been shown from the NGC 1156, NGC 7332, A1367 and now the VC1 and VC2 cubes (chapters 5 and 6) to agree well with observations in the literature.

GRIDZILLA also applies beam normalisation. The flux densities are scaled by the inverse of the beam response, which gives the correct flux for a point source at the pixel center (Minchin et al. in preparation). Although this results in the flux of extended sources being overestimated, in practice this effect appears to be negligible in the VC1 and VC2 cubes (again see chapters 5 and 6).

Radio observations, by convention, measure intensity as a function of frequency. Different conventions have arisen in optical and radio astronomy as to how to convert the measured wavelength or frequency shift respectively into a recessional velocity. The optical convention is that at non-relativistic velocities (i.e. z $\ll$ 1.0), $v = cz$ where c is the speed of light and z is the redshift, i.e. :

$$v_{optical} = c\frac{\lambda - \lambda_0}{\lambda_0} \tag{2.1}$$

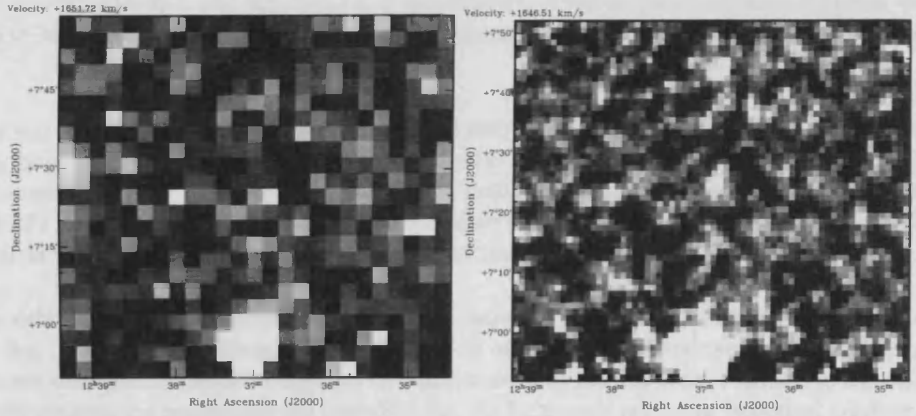
[2]http://www.naic.edu/~rminchin/alfacal.shtml

Where λ is the measured wavelength and λ_0 is the unshifted wavelength. In contrast the radio convention is to use the frequency shift so that :

$$v_{radio} = c\frac{f_0 - f}{f_0} \tag{2.2}$$

Where f is the measured frequency and f_0 is the unshifted frequency (again this is for the non-relativistic regime, i.e. $(f - f_0)/f_0 \ll 1.0$). The radio velocity $v_{radio} \neq cz$. However, since the data reduction programs easily allow either convention to be used, the data are gridded using the optical convention. Each channel in the receiver is of constant frequency width, which is not quite the same as a constant (optical convention) velocity width, and this is not corrected in the final output cubes. The difference is however only marginal, varying by less than 1 km/s over the entire length of a cube.

The standard procedure is to grid the data into 1' pixels (with each frequency (or velocity) channel at the nominal resolution of the instrument). This is oversampling, compared to the 3.5' Arecibo beam, and is useful in source extraction. Each pixel is not independent. This means that any real point source must span more than 3.5 (adjacent) pixels, whereas a spurious detection is likely to appear in less. If the data was binned into 3.5' pixels, both weak point source detections and spurious detections alike would only appear in single pixels. This is illustrated by figure 2.6. Spurious detections can appear with overbinning, but there are other reasons to reject these - see chapter 4.



(a) Pixel size of 3' **(b)** Pixel size of 1'

Figure 2.6: Comparison of the appearance of a weak source (center) in the VC1 cube with different pixel sizes. While the bright source to the south is clearly visible in both cases, the weak source is less visible when using approximate beam-sized pixels (3', left panel) than compared with the standard overbinning (1' pixels, right panel).

The baseline along the entirety of each spectrum (that is, the flux as a function of velocity) is approximated as a first-order polynomial by GRIDZILLA and removed. This proves to be a generally very good approximation - a slice through the VC1 cube, showing the uniformity of the baselines, is shown in figure 2.7. Although this uniformity is not perfect - as with any data set - it is only when more complex analysis is performed that its defects become readily apparent (see chapter 7). The major sources of variation are RFI, the non-uniformity at the field edges due to the beam pattern, and also the ~1,000 km/s edges of the baseline, due to the sensitivity decrease of ALFA at the edges of its frequency range. For VC1 the integration time varies across the cube which is another contributing factor in this case. Higher order polynomials are not used since that would risk creating artificial structures, as well as suppressing real sources.

The two major sources of RFI present are the San Juan FAA radar and the L3 GPS satellite which occasionally passes overhead. These typically reduce sensitivity over the velocity ranges ≈15,300-15,900 and ≈8,200-8,800 km/s respectively. Additionally the signal from the Milky Way is so bright that LIVEDATA is not able to process it correctly, resulting in the survey being virtually blind from -50 to +50 km/s.

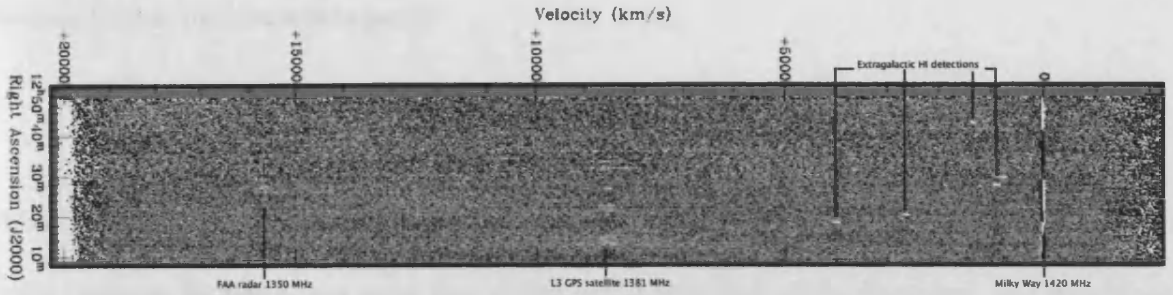


Figure 2.7: A slice through the VC1 cube at a declination of approximately 7:40:00, showing the very uniform background achieved by AGES observation strategies and data reduction. The Milky Way, RFI, and several extragalactic detections are visible.

No smoothing is done during the data reduction phase by GRIDZILLA, but after griding the data are Hanning smoothed to $\approx$10 km/s velocity resolution. This is done for several reasons. Firstly, it greatly reduces Gibbs ringing (see below). Secondly, it improves sensitivity to all but the narrowest of sources. Thirdly, for these reasons it is the standard practice of the ALFALFA survey. For a fair comparison of the two datasets AGES employs the same level of Hanning smoothing.

Gibbs ringing occurs as the spectrometers work in Fourier transform space. While the actual shape of a channel is described by a tophat function, the Fourier transform of this is a sinc function. A source much narrower than the channel width appears as a delta function, which is a constant in Fourier space. Thus the source will be measured as a constant convolved with a sinc function, i.e. a sinc function of the height of the delta function, with the wavelength equal to half the channel separation. Successive channels alternate between peaks and troughs of this sinc function.

To cause any significant Gibbs ringing a source must not only be narrow (or rather, have a sharp edge in velocity) but also very bright, since a sinc function decays rapidly (to a few percent of maximum within 20 channels). In practice this means that only the Milky Way causes any noticeable ringing - other galaxies are simply not bright enough, and RFI is not narrow enough. The affected channels could just be ignored, but it is preferable to greatly reduce the effect with frequency smoothing, which also increases sensitivity.

Hanning smoothing effectively removes the ringing and increases sensitivity, but at the expense of velocity resolution. The flux in a particular channel is set to be half its original value, plus one quarter of the flux in each of the two adjacent channels. Since this combines an equal amount of the flux from channels in a peak and a trough from the Gibbs ringing, the oscillations are cancelled out. As 3 channels are being smoothed together, this reduces the rms by a factor $\sqrt{3}$. This will give a real increase in sensitivity for sources spanning 3 or more channels - there will be signal in each channel averaged together, thus increasing the S/N. For a source that spans fewer channels, signal from the source will be averaged with noise, so sensitivity to these sources will be reduced.

Standard data analysis is performed using the ATNF program MIRIAD. This was designed for radio interferometry data reduction, but has a number of tasks useful for single-dish 21 cm analysis. The most important of these for AGES is *mbspect*, which measures all of the key parameters of an HI detection : the recessional velocity, velocity width, total flux (with all associated errors), rms, and S/N. It also performs position fitting. The user first estimates the position by a visual inspection of the data, then supplies this to *mbspect*, along with the approximate spatial width of the source in pixels. Normally sources are unresolved so this will be 3. MIRIAD is able to fit a 2D (spatial) Gaussian to the source and use the center of this Gaussian as the coordinates of the source.

The flux values output by *mbspect* at each channel are computed by spatially integrating within the specified box. Ordinarily the source is unresolved, so the contribution from each pixel is weighted by the beam parameters. It is also possible to simply sum or average the flux in each pixel with no weighting, which can be useful in detecting very extended sources, as discussed in chapter 7.

A slight baseline ripple often remains despite the data reduction. This is removed by an additional first-order polynomial correction over a shorter velocity range (typically 4,000 km/s). The user also inputs the approximate velocity width of the source. Within the 3-dimensional box described by the spatial and velocity width, *mbspect* integrates the flux, after the additional baseline correction, to give the total flux of the source. This is then used to

estimate HI mass via the standard equation :

$$M_{HI} = 2.36 \times 10^5 \times d^2 \times F_{HI} \tag{2.3}$$

Where M_{HI} is the HI mass, d is the distance in Mpc, and F_{HI} is the integrated HI flux in Jy km/s.

When computing the baseline and rms of the spectrum, part(s) of the spectrum can be masked. This includes the source itself, any other HI sources that may be present, the Milky Way and RFI. These features would bias the estimate of the baseline, leading to inaccurate baseline estimation, and could also inaccurately increase the rms, leading to an artificially low S/N estimate. Such features are identified by visual examination of the data cube and of the spectrum itself.

Mbspect also finds the minimum and maximum velocity width at 50% and 20% of the peak flux. For low S/N sources, the minimum and maximum widths at each flux level can be significantly different. An example is shown in figure 2.8. In this case, the minimum velocity width at 50% of the peak flux (29 km/s) is 177 km/s lower than the maximum width (206 km/s). The minimum width is often a clear underestimate of the true velocity width - in figure 2.8 for example, the minimum width picks out a single peak of the double-horned profile. For this reason velocity widths reported by AGES are always the maximum widths (which are constrained by the user's initial estimate of the velocity width).

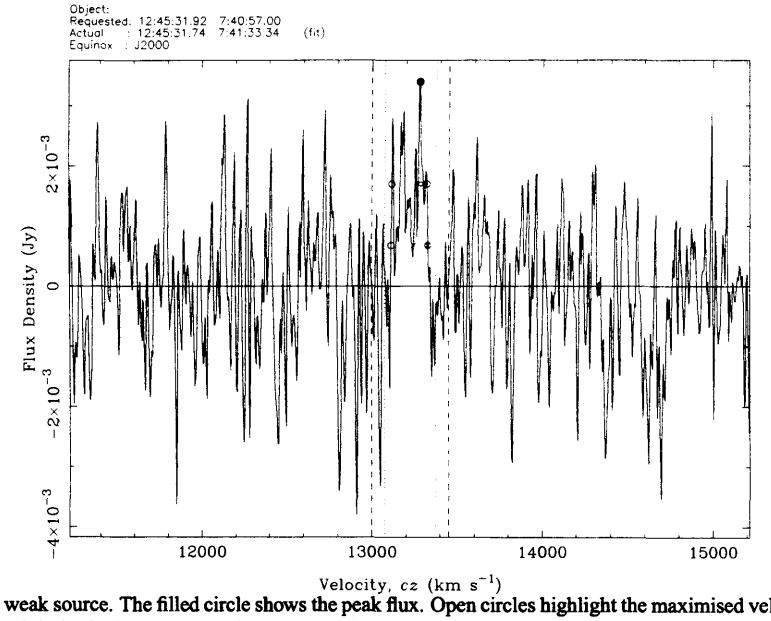


Figure 2.8: Example of a weak source. The filled circle shows the peak flux. Open circles highlight the maximised velocity width of the source, with crosses showing the minimised width. Dashed lines indicate the region masked when estimating the baseline (a first-order polynomial); dotted lines indicate the user's initial estimate of the velocity width.

While the Hanning smoothing applied as standard does increase sensitivity slightly, additional smoothing can also be applied to individual spectra with *mbspect*. This is seldom necessary, but can be useful for more accurately estimating the velocity profile of very faint sources. Once this is achieved, the spectrum is measured and output without this extra smoothing for consistency with the more typical sources measured.

Many other MIRIAD tasks are also used and these are described in more detail in the rest of this thesis where appropriate. These include *imbin* for rebinning images, *immask* for masking sources, and *moment* for making source maps integrated over a specified velocity range. Some of the resulting highlights from all these processes are summarised in the next section.

2.6 Previous results of AGES

Prior to commencing observations of the Virgo Cluster, the main subject of this thesis, AGES had already observed several fields to their full target depth. These are the NGC 628 group, the NGC 1156 isolated galaxy, and the Abell 1367 cluster. In this section the main results of these surveys are described. At the time of writing observations had also been completed of the galaxy pair NGC 7332/7339 and the isolated galaxy UGC 2082, with observations of several other fields ongoing. A paper on the NGC 7332/7339 pair is in preparation (Minchin et al. expected 2010), but as these results have not yet been published in a refereed journal they are not included in this section.

2.6.1 The NGC 628 group

The group associated with the spiral galaxy NGC 628 was selected for the AGES precursor observations. What follows here is a summary of A06. As the primary goal of this was to test the observation and data reduction procedures, the depth obtained was less than the normal target depth of AGES - 192s of integration time, a noise level of typically $\approx$1.1 mJy rms. By combining different scans to give different integration times, A06 was able to show that the noise level achieved was in agreement with expectations. This has since been borne out by fields observed to full depth. The baselines were shown to be stable over the course of each 20 minute scan to within 0.1%.

The field selected spans 5×1 degrees and is centred halfway between NGC 628 and UGC 1176. NGC 628 is the brightest member of a group of objects, which includes the peculiar spiral NGC 660 (a polar-ring galaxy). These objects have a further 7 associated companions, at a distance of $\approx$10 Mpc. These are mostly gas-rich, star-forming dwarf irregulars. The close proximity allows for an AGES sensitivity, even at the reduced depth, with an M_{HI} of just 2×10^6 $M_\odot$. The group is about the same size as the Local Group, but rather richer.

The precursor observations served to highlight the great improvements over HIPASS. AGES was able to detect the dwarf galaxy dw0137+1541, which is so close to UGC 1171 that HIPASS could not differentiate between the two (AGES observations also suggest that these two objects are mutually embedded in a gas envelope). Whereas HIPASS could not resolve any of the group members, AGES could resolve 4 objects sufficiently well as to map their velocity fields.

The high resolution of AGES makes it well-suited to searching for intragroup gas, since it would be able to distinguish between this and gas associated with previously uncatalogued galaxies. However, AGES did not detect any intragroup medium to a column density sensitivity of 2×10^{18} cm^{-2}. A excess of flux (deviation from the expected Gaussian distribution) was found in a blank region of the cube, suggesting that there could be sources below the detection limit. This is an issue discussed in detail in chapter 7, but in short it seems that there are too many errors associated with this technique for it to provide any reliable evidence.

Extended emission was detected from NGC 628, so the lack of an intragroup medium is not a consequence of data reduction techniques. Indeed, half of all the HI detected in the group by AGES is associated with NGC 628. Excluding the immediate vicinity of NGC 628, and assuming the presence of the upper limit for the HI column density, gives an upper limit of 5×10^8 $M_\odot$ of intragroup HI distributed across the cube between NGC 628 and the dwarf KDG 10. This is less than 3% of the HI in the galaxies. A06 conclude that either tidal interactions are poor at removing gas from galaxies, or perhaps that the group is simply too young for enough interactions to have taken place.

Although AGES did not detect any new features within the group, it did find 9 new galaxies in the volume behind it (as well as 13 previously catalogued objects). These are entirely new, and were not previously listed in NED. Most of the other detections had only been previously identified in automated surveys (i.e. they have been catalogued by an automatic source-finding routine and were not previously well-known objects), principally the SDSS and 2MASS, and are without optical redshift data. Despite the relatively low sensitivity, AGES still uncovered a number of interesting detections.

Those detections which are entirely new include AGESJ013827+154728, an optically irregular galaxy but with the HI double-horn profile typical of a rotating disc, and AGESJ014834+152756 which is a one-armed spiral. Such objects are interesting as a single prominent spiral arm is often a sign of a perturbation and may have associated HI streams (Minchin et al. 2007). Interferometric observations are required to confirm this.

AGESJ013956+153135 was also not listed in any previous catalogues - identifying the optical counterpart here is difficult, and it remains as a dark galaxy candidate. This is an especially interesting candidate since its HI mass is very large at 5×10^9 $M_\odot$, more than 100 times that of VIRGOHI21, for example. Follow-up observations have confirmed that this is a real HI detection. Since it is very distant at 231 Mpc, deeper optical imaging is also necessary to confirm if the optical counterpart is merely of low surface brightness, rather than entirely absent.

A detection such as this is an example of an object that could easily be entirely missed through optical surveys. The HI content of some of the objects previously catalogued is also interesting. AGESJ013204+152947, though HI massive at 8×10^9 $M_\odot$, has a red optical colour (g - r $\sim$ 0.75) and no emission lines in the optical spectrum, which would indicate ongoing or recent star formation. The HI profile of AGESJ013827+154728 is a double-horn, suggesting a rotating disc, but the optical appearance is irregular. Similarly, the narrow HI velocity width of AGESJ013953+151955 suggests a face-on disc (given its high mass of 5×10^9 $M_\odot$), yet the optical image suggests an edge-on disc.

Finally, the number of detections in this volume (27 in total) may be rather low. Simulations based on the HIMF from Parkes (Zwaan et al. 2003) suggest an expected number of detections of 49. This is actually the number of detections in the similar sized VC2 region (chapter 6) although a few are awaiting follow up observations. The significance of the difference between the predicted and actual number of detections is only about 3σ, so this may just be an effect of cosmic variance.

2.6.2 The isolated galaxy NGC 1156

This galaxy is amongst the most isolated galaxies known, with no companions at similar redshift within 10 degrees (Karachentseva, Lebedev & Shcherbanovskij 1986). AGES has fully mapped a 2.5×2 degree field around this star-forming, irregular galaxy, and the data is fully described in Auld 2007 (see also Minchin et al. 2010 in preparation) and summarised here. Its redshift is 375 km/s and its distance, determined from measurements of the brightest stars, is 7.8 Mpc (Karachentseva, Musella & Grimaldi 1996). The AGES field spans a projected length of 0.34 Mpc in R.A. by 0.27 Mpc in declination.

Being such an isolated and therefore interesting galaxy, many previous observations have been performed, and the flux measured by AGES has been shown to be consistent with the previous data. Most notably, Hunter et al. 2002 discovered a possible small HI tail extending to the north-east, ending in an HI complex of a high velocity width. This structure cannot be spatially resolved by AGES. However, AGES did discover a new companion object, AGESJ030039+254656, on the same side of the galaxy as the possible tail.

This discovery is a dwarf galaxy, with an HI mass of just 2×10^6 $M_\odot$ - such low masses are detectable thanks to the close proximity of the objects. Using Hα and R-band measurements, its star formation rate was estimated at 9×10^{-4} $M_\odot$ yr^{-1}. This compares with the estimated rate of 0.71 $M_\odot$ yr^{-1} for NGC 1156 - both its mass (1×10^9 $M_\odot$) and star formation rate are about 3 orders of magnitude higher than its dwarf companion.

The single companion is well below the expected number from the HIPASS HIMF, which predicts about 18 dwarf galaxies for each galaxy more massive than 10^9 $M_\odot$. However, given the associated errors this difference is only significant to 3σ, so is perhaps again only the result of cosmic variance. Interestingly, the companion appears to be too small and too far away to create significant tidal forces that might explain the presence of the HI tail. It is possible the two galaxies were once in more suitable locations for a stronger encounter, which might explain the tail.

36 galaxies were detected in HI in the background region. These are all eliminated as other possible companions of NGC 1156 by their redshifts, with the first galaxies not appearing until a redshift of 3,260 km/s. 16 of these have not been previously catalogued. Although 3 sources cannot be assigned a definitive optical counterpart, as multiple optical sources are present within the Arecibo beam, there are no cases without at least 1 clear optical counterpart candidate.

2.6.3 The Abell 1367 cluster

This cluster is sufficiently distant (93 Mpc) that AGES will eventually map its entirety in a 5×4 degree field, while close enough to still be sensitive to relatively low HI masses ($\sim 2\times10^8$ $M_\odot$, about the same as HIPASS at the distance of Virgo). Observations are ongoing but the central 5×1 degree swathe is complete to the standard target depth of AGES. This region, which includes the X-ray emitting cluster core, is described fully in Cortese et al. 2008 (C08 hereafter), and that paper is summarised here.

The area surveyed was approximately the same as the precursor observations, but with the full integration time and very different environment, 100 HI sources were detected - far more than the total of 27 for the NGC 628 group (incidentally HIPASS does not detect anything at all in this part of A1367). Given this, it seems that cosmic variance can easily explain the relatively low number of detections in the precursor study. 54 of the detections belong the the A1367 cluster alone, leaving 46 in the background volume, more consistent with the expectation from the Parkes HIMF.

While 54 of the detections are within the cluster's velocity range (4,000-9,000 km/s), identifying the cluster in the HI data alone is not straightforward. Optical data shows a clear clustering of objects around X-ray gas emission in the center of the surveyed region, as well as the classic "finger of god" in a position-redshift wedge diagram. The HI detections do not follow these trends, as seen in figure 2.9. This may be partially a selection effect - at the cluster distance, it is difficult to detect galaxies in HI nearer the cluster center, as these have probably become HI deficient.

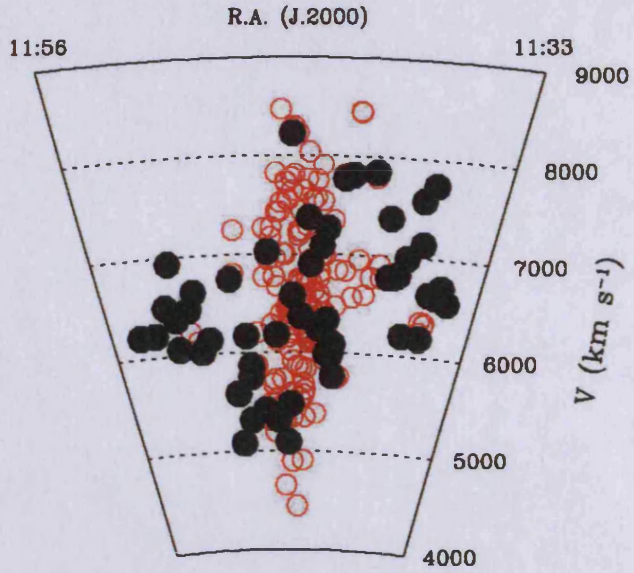


Figure 2.9: The wedge diagram of the AGES-A1367 region in the velocity range 4,000 < V < 9,000 km s^{-1}. Filled circles indicate the HI detections while empty circles are galaxies belonging to the SDSS optically selected sample (g < 17 mag). Reproduction of figure 8 from Cortese et al. 2008.

This is explained by the fact that most of the detections here are not significantly HI deficient, with only 5 objects more deficient than 0.3. The non-deficient objects are probably galaxies still infalling into the cluster that have not yet experienced its full force of its environment - hence their positions and velocities do not correlated with the optically identified cluster members. Although some galaxies are detected within the observed X-ray gas at the heart of the cluster, since the velocity dispersion is so high there is no way to tell if these are really in the cluster center, or just appear so by a projection effect.

A further example of the lack of detection of the cluster in HI is provided by the non-detections of blue galaxies. Virtually all of the HI detections in this region correspond to blue sequence objects (see chapter 1), but about 30% of optically-selected blue sequence cluster members are not detected in HI. C08 show that after rejecting those galaxies within RFI contaminated regions, and objects close to HI-bright galaxies, the remaining objects are not so extraordinary. They need only have an HI deficiency exceeding 0.4 to escape detection due to the AGES sensitivity limit. Thus, those objects actually subjected to the cluster environment have not been detected.

C08 were able to reproduce many well-known relations such as colour-mass and surface brightness-mass for the sample, but also found that the baryon fraction (the fraction of total mass that is baryonic) is a constant - the first time this has been shown for an HI-selected sample. This poses problems for some models of galaxy evolution, which use feedback mechanisms dependent upon galaxy mass to remove baryons - for example as an explanation of the missing satellite problem.

Several individual detections are also of note. These include the CGCG97-027 and CGCG97-041 groups, which are at projected distances of 3.2 and 1.6 Mpc from the cluster respectively. Both show extended emission spanning as much as 200 kpc in projection from the group members. There are also two detections without obvious optical counterparts, AGESJ113939+193524 and AGESJ114809+192109. In both cases potential candidates are visible but no optical redshift data is available to confirm the association with the HI.

Chapter 3

Other Data

3.1 Introduction

As described in the previous chapter, AGES is designed to search for neutral hydrogen. At the most basic level, optical data is necessary to identify which - if any - of the optical galaxies corresponds to the HI detection. Many HI-detected objects are not listed in optical survey catalogues (see chapters 5 and 6) as they are gas rich but optically faint. This means it is important that the optical data be as sensitive as possible, to give the best chance of detecting an optical counterpart.

Neutral hydrogen is often purported to be the ultimate fuel for star formation, but current models indicate the process is complex, with the atomic gas needing to cool to the molecular state (possibly via accretion on dust grains) before stars actually form. To study correlations between the gas and star formation, multi-wavelength data is essential. While optical data can be used to indicate the presence of star formation (by showing if a galaxy is very blue), other wavelengths (such as UV) have been shown to more directly trace star formation via young stars.

It is also desirable, wherever possible, to have other HI data for the same galaxies. Comparisons with the existing data serve to check the validity of the new data, ensuring that it has been accurately calibrated. Occasionally the new, generally more sensitive data can give quite different measurements to previous data. In these cases follow-up may be necessary to determine which is accurate. This is particularly true for very weak sources, where, in a few instances, previous pointed observations may have been considerably more sensitive than the new blind survey.

The Virgo Cluster forms the main subject of this thesis, and this historically important region has already been subject to many surveys at different wavelengths. The key data sources used here include the famous Virgo Cluster Catalogue (hereafter VCC) of Bingelli, Sandage and Tammann (1985), the Sloan Digital Sky Survey (SDSS), the Isaac Newton Telescope Wide Field Survey (INT WFS), data from the Galaxy Evolution Explorer satellite (GALEX), and the Arecibo Legacy Fast Arecibo L-band Feed Array (ALFALFA) survey.

The above sources are all uniform in sensitivity over their respective areas, though by their nature they are of different sensitivities to each other. Where necessary, these surveys are also supplemented with data from other catalogues, primarily the Galaxy On Line Database Milano Network (GOLDMine) and the NASA Extragalactic Database (NED). These provide access to much more data but without such uniformity. The remainder of this chapter serves to describe all these catalogues and survey data, outlining why some data sources may be preferable to others and the data analysis techniques used in this thesis.

3.2 Optical surveys

3.2.1 The Sloan Digital Sky Survey

The SDSS is an ongoing project to map over 12,000 square degrees of the sky, begun in 2000 and scheduled to continue through 2014. This large area is combined with multiple filters (u,g,r,i and z, see figure 3.1) as well as high sensitivity (a surface brightness limit of $\sim$ 26 mag arcsec^2 in the g band, Kniazev et al. 2004). Its two spectrographs provide optical spectra of many of its detections. As a consequence of the size and quality of the data set, the science goals and capabilities of the survey are very wide-ranging, from detecting solar system asteroids

to distant quasars. The survey uses a dedicated 2.5m telescope at Apache Point Observatory, New Mexico, and achieves its high depth despite a mere 54s exposure time.

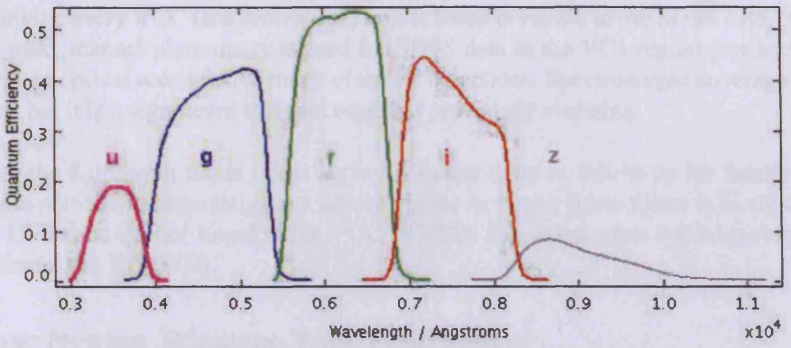


Figure 3.1: Wavelength ranges of the SDSS u,g,r,i,z filters.

Besides the instrumental performance and quality of the observing site, one other key factor giving the survey its high sensitivity is its use of drift scanning, similar to that used by AGES - the difference being that thousands of CDD pixels are used in place of 7 receiver horns. By using each pixel to estimate the background level of the same part of the sky, the resultant images have the equivalent of extremely good flat-fields (see below), i.e. a very uniform background. This gives greater confidence that faint sources are real structures, and not merely artifacts. Smoothing of the images can be employed with little fear of generating artifacts, increasing the sensitivity to a level comparable with much deeper surveys.

The footprint for the SDSS is shown in figure 3.2. Both the VC1 and VC2 areas (chapters 5 and 6) have been entirely observed and spectroscopy is also complete for these areas (complete in the sense that all the spectroscopic targets in the regions have been observed). Due to the size of the survey, these targets are necessarily identified by automatic source-finding programs (as are those for which automatic photometry is performed). Many HI-detected but optically faint galaxies thus remain without optical spectral measurements, being too faint for automatic identification.

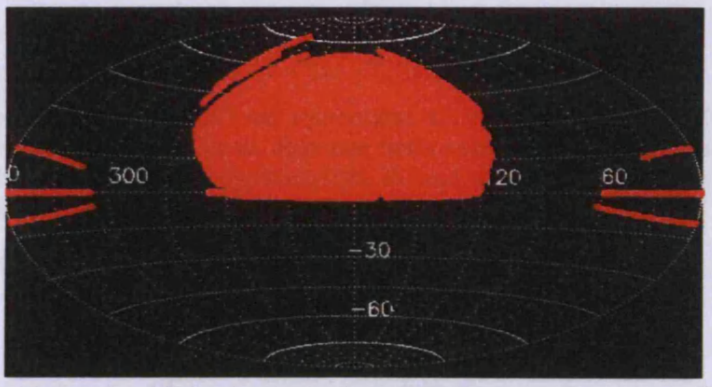


Figure 3.2: Footprint of the imaging coverage for the SDSS Data Release 7, galactic coordinates.

Similarly, a few galaxies identified by the HI survey are so optically faint that they have escaped the automatic SDSS catalogues altogether, though they may be visible in the images. Most, however, have at least an automatic catalogue number and automated photometry measurements. This photometry is usually in reasonable agreement with that from other methods (see section 3.5.2), but at low S/N levels, the area in which the light of the source is integrated can become confused with other sources, particularly foreground stars. For this reason automatic photometry is never used for data analysis.

One of the most useful aspects of the SDSS for AGES is the ease of access to the data. The ability to instantly view fully-reduced images of the appropriate size (i.e. the Arecibo beam size) centered on the coordinates of the HI, for any detection in Virgo, makes it the primary means of identifying optical counterparts. Identification of these

is described in section 3.5, but relies on cross-correlation with spectra from the SDSS and NED (see section 3.4), both of which are aided by the SDSS web interface.

The most important aspect of the SDSS is its uniform complete coverage for the regions studied in this thesis. Thanks to its sensitivity, every VCC (see section 3.4) object listed is visible in the SDSS data. Although automatic photometry is not used, manual photometry is used for SDSS data in the VC1 region (see section 3.5). It is also invaluable in providing optical redshifts for many of the HI detections. Spectroscopic coverage is not complete for every single galaxy, but it is a significant increase over that previously available.

A benefit of having the 5 different filters in the survey is to use them as follow-up for fainter sources. The scan through each filter is performed separately, so a source visible in two or more filters is likely to be genuine. As a further check, for HI detections not found in the VCC or NED, data from other optical surveys is also available. Most notable of these is the INT WFS.

3.2.2 The Isaac Newton Telescope Wide Field Survey

The INT WFS is a much smaller, more focused survey than the SDSS. Like AGES, its total area is about 200 square degrees, consisting of selected, separate areas chosen for scientific interest (such as individual galaxies, nearby and relatively high-redshift clusters). In addition to general studies of the Virgo Cluster its goals include galactic studies, identification of type Ia supernovae at intermediate redshifts, and generating the Millennium Galaxy Catalogue of local galaxies. Its footprint is shown in figure 3.3.

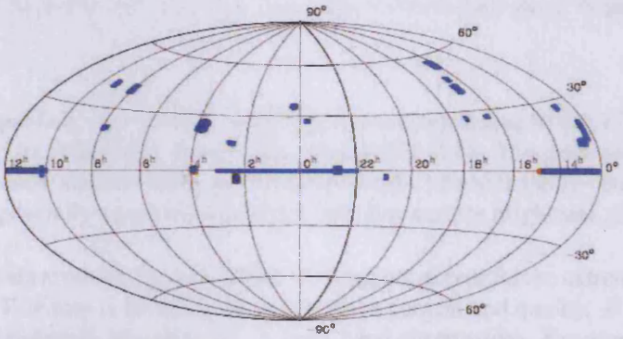


Figure 3.3: Footprint of the imaging coverage for the INT WFS, galactic coordinates.

The survey is performed using the 2.5m INT and commenced in 1998, running through 2005. Unlike the SDSS, different filter sets were used in different areas. Exposure times were generally much greater than for the SDSS, typically 750s in the B-band and 1,000s in the i-band (the B-band is roughly equivalent to the SDSS g-band, see figure 3.4). In most cases, this resulted in a real increase in sensitivity.

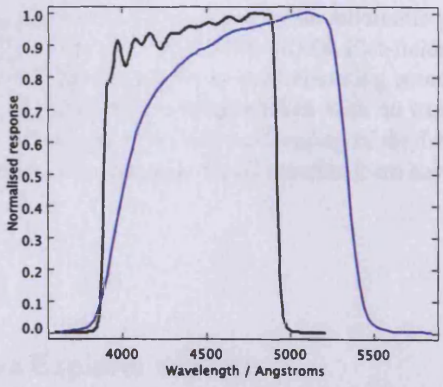
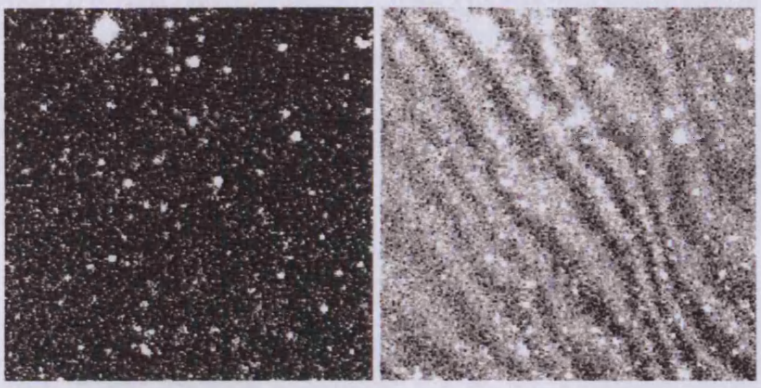


Figure 3.4: Comparison of the SDSS g filter (blue) with the INT B band filter (black).

The quality of the data for this survey is, however, much more variable than that of the SDSS. Observations in the B-band are generally of a very high quality, but the i-band filter causes visible "fringing" effects of vary severity. A particularly strong example is shown in figure 3.5. This problem can obviously negate the much longer integration times in this band (it should be noted that the fringing problem is very much reduced thanks to the data reduction (Irwin & Lewis 2001)). B-band sensitivity is estimated at 27 mag arcsec^2.



(a) B band **(b)** i band

Figure 3.5: Example of images from the INT WFS, of the same patch of sky, with particularly severe "fringing" effects in the i-band. Images are 3.5' across.

Most of the survey data is publicly accessible, including a strip corresponding to the VC2 region - one reason this particular area was selected for AGES was the overlap with the INT WFS. This area was fully observed as part of the survey, in the B and i bands, and is of high quality except for the i-band fringing - but this not usually as severe as in figure 3.5, and is not generally a problem except for very low surface brightness galaxies.

There is also some private data available for most of the VC1 region, except for the extreme south of the cube (south of 7 degrees declination). This data is far more variable both in content and quality. B-band data is available for the entire region, but some pointings also have Hα, R and i band observations. Exposure times vary. Much of the data is also not properly calibrated and lacks a zero point, and some images are dominated by obvious artifacts that render them unuseable. The data for this region has only been used to supplement the SDSS - to check if low surface brightness features are real, and to help search for faint optical counterparts.

Since the VC2 region has good quality, deep INT imaging available for the entire area, and was the first to be studied in this thesis, optical photometry for this region uses the INT WFS data. Optical photometry for the VC1 region uses the SDSS, owing to the many quality issues relating to the INT WFS for that region (it is also not uniformly available as the INT data does not extend south of 7 degrees declination, unlike AGES).

The publicly available data, at least, has been fully reduced by an automatic pipeline. The survey is pointed, not drift-scan, so the data reduction is quite different to that of the SDSS. Flat-fields are taken of the sky during twilight - these generate an image of uniform brightness with no contaminating sources present, and are used to correct for any variance across the CCDs. Bias frames are images taken with no exposure time, and are used to correct for instrumental noise. Algorithms are also used to remove fringing in the I-band images (Irwin & Lewis 2001). Although this is not as sophisticated as drift scanning, it still benefits from having a longer exposure time.

3.3 Other surveys

3.3.1 The Galaxy Evolution Explorer satellite

The GALEX satellite was launched in 2003 to survey the whole sky in the ultra-violet wavebands, specifically the far UV and near UV. These wavelengths cannot be observed from the ground due to atmospheric absorption. UV

emission is generally associated with young stars that are not hot enough to cause actual ionisation (e.g. Hα emission) - UV photons arise from electron transitions between bound states within the atom. The GALEX satellite has a 0.5m primary mirror giving resolutions of 4.3" and 5.3" in the FUV and NUV respectively.

The GALEX telescope is both a survey instrument and an active observatory, performing scientific studies ranging from high-redshift galaxy formation to observing individual star clusters at low redshift. Consequently the depth of the observations available of AGES detections varies. Typical exposure times are approximately 3,000s in the NUV and 1,500s in the FUV. This gives a sensitivity to about 23.5 mag. The field of view of the telescope spans 1.2° and both areas in Virgo targeted by AGES have been completely observed.

Like the SDSS, GALEX provides automated photometric catalogues but these are not considered sufficient for scientific analysis. Aperture photometry is instead performed as described in section 3.5. Unlike the SDSS, since UV emission generally comes only from young ($\lesssim 10^8$ Myr) stars it is not useful for identifying counterparts to the HI detections, as galaxies are generally less visible in the UV than the optical. Nor can it provide much morphological information on many of the fainter galaxies, owing to its poorer resolution. Instead its independent observations in two wavebands serve as an additional check on the colour of the detected objects.

3.3.2 The Arecibo Legacy Fast ALFA survey

As described in chapter 2, AGES represents the middle tier in a "wedding cake" of HI surveys at Arecibo. ALFALFA is the lowest tier, the largest area but lowest sensitivity. This sensitivity of 2 mJy is still much better than previous large-area HI surveys, as described in chapter 2. The survey will eventually cover all 7,000 square degrees of sky visible to the Arecibo telescope, with of course the same 3.5' spatial and 5 km/s velocity resolution as AGES. At present, in Virgo it has observed the entire cluster to a southern declination limit of +8°.

ALFALFA has completely observed the VC2 field, making it a very useful check that AGES source extraction procedures detected everything that they should be capable of (see chapter 4). ALFALFA's publicly available on-line catalogues do not include all of their lower S/N detections (see below) so this was supplemented with the catalogue data still requiring follow-up observations (courtesy Brian Kent and Martha Haynes, private communication). For VC1 there is only a narrow overlap region between the two surveys, but this is still enough for some interesting comparisons (see chapter 5).

For AGES, ALFALFA is inherently a very useful comparison survey since it is another blind HI survey using the same instrument on the same telescope. This makes it ideal for comparing the measured HI parameters such as flux and velocity width. As is discussed throughout chapters 4, 5 and 6 it is also interesting to search for sources that either survey may have missed. This is interesting both as a source extraction comparison and to see the effects of increased sensitivity on the number of detections. ALFALFA's detection "limit" is to a S/N level of 6.5, where S/N is defined (Saintonge 2007, Giovanelli et al 2007) :

$$S/N = \frac{1000 F_c}{W_{50}} \frac{w_{smo}^{1/2}}{rms} \tag{3.1}$$

Where for W50 < 400 km/s, $w_{smo} = W50/(2 \times v_{res})$, where v_{res} is the velocity resolution in km/s, and for W50 > 400 km/s, $w_{smo} = 400/(2 \times v_{res})$. F_c is the total flux in Jy; rms is the rms across the spectrum in mJy. This quantity is relatively complex compared to, say, peak S/N, which is simply peak flux divided by the rms of the spectrum (for this reason, S/N is taken to refer to peak S/N throughout this thesis unless otherwise stated). The advantage is that it is a quantitative estimate of the quality of the detection, providing an accurate measure of the average S/N.

Since both AGES and ALFALFA Hanning smooth the data to 10 km/s, to compare the relative S/N of a detection in either survey it is necessary only to substitute the rms value. As described in chapter 2, this does vary, being significantly higher at the edges of the data cubes for AGES. It is also important to note that their detection "limit" of 6.5 is not a strict cut-off - some weak sources are detected and catalogued below this limit. Details of the results of the comparisons between AGES and ALFALFA are described in chapters 4, 5 and 6.

3.4 Catalogues

3.4.1 The Virgo Cluster Catalogue

This indispensable catalogue of Binggeli, Sandage and Tammnann 1985 was constructed from a visual inspection of photographic plates. Cluster membership, which includes infalling clouds, is based on morphology. It covers an area of roughly 140 square degrees and is "essentially complete" for certain and possible cluster members to a B-band magnitude of 18. A total of 2,096 catalogued members are listed, though as discussed in chapter 1, only 854 have been spectroscopically confirmed as members; 904 remain as possible members without redshifts (assumed to be members throughout); the remaining 338 are spectroscopically confirmed as background objects.

The goals of compiling such a catalogue are numerous. The high sensitivity of the survey enables a very complete analysis of the cluster's components, from the giant galaxies to the far more numerous dwarfs. As outlined in chapter 1, knowledge of the distribution of the components of the cluster is essential in understanding its environmental effects. Crucially, the catalogue allows for studies at other wavelengths to obtain more detailed information (e.g. velocity structure, gas content, colour distribution etc.)

The exact area the catalogue covers has been shown in chapter 1. This area was observed on 67 photographic plates, each measuring 50cm$\times$50cm and spanning $1.5° \times 1.5°$ on the sky, taken with the du Pont 2.5m telescope. Exposure time was about 1 hour. The plates were repeatedly inspected by eye both with and without the use of an eyepiece. Cluster membership was determined by morphology and velocity data where available (which at the time was for 574 galaxies). Velocity data was given priority in the very few cases where it differed with the assignment based on morphology. The catalogue has 4 nominal cluster membership assignments : certain, possible, background and foreground, but in practice no cases of foreground objects were discovered.

Three morphological criteria were used to establish membership : surface brightness, resolution into knots, and luminosity class. Surface brightness can be used as there is a well-defined relation between surface brightness and luminosity. A low surface brightness object is much more likely to be a cluster object since it is also likely to be of low luminosity, and therefore would not be readily detected in the background. Resolution into knots is a straightforward effect of distance : star-forming galaxies appear much less resolved in the background than in the cluster.

Luminosity class is specific to spirals and depends on the thickness, strength and prominence of the spiral arms. A small, faint spiral with clearly defined arms cannot be a cluster member since such objects are generally massive - it is more likely to be a distant background object. Conversely a large, bright spiral with poorly defined arms must be in the cluster since such objects are usually less massive; in the background it would appear small and faint.

The catalogue itself lists only limited data : the position on the sky, apparent magnitude, morphological type, angular diameter and heliocentric velocity. This is useful to identify whether an HI detection is a known cluster member or not. Thanks to its high sensitivity, HI detections that have not been previously identified in the VCC are automatically interesting, demonstrating that not all cluster components can be found solely by optical surveys. Data from the VCC is supplemented with the GOLDMine database, which compiles information from multiple wavelengths.

3.4.2 The Galaxy Online Database Milano Network

This online database was begun in 2003 and is run by Gavazzi & Boselli. It holds data for the Virgo, Coma and Abell 1367 clusters. For Virgo, it contains not only the VCC data but also much other data from observations from various sources, mostly by those directly involved with GOLDMine. It provides access to reduced images at many different wavelengths, including optical, infra-red and Hα, and also spectral energy distributions and light profiles. These are of course all subject to availability - not every galaxy has been observed (or detected) in every waveband.

Use of GOLDMine has been key to several aspects of this thesis. Firstly it provides a morphological classification for each of the VCC galaxies. The classification scheme is based on that used in the VCC and is shown in table 3.1. Early-type galaxies are generally considered to be of type 1 or below (Alighieri et al. 2007) with the remainder as late-types. Type 20 is unclassified but seems generally to indicate a late-type morphology (see chapter 5). If a galaxy is given a type 20 classification, NED (see below) is searched to see if other authors have assigned a more

definite classification.

Table 3.1: Morphological classification scheme used in GOLDMine

Assigned number	Morphology
-3	dS0
-2	dE/dS0
-1	dE
0	E - E/S0
1	S0
2	S0a - S0/Sa
3	Sa
4	Sab
5	Sb
6	Sbc
7	Sc(dSc)
8	Scd
9	Sd
10	Sdm-Sd/Sm
11	Sm
12	Im(Im/S)
13	Pec
14	S/BCD (dS/BCD dS0/BCD Sd/BCD)
15	Sm/BCD
16	Im/BCD
17	BCD
18	S(dS)
19	dIm/dE
20	Unclassified

Another benefit of GOLDMine is that it gives distances to cluster members in a way that at least attempts to account for the complex structure of infalling clouds. As shown in chapter 1 the boundaries between each cloud are not obvious, so there will inevitably be some error in this process. Observations by Gavazzi et al. 1999 have shown that the actual distance is more subtle than the mean redshift of each cloud. As it is impractical to estimate distances for each of the 2,096 VCC members, Gavazzi et al. determined distances for a total of 134 galaxies, 59 early-types (using the fundamental plane) and 75 late-types (using Tully-Fisher). Distance to each cloud (given in the GOLDMine database) is based on distance determinations of the individual galaxies within each cloud for which such measurements exist.

Membership of each cloud was determined, as with cluster membership, by spatial distribution and morphological classification (for example de Vaucouleurs 1961). Such methods are obviously subjective. Given the limited number of galaxies with individual distance determinations, it is difficult to estimate the distance each cloud spans along the line of sight. Mei et al. 2007, for instance, estimate a depth of the main cluster of 2.4 Mpc, and find that the A and B subclusters are both at around 16.5 Mpc distance while the W cloud is at 23 Mpc. This is in contrast to GOLDMine, which, as shown in chapter 1, assigns distances of 17, 23 and 32 Mpc for the A, B and W subclusters respectively.

For these reasons the error on the distance to each galaxy is considerably uncertain. In this study the distance assigned in the GOLDMine database is used, except where otherwise stated (Mei et al. were primarily concerned with the main body of the cluster and do not attempt to define cloud boundaries). When a galaxy is detected in AGES that is not listed in the VCC, cloud membership is assigned (and hence distance) based on its spatial location. Its location in position-velocity space is also considered within the VC1 region, which spans 3 clouds - see chapter 5.

GOLDMine is also useful by providing comparison data, both for optical photometry and for HI. As discussed in chapter 2 there is some subjectivity involved in the analysis of HI data and the same is true of optical photometry (see section 3.5). Thus to avoid any conflicting observational biases by the use of different observers, only the AGES data is used in the actual analysis, to ensure consistency.

In addition to the HI data and optical data, the most useful data from GOLDMine are the spatial position and velocities. There is occasional conflict between the velocity data from AGES, GOLDMine, SDSS and NED. Determination of the optical velocities of these galaxies is discussed on an individual basis (see chapter 5) but largely

depends on which catalogues agree most closely.

3.4.3 The NASA Extragalactic Database

NED is a very much larger project than GOLDMine. It contains data from galaxies and other objects across the entire sky - 163 million in total - from any available sources. This means, for example, that while it may contain some basic photometry that data may be from observations taken many years ago that have been superseded by better quality observations. Since it contains objects as wide-ranging as nearby nebulae to high-redshift galaxies, and many objects are found in different catalogues, its data coverage is far from uniform in any sense.

This, however, is not necessarily a disadvantage. While the VCC (and thus GOLDMine) only lists galaxies in Virgo to a limiting photographic magnitude, NED, essentially, lists everything that anyone has ever found in that region by any survey - or at least those that the NED team have been able to access. This makes it an important part of identifying possible optical counterparts to AGES detections, since many HI detections are too optically faint to be recorded in the VCC (see section 3.5 and chapters 5 and 6). This is especially important when identifying objects behind the cluster, for which the VCC was not designed and no comparable, well-defined catalogue of objects exists.

As has already been mentioned NED is sometimes used to assign a morphological classification for objects remaining unclassified in the VCC. It is also used to check velocity information, on the occasions when no spectra have been recorded in the SDSS or GOLDMine. Velocities from NED are only used in those instances where other, more uniform sources fail. Photometry from NED is not used since there is too much variation in where the measurements originate. NED also contains some redshift-independent distance measurements, but these are too few in number to be very useful. It is however invaluable in finding previously published papers referencing individual objects.

3.5 Analysis methods

3.5.1 Identifying optical counterparts

The coordinates of the HI detections are found using *miriad* as described in chapter 2. These are used to generate finding charts of 3.5' radius (the Arecibo beam size) centered on those coordinates from the SDSS, which are then inspected by eye for any possible optical counterparts. If available images from the INT WFS are also searched. Candidates are searched through GOLDMine, NED and the SDSS for any corresponding optical redshifts, and accepted if that agrees with the HI redshift to within 200 km/s. This is a fairly high velocity width, with about 75% of detections being less than this. This means it is large enough to account for the velocity width of the HI detection while small enough to avoid spurious, coincidental matches. In fact the difference is usually much less, as shown in chapters 5 and 6. Object with redshifts differing by more than this are entirely excluded as being associated with the HI (except in a very few cases discussed in chapter 5).

In most cases there is only a single counterpart with a matching redshift within the Arecibo beam and these are accepted as being uniquely associated with the HI. In a very few cases there may also be possible counterparts without redshift measurements - the association is flagged as uncertain. This is also done for those few cases where there are multiple counterparts with matching redshifts. These cases are discussed individually in the results chapters where appropriate.

The positional match of the optical and HI is usually very good, but there are occasional discrepancies when the HI S/N is poor and/or the HI detection itself is unsure. Again these cases are discussed individually. In a very few instances, the HI may have no clear optical counterparts. In this case NED and GOLDMine are used to generate lists of any galaxies within the Arecibo beam known to be present but not clearly visible in the optical, and these candidates are then inspected. Sometimes there may be objects which are no more than faint smudges usually attributed as background objects, but are the only possible counterparts. The association is flagged as uncertain. Uncertain associations are treated in the analysis as genuine (there are so few of these that they do not greatly matter to the overall conclusions).

The flagging notation for the optical counterparts is based around a search for dark galaxies. Thus, 0 indicates that there is at least 1 counterpart with a matching optical redshift; where multiple matches exist this is indicated in the data tables in the name of the source from existing catalogues. 1 is used to denote a source where these are obvious possible counterparts but none have available optical redshift measurements. These are usually unique counterparts and so are treated as genuine throughout. 2 indicates that the association is very uncertain with possibly no real optical counterpart at all.

It is important to note that the optical counterparts are almost entirely identified on a visual and therefore subjective basis. Similarly the choice of whether an association is very uncertain is entirely subjective and cannot be readily quantified - there is no precise definition on what constitutes a small faint smudge.

3.5.2 Photometry

Aperture photometry

Once the optical counterpart(s) has been identified, its optical magnitude can be computed. This is generally done through aperture photometry using *ds9* with *funtools*. An area, or aperture, is defined which encloses the optical source and background apertures are set which only contain sky. If it is not possible to avoid contaminating objects such as foreground stars, apertures are defined around those objects as masks, and the counts within these areas ignored. The net photon count is that within the source aperture minus the mean count in the background apertures, ignoring those regions which have been masked. An example is shown in figure 3.6.

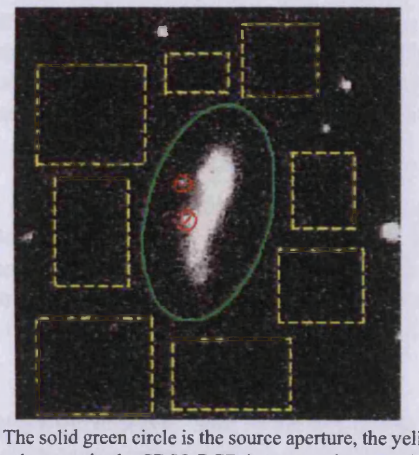


Figure 3.6: Example of aperture photometry. The solid green circle is the source aperture, the yellow dashed boxes are the background regions. The red circles with diagonal lines are seen to be stars in the SDSS RGB images and are masked (the image contrast may be freely varied, making the stars easier to discriminate from the source than is apparent here).

The major disadvantage to this method is that it is subjective, since the size of the source aperture is defined by where the observer thinks the source ends. The approach taken is to generally increase this aperture slightly as enclosing slightly more than the source should not increase the net photon count, provided no additional sources are included and the flat-fielding is good. Masking can be more problematic since it can be difficult to judge where a bright foreground star ends and the extragalactic source begins. It is often necessary to compare the fits files with the SDSS RGB images, which show more obviously which objects are distinct from the main galaxy (e.g. by differences in colour).

There are however several compensating advantages. The method is very fast and generally easy, which is important with a large sample size (well over 300 objects in this study). There is no lower limit to the magnitude of a source that can be measured (though obviously the errors will be greater for faint sources since it is harder to define where the source may end). Only 1 observer performs the photometry, eliminating different biases inherent to different observers. Colour should not suffer significantly since the apertures defined in 1 waveband can be used for the other, only requiring slight adjustments where the coordinates of the images are offset. In any case, calculating the HI masses is inherently slightly subjective both from choosing the data profile, and, within the cluster, the chosen distance assignment, and there is no way to eliminate this.

Once the net count has been estimated it is converted into a magnitude by the following equation :

$$m = -2.5 \times log\left(\frac{counts}{exptime} \times 10^{0.4\times(aa+kk\times airmass)}\right) \tag{3.2}$$

Where $counts$ is the integrated counts in the aperture after subtracting the background value, $exptime$ is the exposure time in seconds of the image, aa is the zero point of the image, kk is the extinction coefficient, and $airmass$ is simply the airmass. For GALEX, the images are already corrected for the exposure time, and kk and $airmass$ are of course zero. The equation is then simplified to :

$$m = -2.5 \times log\left(counts \times 10^{0.4\times aa}\right) \tag{3.3}$$

Aperture photometry was performed on HI detections within and behind the cluster and for non-detections within it, using SDSS, INT and GALEX data. For VC2, where the samples were much smaller, surface brightness profiles were constructed. These can give less subjective and more accurate results by measuring where the light of the sources fades into the sky background. The disadvantages are that it can be much more difficult to measure faint sources since it is harder to fit a profile, and the procedure is generally a lot more time consuming than aperture photometry.

Surface Brightness Profiles

Profiles are constructed using IRAF. The tasks *marksky* and *skyfit* are used to subtract the sky background and *cphot*, *isoimap* and *fitprofile* are used to fit the profile after masking contaminating sources. The user specifies the approximate center, ellipticity and position angle of the galaxy and specifies whether these should be allowed to vary or not. IRAF then displays the measured light curve and the user selects what type of profile should be fitted : an exponential, de Vaucouleurs, truncated de Vaucouleurs or a bulge plus disc (de Vaucouleurs for the bulge with a truncated exponential for the disc).

Once the profile is determined there are a number of choices as to how to compute the magnitude. The simplest is to integrate the observed flux out to the last isophote, in a sense reducing the procedure to aperture photometry (isophotes are generally computed out to a S/N ratio of 1.0). Another is to take the integrated observed magnitude and use the profile to calculate how much light is beyond the last isophote where actual measurements are impossible. A third method is to use only the profile to compute the magnitude, which ignores any asymmetries in the observed galaxy.

The method used was to simply integrate the observed light to the last isophote. Many galaxies detected in HI are optically faint and irregular making an accurate estimate of the profile difficult, especially if close to an optically bright galaxy. In a few cases the extrapolated and integrated magnitudes differ by as much as 3 magnitudes (these cases are faint dwarf galaxies or galaxies in close proximity to one another, where the profile cannot be accurately fitted), while the typical difference is only about 0.1 magnitudes - so small as to be practically negligible. This method still has the advantage over aperture photometry of being computed to an objective limit (S/N of 1.0) and also gives the radius of this limit.

The magnitude calculation is done by the same calculation as for aperture photometry, the only difference that this is done directly by IRAF which finds the parameters through the fits header. It should be noted that while surface brightness profiles are more objective than aperture photometry, the choice of what profile to fit and what method to use to estimate the magnitude mean that this procedure is not entirely free of subjectivity either.

Absolute Magnitudes and M_{HI}/L ratios

The absolute magnitudes are calculated from the distance modulus equation which can be re-arranged to give:

$$M = -(5 \times log(d) - 5)) \tag{3.4}$$

Where d is the distance in parsecs, as described in section 3.4.2 for cluster members. Objects behind the cluster

are assumed to be in pure Hubble flow, so :

$$d = \frac{v}{H_0} \tag{3.5}$$

Where H_0 is assumed to be 71 km/s/Mpc.

The HI mass calculated from equation 3 in chapter 2 is in solar masses. The M_{HI}/L ratio is not intended as a very accurate estimate of the HI mass to true luminosity; it is only designed to indicate whether a galaxy is very gas rich or poor. The luminosity is thus estimated from the absolute magnitude, assuming all of the stars to be identical to the Sun. The absolute magnitude of the Sun is assumed to be 5.45 for the B-band and 5.33 for the g-band (Willmer, http://www.ucolick.org/~cnaw/sun.html). Since the absolute magnitude of the detections is known, their absolute luminosities can be calculated by :

$$\begin{aligned} M &= k - 2.5log(L) \\ \Longrightarrow L &= 10^{\frac{M-k}{-2.5}} \end{aligned} \tag{3.6}$$

It then follows that the ratio $L/L_\odot$ is :

$$\frac{L}{L_\odot} = 10^{\frac{M-M_\odot}{-2.5}} \tag{3.7}$$

Identification of the optical counterparts is relatively straightforward, particularly given the numerous data sources. Searching optical images within a 3.5' radius is a relatively painless exercise. This typically involves looking at a finding chart image about 640×640 pixels in size (400 thousand pixels), where there are usually only a handful (< 10) of obvious galaxies detected. In contrast, a 5 square degree AGES data cube - despite the much lower spatial resolution - contains upwards of 10 *million* pixels, and has the added complication of the third axis. Many detections lack any corroborating HI data. The source extraction procedures are thus quite different, and these are discussed in the next chapter.

Chapter 4

Source Extraction

4.1 Current methods of source extraction

Source extraction from 3-dimensional data cubes remains a difficult process. There is no program equivalent to SEXTRACTOR (widely used in optical data) that can be used to give an unambiguously complete, reliable catalogue from an HI data cube. "Complete" here means that all sources of specified criteria are detected, "reliable" that only genuine sources are detected. While automatic extractors do exist, their output is often neither complete nor reliable, especially at low S/N ratios, which can include some of the more interesting sources (such as those without optical counterparts).

Estimating completeness and reliability can only be done approximately for real data, as there is no existing method proven to extract all HI sources, even to some well-defined criteria (the reasons for which will become apparent in the discussion below). This means that even the number of potentially recoverable sources can never be known exactly. There are also some spurious detections whose characteristics so closely resemble real sources (see figure 4.5, discussed in more detail in section 4.2.1) that if they were missed, real sources would also be missed. In this sense they could be included as potentially recoverable sources. This is generally assumed to be the case in this analysis, unless otherwise stated.

An automatic extractor, in principle, is able to detect every source above some specified threshold, ideally total flux (unfortunately this does not always work well in reality, as discussed below). This enables construction of catalogues that are complete to some mass limit, just as optical catalogues are complete to a given apparent magnitude (or absolute magnitude at a particular distance). In contrast, a visual extraction can never be said (rigorously) to be complete in any sense. Examination of a cube by eye and deciding what should be included as a source, is, fundamentally, a subjective process. The benefit of a visual extraction is that it is often much more reliable than an automatic extractor, the human eye being good at distinguishing real sources from spurious ones.

AGES, to date, has used a combination of strategies in generating source catalogues. A visual extraction is performed by two observers, and the automatic extractor Polyfind and/or Duchamp is employed. A source is counted as "sure" if it is detected by two or more methods, i.e. detected by both observers, or one observer and the automatic extractor. If only detected by one method a source is considered "unsure" and scheduled for follow-up observations. Follow-up is also allocated for "sure" sources if there seems a good reason to do so, for example a detection with no obvious optical counterpart. Sources from the automatic extractor must also be examined individually by eye, as both automatic extractors generate large numbers of spurious sources (discussed below). Thus the whole process of source extraction is only semi-automatic at best.

Another method, less commonly used, is to use optical catalogues to search the positions of galaxies, effectively transforming AGES into a pointed survey. This is of particular benefit when an optical redshift measurement is available. In this case, it is sometimes possible to identify a likely signal that would be too weak to detect either visually or automatically. The focus of this chapter is generally on source extraction independently of other available data, and detections made in this way are not considered unless stated otherwise.

4.1.1 Visual Extraction

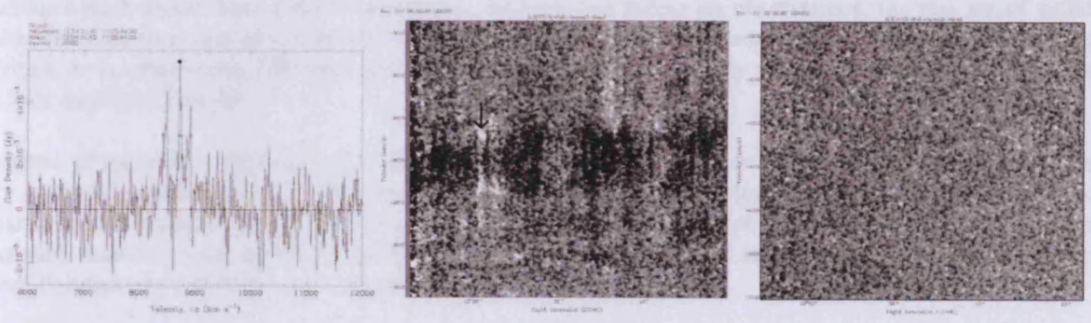
A visual extraction is a straightforward if time-consuming procedure. The three projections of the cube are searched independently by scanning over the appropriate axis. For example, in the right ascension - declination projection, the user will scan through each velocity channel. This is typically done at least twice for thoroughness. A source list is generated simply by recording the position and velocity of any apparent source as they become visible. The source lists for each projection are then cross-matched. Any source that is not detected in at least two projections is checked and removed if it is judged to be false. Checking is done by a repeat examination of the cube and also extracting the spectrum of the source using *mbspect* - as this integrates the source spatially it can increase its S/N.

Searching the three projections ensures some measure of independence to the searches, and also means the cube is searched many times. This increases the chances of detecting the weaker genuine sources (as they may be missed in a cursory glance at the cube), while also highlighting sources that may be spurious. If a source is only detected in one projection, it is usually weak and often found to be spurious on a more detailed examination. Conversely if it is detected in all 3 projections it is likely a bright source, which are usually genuine. Having two observers generate source catalogues independently helps to reduce biases that individuals may have in determining whether a source is real. It also provides even greater thoroughness to the search.

A visual extraction possess an intrinsic advantage over any automatic extractors (completeness and reliability aside). A human examining the cube will look at the context of each possible detection, not just the individual spectra. It is therefore much easier for a human to distinguish false detection from RFI than for any automatic extractor. This is illustrated by figure 4.1, which shows the spectrum of a spurious source and a position-velocity slice through the cube containing the apparent source. A slice through a region uncontaminated by RFI is also shown for comparison.

The difference is obvious - the apparent source spans the same frequency range as an area of RFI, and the whole area is much less uniform than the blank area. When scanned in declination, it is even more obvious that the"source" cannot be real, as innumerable others also appear with similar characteristics (for example, spanning exactly the same frequency range and appearing very strongly double-horned). It is very much harder to determine that the source is actually RFI from the spectrum alone, as all automatic extractors must do. Since this is so difficult, regions contaminated by RFI are usually masked when using an automatic method.

Figure 4.1: A spurious source resulting from RFI.



(a) Spectrum of a spurious source

(b) Position-velocity slice containing the source, highlighted by the arrow

(c) Position-velocity slice with no RFI present

A visual extraction can be more successful in finding real sources in regions contaminated by RFI. These tend to be relatively isolated and sharply defined, whereas the spurious sources often appear in clumps and appear more diffuse - none of which is obvious from spectra alone. Overlapping sources are also significantly easier to distinguish by a visual examination of the cube rather than from spectra. An example of both of these effects is shown in figure 4.2. These particular sources were identified as only a single source by ALFALFA (which relies on an automated extractor). That they are in fact distinct is confirmed by the fact that both have sure, separate optical counterparts.

For a cube $\sim$10 square degrees, a visual extraction will generate a list of $\sim$100 possible sources and take several

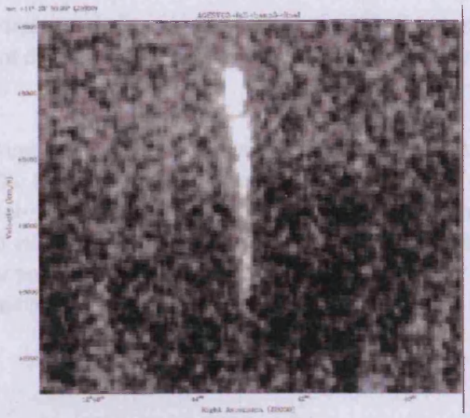


Figure 4.2: Position-velocity diagram of two overlapping sources in VC2. The wider source extends significantly into a region of RFI, yet is clearly (to a human observer) a real source.

days to complete. For VC1, a total of 128 candidates were listed, of which 123 were judged to be sure detections. This compares to a final list of 150 possible sources after other extraction methods had been performed. This is far superior in terms of reliability than the automatic extractors : Polyfind generating a list of ~3,000 sources, Duchamp a few hundred (most of these, as discussed below, are spurious).

Strictly speaking, to be certain a source is genuine or spurious requires follow-up observations. However, there are many cases where follow-up observations would be a waste of telescope time. A 10σ double-horned HI profile with a sure optical counterpart is very unlikely to be spurious. In contrast, an HI profile with a very narrow velocity width and, say, of only 4σ strength with no optical counterpart is unlikely to be genuine. This describes the vast majority of detections from the automatic extractors - their comparatively large output list is due to unreliability, rather than detecting a population missed in the visual extraction. This poor reliability is in large part due to the sheer number of pixels in an AGES data cube : $\sim 1 \times 10^8$ for VC2. Assuming the noise to be Gaussian, more than 3,000 pixels should be above 4σ by chance alone.

A visual extraction is of course not 100% reliable either. Only a few detections in VC2 have yet had follow-up observations, though all of the detections in VC1 requiring follow-up (25 sources) have been re-observed. 122 candidates were found from a visual inspection. 15 have had follow-up observations, the rest are of sufficient quality (or possess a sure optical counterpart) that follow-up is deemed unnecessary. Only 5 were found to be spurious, or in other words 96% are reliable. A visual extraction is clearly vastly more reliable than automated methods employed thus far.

Estimates of the visual extraction reliability are very similar for the smaller VC2 cube. 41 extragalactic sources were extracted visually, of which 67% have sure optical counterparts. If detections with probable optical counterparts are also included (i.e. an obvious galaxy close to the HI coordinates but without an optical redshift measurement), reliability rises to 92%. Considering the genuine sources (by the same criteria) the automatic programs detect, Polyfind has a reliability of $\leq$5%, Duchamp rather better at $\leq$20%.

Completeness of the visual extraction is harder to estimate, since there are no well-defined criteria for what the human eye can detect. The simplest approach is to consider the fraction of the final source list (after using both visual and automatic techniques) that were detected visually. For VC1 this is 81%, for VC2 77%. These completeness levels compare well with an independent assessment of the completeness of AGES. Schneider et. al. 2008 (hereafter S08) introduced artificial sources into a real data cube in order to more accurately quantify completeness and reliability of the visual extraction method.

400 sources were introduced into a set of real data and two observers performed a visual extraction. An automated extractor was also employed. 162 sources were recovered by at least one detector. This, however, does not mean that the completeness of the survey is no more than 41%, as not all the sources are potentially detectable. In any real survey area, there are undoubtedly some HI sources present that are below the sensitivity limit and thus never detectable. The 162 recovered sources may therefore be treated as being all of the detectable sources.

In this sense, visual extraction performs very much better, with about 60% of the injected sources detected. 65%

of the sources recovered overall corresponded to injected sources, with the remainder being spurious detections. With a requirement of 3 independent detections (2 human observers and one automatic program), reliability reaches 95%, with completeness of 68%.

S08 found that the best method overall is to use a combination of a single human and an automatic extractor. In this way, completeness rose to 80%, though reliability reduces to 80%. It is of course much harder to accurately estimate the completeness and reliability of real data, and clearly, achieving something closer to 100% completeness and reliability is very difficult except at high S/N levels. For a visual extraction, there are many reasons why sources may be missed, bias on the part of the human observer by no means the least of these. To understand why automatic extractors do not fare much better, and to try and improve them, it is necessary to understand how they operate.

4.1.2 Polyfind

Polyfind is a source extraction program written by J.I. Davies, R. Minchin and D. Zambonini (Davies et. al. 2001, hereafter D01). Its basic operation is as follows. Each spectrum in the cube is extracted with a first-order baseline calibration applied and its rms measured (regions of RFI are masked). The individual spectra are searched for peaks of a specified S/N ratio, usually 4. If a spectrum contains such a peak, it is convolved with various model spectra which have the form of Gaussian profiles of varying widths. The template that fits the data best is used to derive the central velocity, velocity width and total flux, provided there is a template for which the correlation coefficient exceeds 0.75.

Polyfind has been employed successfully on HIPASS data for which it was designed. D01 describe how it is able to reduce the initial 2,435 spectra to a list of 155 candidate HI detections (6.4% of the original sample). AGES data cubes, however, typically contain a factor 10 more spectra. For VC2, with 320×80 pixels, there are 25,600 spectra. Polyfind reduces this to 1,619 candidate detections, a 6.3% detection rate consistent with that for the HIPASS data. One motivation for an automatic extractor for D01 was to avoid a visual inspection of 2,435 spectra. With AGES data, the number of candidates remaining for a visual inspection is little better than this.

Polyfind's reliability is very poor - only about 2% of the candidate list corresponds to detections verified as real (see Discussion). Although time-consuming, it is nonetheless worthwhile to search the cube both visually and with the aid of Polyfind, as both methods detect potential sources that the other misses. The ideal automated extractor would detect all sources to some flux limit. While this is very difficult to achieve, a more practical goal would be to have an automatic extractor that detects every source found by eye as well as other potential candidates. This would at least avoid the need for a visual extraction, even if the source list to search was still large. Unfortunately, Polyfind is not able to fulfil this requirement.

Polyfind and a visual extraction both found 6 possible sources that the other method missed. It is relatively easy to explain the detections made by Polyfind that were missed by eye, since a visual extraction is subjective and the sources missed are relatively weak. However, 4 sources missed by Polyfind have a S/N greater than 4 and so should have been detected (incidentally, 3 of these have sure optical counterparts and the 4th a single probable counterpart, so they are all very likely real HI detections). The problem is the S/N measurement.

Polyfind treats each spatial pixel as containing an independent spectrum, though in fact AGES data has pixels of 1' but a beam size of 3.5'. Consequently, Polyfind overestimates the rms, typically reporting a value of 0.8 mJy whereas *miriad* usually reports 0.5 mJy. In effect, what Polyfind measures to be a 4.0σ peak may in fact be a 6.4σ peak (with some variation due to natural variance in the rms of each particular spectrum). The strongest source missed by Polyfind has an actual peak S/N of 6.0, so the effect of overbinning is sufficient to explain all of the sources that should have been detected.

Polyfind is thus able to detect most sources that a visual extraction can make as well as several others. It could be improved for AGES data by correcting its measurements of the S/N - in this way, only 2 sources in VC2 found by visual extraction would have been missed by Polyfind. Both of these are very weak and probably spurious. 4 sources would still be missed in VC1, but this is less than 3% of the sample for that region. 3 other sources would also have been missed in VC2 - 2 are in a region masked due to RFI, and another 2 overlap in space and velocity and are hard to distinguish in any automatic extractor. The main practical issue with Polyfind is its poor reliability.

Polyfind's template matching is perhaps unnecessary. Its purpose is to enable a better automatic estimate of source parameters (e.g. total flux, velocity width, etc.). In practise, it is almost always necessary to adjust the velocity profile Polyfind suggests, as it often truncates wider sources and overestimates the width of narrow detections. Single-channel spikes, for instance, are measured by Polyfind to have a width of 40 km/s, though the channel width is actually 5 km/s (resolution of 10 km/s with hanning smoothing). Template matching is also of little use in source extraction, as altering the minimum correlation coefficient required makes no appreciable difference to the number of sources detected.

Since the program searches for the template with the best match, one or more model will always be a reasonable fit to almost any possible detection (the distribution is shown in figure 4.3). This includes "spikes" in the noise that are typical of the spurious sources reported by Polyfind. With a coefficient of 0.75, the number of sources reported when Polyfind is run on the standard VC2 cube was 1619. With the coefficient lowered to 0.00, the number increased only by 4. However, requiring a better match and raising the coefficient to 0.9 only reduces the number of potential detections to 1,244, and also excludes 18 real sources that were previously detected.

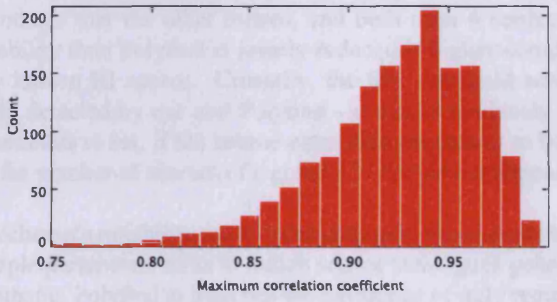


Figure 4.3: Distribution of the maximum correlation coefficient from template matching by Polyfind.

However well any automatic extractor performs in comparison with a visual extraction, it is still useful to compare its results with an independent program. If different methods of source extraction detect different populations, then clearly both are in need of improvement. If, however, one method does not detect all sources found by another and does not detect any new sources, that method may be rejected.

4.1.3 Duchamp

Duchamp is a program written by Matthew Whiting. Its basic procedure is rather different to Polyfind and it possesses a number of advantages and disadvantages. At the simplest level, the program first reads in the cube and establishes the rms. The cube is then searched two-dimensionally (one channel map at a time) for peaks of some S/N threshold based on the rms. If a peak is above the threshold, pixels adjacent to it are also checked. A source is accepted provided there are more than a specified number of contiguous pixels all above the S/N criteria (or, optionally, above an absolute flux threshold instead). Pixels may be contiguous in space and/or velocity, but no baseline calibration is performed[1].

One problem that becomes immediately apparent is the calculation of the rms. Since Duchamp operates on channel maps rather than spectra, the rms is not calculated for each spectrum as Polyfind does. Rather, the rms is estimated either from the flux distribution in the entire cube or in a user-specified subset region - neither of which provide much accuracy when estimating S/N, since they do not account in any way for spatial variation in the rms. If the entire cube is used, the rms will be overestimated, as it contains not only blank areas but also HI sources, continuum sources, and RFI. This means that all sources will have their S/N underestimated - a particular problem for sources close to the 4σ detection limit.

The alternative approach of selecting a blank subset region with a more typical rms throughout is little better. In this case, regions with a particularly high rms will have very high measurements of S/N. This is a severe problem

[1] It is possible to have Duchamp calculate and remove the baselines, but this is not done by a simple, first-order approach and often removes real sources.

for the edges of the cube. Since ALFA's beam pattern is hexagonal, the edges are of significantly higher rms than the interior of the cube. This means essentially every pixel at the edges will be calculated to have a very high S/N.

In this regard Polyfind's method of measuring the rms along each spectrum is far superior. In order for Duchamp to even operate at a practical speed, it is necessary to mask the edges of the cube quite heavily (if this is not done, the program not only operates far more slowly but also produces a source list with many thousands of candidates). In a very large cube where few of the sources are found near the edges, this would be of little consequence. In the case of VC2, 5 sources are missed (10% of the sample) including very bright sources that are instantly obvious to the eye.

Due to the problems with S/N, better results are obtained with a fixed absolute flux threshold. If a 4σ cut is applied, Duchamp generates a candidate list of 119 sources, of which 22 were detected by other methods. It does not detect any new sources, most of the remaining 97 candidates being spikes in the noise or baseline ripple[2]. Using a flux threshold of 3 mJy is more successful - a candidate list of 492 objects is produced, of which 39 correspond to known HI sources.

Duchamp does not detect all of the sources found by Polyfind or by eye, even in the regions it can search. Duchamp and Polyfind both detect 2 sources that the other misses, and both miss 4 sources that are found only by eye. Duchamp's much higher reliability than Polyfind is greatly reduced if higher completeness is required, and even then it does not detect every known HI source. Crucially, the flux threshold was based on the peak flux of a relatively weak source already detected by eye and Polyfind - it would obviously be preferable not to have any prior knowledge of what parameters to set, if the source extraction process is to be truly automatic. S/N ratio is preferable to absolute flux as the number of sources of a given S/N should not depend on the actual rms of the data.

It is possible to improve on Duchamp's reliability and/or completeness, for example by employing hanning smoothing. However, adjusting multiple parameters so as to match source catalogues generated by other methods makes the process very far from automatic. Polyfind at least has the advantage of only requiring one input parameter - the S/N threshold, which can be set without any prior knowledge of the contents of the cube.

It is also worth commenting that searches of the cube by any method can never be entirely independent. For example a visual extraction will easily detect a high velocity cloud due to its extended nature, though its spectral profile may resemble the "spike" typical of a spurious source. Due to the rarity of extended features, and the unusually low velocity, a human may well remember the source when viewing its spectrum from an automatic extractor and choose to accept it - even though such a source might normally be rejected.

There are also other reasons why subjectivity cannot be eliminated even using automatic extractors. The format of the output is more than a matter of aesthetics - it can actually influence the choice as to whether a source is accepted or not. Duchamp produces a postscript file containing all of the candidate's spectra. While this makes viewing the candidates straightforward (the user need only open a file), the images are small and no baseline calibration is performed. This can make it difficult to judge if a candidate is real. Indeed, sources found by other methods are occasionally missed when viewing Duchamp's output, even though Duchamp does actually detect them.

Polyfind does not produce output spectra directly but generates a text file containing parameters for input into *mbspect*. This makes it relatively easy for a user to alter the inputs, if necessary, to determine if a source is real. It suffers from the sheer number of candidates generated. When searching a list of 1,600 candidates, the user falls into a copy-and-paste routine where the spectra are often only given a cursory glance before the next candidate is examined. This can and does lead to some sources being missed, as they were for Duchamp, though only due to the number of candidates rather than the output format.

4.2 New methods of source extraction

4.2.1 Individual polarisations

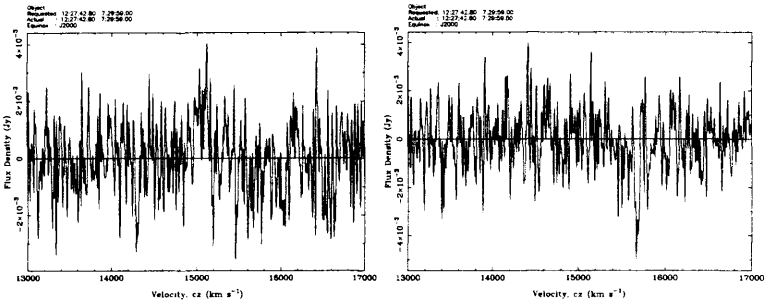
ALFA has 7 beams each receiving 2 linear polarisations separately. The standard approach with AGES is simply to grid the average of both, but it is also possible to grid each polarisation individually. Since the 21cm line is

[2] As with Polyfind, the automatic estimate of total flux and velocity width is not at all accurate, making it difficult to quantitatively assess the spurious detections.

not inherently polarised emission, a source that only appears in one polarisation must be spurious. If sufficiently strong it may nonetheless appear in the averaged data cube. Griding the polarisations separately can then be of help if not with source extraction, then source verification.

At the most basic level, each source extracted by other methods can be examined in each polarisation as a check on its validity. This can certainly help eliminate spurious detections - with this approach, 7 sources were rejected from a preliminary catalogue of VC1 while 3 were rejected from VC2. As an extra check, a search was made for possible optical counterparts, but no plausible candidates were visible. An example of a source rejected due to polarisation is shown if figure 4.4.

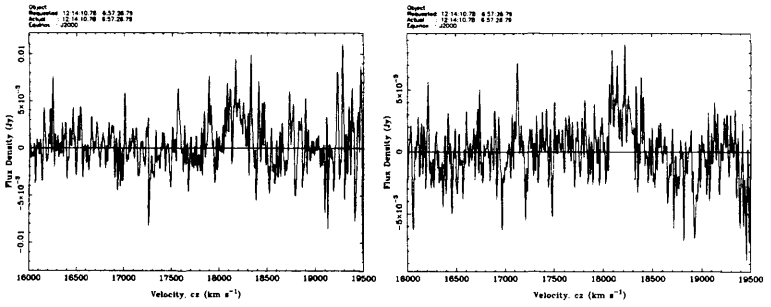
Figure 4.4: A candidate detection rejected due to polarisation.



(a) Polarisation A **(b)** Polarisation B

There are caveats to this approach. It is possible for a spurious source to be clearly visible in both polarisations, and conversely, real sources may occasionally appear quite different in the individual polarisations. An example of the first is shown in figure 4.5. This source was identified both visually and by an automatic extractor, and given its strength (peak S/N of 6.1) and velocity width (W50 of 350 km/s) it would ordinarily be regarded as a secure detection. However, although it is apparently very HI massive ($\sim 2 \times 10^{10}$ $M_{\odot}$) there is no sign of a plausible optical counterpart - even the most likely smudge would still require an extraordinarily large M_{HI}/L_G ratio. For this reason follow-up observations were obtained and the source was not visible. Since there is no way to tell from the spectra alone that this source is spurious, follow-up observations will always remain essential - 100% reliability cannot be achieved except at much higher S/N levels.

Figure 4.5: A spurious source visible in both polarisations.



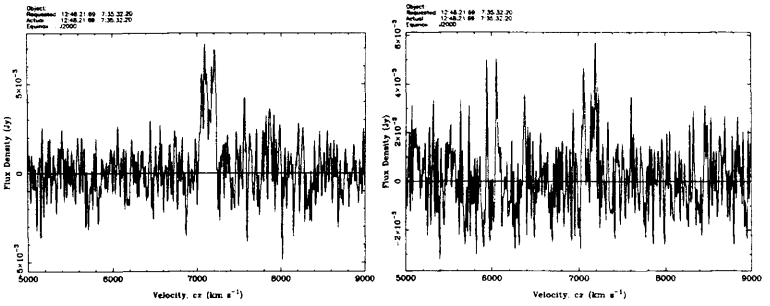
(a) Polarisation A **(b)** Polarisation B

An example of the converse situation is shown in figure 4.6. This source contains almost 4 times as much flux in one polarisation than the other, though it is visible in both. Although it displays a double-horn profile typical of a rotating disc, the possible optical counterpart identified from the SDSS images does not show an obvious disc morphology. However, follow-up observations confirmed that this is a real source, despite the difference in flux in each polarisations. Examination of deeper optical images revealed that the optical counterpart is more extended than is apparent in the SDSS images (this is discussed further in chapter 5).

Examination of the individual polarisations cannot be used blindly to verify a source, though it can aid in ruling out spurious sources. This may be one method to improve automatic extractors. Polyfind in particular produces

spurious detections that are single-channel spikes in the noise. The noise in the individual polarisations is not correlated (as may be seen in the above figures), and so the majority of such spikes are only visible in one polarisation. The most obvious approach is therefore to run Polyfind on each polarisation and cross-correlate the candidate lists produced.

Figure 4.6: A source that appears significantly different in each polarisation but was confirmed by follow-up observations.



(a) Polarisation A **(b)** Polarisation B

This approach is not without issues. The velocity reported by Polyfind of a real source may differ in each polarisation, since the noise is different. This is especially important for weak, wide sources, in which the peak of the distribution may lie in different horns in each polarisation. Some velocity tolerance must be set when cross-correlating the source lists, or real sources may be missed.

A catalogue was generated in this way for VC2, using Polyfind for source extraction and Topcat for cross-correlation. The velocity tolerance was adjusted by trial and error, each time checking against the catalogues that were produced from the previous methods described above. It was found that a tolerance of 100 km/s is sufficient for all known sources (detectable by this method) to be correctly matched. A tolerance of 150 km/s was used to err on the side of caution. The catalogue created contained 438 members, a factor 3.7 less than using Polyfind only on the averaged cube.

Tests found that many of the weaker sources were not detected in one or both polarisations with the standard 4σ criteria. This is unsurprising - since the normal data is an average of 2 polarisations, spectra in each polarisation typically have an rms a factor $\sqrt{2}$ higher than in the averaged cube. To counteract this, additional hanning smoothing of width 3 was applied to each polarisation. 37 known HI sources were detected, or a reliability of about 8% - comparable to Duchamp's 7% using only the averaged cube. This is also the same number of sources Polyfind detected in the averaged cube, so completeness has not been compromised.

This approach could also be applied to Duchamp, although this would require choosing separate flux thresholds for each polarisation. Determining this threshold is not straightforward, requiring a combination of adjusting the flux thresholds and employment of hanning smoothing before the majority of sources can be detected without destroying reliability. Such trial and error might be acceptable provided AGES data were all consistently of uniform quality. In practise, data may be reduced before all observations are complete. This means flux thresholds must be established for each polarisation in each cube - which relies on other methods of source extraction to detect the weaker sources.

A more sensible approach, perhaps, is to combine the advantages of Polyfind and Duchamp while reducing the disadvantages as much as possible. Polyfind operates on spectra, so that its estimate of the rms is at least local (though not correct with the standard pixel size). Duchamp's approach of searching for contiguous pixels (without any kind of template matching) appears to give much greater reliability, but tends to merge close sources, and its assumed uniform rms creates many problems. These issues suggest an extractor that operates on spectra and searches for sources consisting of contiguous pixels above some S/N threshold. Before exploring this possibility, it is worthwhile to examine the final method of source extraction.

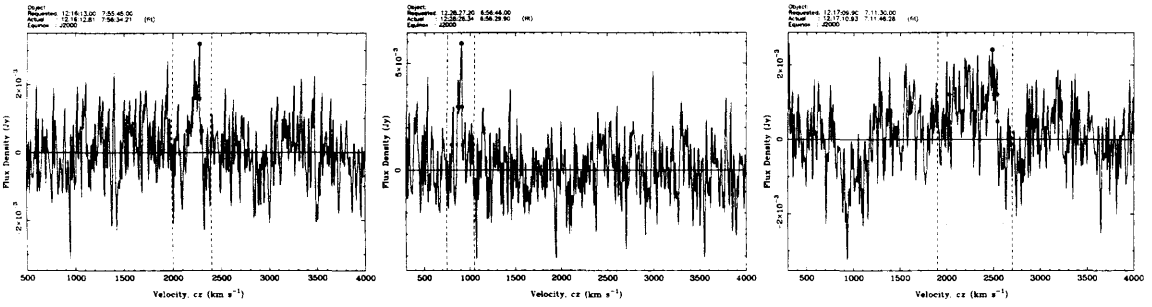
4.2.2 Optical selection

One advantage to studying the Virgo cluster is existence of the Virgo Cluster Catalogue of Binggeli, Sandage and Tammann 1985, an optically-selected catalogue complete to B_T = 18. This makes it possible to select galaxies from the optical catalogue and search for a corresponding detection in the HI data cube. Of course, this somewhat defeats the purpose of a blind HI survey, but can provide a useful third, independent method of source extraction.

The HI spectra were examined for each VCC object not detected in the blind search. Where an optical redshift was not available, the entirety of the spectra was searched - the VCC itself only classes objects as being cluster members or background objects, and is occasionally incorrect. For those objects with optical redshift measurements, stacking was also performed - this is discussed in more detail in chapter 7. For the 62 VCC objects not detected in HI by AGES in VC2, no new candidates were found. For VC1, 3 galaxies were detected from a list of 143. A small caveat is that 14 of the undetected galaxies lie close to other HI detections or are at the edges of the cube where the noise is stronger, where a detection would be less likely.

Spectra of the 3 objects are shown in figure 4.7. All are weak sources. Although VCC 180 has a spectrum of reasonable quality it is doubtful it would be regarded as a secure detection without an optical redshift measurement - it is certainly no better than the spurious source shown in figure 4.5. The narrow velocity width of VCC 1102 makes it difficult to distinguish from the noise despite the relatively high peak S/N. Conversely, the very low S/N of VCC 222 makes it nearly impossible to identify the source without an optical redshift, despite the exceptionally high velocity width (the second highest in the entire sample).

Figure 4.7: Sources detected by optical targeting in VC1



(a) Source 145, VCC 180. Peak S/N 4.7, W50 63 km/s

(b) Source 146, VCC 1102. Peak S/N 5.6, W50 49 km/s

(c) Source 147, VCC 222. Peak S/N 3.7, W50 513 km/s

While VCC 180 might conceivably be identified from its spectrum alone, that there is a sure optical counterpart makes the detection far more secure. VCC 222 would almost certainly never be considered a detection either from a visual inspection or an automatic extractor, and VCC 1102 would also be dubious. Fortunately, the detection rate by optical targeting of galaxies that would not (indeed, could not) be identified from a blind search is very low, no more than 2%. While this method is clearly useful where available (all three detections are of interesting objects, as discussed in chapter 5), its significance to blind searching appears limited.

4.2.3 A new automatic extractor

The above discussion indicates that there are three basic criteria a source must satisfy : a sufficient S/N, a sufficient velocity width, and satisfy both these conditions in both polarisations. Polyfind is able to more accurately measure the S/N of candidates since it operates on individual spectra. Duchamp gives greater reliability, since it requires more than 1 pixel for a detection. In this section both approaches are combined and applied to both polarisations.

A Fortran program was written that feeds parameters into the *miriad* task *mbspect*, the same task as used in analysis of the HI data. Thus, S/N measurements will be virtually identical to that measured when manually fitting a profile. This eliminates the problem with Polyfind measuring the S/N incorrectly. The cube is searched by extracting spectra along each spatial pixel over some user-specified velocity range (this avoids RFI). Spectra are extracted for the whole cube by iterating over the right ascension and declination axes.

The program proceeds by reading the output data from *mbspect* and checking the highest S/N in the spectrum. A 4σ S/N level is required for a detection - in the full $\sim$ 200 candidate list from the Virgo cubes, only 5 sources have a peak S/N below this. 1 of these was identified optically and 2 more are uncertain and require follow-up. A 4σ cut will therefore miss only 1% of the sources detected by other methods while preventing thousands of weak, spurious detections.

The baseline is corrected by a first-order fit, reducing the baseline ripple sometimes counted as a detection by Duchamp. Once a 4σ peak has been found, the velocity of the peak is extracted (*mbspect* outputs the velocity of the strongest peak in the spectrum). This is then used to establish a velocity profile over which *mbspect* measures the parameters of the source, and also masks this region from the calculation of the rms. Real sources, by their nature, tend to be much wider in velocity (so span a significant fraction of the baseline) than spurious detections, which are usually single-channel spikes. If a real source is not masked, the rms may therefore be incorrectly calculated and the S/N estimation significantly different. If, however, the masked profile only contains a narrow spurious source the rms calculation is not likely to be much affected.

The width of the profile may be specified by the user - 1,200 km/s is used here, approximately double the width of the widest source in the sample. This is wide enough to ensure the profile covers the entire source without knowing whether the velocity of the peak corresponds to the central velocity, or the velocity of one horn of the profile. It is narrow enough to reduce the amount of noise in the profile significantly (a factor of a few), which helps to give more accurate measurements of the source.

Once the profile to use has been established the second check on the detection is performed. *mbspect* is able to estimate the minimum velocity width of a peak. This is done at 50% of the peak flux and has sufficient accuracy as to indicate whether the source is resolved in velocity or not. The number of single-channel spikes in the noise that plague Polyfind can be greatly reduced by setting some minimum velocity width. A minimum 20 km/s velocity width is required for a detection, as this corresponds to 2 resolution elements with the standard hanning smoothing applied to AGES data.

The velocity width estimation is not without errors. The minimum width at 50% of the peak flux is used because it is usually a much better estimate than that at 20%. At 50%, the measured width will usually be above the noise and a genuine measurement on the source. At 20%, especially for weak sources, measurements can be confused by the noise and give an overestimate. Measurements at 50% can, however, also be inaccurate, being subject to noise confusion if sufficiently weak (again leading to an overestimate), but also underestimation is possible if there is a strong, narrow peak (such as a spike in the noise) overlaid on a wider, weaker source. In practise, while such errors prevent automatic *accurate* estimation of the velocity width, they only rarely confuse resolved from unresolved sources.

On finding a 4σ 20 km/s velocity width candidate, the program performs the same check in the individual polarisations. The same velocity profile as was established for the averaged cube is used in each - this is another reason for establishing a velocity profile. Were this not done, there is a risk that a stronger, polarised source in some other part of the spectrum might be identified as the source, even if the real source still satisfies the detection criteria. The profile does not eliminate this risk, but does alleviate it.

The increased rms in each polarisation creates a problem for weaker sources, if the same S/N criteria is required for a detection. A different S/N could be specified for each polarisation, or hanning smoothing applied to decrease the rms. Neither approach is ideal, as both the S/N level and/or the amount of smoothing applied must be adjusted so that weaker detections satisfy the criteria - it is difficult to predict analytically what values should be chosen to ensure a detection. Hanning smoothing has the extra complication of sometimes altering the reported velocity width. This is, however, often beneficial as it tends to widen and strengthen the S/N of real sources, while reducing the S/N of spurious single-channel spikes.

To ensure the extractor will recover as many sources as possible, the extractor is tested on weak detections that are definitely real, either with follow-up observations or possessing a sure optical counterpart[3]. In order to detect the weakest sources, it is necessary to use hanning smoothing of width 5 and a polarised S/N threshold of 3.5. In this way, for VC2 *every* confirmed source detected by the other methods is detected by this extractor (barring those that were in regions of RFI not searched by the extractor). Indeed, of the unsure sources, only 2 were missed, one of

[3]"Calibrating" the extractor in this way is certainly not ideal. However, unlike for Duchamp, which works in flux, this should not have to be recalibrated for other cubes with a different rms, since the extractor works in S/N.

which is below the 4σ threshold, and both since determined to be spurious.

The reliability of the extractor is also very high, provided some additional steps are taken. The raw list from the extractor contains 4,208 candidates, but the majority of these are multiple detections of the same objects. Topcat is used to match detections that are close or contiguous in space (within the beam width of 3.5') and velocity (using a tolerance of 150 km/s, as was found to be suitable for Polyfind). This reduces the candidate list to 268 objects (excluding those objects detected only in 1 pixel), a factor 6 better than Polyfind and nearly 2 times better than Duchamp, with no loss of completeness. In each group of pixels only that with the highest S/N is retained.

There is a small caveat to the matching procedure. Sources that are close in space and velocity may be merged together. The easiest approach is simply to retain the original detection list prior to merging and overlay a map of all the original candidates with the final, merged list. An example of this is shown in figure 4.8. Two bright galaxies are detected over pixel ranges that just overlap. It is obvious to a human eye that two distinct sources are present, but the merging process combines them (the second candidate detection after merging (in the east) is not at all associated with either of the two objects, it is actually a spurious source).

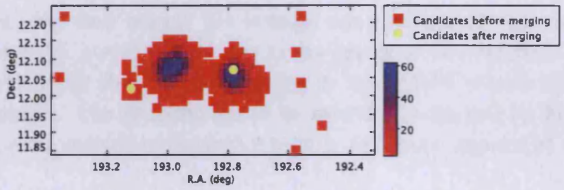
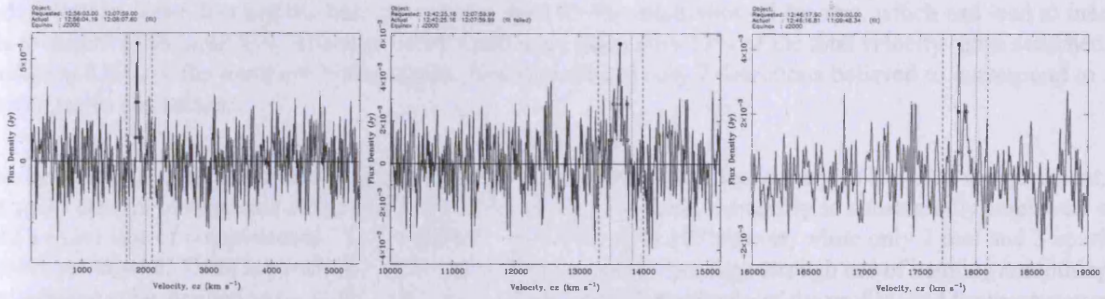


Figure 4.8: Spatial map of the output of the automatic extractor, showing two separate sources. Red-blue squares indicate the raw output, with colour indicating S/N. Yellow circles show the peak S/N pixels retained after merging.

As well as detecting virtually all good candidates from other methods (visual, Polyfind and Duchamp), the extractor also finds other sources that all three previous methods failed to detect. Two sources are found which are definitely real (as they have been detected by ALFALFA) and one more candidate source found that requires follow-up observations. Spectra of the 3 are shown in figure 4.9.

Figure 4.9: Sources detected by the new extractor missed by other methods



(a) Also detected by ALFALFA. Peak S/N 5.33, W50 36 km/s

(b) Also detected by ALFALFA. Peak S/N 5.19, W50 216 km/s

(c) Also detected by Polyfind. Peak S/N 6.00, W50 578 km/s

The first source, detected by ALFALFA, is of low S/N and narrow, and therefore it is not surprising it was missed in a visual extraction. With a peak S/N of 5.53 and a W50 of just 36 km/s, this is probably one of the hardest sources to detect by any method. It is too far north to be detectable by Duchamp. This also explains why it was detected by ALFALFA despite being so weak - it is near the edge of the cube, where the noise for AGES is abnormally high. Problems in Polyfind's method of calculating the rms explain why that extractor also failed to detect it. Admittedly, however, which such a narrow profile this source might easily be dismissed as a spurious spike were it not for the independent observations.

The second, also detected by ALFALFA, is also very near the northern edge of the cube. This probably explains

its absence from a visual extraction, since its S/N and velocity width are comparable to other galaxies detected visually. Its northern position certainly makes it impossible to detect with Duchamp. Its S/N of 5.2 is low enough that problems with Polyfind's estimation of the rms likely explain why it was not detected by that extractor.

The third source was actually previously detected by Polyfind, but rejected. This illustrates the subjective nature of selecting candidates from an automatic extractor - which is especially difficult given an initial list of 1600 potential sources. With less than 300 objects, the user can spend considerably longer viewing each detection. Nonetheless this source still requires follow-up. Two potential optical counterparts are visible in the SDSS, but neither have optical redshift measurements.

While the extractor is able to detect close to 100% of the known recoverable sources (outside of those areas contaminated by RFI), much scope for improvement remains. The initial step is to search for the strongest peak in a spectrum. Should this be a strong, polarised source the extractor will reject it, along with any real weaker sources that may be present in the same spectrum. Similarly, if two real sources are present in the same spectrum, the extractor will only detect one source (this is less problematic as a human examining the spectrum would see the second object).

In reality these difficulties are only very minor. No sources were missed in this way in VC2, and only 1 of the approximately 150 candidates in VC1 was missed due to the presence of a spurious source. This is partly due to the necessity to search over relatively short velocity ranges to avoid RFI, which reduces the chance of detecting two sources in the same spectrum. The problem could be entirely overcome by having the extractor search the velocity ranges within each spectra outside of the initial profile, i.e. those regions of the spectra that do not contain the strongest source.

A visual search is still vital, if only to search areas where RFI is present. The extractor detects 42 of the 51 sources in VC2 known from other methods. Of those missed 2 are spurious while 7 are present in a region contaminated by RFI. Moreover, those regions must first be located by a visual search, as some sources of RFI (such as the L3 GPS satellite) are not always present. A visual check is also essential in case some particular problem exists that might cause problems for an automatic extractor, such as the "wandering birdie" in the NGC 628 data cube (Auld et al. 2006, see also http://www.naic.edu/~ages/ages_results.html).

The presence of RFI can cause problems indirectly for those regions where it is not present. For example, the velocity range 6,000-8,000 km/s in VC1 is sandwiched between two RFI regions. Searching a relatively narrow velocity range (2,000 km/s) such as this can be a problem, since it is necessary to mask a 1,200 km/s range around each detection. Less than half the baseline is being used for the calculation of the rms, which can lead to inaccurate estimates of the peak S/N. Although 6,000-8,000 km/s spans only 13% of the total velocity range searched, 74 candidates (28% of the total) are in this region, though there are only 7 detections believed to correspond to real HI sources in this volume.

One approach is simply to avoid masking altogether and accept some inaccuracy in the S/N measurement. If the same criteria of requiring a detection in 3 or more pixels is used, reliability is substantially improved with only a small loss of completeness. The candidate list is reduced to 107 objects, while only 2 real and 3 spurious sources are missed. There is clearly scope for further improvement - perhaps through use of hanning smoothing on the averaged cube, or altering the width of the masking applied independently of the profile used for measurements.

One of the real sources missed could be recovered if a velocity width criteria is used - accepting, for example, any source with a measured width greater than 50 km/s regardless of the number of pixels it was detected in - without significantly increasing the number of candidates. This would then essentially mimic Duchamp, by searching for detections present over a number of pixels contiguous in space or velocity.

The wide velocity profile used is not ideal. Only 5 sources have velocity widths greater than 500 km/s, whereas about 80% of the sample have velocity widths less than 250 km/s. A better way to estimate the profile needed should considerably improve the accuracy of the estimated S/N and velocity width of the source. This could perhaps be done by comparing the minimised and maximised velocity widths, which for strong sources are equal and a good estimate of the true width of the source.

4.3 Summary and Discussion

All methods of HI source extraction available suffer from a fundamental problem - there exist spurious detections indistinguishable from real sources, as shown in figure 4.5. This can only be overcome with follow-up observations on any source whose validity is in doubt, even if that doubt is due (as in the example of figure 4.5) to optical data rather than the quality of the HI detection. A perfect extractor with 100% completeness and reliability cannot be achieved except at very high S/N levels, and even then RFI contamination must be overcome. A more practical goal is simply to increase completeness and reliability as much as possible.

All methods of source extraction are also subjective to some degree. Candidate lists generated automatically inevitably include far too many spurious sources for follow-up to be obtained for each detection. Combining all the extraction techniques described above, in VC2 alone about 2,300 follow-up observations would be required, equivalent to about 390 hours of telescope time. Inevitably, culling the candidate lists by any process will miss some real objects, since in some cases it is simply impossible to tell which are real and which are spurious.

However, the detection rate from those objects for which follow-up was obtained indicate that this process is fairly successful. Of the 30 Virgo objects with follow-up observations, 16 were detected. This does not imply by any means that of the 2,300 other candidates already rejected, half may be real. The vast majority of those candidates are of much worse quality than those selected for follow-up. Most of them can be easily dismissed as they appear only in one polarisation. A 50% detection rate for dubious sources indicates that the human eye is generally good at discriminating between real and spurious sources.

Ideally, an extraction technique should be complete to some mass limit at some distance. This is not possible for any technique investigated here - it is fundamentally difficult to achieve this in a 3-dimensional survey. All automatic techniques require (essentially) a S/N threshold, but this places no restriction on the total flux of sources that may be present. Sources below the S/N limit may (in principle at least) be of any velocity width, and hence of any total flux and mass. This is possible even for the new extractor, since the velocity width estimation is far from perfect. Sources narrower than the velocity width limit may be of any S/N, so these too may be of any mass. Additionally, rms varies through the cube, so a S/N based extractor will not be complete to the same limit in different areas. Conversely a total flux based extractor will (as Duchamp does) suffer hopelessly poor reliability in regions of higher rms.

The completeness and reliability of each source extraction method explored here is estimated in table 4.1. Completeness is estimated by assuming that the final source list generated by a combination of all methods contains all of the potentially recoverable sources in the cube. This excludes sources rejected due to polarisation (as they can be ruled out immediately), but includes those spurious sources that required follow-up (as these are nonetheless detectable sources). Completeness is then estimated by considering the percentage of this final list detected by each particular method. For reliability, it is assumed that a source is real if it possesses a sure optical counterpart, is confirmed by follow-up observations, or possesses an obvious possible counterpart and is clearly visible in both polarisations.

The first column indicates the method used. "Polyfind cross-pol" refers to running Polyfind on both polarisations and cross-correlating the candidates lists. "Duchamp 4σ" refers to using Duchamp with a 4σ criteria for detection, as estimated by Duchamp itself. "Duchamp threshold" refers to manually setting a flux threshold. "New" refers to the new extractor, "New no masking" refers to that extractor without masking a velocity range around each source. The second column gives the number of candidates each method produces - for a visual inspection this is the total number of independent candidates found after searching the 3 projections. The third column indicates completeness as described above. The bracketed number estimates the completeness by only considering sources in regions searched by the extractor (i.e. sources in regions of RFI are not included). Finally, the fourth column estimates reliability as described above.

Reliability of the source extractors varies considerably, with a visual inspection being far higher than automatic methods. Despite the inherent subjectivity, the human eye appears much better at discriminating real from spurious sources than automatic techniques. Polyfind's very poor reliability is likely due to its simple source selection criteria (peak S/N only). Duchamp and the new extractor perform rather better, because they require a higher total flux (and in multiple pixels and channels) for a detection than Polyfind does.

The new extractor gains significantly in reliability if no masking is applied to the velocity profile when estimating

Table 4.1: Reliability and completeness estimates for various source extraction methods

Method	Initial list	% Completeness (in searched volume)	% Reliability
Visual	75	78 (77)	51
Polyfind	1619	70 (73)	2
Polyfind cross-pol.	438	70 (73)	8
Duchamp 4σ	119	42 (54)	18
Duchamp threshold	492	72 (90)	7
New	268	83 (95)	16
New no masking	107	77 (89)	38

the rms. This may be due to the majority of spurious detections being narrow spikes. In this case, masking 1,200 km/s of velocity would significantly reduce the number of channels used when estimating the rms, lowering the accuracy. Without masking, only 1 channel will be significantly greater in flux than the rest, so including this will not substantially alter the calculated rms. For the majority of cases then, it appears that not masking the velocity profile gives more accurate results[4].

Completeness of the extractors is more consistent, with the exception of allowing Duchamp to estimate a 4σ threshold, which is far worse. This is unsurprising as Duchamp's estimate of 4σ is not local to each spectra and thus always inaccurate. Interestingly, the remaining methods all achieve comparable completeness despite the different techniques used and areas masked. However, Duchamp and the new extractor perform significantly better when considering only the volumes they are able to search. The new extractor even begins to approach 100% completeness, considering that the only 2 sources missed were both discovered to be spurious.

Of the automatic extractors the new method is clearly superior, with both completeness and reliability always greater than for the other automatic techniques, with completeness overall at least as high as for a visual extraction. It does not suffer the need to mask the spatial edges of the cube as Duchamp does, and even within the volumes each extractor is able to search, its completeness is still higher than that of Duchamp. Although Duchamp with a 4σ cut is able to achieve superior reliability, this is at the expense of completeness. The new technique, overall, gives a higher completeness while maintaining a high level of reliability.

The principle advantage of a visual extraction is its ability to discriminate between RFI and genuine sources. Within the volumes they are able to search, automated methods perform better, being faster and more complete, though less reliable. The new extractor in particular detected new candidate sources missed by a visual extraction, whilst only missing 2 spurious sources detected visually. All automated extractors are considerably faster - a visual inspection taking approximately 1 week, Polyfind about 1 hour, Duchamp a few minutes and the new method a few hours. A visual inspection may perhaps be dispensed with except for areas contaminated by RFI. As S08 found, the overall best method of source extraction appears to be a combination of a single human and an automatic extractor.

Visual inspection is unavoidable with regards to the output of automatic extractors, given their low reliability. The 51% reliability quoted for visual inspection in the table is perhaps an underestimate, as this is derived from the initial candidate list of 75 objects. The second stage of visual inspection, in which spectra of each candidates are examined, reduces the candidate list to 41 objects, of which 95% appear reliable. Automatic extractors cannot approach this except at high S/N levels in regions devoid of RFI. All source extraction techniques must, therefore, remain subjective to some degree.

[4]This emphasises that the rms of the data can only be estimated, so even source extraction based on S/N is not without issues.

Chapter 5

Results 1 : The VC1 Area

5.1 Introduction

Two areas of the Virgo Cluster have been selected for study with AGES, VC1 and VC2, shown in figure 5.1. This chapter examines the VC1 area while the next considers VC2 and compares the results of the two areas. The areas were selected for study as they span very different regions of the cluster. VC2 spans 5 square degrees, 5 degrees of right ascension by 1 degree of declination, and observations are complete for that region. The VC1 regions covers 20 square degrees, 10 degrees of R.A. across the cluster and 2 degrees of declination. Currently only the southern half (10 degrees of R.A. by 1 degree of declination) has been observed, and throughout this chapter only observations in that area are considered. Observations for the northern half of the cube have begun but at this stage are very incomplete. In this study the focus is on cluster members, though background objects are used where appropriate as a comparison sample.

It is evident from figure 5.1 that VC1 includes a much denser region of the cluster. Per square degree, there are are 17.4 VCC galaxies (classified as cluster members) in VC1 (in the part that has been observed thus far) and 15.0 in VC2. Although not much different in overall galaxy number density, the regions are quite different in terms of HI. Extensive pointed HI observations of VCC galaxies have been performed, and from these, there are 4.2 detections per square degree in VC1, but only 2.2 in VC2 (this data is from the GOLDMine database). VC1 is thus interesting as a richer area of the cluster to study. As evident in figure 5.1, it spans both the low density cluster outskirts and its higher density interior.

Galaxies within the VC1 region are not distributed uniformly but in 3 separate groups or clouds - the A and B subclusters at 17 and 23 Mpc distance respectively, and the infalling W cloud at 32 Mpc. Distances measurements are as described in chapter 1, but it is worth re-iterating that there the boundaries between each cloud cannot be precisely defined, nor can the depth of each cloud. There is thus likely a significant, but difficult to quantify, error in the distance assigned, but as described in section 5.3.2 it should not alter the HI mass by much more than a factor 2. In this study the distance assigned in the GOLDMine database is used, except where otherwise stated. Galaxies detected that are not listed in the VCC are assigned membership to a cloud (and hence distance) based on their spatial location.

5.2 Data analysis overview

The observations, data reduction and analysis procedures have been fully described in chapter 2. Observations were taken in January-June 2008 and February-June 2009, using the ALFA instrument on the Arecibo telescope in spectral line mode. The field, when fully observed, will span 10 degrees of R.A. by 2 degrees of declination centered on M49. Currently, observations are complete for one quarter of this area, a 5×1 degree field in the south-west of the cube, with observations to 2 out of the 3 required passes complete in the adjoining eastern quadrant of the cube. Due to ALFA's hexagonal beam, a small area outside these is also included. The full spatial range of the data considered here is from 12:08:36 to 12:49:36 in R.A, and from +06:52:55 to +08:06:55 in declination.

As is standard for AGES data, hanning smoothing is applied to the data to give a velocity resolution of 10 km/s. The minimum noise level - rms - reached is approximately 0.5 mJy/beam from measurements of individual sources

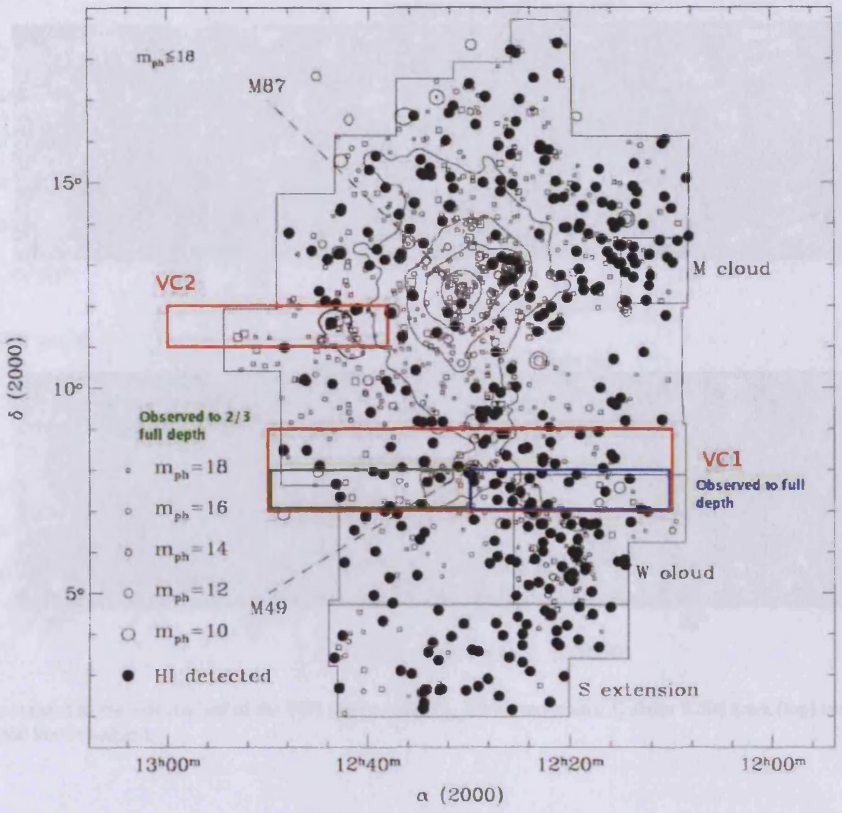


Figure 5.1: The Virgo Cluster region, highlighting the AGES and VCC areas and detections. Squares are early-type while circles are late-type galaxies. Filled circles are those with known HI detections from previous pointed observations. The contours show X-ray detection from the ROSAT satellite. The current status of the observations is highlighted; observations of the northern part of VC1 are not used in this thesis at all. Physically, the area described in this chapter spans 2.97×0.30 Mpc at 17 Mpc distance.

using *mbspect*, with a median of 0.6 mJy/beam. As a sensitivity limit, for an rms of 0.6 mJy a 4σ detection with a tophat profile, 50 km/s velocity width, would have an HI mass at 17 Mpc distance of 8.2×10^6 $M_\odot$. As detailed in chapter 4 there are various reasons why this is only an estimate of the sensitivity limit, and should not be taken too literally (largely because of the shape of the HI profile). This would correspond to a total flux of 0.1 Jy, but there are actually 5 confirmed detections below this limit.

Since observations are not yet complete sensitivity is unfortunately not uniform for the entire area (though the field is fully sampled). The western half (west of about 12:30:00.00 R.A.) has been scanned to full depth, with the sensitivity estimated above. For the eastern half, measurements of detections gives a median rms of about 0.8 mJy/beam, which is equivalent to a mass sensitivity of approximately 1.1×10^7 $M_\odot$ for the same parameters as above. The areas of different sensitivity are indicated in figure 5.1.

As with other AGES data cubes there are several regions contaminated by RFI where sensitivity is significantly affected. LIVEDATA cannot correctly subtract the strong extended signal from the Milky Way, which effectively blinds the survey over the velocity range -50 to +50 km/s. The other two dominant sources of RFI, as typical in AGES cubes, are the San Juan F.A.A. radar at 1,350 MHz (about 15,600 km/s) and the L3 GPS satellite signal at 1,381 MHz (about 8,500 km/s). The survey is not blind within these regions, but sensitivity is greatly reduced. The GPS satellite unfortunately manifests itself in the data as false positive signals, as shown in figure 5.2.

Source extraction procedures for AGES have been described in detail in chapter 4. In this case, two key methods of extraction were used : a visual inspection of the data and the new automatic method described in chapter 4. Polyfind was not used, owing to its poor reliability - 1,600 detections for the VC2 area, half the size of the VC1 area observed (see figure 5.1). Examination of an expected 3,200 detections for this area becomes impractical.

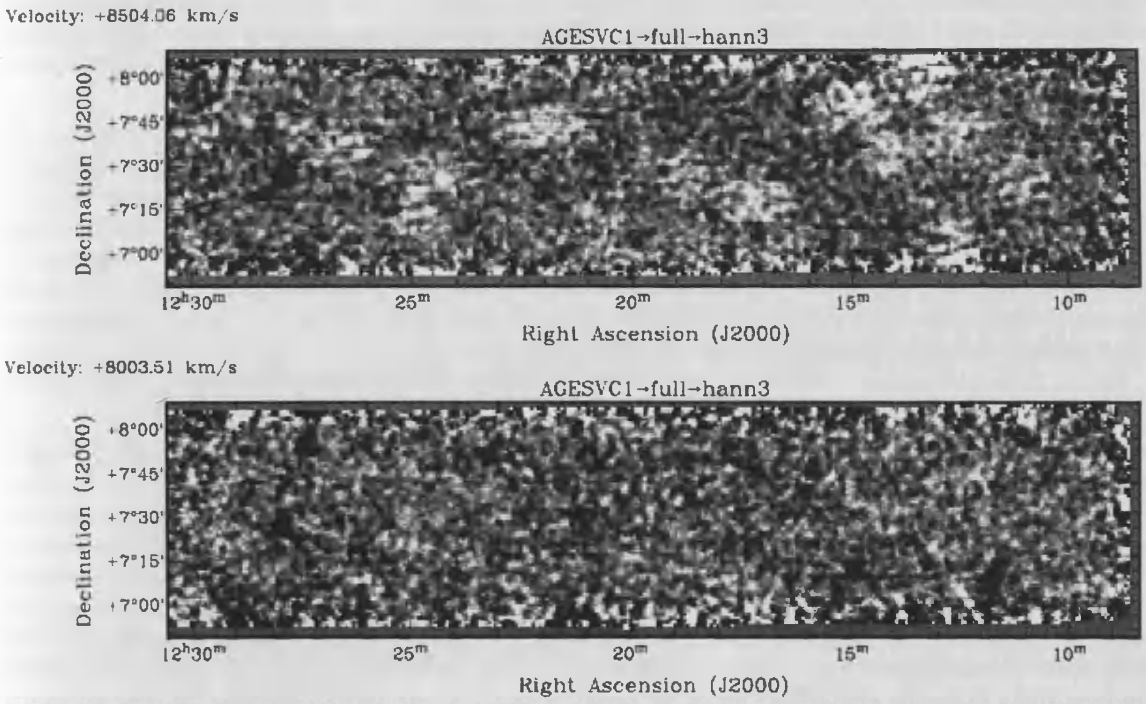


Figure 5.2: Channel maps of the western half of the VC1 region, showing GPS interference at about 8,500 km/s (top) and a more typical blank region at about 8,000 km/s (bottom).

A third method of extraction was also used as described in chapter 4 - inspection of the spectra of optically selected (i.e. from the VCC) galaxies that the other two methods failed to detected. The HI spectra were visually examined along the entire length of the cube where optical redshift measurements were not available. This was done for both early and late-type objects, 128 in total (most of which are early types so it is not surprising that they are not detected). Stacking was performed on the undetected early-types where optical redshifts were available, this is described in detail in chapter 7.

The visual and automatic extractor were employed as described in chapter 4. The only significant difference with regards to the visual extraction was the use of the *miriad* task *immask*. This masks selected volumes of the cube, in this case pixels identified as HI detections. With the high density of sources (a visual inspection produced 68 candidates in the Virgo Cluster alone) in this region, this is necessary to avoid confusing previously identified candidate sources.

The combination of source extraction methods produced a combined candidate list of 151 sources for the entire volume of the cube. 7 of these were rejected on inspection of the individual polarisations (none had any likely optical counterparts). 22 of the remaining 144 were selected for follow-up observations owing to their low signal to noise ratio, detection by only 1 method, or lack of a clear optical counterpart. 11 of those re-observed were found to be spurious. Of the remaining 133 there remain 3 candidates for which HI status is in doubt. One of these requires follow-up observations, the others have been re-observed but in one case the spectrum is dominated by a 1 MHz ripple, and the other may be confused by continuum emission. These objects are not considered further unless otherwise stated, giving a final sample size for this study of 130 objects.

5.2.1 Other data

HI

Several other sources of data have been used in the analysis, see chapter 3 for more details. These can be used as an accuracy check on the data, to ensure that basic parameters such as sky position, mass and velocity width have been computed correctly. For these comparisons, background objects are also included. The most useful source of independent HI data is the online GOLDMine database, which lists information at various wavelengths

(including 21cm and optical) for all VCC galaxies. 62 VCC objects were detected by AGES (including two galaxy pairs where the precise galaxy containing the detected HI cannot be identified), of which 47 have HI measurements listed in GOLDMine.

Most of the remaining VCC objects detected in AGES do not have previous HI observations, though 2 are listed in GOLDMine as observed but not detected. These observations were probably less sensitive than AGES. Searching GOLDMine within the maximum spatial extent of the cube, 10 galaxies are listed with previous HI detections that AGES did not detect. 6 of these are actually just outside the cube, 1 is at too high a redshift to be detectable, and the HI spectra of 2 are described as quality 5, meaning unpublished and not necessarily reliable. The single remaining object is VCC628. Given the reported HI mass and velocity width, and the rms for this spectra of 0.5 mJy, AGES should be able to detect this source at a 6σ level. The peak S/N in a 100 km/s velocity profile centred on the optical redshift of VCC628 is only 1.01. Since the quality of the detection in GOLDMine is listed as 4 (meaning poor), it is likely that this previously claimed detection is spurious.

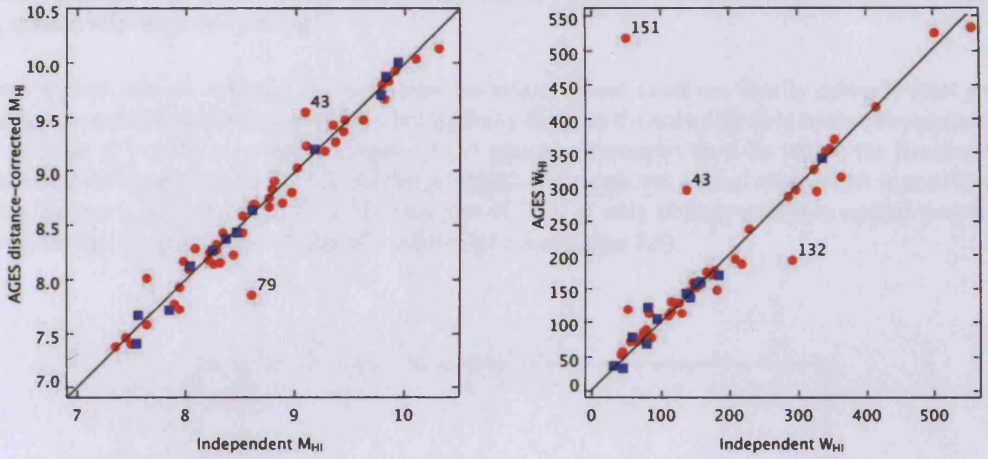
A second useful, but currently smaller, independent data set is the ALFALFA survey (see chapter 3). At present this has only reached south as far as approximately declination +07:50:00. Though a substantially smaller sample than the comparison with objects listed in the GOLDMine database, it has the advantage of comparing two blind surveys carried out on the same telescope with the same instrument (albeit with different data reduction techniques and observation strategies). 14 sources have been detected by ALFALFA in the narrow overlap region between the two surveys in this area. AGES detects 13 of these, the one non-detection is at the extreme northern limit of the cube. Moreover, AGES detects 24 sources in the overlap region that ALFALFA does not. A caveat is that this is the southern limit of ALFALFA just as it is the northern limit of AGES. These non-detections are, however, broadly consistent with the expectation based on ALFALFA's sensitivity, given ALFALFA's criteria of a S/N level of 6.5 (see equation 3.1) for a detection.

The rms level for ALFALFA is assumed to be a constant 2.34 mJy - the mean noise level of the 14 sources detected in the overlap region. The total flux and velocity widths are those measured by AGES. The ALFALFA criteria for a detection is that the S/N measured via equation 3.1 must exceed the (approximate) limit of 6.5. The predicted S/N for the non-detections is below this in all but 2 cases, which have predicted S/N values of 9.4 and 10.9. For their S/N to be below 6.5 the strongest source would need to be in a region of rms 4.0 mJy, which can occur - some ALFALFA detections of stronger sources have been made in regions approaching 5 mJy rms. The AGES detections missed by ALFALFA are thus due to sensitivity effects.

The total M_{HI} as computed by AGES is compared with that from ALFALFA and GOLDMine in the left panel of figure 5.3. The distance used for calculating the MHI from AGES measured flux is that used in the other data sets for consistency. There is generally very good agreement with previous observations, but there are 2 clear outliers. Source 43 is identified with VCC 265, but its HI profile overlaps that of source 42 so inaccuracies are not surprising. The other outlier is source 79, VCC 758, which is discussed in section 5.3.3. The AGES rms for the spectrum is 0.6 mJy. The velocity width (see below) measurements agree very well : 325 km/s for AGES, 322 km/s from GOLDMine. Yet if the HI mass of 8.6 is correct, AGES should detect the source with a peak S/N of 16.0, whereas it is actually closer to 6.0. There is no obvious problem with the AGES data cube that could explain this (e.g. RFI or contamination by a nearby bright source), indeed, sources in close proximity agree well with GOLDMine measurements. The AGES data is assumed to be correct throughout the rest of this study.

A comparison of measured velocity widths is shown in the right panel of figure 5.3. For ALFALFA, the velocity width used for comparison is simply the width at 50% of the peak flux (the W50), as that is the width reported in the online catalogue. For GOLDMine the situation is slightly more complex. GOLDMine gives the "width of the HI line (in km/sec) obtained by averaging the value at 20% of the peak flux with the one at 50% of the mean flux." This is approximated in the figure by averaging the W50 and W20 as measured by AGES.

Again there is generally very good agreement with the independent measurements but with 3 clear outliers. Source 43, as mentioned above, overlaps the profile of another source. Source 132 has an averaged velocity width of 291 km/s in GOLDMine, but only 191 km/s in AGES. This is a weak detection in AGES (5.11 peak S/N) so it is possible some flux has been missed that would extend the profile. The last outlier is VCC 52, source 151. The measured width by AGES (496 km/s) is nearly a factor 10 greater than that listed in GOLDMine (48 km/s) - this object is discussed further in section 5.3.7.



(a) Logarithmic HI mass, with the distance used being that from the other surveys for consistency

(b) Velocity width, using W50 for ALFALFA data and an average for W20 and W50 for GOLDMine

Figure 5.3: Total HI mass and velocity widths from AGES compared with data from ALFALFA (blue squares) and GOLDMine (red circles). Lines show the expected one-to-one correlation. Outliers are labelled with the source number in this catalogue, see tables.

Optical

Optical data available for this region includes the SDSS and the INT WFS. As described in chapter 3 the INT data is too variable in quality to be of much use in this region. It is only used for imaging purposes in a few cases (e.g. searching for faint optical counterparts), and never for photometry. Instead, optical photometry is performed using the SDSS data which is uniformly available in this region, by the aperture photometry procedure described in chapter 3.

Five objects are not considered in any analysis of the photometry unless otherwise stated. Objects 67, 128 and 129 have very uncertain optical counterparts and are possible dark galaxies (see section 5.3.7), and object 121 is severely contaminated by a bright foreground star. Additionally, object 149 has multiple possible counterparts, so it difficult to identify any single source as being associated with the HI (see chapters 6 and 8).

The search for optical candidates is described in chapter 3, where it was stated that the difference in HI/optical redshifts is usually much less than the required 200 km/s for a detection. 81% agree to within 30 km/s, and 93% agree to within 50 km/s. There is however a slight systematic bias - the measured HI redshifts have a mean of 19 km/s greater than the optical measurements. This is shown in figure 5.4.

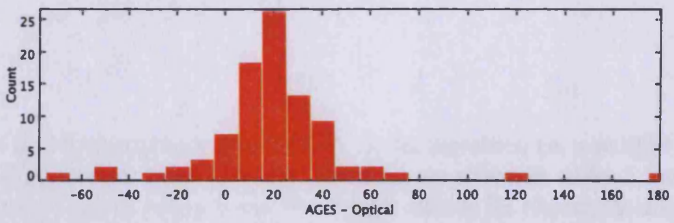


Figure 5.4: Distribution of the difference between measured HI and optical recessional velocity with a bin size of 10 km/s.

Two clear outliers are seen, the most extreme with a difference in HI and optical redshift measurements of 175 km/s. This is VCC 52, source 151 in this catalogue, also an outlier in terms of HI velocity width (see above). The object is discussed more fully in section 5.3.7. The second, with a difference of 117 km/s, is the giant spiral VCC

1555. This is an exceptionally large object with an optical radius of 4.17'. The center of the object actually lies further north than the edge of the cube; AGES is only able to detect part of the HI disc. Since part of the profile is missing, inaccuracies are unsurprising.

Candidates without optical redshifts are considered uncertain. These cases are usually optically faint galaxies for which an optical redshift is difficult to obtain, but typically they are the only plausible optical counterpart and very near the position of the HI. As stated in chapter 3 the optical counterpart must lie within the Arecibo beam, but the positional difference between the HI and the position of the apparent optical counterpart is usually much less than this. The mean difference is 37", with a median of 23". If only objects with sure optical counterparts are considered, the figures reduce to 34" and 18" respectively (see figure 5.5).

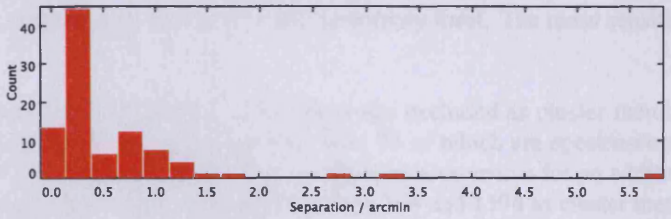


Figure 5.5: Distribution of the difference between measured HI and optical spatial coordinates with a bin size of 0.1'.

As with previous comparisons with independent data, the agreement is generally very good except in a few cases. The most extreme outlier is VCC 1555 with a positional difference of 5.7', for reasons explained above. VCC 1249 is also an outlier with a difference in position of 3.2', again, see section 5.3.7. VCC 1182 (source 27) has a deviation of 2.7', and this is perhaps caused by a combination of low S/N (5.65 peak) and being on the northern edge of the data cube.

UV

Far UV and Near UV data is obtained from the GALEX satellite data archive as described in chapter 3. A total of 172 VCC cluster member galaxies are present in this region, though only 69 of which have been clearly detected in both UV bands (the majority of VCC members here are early-type galaxies). For these objects UV photometry has been provided by Hughes and Cortese (private communication but see Hughes & Cortese 2009 for more details). Photometry was performed on an additional 12 objects detected in HI present within the Virgo Cluster but not listed in the VCC. UV photometry was performed only for cluster members (velocity less than 3,000 km/s), for a total of 57 objects. Tables of the UV properties of HI detections and non-detections are shown in tables 5.5 and 5.6 respectively.

5.3 Results

This study is focused on the HI detections that are clearly cluster members, i.e. a redshift of less than 3,000 km/s. For completeness, the HI parameters of all 130 sure detections are shown in tables 5.1 and 5.2, which detail the Virgo Cluster and background objects respectively. For objects outside the cluster, the distance used is determined simply from the Hubble velocity, assuming a Hubble constant of 71 km/s/Mpc. For objects within Virgo, the distance used is largely that given in the GOLDMine database, with 11 exceptions - from position-velocity space these objects seem to belong to a different cloud than that assigned in GOLDMine. This is explained further below. The optical properties of all detections are given in tables 5.3 and 5.4. UV measurements were only performed on definite cluster members, these are shown in tables 5.5 and 5.6, which are for HI detections and non-detections respectively.

5.3.1 Overall statistics

A total of 60 HI detections were made within the Virgo Cluster. Of these, 15 are not members of the VCC, though one of which (object 128) remains an uncertain detection owing to the presence of continuum emission. The remainder are all secure detections. The VCC is supposedly complete to a photographic magnitude of 18. Taking this to approximately equal the g band magnitude, 8 of these detections are actually brighter than the completeness limit (4 of these are marginal, with $g > 17.5$). This is less than 5% of the total of 172 VCC objects thought to be cluster members (see below) in this region, so the completeness magnitude of the VCC seems reasonable.

Though AGES does detect a population of objects missed by the VCC, this population is small in comparison with the VCC cluster members. Some basic assumptions show that this is at least partly due to the sensitivity effects of each survey. As described earlier, the sensitivity limit of AGES is approximately 8.2×10^6 $M_\odot$ at 17 Mpc distance. The completeness limit of the VCC is a photographic magnitude of 18.0, but it does contain some objects (in the VC1 region) as faint as 20.0, or 19.6 in the g band. Assuming all the stars to be of 1 solar mass and luminosity, this equates to 5.7×10^6 $M_\odot$, within a factor of 2 of AGES sensitivity limit. The mass sensitivities of the two surveys are thus comparable.

The VCC lists 203 objects within this region. 29 of these are excluded as cluster members by previous redshift measurements. The remaining 174 are classed as members, 95 of which are spectroscopically confirmed. AGES finds new redshifts for 5 VCC cluster member objects, while measurements for an additional 2 are found in NED that are not given in GOLDMine. AGES confirms VCC 190, 368 and 1394 as cluster members, while NEDS confirms VCC 1866. Likewise AGES finds that VCC 724 and 694 are background objects, with NED demonstrating that VCC 2068 is behind the cluster. In summary, there are 99 spectroscopically confirmed VCC cluster members in this region. 72 remain without any redshift data and are assumed to be cluster members throughout, while 30 VCC galaxies have been confirmed as background objects and are disregarded unless otherwise stated.

AGES detects 130 sure sources, 59 within the cluster and 71 behind it. The majority of the detections behind the cluster do not correspond with objects listed in any major catalogue. From a NED search, 54 have no previous HI observations, with AGES providing 25 new redshifts. Within the cluster, searching GOLDMine and NED, 20 are new HI detections, with 10 new redshifts being provided by AGES.

5.3.2 Distribution

The spatial distribution of the AGES detections and non-detected VCC galaxies is shown in figure 5.6. The distributions are broadly similar, with the western region being richer in both HI-detected and non-detected galaxies. The western half contains approximately 22 non-detected and 9 detected objects per square degree, the eastern half just 10 non-detected and 3 detected objects per square degree. Qualitatively, the structures defined by the objects are very similar. Gas-rich galaxies appear to trace roughly the same population of objects as the non-detected galaxies.

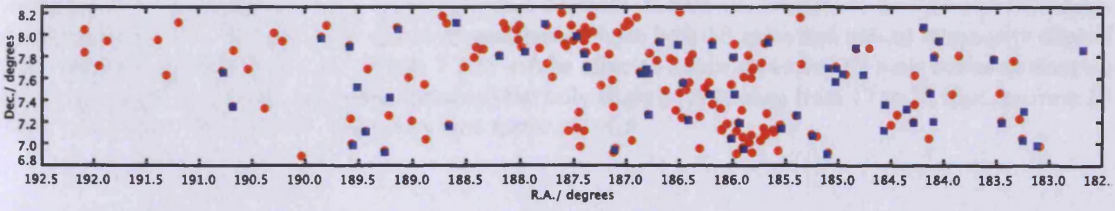
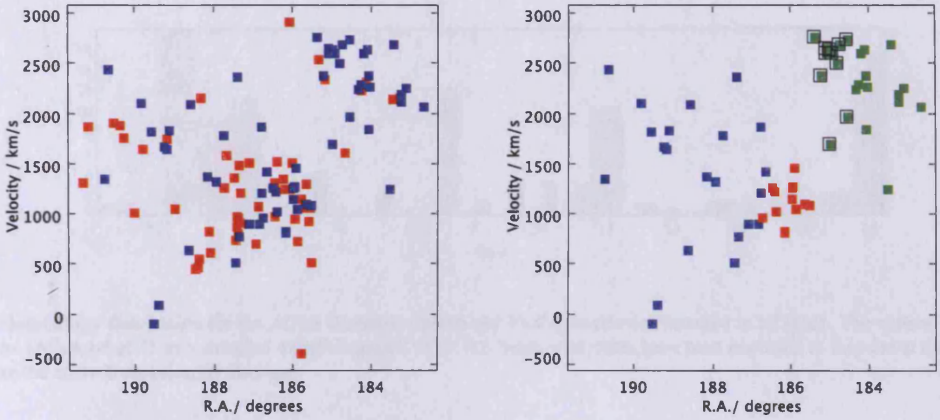


Figure 5.6: Spatial distribution of the 59 AGES detections (blue squares) with $cz < 3{,}000$ km/s and the 135 VCC non-detected galaxies (red circles) identified as cluster members.

This is more clearly evident in the left panel of figure 5.7, which shows the distribution in position-velocity space. Three distinct clouds are evident, one at a redshift of ~1,000 km/s, another at ~1,700 km/s and a third at ~2,200 km/s. The populations of HI detected and non-detected galaxies reside in the same areas of position-velocity space.

HI rich objects belong to the same population as non-detected galaxies, i.e., infalling clouds.



(a) 59 AGES detections (blue squares) together with 55 VCC non-detections (red circles) with available redshift measurements

(b) HI detections only, colour indicating distance. Blue for 17 Mpc, red for 23 Mpc, green for 32 Mpc. Black squares highlight objects which in GOLDMine have been assigned a distance of 23 Mpc.

Figure 5.7: Position-velocity maps for the VC1 region

There is considerable uncertainty as to the exact boundaries of each cloud. The right panel shows the same plot but with the colour of each point indicating its adopted distance. As discussed in section chapter 3, this distance is based on spatial position, i.e. membership of a particular cloud. Galaxies in each cloud are assumed to be at the same distance, which is as determined in Gavazzi et al. 1999, from a combination of Tully-Fisher and fundamental plane measurements.

While the cloud at $\sim$ 2,200 km/s appears quite distinct from the others in position-velocity space, two different distances have been assigned to galaxies within it in the GOLDMine database. This is because only spatial position was used to assign cloud membership, based on proximity to galaxies with individual distance determinations. In position-velocity space the difference between the 23 and 32 Mpc clouds is obvious, and based on this the distances of 11 of the galaxies supposedly in the 23 Mpc cloud (highlighted in the figure with black squares) have been re-assigned to 32 Mpc (the same correction has been made to non-detected objects in this region).

The difference between the 17 and 23 Mpc clouds is somewhat more subjective and so their original distance assignments are retained. The impact of the distance chosen is in any case relatively minor. HI mass scales as distance squared, so a change from 17 to 23 Mpc, or from 23 to 32 Mpc, increases the HI mass by a factor 1.8 and 1.9 respectively. Equivalently, this would result in a decrease of absolute magnitude by 0.66 and 0.72 magnitudes respectively. The HI mass-to-light ratio is unaffected, since both HI mass and optical luminosity depend on distance squared. HI deficiency (see section 5.3.6) will be affected (since measured HI mass scales as distance squared but predicted HI mass scales just as distance) but only slightly. A change from 17 to 23 Mpc, or from 23 to 32 Mpc, will cause a change in HI deficiency by a factor of $\sim$0.9.

5.3.3 Morphology

Late-type galaxies

The morphological distribution for galaxies detected and not detected in HI is shown in figure 5.8. The classification scheme is fully described in chapter 3 - briefly, early-types are from -3 to 1, with late-types (ranging from spirals to dwarf irregulars) from 2 to 19. 20 represents unclassified objects. Unsurprisingly, most of the detections are of late-type objects, and most of the non-detections of early-type galaxies (particularly dwarf ellipticals). A visual inspection shows that the unclassified objects detected in HI are definitely late-type objects, most of them

probably dwarf irregulars or BCDs. One exception is source 51, which is clearly a very flat, edge-on disc. None could be described as early-type objects.

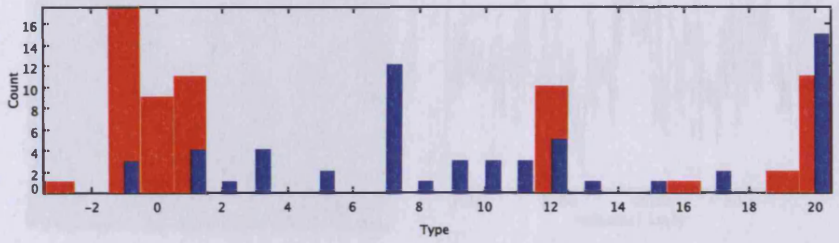


Figure 5.8: Morphology distribution for the AGES detections (blue) and VCC galaxies not detected in HI (red). The vertical scale has been truncated, there are a total of 75 non-detected dwarf ellipticals (type -1). 9 non-detections have been excluded as they are at the very edges of the cube where the noise is significantly stronger.

From the histogram it would appear that there are a significant number of late-type objects not detected in HI, up to 19 if unclassified objects are again assumed to be late-type. A more detailed examination shows that this is probably not the case. The undetected unclassified objects are all very faint and most could easily be early-type galaxies (such as faint dwarf ellipticals). To ensure no detections were missed, the spectrum at the position of each was re-examined. Only VCC 201 shows any sign of a possible detection - a 3.5σ, 40 km/s velocity width source 51 km/s from the optical redshift (this is not included further in this study as follow-up as not yet been obtained).

Of the rest, most are very faint objects that could be misclassified. 6 are close to strong HI detections so could be confused with the brighter objects. 2 have previously claimed HI detections in GOLDMine, though 1 is probably confused with VCC 571 (discussed below). The other, VCC 628, is given a quality of 4, and has been discussed in section 5.2.1. Since there are no other examples of a previously detected object being missed by AGES, it is likely that the claimed poor quality detection is actually spurious and this object is a dwarf elliptical rather than an irregular.

There is only one clear example of a late-type galaxy not detected by AGES, VCC 1435. AGES can place an HI mass limit of 7.3×10^6 $M_\odot$ for this galaxy, based on the rms of the spectrum and the peak S/N in a profile centred on its velocity (from optical redshift measurements in NED). Even deeper observations in GOLDMine place an upper mass limit of just 3.8×10^6 $M_\odot$. This is discussed further in section 5.3.4.

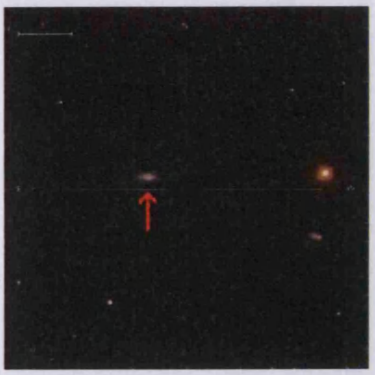
Early-type galaxies

There are a handful of early-type galaxies detected in HI. Finding charts and HI spectra are shown for these 7 objects in figure 5.9. The large red arrow indicates the candidate object selected, on which photometric measurements have been performed. Where appropriate, a smaller arrow indicates other potential candidates. These 7 early-type galaxies are clearly very unusual, given that a further 84 early-type objects listed for this region in the VCC remain undetected[1].

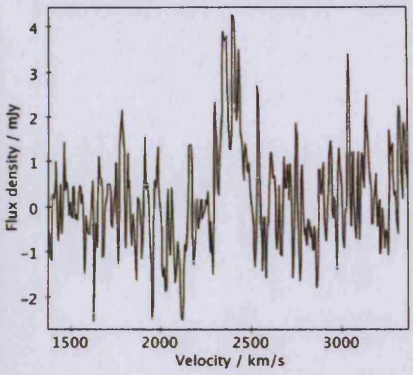
This detected fraction of 7.7%, is somewhat higher than the reported 2.3% from the ALFALFA survey (Alighieri et. al. 2007, Alighieri et. al. 2008). In some cases the identification of the optical counterpart is ambiguous. For source 47 both VCC 450 and VCC 479 are visible within the Arecibo beam. However, VCC 450 is the better match in both redshift (a difference of 40 km/s in the optical and HI, compared with 163 km/s for VCC 479) and spatial position (0.7' compared to 2.4' for VCC 479). Furthermore, VCC 479 is itself an early-type object, so the HI in this case is definitely associated with at least one ETG.

VCC 1394 is 2.4' from the coordinates of the HI, and no optical redshift is available. Fortunately, there are no other galaxies present within the Arecibo beam (except for some objects at much higher redshifts) so this object

[1]This is the number of undetected ETGs within the declination range +7° to +8°, the region of uniform noise coverage. Within the full limits of the cube there are actually 109 undetected ETGs.

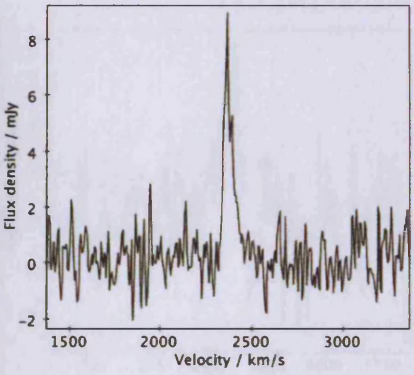


(a) Source 47 RGB, VCC 450. VCC 439 is visible to the west.



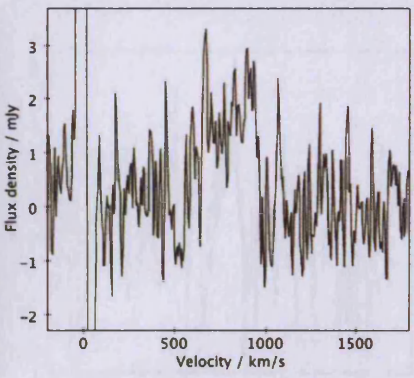
(b) Source 47 HI profile

(c) Source 56 RGB, VCC 190.



(d) Source 56 HI profile

(e) Source 79 RGB, VCC 758.



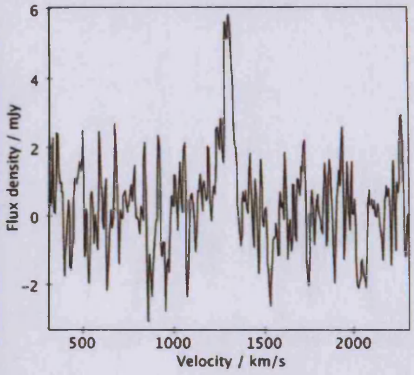
(f) Source 79 HI profile

Figure 5.9: SDSS RGB images and spectra for the early-type galaxies detected in Virgo. The finding charts are centred on the HI coordinates, the green line is 1' in length.

can probably be reliably identified as the optical counterpart. As mentioned in section 5.2.1, there is one other object with a sure optical counterpart with an even greater separation of the optical counterpart from the HI - although high, such a large separation of the HI and optical coordinates does occasionally occur.

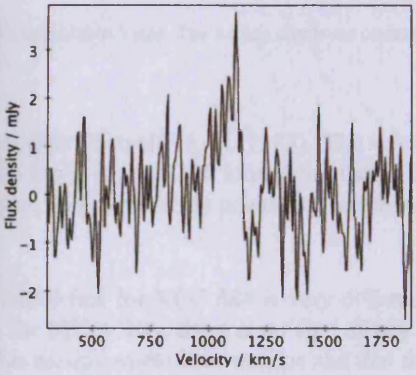
For source 133 the situation is more complex. The HI coordinates place the gas between VCC 571 and VCC 584 (a late-type object), 1.8' and 1.4' spatial separation respectively. As an extra complication, an uncatalogued galaxy is

(a) Source 117 RGB, VCC 1394. The blue disc to the west is at 11,209 km/s (SDSS).



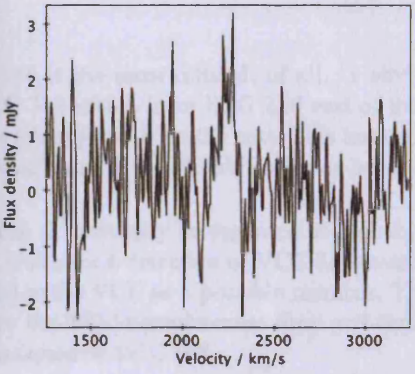
(b) Source 117 HI profile.

(c) Source 133 RGB, VCC 571 with VCC 584 to the south.



(d) Source 133 HI profile

(e) Source 145 RGB, VCC 180.

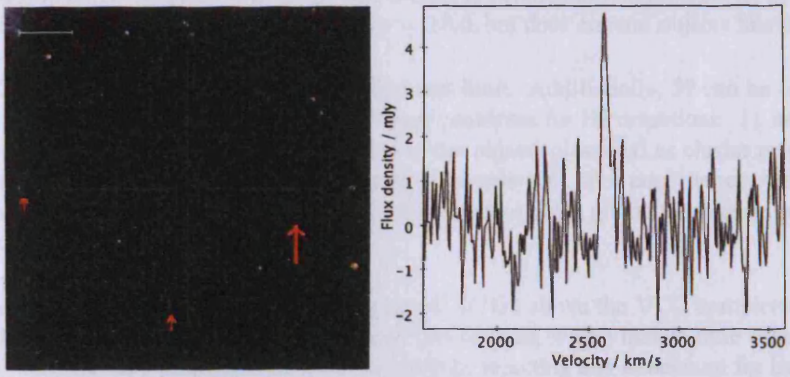


(f) Source 145 HI profile.

Figure 5.9: SDSS RGB images and spectra for the early-type galaxies detected in Virgo. The finding charts are centred on the HI coordinates, the green line is 1' in length.

present at the HI coordinates, but this appears to be a background object (though no optical redshift measurement is available). With an optical diameter of no more than 10" it would be exceptionally compact compared to most cluster members, especially considering its W20 of 161 km/s (but see chapters 6 and 8 for more discussion on this).

The measured velocity of the HI from AGES is 1,090 km/s, for which optical redshifts of VCC 571 agree very



(a) Source 148 RGB, VCC 368. Other candidates visible are not identified in any major catalogues. The bright VCC 393 is just visible to the north east.

(b) Source 148 HI profile.

Figure 5.9: SDSS RGB images and spectra for the early-type galaxies detected in Virgo. The finding charts are centred on the HI coordinates, the green line is 1' in length.

well - 1,080 km/s from the SDSS, and 1,047 km/s from NED (Bingelli et. al. 1985). The velocity of VCC 584 is more problematic. The redshift from the SDSS is 1,290 km/s, exactly 200 km/s from the HI redshift - the usual cutoff at which to reject a source as an optical counterpart. In this case, the positional match to the HI favours the late-type object, but the velocity favours the ETG.

To further complicate matters, the velocity listed in GOLDMine for VCC 584 is very different again : 56 km/s. Although this is in the region heavily contaminated by the Milky Way, there is no sign of any HI emission in the AGES cube at this location. It seems more likely that this measurement is erroneous and that the SDSS is correct. Given this, it still seems that VCC 571 is the optical counterpart. Although the velocity width of the profile is 160 km/s, it does not extend to 1290 km/s, as would be expected if VCC 584 were the candidate or the HI were shared between the two. A third possibility is that the HI originated in one or both galaxies, but has been completely removed and is now associated with neither of the two objects. This seems unlikely, given the very good velocity match to VCC 571.

Finally, identifying the optical counterpart of source 148 is the most difficult of all. 3 obvious candidates are visible, none with optical redshift measurements : VCC 368 which is an ETG 2.0' east of the HI coordinates, a late-type object 2.2' south of the HI, and a second late-type object 3.0' to the east. This last can probably be ruled out as being too far from the HI, but for the remaining candidates it is difficult to choose between them.

There are therefore 4 unambiguous early-type detections. 2 more may be regarded as probable as they seem the most likely match to the HI (VCC 1394 and 571). The claim for a detection of VCC 368 rests on the fact that the HI is within the cluster and VCC 368 has been assigned in the VCC as a possible member. The other candidates visible are not VCC members, but these are both below the VCC completeness limit and cannot be excluded as cluster objects. This detection can only be tentatively assigned to VCC 368.

Despite this, the detection rate is still slightly higher than that reported by ALFALFA, at least 4.4% and probably closer to 6.6%. If all ETGs contain some gas, then it is to be expected that a more sensitive survey will detect more of them, and the higher detection rate is not surprising. In this scenario, the detected ETGs are simply the most HI massive of the population - increasing the sensitivity further would result in more detections of ETGs. However, given the very low detection rates, an alternative scenario is also possible. The ETGs detected by AGES and ALFALFA could be very unusual objects, with the majority of ETGs containing no HI at all. In this case, increasing sensitivity will not necessarily increase the fraction detected in HI.

To test which of these possibilities is actually the case it is necessary to make a like-for-like comparison of the detection rates. The figure quoted by ALFALFA of 2.3% depends on the statistics used. ALFALFA actually detect 13 ETGs in the declination strip from +8° to +16°. This region contains 939 galaxies classed as early-type in the

VCC, which would give an HI-detected fraction of just 1.4%. As detailed in Alighieri et. al. 2007, this would not be a fair comparison as the VCC is only complete to $B_T = 18.0$, but does contain objects fainter than this.

Only 457 of the 939 are brighter than the VCC completeness limit. Additionally, 59 can be rejected as they are within 1° of M87, a strong continuum source which causes problems for HI detections. 11 more are rejected to account for the percentage of background contamination in the objects classified as cluster members, but lacking a velocity measurement. This leaves 387 ETGs as the total sample size. For consistency, 4 ETG HI detections must be rejected as they are of galaxies below the VCC completeness limit, giving 9 detections out of 387, i.e., the quoted detection rate of 2.3%.

Applying the same criteria to the AGES sample, there are 45 ETGs above the VCC completeness limit, using g band photometry (there are 47 if the photographic magnitude is used, which makes little difference). Only 3 of these do not have a velocity measurement, too few to necessitate rejecting any to account for background contamination. Of the sample detected in HI 5 are above the VCC completeness limit, a detected fraction of 11.1%. Even rejecting the (slightly) ambiguous case of VCC 571 still gives a detection rate of 8.9%. This is much higher than that of ALFALFA, though lower than some previous estimates, such as the 15% reported by Conselice et. al. 2003 (hereafter C03).

Sensitivity effects can account for the difference in detection rates between AGES and ALFALFA. Assuming the typical ALFALFA rms noise of 2.2 mJy, all of the detected ETGs would have a S/N below ALFALFA's 6.5 criteria - the highest would only be 4.99, the lowest, 0.45. Additionally, AGES has a typical rms a factor 3.7 lower than ALFALFA, and a detection rate of ETGs at least a factor 3.8 higher. This implies that it is indeed only a sensitivity effect that ALFALFA's detection rate is lower than that from other studies.

A caveat is that the results of stacking argue against this - increasing the sensitivity further does not appear to result in a higher detection rate (but see chapter 7 for details and the many caveats of this technique). The results of C03 also give a more direct argument for this. Many observations in C03 are significantly more sensitive than AGES, with a sensitivity limit of approximately 2×10^6 $M_\odot$, but the weakest detection listed by C03 is more than 10 times more massive than this, at 3×10^7 $M_\odot$. This is higher than the AGES sensitivity limit and the same as ALFALFA's sensitivity limit. Going deeper in this case has not resulted in the detection of less massive galaxies.

This makes it difficult to explain the much higher detection rate of C03 by sensitivity alone, since both AGES and ALFALFA have mass sensitivity limits comparable to the lowest mass actually detected in ETGs by the much more sensitive observations. However, the 15% quoted by C03 includes archival detections from the literature, with unknown selection effects and biases. From the original observations performed by C03, the detection rate is actually 9%. The sample observed by C03 was a random set of dEs with apparent B magnitude < 17.0, or $M_B < -14.0$. None of the dEs detected by AGES are as bright as this - this means a comparison of AGES detection rates with C03 is actually 0% compared to 9%.

A major problem here is of course the small number statistics, since AGES detects 7 ETGs of which only 3 are dEs. Of the 92 dEs present in the region, just 21 are brighter than $g = 17.0$. C03 observed such 22 dEs and detected 2. Such small numbers make it difficult to draw any firm conclusions. In short, it is not yet possible, given the state of observations, to say whether those ETGs detected in HI are simply the most HI-rich galaxies of their population, or are truly exceptional objects with most ETGs being really devoid of gas. However, it is fair to say that the ALFALFA detection rate (2.3%), which includes ETGs other than dEs, is lower than that of AGES (11.1%), and this is probably due to sensitivity.

One final note is that the AGES detected fraction of dEs (3.3%) is very much lower than the fraction of S0s detected in HI (20.0%, again by AGES). This further highlights the differences between such objects (as will be further discussed throughout this chapter), and taking all ETGs to be a single class of objects is a mistake.

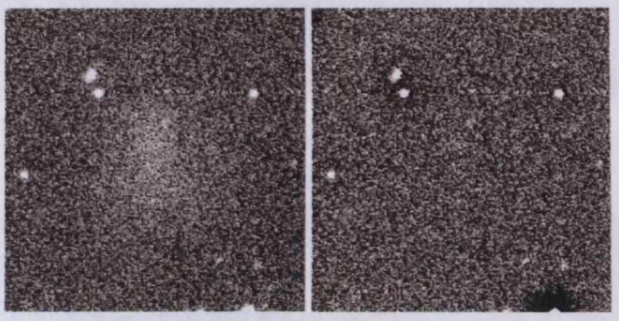
Having established that most of these 7 supposed early-type detections are indeed most likely associated with objects classified as early-type, it is also important to confirm that those objects really are ETGs and not misclassified objects (although C03 argue that this may not matter, provided the classification of the VCC is consistent). ALFALFA argue that several of their detected ETGs show peculiar morphological characteristics and are only borderline ETGs. For example, M86 "has a complex structure of filaments and arcs" in the H-alpha filter. This does not appear to be the case for the AGES detections, which, with the exception of VCC 758 with its clear dust lane/disc, appear smooth and featureless in optical, UV and H-alpha wavebands.

Following Lisker et. al. 2006 (hereafter L06), two methods are employed to enhance any possible structures. The first is to create a residual image. A symmetrical model of each galaxy is created using the IRAF tasks *ellipse* and *bmodel*, which is subtracted using *imarit* to reveal any asymmetrical features. The position angle and ellipticity are fixed, otherwise, as shown in L06, there is a risk the model image will also incorporate asymmetrical features. This technique is only possible for 3 of the early-type objects detected by AGES. 3 more are too small and too faint for *ellipse* to create a successful model. For VCC 758, there seems little point attempting such a procedure given its obvious central dust lane or disc.

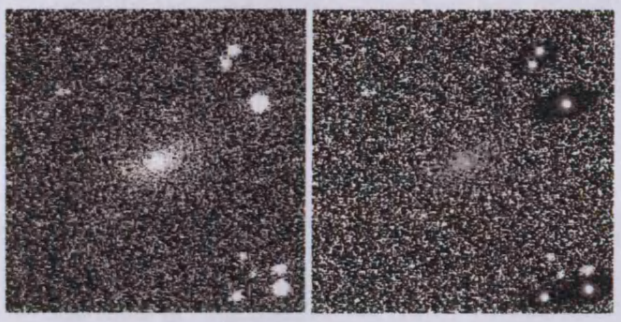
The second method is to create an unsharp mask. This is done by Gaussian-smoothing the image with various kernel sizes (using the the *gauss* task in IRAF) in order to search for features of different sizes. Kernel sizes of 3, 5, 10 and 20 pixels are used. The original image is then divided by the smoothed image using *imarit*. This has the advantage of being possible for all objects (though of course there is still no benefit in attempting this for VCC 758). Adopting the procedure in L06, both circular and elliptical unsharp mask images are created. This is because both can give rise to artificial structures and serve as a complimentary check on the other. For elliptical kernels, the ratio of the minor to major axis and position angle is found using IRAF. Residual images and unsharp masks are shown in figures 5.10 and 5.11.

Most of the detections show little evidence of structure. As a comparison, the type 2 galaxy VCC 94 is also shown. With clear spiral structure evident in both the unsharp mask and the residual image, this demonstrates that the techniques used here are able to reveal extended features not visible in the SDSS RGB images. A caveat is the low luminosity of the dwarf ellipticals, but this is not an issue for the lenticulars.

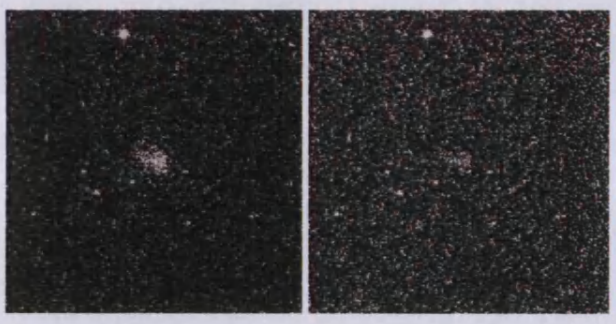
VCC 180 does show a bar-like feature in its center, which is seen most clearly in the unsharp mask but is also visible in the residual image. Only 2 out of the 7 detected ETGs seem to have any kind of unusual morphological feature, and that within VCC 180 is relatively small. There is certainly nothing approaching that seen in L06 figure 7, or VCC 94, which shows clear spiral structure extending over the whole diameter of the galaxies. As a whole, the ETGs detected here are not even very unusual early-type objects, and certainly not misclassified late-type objects.



(a) VCC 190 g band **(b)** VCC 190 unsharp mask

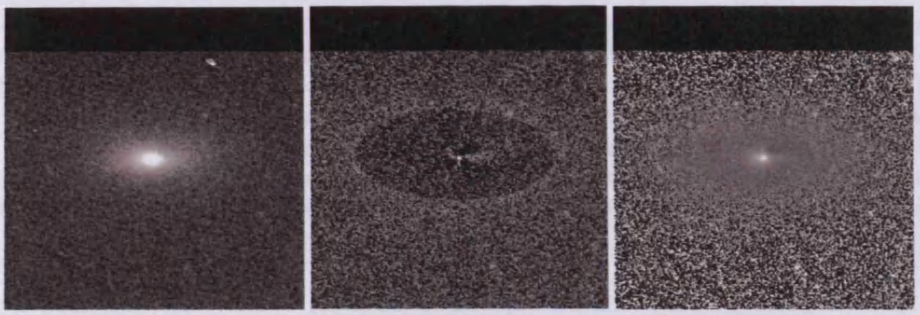


(c) VCC 571 g band **(d)** VCC 571 unsharp mask

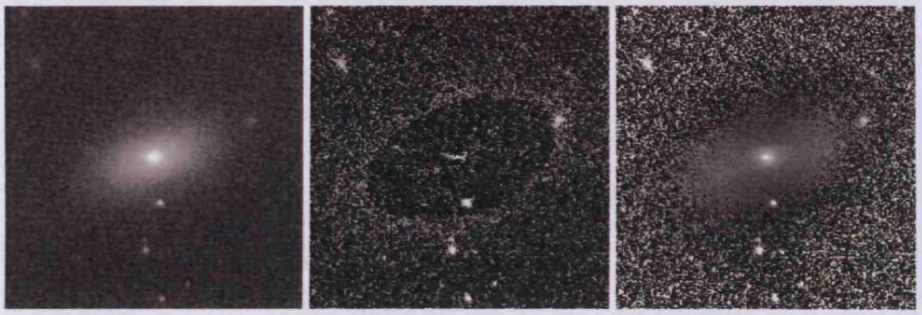


(e) VCC 180 g band **(f)** VCC 180 unsharp mask

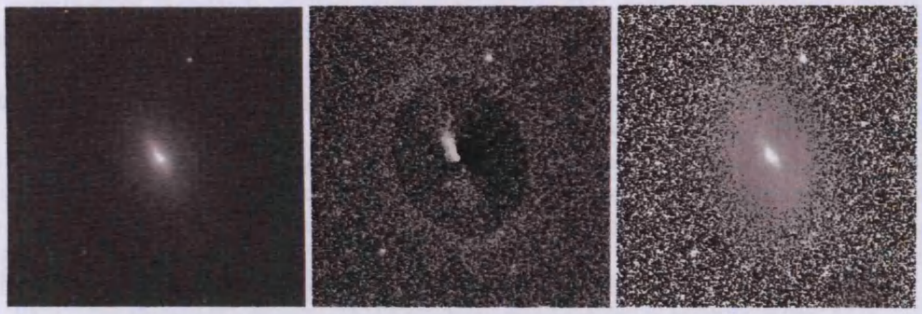
Figure 5.10: SDSS g-band and unsharp masks for the dwarf ellipticals, using elliptical profiles with kernel sizes of 10 pixels. Images are 1.7' across. These objects are too faint for IRAF to create a usable model for creating a residual image, hence only unsharp masking was attempted.



(a) VCC 450 g band (b) VCC 450 residual (c) VCC 450 unsharp



(d) VCC 571 g band (e) VCC 571 residual (f) VCC 571 unsharp



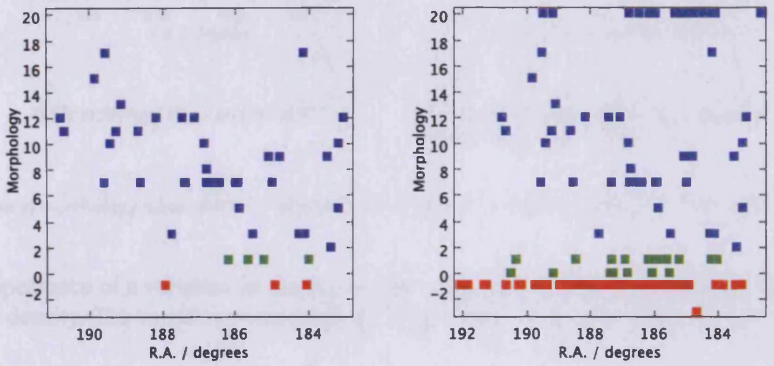
(g) VCC 180 g band (h) VCC 180 residual (i) VCC 180 unsharp

(j) VCC 94 g band (k) VCC 94 residual (l) VCC 94 unsharp

Figure 5.11: SDSS g-band, residual images and unsharp masks (using elliptical profiles with kernel sizes of 20 pixels) for the S0s detected in HI. Images are 1.7' across. The model from IRAF does not quite extend to the outer edge of each galaxy, hence the appearance of a cutoff ring in the residual images. The type 2 VCC 94 is shown for comparison.

Distribution

The well-known morphology-density relation describes the relative abundance of ETGs in the dense cluster compared with the dominance of late-type objects in the field (Dressler 1980). Interestingly, a similar relation[2] is perhaps evident in the HI-detected objects in this region - see the left panel of figure 5.12. Galaxies of earlier morphology are preferentially found in the denser (68% of the VCC cluster members are found here) western half of the cube. If, however, the unclassified objects and the non-detections are included this possible trend becomes far less apparent (right panel of same figure) to the human eye. Calculations of Pearson's product-moment correlation coefficient show that the trend in the left panel is significant at the 99.6% level, that in the right panel at just 8.4%.



(a) Morphology of the HI detections, with an existing classification, varying across the cube.

(b) Morphology of all known objects varying across the cube

Figure 5.12: Variation in morphology across the VC1 cube. Blue for late-type objects, green for S0s and red for dEs.

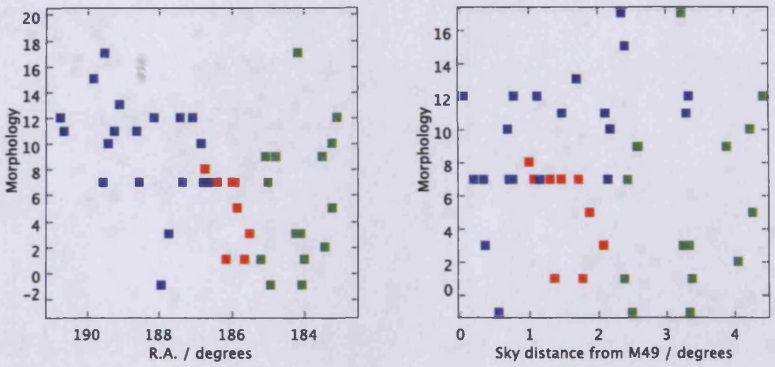
Of those galaxies with HI detections and an existing morphological classification, just 36% are in the eastern half of the cube, with 64% in the western side. The difference in morphology is also clear. Of those of type 10 or later (dwarfs and irregulars), 67% are in the eastern half. For those below type 10 (spirals and early-types), only 20% are in the east with 80% in the west. The difference is most clear in terms of the true early-type galaxies, with only 15% in the eastern half and 85% in the west. The cube thus covers a region divided into a low density volume populated mainly by late-type objects, and a high density volume richer in earlier-type galaxies. Interestingly, the true ETGs detected in HI seem to follow the relation, whereas those without HI do not.

This apparent trend is complicated as the galaxies are distributed in 3 different clouds at 3 different distances (see figure 5.7). Whether it really represents a morphology-density relation is difficult to determine. Estimating the projected density is not straightforward, and estimating the true volume density is impossible without individual distance measurements. R.A. may only be a convenient but crude proxy for density, if it even relates to true volume density at all. Figure 5.13 shows the same relation but with colour indicating the supposed distance. Morphology as a function of distance across the sky from M49 is plotted for comparison.

There is at least evidence for a trend for those galaxies at 17 and 23 Mpc distance when plotting R.A. There are too few objects to say whether those at 32 Mpc also follow this trend, but they are perhaps more scattered. Although weak, the trend does appear to be real - no such relation (to 99.8% confidence) is observed with distance across the sky from M49, as shown in figure 5.13, or with declination or velocity. While there is a trend, whether it has any physical relevance is another matter.

Were no distance information available, it would seem that galaxies in the denser western region are of earlier morphological type, which could be interpreted to suggest evolution with density. The estimates of distance suggest that it is instead only a curious projection effect. Going across the cube, galaxies are found in clouds at different distances. Each cloud is populated with galaxies of different morphologies. It is the position of these clouds that

[2]This is not exactly the same as Dressler's relation, which describes the *fraction* of galaxies that are of a particular morphology, but the principle is the same.



(a) Morphology as a function of R.A.

(b) Morphology as a function of projected distance from M49 instead of R.A.

Figure 5.13: Variation in morphology across the VC1 cube with colour indicating distance - blue for 17 Mpc, red for 23 and green for 32.

gives rise to the appearance of a variation in density and morphology, but this sky density probably has little relation to true volume density. The variation in morphology in this case cannot be interpreted as evidence of evolution.

5.3.4 Colours

Colour-magnitude diagrams for the HI detections are shown in figure 5.14. As a comparison, photometry was also performed for the undetected galaxies with optical redshift measurements (54 objects) as well as the background HI-detected objects (73 objects). The requirement for an optical redshift ensures the distance to the object can be assigned as described in section 5.3.2, for consistent absolute magnitudes. The majority of the non detections are unambiguously early-type objects.

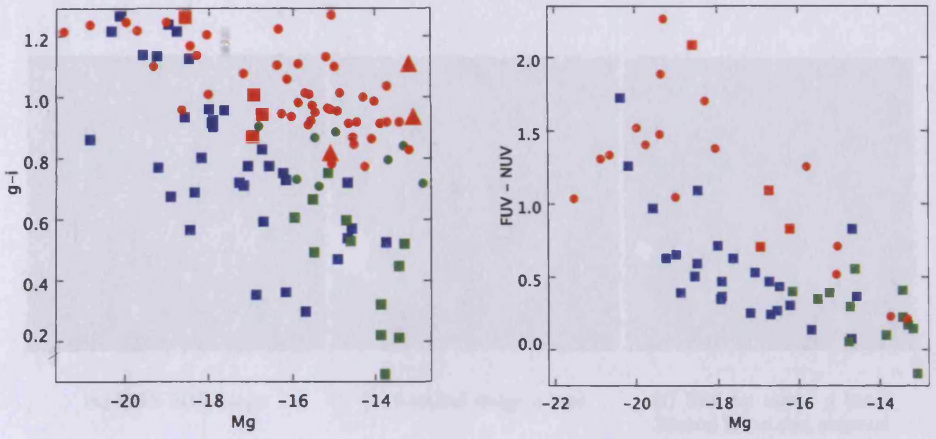
At the most basic level, the well-known result of brighter galaxies being redder than fainter galaxies is clearly seen. Objects in Virgo appear systematically redder than those in the background of the same luminosity, which could suggest an environmental influence of the cluster. It is difficult to disentangle this possibility from selection effects - objects in the background are selected by HI, and being more distant, must be richer in HI and therefore more likely to be bluer in comparison to *any* nearby detections, not just those in the cluster (see next section).

Separating the cluster objects by morphological type shows a red and blue sequence. The unclassified objects detected in HI are seen to lie on the blue sequence, indicating that they are indeed late-type objects, as a visual inspection indicated in section 5.3.3. Conversely, the unclassified objects not detected in HI lie on the red sequence, suggesting they are early-type objects, again as previously indicated.

The UV data provides a useful independent check on the photometry and broadly confirms the trends seen in the optical colour-magnitude diagrams. The red sequence is more difficult to discern in the UV due to the fewer number (only 16) of early-type galaxies that were detected in both UV bands. This is not surprising, since UV is believed to be generally associated with young stars and few early-type galaxies show ongoing star formation.

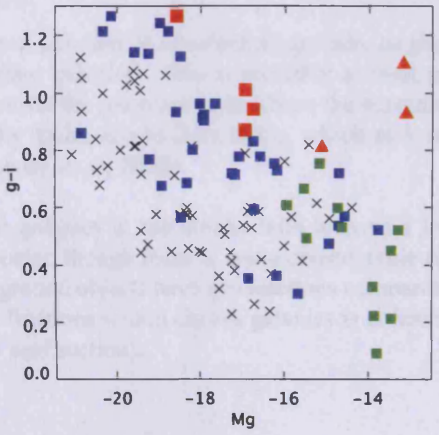
Only 3 unclassified objects not detected in HI have optical redshift measurements. 2 of these lie clearly on the red sequence with the third in the blue sequence (VCC 479). This does indeed appear to be a late-type object, but is probably too far north for AGES to detect it (an HI detection is given in GOLDMine). All of the HI-detected unclassified objects are found on the blue sequence - they only lack a morphological classification as they are optically faint.

There are 5 undetected objects classed as late-type. Only VCC 2007 is definitely on the blue sequence, but is too far north for AGES to detect (again an HI detection is actually recorded in GOLDMine). The 2 faintest objects, VCC 867 and VCC 628, are on the red sequence. The classification of irregular may be erroneous as both of these objects are very faint - their red colour and lack of HI suggests they are actually early-type objects. The remaining



(a) Optical colour-magnitude diagram for cluster members only.

(b) UV colour-magnitude diagram for cluster members only.



(c) Optical colour-magnitude diagram for cluster members and background objects detected in HI.

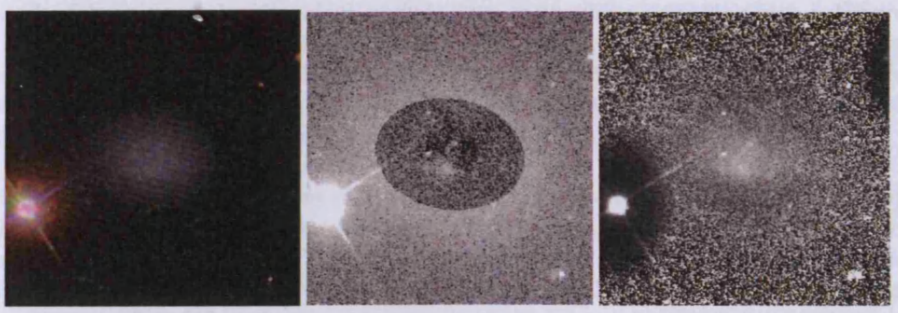
Figure 5.14: Optical and UV colour-magnitude diagrams for cluster members and background objects. Blue indicates late-type objects, red early-type and green unclassified. Squares indicate that HI is detected, circles than it is not. Red squares are detected S0s, red triangles are detected dEs, and black crosses represent background detections. No correction was made for internal extinction.

2 galaxies are VCC 584 and VCC 1435. VCC 584 has been discussed previously - see section 5.3.2 and figure 5.9.

VCC 1435 has also been previously mentioned as the only definite LTG not detected in HI. Its colour places it on the border between the red and blue sequence, while its appearance in the SDSS RGB image is of a smooth, featureless blue galaxy. This is shown with a residual image in figure 5.15, which reveals that some structure is definitely present. Together with its reasonably blue colour, this confirms the galaxy as late-type, despite its lack of detected HI.

The 4 detected lenticulars are at the boundary between the red and blue sequence. They are certainly redder than most late-type detections of similar magnitude, but not as red as the undetected early-types. They may be described as transition objects. In contrast, at least 2 of the 3 early-type detections (VCC 1384 and VCC 368) are unquestionably red sequence objects. The other, VCC 190, seems to be in the transition region. The UV data clearly shows that the detected ETGs are redder than the late-type objects. Since there are so few red sequence objects detected

in both UV bands, it is difficult to use this to confirm if the detected ETGs are red sequence or transition objects.



(a) SDSS RGB image **(b)** Residual image, g band **(c)** Unsharp mask, g band, 20 pixel kernel size, elliptical profile

Figure 5.15: Images of VCC 1435 (100" across), the only definite late-type not detected in HI by AGES.

5.3.5 HI mass-to-light ratio

The HI mass-to-light ratio varies as a function of absolute magnitude, as shown in figure 5.16. Fainter galaxies tend to be more gas-rich than brighter galaxies. This is probably at least partially a sensitivity effect - fainter galaxies require a higher gas fraction for the gas mass to be above the sensitivity limit. This does not explain why brighter galaxies do not have equally high mass-to-light ratios, which may relate to a "saturation" density of HI above which all gas is molecular (Leroy et. al. 2008).

It is apparent from figure 5.16 that galaxies in the cluster tend to have a lower gas mass-to-light ratio (or gas fraction) than objects behind the cluster, though there is some considerable scatter and in large part likely due to sensitivity effects. Only a few background objects have gas fractions comparable to cluster objects, and the reverse is also true. The relatively low gas fractions within cluster galaxies is at least partly a real effect of environment, and not just a sensitivity effect (see next section).

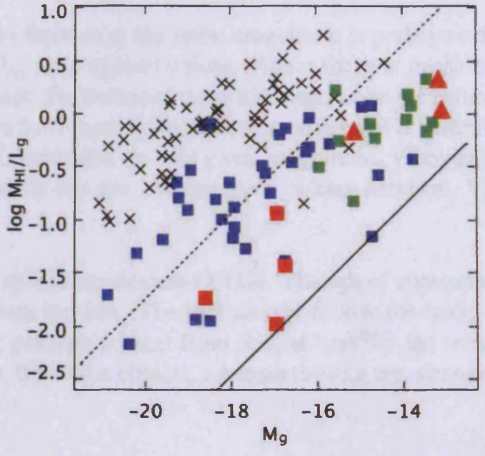


Figure 5.16: HI mass-to-light ratio. The lines indicate the sensitivity limits for a 4σ, 50 km/s tophat profile at 0.5 mJy rms at 17Mpc (solid line) and 100 Mpc (dashed line). The y-axis is in logarithmic units. Blue squares are late-types, red are S0s and green are unclassified objects. Red triangles are dEs and black crosses are background detections.

There is a clear split in the detected ETGs. The dEs are relatively faint and gas-rich, of comparable M_{HI}/L_g ratio to the late-types of similar magnitude (i.e. dwarf irregulars). The detected S0s are brighter and perhaps also rather

poorer in gas content compared to the other detections of similar magnitude (i.e. spirals).

Two background objects are detected with gas fractions comparable to those in the cluster - sources 31 and 38 (VCC 694). Both of these have velocities which exclude them as possible cluster members, 6,436 and 4,210 km/s respectively. Both also appear reasonably isolated, with no obvious interactors that could account for the removal of their gas content. They are also both relatively weak detections, with S/N peaks of 5.28 and 7.76 respectively. The existence of these objects implies that the apparently simple linear trend in gas fraction with luminosity might be radically changed with deeper observations, which would be better able to detect such galaxies.

Two cluster objects can be readily identified as having gas fractions more typical of background objects, sources 63 (VCC 827) and 64 (VCC 975). These are very similar objects - large, bright, relatively blue spirals with very similar gas mass-to-light ratios of about 0.77, and both have comparable absolute magnitudes of about -18.5. The nearest match to this within the cluster is VCC 105 (absolute magnitude -18.9) with a ratio less than half of this. In a more extreme case, VCC 534, with again a magnitude of -18.5, has a gas fraction nearly 80 times lower than these two sources.

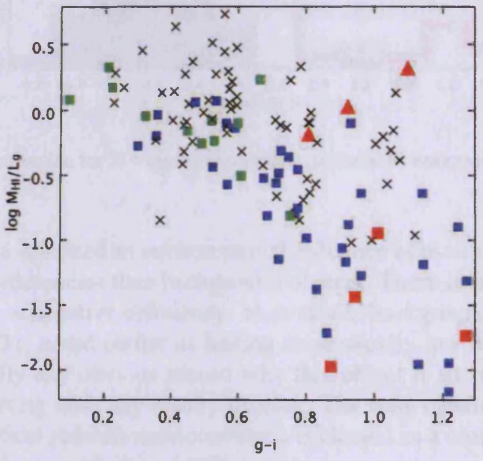


Figure 5.17: HI mass-to-light ratio (logarithmic scale) varying with g-i colour. Blue squares are late-types, green are unclassified and red are S0s. Red triangles are dEs, and black crosses are background galaxies

The effect of such variation in gas fraction at the same magnitude is perhaps reflected in the colours of the galaxies. Figure 5.17 plots the M_{HI}/L_g ratio against colour. Again there is much scatter but a clear trend - the most gas-rich galaxies are also the bluest. Furthermore these also tend to be the faintest objects with the most late-type morphology. In contrast to figure 5.16 there appears little evidence of a systematic difference in gas fraction between the Virgo and background detections. At any given magnitude, Virgo galaxies seem to be redder than their background counterparts and contain less gas. At any given colour, however, Virgo galaxies possess just as much gas as equivalent field galaxies.

As in figure 5.16, there is a clear split in the detected ETGs. Though of approximately the same g-i colour, the dEs are significantly more gas-rich than the S0s. The S0s clearly follow the basic trend of redder colour with lower gas fraction, whereas the dEs are perhaps deviant from this, at least for the trend shown by cluster members. The S0s may be described as gas-poor, transition objects, whereas the dEs are, strangely, gas-rich red sequence galaxies.

5.3.6 HI deficiency

As described in chapter 1, the HI deficiency is a measure of how much gas a galaxy of given morphological type and optical diameter has lost in comparison to a similar field galaxy. Since this has been calibrated on a field sample it is a more precise measure of actual gas loss than comparing relative M_{HI}/L_g ratios. The calculation of HI deficiency is shown in equations 1.1 and 1.2. The a and b parameters of field galaxies have only been calculated for galaxies of morphological type 2 through 7 (see table 3.1).Following Gavazzi 2008, for later-type objects the

parameters for an Sc galaxy are used (deficiency for these objects should be considered highly uncertain, since spirals are fundamentally different objects to irregulars). Early-type galaxies are rarely detected in HI so it makes little sense to speak of deficiency for these objects.

The distribution of HI deficiency is shown in figure 5.18. The deficiency for an isolated object is typically -0.3 < HI_{def} < +0.3. The mean deficiency in Virgo is 0.7, with 83% greater than 0.3 (89% if the objects without a morphological classification are excluded). In comparison, the mean deficiency in the background objects is 0.2. The distribution of deficiencies in cluster galaxies is bimodal, with 67% less deficient than 0.8, but with the remaining 33% extending up to deficiencies as high as 1.8 (again there is little change if the unclassified objects are excluded, and these fractions do not include the detected ETGs).

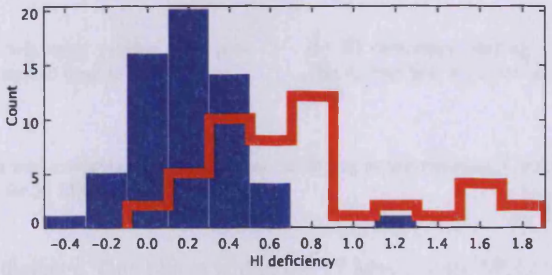


Figure 5.18: HI deficiency distribution for 52 Virgo cluster objects (red) and 61 background objects (blue), bin size 0.2.

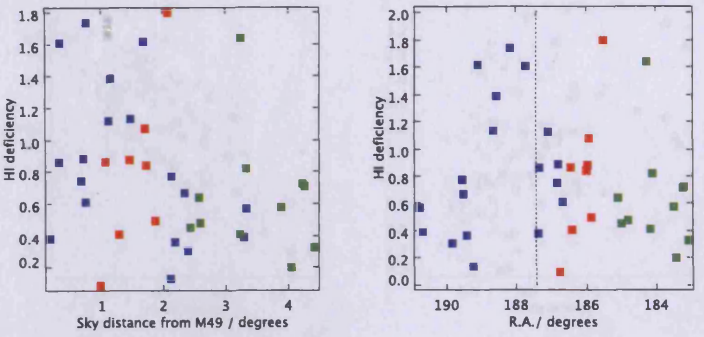
The deficiencies confirm that there is indeed an environmental influence of the cluster acting to remove gas. Cluster galaxies typically show greater deficiencies than background objects. There is no suggestion of any cluster object with more gas than expected, i.e. a negative deficiency. In contrast, background objects have deficiencies as low as -0.4. The background source 31, noted earlier as having an unusually low M_{HI}/L_g ratio, is indeed unusually deficient at 1.2. There is not really any obvious reason why this object is so strongly deficient since it does not appear to be associated or interacting with any nearby objects. The only candidate is VCC 744 (1.1' away), but although this does not have an optical redshift measurement it is classed as a cluster member, and so unlikely to be associated with object 31 (which is at a velocity of 6436 km/s).

While there is a bimodality in the distribution of HI deficiencies of cluster galaxies, this is not reproduced in their sky location or in position-velocity space. Galaxies of a wide range of deficiency are found in all 3 clouds within the surveyed volume. However, it does seem to be visible in their distance from M49, the bright elliptical at the center of the cube (when observations are complete) - see figure 5.19. The cube extends to about 5 degrees to the east and west of M49. 15 galaxies in total have deficiency above 0.8, of which, 87% are within 2.5 degrees. This bias in the distribution is not reproduced in the galaxies of lower deficiency, which are distributed evenly with distance from M49 - of the 22 objects below 0.8, 54% are within 2.5 degrees of M49.

Deficiency is seen to remain high even as far as 4 degrees from M49, which indicates that the environmental influence of the cluster probably continues beyond the edge of the region. The higher deficiency closer to M49 may be a result of the denser X-ray gas (see contours in figure 5.1). As with the possible variation in morphology discussed in section 5.3.3., the situation is complicated by the 3 separate galaxy clouds - projected distance from M49 is at best a very crude proxy for X-ray emitting gas density. It is interesting to note that while there seems to be a trend in morphology with R.A., and specifically not distance from M49, the opposite is true for deficiency.

The trend with distance from M49 is perhaps enhanced by considering the galaxies according to the supposed distance of the cloud they inhabit. At 17 Mpc, galaxies are seen to be of high deficiency, but only within about 1.5 degrees of M49. Beyond this they are of relatively low deficiency. Between 1 and 2 degrees from M49, galaxies are at 23 Mpc and mainly of intermediate deficiency - lower than the strongly deficient objects close to M49, but still higher than those further away. The 32 Mpc group lies furthest from M49 and is of generally low deficiency.

There are of course outliers from this general trend. One galaxy at 23 Mpc is actually the most deficient in the entire sample at 1.79 (VCC 534), though it is still relatively close on the sky to M49 at 2.0 degrees projected distance. Another, VCC 222, is also strongly deficient (1.6) but is not only in the 32 Mpc group, but also far from



(a) HI deficiency varying with projected distance from M49.

(b) HI deficiency varying with R.A. The dashed line indicates the R.A. of M49.

Figure 5.19: HI deficiency of objects with a morphological classification varying across the cube. Colour indicates line-of-sight distance, blue for 17 Mpc, red for 23 Mpc and blue for 32 Mpc.

M49 at 3.2 degrees projected distance. One object within the 17 Mpc group, VCC 1205 is very close to M49 (0.18 degrees) but is only borderline deficient at 0.38.

The M_{HI}/L_g ratio correlates with HI deficiency, as shown in the left panel of figure 5.20. This is to be expected given that both HI deficiency and M_{HI}/L_g ratio are related to the current M_{HI}. Since $log(M/L)$ is the same as $log(M) - log(L)$, HI_{def} can also be written as :

$$HI_{def} = -log\frac{HI_{obs}}{L_g} + log\frac{HI_{ref}}{L_g} \tag{5.1}$$

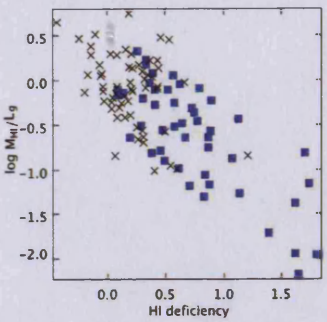
This gives a slope of -1 and indicates the relationship will be strongly scattered, since the intercept depends on the reference galaxy which is unique to each object. A linear regression fit to the data finds that in practise, the HI deficiency is related to the M_{HI}/L_g ratio by the following :

$$HI_{def} = -1.01 \times M_{HI}/L_g + 0.10 \tag{5.2}$$

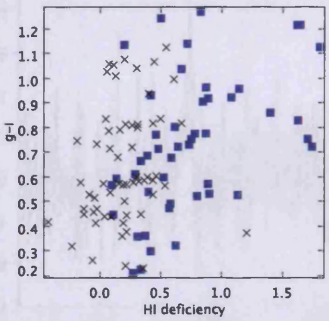
Where M_{HI}/L_g is in logarithmic units. This can then be used to estimate deficiencies for galaxies for which the standard method may be difficult, i.e. an unclear or early morphological type. Since it has already been shown in figure 5.17 that galaxy colour is related to its gas fraction, it might be expected that there should also be a colour-HI deficiency correlation. In fact there is little to suggest that this is the case, as shown in the right panel of figure 5.20. While the M_{HI}/L_g ratio does relate to colour, the fraction of gas a galaxy has lost does not. Physically this makes intuitive sense - it is only the amount of gas within a galaxy that can influence its star formation and hence colour, not the amount of gas it has lost.

VCC 827 and VCC 975 were earlier identified as having unusually high M_{HI}/L_g ratios. VCC 827 is borderline deficient at 0.40, while VCC 975 is non-deficient at 0.08. This is consistent with the scatter seen in figure 5.20. The M_{HI}/L_g ratio of the ETGs detected in HI have a high range from (in logarithmic units) -1.98 to +0.29. If these follow the trend of the late-type galaxies shown in the figure, equation 5.2 predicts a correspondingly high range of deficiencies from 2.05 to -0.27. There is of course much scatter in the relation, as expected.

Deficiencies as large as 2.0 are however too high to attribute solely to scatter in the M_{HI}/L_g ratio-HI deficiency trend. This indicates that some of these ETGs have probably been affected by the cluster environment. Interestingly, only the S0s appear to be deficient from this equation, with a minimum predicted deficiency of 1.1 (strongly deficient). In contrast the dEs are predicted to be non-deficient, with a maximum predicted deficiency of 0.3 (borderline non-deficient). The S0s and dEs are fundamentally different : the S0s are brighter, of lower M_{HI}/L_g ratio, probably HI deficient but not uncommonly detected in HI, and from the colour-magnitude diagrams appear to be transition objects. The dEs are fainter, of M_{HI}/L_g ratios comparable to dwarf ellipticals, probably non-deficient



(a) HI mass-to-light ratio in logarithmic units, varying as a function of estimated HI deficiency.



(b) Colour as a function of HI deficiency

Figure 5.20: Relations between HI deficiency, M_{HI}/L_g ratio and colour. Blue squares are Virgo detections while black crosses are background galaxies.

but rarely detected in HI, and seem to be red sequence objects.

5.3.7 Dark galaxies and other exotica

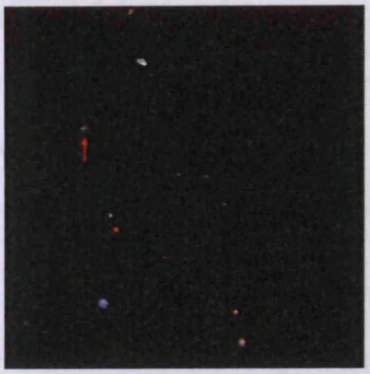
In this survey, 82 objects have within the Arecibo beam optical counterparts for which the HI redshift matches an available optical redshift measurement. For the remaining 51, in most cases only a single possible counterpart is visible, usually a blue object close to the coordinates of the HI. In these cases the candidates thus seem very probably associated with the HI, despite the lack of optical redshift measurements. In a very few cases the optical counterpart cannot easily be identified, for example, source 148 which has been tentatively identified as VCC 368, but for which several other possibilities exist (see section 5.3.3).

Three objects remain even more problematic. Finding charts and spectra for these are shown in figure 5.21. The first of these, source 128, has follow-up observations but there is some weak continuum emission at its location. It is possible this source is actually just baseline ripple caused by the continuum source. However, this purported detection corresponds to the highest peak S/N along the entire spectrum, including the background region. If it is ripple, other, similar features should also be found, but none exist. Strong continuum sources also cause a clear low-frequency structure in the baseline, which is not seen in this spectrum.

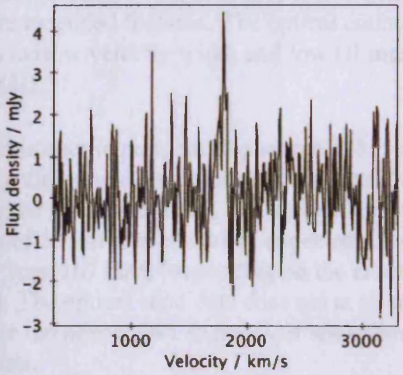
If it is really HI, it would have a M_{HI} of just 4.16×10^6 $M_\odot$, making it the least massive detection in the sample. A single possible candidate has been identified. If this is indeed the counterpart, it would have a M_{HI}/L_g ratio of 0.48 and a g-i colour of 0.72, neither of which seem unreasonable given its small size. Its distance of 1.1' from the HI coordinates is also perfectly normal.

This candidate would, however, be extremely faint, with an absolute g magnitude of -12.0 making it the faintest object in the sample - if not a dark galaxy then at least a very dim one. It would also be the second most compact object detected, with an optical diameter just below 1 Kpc. While this is consistent with its small mass, its velocity width of up to 180 km/s would more ordinarily be associated with larger, more obviously disc-like objects (a caveat is the poor S/N ratio, which makes an accurate measure of the velocity width difficult). Deeper INT images are available for this object, but neither they nor smoothing the SDSS data reveals any kind of larger structure. Without at least spectroscopic observations to confirm the optical redshift of the candidate, it must remain enigmatic.

A second dark galaxy candidate, source 129, has really no plausible candidates at all. That shown in figure 5.21 and described in table 5.3 corresponds to a random smudge that would be the smallest object in the sample as well as one of the faintest. This weak detection (peak S/N 5.04) has the fourth lowest HI mass in the sample at 1.0×10^7 $M_\odot$. It has been confirmed with follow-up, and unlike source 128, there seems no chance it could be confused with continuum emission. The nearest detection is actually source 128, itself uncertain. The nearest sure source is 122 (VCC 1726) but this is over 23' away and at 1,549 km/s lower velocity. It is difficult to see how this object could

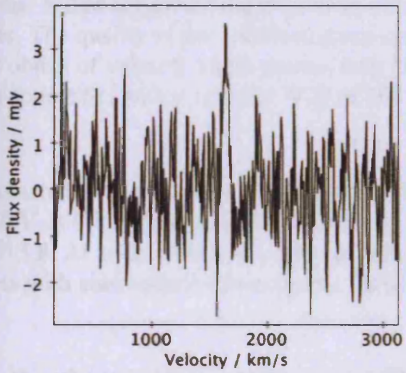


(a) Source 128

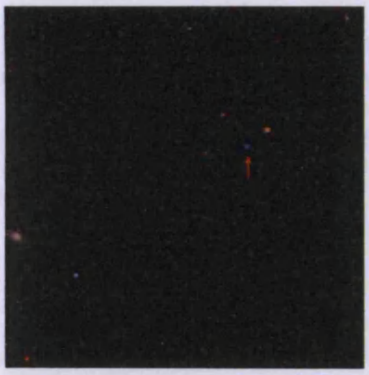


(b) Source 128

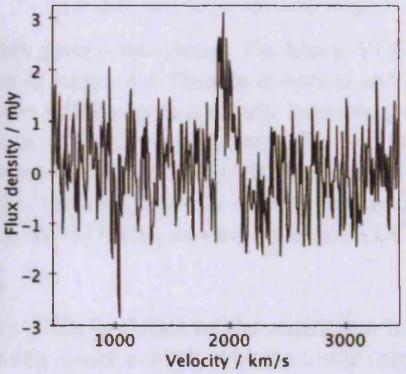
(c) Source 129



(d) Source 129



(e) Source 67



(f) Source 67

Figure 5.21: Dark galaxy candidates - SDSS RGB images (3.5' across, possible optical counterparts are marked with an arrow) and HI spectra (each spans 3,000 km/s of baseline).

have originated from any of the nearby galaxies, particularly given the complete lack of any extended HI features that might indicate a tidal interaction.

While the velocity width of this source of no more than 110 km/s is low, this is not enough to preclude it as a dark galaxy. Indeed, simulations of Davies et al. 2006 show that the majority of dark galaxies will be of extremely narrow velocity widths, no more than 40 km/s. Observationally, object 66, with a W20 of just 86 km/s, resembles a small edge-on disc. However, to properly asses the dynamical mass requires aperture synthesis observations to

determine if object 129 really is a rotating disc or not. As with source 128, both deeper INT imaging and smoothing the SDSS images for this region fails to reveal any more extended features. The optical counterpart of this object could be even more difficult to identify, as with such a narrow velocity width and low HI mass, even the faintest optical smudge could potentially be associated with the HI.

The third candidate is source 67. This has a single possible counterpart, but like source 128 it would be exceptionally compact at 1.7 Kpc diameter - the fourth smallest in the sample, assuming that the candidates for sources 128 and 129 really do correspond to the HI (it would also be the seventh faintest). The only other source more compact is 77 at 1.5 Kpc. This has a very narrow velocity width of 57 km/s, as would be expected for such a small optical source. In contrast, source 67 has a velocity width of at least 167 km/s (depending on the exact parameters used in *mbspect* the measured W20 can be as high as 208 km/s). The optical candidate does not at all resemble any kind of a disc. As with the other dark galaxy candidates, neither the deeper INT images nor smoothing the SDSS images reveals any more extended alternative optical counterparts.

This is perhaps the most intriguing of the dark galaxy candidates, and even if not a genuine dark galaxy, is certainly an unusual object. Its HI mass is low at 5.4×10^7 $M_\odot$, but not extraordinarily so - it is only the 15th lowest in the sample and has been confirmed by follow-up observations. While not exhibiting the classic double-horn profile, its velocity width is still sufficient to suggest a rotating disc. The quality of the spectrum gives confidence that this is an accurate estimate of the velocity width. Every other object of velocity width greater than 167 km/s is clearly a rotating disc - for comparison, the next widest source is source 51, with a velocity W20 of 168 km/s but an optical diameter of 10.1 Kpc.

Even assuming the potential optical counterpart is associated with the HI, besides its unusual compactness, it is also incredibly blue, with a *g-i* colour of 0.09 and a *FUV-NUV* colour of -0.17. This is the bluest object in the entire sample, as well as one of the faintest with M_g = -13.8. At 1.18, its M_{HI}/L_g ratio is high, but this in keeping with the trend in M_{HI}/L_g ratio with M_g seen for objects with sure optical counterparts. Its deficiency, as may be expected, would be negligible at 0.09.

A tidal origin seems more likely for this object than for the other candidates, as the nearest HI detection is source 51, 12' away in projected distance across the sky, and only 264 km/s difference in velocity. In fact, within half a degree of the HI detection there are 7 other detections, all within 800 km/s velocity difference. There are, however, no detected HI streams, so it is difficult to ascribe the origin of this object to any particular interaction.

Two other objects are also noteworthy though are not dark galaxy candidates. The first is VCC 1249, source 82 in this catalogue. This object is one of the extreme outliers in terms of difference in optical and HI sky coordinates, at 3.3'. This blue, irregular, well-studied (60 references in NED) object is clearly interacting with M49, showing an optical tail extending north towards that galaxy (which is actually significantly closer to the HI coordinates at 2.3'). In terms of velocities the situation is also complicated. AGES gives an HI velocity of 499 km/s. Velocities for VCC 1249 and M49 are given in NED as 276 and 997 km/s respectively - the HI thus lying between the two, but closer to VCC 1249 at 223 km/s velocity difference. GOLDMine, however, assigns VCC 1249 a velocity of 473 km/s.

AGES is just able to resolve an HI tail from this object. This is shown by the renzogram in figure 5.22, which shows that the HI detected at higher velocities is marginally closer to M49, with the lower velocity material closer to VCC 1249. The higher velocity material is as much as 5' away from that at the lower velocity. AGES does not have the resolution to say if the HI is at all associated with VCC 1249 or is entirely separate, though previous observations by Henning and Sancisi, 1993, find it to be separate. Given the almost ubiquitous presence of HI in late-type objects, and the general absence in early-types, this is a clear case of the gas being removed by tidal interactions, and explains why it is found so far from the optical counterpart.

The second unusual object is source 151, VCC 52. The HI spectrum and SDSS image are shown in figure 5.23. The object is unusual because of its extreme velocity width, a W20 of 565 km/s. This in itself is unusual but not remarkable - 3 other objects in this sample have W20 in excess of 500 km/s, but these are all obviously rotating discs. VCC 52 shows no evidence of a disc in optical images (again, including the deeper INT images and smoothing the SDSS images), and its HI velocity width in GOLDMine of 48 km/s seems more typical of its irregular classification. Its velocity, according to AGES, is 2,245 km/s, whereas GOLDMine gives 2,088 km/s (157 km/s difference), the SDSS 2,070 km/s (170 km/s difference) and NED 2,060 km/s (185 km/s difference, using earlier SDSS data). The positional deviation of the HI from the optical is also high at 1.8'.

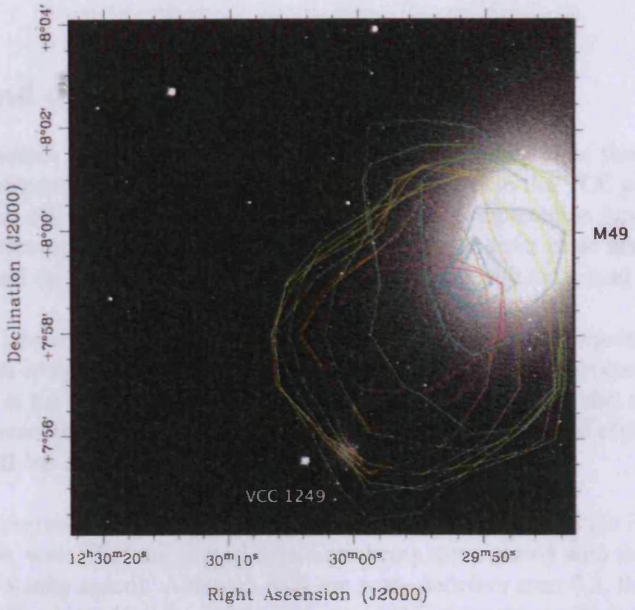
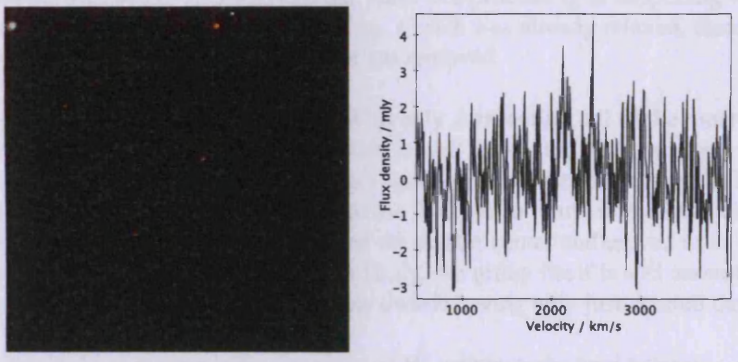


Figure 5.22: Renzogram for AGES detection of HI associated with VCC 1249 and M49. Contours are all at 5 mJy/beam with colour indicating velocity - red the lowest at 468 km/s, through orange, yellow, green and finally cyan at 533 km/s.



(a) SDSS RGB image, 6' across. VCC 52 is the blue object west of center.

(b) HI spectrum

Figure 5.23: HI profile and SDSS RGB image for unusual object source 151, VCC 52.

The key difference is the measured W20, nearly a factor 10 larger in AGES than in GOLDMine. The difference in recessional velocity, and perhaps also the positional deviation due to the profile fitting, stems from this. AGES reveals an apparent, though weak, double-horn profile. The lower velocity peak corresponds far more closely with the velocities quoted from other data, at 2,105 km/s, but itself has a velocity width of as much as 262 km/s - still far higher than the 48 km/s in GOLDMine. The second peak, which has a slightly higher peak S/N, is at 2,472 km/s, and though narrower than the first peak is still as wide as 156 km/s.

One simple explanation is that this second peak is spurious. The detection as a whole is quite weak, with a peak S/N of 5.03 - yet this actually corresponds to the second peak. However, if this peak is real, and is of higher S/N than the lower velocity peak, it should have been detected in previous observations - indeed it should have been more readily detected than the lower velocity peak. This still would not explain the greater width of the first peak than previous observations would indicate. Only HI follow-up observations will confirm the nature of this detection and allow further speculation.

5.4 Summary and discussion

AGES detects 60 objects within the Virgo Cluster in a 10 square degree area. 21 of these are new HI detections, while 15 of these are not members of the VCC. Of these, 8 are brighter than the VCC completeness limit. Given that this represents less than 5% of the 172 VCC detections the VCC completeness limit seems reasonable. The fact that AGES does not uncover many more new detections is likely due to its mass sensitivity, which is comparable to that of the VCC, and the fact that HI is almost always associated with an optical galaxy.

The majority of HI detections are of late-type objects, typically blue spirals or irregulars. Only 1 unambiguous late-type object (far enough away from other detections to rule out any possibility of confusion) is not detected in HI (VCC 1435), although in the exceptional instance of VCC 1249, it seems clear that the gas has been removed from that object by its interaction with M49. In most cases, though the majority of objects are HI deficient, it is not at all clear where the HI has actually gone.

The basic structure of this region is of 3 clouds at 17, 23 and 32 Mpc distance. Both the HI and optical studies (i.e. the VCC) support this view, with all of the new HI detections being interspersed with the optical detections (both spatially and in position-velocity space). Although 83% are more deficient than 0.3, there are still 17% that are effectively non-deficient. This includes many of the new, non-VCC detections, which are generally small, blue, and gas-rich (M_g fainter than -16, g-i bluer than 0.7, and with M_{HI}/L_g ratios greater than 0.3).

Simulations by Vollmer et. al. 2001 and Smith 2009 have shown that dwarf galaxies will lose all their gas through ram-pressure stripping on their first pass through the cluster center. Paradoxically, AGES detects both giant galaxies which are strongly deficient in HI and gas rich dwarfs (and even HI objects with no obvious optical counterparts at all) in the same sub-clusters. This is even true for the 32 Mpc group, which is too distant to be interacting with the X-ray gas associated with M49. This suggests that some pre-processing is happening within each sub-cluster (i.e. gas removal), and that each is itself still assembling. If each was already relaxed, there would be no gas-rich dwarf galaxies present as they would have had all their gas removed.

There is at least a very broad trend in HI deficiency. With only 1 exception, all of the most deficient (greater than 0.8) galaxies are found close to M49. That this exception is well within the 32 Mpc group (there is no possibility it is actually a member of one of the other populations) again supports the idea that at least some gas removal is occurring in situ within the groups, and not only due to the X-ray halo found in the M49 cloud. Were this not the case, the 32 Mpc cloud would have had to have entered the cluster center and moved to its present location while somehow preserving the gas within the dwarfs. More likely, the group itself is still assembling, with some giant galaxies having already lost much of their gas, but some dwarfs having only just entered the group.

Such an interpretation could also explain the presence of HI within early-type galaxies. ALFALFA have found that the fraction of early-types detected in HI is much higher in the field, as much as 44% for objects brighter than $M_B < -17$. The detected ETGs are much more likely to contain HI if they have only very recently entered the cluster. This view is supported by the fact that the AGES detections are (arguably) not, like those found in the ALFALFA survey, borderline cases of the early-type classification. Rather, they are all red both optically and in UV, and with no prominent signs of structure except in the case of VCC 758, which has a clear dust lane or disc. They also contain as much gas as late-type objects of similar magnitude, suggesting they have not lost gas and have not been greatly affected by the cluster environment.

Morphology varies with R.A. but not distance from M49, whereas the opposite appears true for deficiency (at least, the variation is clearer in terms of distance from M49 rather than R.A.). This raises the question of what role environment plays in influencing galaxy morphology - the old "nature or nurture" debate. That the trends in deficiency and morphology are related to different spatial parameters suggests that while galaxy density may perhaps be important, gas loss may not have a great deal of impact on morphology. The ETGs can also be interpreted to support this. They have as much gas as late-type objects, but are clearly early-type. If they have lost gas and were of originally a different morphology, they show few remnant features of their earlier incarnations.

This interpretation is challenged when considering that the detected early-types are of two quite distinct morphologies, lenticulars and dwarf ellipticals. The characteristics of these morphologies are very different to one another. The lenticulars appear to be bright objects of relatively low gas content, consistent with being gas deficient, but a

significant fraction (20%) of the S0s are detected in HI. Although they are redder than most HI detections, they are not as red as the non-detections, and are probably transition objects. Two of the detected S0s (50%) do show signs of structure, VCC 758 with its dust feature and VCC 180 with its small central bar. These objects then do show some remnant features of an earlier morphology : their gas, structure, and colour.

In contrast, the detected dEs are faint, red sequence objects that are as gas rich as dwarf irregulars, and are probably not gas deficient, but only 3% of the dEs are detected in HI. None of them show signs of structure, with the caveat that they are optically faint. The low detection rate combined with their high M_{HI}/L_B ratios, which implies non-deficiency, suggests that these are peculiar objects. If they have evolved from more gas-rich objects, the expectation would be to detect perhaps a greater fraction of the population but certainly less rich in gas, like the S0s. They are also truly red-sequence rather than transition objects. The dEs then are quite different to the S0s, and may have a different origin.

Further differences are also apparent. Morphology is correlated with magnitude, colour, and also M_{HI}/L_g ratio and velocity width (though these last two are weak trends), as shown in figure 5.24. Firstly, the early-type spirals tend to be brighter than the very late-type objects. The detected S0s fit roughly on this trend, but are perhaps slightly fainter than the true spirals. The dEs are quite distinct, being much fainter than the overall trend would suggest. Secondly, the early-type spirals tend to be redder than the later-type objects. All the detected ETGs lie approximately on this trend, perhaps slightly bluer than expected. Earlier-type objects seem to have the lowest M_{HI}/L_g ratio, with the S0s fitting this relation, but the dEs more gas-rich than expected, comparable with very late-type objects such as dIs. Finally, earlier-type objects have the highest velocity widths, except for the detected dEs which are again more similar to very late-type objects.

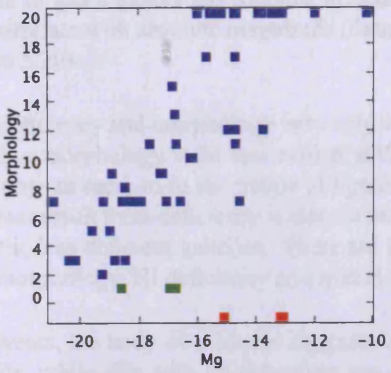
There is then good evidence that the S0s have morphologically evolved from later-type objects, with gas loss playing at least some role in the process. Being disc-like objects, the natural suggestion is that their progenitors would be spiral disc galaxies. They are broadly similar to these objects, but with less gas, redder than spirals of similar magnitude (but transition objects rather than true red-sequence galaxies), and perhaps fainter than the morphological trend would suggest. Though they have gas fractions consistent with being gas deficient, a significant proportion are nonetheless detected in HI. This suggests the population as a whole is loosing gas, with a few still retaining the last vestige of their original gas content.

The detected dEs appear to markedly differ from most trends shown in figure 5.24. In most regards - magnitude, velocity width and gas fraction - they are very similar to much later-type objects such as dIs. Yet their colours are very different, and they lie firmly on the red sequence (with the exception of VCC 190, which could be described as a transition object). This indicates that they cannot simply be morphological misclassifications, but it seems unlikely that they have experienced or are experiencing morphological evolution. C03 argue that as detected dEs are far more gas-rich than Local Group dEs, they cannot simply be analogues to those galaxies, and might instead be morphologically evolving dIs. However, the very similarities between the detected dEs and very late-type objects poses a problem for this interpretation. If this were the case, it does not explain why some dIs become much redder than others, despite being of similar luminosity, gas fraction, and gas velocity width.

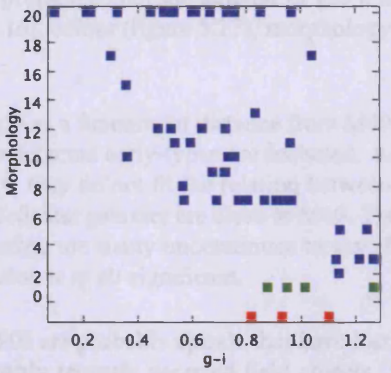
Another possibility is that the objects identified as dEs detected in HI are outright misidentifications. This is certainly plausible for VCC 368, as has been discussed. It may just be possible for VCC 1394, since the HI is 2.4' from the optical galaxy. However, there are simply no other candidates within the Arecibo beam, so in this case the solution would be in a sense worse than the problem. Finally, there is almost no possibility of this for VCC 190, which is close to HI coordinates (0.9' separation) and the only plausible candidate within the beam.

It is also possible that while the S0s are victims of morphological evolution caused by gas loss, the dwarf ellipticals are recently accreted field galaxies. The numbers are too small and the variation too large to say definitively if the M_{HI}/L_B ratios are consistent with the ALFALFA detected faint field ETGs. For early-type objects fainter than M_g = -16.0, ALFALFA find a range of M_{HI}/L_B ratios from -0.66 to +0.13 in logarithmic units, for AGES they range from -0.21 to +0.29. This does not rule out the possibility of field dEs with M_{HI}/L_B ratios comparable to the objects detected in this study (which would make the argument of C03, that the detected dEs cannot be Local Group analogues, a moot point). Their high gas fractions, coupled with their low detection rate, indicates that these are exceptional objects and very different to the majority of Virgo dEs. The idea of recent field accretion is at least plausible, and perhaps more likely than morphological evolution from dIs.

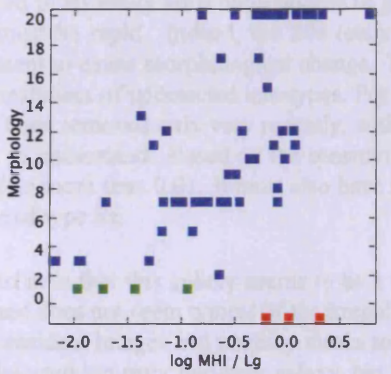
It is interesting to note that the objects without a proper morphological classification seem to fit broadly on the



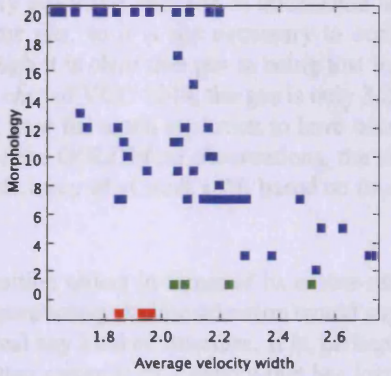
(a) Morphology as a function of absolute g magnitude



(b) Morphology as a function of optical g-i colour



(c) Morphology as a function of M_{HI}/L_g in logarithmic units



(d) Morphology as a function of the averaged velocity width (W50 and W20) in logarithmic units

Figure 5.24: Morphology as a function of optical parameters for HI detections with morphological classifications. Blue indicates late-types and unclassified objects, green S0s and red dEs.

trends shown in figure 5.24, despite simply being assigned as type 20. Perhaps these small, blue objects represent a truly different morphology to conventional dwarf irregulars and blue compact dwarfs, or perhaps a more accurate classification would reveal that they actually follow the trends more closely (see chapter 6 for more discussion). They are unlikely to be widely deviating from the relations. A visual inspection shows that only 1 object could be classed as early as type 10 (Sdm) with most being from 12-16 (Im to Im/BCD). Shifting their classification down to these values would have little enough impact on the apparent trends.

There is a major caveat to the idea that gas loss is driving morphological evolution in the spirals and lenticular galaxies. This is partly based on the trends in figure 5.24, but the morphological classification scheme is subjective and its ordering arbitrary. For example, the trends include the very late-type dIs. Accepting that S0s evolve from spirals, the logical implication from the trends is that spirals evolve from dIs. This makes no sense in a cluster environment where galaxies are losing gas, in particular for small galaxies which should lose all their gas on the first pass through the cluster center. While gas loss might cause reddening, there is no way it could cause a dI to brighten by 6 magnitudes into a giant spiral. The fact that there is a trend in morphology does not necessarily imply morphology evolution.

Another caveat is that as shown in figure 5.13 and 5.19, morphology varies with R.A. but deficiency varies with

distance from M49. The situation is complicated by the 3 different sub-clusters. However, as has been shown, it is not the amount of gas a galaxy has lost that affects its properties, but the amount of gas it still has. This has been shown to correlate with absolute magnitude (figure 5.16), colour (figure 5.17), morphology (figure 5.24) and deficiency (figure 5.20).

Moreover, both deficiency and morphology vary only weakly as a function of distance from M49 and R.A. respectively. The trend in morphology is far less evident if the undetected early-types are included. Additionally, while the unclassified objects seem to fit the trends of figure 5.24, they do not fit the relation between morphology and R.A. The only clear result from deficiency is that the most deficient galaxies are close to M49. There is no evidence of any trend in the less deficient galaxies. There are therefore too many uncertainties to say if the difference in trends between morphology, HI deficiency and spatial location is at all significant.

Despite these caveats, the body of evidence suggests that S0s are probably spirals that have lost most of their gas and begun to fade, while dEs with HI detections are probably recently accreted field objects. The S0s are very similar to the spirals except for their relatively smooth appearance and their lower gas content. The dEs are similar to late-type objects in all regards except colour. It is difficult to see how any gas loss could have caused reddening in these cases, given that they still retain as much gas and in such a similar state (as evidenced from the similar velocity widths) to the late-types.

If the S0s detected in HI really are a consequence of galaxy evolution, the lack of undetected late-types suggests such evolution must be rapid. Indeed, the S0s retain some gas, so it is not necessary to completely deplete a galaxy's gas content to cause morphological change. Though it is clear that gas is being lost in the cluster, there are only 2 clear instances of undetected late-types. For the case of VCC 1249, the gas is only 3.2' from the galaxy so has probably been removed only very recently, without time for much evolution to have occurred. VCC 1435 is more difficult to understand. Based on the sensitivity of the GOLDMine observations, the object can have an M_{HI}/L_g ratio of no more than 0.01. It must also have a deficiency of at least 1.77, based on its optical radius and assuming it to be of type Sa.

It is interesting to note that this galaxy seems to be a transition object in terms of its colour-magnitude, has lost most of its gas, and does not seem typical of the irregular morphology its classification would suggest. It was even necessary to use residual images and unsharp masks to reveal any kind of structure. It is, perhaps, not an example of a rare HI-undetected but truly late-type galaxy, but another example of a galaxy that has lost most or all of its gas and is consequently beginning to undergo a morphological transformation.

This region of the cluster is rich in HI-detected galaxies of all morphological types, many of which appear to be entering the cluster for the first time. This makes it an ideal location to search for dark galaxies. No more than 3 candidates are found, 5% of the HI sample and less than 2% of the total number of galaxies present (from the number of VCC galaxies and non-VCC objects detected in HI). This is certainly well below the 23% predicted by Davies et. al. 2006. Even these candidates may not be dark galaxies at all.

Object 129, while having no apparent optical counterparts at all, has too low a velocity width to be described as a dark galaxy. Without knowing the spatial extent of the HI, its dynamical mass cannot be estimated, and the presence of dark matter cannot readily be inferred. This would be true in principle for any dark galaxy candidate, but to claim the detection of such an extraordinary object requires at least a hint of a high dark matter content, i.e. a reasonably high velocity width. Object 128 also lacks an optical counterpart and has a much higher velocity width, potentially invoking the need for a rotating disc, but could possibly be confused with continuum emission. Finally object 67, which also has a high velocity width, does seem to have a possible optical counterpart, even if a very unusual object.

Even if none of these objects are really dark galaxies, they are still interesting objects. All would be amongst the faintest objects in the sample, assuming they do have optical counterparts at all. None are easy to explain as debris from tidal interactions, particularly object 129 which is 23' from the nearest HI detection. The only plausible counterpart for object 67 would not only be the bluest object in the sample, but also one of the smallest, despite its ~200 km/s velocity width, normally indicative of a rotating disc.

If they exist, dark galaxies must be rare and any candidates will invariably be controversial. A survey as sensitive as AGES in a region as dense as the Virgo Cluster has detected, at best, 3 candidates, and at worst (quite plausibly) none at all. Those candidates all have candidate optical counterparts, even if very faint, and to properly rule them

out requires optical spectroscopy. Determining that they are truly primordial objects and not debris from some interaction is even harder. This would require aperture synthesis observations to correctly measure the spatial extent of the HI disc, to estimate its dark matter content. Even this relies on the assumption of gravitational stability, and proving this would be harder still.

The Virgo Cluster is a complex and dynamic environment. It is clear that galaxies within the cluster are suffering gas loss and probably quenching of star formation as a result. The weight of evidence suggests that at least the S0s detected in HI are the result of morphological evolution, in this case from spirals. They are broadly similar to these objects, but fainter and with less gas. In contrast the detected dEs seem to be recently accreted field objects, being as gas-rich as dIs but considerably redder. Such gas-rich red galaxies can exist in the cluster environment, since external forces need not immediately act to cause star formation. This is evidenced by the 3 intriguing dark galaxy candidates detected. Proving that these are dark galaxies, or some other exotic form of HI, requires many further observations.

Table 5.1: HI properties of objects detected within the Virgo Cluster. Bracketed values indicate errors as computed by miriad. Columns : (1) Source number in this catalogue, an asterisk indicates the detection is uncertain (2) Right ascension J2000, error in seconds of time (3) Declination J2000, error in seconds of arc (4) Heliocentric velocity km/s (5) Maximum velocity width at 50% and (6) 20% of the peak flux (7) Total flux Jy (8) Distance Mpc (9) Estimated HI mass, log $M_{\odot}$ (10) Peak signal to noise (11) R.m.s. mJy

(1) No.	(2) R.A.	(3) Dec.	(4) Velocity	(5) W50	(6) W20	(7) Flux	(8) Dist	(9) M_{HI}	(10) S/N	(11) rms
45	12:20:28(3.1)	+06:53:50(24)	2588(5)	89(11)	134(16)	1.461(0.247)	32.0	8.55	9.28	2.0
46	12:13:03(0.7)	+07:02:02(10)	2103(3)	403(6)	428(9)	2.533(0.176)	32.0	8.79	13.15	0.7
47	12:21:01(1.0)	+07:04:26(11)	2373(9)	140(18)	197(27)	0.381(0.089)	32.0	7.96	5.86	0.7
48	12:18:18(0.8)	+07:07:17(11)	2726(4)	71(8)	108(12)	0.424(0.069)	32.0	8.01	11.32	0.6
49	12:13:50(0.7)	+07:11:52(10)	2682(2)	322(4)	349(6)	6.450(0.287)	32.0	9.19	44.1	0.6
50	12:13:00(0.7)	+07:17:50(10)	2189(4)	128(8)	167(11)	0.805(0.092)	32.0	8.29	13.27	0.6
51	12:17:27(0.7)	+07:19:15(10)	2227(2)	143(5)	168(7)	1.563(0.123)	32.0	8.58	21.57	0.6
52	12:16:34(0.7)	+07:27:27(10)	2624(3)	519(5)	544(8)	3.232(0.179)	32.0	8.89	15.62	0.6
53	12:20:07(0.7)	+07:41:17(10)	2631(2)	123(4)	151(5)	4.411(0.274)	32.0	9.03	80.23	0.5
54	12:19:06(0.7)	+07:37:53(11)	2685(2)	62(4)	76(6)	0.486(0.069)	32.0	8.07	16.46	0.6
55	12:29:38(0.7)	+07:49:11(10)	2363(2)	137(4)	172(6)	6.591(0.386)	17.0	8.65	64.28	0.8
56	12:16:25(0.7)	+07:47:12(11)	2373(3)	50(7)	91(10)	0.388(0.062)	32.0	7.97	15.72	0.6
57	12:19:22(0.7)	+07:52:16(10)	2491(3)	69(5)	107(8)	1.154(0.108)	32.0	8.45	28.03	0.6
58	12:10:46(0.8)	+07:51:30(13)	2060(7)	88(13)	134(20)	0.234(0.059)	32.0	7.75	7.21	0.6
59	12:16:50(0.8)	+08:04:06(17)	2588(7)	78(15)	145(22)	0.556(0.114)	32.0	8.13	7.95	1.0
60	12:21:36(1.3)	+07:15:23(15)	2754(9)	91(18)	156(27)	0.100(0.041)	32.0	7.38	6.43	0.6
61	12:23:38(0.7)	+06:56:35(10)	1034(3)	327(6)	374(9)	5.400(0.311)	23.0	8.83	24.04	1.1
62	12:24:05(0.7)	+07:02:14(10)	1253(2)	62(4)	83(5)	1.385(0.135)	23.0	8.24	35.88	0.7
63	12:25:42(0.7)	+07:12:51(10)	1015(2)	281(3)	306(5)	23.104(0.953)	23.0	9.46	161.44	0.6
64	12:27:10(0.7)	+07:15:43(10)	953(2)	145(3)	165(5)	20.948(1.210)	23.0	9.42	289.67	0.6
65	12:23:47(0.7)	+07:11:11(10)	1444(2)	164(4)	187(7)	2.102(0.142)	23.0	8.42	22.19	0.6
66	12:16:25(0.7)	+07:12:19(11)	1839(3)	73(6)	86(9)	0.346(0.065)	32.0	7.92	9.55	0.6
67	12:18:16(1.0)	+07:21:23(14)	1963(8)	121(16)	167(24)	0.224(0.062)	32.0	7.73	6.01	0.6
68	12:23:57(0.7)	+07:27:00(11)	1257(2)	48(5)	76(7)	0.581(0.075)	23.0	7.86	21.25	0.6
69	12:25:53(0.7)	+07:33:11(10)	1212(2)	172(5)	211(7)	3.972(0.230)	23.0	8.7	47.16	0.6
70	12:20:08(0.8)	+07:33:55(19)	1682(4)	29(9)	70(13)	0.183(0.041)	32.0	7.65	11.68	0.5
71	12:27:12(0.7)	+07:38:28(10)	1197(2)	50(3)	68(5)	1.019(0.106)	17.0	7.84	34.49	0.6
72	12:27:16(0.7)	+07:40:25(11)	1863(3)	53(6)	88(10)	0.361(0.055)	17.0	7.39	15.32	0.5
73	12:27:28(0.7)	+07:38:29(10)	887(2)	130(4)	158(6)	2.720(0.177)	17.0	8.27	37.33	0.6
74	12:14:09(0.7)	+07:46:19(10)	1239(2)	114(4)	140(5)	6.110(0.374)	32.0	9.17	91.22	0.6
75	12:23:59(0.7)	+07:47:13(10)	1135(3)	126(6)	166(10)	1.108(0.104)	23.0	8.14	17.72	0.6
76	12:26:46(0.7)	+07:55:13(10)	1408(2)	142(4)	175(6)	3.826(0.222)	17.0	8.42	44.69	0.6
77	12:26:04(0.8)	+07:54:06(12)	1249(4)	32(8)	57(12)	0.181(0.049)	23.0	7.35	8.76	0.6
79	12:24:51(0.9)	+07:26:42(12)	802(9)	296(19)	353(28)	0.555(0.100)	23.0	7.84	5.72	0.6
80	12:24:47(0.7)	+07:54:27(10)	816(2)	37(5)	47(7)	0.188(0.047)	23.0	7.37	10.58	0.6
81	12:29:31(0.7)	+07:41:44(11)	771(3)	162(6)	206(10)	2.380(0.185)	17.0	8.21	19.56	0.9
82	12:29:54(0.7)	+07:58:31(11)	499(2)	36(5)	73(7)	0.839(0.107)	17.0	7.76	27.99	0.8
111	12:42:45(0.7)	+07:20:09(10)	2435(2)	64(3)	82(5)	1.775(0.160)	17.0	8.08	41.62	0.7
112	12:39:24(0.7)	+07:57:31(10)	2097(2)	105(5)	148(7)	7.001(0.451)	17.0	8.68	86.02	0.8
113	12:34:24(1.3)	+08:06:13(11)	2078(3)	77(7)	164(10)	8.622(0.664)	17.0	8.77	44.16	2.4
114	12:37:02(0.7)	+06:54:60(11)	1659(2)	89(4)	129(7)	6.954(0.500)	17.0	8.68	62.68	1.3
115	12:43:07(0.7)	+07:39:02(10)	1339(2)	57(4)	82(5)	2.161(0.192)	17.0	8.17	53.78	0.7
116	12:32:46(0.8)	+07:48:10(14)	1362(13)	58(27)	179(40)	0.106(0.051)	17.0	6.86	5.74	0.8
117	12:31:50(2.0)	+07:50:06(35)	1309(9)	52(17)	117(26)	0.349(0.092)	17.0	7.38	6.64	0.9
118	12:38:21(0.7)	+07:53:10(10)	1809(2)	155(4)	184(6)	2.957(0.191)	17.0	8.3	31.45	0.7
119	12:36:33(1.2)	+08:02:38(14)	1821(4)	31(9)	72(13)	0.412(0.092)	17.0	7.45	11.1	1.1
120	12:31:04(0.9)	+08:05:37(14)	1770(22)	170(45)	301(67)	0.756(0.298)	17.0	7.71	3.5	2.5
121	12:38:14(0.8)	+06:59:12(11)	-99(6)	67(12)	98(17)	0.342(0.095)	17.0	7.36	6.76	1.0
122	12:37:46(0.7)	+07:06:22(10)	85(2)	61(4)	89(5)	5.460(0.468)	17.0	8.57	122.22	0.8
123	12:34:40(0.7)	+07:09:28(10)	624(4)	89(7)	137(11)	1.213(0.134)	17.0	7.92	16.01	0.9
128*	12:38:04(0.8)	+07:30:40(17)	1815(13)	40(26)	180(39)	0.061(0.034)	17.0	6.62	6.34	0.7
129	12:36:52(1.1)	+07:25:28(19)	1634(9)	69(18)	110(27)	0.146(0.059)	17.0	7.0	5.04	0.7
132	12:22:12(1.4)	+07:08:20(19)	1071(6)	182(11)	199(17)	0.296(0.077)	23.0	7.57	5.11	0.6
133	12:22:43(1.0)	+07:55:18(11)	1090(12)	75(24)	161(36)	0.198(0.066)	23.0	7.39	5.4	0.7
145	12:16:13(0.9)	+07:56:34(45)	2252(14)	63(28)	154(42)	0.036(0.030)	32.0	6.94	4.66	0.7
146	12:28:28(0.7)	+06:56:30(10)	889(17)	49(34)	235(51)	0.233(0.089)	17.0	7.2	5.56	1.1
147	12:17:11(0.9)	+07:11:46(20)	2282(5)	513(9)	519(14)	0.537(0.130)	32.0	8.11	3.68	0.7
148	12:19:53(3.6)	+07:38:11(21)	2597(9)	46(18)	135(28)	0.207(0.055)	32.0	7.7	7.33	0.6
149	12:20:26(2.8)	+07:41:02(28)	2637(7)	112(14)	143(22)	0.155(0.059)	32.0	7.57	5.4	0.7
151	12:12:30(3.0)	+06:59:08(13)	2245(14)	467(27)	565(41)	0.414(0.112)	32.0	8.0	5.03	0.9

Table 5.2: HI properties of galaxies behind the Virgo Cluster. Columns are as for table 5.1.

(1) No.	(2) R.A.	(3) Dec.	(4) Velocity	(5) W50	(6) W20	(7) Flux	(8) Dist	(9) M_{HI}	(10) S/N	(11) rms
4	12:24:36(0.7)	+07:32:29(10)	14640(7)	216(13)	301(20)	0.921(0.103)	206.0	9.97	10.7	0.6
7	12:22:19(0.7)	+07:37:32(10)	11301(3)	163(5)	186(8)	1.020(0.092)	159.0	9.79	15.3	0.5
8	12:25:28(1.3)	+07:38:50(11)	11387(7)	133(14)	204(21)	0.449(0.069)	160.0	9.44	9.02	0.5
9	12:22:23(0.7)	+07:53:07(10)	11380(5)	76(11)	135(16)	0.505(0.076)	160.0	9.49	11.1	0.6
10	12:13:19(3.0)	+07:51:35(11)	11333(8)	111(16)	164(24)	0.409(0.096)	159.0	9.39	6.28	0.8
11	12:12:35(0.7)	+08:02:18(10)	11168(11)	107(22)	221(33)	0.681(0.134)	157.0	9.6	6.81	1.0
12	12:19:01(0.7)	+07:48:00(11)	10861(7)	208(15)	319(22)	0.884(0.101)	152.0	9.69	10.83	0.6
13	12:18:42(0.9)	+07:55:14(19)	10801(13)	80(26)	210(39)	0.279(0.069)	152.0	9.18	6.2	0.6
15	12:19:50(0.7)	+06:59:07(10)	7221(2)	54(3)	80(5)	4.887(0.437)	101.0	10.08	104.59	0.9
16	12:25:42(0.7)	+07:09:47(10)	7176(3)	277(6)	302(9)	1.346(0.106)	101.0	9.51	14.72	0.5
17	12:23:24(0.9)	+07:09:08(11)	7207(13)	178(25)	276(38)	0.325(0.078)	101.0	8.9	5.51	0.6
18	12:31:34(0.7)	+07:18:10(10)	7339(6)	194(11)	268(17)	1.256(0.126)	103.0	9.5	11.96	0.7
19	12:20:28(0.7)	+07:30:08(10)	7337(2)	59(4)	82(6)	1.178(0.114)	103.0	9.47	34.59	0.6
20	12:28:35(0.7)	+07:29:26(11)	7553(8)	184(15)	287(23)	0.810(0.090)	106.0	9.34	10.14	0.5
21	12:28:14(0.7)	+07:36:26(11)	7447(3)	119(6)	145(8)	0.947(0.097)	104.0	9.39	15.78	0.6
22	12:19:28(0.7)	+07:47:29(10)	7292(3)	158(6)	200(10)	1.545(0.122)	102.0	9.59	18.04	0.6
23	12:28:59(2.1)	+07:51:02(17)	7471(9)	316(19)	387(28)	0.706(0.118)	105.0	9.27	6.24	0.7
24	12:24:13(0.7)	+07:57:05(11)	7231(2)	160(4)	188(6)	3.034(0.195)	101.0	9.87	32.29	0.7
25	12:24:58(0.0)	+07:53:34(00)	7270(0)	96(0)	96(0)	0.967(0.000)	102.0	9.38	14.26	0.8
26	12:28:20(0.0)	+08:10:16(00)	7450(0)	245(0)	381(0)	0.772(0.000)	104.0	9.29	6.58	0.9
27	12:29:25(1.0)	+08:04:13(13)	7495(7)	205(14)	238(21)	0.629(0.158)	105.0	9.22	5.65	1.3
28	12:11:13(1.1)	+07:29:23(18)	6239(12)	79(24)	187(36)	0.283(0.070)	87.0	8.71	6.15	0.6
29	12:28:29(0.8)	+07:41:53(14)	6411(9)	63(17)	119(26)	0.120(0.046)	90.0	8.36	6.03	0.6
30*	12:29:04(0.9)	+07:48:48(12)	6233(6)	90(11)	102(17)	0.141(0.058)	87.0	8.41	4.13	0.6
31	12:24:50(0.7)	+07:53:29(13)	6436(8)	78(16)	114(24)	0.033(0.025)	90.0	7.81	5.28	0.6
32	12:15:26(0.7)	+07:46:23(10)	5961(3)	193(6)	221(9)	1.245(0.112)	83.0	9.32	14.29	0.6
33	12:12:25(0.7)	+07:08:07(12)	5670(4)	149(8)	181(12)	0.830(0.097)	79.0	9.1	11.16	0.6
35	12:24:25(0.7)	+07:08:04(10)	4238(3)	82(5)	104(8)	0.669(0.081)	59.0	8.75	15.9	0.6
36	12:23:23(0.7)	+07:28:15(10)	4265(2)	270(4)	299(6)	3.862(0.200)	60.0	9.52	29.6	0.6
37	12:26:03(0.7)	+07:25:58(10)	4567(3)	55(7)	81(10)	0.318(0.060)	64.0	8.49	11.28	0.6
38	12:24:05(2.2)	+07:32:22(10)	4210(5)	23(9)	47(14)	0.096(0.037)	59.0	7.9	7.76	0.6
39	12:18:11(0.7)	+07:39:31(10)	3974(2)	137(4)	163(5)	5.566(0.311)	55.0	9.61	81.88	0.5
40	12:18:51(0.7)	+07:32:32(10)	3926(7)	67(14)	130(20)	0.263(0.059)	55.0	8.28	8.53	0.6
41	12:17:23(0.7)	+07:02:14(10)	3753(2)	145(5)	166(7)	1.355(0.124)	52.0	8.95	16.83	0.7
42	12:17:58(0.7)	+07:11:10(10)	3744(2)	341(4)	368(5)	18.47(0.717)	52.0	10.08	112.26	0.6
43	12:17:58(0.7)	+07:16:21(10)	3933(3)	266(6)	321(9)	5.208(0.256)	55.0	9.58	32.81	0.7
44	12:19:52(0.7)	+07:43:39(10)	3799(2)	123(4)	153(7)	1.758(0.144)	53.0	9.07	28.03	0.7
83*	12:38:10(0.8)	+07:01:30(14)	15183(16)	235(32)	351(48)	0.403(0.108)	213.0	9.64	4.62	0.8
84	12:43:36(0.8)	+07:26:14(11)	14031(7)	279(13)	317(20)	0.846(0.126)	197.0	9.89	6.61	0.7
85	12:48:05(0.7)	+06:58:53(10)	13458(3)	122(5)	148(8)	1.836(0.182)	189.0	10.19	16.32	1.1
86	12:37:46(0.7)	+07:43:02(11)	13496(7)	39(14)	145(21)	0.342(0.071)	190.0	9.46	11.58	0.8
87	12:43:14(0.9)	+07:59:38(13)	13265(8)	313(15)	350(23)	0.872(0.165)	186.0	9.86	5.5	1.0
88	12:45:32(0.7)	+07:41:33(12)	13218(5)	206(11)	216(16)	0.338(0.105)	186.0	9.44	4.1	0.8
89	12:44:47(0.8)	+07:21:32(11)	11517(7)	261(13)	305(20)	0.739(0.115)	162.0	9.66	6.99	0.7
90	12:38:17(0.7)	+07:48:38(10)	11365(10)	269(20)	414(30)	1.382(0.155)	160.0	9.92	8.76	0.8
91	12:31:37(0.8)	+07:51:01(11)	11193(13)	209(27)	309(40)	0.513(0.123)	157.0	9.48	5.24	0.9
92	12:33:27(0.7)	+07:51:02(10)	11461(10)	44(20)	125(30)	0.156(0.055)	161.0	8.98	6.47	0.7
94	12:36:36(0.7)	+07:58:31(10)	11391(4)	341(9)	370(13)	1.670(0.171)	160.0	10.01	8.76	0.8
95	12:33:51(0.8)	+08:00:46(13)	11386(3)	122(6)	135(9)	0.832(0.138)	160.0	9.7	8.23	1.0
96	12:34:07(0.7)	+08:00:49(11)	9221(11)	292(23)	374(34)	0.507(0.126)	129.0	9.3	5.52	1.0
97	12:36:06(0.7)	+07:42:28(11)	8802(3)	89(6)	116(8)	1.072(0.119)	123.0	9.59	16.07	0.8
98	12:38:53(0.7)	+07:06:35(11)	7205(8)	311(17)	423(25)	1.007(0.121)	101.0	9.39	9.36	0.7
99	12:39:22(1.0)	+07:07:48(13)	7242(8)	125(16)	174(24)	0.468(0.103)	102.0	9.06	6.24	0.8
100	12:39:39(0.7)	+07:09:55(11)	7287(4)	264(7)	318(11)	2.789(0.177)	102.0	9.84	18.58	0.7
101	12:36:09(0.7)	+07:31:56(10)	7233(6)	295(11)	369(17)	1.806(0.162)	101.0	9.65	11.9	0.8
102	12:48:21(0.7)	+07:35:35(11)	7129(5)	168(11)	207(16)	0.650(0.100)	100.0	9.19	8.61	0.7
103	12:32:40(1.0)	+07:55:50(11)	7260(5)	141(10)	166(15)	0.540(0.097)	102.0	9.12	7.21	0.7
104	12:48:19(2.1)	+08:03:04(54)	7408(8)	138(16)	182(24)	0.606(0.147)	104.0	9.19	5.88	1.2
105	12:30:18(0.8)	+06:56:52(14)	6467(12)	182(24)	244(36)	0.406(0.128)	91.0	8.9	4.56	1.1
106	12:37:38(0.8)	+07:09:22(12)	6544(8)	132(15)	184(23)	0.479(0.095)	92.0	8.98	6.65	0.7
107	12:31:42(3.9)	+07:20:14(16)	6453(10)	37(21)	149(31)	0.182(0.056)	90.0	8.55	7.29	0.7
108	12:40:49(0.7)	+07:26:53(11)	6526(3)	59(6)	76(8)	0.444(0.076)	91.0	8.95	11.67	0.7
109	12:37:00(0.7)	+07:28:07(11)	6484(8)	70(17)	148(25)	0.384(0.086)	91.0	8.88	7.52	0.8
110	12:35:13(0.7)	+07:08:11(10)	5663(4)	169(8)	192(11)	0.710(0.102)	79.0	9.03	9.3	0.7
126	12:32:36(0.8)	+07:53:07(12)	6544(9)	131(17)	188(26)	0.431(0.092)	92.0	8.94	6.25	0.7
135	12:49:15(0.9)	+07:30:27(13)	11577(9)	37(18)	132(27)	0.465(0.116)	163.0	9.47	7.65	1.2
136	12:33:18(0.8)	+07:06:52(13)	14732(3)	46(6)	57(9)	0.149(0.045)	207.0	9.18	8.43	0.6
137	12:48:32(0.8)	+07:18:11(14)	14793(13)	167(25)	299(38)	0.464(0.095)	208.0	9.68	6.37	0.7
138	12:15:40(0.7)	+08:06:44(18)	11142(11)	483(22)	564(33)	2.498(0.416)	156.0	10.16	5.75	2.3
139	12:48:23(0.8)	+07:06:00(13)	12803(11)	73(23)	172(34)	0.258(0.072)	180.0	9.3	6.11	0.7
140	12:11:25(1.1)	+07:56:00(13)	13725(11)	52(21)	129(32)	0.191(0.068)	193.0	9.23	5.76	0.8
142	12:17:32(1.1)	+07:04:07(13)	12124(4)	297(8)	304(12)	0.324(0.084)	170.0	9.35	4.66	0.6
150	12:22:59(1.0)	+07:27:24(15)	4139(15)	129(30)	204(45)	0.234(0.083)	58.3	8.27	3.98	0.7

Table 5.3: Optical properties of HI detections in the Virgo Cluster. Columns : (1) Source number in this catalogue, an asterisk indicates the association of the HI with the optical object is highly uncertain (2) Name in a major catalogue if present (3) Flag for status of the optical counterpart used here - 0 indicates the object has a matching optical redshift, 1 indicates no optical redshift is available, 2 indicates the match to the HI detection is very uncertain (4) Morphological type, using the system of the GOLDMine database (5) Difference between optical and HI centres in arcminutes (6) Sky distance from M49 in degrees (7) Absolute magnitude in the g and (8) i bands (9) g-i colour (10) HI mass to light ratio, g band (11) Optical diameter in Kpc (12) Estimated HI deficiency

(1) No.	(2) Name	(3) Flag	(4) Type	(5) Sep.	(6) Dist. M49	(7) M_g	(8) M_i	(9) g-i	(10) M_{HI}/L_g	(11) Diam.	(12) Def.
45	VCC415	0	9	1.0	2.55	-17.28	-18.0	0.71	0.318	14.52	0.63
46	VCC73	0	5	0.3	4.25	-19.6	-20.74	1.13	0.064	21.78	0.7
47	VCC450	0	1	0.74	2.36	-16.95	-17.96	1.0	0.112	17.31	
48	None	0	20	0.0	2.97	-15.97	-16.58	0.6	0.306	4.46	0.28
49	VCC94	0	2	0.2	4.03	-19.27	-20.41	1.13	0.224	23.64	0.19
50	FGC1384	0	10	0.14	4.21	-16.18	-16.91	0.72	0.482	11.54	0.71
51	None	1	20	0.0	3.12	-15.55	-16.22	0.66	1.666	10.05	0.32
52	VCC199	0	3	0.28	3.31	-20.19	-21.45	1.26	0.048	44.49	0.81
53	VCC393	0	7	0.23	2.41	-19.23	-19.99	0.76	0.159	21.4	0.44
54	None	0	20	0.33	2.66	-15.5	-15.99	0.48	0.545	7.07	0.56
55	VCC1205	0	7	0.22	0.18	-18.35	-19.04	0.68	0.15	11.86	0.37
56	VCC190	0	-1	0.97	3.31	-15.12	-15.93	0.8	0.615	8.01	
57	VCC343	0	9	0.03	2.58	-17.2	-17.91	0.71	0.27	10.23	0.47
58	None	0	20	0.65	4.7	-14.65	-15.17	0.52	0.575	7.07	0.88
59	VCC207	0	17	1.25	3.2	-15.7	-16.0	0.29	0.516	6.14	0.4
60	VCC488	0	20	0.84	2.15	-15.18	-15.93	0.75	0.15	12.65	1.69
61	VCC656	0	5	0.69	1.85	-19.01	-20.25	1.23	0.122	15.65	0.48
62	VCC697	0	7	0.3	1.7	-17.94	-18.84	0.9	0.085	12.57	0.83
63	VCC827	0	7	0.18	1.27	-18.59	-19.52	0.93	0.773	35.72	0.4
64	VCC975	0	8	0.18	0.98	-18.48	-19.04	0.56	0.78	22.21	0.08
65	VCC667	0	7	0.27	1.69	-17.93	-18.85	0.92	0.129	21.81	1.07
66	None	1	20	0.4	3.4	-14.74	-15.34	0.59	0.777	5.4	0.51
67*	None	2	20	0.72	2.92	-13.82	-13.91	0.09	1.184	2.42	0.09
68	None	0	20	0.28	1.54	-13.89	-14.11	0.21	1.485	3.88	0.32
69	VCC851	0	7	0.21	1.06	-18.02	-18.98	0.95	0.225	23.68	0.85
70	None	1	20	0.42	2.42	-13.91	-14.23	0.31	0.887	4.28	0.61
71	None	0	20	0.27	0.73	-13.47	-13.68	0.2	2.091	3.46	0.25
72	VCC989	0	7	0.35	0.7	-14.59	-15.16	0.56	0.263	4.35	0.87
73	VCC1011	0	10	0.22	0.67	-16.25	-17.0	0.74	0.431	11.57	0.74
74	VCC105	0	9	0.28	3.87	-18.92	-19.59	0.67	0.292	30.53	0.57
75	VCC688	0	7	0.24	1.44	-17.96	-18.88	0.91	0.066	11.64	0.87
76	VCC938	0	7	0.14	0.74	-18.2	-19.0	0.8	0.1	11.76	0.6
77	None	1	20	0.07	0.92	-13.47	-13.92	0.44	0.679	1.47	0.09
79	VCC758	0	1	0.89	1.34	-18.61	-19.87	1.26	0.018	15.78	
80	None	1	20	1.16	1.24	-13.33	-13.85	0.51	0.805	3.75	0.78
81	VCC1193	0	7	0.08	0.31	-17.09	-17.86	0.77	0.174	12.46	0.85
111	VCC1933	0	11	0.14	3.28	-14.71	-15.24	0.53	1.164	5.64	0.38
112	VCC1791	0	15	0.47	2.38	-16.87	-17.22	0.35	0.625	11.07	0.29
113	VCC1555	0	7	5.69	1.14	-20.88	-21.74	0.85	0.019	52.41	1.38
114	VCC1699	0	11	0.66	2.1	-16.72	-17.31	0.58	0.713	8.8	0.12
115	VCC1952	0	12	0.28	3.32	-14.95	-15.42	0.47	1.13	8.01	0.56
116	VCC1455	0	12	0.28	0.76	-14.71	-15.43	0.71	0.069	6.63	1.72
117	VCC1394	2	-1	2.41	0.53	-13.11	-14.03	0.92	0.997	5.74	
118	VCC1758	0	7	0.33	2.12	-16.58	-17.35	0.77	0.345	12.56	0.76
119	VCC1675	0	13	0.81	1.67	-16.76	-17.59	0.82	0.04	12.36	1.61
120	VCC1330	0	3	1.36	0.33	-18.82	-20.03	1.21	0.011	20.76	1.6
121	VCC1750	0	17	0.58	2.32	-14.7	-15.74	1.03	0.225	3.16	0.66
122	VCC1726	0	10	0.24	2.17	-16.19	-16.55	0.35	0.912	10.38	0.35
123	VCC1575	0	11	0.15	1.47	-17.65	-18.61	0.95	0.052	12.16	1.13
128*	None	1	20	1.13	2.1	-12.0	-12.72	0.72	0.484	0.99	0.53
129*	None	2	20	1.31	1.84	-10.21	-11.17	0.95	6.043	0.92	0.09
132	VCC534	0	3	0.31	2.06	-18.49	-19.61	1.12	0.01	22.47	1.79
133	VCC571	0	1	1.78	1.75	-16.77	-17.71	0.94	0.036	10.1	
145	VCC180	0	1	0.81	3.35	-16.97	-17.84	0.86	0.01	13.96	
146	VCC1102	0	12	0.39	1.1	-13.77	-14.29	0.52	0.361	4.65	1.11
147	VCC222	0	3	0.35	3.22	-20.39	-21.6	1.21	0.0060	47.47	1.63
148	VCC368?	1	-1	2.04	2.47	-13.18	-14.28	1.09	1.965	3.72	
149*	None	1	20	1.27	2.33	-13.11	-13.96	0.84	1.565	2.79	0.36
151*	VCC52	0	12	1.79	4.4	-14.8	-15.43	0.63	0.882	4.65	0.32

Table 5.4: Optical properties of HI detections behind the Virgo Cluster. Objects with multiple or unclear optical counterparts are omitted. Columns are as for table 5.3 except that the previous column (6), sky distance from M49, is omitted.

(1) Source	(2) Name	(3) Flag	(4) Type	(5) Sep.	(6) M_g	(7) M_i	(8) g-i	(9) M_{HI}/L_G	(10) Diam.	(11) Def.
4	None	0	20	0.17	-19.41	-19.93	0.51	1.16	38.35	-0.05
7	None	0	20	0.1	-19.02	-19.59	0.57	1.1	25.9	-0.16
8	None	0	18	1.09	-20.83	-21.76	0.93	0.09	34.44	0.4
9	None	0	20	0.25	-18.61	-19.32	0.71	0.8	24.2	0.08
11	None	1	20	0.36	-17.52	-17.26	-0.26	2.86	18.26	-0.24
12	None	0	20	0.47	-18.17	-18.74	0.57	1.9	29.18	0.02
15	VCC377	0	6	0.44	-20.22	-20.96	0.74	0.7	47.0	-0.19
16	AGC221659	0	20	0.3	-18.85	-19.9	1.05	0.68	9.99	-0.61
17	None	1	20	1.35	-15.99	-16.45	0.46	2.31	27.61	0.77
18	None	0	20	0.37	-18.67	-19.11	0.44	0.78	34.15	0.33
19	None	0	20	0.1	-18.25	-18.71	0.45	1.08	19.77	-0.05
20	None	0	20	0.28	-18.47	-19.03	0.56	0.64	21.58	0.14
21	Two	0	20	0.05	-19.89	-20.69	0.79	0.19	28.43	0.3
22	IC3150	0	20	0.46	-19.24	-19.71	0.46	0.55	23.73	-0.02
23	VCC1152	0	5	0.26	-20.35	-21.34	0.99	0.09	43.37	0.6
24	VCC712	0	4	0.26	-20.4	-21.08	0.67	0.36	43.02	0.14
25	None	1	20	1.72	-16.52	-16.93	0.41	4.29	10.68	-0.42
26	VCC908	0	20	0.85	-19.53	-20.34	0.8	0.22	25.41	0.31
27	VCC1182	0	20	2.7	-19.4	-20.11	0.71	0.2	21.38	0.26
28	None	0	20	0.29	-17.26	-17.69	0.43	0.46	8.1	0.03
29	None	3	20	0.7	-14.39	-14.99	0.6	2.94	8.96	0.45
31	None	1	20	0.5	-16.3	-16.67	0.36	0.13	11.51	1.2
32	None	0	20	0.14	-18.29	-18.85	0.56	0.72	22.69	0.21
33	None	0	20	0.11	-16.87	-17.45	0.58	1.6	20.68	0.36
35	VCC724	1	-1	0.24	-15.98	-16.73	0.75	1.63	15.78	
36	VCC638	0	5	0.21	-20.89	-22.01	1.11	0.1	62.13	0.54
37	None	1	20	0.27	-15.55	-15.96	0.4	1.35	7.45	0.18
38	VCC694	1	-1	0.47	-15.24	-15.85	0.61	0.46	7.55	
39	VCC277	0	12	0.12	-18.0	-18.44	0.44	1.84	21.11	-0.13
41	None	0	20	0.27	-17.24	-17.64	0.4	0.8	9.98	-0.03
42	VCC264	0	7	0.07	-20.5	-21.33	0.83	0.54	49.61	0.03
43	VCC265	0	3	0.12	-20.87	-20.67	-0.2	0.12	23.99	0.19
44	VCC380	0	17	0.0	-17.22	-17.45	0.22	1.1	19.73	0.34
83	None	1	20	2.21	-20.11	-20.91	0.8	0.28	37.17	0.25
85	CGCG043-031	0	20	0.27	-20.47	-21.19	0.72	0.73	52.77	-0.03
86	None	0	20	0.09	-18.33	-18.91	0.57	0.99	32.05	0.31
87	None	0	20	0.7	-19.57	-20.38	0.8	0.77	46.53	0.2
91	None	0	20	0.71	-18.28	-18.88	0.59	1.06	35.62	0.38
92	VCC1506?	1	20	1.46	-16.6	-17.18	0.58	1.6	19.66	0.43
94	None	0	20	0.0	-19.38	-20.44	1.06	1.31	77.25	0.44
95	VCC1525	0	6	0.71	-21.06	-21.84	0.78	0.13	38.16	0.06
96	None	0	20	0.28	-18.19	-19.0	0.81	0.77	30.01	0.43
98	VCC1774	0	18	0.3	-19.57	-20.61	1.03	0.26	28.2	0.29
99	None	0	20	0.0	-17.4	-17.9	0.49	0.92	16.02	0.19
100	VCC1802	0	18	0.21	-19.78	-20.8	1.02	0.61	36.79	0.04
101	None	0	20	0.14	-19.44	-20.44	1.0	0.53	31.14	0.12
102	None	3	20	0.31	-15.81	-16.37	0.56	5.34	6.4	-0.62
103	None	0	20	1.21	-17.17	-17.55	0.37	1.31	17.8	0.21
104	VCC2067?	1	20	0.72	-16.97	-17.51	0.53	1.84	16.33	0.08
105	None	0	20	1.55	-17.53	-17.89	0.35	0.56	12.7	0.18
106	None	1	20	0.54	-16.71	-16.97	0.25	1.45	10.16	-0.06
107	None	1	20	1.21	-14.93	-15.49	0.56	2.73	12.56	0.53
108	None	1	20	0.36	-16.33	-16.56	0.23	1.87	13.76	0.2
110	None	0	20	0.3	-17.42	-17.94	0.52	0.82	25.73	0.59
126	None	0	20	0.86	-16.78	-17.37	0.59	1.22	15.52	0.29
135	None	1	20	0.78	-19.4	-19.83	0.43	0.37	23.7	0.08
136	None	1	20	0.66	-17.94	-18.38	0.43	0.73	19.26	0.21
137	None	1	20	0.09	-18.47	-19.04	0.57	1.42	39.93	0.27
138	VCC161	0	6	1.36	-20.86	-21.9	1.04	0.47	96.2	0.1
139	None	0	20	0.95	-19.36	-20.19	0.82	0.26	32.46	0.49
140	None	1	20	1.25	-16.92	-17.7	0.78	2.11	19.08	0.16
142	None	1	20	0.58	-16.77	-17.29	0.52	3.17	15.82	-0.09
150	None	0	20	1.1	-15.43	-16.25	0.81	0.92	10.51	0.66

Table 5.5: UV properties of HI-detected objects detected in both UV bands within the Virgo Cluster. Columns : (1) Source number in this catalogue (2) Name in a major catalogue if present (3) Distance in Mpc (4),(5) Apparent magnitude in the FUV and NUV wavebands (6), (7) Absolute magnitude in the FUV and NUV (8) FUV - NUV colour

(1) No.	(2) Name	(3) Distance	(4) mag_{FUV}	(5) mag_{NUV}	(6) M_{FUV}	(7) M_{NUV}	(8) $FUV - NUV$
45	VCC415	32.0	17.47	17.21	-15.05	-15.31	0.26
46	VCC73	32.0	17.21	16.24	-15.31	-16.28	0.96
47	VCC450	32.0	19.73	18.9	-12.79	-13.62	0.82
48	None	3.0	18.57	18.18	-13.23	-13.62	0.38
49	VCC94	32.0	18.06	17.44	-14.46	-15.08	0.61
50	FGC1384	2.0	18.99	18.6	-13.53	-13.92	0.39
51	None	2.0	19.22	18.88	-13.3	-13.64	0.34
52	VCC199	32.0	18.64	17.38	-13.88	-15.14	1.26
53	VCC393	32.0	16.1	15.51	-16.42	-17.01	0.59
54	None	3.0	18.7	18.64	-13.1	-13.16	0.05
57	VCC343	32.0	17.86	17.43	-14.66	-15.09	0.43
58	None	2.0	20.47	19.92	-12.05	-12.6	0.55
59	VCC207	32.0	17.92	17.79	-14.6	-14.73	0.12
61	VCC656	23.0	17.34	16.69	-14.46	-15.11	0.64
62	VCC697	23.0	16.77	16.41	-15.03	-15.39	0.36
63	VCC827	23.0	16.19	15.69	-15.61	-16.11	0.5
64	VCC975	23.0	15.04	14.79	-16.76	-17.01	0.25
66	None	2.0	19.63	19.33	-12.89	-13.18	0.29
67	None	3.0	19.15	19.31	-12.65	-12.48	-0.16
68	None	3.0	18.76	18.67	-13.04	-13.13	0.09
65	VCC667	23.0	17.08	16.61	-14.72	-15.19	0.46
69	VCC851	23.0	17.47	16.76	-14.33	-15.04	0.7
70	None	3.0	20.16	20.02	-11.64	-11.78	0.13
71	None	7.0	19.08	18.86	-12.06	-12.28	0.22
72	VCC989	17.0	18.84	18.48	-12.31	-12.67	0.36
73	VCC1011	17.0	17.49	17.19	-13.66	-13.96	0.29
74	VCC105	32.0	16.21	15.82	-16.31	-16.7	0.38
75	VCC688	23.0	16.98	16.64	-14.82	-15.16	0.34
77	None	3.0	20.04	19.63	-11.76	-12.16	0.4
79	VCC758	23.0	19.72	17.64	-12.08	-14.16	2.08
80	None	3.0	20.19	20.02	-11.61	-11.78	0.17
81	VCC1193	17.0	16.97	16.44	-14.18	-14.71	0.52
114	VCC1699	17.0	16.62	16.38	-14.53	-14.77	0.24
116	VCC1455	17.0	19.1	19.03	-12.05	-12.12	0.06
119	VCC1675	17.0	17.72	17.25	-13.43	-13.9	0.46
121	VCC1750	17.0	14.81	13.98	-16.34	-17.17	0.83
123	VCC1575	17.0	16.7	16.08	-14.45	-15.07	0.62
132	VCC534	23.0	18.89	17.8	-12.91	-14.0	1.09
133	VCC571	23.0	19.76	18.67	-12.04	-13.13	1.09
145	VCC180	32.0	19.34	18.64	-13.18	-13.88	0.7
147	VCC222	32.0	18.95	17.23	-13.57	-15.29	1.72

Table 5.6: UV properties of HI non-detected objects detected in both UV bands within the Virgo Cluster. Columns are as for table 5.5 except that the previous column 1 is omitted as it is not applicable.

(1) No.	(2) Distance	(3) mag_{FUV}	(4) mag_{NUV}	(5) M_{FUV}	(6) M_{NUV}	(7) FUV-NUV
VCC77	32.0	21.22	20.52	-11.3	-12.0	0.69
VCC80	32.0	21.03	20.83	-11.49	-11.69	0.2
VCC220	32.0	19.1	17.59	-13.42	-14.93	1.51
VCC255	23.0	19.74	19.52	-12.06	-12.28	0.21
VCC657	23.0	18.31	17.64	-13.36	-14.4	1.03
VCC731	23.0	16.23	15.2	-15.57	-16.6	1.02
VCC1178	17.0	19.47	18.1	-11.68	-13.05	1.36
VCC1199	17.0	21.95	21.44	-9.2	-9.71	0.51
VCC1254	17.0	20.76	19.51	-10.39	-11.64	1.25
VCC1535	17.0	16.73	15.43	-14.42	-15.72	1.29
VCC1692	17.0	17.89	16.49	-13.26	-14.66	1.39
VCC1883	17.0	19.13	16.87	-12.02	-14.28	2.25
VCC1903	17.0	16.73	15.4	-14.42	-15.75	1.32
VCC1913	17.0	19.54	17.84	-11.61	-13.31	1.7
VCC1938	17.0	18.65	16.77	-12.5	-14.38	1.87
VCC2000	17.0	18.16	16.69	-12.99	-14.46	1.46

Chapter 6

Results 2 : The VC2 Area

6.1 Introduction

As described at the start of chapter 5, two areas of the Virgo Cluster have been selected for study by AGES. The first, a larger and richer area, has been described in chapter 5. The second, a smaller area in a lower-density region of the cluster outskirts, is described here. The exact positions of the two areas are shown in figure 1 of chapter 5.

Study of a second area is essential as a comparison region, given the large size of the cluster and the relatively small area AGES is able to survey. Observing the entire cluster to full depth is unfortunately impractical, necessitating over 1,000 hours of telescope time (more than half of the total time for the whole survey). The VC1 area is composed of different infalling clouds, but large areas of the cluster are thought to contain galaxies of a single population (as shown in chapter 1). The results of chapter 5 might, therefore, not be typical of the cluster as a whole, hence the need for a comparison region.

Galaxies within the VC2 region are all thought to be part of subcluster A and are assigned a distance of 17 Mpc in GOLDMine, eliminating any confusion with infalling clouds. Any trends detected should therefore only be trends within a single population of galaxies. The central giant elliptical galaxy M87 is avoided, as its strong continuum emission can cause severe interference by setting up standing waves in the Arecibo dish. The field extends beyond the edge of the VCC plates, so it is hoped that changes in galaxy properties from the cluster interior to the field may be seen. Physically the area spans 1.5×0.3 Mpc at 17 Mpc distance.

This area, then, is rather simpler than VC1. While VC1 includes separate subclusters, VC2 observes a single cluster population over a range in densities. In principle this should make any visible trends easier to understand, without having to disentangle the effects of a mixed population. The disadvantage is that the region is both less dense than VC1 and smaller, which is reflected in the lower number of HI detections.

Chapter 5 was almost entirely concerned with the cluster detections within VC1. As the sample within the cluster is much smaller in VC2, this chapter also serves to describe some of the interesting sources found behind the cluster in both VC1 and VC2.

6.2 Data analysis overview

The observations, data reduction and analysis procedures have been fully described in chapter 2. The field consists of an area of uniform sensitivity 5 degrees in R.A. by 1 degree in declination; due to the hexagonal beam pattern of ALFA a small area outside this field is also included. The maximum range of the field in R.A. is from 12:37:35 to 12:58:52; in declination it ranges from +10:56:00 to +12:13:00. Observations are complete for this region.

The sensitivity attained is very similar to that of the fully-observed area of VC1 - a minimum of 0.4 mJy/beam, a median of 0.6 mJy/beam and a mean of 0.8 mJy/beam based on measurements of the individual HI detections. As for VC1 a sensitivity of 0.6 mJy corresponds to an HI mass sensitivity of 8.2×10^6 $M_\odot$ at 17 Mpc (for a 4σ, 50 km/s velocity width source with a tophat profile). This is significantly lowered in regions contaminated by RFI, which are shown in figure 6.1. The usual major sources - the Milky Way, the FAA radar at San Juan airport, and the L3 GPS satellite - are labelled. Other more minor sources (mostly harmonics of the radar) are visible as white

bands.

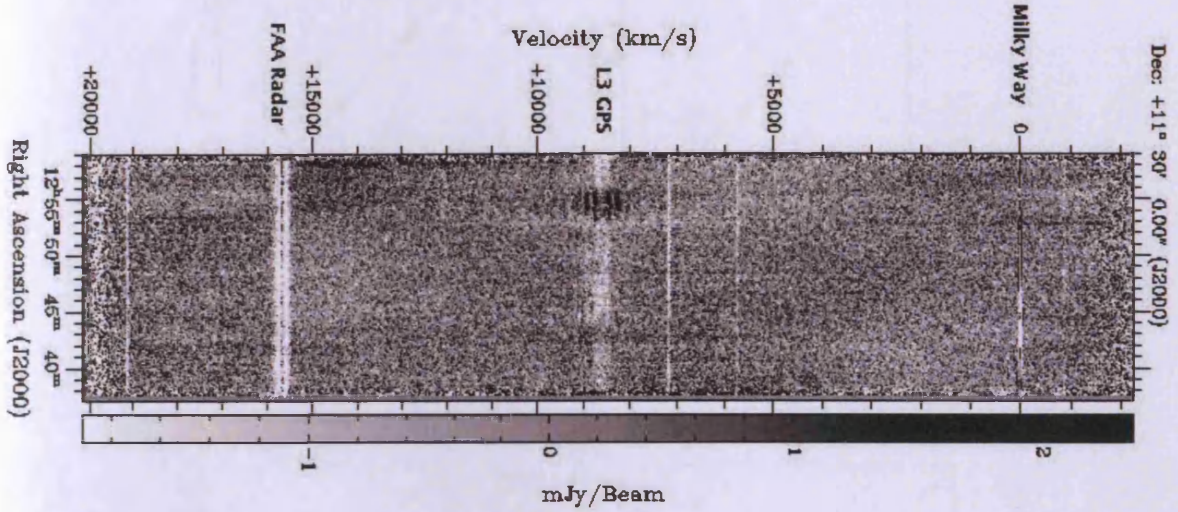


Figure 6.1: Position-velocity slice through the VC2 cube showing the major sources of RFI.

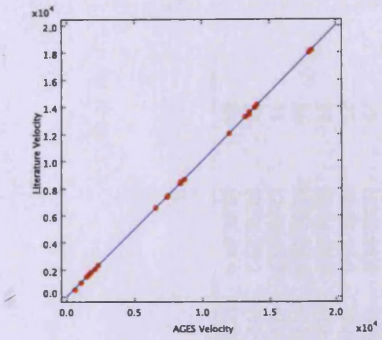
Source extraction is performed as described in chapter 4 : visual inspection by a single observer, Polyfind, Duchamp and the new extraction method described in that chapter, as well as a visual search based on the positions of VCC galaxies. After L-wide follow-up observations (see chapter 2) a total of 49 detections remain, of which 7 still require follow-up observations. 13 objects are classed as Virgo Cluster members ($v < 3{,}000$ km/s), 2 are high velocity clouds and the remaining 34 are galaxies behind the Virgo Cluster. 7 cluster member detections are listed in the VCC - 2 more could not have been listed since they are outside the VCC survey area, the other 4 escaped detection by the optical survey (see section 6.4).

Optical data is described in chapter 3 - here, INT WFS data is used to generate surface brightness profiles in the B and i bands except when a source is too faint to fit a profile, in which case aperture photometry is performed.

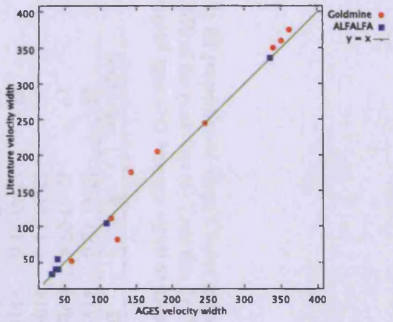
6.3 Comparisons with other data

From the GOLDMine database there are a total of 22 objects in this region listed as having previous (pointed) HI observations, half of which are claimed as detections and half as non detections. Of the detections, 2 are listed as quality 5 meaning unpublished and not necessarily reliable. Both of these galaxies (VCC 1889 and VCC 2095) are listed as having sufficient HI that they should be easily detected by AGES if genuine, but neither are actually detected (the rest of the GOLDMine detections are also detected by AGES). The remaining 11 non-detections with observations have limits placed on the amount of HI present, based on the noise level of the GOLDMine observations. None of these objects were detected by AGES. In only 3 cases AGES does not go deeper than previous studies - i.e. it is not sensitive enough to place even lower limits on the possible amount of hydrogen present than the previous observations. With one exception that is just beyond the velocity range detectable by AGES, these are all early-type galaxies, therefore it is not surprising that none of them were detected.

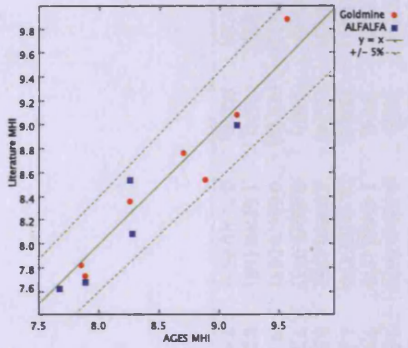
Comparisons with previous observations are shown in figures 6.2, for those objects detected both in AGES and previous studies. The width comparison uses the FWHM (i.e. the W50) for ALFALFA but for GOLDMine this is not listed. Instead, GOLDMine gives the velocity width as the average of the W50 and W20, so in the comparison the AGES velocities in the figure are also the average of the W50 and W20. For the optical data, the comparison is between the extrapolated B magnitudes from these measurements with those listed on GOLDMine (elsewhere the last integrated magnitude is used to avoid any contamination introduced in the extrapolation - see chapter 3).



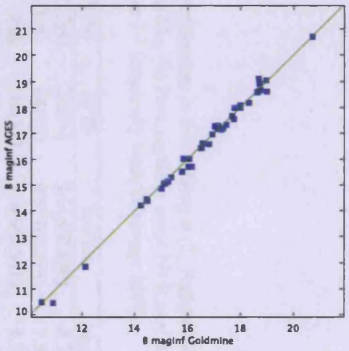
(a) Comparison of AGES measured recessional velocities with literature values from NED for optical counterparts



(b) AGES HI W50 compared with literature values



(c) AGES HI mass compared with literature values



(d) Comparison of our measurements of magnitudes of VCC galaxies against those listed in GOLDMine.

Figure 6.2: Comparisons of optical and HI data with existing measurements from the literature. The solid line shows the one to one correlation.

From figure 6.2, in general it is apparent that agreement between data sources is very good. Quantitatively, the mean difference between measurements are 4 km/s for recessional velocities (maximum 120 km/s), 4 km/s for velocity width (maximum 42 km/s), and 0.3 magnitudes for the extrapolated B-band magnitudes (maximum 0.6 magnitudes). For the HI masses, the two largest deviations are due to RFI and one source being at the very edge of the cube (and so some flux being missed), but there are no significant or apparent significant systematic offsets. The mean positional offset between HI spatial coordinates and the optical counterpart is 20" with a maximum of 1.4' (for detections with sure optical counterparts as described in chapter 3).

6.4 Virgo Cluster Objects

Table 6.1 lists the HI parameters of the objects identified as Virgo cluster members. All quantities are measured/calculated as described in chapters 2 and 3. Table 6.2 lists all optical parameters for Virgo objects as described in chapter 3 - those not included are either because they are outside the INT area, overlap a CCD gap or have multiple possible optical counterparts.

Comparing table 6.1 with table 5.1 (in chapter 5), it is immediately obvious that there are far fewer detections in VC2 than in VC1. This is not simply because VC1 is larger : there are about 2.6 HI detections per square degree in VC2 compared with 6.0 per square degree in VC1. It may also be noted from table 6.1 that the S/N estimate

Table 6.1: HI properties of Virgo Cluster galaxies. Columns : (1) Source number in this catalogue (2) Right ascension J2000 (3) Declination J2000 (4) Heliocentric velocity km/s (5) Maximum velocity width at 50% and (6) 20% of the peak flux (6) Total flux Jy (7) Estimated HI mass, log $M_{\odot}$ (8) Peak signal to noise (9) R.m.s. mJy (10) Reliability flag (11) Optical counterpart flag (12) Name of counterpart in major catalogues (13) Morphological type (14) Average signal to noise estimated using equation 3.1 (originally from Saintonge 2007).

Source no.	R.A.	Dec.	Velocity	W50	W20	TotFlux	M_{HI}	S/N	rms	Flag	Optical Flag	Other Names	Mtype	ALFALFA S/N
18	12:41:12.0	+11:53:13	2259(2)	227(4)	248(6)	3.614(0.190)	8.34	26.69	0.6	0	0	VCC1868	8	88.05
19	12:43:07.4	+12:03:21	2026(3)	41(6)	84(10)	0.537(0.078)	7.51	19.31	0.7	0	0	VCC1955	14	26.46
20	12:48:11.0	+10:58:19	1163(4)	78(8)	116(12)	3.785(0.531)	8.36	12.62	4.2	0	0	VCC2066	13	20.01
22	12:43:31.6	+11:34:51	1418(2)	173(3)	195(5)	6.067(0.317)	8.56	62.7	0.6	0	0	VCC1972	7	165.32
23	12:37:44.0	+11:49:20	1537(3)	331(6)	368(9)	7.906(0.551)	8.68	19.36	2.4	0	0	VCC1727	4	39.81
24	12:44:58.1	+12:01:56	1544(2)	37(3)	52(5)	0.538(0.067)	7.51	27.53	0.5	0	1	None	20	29.08
25	12:51:05.9	+12:03:33	1793(1)	39(3)	54(4)	1.466(0.157)	7.95	63.8	0.6	0	0	None	19	62.65
27	12:51:55.0	+12:05:00	1785(2)	334(3)	358(5)	12.151(0.476)	8.87	59.9	0.7	0	0	NGC4764	5	207.89
29	12:44:08.9	+12:07:05	1013(1)	106(3)	122(4)	4.039(0.279)	8.39	44.53	0.9	0	0	VCC1992	12	94.67
30	12:38:39.8	+11:58:46	1041(3)	42(6)	74(9)	0.402(0.063)	7.39	16.67	0.6	0	0	None	19	22.56
33	12:40:55.2	+11:55:03	1653(6)	115(12)	195(18)	0.666(0.085)	7.6	11.8	0.6	0	0	VCC1859	3	21.58
35	12:56:43.2	+11:55:52	569(1)	57(3)	76(4)	1.847(0.169)	8.05	92.58	0.4	0	0	UGC08061	12	114.56
63	12:56:04.9	+12:09:07	1869(10)	32(20)	89(30)	0.219(0.092)	7.55	5.25	1.2	0	1	None	20	7.18

Table 6.2: Optical properties for all Virgo Cluster HI detections where INT data is available and only one optical counterpart is visible. M_{HI}/L_B ratios are in $M_\odot/L_\odot$, R75 is in arcseconds. Sources 30 and 63 were measured using aperture photometry - very low surface brightness made fitting a surface brightness profile impossible, and fringing prevents measurements in the i band.

Source no.	OtherNames	Mtype	m_B	m_i	B-i	M_B	M_{HI}/L_B	$R75_B$ (Kpc)	HI def.
VC219	VCC1955	14	14.0	12.34	1.67	-17.02	0.03	15.8	1.87
VC222	VCC1972	7	11.53	10.04	1.49	-19.49	0.04	33.5	0.87
VC223	VCC1727	4	10.52	8.6	1.92	-20.5	0.02	38.5	0.98
VC224	None	20	19.46	18.62	0.83	-11.56	5.06	2.3	-0.03
VC225	None	19	17.42	16.44	0.97	-13.61	2.13	5.59	0.28
VC229	VCC1992	12	15.94	15.0	0.95	-15.08	1.51	4.31	0.52
VC230	None	20	20.48	-	-	-10.54	9.86	-	0.90
VC233	VCC1859	3	12.75	11.39	1.36	-18.27	0.01	20.3	1.81
VC263	None	20	20.64	-	-	-10.51	6.14	-	-0.15

- either using peak S/N or the ALFALFA measurement - is high for all Virgo detections. The lowest peak S/N is 5.25, but the next lowest is 11.80. Yet it is very difficult to ascribe this result to poor source extraction procedures; as described in chapter 4 this particular cube has been searched by many different, independent methods. However, lower S/N detections in the cluster were originally recorded, but all were rejected based on follow-up observations (some in the background have been confirmed, see section 6.5).

The relatively low quantity of HI detections in VC2, in comparison with VC1, is not a source extraction or data reduction problem. Rather it is simply because VC2 is, overall, a less dense region of the cluster. In the optically-selected VCC, there are 13.4 cluster members per square degree compared with 17.4 in VC1. Moreover, the difference in galaxy numbers between VC1 and VC2 is not unusual but simply the varying nature of galaxy density in the cluster - if we consider the 5×1 degree strip immediately west of VC2, there are 23.6 VCC galaxies per square degree in that region. The key point of the lower number of detections is that any kind of detailed comparison between VC1 and VC2 will suffer from small number statistics.

6.4.1 The HI mass function

As mentioned above the S/N of the HI detections in VC2 are all relatively high. This is reflected in the high HI masses of the galaxies detected in this region, with only 1 lower than 3×10^7 $M_\odot$ (7.5 in logarithmic solar units). 13 galaxies below this limit were detected in VC1 although it covers an area only twice the size of VC2. Given the higher galaxy density in VC1, more low-mass galaxies should be expected there simply because there are more galaxies present overall. If galaxies of low HI-mass were detected in VC2 in the same proportion as in VC1, 3 detections would be expected as opposed to the 1 that is actually found - so the number of low-mass detections in the two areas is consistent given $\sqrt{n}$ errors.

Directly comparing the HI mass distributions between VC1 and VC2 is very difficult since there are only 13 objects in VC2, and clouds at 3 different distances (and so differing mass sensitivity limits) within VC1. It is more useful to collate all the HI detections and produce an HI mass function, using, for consistency with VC2, only those objects at 17 Mpc from VC1. This gives a total sample size of 36 objects. The mass function is shown in figure 6.3, which compares the AGES result with that of the VIRGOHI survey of Davies et al. 2004 and the currently published data from ALFALFA in the Virgo region (the surveys are made consistent by rejecting galaxies in infalling clouds and assuming a distance of 17 Mpc for the rest).

The sensitivity limits shown in figure 6.3 assume a 4σ, 50 km/s velocity width with a tophat profile for AGES and VIRGOHI. These surveys employ(ed) visual inspection, so quantifying a precise, accurate sensitivity (let alone completeness) limit is difficult - particularly given the 3-dimensional nature of the data cubes, as discussed in chapter 4. Consequently some detections are below the sensitivity limits shown. For ALFALFA the sensitivity limit assumes a 4.5σ limit as estimated by equation 3.1 - Giovanelli et al. 2007 estimate this as a lower limit for their detections.

Figure 6.3 shows evidence for differences in the estimates of HIMF by each survey. At the high mass end ($> 10^9$ $M_\odot$), ALFALFA and VIRGOHI are in agreement but AGES detects relatively few objects. AGES detects only 2 galaxies more massive than 8.75 in logarithmic solar units (i.e. in the bins centered at 9.0 and 9.5), or 0.13 per square degree. ALFALFA and VIRGOHI detect 0.79 and 0.26 per square degree respectively. Yet even with the 4 mJy sensitivity of of VIRGOHI21, detecting objects with $M_{HI} > 8.75$ should not be difficult except for the very

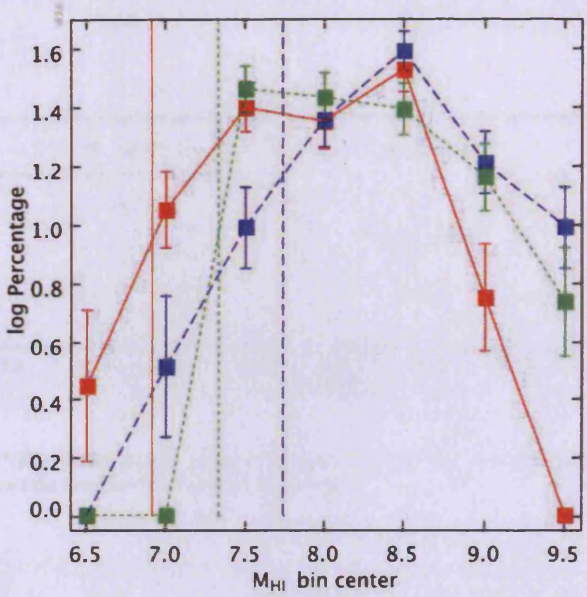


Figure 6.3: HI mass functions for 36 AGES galaxies at 17 Mpc (solid red line), the 31 galaxies of VIRGOHI at 17 Mpc (blue dashed line), and the 167 galaxies in ALFALFA within the limits of the VCC (green dashed line). The x-axis is the HI mass bin center in logarithmic solar units with a bin size of 0.5 (i.e. the 7.0 bin includes galaxies from 6.75 to 7.25). The y-axis is the logarithm of the percentage of the galaxies in each sample in each bin. Errors are root n. The vertical lines of the corresponding colour and dash show the approximate sensitivity limits for each survey (see text).

widest sources - to fall below a S/N of 5.0, a source of this mass would need a velocity width exceeding 400 km/s at 4 mJy sensitivity. The more sensitive AGES cannot be missing such objects due to its sensitivity. Therefore, the differences at the high mass end must be due to cosmic variance in the distribution of massive sources.

At the low mass end of the HIMF (< 7.5 in logarithmic solar units) there are also differences between the surveys due to incompleteness. As the deepest survey AGES detects a greater proportion of low HI-mass objects than the others, despite their larger area. However, it is interesting to note the detection of a single object in the 7.0 bin in the VIRGOHI HIMF - the shallowest survey in this comparison - but none in the deeper ALFALFA survey. The object in question is VCC 963, a dwarf irregular, which falls within the ALFALFA survey area. It is actually detected by ALFALFA, but at a mass nearly 3 times greater - hence it is included in a different mass bin. Small number statistics means that even individual objects can have a significant impact on the measured HIMFs. Thus, while it may appear that the shape of the AGES HIMF is different to the others at low masses, this result should be treated with caution at least until AGES is complete within Virgo (even then incompleteness due to sensitivity will still be an issue).

The importance of cosmic variance must also be stressed. Of the AGES objects below 7.5 in mass, 13 out of the 14 detections are in VC1. When ALFALFA covers this area, it may detect some low-mass objects, but it seems unlikely that this fraction will increase drastically given the sensitivity limit. 25% of AGES detections are less massive than 7.5, so ALFALFA should have detected about 40 objects if it detected the same proportion. Also, as noted in chapter 5 there are many AGES detections in the region also covered by ALFALFA that ALFALFA did not detect.

6.4.2 Distribution

In chapter 5 the HI detections in VC1 were shown to broadly follow the same spatial and velocity distributions as the VCC non-detections. The same cannot be said for VC2, as shown in figure 6.4. Here it appears that the HI detections are found primarily in a narrow band in declination which is quite contrary to the result from the VCC non-detections (figure 6.5), although both samples show similar clustering in right ascension. There is also a suggestion that the HI detections are at higher velocities than the optically selected sample (panel (c) of the same

figure), although the result is not as clear as in Davies et al. 2004. This is also shown in figure 6.6.

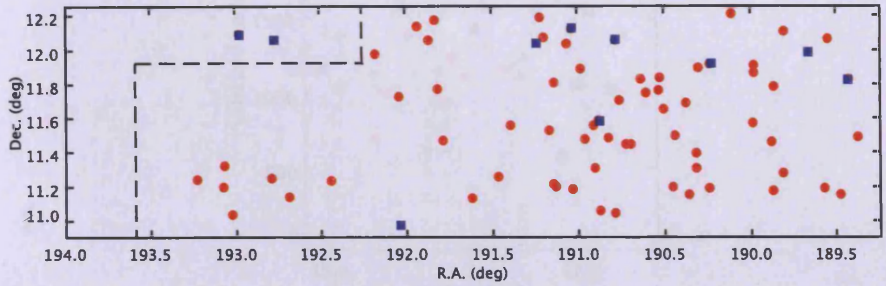
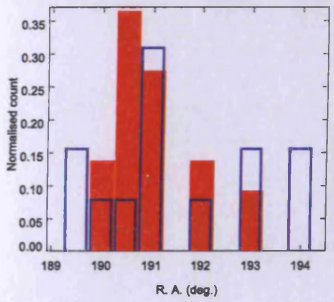
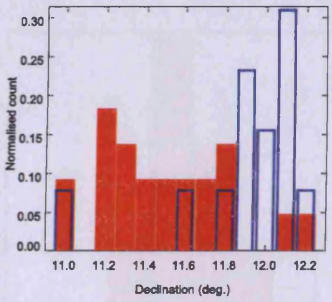


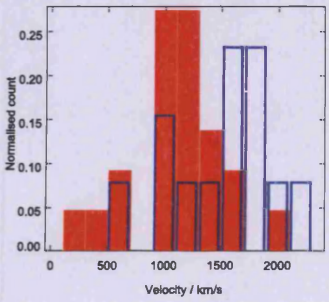
Figure 6.4: Spatial distribution of Virgo Cluster galaxies in the VC2 region. Red circles are the 67 VCC objects, blue squares are the 13 AGES HI detections. The dashed line shows the boundaries of the VCC survey area.



(a) R.A. distribution, bin width of 0.5 degrees



(b) Declination distribution, bin width of 0.1 degrees



(c) Velocity distribution, bin width of 200 km/s

Figure 6.5: Distribution histograms of HI detections with $cz < 3{,}000$ km/s (blue) and VCC cluster members not detected in HI (red). All 60 VCCs without AGES detections are used in the spatial histograms, only those 23 with optical redshifts are used in the velocity distribution. In all histograms the count of both samples has been normalised to give a total of 1.0 for each.

From the raw numbers, the mean declination of the HI detections is $+11.89 \pm 0.30°$ (1σ error); for the non-detections it is $+11.55 \pm 0.35°$. The mean velocities are $1{,}513 \pm 449$ and $1{,}136 \pm 393$ for the detections and non-detections respectively. This means the distributions actually agree to within 1σ. However, this does not account for the different number of detections and non-detections. A Kolmogorov-Smirnov test shows that the declination distributions are different at greater than 99.9% significance level, the velocity distributions different to 97.5% significance.

This result suggests that this part of the cluster is not a single, relaxed population. In VC1 HI detections were seen to be mixed with non-detections with no obvious bias in their distributions, with objects of varying degrees of environmental influences being found in close proximity to one another. In VC2 it appears that there may be two distinct populations beginning to coalesce. This is further evidence that the individual subclusters of Virgo are themselves still assembling (see chapter 5). The difference in distributions is not at all evident (in this small area) from previous pointed observations (see figure 5.1).

This difference in velocity distribution of HI detections and non-detections was also noticed on the much larger scale of the whole cluster - see figures 1.9 and 1.10. In chapter 5 it was found that the HI detections are present in the same clouds as the VCC non-detections - but in fact there is a difference in their velocity distributions, as shown in figure 6.7 (this is because many of the HI detections in VC1 are found in the W cloud). The evidence is clear that HI detections within the whole cluster are generally at higher velocities than non-detections - again,

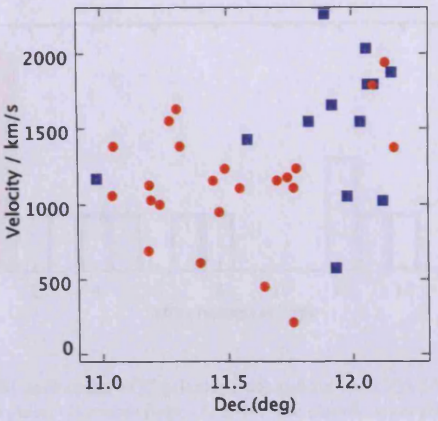


Figure 6.6: Declination-velocity plot for HI detections (blue squares) and VCC galaxies not detected in HI (red circles) in the VC2 region.

assembly of the Virgo Cluster is an ongoing process.

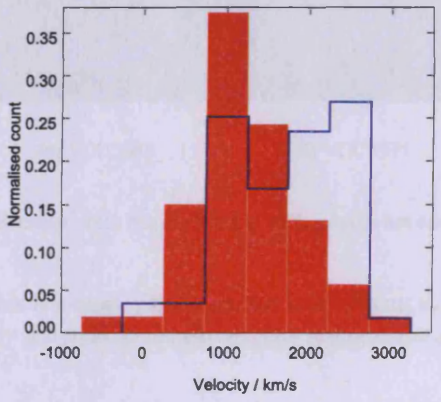


Figure 6.7: Velocity distribution of AGES detections (blue) and non-detected VCC galaxies (red) in the VC1 region, bin width of 500 km/s.

6.4.3 Morphologies

In the VC1 region it was found that most HI detections were of late-type objects, even if not members of the VCC, with virtually all late-type VCC objects being detected. This is also the case in VC2, as shown by figure 6.8. The two "undetected" unclassified objects are VCC 1977 and 2085 (see figure 6.9). VCC 1977 has a colour ($B - i = 1.6$) typical of the sample and it is a genuine non-detection in HI, but is so faint it could easily be an early-type object such as a dwarf elliptical. VCC 2085 is clearly late-type but there is an uncertain detection at this spatial position (still awaiting follow-up) at 7,300 km/s. It is likely that every late-type VCC cluster member in this region is detected in HI.

Two galaxies have not been shown in the morphological distribution histogram as the HI appears to be associated with both of them, linked by a stream. This is the only example of an extended extragalactic HI tail detected in Virgo by AGES, and may be associated with early-type objects. VCC 2066 is assigned a type 20 (unclassified) in GOLDMine, but is listed as a peculiar S0. The nearby VCC 2062, which is projected against the bulk of the gas but has no optical redshift measurement, is classed as a dwarf elliptical.

In VC1 about 10% of the HI sample correspond to early-type objects, and in chapter 5 it was suggested that the detected S0s are being evolved via gas loss from spirals and the dEs recently accreted field objects. In VC2 there are only 13 HI detections so if the detection rate of early-types is consistent with that of VC1, a single early-type

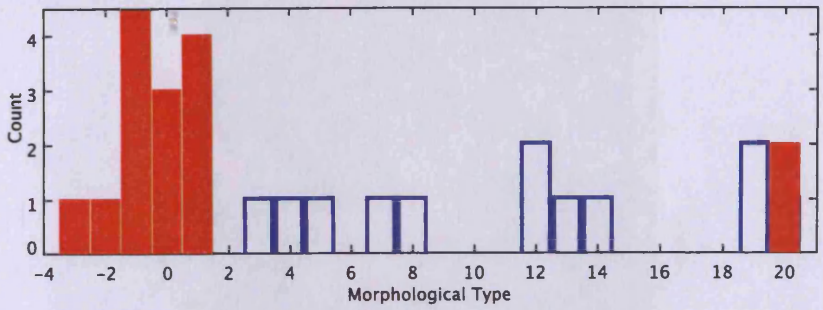
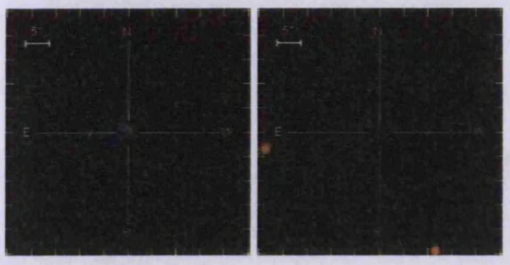


Figure 6.8: Morphology distribution of the 61 undetected VCC galaxies (red) and the 13 AGES detections (blue). The y-axis has been truncated for clarity; the total number of non-detected dwarf ellipticals (type -1) is 49. The classification scheme is described in chapter 3.



(a) VCC 2085 **(b)** VCC 1977

Figure 6.9: SDSS RGB images of the two unclassified VCC galaxies not detected by AGES in VC2.

HI detection might be expected. This is arguably the case, but it is difficult to precisely define the morphology of the detection here. It is clearly very different to the early-types discussed in chapter 5, none of which show any evidence of extended HI features.

Figure 6.10 shows the HI of detection 20 in position-velocity space. The bulk of the gas is actually found within the HI tail rather than directly associated with the galaxies. No optical redshift exists for VCC 2062, but the HI tail does appear to terminate at that position. Higher resolution Westerbork observations by van Driel & van Woerden 1989 show this far more clearly, as do archival VLA observations from the VIVA survey. Duc et al.'s 2007 figure 1 is reproduced in figure 6.11.

VCC 2062/2066 is clearly a very unusual system. Van Driel & van Woerden considered several possible origins for the HI : interaction with another massive galaxy, interaction with a dwarf galaxy, or gas stripping of VCC 2066. Firstly, it is important to establish the nature of the galaxies involved, since late-types are almost always detected in HI while early-types only rarely (see chapter 5 and also Alighieri et al. 2007). VCC 2066 appears morphologically like any other S0, though has been noted to be unusually blue (van Driel & van Woerden estimate the B-V colour at 0.62). VCC 2066 is classified as a dE in GOLDMine, but van Driel & van Woerden's imaging showed it to be irregular (Duc et al. 2007 and Sabatini et al. 2005 agree).

The total HI mass associated with VCC 2062/2066 is about 2.3×10^8 $M_\odot$ from AGES, although some of the gas may be truncated by the southern edge of the cube. Van Driel & van Woerden estimate it slightly higher at 5×10^8 $M_\odot$. This is more than double the amount in any ETG detected by AGES in VC1, though the borderline early-type (assigned a classification of S0a - S0/Sa in GOLDMine) VCC 94, detected in VC1, has more than 6 times more HI than this system (see chapter 5), and the S0 VCC 355 is also more HI massive (for example as shown by ALFALFA in Alighieri et al. 2008).

Unless there has been a substantial morphological and luminosity evolution, it seems unlikely that VCC 2062 could have provided all of the gas in the system. Van Driel & van Woerden note that this would give it an original

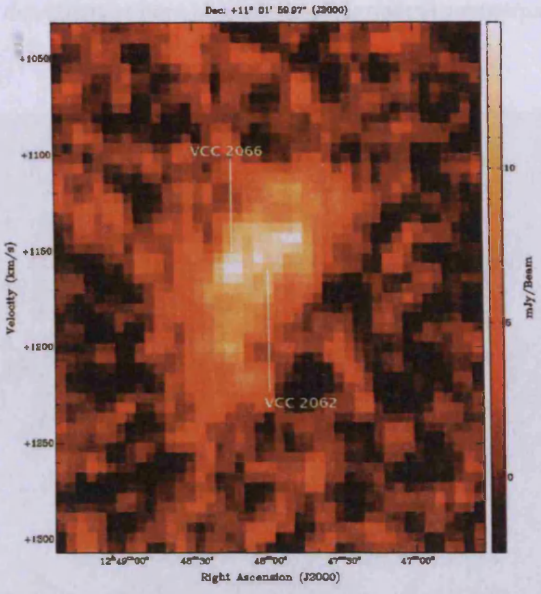


Figure 6.10: AGES position-velocity map of the HI complex associated with VCC 2066 and VCC 2062, 1' pixels.

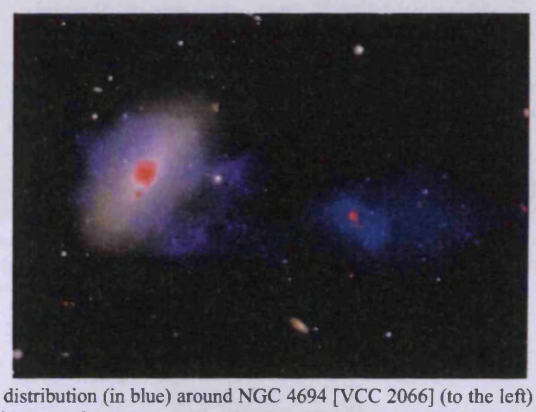


Figure 6.11: VLA map of the HI gas distribution (in blue) around NGC 4694 [VCC 2066] (to the left) and VCC 2062 (to the right) superimposed on a true colour (BVR) optical image of the system. The GALEX-FUV emission, tracing regions of recent star formation, is overlaid in red. The field of view is 9'×6'. North is up and east to the left. Figure and caption reproduced from Duc et al. 2007.

M_{HI}/L ratio of 40, far higher than any other dwarf irregular detected in Virgo (or indeed of galaxies of any other morphology). Even if stellar stripping did occur, which could change this ratio, it is difficult to see why there are no stellar streams linking VCC 2062 with 2066 along the HI tail (to the surface brightness limit of the SDSS of 26 mag arcsec^2 in the g-band)[1]. While it is possible that VCC 2062 is only coincidentally aligned with the HI and actually at a quite different distance, another possibility is that it is involved in the interaction without being the source of the HI.

While AGES observations help to rule out the origin of the gas as from an external dwarf galaxy (since they confirm that dwarfs do not possess sufficient gas), they are not of high enough spatial resolution to contribute further. A detailed analysis is given by Duc et al. 2007, using the high resolution VIVA images and data at other wavelengths (optical and UV). The authors conclude that the VCC 2062 is likely to be a tidal dwarf galaxy, with the HI provided either by VCC 2066 itself or by a gas-rich object interacting or even merging with the S0.

4 galaxies are detected which are not members of the VCC, though 2 of them are outside the VCC area. As with many of the new detections in VC1, most are small, optically faint dwarf irregulars best described as fuzzy blue blobs in optical images. SDSS RGB images for all of the non-VCC detections in both VC1 and VC2 are shown in figure 6.12 (excluding object 30 in VC2, which is too faint for the SDSS and is shown separately in figure 6.13).

[1] This, however, opens a whole new level of complexity which is beyond the scope of this study - the streams might be very dispersed, depending on their age, or their stars have been recaptured by the S0 - such speculations require a detailed study in their own right.

Objects which are uncertain HI detections or have no clear possible optical counterpart are excluded and have been discussed separately in chapter 5.

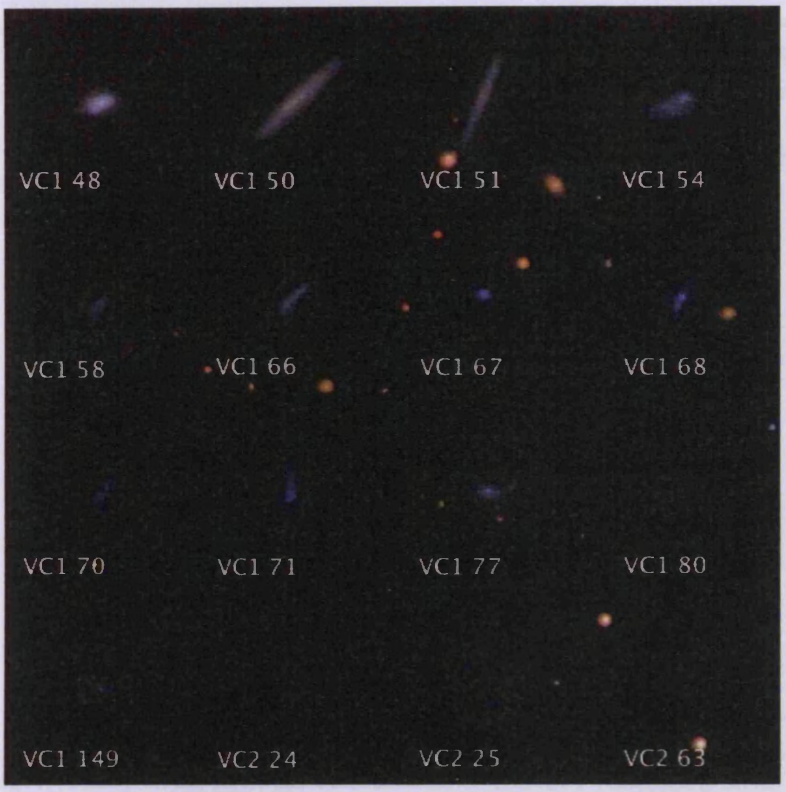


Figure 6.12: SDSS images of galaxies detected by AGES not listed in the VCC. Images are all 1' across. Only those objects which have a single obvious counterpart are shown (though the association of the HI with VC1 67 is considered uncertain and discussed in chapter 5), but not all have optical redshift measurements.

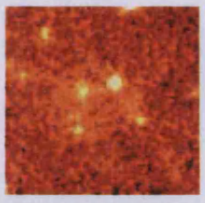


Figure 6.13: INT B-band image of object 30 in the AGES VC2 catalogue. This is not visible in the SDSS.

With the exceptions of VC1 50 and 51, which are probably edge-on spiral discs, the non-VCC galaxies detected by AGES all appear to be blue compact dwarfs or dwarf irregulars. Many resemble edge-on discs - which is interesting since face-on discs would be of lower surface brightness and therefore easier to miss in a visual inspection. Only VC2 25 and 63 are outside the VCC area - it is very doubtful that VC2 63 could have been detected by a visual inspection in any case, with an apparent B magnitude of 20.6. All are late-type, probably reflecting the preponderance of late-types in the field (as described in chapter 1) from which they are infalling (of those in VC1, all bar 1 are assigned to an infalling cloud rather than subcluster A).

The colours of the new detections are blue but consistent with their low magnitudes (g-$i \sim 0.4$, $M_g <$ -16.5 for the VC1 objects) for late-type objects. They are not of particularly low surface brightness but they are all small, with 75% less than 5 Kpc diameter. They are also of fairly high M_{HI}/L ratios (> 0.3) but usually consistent with their low optical luminosities. They are thus described as a population of small, blue, gas-rich objects. The properties of the rest of the sample are described in the remainder of the chapter.

Object 30 was previously detected in a search for low surface brightness objects (using the same INT data) by

Sabatini et al. 2005. It is a very unusual object, exceptionally faint optically but gas-rich at 2.5×10^7 $M_\odot$. Its colour cannot be accurately measured due to i-band fringing of the INT data and lower sensitivity of the SDSS. Its M_{HI}/L_B ratio (see section 6.4.5) is very high but consistent with its low optical luminosity.

Two HI non-detections are worth mentioning due to unusual morphological features: VCC 2095 and VCC 2000. VCC 2095 is clearly an edge-on disc (classed as an S0 in GOLDMine) but shows obvious disturbances at the edges - these are well known and described as "brushes" in NED notes. The SDSS RGB image is shown in figure 6.14. They appear to be the result of a tidal interaction, possibly involving the nearby elliptical VCC 2092 (though there are 245 references to VCC 2095 listed in NED, a detailed study of the prospect of an interaction appears lacking). There is no HI detection either in AGES or the literature.

Figure 6.14: VCC 2095 RGB image from the SDSS, 12' across.

The optical tail of VCC 2000, figure 6.15, was discovered serendipitously after rebinning an INT image by a factor 3. It is also clearly visible in the g, r and i bands of the SDSS after smoothing or rebinning, though not in the standard RGB images. Though VCC 2000 has been extensively studied (243 references in NED) the tail is at approximately 27 B magnitudes arcsec^{-1} and it appears to have remained undetected. The tail stretches about 6' from the center of VCC 2000 with a length along its curve of about 17', 35 Kpc at 17 Mpc distance.

(a) Standard SDSS RGB image of the region

(b) Rebinned SDSS RGB image of the region (pixels are a factor 5 larger than standard)

(c) 2 stacked B-band images from the INT WFS

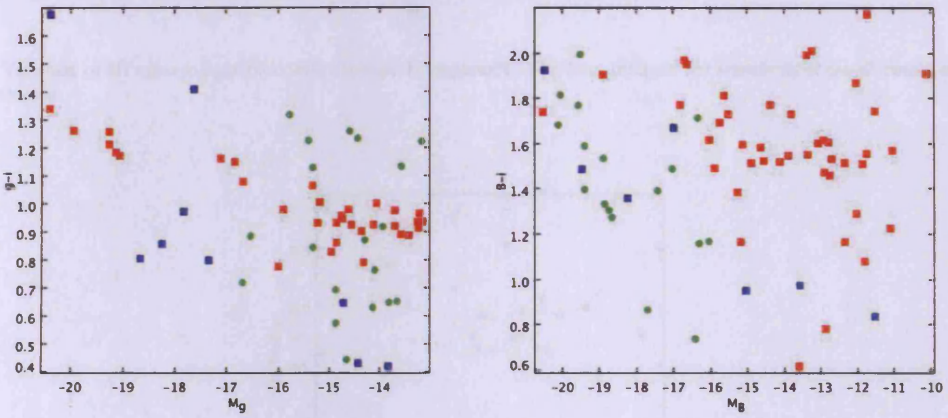
Figure 6.15: Images of VCC 2000 and environs, all 12' across.

It is very difficult to accurately estimate the colour of the tail, since it is barely detected in either the SDSS g band or the INT i band due to fringing. From aperture photometry the B-i colour of the galaxy is estimated to be 1.77, the nearby elliptical VCC 1991 to be 1.71 and the tail at 1.55. In the SDSS rebinned RGB image it appears to have the same colour as the elliptical galaxies. The dwarf galaxy VCC 2002 does not seem to be connected with the tail. This is confirmed by Hubble ACS images from Ferrarese et al. 2006, who note the detection of a possible stripped satellite apparently quite independent of VCC 2002. Unfortunately the ACS image does not include the tail, but it

may be this is a relic of the merger of a (presumably) gas-poor dwarf.

6.4.4 Colours

Colour-magnitude diagrams for the VC2 detections are shown in figure 6.16. The SDSS data comes from an automatic script developed by Petrolli & Gavazzi (provided by Luca Cortese, private communication) which was used to try to reduce the problems associated with the automatic photometry (see chapter 3), but it is likely this is still the reason for the high scatter. The INT WFS data is particularly scattered at the faint end where i-band fringing becomes significant. Overall, while HI detections are probably bluer than non-detections at the same magnitude, the result is nowhere near as clear as the results shown in chapter 5.



(a) SDSS

(b) INT WFS

Figure 6.16: Colour-magnitude diagrams for VC2. The SDSS uses an automatic script; the INT WFS uses surface brightness profiles. Squares are Virgo Cluster members; red for HI non-detections and blue for AGES detections. Green circles are HI detections behind the Virgo Cluster.

6.4.5 HI mass to light ratio

In chapter 5 there was shown to be a clear relation between M_{HI}/L_B ratio and absolute magnitude (see also Gavazzi et al. 2008). Inferring such a relation in the VC2 region is severely hindered by the small number statistics, as shown in figure 6.17. From this figure it may appear that faint ($M_B < 16.0$) objects in the cluster follow a different trend to the other cluster members, being more similar to the background detections. This would leave the 4 brightest objects as a completely separate group that perhaps follow a different trend entirely.

However, figure 5.16 demonstrates that the relation as measured in VC1 has considerable scatter, making inferring any kind of trend from this more limited data set very difficult. What is apparent, as in chapter 5, is that Virgo galaxies have lower M_{HI}/L_B ratios than field objects, particularly so at the high mass end. Whether the background objects really follow a relation with a different slope is hard to determine as they are found in several different populations at different distances (see section 6.5) making sensitivity effects important.

In contrast to VC1 (see figure 5.17), it seems that here the reddest HI-detected galaxies in the cluster *do* have significantly less gas than their background counterparts. While there is at best a very weak offset between background and cluster galaxies in figure 5.17, in figure 6.18 the difference is clear - the reddest cluster HI detections have much lower gas fractions than background objects of the same colour. Yet if it is really the HI content which determines star formation rate, and ultimately colour, galaxies of the same size and gas content should have approximately the same colour regardless of environment. This is indeed the case for VC1, but apparently not for those galaxies in VC2.

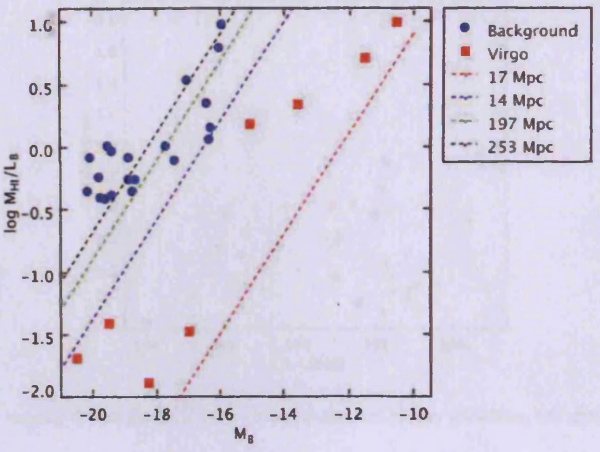


Figure 6.17: Variation of HI mass to light ratio with absolute B magnitude. The lines indicate the sensitivity at the distances of the various populations detected.

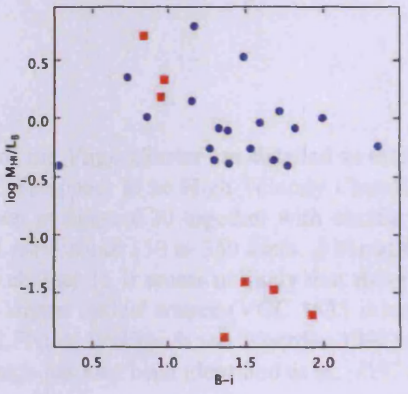


Figure 6.18: Variation of HI mass to light ratio, logarithmic solar units, with $B - i$ colour. Red squares are cluster members, blue circles are background detections.

As with the apparent trends in M_{HI}/L_B ratio, however, great care must be taken because of the low number of detections. Figure 5.17 shows a very strong scatter and as shown in the previous section the colour-magnitude relation is more scattered than in chapter 5. If, however, the difference between cluster and background objects is real, then there are any number of possible explanations. For example, if the gas density in the cluster galaxies were to drop below the supposed threshold for star formation (e.g. Schaye 2004, Kennicutt 1989) before being entirely removed, then further gas loss could not cause further reddening. The effects of environment on the stellar component might also be significant. Such speculations should not be considered further without more observations to give confidence that the difference is real.

6.4.6 HI deficiency

As discussed in chapter 5 the M_{HI}/L_B ratio is closely linked to HI deficiency - how much gas a galaxy has lost compared to a field object of the same morphology and optical diameter (see equations 1.1 and 1.2). There are insufficient objects here to show the distribution of deficiencies as in figure 18 of chapter 5, but the deficiency as a function of R.A. is shown in figure 6.19 with the background objects for comparison.

Again numbers are poor, but as with VC1 there does appear to be a weak radial trend - all 6 detections with deficiency above 0.8 are found in the most westerly 2 degrees of the cube (about 1 Mpc from M87). As described in chapter 5, in VC1 83% of the HI detections are more deficient than 0.3, which is consistent with the fraction

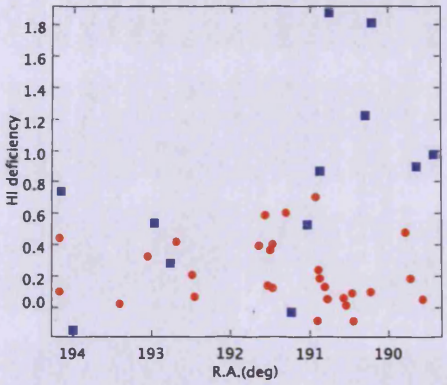


Figure 6.19: HI deficiency varying across the VC2 cube - blue squares are cluster members, red circles are background objects.

for VC2 (75%). Significant deficiencies (> 0.4) are found even at the eastern edge of the cube at about 2 Mpc from M87, indicating that this area does not reach completely beyond the cluster's influence. A few objects in the background are seen to be HI deficient, these are discussed in the next section.

6.5 Other detections

Objects which cannot be members of the Virgo Cluster are detailed in tables 6.3 and 6.4. Not all of these correspond to galaxies. Objects 31 and 32 appear to be High Velocity Clouds associated with the Milky Way. An integrated flux map of these is shown in figure 6.20 together with similar features from VC1. Both clouds in VC2 lie at ~250 km/s, those in VC1 from about 150 to 350 km/s. Although some Virgo cluster galaxies do have velocities much lower than this (see chapter 1), it seems unlikely that these are cluster members. Each source is up to 19' across, far larger than the largest optical source (VCC 1555 is approximately 8.3' across according to GOLDMine). A literature search (ALFALA, Wakker & van Woerden 1991) finds that all these detections are new, except the northern cloud in VC2 which has also been identified as an HVC in ALFALFA.

If the clouds shown in figure 6.20 were at the Virgo Cluster distance of 17 Mpc, the projected length of the clouds would be about 90 Kpc, but none has any optical counterpart. The fact that these are the lowest velocity sources detected, coupled with the fact that they are by far the largest sources in angular extent (well resolved by the Arecibo beam) but with low velocity widths (30 - 100 km/s) makes it much more likely that these are clouds at galactic rather than extragalactic distances. Diffuse objects at substantially higher velocities ($\geq$ 500 km/s) would be far more promising dark galaxy candidates or evidence for gas stripping in the cluster. Still, they do show irrefutably that AGES is capable of detecting significantly extended HI emission (see chapter 7).

Beyond 3,000 km/s there are a remarkably consistent 7.2 detections per square degree for VC2 and 7.3 for VC1. This demonstrates that the source extraction procedures are producing consistent results. Although the S/N of many background detections may seem high (see table 6.3) this is only after follow-up observations have rejected most of the very low S/N detections. Galaxies are detected up to about 18,000 km/s, as shown by the wedge diagrams in figure 6.21.

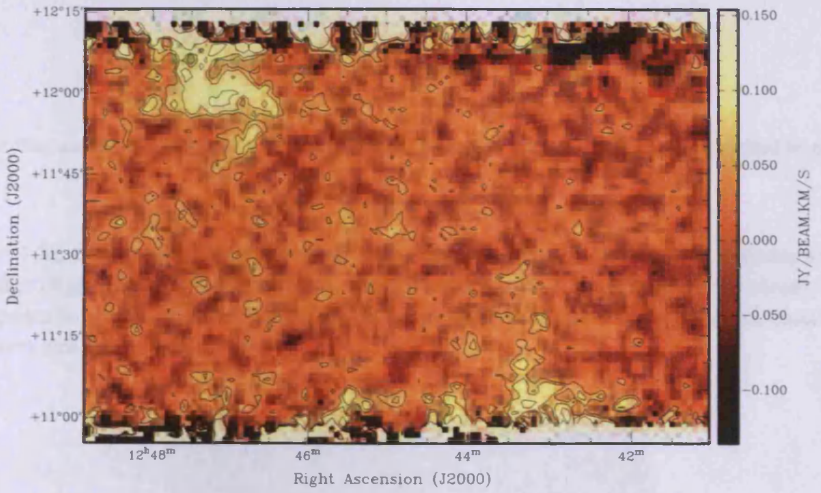
In both cases the gap between the Virgo Cluster and other populations is clear. In VC1 there is a gap of about 1,000 km/s until the next HI detection; in VC2 it is over 4,000 km/s. There is thus little chance that there are any background detections which are actually Virgo Cluster members. As noted in chapter 1, galaxies are present even at redshifts of -1,000 km/s, and Binggeli, Sandage & Tammann 1985 believed that these had been accelerated by the cluster's potential well. While this could mean that, by extension, galaxies at 4,000 km/s might also have been accelerated by the cluster, the blueshifted galaxies are not an isolated population but part of an extension of the cluster - see figures 1.8 and 1.9. Gaps in the velocity distribution of 1,000 km/s are thus significant. While analysis of these field galaxies behind the cluster is restricted by sensitivity effects, several of the background galaxies are, however, very interesting as individual objects.

Table 6.3: HI properties of objects beyond 3,000 km/s, columns as in table 6.1.

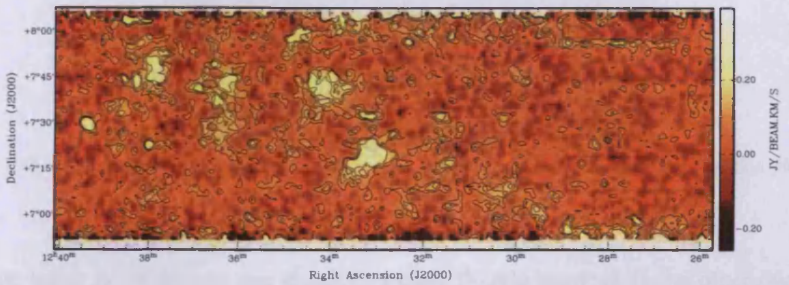
No.	R.A.	Dec.	Velocity	W50	W20	TotFlux	M_{HI}	S/N	RMS	Flag	Optical Flag	Other Names	Mtype	ALFALFA S/N
VC21	12:44:17.80	+11:50:20.0	18167(3)	250(7)	262(10)	0.767(0.107)	10.07	7.29	0.6	0	0	Two	20	17.71
VC22	12:43:49.70	+11:02:12.0	13936(5)	63(9)	149(14)	1.393(0.157)	10.1	17.55	1.1	0	0	None	20	34.82
VC23	12:54:48.00	+12:05:17.0	13896(9)	123(17)	167(26)	0.409(0.102)	9.57	5.37	0.8	0	1	None	20	10.18
VC24	12:45:53.30	+11:52:12.0	14101(11)	141(21)	202(32)	0.218(0.066)	9.31	5.09	0.6	1	0	None	20	6.81
VC25	12:43:33.20	+11:17:13.0	13902(7)	384(14)	472(21)	1.056(0.113)	9.98	9.99	0.6	0	0	None	20	19.98
VC26	12:53:43.20	+11:19:08.0	13890(5)	300(9)	339(14)	1.109(0.115)	10.0	9.97	0.6	0	0	None	20	23.86
VC27	12:38:20.20	+11:43:48.0	13237(4)	109(8)	144(12)	0.895(0.115)	9.87	11.69	0.8	0	0	IC3604	20	22.95
VC28	12:46:09.90	+11:55:15.0	13503(6)	206(12)	263(19)	0.816(0.102)	9.84	8.99	0.6	0	0	IC3756	20	20.88
VC29	12:43:08.90	+11:12:32.0	8362(7)	163(14)	345(21)	1.381(0.106)	9.66	17.35	0.5	2	0	VCC1954	6	47.93
VC210	12:42:11.50	+11:26:58.0	8340(2)	40(5)	64(7)	0.389(0.056)	9.1	18.1	0.5	0	1	None	20	27.51
VC211	12:41:48.60	+11:35:37.0	8345(4)	89(9)	126(13)	0.572(0.081)	9.27	10.84	0.6	0	0	None	20	21.42
VC212	12:43:41.90	+11:41:05.0	8638(9)	73(17)	175(26)	0.283(0.061)	8.99	8.62	0.6	2	0	None	20	11.19
VC213	12:49:57.80	+11:16:00.0	7435(2)	238(5)	260(7)	1.361(0.094)	9.55	17.73	0.4	0	0	VCC2082	18	47.42
VC214	12:41:54.40	+11:46:22.0	7321(2)	178(4)	199(6)	1.701(0.128)	9.63	19.8	0.6	0	0	IC3670	20	45.05
VC215	12:56:45.70	+11:05:31.0	6474(2)	343(4)	357(5)	3.543(0.226)	9.84	19.64	0.9	0	0	UGC08064	5	47.32
VC216	12:45:12.00	+11:33:17.0	6510(4)	209(8)	230(12)	0.552(0.087)	9.04	8.14	0.6	0	0	None	20	13.92
VC231	12:43:22.20	+11:03:56.0	255(11)	36(21)	106(32)	0.186(0.072)	5.75	5.5	0.9	0	2	None	HVC	5.74
VC232	12:47:04.40	+12:00:33.0	255(2)	31(4)	42(6)	0.227(0.049)	5.84	13.26	0.6	0	2	None	HVC	12.20
VC237	12:38:56.10	+11:07:55.0	18046(4)	389(7)	401(11)	0.844(0.127)	10.11	6.52	0.7	0	0	None	20	13.51
VC238	12:38:34.60	+11:37:36.0	15074(5)	213(9)	220(14)	0.177(0.072)	9.27	3.98	0.7	1	1	None	20	3.75
VC239	12:39:10.60	+11:07:22.0	13188(9)	142(17)	174(26)	0.290(0.088)	9.37	4.45	0.7	0	0	None	20	7.72
VC240	12:52:15.50	+11:26:47.0	13462(4)	156(8)	173(13)	0.359(0.067)	9.48	7.15	0.5	0	1	None	20	12.18
VC241	12:49:50.70	+11:26:47.0	17992(10)	336(20)	386(31)	0.530(0.106)	9.9	4.79	0.6	0	0	None	20	10.52
VC242	12:45:55.50	+11:28:35.0	13404(10)	302(21	370(31)	0.458(0.084)	9.59	5.59	0.5	1	0	None	20	11.75
VC243	12:46:32.20	+11:23:57.0	12221(24)	83(47)	231(71)	0.038(0.030)	8.42	3.51	0.5	1	1	None	20	1.83
VC244	12:40:58.20	+11:46:51.0	12901(6)	193(12)	227(18)	0.428(0.075)	9.52	6.77	0.5	1	1	None	20	13.31
VC245	12:46:01.30	+11:47:09.0	11977(12)	250(23)	349(35)	0.439(0.080)	9.47	6.0	0.5	1	0	None	20	12.17
VC247	12:43:16.80	+11:12:35.0	8297(3)	50(7)	126(10)	1.109(0.114)	9.55	38.64	0.6	0	0	VCC1961	7	52.28
VC249	12:50:46.90	+11:07:51.0	7368(6)	115(12)	142(18)	0.277(0.062)	8.85	6.34	0.5	1	1	VCC2085	20	10.77
VC251	12:46:21.10	+11:48:34.0	14024(6)	145(12)	186(19)	0.371(0.074)	9.53	7.28	0.6	1	1	None	20	11.28
VC252	12:43:34.10	+11:47:43.0	12142(7)	40(14)	75(21)	0.137(0.053)	8.98	5.88	0.7	1	1	None	20	6.92
VC255	12:43:40.80	+11:22:58.0	13983(6)	146(12)	177(19)	0.267(0.061)	9.39	6.26	0.5	1	1	None	20	9.68
VC256	12:56:47.20	+11:10:08.0	13860(4)	586(8)	596(13)	0.701(0.126)	9.8	5.3	0.7	1	0	AGC221140	20	9.20
VC257	12:46:37.20	+11:54:02.0	13926(11)	319(23)	371(34)	0.398(0.096)	9.56	4.38	0.6	0	0	Two	20	8.05
VC260	12:52:35.20	+11:46:33.0	11971(11)	94(22)	166(33)	0.220(0.065)	9.17	5.34	0.6	1	1	None	20	7.80
VC261	12:46:16.00	+11:09:46.0	17872(9)	77(17)	130(26)	0.219(0.063)	9.52	5.82	0.6	1	1	None	20	9.30
VC262	12:42:22.10	+12:08:12.0	13643(5)	220(10)	234(15)	0.59(0.147)	9.71	5.11	1.1	0	0	None	20	8.09
VC263	12:56:04.90	+12:09:07.0	1869(10)	32(20)	89(30)	0.218(0.092)	7.55	5.25	1.2	0	1	None	20	7.18

Table 6.4: Optical properties for all non-cluster member detections where INT data is available and only one optical counterpart is visible. M_{HI}/L_B ratios are in $M_\odot/L_\odot$, R75 is in arcseconds.

Source no.	Other Names	Mtype	m_B	m_i	B-i	M_B	M_{HI}/L_B	$R75_B$	HI def.
VC24	None	20	17.74	16.44	1.3	-18.74	0.43	4.82	0.39
VC25	None	20	16.94	14.95	1.99	-19.51	0.99	4.1	0.17
VC27	IC3604	20	16.21	14.54	1.68	-20.13	0.43	4.52	0.04
VC28	IC3756	20	16.61	14.27	2.34	-19.78	0.56	3.44	0.13
VC29	VCC1954	6	15.6	15.32	0.28	-19.74	0.38	4.31	0.52
VC210	None	20	18.89	18.16	0.73	-16.45	2.19	3.61	0.01
VC211	None	20	17.61	16.75	0.86	-17.73	1.0	2.18	-0.09
VC214	IC3670	20	16.19	14.87	1.33	-18.87	0.8	6.17	0.08
VC216	None	20	17.37	15.98	1.39	-17.43	0.77	5.2	0.59
VC237	None	20	16.95	15.14	1.81	-20.07	0.8	4.25	0.18
VC239	None	20	17.68	16.41	1.27	-18.66	0.53	2.76	0.47
VC241	None	20	17.61	16.03	1.59	-19.41	0.91	2.16	0.06
VC242	None	20	16.8	15.03	1.77	-19.57	0.38	3.81	0.19
VC244	None	20	19.25	17.76	1.49	-17.04	3.34	2.78	0.09
VC245	None	20	17.24	15.71	1.53	-18.89	0.54	2.89	0.35
VC247	VCC1961	7	15.92	14.52	1.39	-19.41	0.4	3.82	0.12
VC249	VCC2085	20	18.74	17.59	1.15	-16.32	1.39	2.55	0.41
VC255	None	20	20.4	19.24	1.16	-16.06	6.11	1.81	-0.09



(a) VC2 clouds from 220 to 280 km/s. Contours from 55 mJy/beam increasing in steps of 30 mJy/beam.



(b) VC1 cloud complex from 150 to 350 km/s. Contours are from 75 mJy/beam increasing in steps of 50 mJy/beam.

Figure 6.20: Moment maps of High Velocity Clouds detected by AGES in VC1 and VC2

One obvious difference between VC1 and VC2 from the wedge diagrams is the detection in VC2 of galaxies as distant as 18,000 km/s. The largest of these is VC2 37, which is extremely HI massive at 1.3×10^{10} $M_\odot$. This

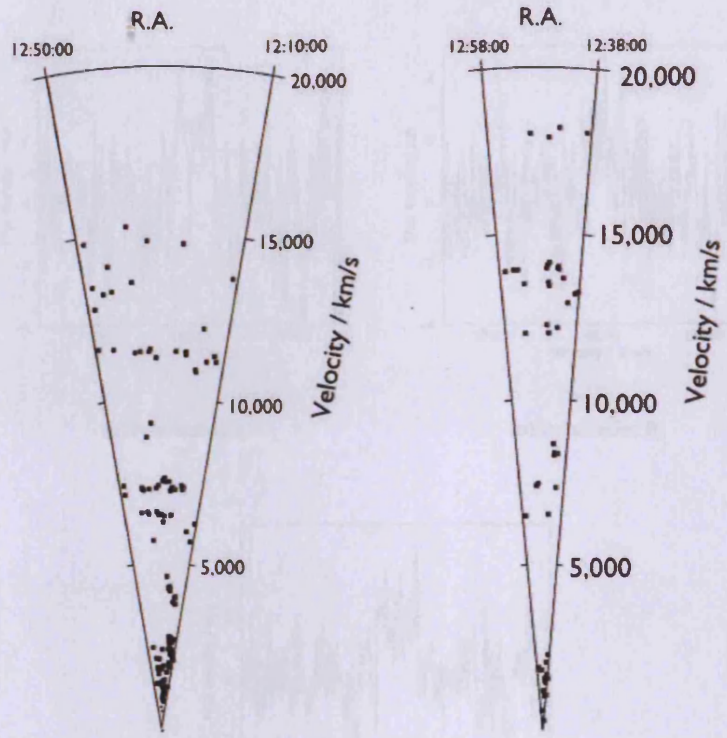


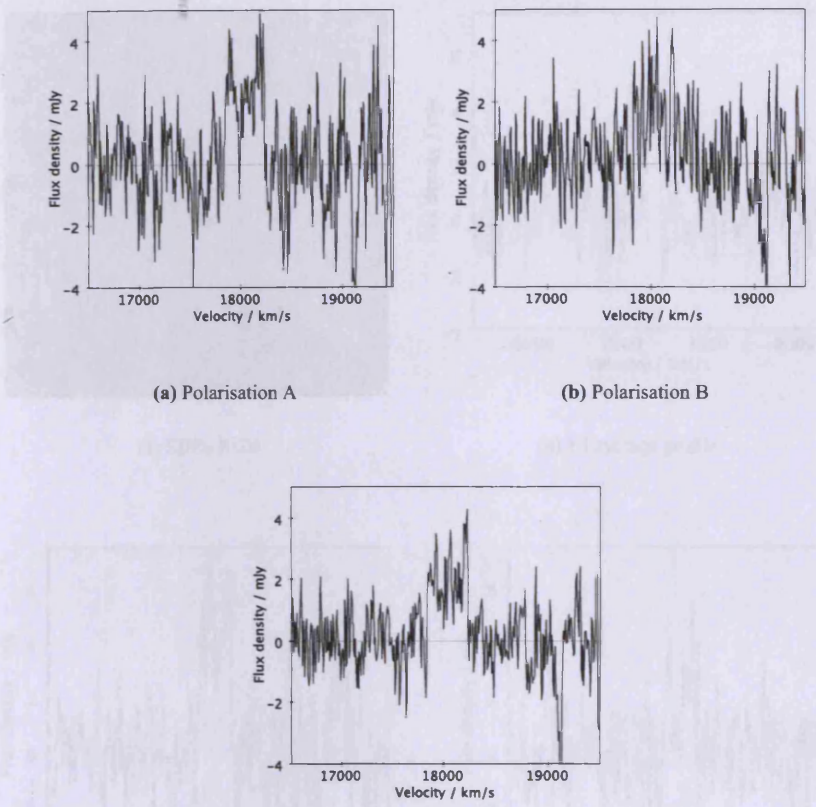
Figure 6.21: Wedge diagrams for VC1 (left) and VC2 (right). The angle each cube subtends has been exaggerated by a factor 2 for clarity.

red object (B - i = 1.8), though spanning only a very small angular size (R_{75} = 4.25") appears to be a disc with a disturbed outer edge (figure 6.22). Its HI profile is rather peculiar - it does not display the classic double horn of a disc, rather, it appears to show 3 distinct peaks (this is in contrast with other low S/N detections such as VC2 57, where a double horn can be discerned).

Figure 6.22: *B* band INT image of detection 37 spanning 6 arcminutes. North is up, east is to the left. Detection 39 is the disc visible to the east.

In the two separate linear polarisations, one displays a clear double horn while the other does not (figure 6.23). This could be the result of noise, although polarisation A, which shows a double horn, actually has a higher rms than polarisation B (1.5 compared to 1.3 mJy). The disturbed morphology and unusual HI profile may perhaps be explained by a merger or interaction, but this is speculative. Other objects are visible within a few arcminutes but redshift measurements are not available to suggest candidate interactors. While this object is at 18,046 km/s, the disc visible to the east in the INT image is detection 39, which is at 13,188 km/s so cannot be an interacting companion.

A similar difference in the polarisations is displayed by object 102 in VC1 (figure 6.24). The SDSS RGB image



(a) Polarisation A

(b) Polarisation B

(c) Average

Figure 6.23: Spectra of source 37, showing a different appearance in different polarisations

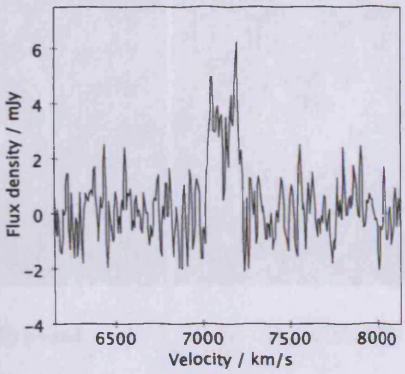
shows only a very faint and compact source that appears inconsistent with the strong, ~200 km/s velocity width, double-horned HI profile. Although both polarisations appear to show the same basic shape, the flux in the A polarisation is 2.3 times that of the B polarisation. The rms of the spectra in each polarisation is the same at 1 mJy, so it cannot be a difference in noise that is causing this. A full understanding of the polarisation of these sources is beyond the scope of the present study, though possibly an instrumental effect - both of these unusual sources are weaker in the B polarisation.

The apparent discrepancy between the optical counterpart and the HI profile is resolved by the deeper INT imaging, which shows a ring (or possibly a disc) like structure approximately 0.75' across (figure 6.25). This is also visible in the SDSS g-band after rebinning by a factor 5 and even in the GALEX FUV despite the poorer spatial resolution. This structure would be 22 Kpc across at the distance of the HI detection, which is much closer to what would be expected for the HI profile.

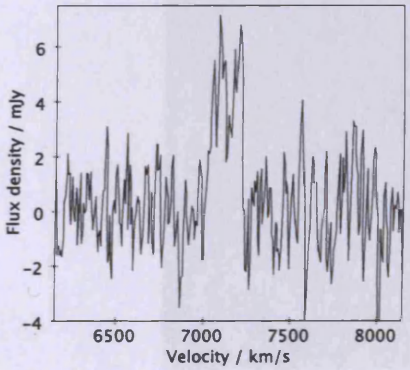
In VC2 a distinct group of 5 galaxies is detected at about 13,000 km/s, all within a radius of 11'. The center of this group is dominated by the cD galaxy VCC 2041 and is shown in figure 6.26 (the other 4 detections include the edge-on spiral visible to the north and 3 other late-type objects outside this image). The galaxy shows optical tails which are just visible in the RGB image but much more clearly seen in the INT (figure 6.27). The circles have radii of 20" and 1.4', the positional offset between the HI coordinates and sure optical counterparts (the median and maximum respectively), centred on the coordinates of AGES detection 57. The optical tails appear to be a new discovery - VCC 2041 has only 9 references and 1 note in NED, none of which mention its disturbed morphology.

The HI detection is quite low signal to noise but given its good positional and velocity match with many possible optical sources it is almost certainly real. The spatial coordinates are midway between the elliptical and an edge-on disc (none of the other galaxies visible are HI-detected). Its average recessional velocity of 13,926 km/s is slightly

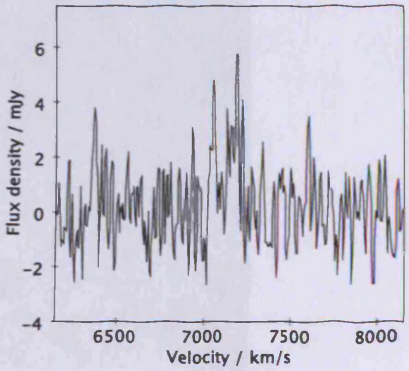
(a) SDSS RGB



(b) HI average profile



(c) HI polarisation A



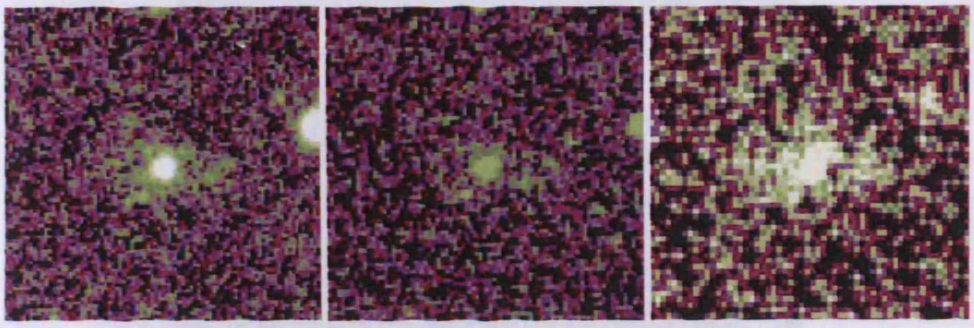
(d) HI polarisation B

Figure 6.24: SDSS RGB image and HI spectra of source 102 in VC1]

closer to the optical redshift measured for the disc (13,980 km/s) than the elliptical (14,040 km/s), though both are well within the 200 km/s required for a sure optical counterpart. In contrast the blue spiral just outside the 1.4' range is more than 1,000 km/s different (12,803 km/s) and may therefore be ruled out as the HI counterpart. The small spiral just to the south of VCC 2041 does not have a redshift measurement. The edge-on disc is the most likely optical counterpart as it displays a double-horn HI profile, typical for discs (but, given the resolution of AGES, it is of course possible that the HI has multiple optical associations).

The northern tail from the elliptical is a narrow, linear feature which extends north-west towards the blue undetected spiral, making a sharp change of direction and then extends in a smooth curve south-west, intersecting the northern spiral. In contrast the southern tail is much shorter and broader, without such a linear shape. The overall appearance is perhaps suggestive of shells, which are believed to indicate recent mergers (Quinn 1984). Given the number of galaxies in the nearby region, it is interesting to try to determine the colour of the tails : are they stripped stars from the elliptical, the undetected spiral, or another galaxy ? This is difficult to answer with certainty as the i-band fringing is quite strong in these images. Fortunately there is another image in which this is far less of an issue, but in this case a CCD gap runs directly through the very center of VCC 2041. Unfortunately the tails are barely visible in the SDSS images and measurements on these would not be accurate.

Aperture photometry was attempted on selected regions of the tails using the relatively infringed i-band images, and the fringed images for the whole galaxy VCC 2041. As the galaxy itself is quite bright the fringing should have relatively little impact. Using this, table 6.5 shows the results for VCC 2041, the tails and the 4 nearest galaxies.



(a) INT WFS B-band (b) SDSS g-band (c) GALEX FUV

Figure 6.25: Optical and UV images of source 102 in VC1

Figure 6.26: Central region of a galaxy group at v ≈ 14,000 km/s, in VC2

Central region of a galaxy group at v ≈ 14,000 km/s. The image is 6.6' across, a projected length of 380 Kpc at that velocity.

Only the southern tail was attempted, as the CCD gap intersects a large part of the northern stream. The edge-on disc was measured using the fringed i-band image as it lies entirely in the CCD gap of the other images.

The tail appears at most marginally bluer than VCC 2041 (as with the tail of VCC 2000), while there is a more substantial difference between the ellipticals and the spirals. If the stars had been stripped entirely from the spiral(s) the tail should be substantially bluer. It is unfortunate that the southern spiral has no redshift measurement as it is quite substantially redder than the northern spiral, almost as red as the tail - although neither tail connects to the southern galaxy. Spatially the sharp change in direction of the northern tail approximately coincides with the northern spiral. Though this galaxy appears not to have been the sole source of the stars in the tail it may have perhaps drawn out the stars from the elliptical tidally or deformed the morphology of an existing tail.

The edge-on disc, though of matching colour, seems to be in the wrong place to create the tails. Since it has an HI detection its red colour is most likely due to internal extinction as it is edge-on. Furthermore, its double-horn profile suggests it is only associated with this galaxy - a contribution from others would not be expected to produce a double-horn shape (since both the other spiral discs are much less inclined in our direction of view). If so then this galaxy is probably undergoing its first interaction with the group and has not yet come close enough to be tidally stripped of its stars or gas.

The edge-on spiral is only barely detectable with a peak signal to noise of 4.4. It is interesting to ask whether the two face-on spirals should be detected in HI given their approximate distance and stellar masses. By assuming a minimum S/N requirement of 4 for a detection, a distance of 200 Mpc (the distance to the elliptical if it is in pure Hubble flow) and a tophat profile for the signal, the HI mass and thus M_{HI}/L_B ratio can be constrained. Additionally the noise is assumed to be the same as for the edge-on detection at 0.6 mJy - this is reasonable since

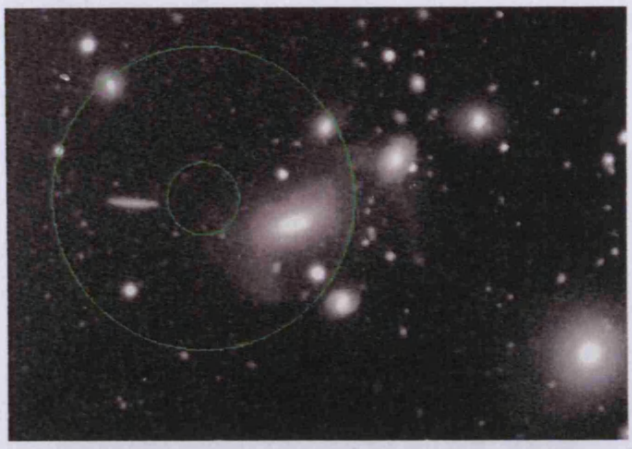


Figure 6.27: B-band INT image of VCC 2041. The circles are centered on the coordinates of an HI detection at 13,926 km/s with radii 20" and 1.4'.

Table 6.5: Colours and apparent magnitudes for objects close to VCC 2041. The objects are labelled in figure 6.26.

Object	RA	Dec	B	i	$B-i$
VCC2041	12:46:33.6	11:53:50	15.48	13.29	2.19
Southern tail	-	-	18.87	16.92	1.95
Northern spiral	12:46:29.8	11:54:23	16.86	15.44	1.42
Southern spiral	12:46:32.0	11:53:04	17.24	15.44	1.8
Edge-on disc	12:46:39.8	11:54:00	18.44	16.38	2.06
Northern elliptical	12:46:32.5	11:54:41	17.63	15.58	2.05

the objects are spatially close. A typical velocity width of 200 km/s is assumed. This gives an HI upper limit of $4.53\times10^9 M_\odot$.

The northern and southern spiral have absolute B magnitude estimates of -19.65 and -19.27 respectively. Using the HI mass limit calculated above places upper limits on the M_{HI}/L_B ratios of 0.4 and 0.6. These are not extreme values (the median for the background sample in VC2 is 0.8, for the edge-on disc here it is 0.6 if all the hydrogen is from the disc alone), therefore, the HI observations are not deep enough to detect these objects even for quite typical M_{HI}/L_B ratios. It is conceivable that these objects are contributing to the HI detection, though other lines of evidence (the close spatial and velocity match and the double-horn shape) suggest the edge-one spiral may be the main contributor.

This a very complex system and it is not easy to explain the disturbed morphology of the elliptical. The tails might be the result of a merger, or merely tidal interactions disturbing or removing its outer stars, or both. Since there are many candidates present it is impossible to distinguish, using the available data, which is responsible - or if the morphology is a result of a combination of interactors, or a galaxy which has since completely merged and is no longer visible.

As mentioned earlier there are several objects in the background which seem to be significantly HI deficient. 3 objects in VC2 and 5 in VC1 are more deficient than 0.5, and 1, object 31 in VC1, as high as 1.2 (see chapter 5). In VC2, object 61 may be interacting with a nearby companion, which would explain the relatively low HI content (which also applies to object 107 in VC1). Object 36 in VC1 (VCC 638) has low surface brightness spiral arms extending as much as 1' from the main disc, so it is difficult to precisely define the optical diameter (see figure 6.28).

For the other background objects of apparently high HI deficiency, the error in the deficiency is likely to be higher in the background since a slight error in the optical diameter can give a significantly different prediction for the expected HI mass. Few of the background detections have any classification more sophisticated than identifying them as spirals or irregulars, since identifying Hubble type is very difficult with lower resolution. As an example, consider a 30 Kpc diameter disc. If this is type Sa, the HI content would be predicted as 9.5 in logarithmic solar masses, but if of type Sc it would be 9.7 - a difference of 0.2 being enough to reclassify a borderline deficient into

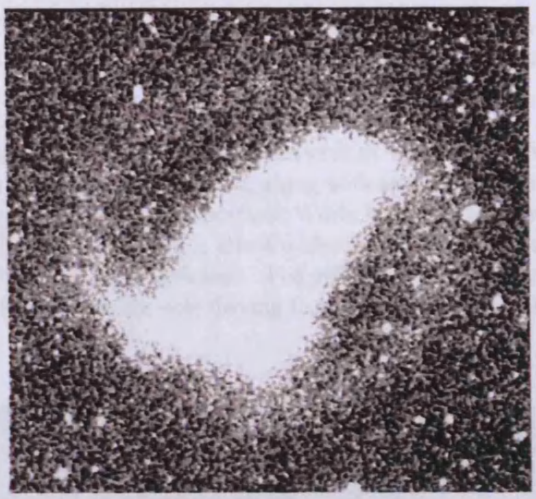


Figure 6.28: SDSS g-band image of object 36 in the AGES VC1 catalogue (see chapter 5), smoothed with a Gaussian kernel of radius 3 pixels. The image is approximately 4.5' across.

a significantly deficient galaxy.

6.6 Summary and Discussion

VC2 is a poorer region than VC1 both in HI and optical. The difference is not so extraordinary that it cannot be explained by normal variance within the cluster, and the number of background HI detections in each region is consistent. Source extraction methods are thus not to blame. Indeed, the HI mass function from AGES shows a greater fraction of objects in the low-mass bins than other shallower surveys, as might be expected for a deeper survey given the problem of incompleteness.

Unlike in VC1, the spatial and velocity distributions of the AGES HI and VCC optically-detected galaxies are clearly different. The HI detections are clustered in declination, whereas the VCC non-detections are uniformly distributed, and tend to be found at higher velocities within the cluster. There is also a suggestion of bimodality in HI deficiency, with all those AGES detections within 1 Mpc of M87 being more deficient than 0.8, and all those beyond it being less. This suggests that subcluster A is still assembling, with the HI detections being recently accreted from the field - perhaps from a single infalling group, since their spatial and velocity distribution is quite different to and distinct from the HI non-detections.

Those galaxies within 1 Mpc of M87, at least, are likely to be on their first pass through the cluster center. The extremely faint object 30 is detected here and is HI deficient, but, as explained in chapter 5, it seems unlikely that such small objects could retain their gas after a single pass through the cluster core. Deficiency in the objects beyond 1 Mpc from M87 is probably less due to the lower density of intracluster gas. If, however, galaxies are still on their first approach into the cluster, low HI-mass galaxies should be detected since some of them will not yet have experienced gas removal. Although far fewer such objects were detected in VC2 than in VC1, this appears to be consistent with the lower number of galaxies present in VC2 overall (HI detections and non-detections alike).

In chapter 5 the detection of several S0 galaxies with low masses of HI was interpreted as evidence for morphological evolution from spirals to lenticulars (i.e. from one Hubble type to another). In VC2 the only detected S0 is VCC 2066, which, as discussed above, is an extremely unusual system which for a proper understanding requires a detailed study in its own right (see Duc et al. 2007). Leaving this exceptional object aside, there is little direct evidence in this region for morphological evolution. This does not affect the results of chapter 5 either way - it is simply that it is very difficult to see any trends at all within a sample of 13 objects.

Though there is no suggestion of galaxies in VC2 changing Hubble type, the influence of the cluster is nonetheless evident in two of the non-detected VCC galaxies. Stellar "brushes" are seen in the outer disc of VCC 2095, and an optical tail is seen around VCC 2000 which may be evidence of a (presumably dry) merger. Additionally, while there are several strongly HI deficient galaxies, these all appear to be quite normal late-types in every other sense.

Indeed, some detections are of much lower M_{HI}/L_B than field objects, but of exactly the same colour. Morphological and/or colour evolution may not be driven solely by gas removal (e.g the harassment scenario of Moore et al. 1999).

Furthermore, there are no examples of undetected late-types such as VCC 1435 in VC1. All of the cluster member late-types listed in the VCC in this region are detected, along with several other late-type galaxies. Whatever the effects of gas loss may be, they are undoubtedly complex. While VCC 1435 is a blue undetected late-type, some much redder late-types were also detected in VC1, albeit with high HI deficiency (such as VCC 534). Even VC2 30, with an exceptionally high M_{HI}/L_B, is deficient. The effects of gas loss may be highly dependent on the individual circumstances, and it cannot be the sole driving force of galaxy evolution (see figure 6.29).

(a) VCC 1435, deficiency > 2.5. **(b)** VCC 534, deficiency 1.8. **(c)** VC2 30, deficiency 0.9

Figure 6.29: Examples of significantly deficient galaxies, showing their widely differing colours and morphologies. Images are all 100" across.

The HI detections within the cluster that are not in the VCC are all generally small, blue, gas-rich objects - not necessarily of low surface brightness, but small and faint. Those in VC2 are all non-HI deficient, except for object 30, but some in VC1 are more deficient than 0.5. This could suggest the presence of an infalling population of even more gas-rich objects but none has been detected (but see below). A caveat is that the HI deficiency relation is not well calibrated for objects other than spirals (Solanes et al. 1996) - though more complex methods, such as measurements of the oxygen abundance, may improve matters (Lee et al. 2003).

There are no dark galaxy candidates in VC2, even in the background. Several low S/N candidates were originally detected but all were found to be spurious on follow-up. The closest object to a dark galaxy is object 30, which has the highest M_{HI}/L_B ratio in the sample at almost 10. Given its HI deficiency, its original M_{HI}/L_B ratio (i.e. before it entered the cluster, and assuming it to have the expected HI mass predicted by equation 1.2) could have been much higher, up to 80 assuming no optical evolution. Even at its current HI mass it is detected at a S/N level above 16, so there is no problem in detecting such objects. Rather, such objects must be very rare - there are only 4 other objects in VC1 and VC2 combined with a M_{HI}/L greater than 2. Reinforcing the conclusion from VC1, if dark galaxies exist they are not at all common in the Virgo Cluster.

It might be argued that a galaxy such as object 30 is the result of gas removal from a more typical dwarf such as those seen in figure 6.12. Simulations by Smith 2009 show that a dwarf irregular can be transformed into a spheroidal. One problem is that the HI deficiency cannot be accurately estimated for dwarfs, so it is uncertain how much gas this object has actually lost. Some of the other non-VCC objects, such as VC1 58 and 70, have equal HI deficiency but are 2.5 magnitudes brighter and are certainly not as red as dwarf red sequence galaxies, as shown in chapter 5. Whether this is because the effects of gas loss vary from object to object or just a result of poor HI deficiency calibration is unknown.

As in VC1, no new extended HI features are detected in the cluster. The known HI complex associated with VCC 2066 and 2062 is detected, as are several HVCs. This means the lack of extended HI detections is not due to data reduction or source extraction procedures. Although it is clear that some Virgo galaxies are significantly HI deficient, and that this varies with clustercentric distance, it remains an open question as to where this HI actually

goes. This problem is discussed in more detail in chapter 7.

While there are key differences between the VC1 and VC2 regions - different infalling clouds and different galaxy densities - in many ways they are also very similar. Both show a mix of highly deficient and non-deficient galaxies in approximately the same proportion. Both show evidence that deficiency varies radially within the cluster. Both indicate that dark galaxies and extended HI features are, at the least, rarities within the cluster environment. Perhaps most importantly, both demonstrate that the effects of gas loss can be far from simple - highly deficient objects may share few other - if any - common features. The key result from this sample is that the effects of gas loss are not always clear, and can be difficult to disentangle from other environmental effects and the varying properties intrinsic to individual objects.

Chapter 7

Results 3 : Extending AGES Through Stacking

7.1 Improving sensitivity through stacking

It is possible to increase the sensitivity of observations at the expense of spatial information, by combining observations of different areas of the sky. This technique is usually referred to as stacking and has been used very successfully at other wavelengths. For example, Zibetti et al. 2005 used SDSS images of galaxy clusters to detect their intracluster light (i.e. from stars that are not bound to any particular individual galaxy), and found that it is strongly correlated with the brightest cluster galaxies. Zheng et al. 2006 used stacking of Spitzer 24μm data to detect objects a factor 10 fainter than the standard detection limit. Stacking has also been previously used to detect HI in galaxies at $z = 0.2$ (Verheijen et al. 2007), a feat very difficult for individual objects (Catinela et al. 2008).

As discussed in chapters 1 and 5, some early-type galaxies do contain HI, though a far smaller fraction than for late-type objects. Given that the vast majority of ETGs are not detected, and the fact that most galaxies in both regions examined in this study are HI deficient, it is natural to ask why any are detected at all. Do most early-types contain some HI below the AGES sensitivity limit, with the detected objects merely representing the high-mass end for early-types, or are most totally devoid of HI, with the few detections present being extremely unusual objects ? Answering this is important in understanding galaxy evolution. If early-types do generally contain some HI, then it is possible that detected and non-detected galaxies alike are basically similar objects. If not, then it is possible that the detected galaxies are fundamentally different objects (perhaps evolved from late-type objects and now in the last stages of losing their gas - see chapter 5), or perhaps have individual reasons why they contain HI but their counterparts do not, such as the tidal capture of gas from other, gas-rich objects.

It has also been shown that most of the HI-detected galaxies are significantly deficient. The environmental mechanisms discussed in chapter 1 generally act to reduce the HI content of galaxies by physically removing the gas into the intragalactic medium, and, as discussed, simulations generally suggest the formation of HI streams. A few examples of these have actually been directly detected, again see chapter 1.

As shown in chapters 5 and 6, however, it is not generally clear where the missing HI has actually gone. If the simulations are correct then why are so few (see chapter 8) HI streams actually detected, when so many galaxies are significantly HI deficient ? Is there an HI equivalent of the faint intracluster light detected by Zheng et al. 2006 (or the stellar streams directly detected by Mihos et al. 2005 as shown in chapter 1) ? Alternatively, does the stripped HI become ionised by the cluster's X-ray gas as Vollmer et al. 2001 suggest, or are the numerous environmental processes at work in the cluster (chapter 1) sufficient to disperse the gas beyond any prospect of detectability ?

By increasing sensitivity, stacking potentially enables the detection of both early-types and, possibly, intracluster HI streams. These two possibilities are explored throughout this chapter. The first relies on stacking spectra selected from the AGES data cubes based on the position of known early-types undetected in HI. The second stacks large areas without any known HI detections to search for low density extended HI features. Both rely on essentially the same basic technique : combining multiple observations to increase sensitivity.

The basic principle of stacking is that if observations of two or more areas are combined by averaging the measurements together, this is equivalent to increasing the observation time. Since sensitivity is proportional to $t^{0.5}$,

where t is the integration time, it follows that the sensitivity from stacking together n spectra increases by a factor $n^{0.5}$. To do this requires that there is some reason to suspect that each spectrum stacked may contain some amount of HI undetectable with the sensitivity of the raw data. In practise, this places an upper limit on n (for example the number of optically-selected galaxies undetected in HI) which in turn limits the maximum sensitivity achievable. It is important to note that "sensitivity" in this context refers only to the rms of the averaged data values.

The loss of spatial information makes this subtly but fundamentally different to the rms of the raw data. The rms of the raw data is literally its sensitivity. However, if the data is averaged then its rms defines an *average* sensitivity limit. To illustrate this, consider the spectrum of a non-detected galaxy in AGES data. For the typical rms this will place an upper limit on the HI mass present of about 2×10^7 $M_\odot$ at the Virgo distance. This literally means that there can be no more than this mass of HI present in the galaxy (for assumed width and signal to noise parameters).

However, suppose we have some collection of non-detected spectra that have been stacked to give an rms equivalent to a mass sensitivity a factor 10 lower. If there is still no detection, it does not mean that the total amount of HI present is no more than 2×10^6 $M_\odot$, as it would for a single spectrum. Rather it means only that the average HI content of every galaxy cannot be more than this. Some individual galaxies, however, might contain more than the apparent mass sensitivity limit, providing a sufficient number of the remainder contain less. This is because the signal present in some objects would be averaged with the noise present in the others, driving the signal to noise down.

The sensitivity limit from stacking is therefore a limit only on the average HI content in each spectrum. This can be converted into a true total mass limit. Suppose we have a collection of n spectra all of rms σ, all non-detections of galaxies at distance d. To place a limit on the HI mass present assume the galaxies have velocity width W_{50} and signal to noise ratio S. From the raw data the total HI mass that could be present is simply the number of spectra multiplied by the mass limit per spectrum :

$$M_{HI} = n \times 2.36 \times 10^5 \times d^2 \times W_{50} \times S{\times}\sigma$$

For the stacked n spectra the noise will be reduced to $\sigma.n^{-0.5}$. The total mass limit will be the number of spectra multiplied by the average mass limit (as this is equivalent to the mass limit per spectra), so :

$$M_{HI} = n^{0.5} \times 2.36 \times 10^5 \times d^2 \times W_{50} \times S{\times}\sigma$$

While this is a true constraint on the total mass (were any more HI present, the average mass would have to be greater than the average mass sensitivity), it is apparent that this may be of limited use. For example the mean HI mass for Virgo cluster member galaxies from VC1, for instance, is 7.9×10^7 $M_\odot$. This on its own gives no indication that some objects may have masses as low as 4×10^6 $M_\odot$ while others are seven hundred times more massive. As will be shown, this great variation is one of many issues that mean stacking is by no means guaranteed to succeed in detecting hydrogen even when present, both when trying to detect intracluster HI but more particularly in the attempt to detect early-type galaxies.

7.2 Early-type galaxies in the Virgo Cluster

It is relatively straightforward to stack the spectra of undetected early-type galaxies. "Early-types" here refer to galaxies of morphological type less than 2 as defined in the GOLDMine database, i.e. ellipticals, spheroidals, lenticulars and their dwarf counterparts (see chapter 3). The optical redshift of the galaxy must be available, so that the extracted spectra are all centred on the galaxy. In this way, when the spectra are averaged together, any signal present in each galaxy will be averaged with signal present in the other galaxies. A baseline of 800 km/s devoid of any detections is required, to ensure baseline ripple is not counted as a detection, and also to ensure any detection is clearly visible (for example objects too close in velocity to the Milky Way are rejected, to avoid its very strong signal contaminating the other spectra).

Spectra are extracted from the cubes with a first-order polynomial fit to reduce baseline ripple. Of the 59 early-type VCC galaxies in the VC2 region, there are 15 with optical redshift measurements that satisfy the above criteria. In VC1 there are 112 undetected early-type galaxies of which 37 are suitable for stacking. Spectra of the non-detections are examined individually to ensure there is no contamination from other sources at similar positions and redshifts, nor any RFI or baseline depressions caused by over-subtraction due to bright galaxies (a consequence

of data reduction - see chapter 2), or the Milky Way.

Although each spectrum is centred in velocity on the galaxy, it is still possible signal is being averaged with noise in some cases, due to the different morphologies (and inclination angles of lenticulars). It has been shown in chapter 5 that lenticulars and dwarf ellipticals greatly differ in their HI contents. We might expect, for example, that an edge-on lenticular galaxy would have its HI in a rotating disc and display a wide and possibly double-horned profile, whereas the profile of an elliptical (dispersion-dominated) might be closer to a narrow Gaussian. Depending on the velocity widths of the galaxies it is possible that the horns of the lenticular profile lie outside the Gaussian of the elliptical. For this reason stacking is attempted not only on all early-type galaxies but also on lenticulars and ellipticals separately, though obviously at the expense of sensitivity.

From B-band photometry available in GOLDMine, the galaxies stacked here span a luminosity range of 8.86 magnitudes. If the M_{HI}/L_G trend for the detected galaxies holds (see chapter 5), this indicates a possible variation of a factor 60 in HI mass - again it is not certain that signal is always being averaged with signal. The number of galaxies available means it is not worthwhile to stack only objects of similar luminosity.

The resulting spectrum from stacking the early-type galaxies are shown in figure 7.1. Galaxies from the two regions are stacked separately and combined, and stacking is also performed by morphology. No detection results by any method. Before commenting on the significance or otherwise of this, it is necessary to understand if the sensitivity is increasing as expected or if the rms is not reaching a substantially lower level. This is useful as it is possible to choose which spectra to stack so as to minimise the rms and/or the actual mass sensitivity, which does not necessarily involve stacking the maximum number of objects.

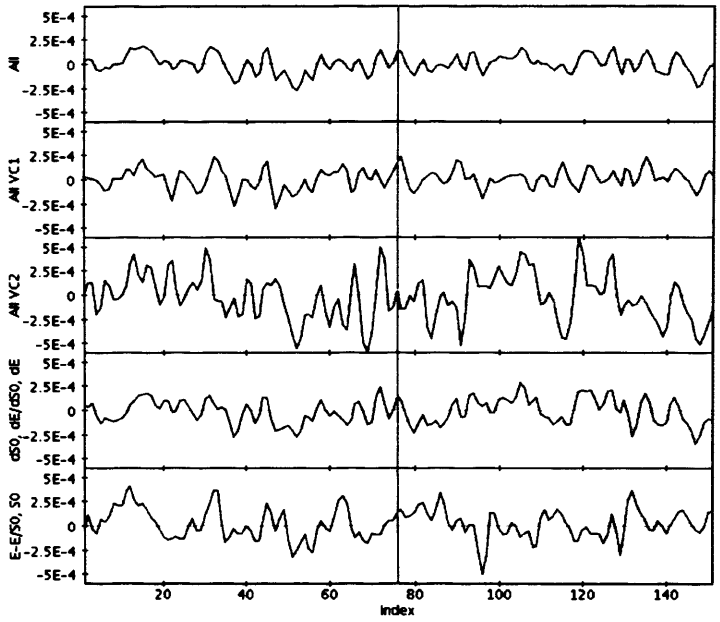


Figure 7.1: Stacked spectra of early-type galaxies with the same vertical scale for each. The vertical line indicates the expected position of any detection. From top to bottom : All 52 galaxies stacked, with an rms of 0.097 mJy; all 37 galaxies from the VC1 region stacked with an rms of 0.101 mJy; all 15 from VC2 with an rms of 0.245 mJy; all 36 galaxies from both regions of morphological type dS0 - dE, with an rms of 0.129 mJy; all 16 galaxies from both regions of morphological type E - S0, with an rms of 0.157 mJy.

7.2.1 Sensitivity

In figure 7.2 the rms is plotted as a function of the number of spectra stacked. The expectation is that that the $rms = rms_1 \times n^{-0.5}$, where rms_1 is the rms of a single spectrum. Therefore :

$$log(rms) = log(rms_1) - 0.5 \times log(n) \tag{7.1}$$

And so a log-log plot for rms as a function of n should have a slope of -0.5 In the figure the value for rms_1 used

is the mean standard deviation for all 52 spectra, 0.608 mJy. A linear regression fit to the data from Topcat gives a slope of -0.533, in approximate agreement with expectation. It is apparent from the graph that the rms does not always decrease at the same rate with increasing n and sometimes even increases. This is because in the ideal case each spectrum would have the same rms whereas in fact some have substantially higher, while some are lower than the average. While the theoretical rms that could be achieved from stacking 52 spectra is 0.084 mJy, the actual value is 0.097 mJy. This is only slightly higher than the theoretical, indicating that stacking is basically performing as expected.

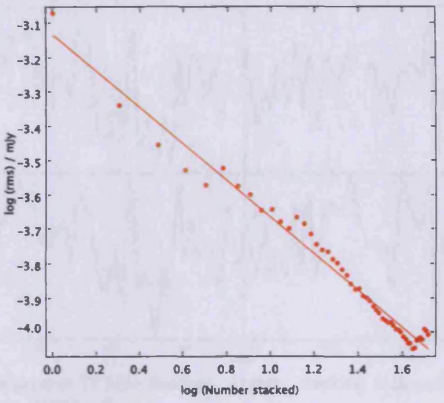


Figure 7.2: The noise level attained by stacking as a function of the number of spectra stacked. The solid line is a linear regression fit to the data, slope -0.533.

It is possible to improve on the sensitivity attained by rejecting those spectra with particularly high rms. Alternatively, when adding the spectra for averaging their contribution can be weighted by rms such that those with the highest rms contribute less - this approach has the advantage of not disregarding any spectra. The results of both methods are shown in figure 7.3. In the first method the worst 5 (10%) spectra are rejected, improving the averaged rms to 0.089 mJy. This is in agreement with the theoretical expectation for stacking 47 spectra. In the second each spectrum is weighted by the inverse of its rms. The spectrum with the lowest rms is weighted to 1.0 and spectra with double this rms are weighted to 0.5. In this way the noise is reduced to 0.078 mJy. That this is lower than the theoretical 0.084 mJy stated above as that figure does not account for weighting the mean rms of the individual spectra, which will be 0.512 mJy, rather than the raw 0.608 mJy (this reduces the theoretical minimum to 0.071 mJy). In any case neither removing the worst spectra nor weighting by noise results in a detection, though both methods improve the rms reached.

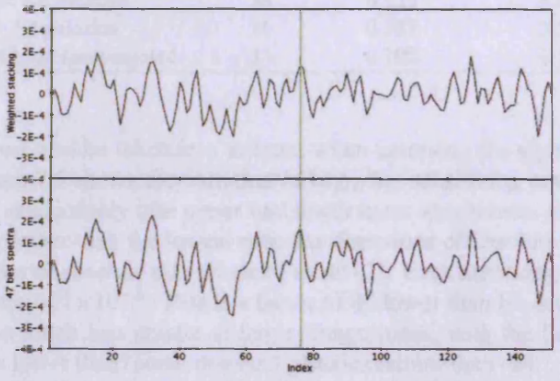


Figure 7.3: Stacked spectra of early-type galaxies. Above : stacking with the worst 5 spectra rejected, rms of 0.089 mJy. Below : stacking by weight according to rms, rms of 0.078 mJy.

7.2.2 Significance of the non-detections

The importance of the non-detections by the stacking procedures depends upon the mass sensitivity achieved. As well as reducing the rms the detectable mass may be reduced by selecting only the closest galaxies in the sample. Galaxies in the Virgo regions considered here vary in distance from 17-32 Mpc (see chapter 3). Since mass sensitivity is a function of d^2, this equates to mass sensitivity variation of a factor 3.5 (with the caveat that the exact range of distance is unknown). In figure 7.4 only the 35 galaxies at a distance of 17 Mpc are considered, both with and without weighting by noise. Weighting causes only a minor improvement here, reducing the rms from 0.130 to 0.100 mJy. Again, no detection results.

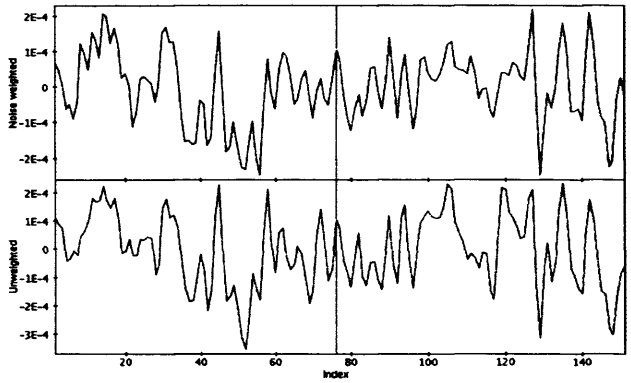


Figure 7.4: Stacked spectra of early-type galaxies at 17 Mpc distance. Above : stacking with each spectra weighted by noise, rms 0.010 mJy. Below : stacked spectra with no weighting, rms 0.013 mJy.

In table 7.1 the rms and mass sensitivity are shown for each procedure described above. The distance used is the mean distance of the galaxies stacked in each case. The velocity width used is, arbitrarily, 50 km/s (since sensitivity is directly proportional to width this is not likely to alter the mass sensitivity by more than a factor of a few) and a tophat function is assumed for a signal to noise level of 5. It is clear that all methods used produce a mass sensitivity of the same order of magnitude and that no approach has resulted in any kind of detection. For comparison, the mean mass of HI in the early-type galaxies which are detected is 5.4×10^7 $M_\odot$, 30 times greater than the lowest limit from stacking. This would seem to indicate that those early-type galaxies which are detected are weird objects with significantly greater HI masses than their non-detected counterparts (but see below).

Table 7.1: Mass sensitivity limits for the stacking procedures attempted on early-type galaxies

Objects stacked	No. objects	rms (mJy)	Mass sensitivity ($M_\odot$)
All galaxies	52	0.097	2.13E6
All VC1 galaxies	37	0.101	2.44E6
All VC2 galaxies	15	0.245	4.18E6
All galaxies noise-weighted	52	0.078	1.72E6
All dS0 - dE galaxies	36	0.129	3.08E6
All E - S0 galaxies	16	0.157	3.50E6
Closest galaxies noise-weighted	35	0.100	1.71E6

The HI mass-to-light ratio must also be taken into account when assessing the significance of the mass sensitivity attained by stacking. Figure 7.5 shows the variation in M_{HI}/L_G sensitivity, assuming a constant HI mass of 4.18×10^6 and 1.71×10^6 $M_\odot$ respectively (the upper and lower mass sensitivities obtained by stacking). In both cases the ratio is significantly lower than the lowest ratio for detections of absolute magnitude brighter than -18. The brightest galaxy stacked has an absolute magnitude of about -21. Even assuming an HI mass of 4.18×10^6 $M_\odot$, this still implies a M_{HI}/L_g ratio of 1×10^{-3}. This is a factor of 80 lower than HI detected galaxies of comparable magnitudes. The difference is much less drastic at fainter magnitudes, with the limits from stacking becoming comparable to (though always lower than) some detected galaxies fainter than -14.

The absolute magnitude of -21.0 is not, however, typical of the stacked galaxies, as nearly 80% of the sample are fainter than -17.0. At this magnitude the sensitivity achieved by stacking is equivalent to M_{HI}/L_G ratios about 5 times lower than galaxies actually detected. As discussed in chapter 5, galaxies of different morphologies

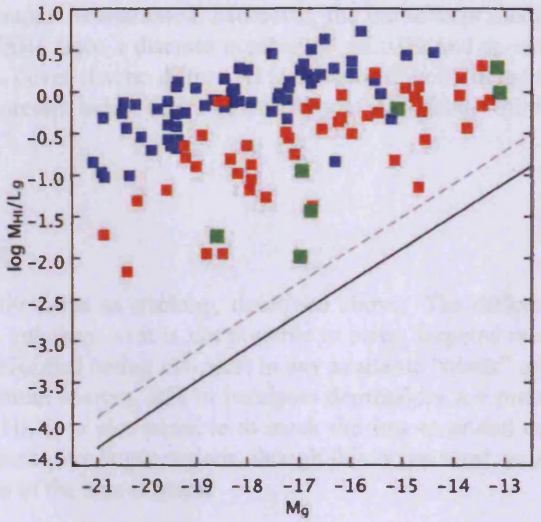


Figure 7.5: Variation in the M_{HI}/L_g ratio (logarithmic axis) as a function of absolute magnitude. Background detections are in blue, cluster members in red with early-types in green. The dashed and solid lines assume a constant HI mass of 4.18×10^6 and 1.71×10^6 $M_\odot$ respectively, the upper and lower mass sensitivities obtained by stacking.

have different amounts of HI and S0s are detected far more frequently than dEs. Also, all the HI-detected S0s are classed as type 1 - exclusively lenticulars, rather than types 0 and 1 (used here for stacking) which also included ellipticals (see table 3.1). Unfortunately only 5 type 1 galaxies are available for stacking here - while these might seem more likely to contain HI, the small numbers prevents a significant gain in sensitivity through stacking.

As mentioned in chapter 5 the conclusion from stacking agrees with that from Conselice et al. 2003 - deeper observations will not uncover more early-type galaxies with neutral gas, at least for ellipticals. This implies that those that are detected are very unusual, perhaps recently accreted field objects as suggested in chapter 5. It does not, however, by any means rule out cluster dwarf ellipticals as being in general morphologically evolved from later-type galaxies. Dwarf ellipticals appear to be much more common in the cluster than the field (Sabatini et al. 2003, Davies, Roberts & Sabatini 2005) and share many properties with late-type dwarfs (Lisker et al. 2006, Boselli et al. 2008). The point here is only that the presence of HI in dwarf ellipticals cannot by itself be used to imply morphological evolution in these objects (see chapter 5 and Conselice et al. 2003). This is discussed further in chapter 8.

7.3 Intracluster Hydrogen

All bar two of the Virgo Cluster member galaxies detected - with the exception of early-types for which it is not possible to estimate - have positive HI deficiencies. Environmental mechanisms are believed to act to remove gas from galaxies. Other surveys within Virgo have detected a variety of HI streams and clouds (see chapter 1) whereas AGES, to date, has detected no such (new) features. The fate of the stripped gas remain relatively unknown - are the streams of gas persistent, or are they rapidly dispersed by tidal interactions or ionised by UV background radiation ? The total amount of gas lost is not insubstantial - for VC2 it is $\sim 1 \times 10^{10}$ $M_\odot$ while for the richer VC1 region it is 9 times greater. However if sufficiently dispersed it could easily fall below the AGES sensitivity limit. In this section two procedures are described to attempt to detect any intracluster HI without knowing its distribution.

The HI lost to the intracluster medium must retain sufficient density to avoid ionisation. Assuming the missing HI estimated from deficiency is spread uniformly over each area, the column densities will be 2.6×10^{18} cm^{-2} and 1.0×10^{19} cm^{-2} for VC2 and VC1 respectively (making the simplistic assumption that all the missing HI in VC1 is at 17 Mpc distance). Column densities are of little relevance, however. In a typical spiral galaxy the HI is distributed in a thin disc; in the intracluster medium it may span at least as much depth as breadth across the sky. Thus the column density cannot indicate whether the HI is dense enough to shield itself from ionisation or not.

The volume densities are more difficult to quantify - as there are not individual distance measurements available, the exact volume of the cube cannot be estimated. Moreover, the intracluster medium is not likely to be uniform. The intracluster HI must originate from a discrete number of galaxies and so may be expected to be clumped, possibly distributed in streams. Given that no diffuse HI is detected directly in the raw data cube, it seems entirely possible that there is some HI present below the raw sensitivity of AGES but sufficiently dense as to avoid ionisation.

7.3.1 Spatial averaging

This procedure is essentially the same as stacking, described above. The difference is that the spatial/velocity range of the intracluster HI is unknown so it is not possible to select targeted areas where HI might be present. Instead spatial averaging is performed (using *mbspect*) in any available "blank" areas of the data cubes where no contaminating galaxies, continuum sources, RFI or bandpass depressions are present, in order to search for very diffuse, extended intracluster HI. It is also possible to mask the less extended contaminating sources (galaxies and continuum sources) and stack over larger regions, though this is not ideal, as *mbspect* occasionally generates spurious detections if too much of the area is blank.

The search for extended features within the cluster is not new - as already mentioned Mihos et al. 2005 and Zibetti et al. 2005 have considered the intracluster optical light, which appears to be a product of tidal interactions between the brightest cluster galaxy and other cluster members. Davies, Roberts & Sabatini 2005 describe a method to search for low surface brightness optical tails that are expected to be produced by the harassment process (see chapter 1 and Moore et al. 1996). Their technique is to smooth optical images using a filter which avoids smoothing the light from point sources (foreground stars). In (apparently) empty regions of HI data cubes there is no problem with point sources but there is of course the complication of the 3-dimensional nature of the data, which is discussed below.

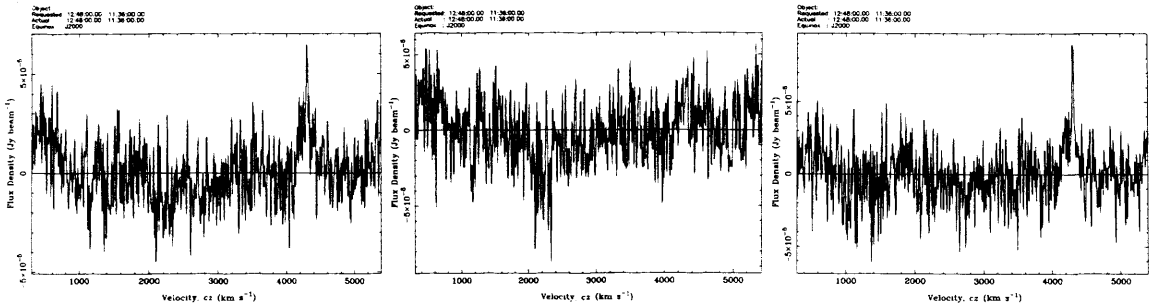
The most obvious method is simply to average over the largest area possible and gain the most sensitivity, equivalent to stacking with the largest n (it is far less straightforward, given the number of spectra being stacked here, to weight each one by its rms). The result of doing this for the whole area of VC2 is shown in figure 7.6. Here the more obvious sources of contamination (galaxies, continuum sources, and the edges of the cube, which are noisy due to ALFA's hexagonal beam pattern) have been masked. Baseline depressions have not been masked as these span the full range of right ascension.

The result is noteworthy for several reasons. Firstly the rms has indeed been significantly lowered to 0.016 mJy, so stacking continues to give sensitivity increases for very large n. Secondly, something strongly resembling a detection is seen at $\approx$ 4,200 km/s. However, this apparent detection is very much weaker (indeed almost invisible) in one polarisation, making it unlikely to be HI. Additionally its redshift is far too high to be in the cluster, equivalent to a distance of 60 Mpc, where there is no reason to expect a diffuse HI medium. The polarisation of the source means it is most likely to originate from weak RFI. Nonetheless, it does show that stacking in this way can detect sources too faint to detect in individual spectra.

Finally, the baseline is far from perfect. Several factors may contribute to this. There may be relatively weak continuum sources present that have escaped masking, and similarly RFI. Continuum sources cause baseline ripple of broadly similar shape (i.e. peaks and troughs occur at approximately the same frequencies), thus, averaging several together will cause an increase in baseline ripple. More fundamentally, galaxies have been masked by the volume they occupy in the cube. This means that at any given channel the number of unmasked spectra present is not constant, because the number of galaxies present varies, and so the number of pixels to average varies as a function of channel.

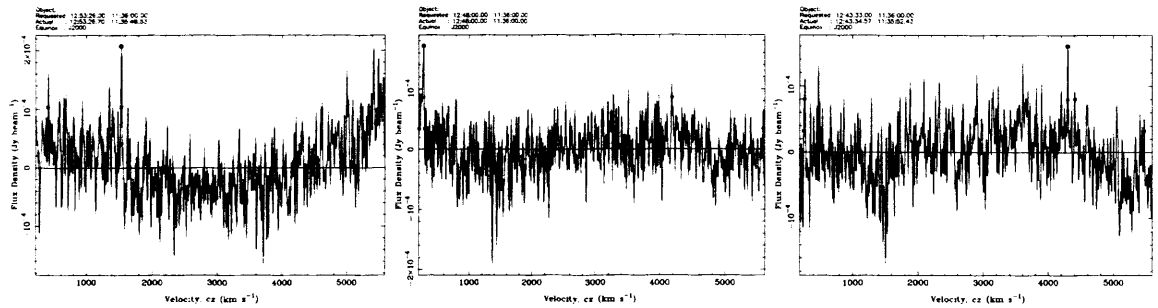
As the distribution of the intracluster HI is unknown, stacking was also performed on four smaller areas within the cube, in case the density of the HI varies significantly. The first 3 areas divide the cube into thirds in right ascension; since the cube spans a significant fraction of the cluster in right ascension and galaxy density appears higher to the west (see chapter 6), it may be expected that there is some variation in the intracluster medium also in this direction. The fourth spans a narrow range of declination containing most of the HI detected galaxies; since their density appears higher to the north it is possible that the intracluster HI is also denser in the northern part of the cube. None of these areas show any sign of a detection (see figure 7.7). The problem of baseline variation is rather more obvious in these smaller areas, where the masked objects span a greater percentage of the stacked area.

Figure 7.6: Spatially averaged spectra of the entire VC2 cube, with galaxies and continuum sources masked.

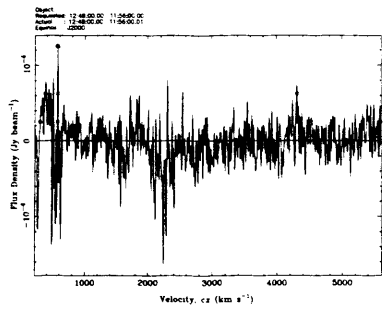


(a) Averaged polarisations (b) Polarisation A (c) Polarisation B

Figure 7.7: Spatially averaged spectra of selected regions of the VC2 cube.



(a) Eastern area RA 12:58:52 - 12:51:41 (b) Central area RA 12:51:41 - 12:44:32 (c) Western area RA 12:44:32 - 12:37:11



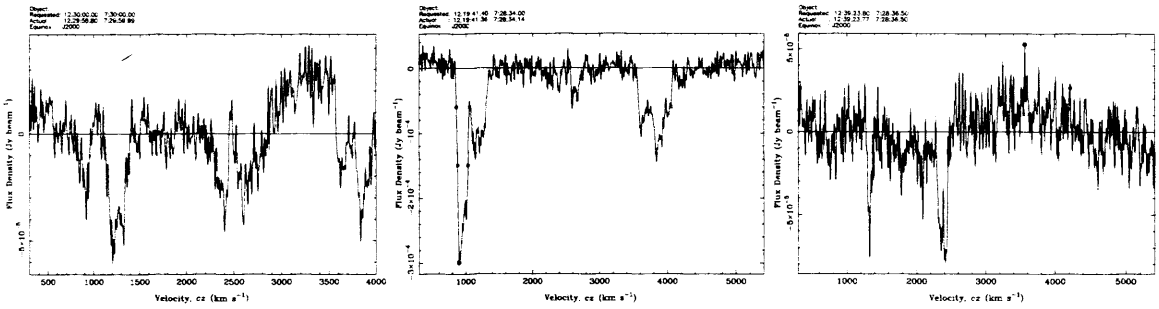
(d) Whole RA, north of 11:48:00

The sensitivity for averaging the whole area of ≈ 0.016 mJy is equivalent to just 2.2×10^5 $M_\odot$ for a 4σ, 50 km/s velocity width tophat signal at 17 Mpc. This is of course a limit on the average HI mass present per spectrum. Given the masking of the edges of the cube, the total number of pixels stacked is approximately 17,110 (an upper limit not accounting for the masking of the continuum sources and galaxies) which equates to 1,779 independent spectra (since each pixel is 1×1' but the beam has a diameter of 3.5'), so the limit on the total amount of HI is 3.8×10^8 $M_\odot$, for the same parameters as above. If the HI spans the full velocity range of the cluster ($\approx$ 3,000 km/s) rather than a mere 50 km/s, the limit is a rather less significant 2.3×10^{10} $M_\odot$.

The HI content limit from stacking is unfortunately not significantly different from the expected estimate from the HI deficiency. It is not obvious as to the exact velocity range the intracluster medium will span, which can affect the sensitivity limit by a factor of a few. Similarly a 5σ S/N limit may be rather high. The estimate from HI deficiency itself is of course only a ballpark estimate at best - gas may flow freely in and out of the region observed, and there it no guarantee it was all removed from the galaxies within this region originally. Thus it is not clear whether this result has any astrophysical significance or is hampered by the sensitivity of the survey.

The VC1 region spans a larger area and has a higher total HI deficiency by a factor 9.1, which, since the same sensitivity should be achievable, means it should be better suited to the stacking procedure. Unfortunately, while the greater number of galaxies present means the total deficiency in the region is greater, it also means a greater fraction of the cube has to be masked. This results in extremely poor baselines, as shown in figure 7.8, which make it very difficult to comment on the possibility of a detection. The eastern area has less masking and so a better baseline, and it is perhaps valid to comment that there is no sign of a detection in this area of the cluster. The noise level reached is 0.020 mJy, corresponding to a total of 2.4×10^{10} $M_\odot$ at 17 Mpc for a 3,000 km/s velocity width, 4σ detection. Again this is only to the same order of magnitude as the calculated missing HI in this area of 3.8×10^{10} $M_\odot$, so cannot be said to be a decisive result either way.

Figure 7.8: Spatially averaged spectra of the VC1 cube, with galaxies and continuum sources masked.



(a) Whole area **(b)** Western half **(c)** Eastern half

7.3.2 Sensitivity

The rms of a small number of stacked spectra has already been shown in figure 7.2. Over this small number of spectra the noise overall appears to agree with expectations. In order for stacking of the intracluster medium to succeed, noise must continue to decrease for much larger numbers of spectra stacked there will be no benefit in stacking. This can be tested within the blank areas (regions apparently devoid of HI or continuum sources) available by averaging over many different sized areas, the result of which is shown in figure 7.9.

It is apparent that there is good agreement with expectation, the linear regression fit to the data giving a slope of -0.44 compared to the ideal of -0.5. The actual rms is never more than a factor of 1.75 greater than the ideal, so the sensitivity achieved is about as good as can be expected. From the previous section, this indicates that even if the noise was reduced to the theoretical minimum, it would still not be sufficiently low as to be a decisive indicator as to the presence (or otherwise) of intracluster hydrogen.

The column density sensitivity reached can be calculated by considering the total mass sensitivity and the size of the area. Assuming that the smallest detectable flux corresponds to a 4σ, 50 km/s velocity width peak, the total mass sensitivity will be :

$$M_{HI} = \frac{N_p}{9.62} \times 2.36 \times 10^5 \times d^2 \times 50 \times 4 \times rms$$

Where N_P corresponds to the number of pixels being averaged and the factor 9.62 converts to the number of independent spectra. Assuming a distance of 17 Mpc and converting this mass into the number of atoms gives :

$$M_{HI} = 1.698 \times 10^{66} \times N_p \times rms$$

The area being stacked is equivalent to the number of pixels, which can be converted to square centimetres by the following (again assuming a distance of 17 Mpc) :

$$Area = (\frac{\sqrt{N_p}}{360} \times \frac{1}{60} \times 2 \times \pi \times 17 \times 10^6 \times 3.09 \times 10^{18})^2$$

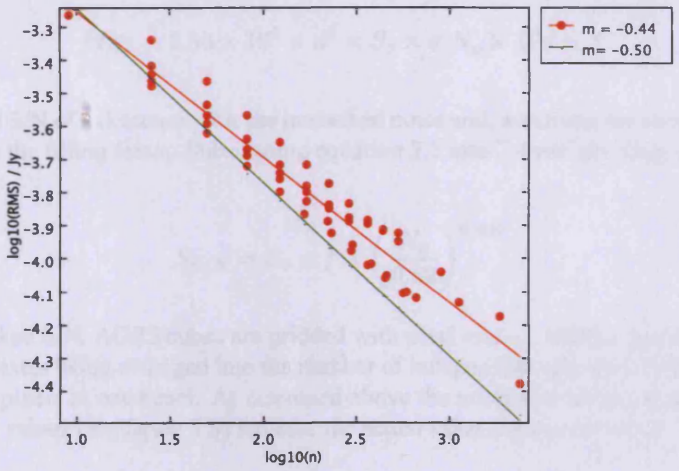


Figure 7.9: Graph to show how stacking reduces the noise level over a large number of spectra stacked.

This simplifies to give :

$$Area = 2.37 \times 10^{44} \times N_p$$

Dividing the total mass sensitivity by the area gives the column density sensitivity, C_s :

$$C_s = 7.29 \times 10^{21} \times rms \tag{7.2}$$

Alternatively, using equation 7.1 to predict the rms, and using a slope of -0.44 as this gives better agreement with the measured values (as opposed to the ideal of -0.5), it can be shown that :

$$C_s = 9.87 \times 10^{18} \times N_p^{-0.44} \tag{7.3}$$

From the HI deficiency estimates, the column density of the intracluster HI in VC1 and VC2 is expected to be 1.1×10^{19} atoms cm^{-2} and $\approx 2.9 \times 10^{18}$ respectively. From equation 7.2, using the lowest rms reached in the blank areas, the column density sensitivity limits reached are 4.0×10^{17} and 4.5×10^{17} atoms cm^{-2} respectively. At face value this indicates the observations are sensitive enough to determine that the intracluster medium is not as dense as expected. However, as discussed in the next section, there are various contributing factors that mean it is difficult to be certain if this is actually the case.

7.3.3 Effects of HI distribution

While the sensitivity achieved places limits on the amount of HI that can *potentially* be detected, it in itself does not guarantee that a detection will actually be made. Both the spatial and velocity distribution of the intracluster HI will affect its detectability (i.e. S/N) through spatial stacking. As a first step it is useful to know how stacking will alter the S/N of a collection of spectra and noise in one channel. Stacking is simply averaging, so the stacked S/N is simply the average intensity (flux) divided by the averaged noise. The average intensity of a source will be given by :

$$I = \frac{M_{HI}}{2.36 \times 10^5 \times d^2 \times N_p \times \Delta V} \tag{7.4}$$

Where Np is the number of pixels being stacked and ΔV is the velocity width of one channel. M_{HI} is the total HI

mass in one channel and is given by :

$$M_{HI} = 2.36 \times 10^5 \times d^2 \times S_o \times \sigma.N_p \times \Delta V \times f \tag{7.5}$$

Where S_o is the original S/N of a detection, σ is the unstacked noise and, assuming for simplicity that all detections have the same S/N, f is the filling factor. Substituting equation 7.5 into 7.4 and dividing by the stacked noise (see equation 7.1) produces :

$$S_{S/N} = S_o \times f \times \left(\frac{N_p}{9.62}\right)^{0.440} \tag{7.6}$$

Where $S_{S/N}$ is the stacked S/N. AGES cubes are gridded with pixel size 1', but the Arecibo beam size is 3.5'. To convert the number of pixels being averaged into the number of independent spectra averaged, N_p must be divided by 9.62, the number of pixels in one beam. As discussed above the noise is expected to decrease with a power of -0.50, but the measured values (see figure 7.9) indicate the actual value is closer to -0.44.

This equation enables us to calculate how the S/N should average for a collection of sources within N_P pixels being averaged. We can now estimate how changing the distribution of the hydrogen may enable some potentially detectable hydrogen to have been missed in the stacking process. The filling factor of the HI over the whole area of the cube and the filling factor for stacking need not be identical. For example, in the case of a cloud filling only a small part of the cube, it would be better to average over a smaller area enclosing only the cloud, rather than a larger area containing the whole cloud but also a great deal of noise.

For instance, suppose that somewhere within the cube is a barely detectable HI cloud, for simplicity's sake with some resolved spatial extent but present over only one channel. Let us suppose the cloud spans 10x10 pixels, each at the 2σ level (unstacked). This would be very difficult, if not impossible, to detect in the raw cube. It would also be undetectable if we were to average over the full 17,110 pixels available, as the equation predicts $S_{S/N}$ = 0.28 (this is due to averaging a small amount of signal with a large amount of noise). However, if we were to average only over those 100 pixels with detections, the $S_{S/N}$ would be 5.03, which could be detected. Thus, averaging over the whole area may not be the best strategy if the HI is clumped. Regardless of the column density constraint, there may be potentially detectable HI below the limit that this method is effectively smoothing out.

Alternatively, the HI may be present not as spatially contiguous clouds, but fluff - very spatially small, isolated clouds again perhaps spanning only 1 channel. If we have such clouds throughout the cluster, i.e. in each of the 500 or so channels that span its approximate 3,000 km/s velocity range, then their detectability again depends only on their S/N and filling factor. Unless these are sufficiently high they will be even harder to detect than spatially extended clouds, since there is no possibility of averaging over only those pixels containing them (except by taking an average over only a few pixels, but this will not increase their S/N). A simple illustration of this is shown in figure 7.10.

It is worth noting that HI has previously been found in both relatively uniform streams and in distributions similar to the HI fluff described above. Some of the streams have been described in chapter 1 (e.g. VIRGOHI21 and a 500 Kpc stream discovered by Koopmann et al. 2008). While nothing exactly like the HI fluff is known within the Virgo Cluster, HI is found in similar distributions around nearby galaxies (for example M33 - Grossi et al. 2008).

The issue of the velocity filling factor must also be considered. A 3σ peak in one channel would undoubtedly be disregarded as noise. Firstly, it does not look at all prominent to the human eye above the surrounding channels. Secondly, the data cube has such a large number of pixels that 3σ peaks will be extremely common simply from statistics. Assuming the noise to be Gaussian, the probability of any individual pixel having an intensity above the 3σ level is 0.00135. But within the whole cube there are $320\times80\times4096$ pixels, so we expect 141,000 such peaks by pure chance (or about 20,000 within the volume of the Virgo cluster itself). However, a sufficient number of 3σ peaks contiguous in space and velocity begin to look more interesting, and is unlikely to happen by chance alone.

From experience, a 3σ detection is probably the lowest that could conceivably be claimed as a detection. As an example the lowest S/N of a secure detection in the whole VC2 cube is 4.38 for a velocity width of 370 km/s. Whether a detection can be claimed as genuine - from a spectrum alone - depends upon the velocity width as well as its S/N. As another example, source 31 (a high velocity cloud) has a narrow ($\approx$100 km/s velocity width) detection with a peak S/N of 5.5. From its spectrum alone, even at 5.5σ it is a borderline detection. Yet there is no

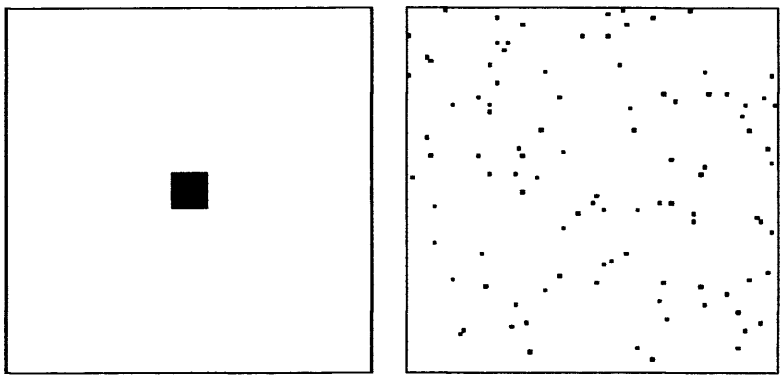


Figure 7.10: Two hypothetical slices through a data cube containing the same M_{HI} per pixel and the same number of pixels containing HI (black points). If distributed in a contiguous cloud (left) then stacking the correct small area will enable a detection. If distributed as widely-dispersed HI "fluff" (right), then stacking a large area will not necessarily result in a detection, even though the final rms may be lower and the amount of HI actually present the same. This is because it is possible to stack the HI on the left with filling factor 1, though it is not with that on the right - even though the HI in both cases has the same filling factor of the full area of the cube.

doubt that this particular source is genuine because it spans a wide spatial range. Unfortunately, whether a peak can be described as a genuine detection depends on more than its S/N alone, and is difficult to quantify objectively (as discussed in chapter 4).

The velocity filling factor may make detecting HI fluff especially difficult, if not impossible. In the example of figure 7.10 it is assumed that each separate cloud is in the same channel. While this seems a very reasonable assumption for spatially contiguous clouds it seems much less likely for a series of independent clouds. Consider an example similar to that of the independent clouds in figure 7.10 but suppose there are now 10,000 pixels at the 2σ level rather than 100. If these are all distributed in one channel, their filling factor will be high and equation 7.6 predicts a stacked S/N of 28.29, easily detectable. However suppose they are now distributed over 100 channels. This means their filling factor per channel will be 100 times lower and correspondingly, the stacked S/N will now only be 0.28, completely undetectable even though the mass of hydrogen actually present is the same. Yet it is also worth reiterating that if the HI is too narrow in velocity, it will be difficult to detect unless its S/N is very high. For a fixed mass of gas, its velocity filling factor must be neither too high (or its spatial filling factor will be too low for it to be detectable through stacking) nor too low (or it will appear as a spike in the noise, which are usually interpreted as spurious).

It is useful to write equation 7.6 to account for the velocity width of the source and in terms of the total mass of hydrogen present. For an area of N_p pixels, there are $N_p/9.62$ independent spectra. Given a total mass of hydrogen present of MHI_T, the average mass of hydrogen per spectrum MHI_S will be :

$$MHI_s = \frac{9.62 \times MHI_T}{N_p}$$

For a tophat profile, the S/N in each spectrum will be given by :

$$S_0 = \frac{MHI_S}{2.36 \times 10^5 \times d^2 \times \Delta V \times rms}$$

Where ΔV is the velocity width of the source. Using these equations, assuming the typical rms of 0.5 mJy together with equation 7.6, gives :

$$S_{S/N} = \frac{3.01 \times 10^{-2}}{\Delta V \times d^2} \times MHI_T \times N_p^{-0.56} \tag{7.7}$$

As discussed in more detail in the next section, stacking was also performed on 3 regions in each cube apparently containing no sources, and therefore not suffering from the problems of masking. The largest of these areas, in VC1, has an area of 2,555 pixels. The expected mass that should be present from HI deficiency in this region is 6.7×10^9 $M_\odot$. Even in the case of the HI spanning 3,000 km/s, equation 7.7 predicts a S/N of 2.9. Recall that the

equation assumes a tophat profile, which will give the lowest peak S/N for any given total flux. Assuming that the HI spans such a large velocity range is also pessimistic with regards to S/N. If there really is that much HI present in the region, this suggests that it could (and indeed should) be detected by this method.

The counter argument is that there may nonetheless be considerable spatial variation in the HI distribution. In the case of the velocity distribution outlined above, it would only require a variation of a factor 3 to make the HI easily detectable or completely undetectable (though admittedly, as stated above, this would require a pessimistic velocity distribution). The point is that while this method can succeed, it relies entirely on the distribution of HI satisfying quite strict criteria.

A perhaps more realistic estimate of the S/N obtainable by stacking may be made by considering the velocity range spanned by the galaxies present. In VC2, the cluster members span approximately 1,700 km/s of velocity. The largest area stacked in this region has an area of 1,235 pixels, and should contain 4.8×10^8 $M_\odot$ of HI from the deficiency of the galaxies. Equation 7.7 predicts a stacked S/N of just 0.5. This is not so low as to make the exercise hopeless - a variation of a factor 6 could make the HI detectable - but clearly it indicates that the lack of a detection is not surprising. If the full area of 17,110 pixels is used, the equation predicts a S/N of 2.6. Given the problems of baseline ripple caused by masking galaxies, it is again not surprising that no detection results.

In VC1 there are 3 separate galaxy clouds spanning different spatial areas, velocity ranges and each at a significantly different distance. An attempt to account for these factors is shown in table 7.2. This is unavoidably subject to considerable uncertainty since the distance determinations are not performed for individual objects, as discussed in chapters 3 and 5. Galaxies at the same distance are treated as associated. The predicted MHI is derived from the total deficiency in each cloud and the fraction each stacked area spans of the area of the cloud as a whole (this assumes gas lost by galaxies in the cloud is distributed over the same spatial area as those galaxies).

Table 7.2: Predicted S/N for stacking of intracluster HI in VC1.

Cloud distance (Mpc)	Expected MHI (log $M_\odot$)	Velocity range (km/s)	Area (pixels)	Predicted S/N
17	9.43	2500	2555	1.37
23	9.63	2000	2075	1.72
32	9.53	1500	931	1.43

The predicted S/N values for VC1 are only at best a factor 3.4 above that for VC2, even though the total deficiency is so much greater. Again, these values are large enough that the exercise has the potential to succeed. However, they are low enough that the lack of any detection cannot really be taken to indicate the presence (or lack) of an intracluster medium sufficiently massive to account for HI deficiency.

In short, the rms reached by stacking may enable relatively small quantities of HI to be detected, but only if stacking is performed correctly and the HI distribution is correct both in space (contiguous clouds) and velocity (contiguous over several channels). Perhaps counter-intuitively, although stacking over the largest area may give the lowest rms, it does not necessarily enable the detection of the lowest mass of HI, as small clouds with low S/N may be smoothed out. The sensitivity achieved by stacking is high enough that a detection is possible, but not high enough that it should be expected.

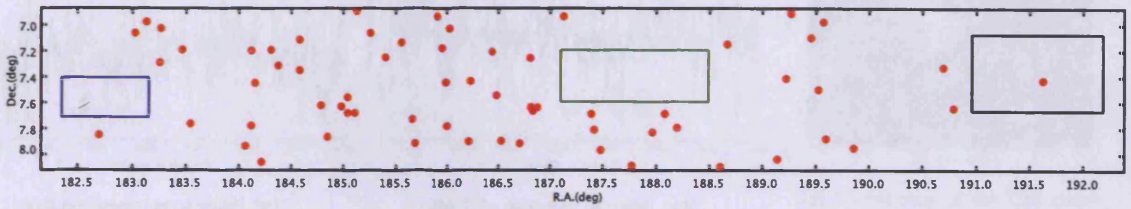
7.3.4 Four hundred million non-detections

Using equation 7.6 and some appropriate constraints may suggest a better strategy for stacking to succeed. The HI fluff of figure 7.10 would seem to be a worst-case scenario for any kind of stacking, as it would be difficult to improve on the S/N of the individual sources - if the HI lies in single-pixel peaks, then averaging over more than a single peak will be averaging signal with noise. A medium uniform in space and velocity will also be difficult to detect, as discussed above. However, if the HI is distributed in discrete, spatially resolved clouds then the chances for success may be improved.

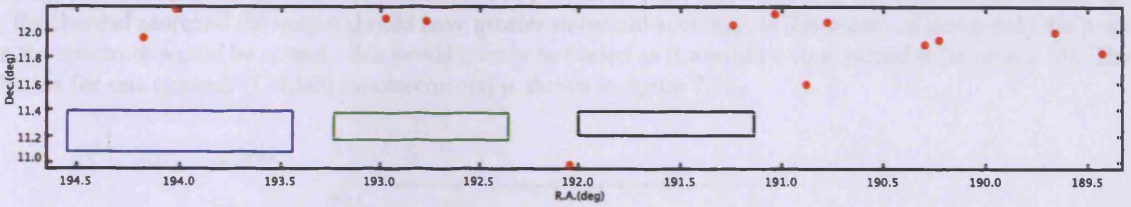
Since contiguous objects with a relatively low overall filling factor may have been missed by stacking large areas, it seems logical to stack smaller areas. As the location and shape of any such objects is unknown the best chance of success will come from using an automated algorithm to search every possible area. It is also evident from the

above discussion that the baseline may be strongly affected by the presence of masked objects. For an automated extraction program this would be a severe problem. To avoid this issue, only areas in which no masking is required are considered (3 in VC1 and 3 in VC2). The areas examined are shown in figure 7.11. The algorithm is a Fortran program that feeds parameters into *mbspect*.

Figure 7.11: Blank areas of the data cubes used in the automated stacking algorithm, showing all HI detections with $cz < 3{,}000$ km/s. For reference, in each cube blue refers to area 1, green to area 2 and black for area 3. Red circles indicate the position of HI detections.



(a) VC1 region. The single "detection" in area 1 is a very weak unsure source.



(b) VC2 region.

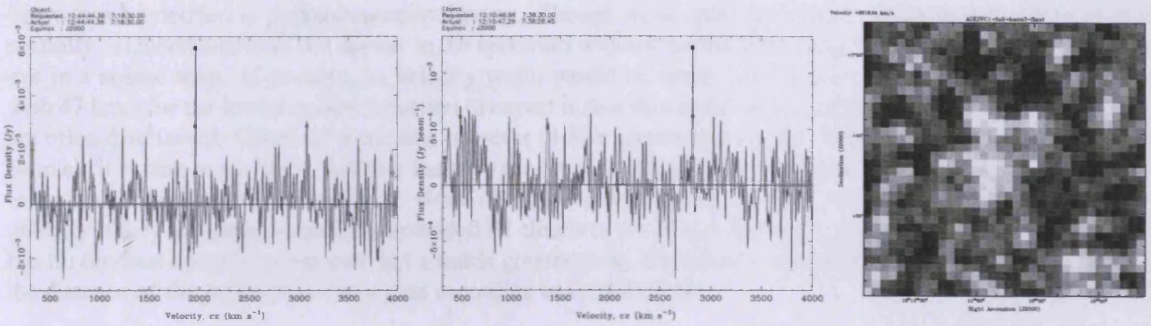
The program proceeds as follows. At each pixel, a box is defined over which the spatial averaging is done by *mbspect*. The maximum size of this box is defined by the position of the pixel - the program prevents the box from ever overlapping with pixels outside the area chosen to examine. The box is initially set to a width of 3 by 3 pixels. The width in the x direction is increased until the box reaches its maximum x-size, then is reset to 3. The size in the y direction is increased and the width in the x direction again iterated. This process is repeated for every possible y width at every pixel, thus, every possible rectangular area is examined. Only the velocity range 300-3,000 km/s is searched, chosen to include Virgo, avoid contamination by the Milky Way, high velocity clouds, and RFI.

This technique searches a vast number of rectangles, in total for all 6 areas about 830,000 (which as 540 channels are searched simultaneously leads to 440 million searches). In order to avoid large numbers of spurious sources and complicated source extraction techniques, the only criteria to examine a source is that it has a S/N greater than 5. There are often detections centred on or overlapping the same spatial coordinates and at the same velocity (but with different spatial widths). Topcat is employed to match these detections, and only the detection with the strongest S/N retained. This reduces the number of potential detections to examine to a manageable level, a few tens of objects.

No obvious, convincing detection was made. Although the extractor did find objects which at first glance appear interesting, the majority are easily dismissed because they are present in only one channel and/or only visible in one polarisation. A very few ($\sim$10) cannot be explained away like this but none of them appear at all convincing as genuine detections (this is of course unavoidably subjective). Two such candidates are shown in figure 7.12. The very narrow width of the detections (< 15 km/s) makes them unlikely to be genuine. Note that these are two of the

most plausible candidates - the majority of the remaining candidates are of even lower quality in that their S/N is lower and/or it is unclear if they are polarised.

Figure 7.12: HI cloud candidates.



(a) 8.6σ detection of width 3x3'

(b) 7.2σ detection of width 7x9'

(c) Spatial map of the 7.2σ cloud candidate

The high S/N of the detections would seem to make them statistically significant. A detection greater than 8.5σ should occur with a probability of only 9.48×10^{-18}, if the distribution of S/N is Gaussian. Even with 440 million combinations of data from the cubes, the number of such detections expected by chance alone is only 4.17×10^{-9}. To better place these apparently high-sigma detections in their statistical context, the algorithm is modified so that the channel searched is fixed and the S/N output for every rectangle, not just when above a specified threshold. By fixing the channel searched the output should have greater statistical accuracy. If this were not done, only the peak S/N in the spectrum would be output - this would clearly be biased as it would be constrained to be above 1.0. The distribution for one channel (124,200 measurements) is shown in figure 7.13.

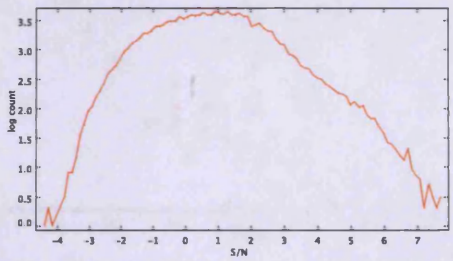


Figure 7.13: Distribution of the measured S/N from stacking every possible rectangular area at one channel, with a logarithmic number count and 100 bins. 124,200 measurements, mean 0.80 with standard deviation 1.49.

Regardless of the reason why, it is clear that the distribution of S/N from this procedure is non-Gaussian[1]. Making the approximation that it is Gaussian but centred on the mean of 0.8 with standard deviation 1.49, the 8.6σ detection is actually only 5.3σ above the mean. This is far less statistically significant, expected to occur with probability 5.83×10^{-8}, or, for 440 million measurements, 26 times in total. That it does not occur this frequently is probably a reflection of the fact the distribution is not really Gaussian. Although the detections are reasonably strong, since they are effectively single-channel peaks in the distribution they can be attributed to occurring by chance alone, rather than HI emission. Additionally, although there are a few tens of objects at (apparently) high S/N ratios, as noted above the majority of these are not independent detections.

The first detection in figure 7.12 is of both low velocity and spatial width (3x3'). Since this is the minimum spatial width it is not really stacking at all, sensitivity is no higher than in the standard method of extracting spectra. It was missed in conventional source extraction techniques because of its very low velocity width - in this regard,

[1] There are several factors that will contribute to this. The distribution of the rms of each spectra cannot be Gaussian, it must be offset from zero. Thus although the flux distribution is Gaussian the S/N distribution need not be. Also, stacking is being performed on a different number of different sized areas each of which will have a different rms. Furthermore, if there is an extended source present (genuine HI or even RFI), it will enhance the stacked S/N in multiple areas, skewing the distribution.

it is typical of the objects detected by this extractor. It is unusual only in its S/N, but, as discussed above, given the number of areas searched even 8.6σ is not statistically very significant. This object has no obvious optical counterpart, and if this is a genuine source it would have an HI mass of just 1.2×10^6 $M_\odot$ at a distance of 17 Mpc (the distance of the VCC galaxies in this region of the cube as given in GOLDMine).

The second detection is perhaps more intriguing. Though of no greater velocity width than the first, it is more spatially extended and does not appear in the spectrum without spatial averaging. However, it is just visible to the eye in a spatial map. If genuine, its velocity width would be exceptionally low : a W20 of 20 km/s, compared with 47 km/s for the lowest secure detection (a caveat is that this could be a selection effect as very narrow sources are often dismissed). Given its weakness, however (4.30σ approximating the S/N in the same way as for the first source), it seems more likely that this merely represents a coincidental alignment of a few above-average peaks in the flux that occur by chance alone. None of the candidate detections have much velocity width, which seems unlikely if they are genuine spatially-extended HI clouds in the Virgo cluster. As with the first detection, this object has no obvious optical counterpart, but a much greater mass if genuine - approximately 1.9×10^7 $M_\odot$ at 32 Mpc, the distance of the galaxies in this region according to GOLDMine.

A statistical argument may also be made against the validity of the cloud candidates as real HI detections. They are not nearly numerous (or massive) enough to explain the HI deficiency in the regions concerned. Even with clouds as massive as the second candidate, over 400 would be required in area 3 alone, yet only 80 candidates over 5σ are actually detected (this is an upper limit, not accounting for the majority being clearly polarised). The total flux from the candidates is a factor of 70 too low to explain the HI deficiency. Even if these clouds are real, they comprehensively fail to explain the HI deficiency in galaxies - the original motivation for the search. Secondly, if they are clouds originating from gas lost from galaxies, they would be expected to have a broadly similar velocity distribution to galaxies in the region. As shown in figure 7.14 this is not the case - the cloud candidates are uniformly distributed in velocity.

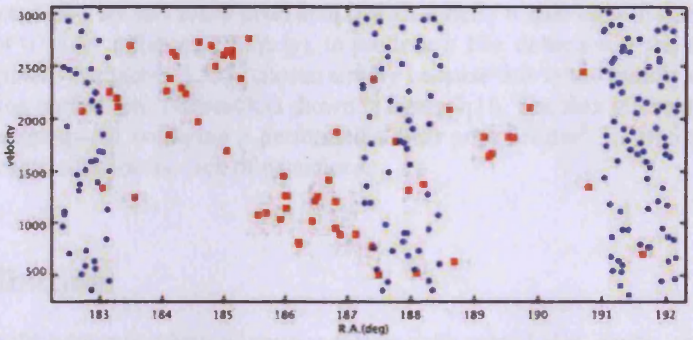


Figure 7.14: Position-velocity distribution of the HI cloud candidates found by spatial averaging (blue circles), in comparison with that of the HI-detected galaxies in VC1 (red squares)

The distribution of the clouds remains uniform regardless of the redshift range searched : a similar number of "clouds" are found in the velocity range -1,100 to -100 km/s as are found in the range 17,000 to 18,000 km/s. In each case the mean size of the clouds in pixels is also very similar, about 100 pixels^2, and their velocity widths extremely low (typically < 15 km/s). It seems unlikely that more distant clouds would conspire to be physically larger so as to be of the same pixel scale as nearby clouds. The "clouds" at higher redshift would be huge, with a mean diameter of 1 Mpc. Finally, it also seems unlikely that there are spatially extended HI features that are strongly blueshifted but of very low velocity width. It is far more plausible that the cloud candidates, in general at least, are artifacts rather than real HI detections.

The procedure adopted here is by no means the ideal way of stacking the cube. *mbspect* only allows control of the width of the stacked area in the x and y axes; diagonal orientation is not possible, and width must be an odd number of pixels. If the HI is present in a stream it may be detected if aligned with the right ascension or declination axis, as its stacked filling factor may be close to 1.0, but if diagonally orientated with respect to the axes then its filling factor must be much less, and its chance of being detected decreases (see figure 7.15). There is no easy way to correct for this. The ideal method would be to stack every possible group of 3x3 pixels in every possible combination, thus avoiding any issues of the shape of the HI clouds, and also allowing the possibility to detect HI "fluff". However this would be computationally prohibitive. Even for the smallest area the number of

combinations exceeds 10^{17}.

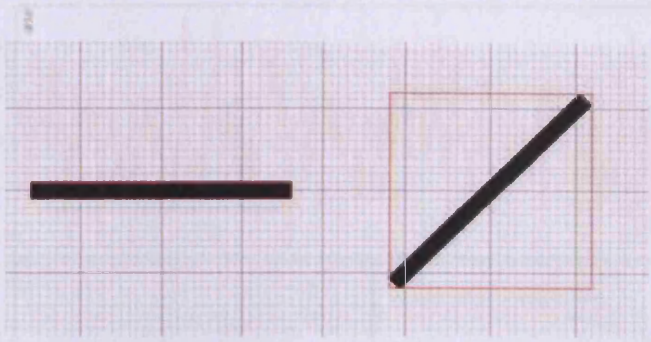


Figure 7.15: Example to show how orientation of an HI stream affects its detectability by the procedure outlined here. Black pixels represent pixels containing HI. Both streams have the same number of pixels. The stream on the left can be entirely enclosed in a rectangular area (red outline) and stacked with filling factor 1.0. To enclose the entire stream on the right, because of its orientation, requires it to have a stacking filling factor of 0.11, reducing its stacked S/N by a factor 9.

Spatial averaging has failed to produce a clear detection despite the millions of different areas examined. There are many reasons why this might be so, both astrophysical (such as the distribution of the HI) and observational (i.e. sensitivity of the data) - see section 7.6 and chapter 8. It is at least possible to prove that the method itself is valid and can indeed be successful provided the right conditions are met. Using equation 7.6 together with the rms of the raw data, and choosing arbitrarily a number of pixels, it is possible to predict the amount of flux (or S/N) that could be recovered by stacking that would be undetectable in the raw data. To test that spatial averaging can recover signals from genuine data, flux can be artificially injected and stacking performed on the modified data.

Suppose for clarity that the S/N required for a detection is 10σ. The flux in area 3 of the VC1 cube has an rms of 0.95 mJy. A signal injected into the full 1,890 pixel area (for simplicity's sake only in one channel) will by equation 7.6 require a S/N of 0.98 (or a flux of 0.93 mJy), to produce a 10σ detection when spatially averaged. This flux is injected into all pixels in channel 3,555 (chosen simply because this is the middle channel in the extracted data) and spatial averaging performed. The result is shown in figure 7.16. The flux injected is quite undetectable in a single spectrum, but when spatial averaging is performed a clear peak results[2]. Therefore, the spatial averaging procedure itself is not responsible for the lack of detections.

7.4 Flux distribution

Spatial averaging is clearly subject to a host of various factors - the spatial size, shape, and orientation of any HI cloud as well as its velocity distribution - which must be quite specifically satisfied to gain a confident detection. Another approach is possible which is far less dependent upon the distribution. Any HI present in the data is by definition an excess of flux. By summing the flux values within a region, the net flux can be measured, and this should correspond to any HI that is present. This assumes that the distribution of flux in the blank pixels is symmetrical (though not necessarily Gaussian). The advantage of this approach is that it is distribution independent - it does not matter if the HI cloud has a specific shape or velocity profile. The disadvantage is that all information besides the flux from the HI is lost, whereas spatial averaging preserves some measure of its spatial location and velocity.

An automated program is employed that extracts the spectra at each pixel in the cube, over the velocity range 300-3,000 km/s, to avoid contamination from high velocity clouds. A first-order baseline correction is applied, again by use of a Fortran program that feeds parameters into *mbspect*. The spectra are then integrated in velocity, and the positive and negative integrated flux values are summed to give the net integrated flux. This is done for 3 areas in the Virgo cluster in VC1 (see figure 7.11).

AGES data have first-order baseline calibrations applied to each spectra during the data reduction process (see chapter 2). However it is important that this step be included while extracting spectra. The reason for this is well-illustrated by integrated flux maps. Figure 7.17 shows maps with and without this extra calibration. Without it,

[2]The actual S/N of the peak is 10.31, rather than the expected 10.00 - this is likely due to the uncertainty in how the rms decreases with number of pixels averaged, see figure 7.9.

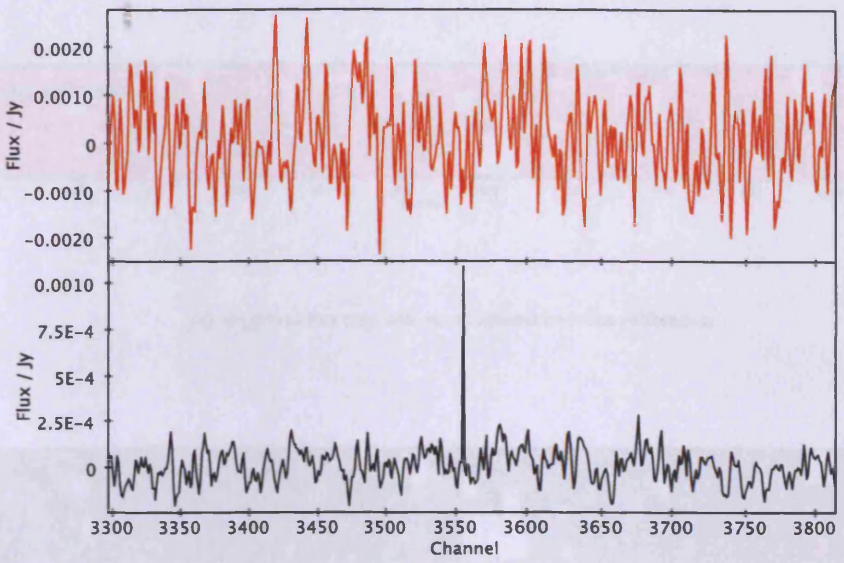


Figure 7.16: Example to prove the validity of spatial averaging. Flux is artificially injected into channel 3555. The amount injected is too small to be seen in a randomly-chosen single spectrum (above), but when spatial averaging over the whole area (below), a clear signal results.

there is clear horizontal striping, but this is virtually entirely removed with the calibration. The calibration applied during data reduction is performed on the entirety of each spectra, i.e. over the full velocity range. It appears that there is some time-dependent noise that this is not able to remove fully, presumably varying over a higher frequency equivalent to a few thousand km/s of baseline. Thus repeating this calibration over the range 300-3,000 km/s is able to remove this variation.

While it is possible to give some quantification as to what constitutes a detection by spatial averaging, it is less obvious as to how much flux is required for a sure detection by this method. The net integrated flux for the three VC1 areas are +3.6, -24.4 and +20.8 Jy km/s.[3] The flux distributions are shown in figure 7.18. To show the symmetry of the distributions, the positive flux values are also shown reflected about the y-axis.

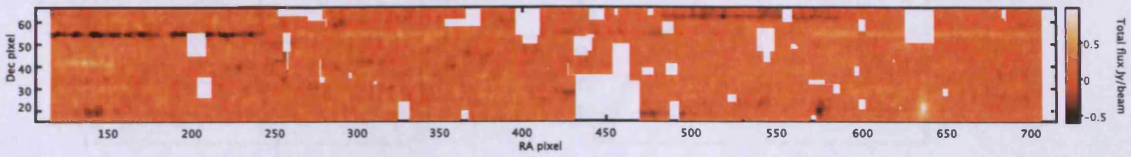
The distributions are clearly very good approximations to symmetrical Gaussians. The measured excesses can be seen to be statistical effects : in each case they are due to the shape of the overall distribution, rather than a few unusually high or low values. Given the total HI deficiency of galaxies in the VC1 region and the size of each area, the expected flux values are 35, 76 and 97 Jy km/s, for areas 1, 2 and 3 respectively.

From these results it is apparent that there is no clear detection of HI from this method, since the negative flux in region 2 is even greater than the positive flux in area 3 - there are clearly many intrinsic errors in this process. However, it is difficult to say if this genuinely places some (significant) constraint on the amount of intracluster HI present or if the errors are simply too high to give any indication at all. To try to address this the flux values are extracted in a further 7 volumes, from the same 3 areas but at higher redshifts. While there is no reason to suspect an intracluster HI medium outside the cluster, there could be undetected galaxies contributing to the flux. Therefore, as an estimate of the error the highest negative flux found by this method was -27.20 Jy km/s, a factor 3.6 below the highest expected flux. Thus it seems that the lack of detections by this method is a genuine result (at least for the two larger areas) and not merely due to the errors inherent in this process.

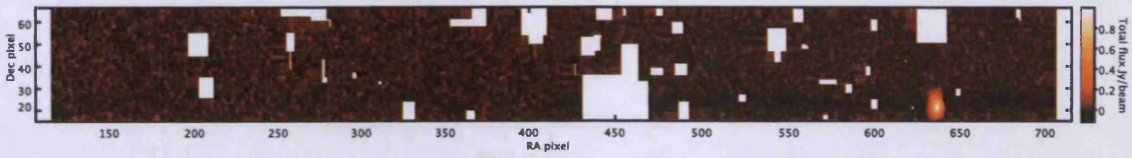
Another way to test this approach is to use the flux from known galaxies to see how the distribution is affected. In figure 7.19, the distribution of the flux values in the VC1 cube are shown using different extraction parameters, which are discussed below. Firstly, a small subset volume of the cube is extracted containing only source 66 (unmasked). This is a relatively weak source of $M_{HI} \approx 8\times10^7$ $M_\odot$, a total integrated flux of 0.346 Jy km/s. A clear positive tail is seen with no corresponding negative excess, thus, the source is detectable. The measured integrated

[3] As each area lies close to a different galaxy cloud, so the net flux in each area is also measured over an appropriate velocity range (i.e. corresponding to that of the associated galaxies). The net fluxes in this way are +2.78, -38.72 and +28.92 Jy km/s.

Figure 7.17: Integrated flux maps for the VC1 cubes, velocity range 300-3,000 km/s.



(a) Integrated flux map with no additional baseline calibration



(b) Integrated flux map with first-order baseline calibration applied to each spectra

flux in this subset volume is 74 Jy km/s, much higher than would be expected from the source alone.

In the second figure the area is greatly increased to include the entirety of the cluster volume (about 15 million values extracted from the cube). The positive tail is much less evident. Simply by chance due to the larger number of pixels, there are now many more pixels of a flux comparable to the source, effectively overwhelming the positive tail. The net flux, however, greatly increases to 1,706 Jy (or an integrated flux of 8,620 Jy km/s). At face value this would suggest there are other sources of flux included that have not been masked. Whether these correspond to actual neutral hydrogen, or are some statistical artifact, is harder to assess. In the third figure, with all sources masked there is no sign of the positive tail, though the net flux remains 1,695 Jy (8,564 Jy km/s).

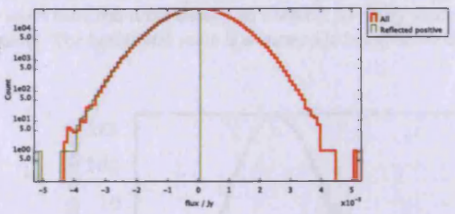
In the fourth figure, all but but the bright galaxy VCC 94 are unmasked. There is now an unmistakable positive tail, which is very much stronger than that of source 66, since VCC 94 source has a total flux nearly 20 times greater (6.45 Jy km/s, or an HI mass of $\approx 2\times10^9$ $M_\odot$). The flux values in this source are far greater than occur by chance alone even with the 15 million values of the the entire cube, thus they lie well outside the Gaussian distribution from the noise and are clearly visible. Similarly in the fifth figure, with no masking applied at all, the positive tail is even stronger and the net flux significantly (58%) higher, 2,668 Jy (13,577 Jy km/s).

Although it is apparent that bright sources can be identified by a visual examination of the flux distribution, it is not clear if diffuse intracluster HI would be manifested in such an obvious way. It is far more likely to be present as an excess of faint flux values (compared with the Gaussian ideal) than a few very strong flux values. The effect on the visual appearance of the flux distribution can be tested by artificially injecting an appropriate amount of flux into, for example, area 1[4] and examining the distribution visually. A total of 35 Jy km/s is injected into 30% of the volume, spanning the full area but only 30% of the channels - if this were part of a uniform medium it would be sufficient to explain the observed HI deficiency.

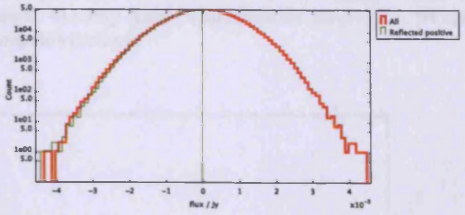
While the measured net flux is correct, it is apparent in figure 7.20 that the shape of the distribution remains very close to a Gaussian, and the flux cannot be identified by visual inspection. This suggests that only a quantitative approach is at all suitable for detecting faint intracluster HI. However, it is not easy to explain the net 1,695 Jy that is present with every known source masked. To attribute this to intracluster HI requires confidence that the net flux

[4]This area is used as it has an excess close to zero, and so will serve as the best approximation for a Gaussian flux distribution. This is assuming that no intracluster HI flux is actually present, but since the measured excess is so small this seems a reasonable assumption.

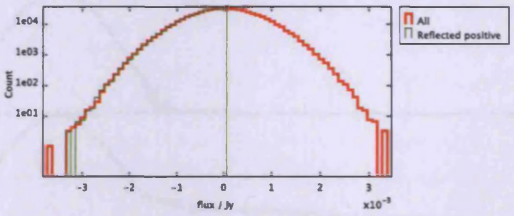
Figure 7.18: Integrated flux distribution for 3 blank areas in VC1.



(a) Area 1 : 0.5 million values, net flux +3.6 Jy km/s



(b) Area 2 : 1.0 million values, net flux -24.4 Jy km/s



(c) Area 3 : 1.4 million values, net flux +20.8 Jy km/s

from the noise alone is truly zero.

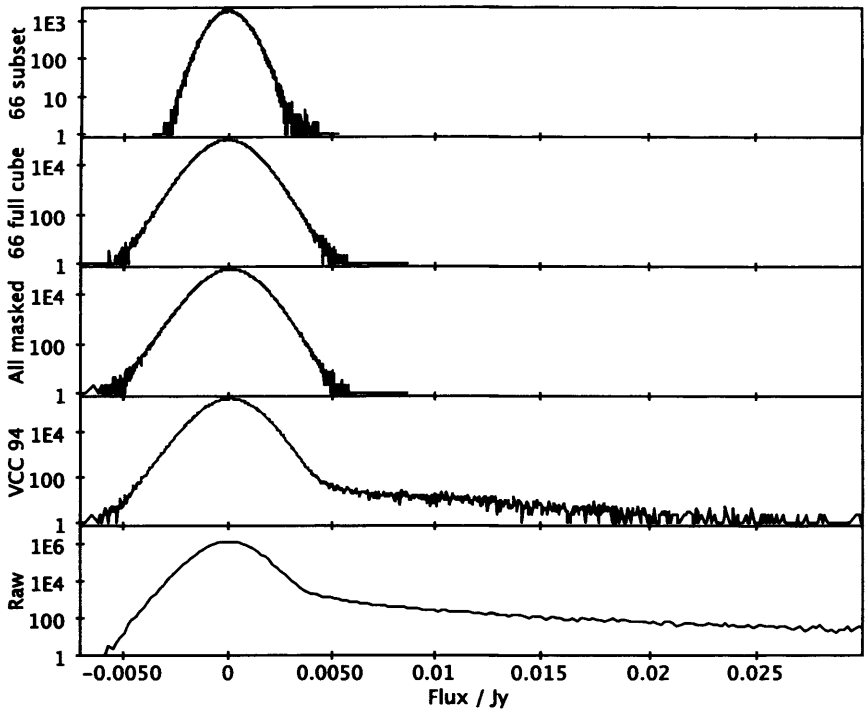
The ideal test would be to have a volume of the cube where it is known that no sources of any kind are present. While this is not possible, the region that comes closest to this criteria is the foreground volume, -1,100 to -100 km/s. Below -1,100 km/s noise in the cube significantly increases while above -100 km/s contamination by the Milky Way becomes important. There is only 1 unsure detection in this region, and only 2 VCC galaxies (without HI detections) with optical redshifts in this range (compared to 174 VCC cluster members in this region overall). Given the paucity of blueshifted galaxies, it seems very unlikely that significant amounts of HI have been blueshifted compared to that present in the more typical velocity range of the cluster, 0-3,000 km/s. Taken at face value, the net flux indicates otherwise. After correcting for the number of flux values extracted being a factor 2.6 lower in this volume (since the velocity range used is necessarily lower), the net flux measured is 2,091 Jy (4,059 Jy km/s).

The conclusion from this must be that this method is simply untenable for detecting diffuse intracluster HI. Bright point-sources are easily detectable by this method, and even quite faint sources are visible provided a small area is chosen. However, detecting a diffuse medium is far more problematic. In the example of figure 7.20, the net flux measured was indeed the expected 35 Jy km/s - but there it was known previously that the net flux from the noise was close to zero. Without knowing for certain that the net flux from the noise is *in general* zero, it is not possible to definitively prove that the measured positive flux originates from HI. Furthermore, the net flux in a likely empty volume of the cube is even greater than that in the cluster. This makes it very difficult to comment on the origin of the net flux, but implies it is a statistical artifact of some kind and does not originate from intracluster hydrogen.

7.5 Effects of data reduction

While it is apparent that stacking can potentially succeed in detecting HI, if it is present and distributed suitability, one final issue of the data reduction process remains. In principle stacking the largest area results in the lowest rms (true enough) and thus the detectability of the lowest column density material. However, a totally uniform medium present throughout the entire cube will never be detectable. The bandpass is calibrated for each scan at each channel by subtracting the median flux of the scan (see chapter 2). While this is perfectly adequate for point sources, which span a small spatial range and so will not affect the median, for sufficiently extended (however

Figure 7.19: Flux distribution from the VC1 cube over the velocity range 300-3,000 km/s, with baseline calibration applied to each spectra. Each distribution has 1000 bins and the same horizontal scale, with a logarithmic number count for the vertical axis. From top to bottom : 1) The flux in a subset area containing source 66, 311,000 values. The remaining distributions are all of the entire cube, 15.2 million values. 2) Every source apart from the weak source 66 masked. 3) Every source masked. 4) Every source apart from the bright VCC 94 masked. 5) Raw data - no masking. The horizontal scale is truncated to better show the above distributions.



faint) sources the median will be affected. Thus any perfectly uniform medium spanning the full spatial area of the cube will be subtracted and not detectable.

It is not straightforward to analytically determine at what filling factor a source will become undetectable due to data reduction, because it is difficult to estimate how the median flux will vary when a flux is injected with some filling factor. Suppose a single random pixel was to be set to a flux equivalent to one million sigma - the median, simply being the middle value, would not be at all affected, whereas the mean (which is the flux recovered when spatially averaging) clearly would be. Both processes of stacking and data reduction must be accounted for. This is straightforward to do empirically.

Consider an AGES data cube of uniform noise containing an extended source. When spatial averaging is performed on the data, the S/N, accounting for baseline subtraction, will be given by :

$$S_{S/N} = \frac{mean\ flux - median\ flux}{stacked\ rms} \tag{7.8}$$

Where the median flux refers to that of a single scan (i.e. a track across the sky in right ascension). The mean flux from the noise will be zero if the noise is perfectly Gaussian. If a source is present, the mean flux should therefore be equal to the flux per pixel of the source. Given a distribution of fluxes from real data, this equation, together with equation 7.6, enables the effect of baseline subtraction on simulated extended sources to be estimated.

An AGES scan in right ascension typically covers 5 degrees and contains 960 pixels (320 pixels of right ascension. and 3 pixels of declination). The fluxes in a single blank spectrum 960 channels in length will serve as an approximation to this. This is only an approximation as real baselines inevitably contain a higher rms at the spatial edges and are rarely totally blank : in addition to real HI sources, continuum sources are often present. However, a blank spectrum allows the equation to be tested in a more controlled way without such effects. Another point is that in reality there are not 960 truly independent points in a baseline, this is an effect of binning into 1' pixels. In reality scans are continuous in time, not a set of discrete measurements (though integration is done once per

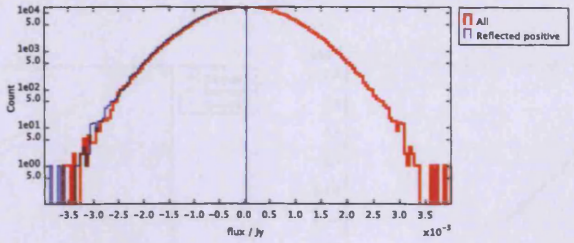


Figure 7.20: Flux distribution for area 1 with a total of 35 Jy km/s injected,

second). Crucially, this exercise uses reduced data rather than raw data. It is intended only to give a very basic estimate of the effect of data reduction.

Both the mean and median of the blank spectrum extracted are very close to zero : 2.12×10^{-5} and 2.46×10^{-5} Jy respectively. That the mean is close to zero allows use of the stated approximation that the mean will be the flux per pixel of any source present. A source can now be artificially "injected" into this distribution by selecting random pixels and adding the required flux to their existing intensity values. For a filling factor f, the first $f \times 960$ pixels have flux injected. This is equivalent to injecting random pixels with flux because the flux in one pixel shows no sign of being correlated with the flux in neighbouring pixels. Stacking can then be simulated by means of equations 7.8 and 7.1[5]. In this way both the change in the median due to the presence of a source, which will lower the S/N recovered by stacking, and the effect of increased sensitivity due to the source spanning a larger area, which will increase the recovered S/N, can be measured simultaneously. Note that the use of equation 7.1 means the stacking process need not be actually performed : rather, its effect is calculated using an equation previously shown to be valid. Only the effect of a source on the median level of the baseline is measured empirically.

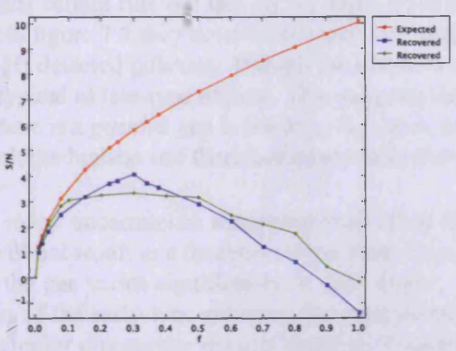
The flux injected is chosen to give an expected value for $S_{S/N}$ of, arbitrarily, 10, if the maximum possible (320x80) pixels contained such a flux level and were stacked. The value for f_s, the filling factor when stacking, is assumed to be 1.0, simulating the ideal case of stacking only over a contiguous uniform cloud (since the flux is injected into effectively random pixels it would be equally valid to assume f_s equal to some other value to simulate HI "fluff"). The number of pixels injected is varied, so that the stacked S/N predicted by equation 7.6 will vary. This has the advantage of exploring only the effect of f_b (the scan length filling factor) on baseline subtraction - alternatively the flux can be altered with f_b so that the predicted S/N remains constant, but this means that two parameters will be affecting the subtraction.

A final caveat is that the amount of scan length the injected source covers and the amount of spatial area available for stacking that it covers are not necessarily equivalent. If, for example, we have a source covering 3x3 pixels it will cover about 0.94% of a scan length, but only 0.04% of the total area of the cube. This means its effect on the baseline will be relatively high with respect to the ability of stacking to recover it. It is preferable to avoid this issue and have the filling factor of the scan length and total area be equal. This requires the simulated HI source to have a variable shape - the 3x3 source in the baseline is then assumed to span 3x80 pixels in the full cube (i.e. when stacked). The result is shown in figure 7.21.

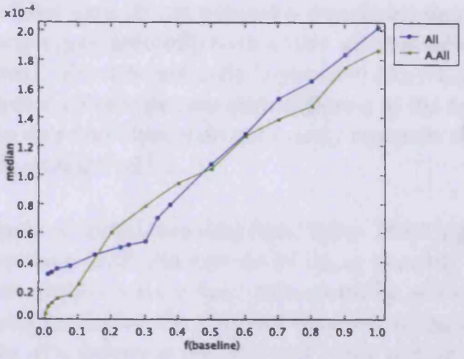
The results for the two spectra appear markedly different. In part (a) the red line indicates the expected recovered S/N without accounting for the median subtraction, while the blue and green indicate results for real data with median subtraction. In one (blue line), while the baseline subtraction clearly affects the S/N recovered at all levels, there is a sharp threshold at a filling factor ≈ 0.3 above which the S/N drops rapidly. Below this filling factor the recovered S/N scales approximately as $f^{0.4}$, rather than $f^{0.5}$ in the ideal case. Above this the S/N decreases in a roughly linear fashion. In the other (green line), for $f < 0.1$ there is good agreement with the ideal case. Above this the S/N plateaus, decreasing for $f > 0.5$ but without the sharp threshold evident in the first case. The difference in the S/N behaviour is reflected in the variation of the median with filling factor : a clear threshold at $f > 0.3$ in the first case, where the median begins to increase more rapidly with filling factor, but no such behaviour in the second

[5]While it is possible to account for the fact that the mean flux is not exactly zero, this is not done. As the filling factor is changed the pixels injected also change and so the mean, inevitably, changes, regardless of the signal being injected. This variation would detract from the main point of the exercise, which is to examine the impact of the baseline subtraction on source detectability, not the typical mean flux of blank spectra.

Figure 7.21: Effect of bandpass calibration on the recovery of sources by stacking.



(a) Recovered S/N as a function of the filling factor. The red line is the expected value without bandpass subtraction, the blue and the green are two spectra where the technique described in the text was applied.



(b) The median as a function of filling factor for the two spectra.

case.

These spectra were extracted from a cube that has already undergone the data reduction process, and so they should be broadly similar. Despite this it is clear that there is a significant difference between the two, although a general trend is apparent. Importantly stacking can succeed provided sufficient flux is present and $f_b \lesssim 0.3$ - 0.5[6]. Even a filling factor of 0.3 enables very long HI streams to be potentially detected, corresponding to 420 Kpc at the Virgo distance. Given the possibility that the the filling factor of a source does not correspond to its filling factor along the baselines, that individual baselines appear subject to significant variation, and that real baselines are often "contaminated" by continuum sources and extragalactic HI detections, it is not possible to accurately quantify the lowest HI column density detectable by stacking. It is nonetheless useful to have a numerical figure at least as an order of magnitude estimate. The values given at the end of section 7.3.2 are approximately correct, provided the HI does not span very much more than the areas in question, as the largest of these spans 26% of a baseline.

The result in figure 7.21 would seem to indicate an optimum source size for stacking to succeed, below which there is insufficient gain in sensitivity, and above which the source spans a significant length of a scan and so is subtracted away. This is true provided the flux per pixel of the source is constant. If the flux per pixel is higher the source may still be enhanced by stacking even if it only occupies a small area. If it is lower, stacking may still gain sufficient sensitivity provided it spans a larger area.

Importantly, if the HI is in a narrow stream oriented north-south, it will occupy only a small fraction of the scan length and thus have no significant impact on the median. In this case stacking would succeed regardless of the total size of the source. Conversely, stacking will be useless for a narrow source oriented east-west (i.e. occupying the full length of a scan), as in this case it would have been completely subtracted during data reduction. This indicates that the automated process described in section 7.3.4 is indeed useful - the shape of the HI clouds will affect their detectability, and since their shape is unknown it is worthwhile to consider multiple areas, rather than only averaging areas of the (apparently) optimum size.

7.6 Summary and discussion

Three techniques have been employed to improve the sensitivity of AGES but none have resulted in convincing detections. By stacking spectra of early-type galaxies, a mass sensitivity limit of 1.7×10^6 $M_\odot$ was reached. This is a factor of 4 lower than the lowest HI mass detection in the VC1 cube, and a factor 5 lower than the lowest HI mass in a detected early-type galaxy. More significantly it is a factor of 30 lower than the mean HI mass in

[6]Note that sources in HIPASS, which uses similar data reduction techniques, are also subtracted if their filling factor is similar (Putman et al. 2002, Barnes et al. 2001).

early-type detected galaxies, though these are admittedly small number statistics.

These results cannot rule out that all the early-types considered here do not possess a significant quantity of HI. As shown in figure 7.5 they do at least imply that bright early-types generally have a very much smaller M_{HI}/L_G ratio than HI detected galaxies. Though the numbers are small, the detected early-types have M_{HI}/L_G ratios that are fairly typical of late-type objects. This suggests the detected early-types are quite different to the non-detected objects - there is a genuine gap in the M_{HI}/L_G ratio, and the detected objects do not merely represent the brightest objects with the highest and therefore most easily detectable quantity of HI.

There are many uncertainties associated with all of the stacking techniques described here. Stacking individual galaxies will not result in a detection if too many do not contain a sufficient amount of HI, or possibly if the kinematics of the gas varies significantly in each object. These problems have been minimised by stacking various subsamples of the early-type galaxies, for example separating lenticular and elliptical galaxies so that more kinematically similar objects are stacked together. Thus the lack of a detection places a real constraint on the amount of HI present. Intracluster HI is far more problematic.

The size of the areas considered in the above analysis are generally only a few percent of the size of the data cubes. The necessity to mask galaxies reduces the area available for the automated program to operate on - for VC1, about 20% of the total area. It is clear that HI on this scale will not affect the baseline significantly and so will be potentially detectable. Whether the intracluster HI is present uniformly over the entire scan length (which would render it undetectable) remains unknown. It seems unlikely that the intracluster medium would be perfectly uniform, given the spatial range spanned by the cubes (at 17 Mpc, 3.0 and 1.5 Mpc for VC1 and VC2 respectively). Gravitational and hydrodynamical interactions with galaxies (whose number density varies spatially) would be likely to perturb such a medium, although it is not clear over what scales clumps in the medium would occur. Analysis of the data reduction process shows that very long streams of order 400 Kpc should be detectable, and since the medium is unlikely to be uniform, the data reduction process cannot account for the lack of detections.

That the stacking procedure itself is capable of detecting gas - with the correct distribution - has been demonstrated by artificially injecting a suitable source and performing spatial averaging. Neither the data reduction nor the spatial averaging procedure can be blamed for the lack of detections. Absolute proof that AGES is capable of detecting extended sources is provided by high velocity clouds, one of which in VC2 spans 19' of right ascension - which would be equivalent to nearly 100 Kpc at the Virgo distance. Thus if there are spatially extended HI structures within the cluster, AGES should be able to detect them but fails to do so.

That the spatial averaging procedure has not made any real detections cannot be proved absolutely, but this seems by far the most likely result. Cloud candidates of similar pixel size, flux, and velocity distribution are found in similar numbers in similar volumes (in terms of pixels), even in the blueshifted region of the cube. There is no correlation with the velocity distribution of the galaxies. The candidates are all of extremely narrow velocity width, most are polarised, and do not have nearly enough flux to account for HI deficiency. Even the strongest source does not appear to be statistically significant given the number of areas searched. As with source extraction of point sources, it can be difficult to decide if a detection is real. The balance of evidence strongly suggests that nothing has been found - certainly nothing that could explain the HI deficiency of galaxies in the cluster.

As has been shown, spatial averaging will only succeed provided the distribution of the HI is suitable. It is able to rule out the intracluster HI as being distributed in clouds of narrow velocity width ($\sim$ 100 km/s), as these would be expected to have a high S/N ratio but are not detected. It is not able to rule out a medium of more uniform spatial and velocity distribution ($\sim$ 1,000 km/s). However, it does not seem plausible that the medium would be this uniform, since it is supposed to originate from a discrete number of galaxies losing gas through processes such as ram pressure stripping, i.e. in streams. If these streams are spatially narrow, then averaging over large areas may not detect them, but an automated algorithm has been developed to reduce this problem.

The simplest explanation as to why no detections are found is that the areas considered do not contain sufficient HI. This does not exclude the possibility of a clumpy intracluster medium massive enough to account for the observed HI deficiencies. The automated stacking algorithm is in no way able to rule out the possibility of intracluster streams present outside the areas examined. As shown in figure 7.11, this may be a particular problem for VC2, in which the blank areas available lie as much as 1 degree from the nearest HI detection. In other words it could be that all the HI lost from galaxies resides outside the areas available for stacking.

It is surprising that all attempts at spatial averaging have failed to produce a single convincing detection, especially in VC1. In VC2 there are only 13 HI-detected galaxies, and the areas available for stacking are up to one degree away from them. In contrast there are 64 detections in VC1, some of which correspond to galaxies that have lost in excess of 1×10^{10} $M_{\odot}$ of HI, and the areas available for stacking are in close spatial proximity to the galaxies. This is especially surprising considering that many streams have been detected in other (less sensitive) surveys of the cluster, including one over 500 Kpc long by ALFALFA, without the need for any kind of stacking.

It does not seem unlikely that streams of HI should be found in this region. Both massive and dwarf galaxies show HI deficiency, but with few undetected late-type objects. Simulations suggest that dwarfs lose all their gas on their first pass through the cluster, so the detected dwarfs must still be undergoing their first pass and thus still losing gas. Since the galaxies reside in distinct clouds, it follows that the more massive galaxies are also still losing gas.

To summarise, spatial averaging can succeed given the correct distribution of HI, but no convincing detections are made. Since data reduction and source extraction procedures appear not to be responsible, there are only two possible reasons for this : that there is not sufficient HI present, or its distribution is not suitable. Both explanations are problematic : large amounts of HI has definitely been lost from galaxies in this region, and stacking achieves enough sensitivity that a detection is at least plausible. The original mystery remains : where has all the HI gone ?

It has been demonstrated that examination of the flux distribution is easily capable of detecting bright sources, or even relatively faint sources provided a small enough volume is chosen. However it is very difficult indeed to detect faint, diffuse HI in this way. Point sources are visible as tails in the distribution of positive flux. Faint, diffuse HI is merely an overall excess of positive flux, and does not create such tails. While it is in principle possible to measure the net flux in a volume to estimate the amount of HI present, in practise this is unworkable.

In the ideal case, the flux distribution from blank (devoid of HI) pixels would be such that the net and mean flux in a blank volume would be zero. In actuality there is no region of the cube that can be guaranteed to be totally devoid of HI. However the foreground region apparently contains even more flux per pixel than the cluster volume. Since (in VC1) there is only a single HI detection in this region compared to 59 in the redshift range $0 < v < 3{,}000$ km/s it is improbable that this excess flux is due to HI. If it were genuine HI, there should still be substantially more in the cluster volume than outside it, since the cluster also possesses many (HI-deficient) galaxies, whereas the foreground volume does not.

The net flux measured in smaller, apparently blank areas of the cube is also difficult to interpret. All three areas considered have a total integrated flux below the value expected from deficiency and the size of each area. However, the errors of this method are of order 30 Jy km/s while the expected HI flux from deficiency, for the largest area, is of order 100 Jy km/s. This expected flux level assumes the HI is uniformly distributed spatially, though as discussed above it is expected to be clumped. Consequently a factor 3 difference between observed and predicted flux levels is not really sufficient to say whether there truly is more or less HI present than deficiency suggests.

In final conclusion, it is not possible to determine whether the HI lost from galaxies resides in the intracluster medium as HI. Both spatial averaging and examination of the flux distribution have the sensitivity to detect the HI if present, but fail to do so. Regardless of sensitivity, there are many other uncertainties associated with both methods. For spatial averaging, the distribution of the HI must be correct. For the flux distribution, the net flux from truly empty pixels must be zero, but this does not appear to be the case. Therefore these methods are unable to constrain the amount of intracluster HI present.

Chapter 8

Summary, Discussion and Future Work

8.1 Key Results

- There are probably no dark galaxies. A very few candidates are detected, but they comprise no more than 2% of the total sample, or 7% of the cluster detections. They all require more observations to better determine their nature - i.e. whether the HI is associated with any optical emission at all. At best, dark galaxies are very rare in the Virgo Cluster. Although they may be more common in more tranquil environments, the predicted 23% of AGES detections being dark is very unlikely to be correct.
- That said, nearly 30% of the detections with v < 3,000 km/s are not members of the optically selected VCC. These are generally small, faint, blue and gas rich, and are probably entering the cluster for the first time. Most are below the VCC completeness limit and are much easier to detect in HI than an optical survey. These new detections are a small population compared to the VCC : 21 new detections from AGES compared to a total of 249 VCC cluster members.
- A very few early-type galaxies are detected in HI, about 8% of the early-types listed in the VCC in the VC1 region. The S0s detected are broadly similar to spirals, but fainter, redder, and possess less gas. They lie in the transition region on the colour-magnitude diagram, seem to be HI deficient, and some show signs of structure. This suggests that they have morphologically evolved from late-type spirals via gas loss. From stacking, the majority of early-types possess much less gas - the detected objects are probably in the transition stage between late and early-types.
- In contrast the detected dEs seem to be normal dEs in every sense, being faint, morphologically smooth and lying on the red sequence. They are only unusual in their gas content, having comparable M_{HI}/L ratios to dwarf irregulars. Since ALFALFA have found similar objects in the field, this suggests that they are recently accreted field objects.
- The Virgo Cluster is still assembling, with non-deficient dwarfs being interspersed with highly deficient giant spirals. HI deficiency is greater closer to the cluster center, but can still be significant even at the edges of the VCC survey. The HI detections in VC2 may be a separate infalling population to the non-detected VCC galaxies; in VC1 it seems that several merging subclusters are themselves still assembling.
- While gas is definitely being lost from galaxies in significant quantities, probably *in situ*, there is no direct evidence for the fate of the stripped gas. It seems unlikely that it all remains neutral as from stacking, it would have to have a column density below $\sim 5 \times 10^{17}$ atoms cm^{-2} to remain undetected. Those streams that are detected within Virgo are thus exotic rarities. A major caveat is that the sensitivity of the stacking technique depends upon the size and orientation of the stream (see chapter 7 for a detailed discussion).
- The effects of gas loss are not always obvious. Some galaxies with significant HI deficiency are morphologically smooth and relatively blue in colour while others are red but with clear spiral structure. Gas removal cannot always cause instantaneous morphological change, nor is is the solve driver of morphological evolution - as evidenced by stellar tails in early-type galaxies.
- The HIMF of the Virgo Cluster depends on the sensitivity of the survey just as much as the area observed. HI observations of VCC-selected objects will not well reproduce the true HIMF, since even above 3×10^8 M$_\odot$ in HI there are objects not included in the VCC. The environment of the cluster is complex with studies of different areas giving different results. Ultimately, only a deep HI survey of the entire cluster will properly determine its HIMF.

8.2 Dark Galaxies

As detailed in chapters 1 and 5, identifying a dark galaxy is never going to be easy. To recap, it requires first a blind HI survey to detect a candidate (which is due to sheer luck as much as anything), very deep optical imaging to confirm the lack of optical emission, and finally, if the original survey was insufficient, high spatial resolution HI imaging to determine the presence of a rotating disc. Even then, there must be good grounds to reject the notion of a tidal or other non-primordial origin. Clearly, any dark galaxy candidate deserves to be treated with a degree of skepticism, but it is also vital not to reject any candidates out of hand.

Within the cluster environment it must be accepted that a really isolated object, where non-primordial origins can be dismissed, is unlikely to ever be found. As was mentioned in chapter 1, Duc & Bournaud 2008 created a numerical model to attempt to explain the formation of VIRGOHI21. This model relied on an interaction between NGC 4254 and another massive galaxy at a speed of 1,100 km/s. Regardless of the success or otherwise of the model in reproducing this particular observation, it shows that it can be very difficult to eliminate tidal interactions due to a lack of obvious (i.e. nearby) possible interactors. The high speed means that the interacting galaxies become widely separated - in their model, the interactor is now over 1 degree from NGC 4254.

Coincidental alignments, where an HI detection is close to an optical galaxy to which it has no physical association, is another hazard. Even leaving aside other cluster members, virtually any patch of sky, if observed to sufficient depth, will be seen to contain innumerable background galaxies (as demonstrated in spectacular fashion by the Hubble Deep Fields). Interferometric imaging can greatly help to reduce these problems, but without this the AGES candidates cannot be considered as anything more than tentative possibilities.

There are many HI detected galaxies which are non-deficient, which suggests that the cluster environment is not instantly lethal to dark galaxies. Assuming their total mass to be the same as optical galaxies, they should therefore be able to enter the cluster and survive for some time before their gas is removed. Unfortunately this does not guarantee that their gas would be completely unperturbed - tidal interactions might induce star formation without actually removing any gas. Given that simulations have predicted such different effects of ram pressure stripping (see chapter 1), it is still unclear how long an infalling dark galaxy could survive in a cluster environment.

A study of the Virgo Cluster, then, cannot hope to definitively answer whether dark galaxies are common in the Universe or not. Certainly though, they are not common in the cluster. If the Davies et al. 2006 model was correct there should be about 17 HI detections without optical counterparts. There are in fact 13 cluster detections without an available optical redshift measurement, so here the association to the optical is not certain. However, it cannot be coincidence that, in except for a few cases, an obvious possible counterpart is visible within an arcminute or so of the HI coordinates. As shown in chapter 6, these are invariably blue dwarf galaxies, and their similarities make it highly unlikely that they have all been misassigned as optical counterparts.

Disney 2008 notes a caveat to the identification of optical counterparts : galaxies are clustered in redshift as well as position. Two galaxies in the field in close proximity on the sky are thus more likely to also be at similar redshifts than completely separate. By extension, if there is an optical galaxy close to the coordinates of an HI detection, the two may be at similar redshifts despite being physically unrelated. Redshift is not always the final answer as to whether the optical counterpart is really associated with the HI. Disney suggests that there may still be many dark galaxy candidates hidden in the low resolution HIPASS data.

This can be better illustrated by assuming that the distribution of HI detections and optical galaxies is entirely separate and random, and considering the probability of an optical and HI detection matching by chance. Let a match mean that the two are within a 3.5' radius and 200 km/s recessional velocity of each other, so approximate each galaxy to have a 3.5' radius and 200 km/s velocity width. The "volume" in position-velocity space occupied by the VC1 galaxies will therefore be about 0.4% of the VC1 cube for the AGES detections, and 1.4% for the VCC galaxies. The probability of a chance match would thus be about 0.06%.

In actuality of course the percentage of matches is very much higher, about 75%, since the two are not physically unrelated and this is manifested in their similar distributions. Disney's point is that we should be cautious of assuming that all matches are due to a physical association since, in general, galaxies are *not* distributed randomly or evenly in space or velocity. Some of the matches may actually be physically unrelated, allowing for the prospect of considerably more dark galaxies.

Against this, galaxies within the specific environment of clusters have a high velocity dispersion, which may be approximated as a random distribution. Objects a few arcminutes apart, though they may be physically interacting, may have recessional velocities differing by hundreds of km/s or even more. Fortunately, if the gas is really gravitationally bound to the optical component, it must be at a very similar redshift. So if an HI detection is discovered that is close in both space and velocity to an optical galaxy, it is more probable that the two are part of the same galaxy than being unrelated objects.

Consider as an example the 95 VCC cluster members in VC1 with optical redshifts. There is only 1 pair of objects where the spatial difference between each member is less than 3.5' and the velocity difference less than 20 km/s (the mean difference between HI and optical redshifts, see chapter 5). Even using the HIPASS beam size of 14' this only increases to 4 pairs. If dark galaxies are clustered to this level of precision then there is no reason why the optical galaxies should not be also, but such close associations are rare. It is very unlikely that there is any kind of large misidentified population of dark galaxies hidden in the AGES data.

Dark galaxies are rare in the cluster beyond reasonable doubt. What, then, is the significance of the dark galaxy candidates described in chapter 5 ? As discussed the various environmental processes acting within a cluster may not make it a very hospitable environment for their survival, so even if they are dark galaxies they do not constrain their existence or cosmological significance elsewhere. Establishing whether they exist at all is still a crucial step.

As described in chapter 5, the possible optical counterparts which are present for these objects are considered very uncertain because they do not fit with expectations based on the HI profiles. More precisely, their velocity widths would normally be associated with disc objects but the optical candidates do not show obvious disc structure. This is shown more quantitatively in figure 8.1, which plots the distribution of the ratio of the velocity width to the optical diameter (physically, the velocity width will relate to the size of the HI disc and/or the dynamical mass but resolved observations are necessary to determine anything further[1]). The key point is that 5 objects (less than 4% of the sample) clearly stand out as having far higher ratios than would be expected from the rest of the sample.

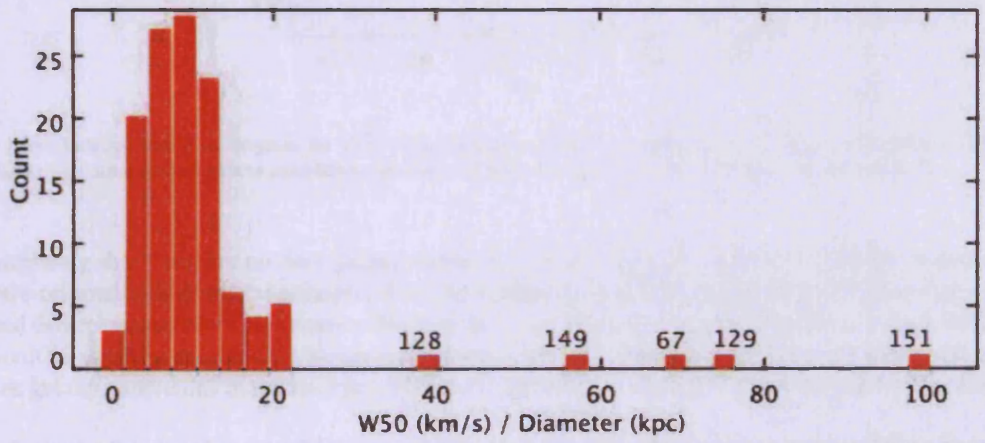


Figure 8.1: Distribution of the ratio of the HI velocity width W50 (in km/s) to the physical optical diameter (in Kpc). The most extreme objects are labelled with their catalogue numbers from chapter 5.

Recall that the 3 suggested dark galaxy candidates were objects 67, 128 and 129. Object 151 was shown to have a strange HI profile in disagreement with previous observations, so requires follow-up. Object 149 is also seen to have an exceptionally high ratio of velocity width to diameter - though the optical candidate resembles many of the non-VCC HI detections (see figure 6.12) its velocity width is too high. The separation between the HI and optical is also rather high at 1.3', so follow-up is also required for this object.

Objects 67, 128 and 129 remain extremely interesting. If the HI is really associated with the purported candidates, then an explanation must be found for their unusual velocity widths. Object 149 may provide a clue - with an

[1]The velocity width is also a measure of the absolute magnitude - the fundamental aspect of the Tully-Fisher relation.

optical/HI separation of 1.2', perhaps the HI is being or has been tidally disrupted, increasing its velocity width. Indeed objects 67 and 128 also have high separations of 0.7 and 1.1' respectively - but this may be only because the HI is not associated with the candidates at all[2]. Still, tidal disruption would certainly be more effective on these low-mass objects.

On the other hand, dark galaxies were predicted to be far more common at low masses, and these candidates have HI masses amongst the lowest in the AGES sample. Their velocity widths are not particularly high, but this is to be expected given their small masses and does not preclude the existence of a dominating dark matter component. Indeed, even if their optical counterparts are real physical associations, the objects remain extremely interesting with such high ratios of velocity width to optical diameter possibly suggesting a high dynamical mass. Additionally, while object 67 is the bluest object in the entire sample, 128 and 129 are clearly red sequence objects (figure 8.2), while 149 and 151 may be transition objects. Further speculation requires more observations, particularly optical spectroscopy.

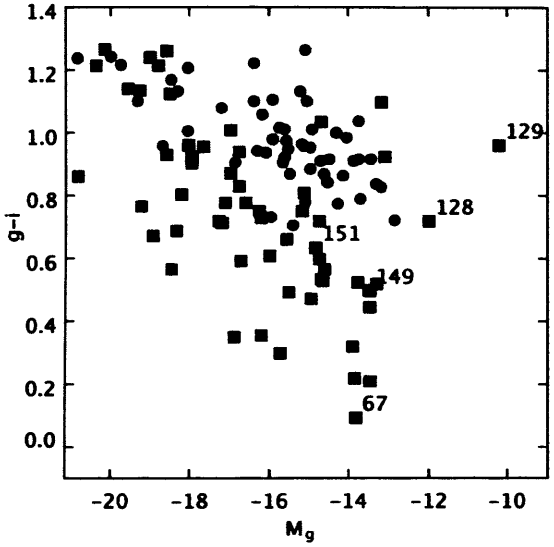


Figure 8.2: Optical colour-magnitude diagram for VC1. Blue squares are AGES detections, red circles are VCC HI non-detections. Black squares highlight the purported dark galaxy candidates, labelled with their number in the AGES catalogue (see chapter 5).

It is not surprising that there are no dark galaxy candidates detected in the volume behind the cluster. A few candidates were originally detected (see chapter 4 for an example), but were rejected by HI follow-up observations. Background detections are more HI-massive due to sensitivity effects, so are less likely to be dark. It also becomes more difficult to identify the optical counterparts since at greater distances they appear fainter and smaller. The lack of dark galaxy candidates in the background cannot be taken as evidence either way for their existence.

Similarly, the lack of dark galaxy candidates in VC2 is probably only because that region of the cluster is less rich than VC1. Assuming the detection rate to be at most 7% of the sample (object 149 might also be included as a candidate, raising the total to 4 candidates out of 60 HI detection in the cluster in VC1), then less than 1 detection is expected in VC2 given the 13 HI detections in that area of the cluster. Source extraction techniques (chapter 4) are not to blame. The cluster detections in VC2 are all high S/N, but like so many dark galaxy candidates, fainter detections in the cluster have faded into nothing on follow-up.

It was mentioned in chapter 1 that some High Velocity Clouds of the Milky Way may have the necessary properties to be dark galaxies and even explain the missing satellites problem. In chapter 6 some of the HVCs found by AGES in the Virgo regions were described and ruled out as extragalactic dark galaxy candidates - it is too much of a coincidence that they are the only extended sources and also among those of the lowest recessional velocities. However, they might still be part of the possible population of "minihalos", tentatively suggested by Giovanelli et al. 2009. The authors note that small HVCs (< 15' HI diameter) of low velocity widths (< 30 km/s) could be

[2]A further caveat is that tidal disruption does not always cause an increase in HI velocity width - VCC 1249, discussed in chapter 6, has a W50 of just 36 km/s.

bound to dark halos in orbit of the Milky Way, and would be small enough that they would not have been detected by existing surveys of objects beyond the Local Group.

Several of the clouds shown in figure 6.20 do satisfy Giovanelli et al.'s criteria, although, as is evident from the figure, a precise estimate of their diameter is difficult in some cases - these clouds are better described as a complex, rather than a collection of discrete objects. This is considerably more evident when viewing the actual 3-dimensional data cube rather than the 2-dimensional moment map, which tends to suppress some of the weaker emission visible between the clouds. The Turn On GALFA Survey (TOGS), which is run commensally with AGES and has 0.2 km/s velocity resolution (Gibson et al. 2008), can correctly process the very bright and extended Milky Way emission, and might be better suited to analysis of the HVCs. Modification of the source extraction techniques described in chapters 4 and 7 could also help to prove if some of the clouds are really discrete entities (and so possible minihalos) or part of larger extended features.

The possible dark galaxy detection rate of 7% is at odds with the Davies et al. 2006 model, which predicted 23% at a 16 Mpc distance (hence the HVCs can be ignored). Either many of the detections have been misidentified, which does not seem likely (but it is still important to obtain optical spectra for those objects where such measurements are lacking), or the model is incorrect. 10 square degrees are still awaiting observations in Virgo, so this fraction may still change. Currently, only 2 other candidates are listed in the other AGES fields completed (see chapter 2), giving an overall detection rate of just 1.7%. This cannot be directly compared to the Davies et al. model, since these fields are studying objects at different distances than the 16 Mpc used for predicting dark galaxy detection rates.

It has been described here and throughout this thesis that dark galaxy candidates are not common, but a few do exist (see chapter 1 for examples of non-AGES detections). This conclusion is supported by other blind HI surveys. Doyle et al. 2005 find that there are no dark galaxy candidates are detected by HIPASS, though for the reasons given in Disney 2008 (see above), Davies et al. 2006 dispute this result as definitive. Doyle et al. also allow optical/HI matches where velocities agree to within 400 km/s, twice that of AGES (though the majority of their matches are to within 100 km/s).

The ALFALFA survey was predicted by Davies et al. to detect dark galaxies at a rate of 4%. From their online database, the current total of ALFALFA detections is 1796, so the survey should already have over 70 candidates if the model is correct. While, unlike HIPASS, some of the detections cannot be assigned an optical counterpart, this detection rate is low though not necessarily inconsistent with the Davies et al. model ("under 3%" - Haynes 2007). More problematically, the ALFALFA observations (and in some cases higher-resolution follow-up observations - Kent et al. 2009) generally indicate that these are formed through tidal and other environmental interactions, rather than being primordial dark galaxies (Giovanelli 2007). Thus, at present, HI surveys have not found a significant population of dark galaxies that could explain the missing satellite problem.

A final note on the issue of dark galaxies in general is that their unambiguous detection (insofar as that is possible) may require even deeper HI observations. The HIMFs shown in chapter 6 indicate that deeper observations can and do detect ever less HI-massive objects. The M_{HI}/L plots shown in chapters 5 and 6 indicate that it is the smallest objects which are the most gas-dominated (even if this is partially a selection effect). Deep, high-resolution observations, ultimately provided by the S.K.A., may be necessary to give the last word on the subject.

8.3 Morphological Evolution in the Cluster Environment

As described in chapter 1, the debate on whether galaxy evolution (particularly with regards to morphology) is dependent only upon the environment at birth (nature) or is more transient and can be strongly influenced by later conditions (nurture) has been ongoing for some time. A myriad of different effects are at work simultaneously - from the gravity of the cluster as a whole to encounters between individual galaxies at various speeds. Gas may be lost within the inner regions of clusters but accreted in the field. These effects must somehow be distinguished from the intrinsic variation in galaxy properties that are not environmentally dependent.

Only a very few early-type galaxies are detected in HI by AGES, about 8% of those present in the VCC. This of course is not surprising : since early-types are not normally star-forming, they should not contain much gas. However, even this small population of objects shows internal variation. Most importantly, the dwarf ellipticals are

fundamentally different objects to the lenticulars.

Firstly, both HI-detected and non-detected S0s are brighter than dEs, with typically M_g < -16.0 while dEs typically have $M_g > -16.0$. At the high mass end, they tend to have similar colours to late-type detections (see figure 8.3) while dEs are more obviously on a separate red sequence. S0s are detected relatively frequently in HI - at least 20% are detected in VC1 by AGES, but possibly as high as 40%[3]. Dwarf ellipticals are far more rarely detected in HI in the cluster, with a detection rate in VC1 no higher than 3%. S0s have M_{HI}/L ratios comparable with (though rather lower than) cluster spirals, whereas the M_{HI}/L ratios of the detected dEs is much higher, being comparable with dwarf irregulars.

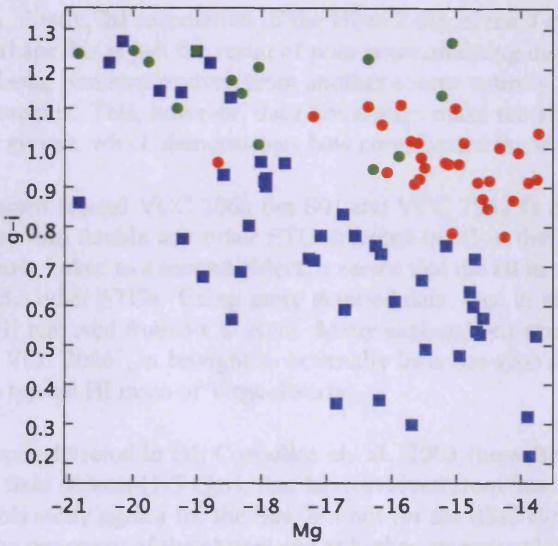


Figure 8.3: Colour-magnitude diagram for objects in VC1. Squares are AGES detections, circles are VCC galaxies. Blue indicates late-type, green are S0s and red dEs.

A comparison with the detection rates of early-types from other studies is not entirely straightforward, as discussed in chapter 5. 3 of the 7 detected ETGs in AGES (42%) are below the ALFALFA sensitivity limit, so, as discussed in chapter 5, the difference in detection rates is likely only due to sensitivity. In GOLDMine, 1,241 ETGs are listed as cluster members, of which just 10% have been observed in HI. Of those, 111 have not been detected, but for nearly 90% of these the observations were not as deep as AGES. 22 ETGs are listed in GOLDMine as detected, but only 10 of these are listed as reliable. This makes it very difficult to say whether the AGES detection rates are consistent with previous observations.

As described in chapter 5, the evidence suggests that the S0s have evolved via gas loss from spirals, whereas the detected dwarf ellipticals are recently accreted field objects which have experienced little of the cluster's effects. The detected S0s are broadly similar to spirals but are fainter and have lower gas fractions (this is significant since fainter objects, on the whole, tend to have *higher* gas fractions) consistent with their being HI deficient, and their rotation velocities imply rotating discs. Two of them (half the sample) show some signs of structure. If they originally had as much gas as spirals, then perhaps they also shared similar morphological characteristics which have faded due to loss of gas and quenching of star formation (see Elmegreen et al. 2002).

In contrast it was shown that the dwarf ellipticals detected in HI are comparable to dwarf irregulars in every way except colour, so it was argued that these are recently accreted field objects that have not yet experienced much influence from the cluster. It may seem contradictory to argue that the brighter S0s are morphologically evolving but the fainter dEs, where gas can be more easily removed, are not. The problem is that while the S0s have less gas than spirals of similar magnitude, the dEs have just as much gas as dwarf irregulars. The velocity widths of the dEs and dIs are similar, so it is not straightforward to explain why one group is clearly star-forming while the other

[3]The morphological classification scheme used (see chapter 3) does not definitively distinguish between lenticulars and ellipticals. All of the detected S0s are type 1 (S0s only), no type 0 (E-E/S0) objects are detected.

is not. Moreover, the current M_{HI}/L of the dEs ranges from 0.6 to 2.0. Since few objects are detected with more gas than this[4], it seems unlikely they have lost a significant proportion of their gas. If that were the case, then since non-deficient objects are detected, a population of even more gas-rich dwarfs should also be detected, but figure 5.16 shows that 2 of the 3 detected dEs have some of the highest M_{HI}/L ratios in the whole sample.

This points to the dEs being field objects, a fact supported by ALFALFA observations of dEs in the field detected in HI. As mentioned in chapter 5, Grossi et al. 2009 find that early-types in general are detected far more frequently in HI in the field. Sample sizes are as yet too small to determine if gas fractions of cluster and field objects are comparable or not, which also applies to S0s. This means that it is still possible that the S0s are also recent additions to the cluster, and only future surveys of field lenticulars will be able to tell if the cluster and field detections have similar gas contents or not.

There are two further caveats. Firstly, the association of the HI to 2 out of the 3 dEs is uncertain, with positional deviations in excess of 2'. Perhaps this is just the result of poor position fitting due to low S/N, or perhaps it indicates that the gas is actually being removed or even from another source entirely. Only observations with higher spatial resolution can determine this. This, however, does not always make the situation any clearer - the second caveat is the VCC 2066/2062 system, which demonstrates how complicated the situation can become.

Extended HI emission is detected around VCC 2066 (an S0) and VCC 2062 (a dE or more probably a dI). The HI mass of the system is more than double any other ETG detected in HI in the VC1 region. Together with the extended nature of the HI, clearly linked to a second object, it seems that the HI in this system is undergoing a very different evolution to that in the other ETGs. Using more detailed data, Duc et al. 2007 suggest that VCC 2062 may have formed out of the HI removed from VCC 2066. Many explanations remain for the origin of the HI - it could be primordial gas from VCC 2066[5], or brought in externally by a late-type massive or dwarf object (though the latter is unlikely given the typical HI mass of Virgo dwarfs).

In an earlier study of early-types detected in HI, Conselice et. al. 2003 (hereafter C03) concluded that they are likely to be recently accreted field objects (1-3 Gyr), that have evolved from later-type galaxies through their interactions with the cluster. This study agrees for the S0s, but not for the dEs. Firstly, C03 infer recent accretion because their objects are on the periphery of the cluster and at higher recessional velocities than most of the VCC galaxies. This is also possible for the AGES detections (see figure 8.4) though neither distribution appears particularly convincing given the small number statistics. A more powerful demonstration of the possibility of recent accretion is shown by the detection of late-type objects with negligible deficiency that cannot yet have experienced the full force of the cluster environment.

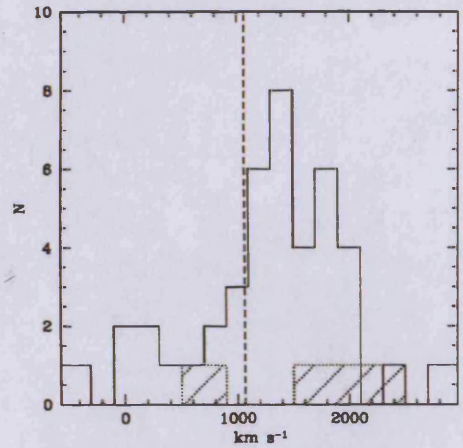
Secondly the authors concluded that the detected ETGs may have evolved from later-type objects infalling from the field, for precisely the reason given in chapter 5 and above that they cannot be - they have as much gas as dwarf irregulars. C03 argued that this meant the detected dE's cannot be the same as Local Group dE's, inferring them to originally have had even higher gas contents. The possibility that C03's detections are of a different population than the AGES detections cannot be ruled out, since the dEs in C03 are somewhat brighter than those detected in AGES. However, given the ALFALFA detection of field dwarf ellipticals in HI at a much higher rate than in the cluster, it now seems unnecessary to invoke morphological evolution of these objects - indeed it would open the question of why even more gas-rich infalling dwarf irregulars are not detected.

Certainly though, the detection rates of S0s and dEs in the cluster are substantially different. Bluntly put, there are a few gas-rich dEs, but many gas-poor (but detected) S0s. This conclusion is partially reinforced by the failure of stacking the non-detected early-types, which are mainly dE's (see chapter 7,) to achieve a detection, indicating that the detected galaxies have M_{HI}/L ratios at least 5 times greater than the non-detections. Also, as described in chapter 5, the lowest HI mass detected in C03 is comparable to the ALFALFA sensitivity limit - despite C03's observations being about 10 times deeper than ALFALFA. C03 did not observe any S0s, and only 5 (definite) S0s were available for stacking by AGES, so it is still possible that many more S0s contain some HI.

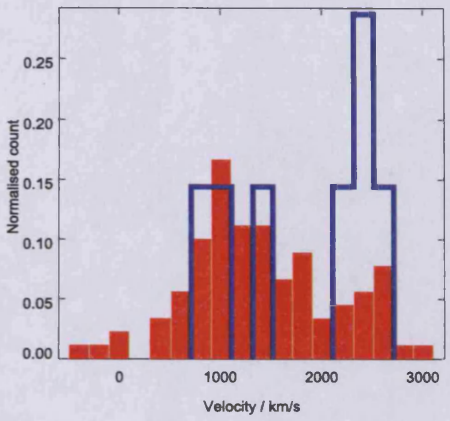
Being in the cluster rules out the possibility of gas accretion. As described in chapter 1, the numerical simulations of Tonnesen, Bryan & Van Gorkum 2007 state that galaxies can only accrete gas beyond 2.4 Mpc from the cluster center. Defining the center of the cluster is not quite so straightforward in reality, but the edges of VC1 are about

[4]Object 30 in VC2, which has more gas yet is seemingly HI deficient, must stand as an interesting and very unusual exception.

[5]Though HI massive, it is also optically bright. Assuming type Sa gives a predicted deficiency of 0.7, so this galaxy may still be losing gas like every other cluster galaxy.



(a) Figure 7 from Conslice et al. 2003, showing the distribution of velocities of dE's observed but not detected in HI (open histogram) and detected in HI (shaded histogram). The vertical line shows the mean velocity of the cluster as estimated by the authors in an earlier paper (Conselice et al. 2001)



(b) Velocities of non-detections (red) compared with HI-detected ETGs (blue) in VC1 from AGES and the VCC. The count for both distributions is normalised to a total of 1.0.

Figure 8.4: Velocity distribution of ETGs from Conselice et al. 2003 (left panel) and from AGES (right panel).

2.1 Mpc from M87 and only 1.5 Mpc from M49 (and as shown in chapter 5 galaxies still show significant deficiencies at the edges of the observed fields). It seems that the dEs have not lost any gas whereas the S0s are deficient (again see chapter 5). If this is so, it raises the following questions : are there any signs in the late-type spirals of morphology evolution via gas loss, since many of these too are deficient, and how do the detected dEs remain on the red sequence while dwarf irregulars with similar gas contents are blue ? The latter requires interferometric observations to better determine the kinematics of the gas - the velocity widths of dEs and dIs are similar (chapter 5).

It would be surprising, if morphological evolution of spirals is really occurring (via gas loss), for the only evidence of this to be a few detected S0s. As spirals lose gas they should migrate from the blue to red sequence or at least the transition region, and also lose structure. The S0s fit these trends, but there might also be evidence in the more ordinary late-types. Indeed, it was shown in chapter 5 that the lower the gas fraction, the redder the colour, though this ignores the effects of colour varying with magnitude. It was also shown that cluster galaxies as a whole are more deficient and redder than background objects, so it seems logical to also search for morphological differences between background and cluster galaxies.

If, to some degree, *all* cluster galaxies are evolving from the blue to red sequence, a simple expectation would be that late-type galaxies closer to the transition region on the colour-magnitude diagram more closely resemble early-types, and vice-versa. Following Kannappan, Guie & Baker 2009 (see figure 5 of chapter 1), figure 8.5 shows the colour-magnitude diagram with SDSS RGB images in place of simple points.

While some individual galaxies also appear to be intermediate between late and early types, there does not appear to be (albeit entirely subjectively) any smooth transition from one class of galaxies to the other. Evidence for migration might perhaps only be found in the transition region which is very narrow in the optical and better seen in the UV (Hughes & Cortese 2009). Also, background spirals have been shown to be much more gas-rich, so loss of gas cannot always cause immediate morphological or colour evolution, and it may not be proportional to the amount of gas lost. As shown in chapter 6, highly deficient galaxies can have very different morphologies.

The is further illustrated by figure 8.6, which plots the M_{HI}/L_g ratio as a function of absolute magnitude in a similar fashion to figure 8.5. If evolution is occurring via gas loss, then since galaxies in the cluster tend to be redder and with less gas than background objects, morphological differences might be seen between the two populations. No such trend is readily apparent. In chapter 6 it was suggested that if gas loss exceeds a threshold then no further

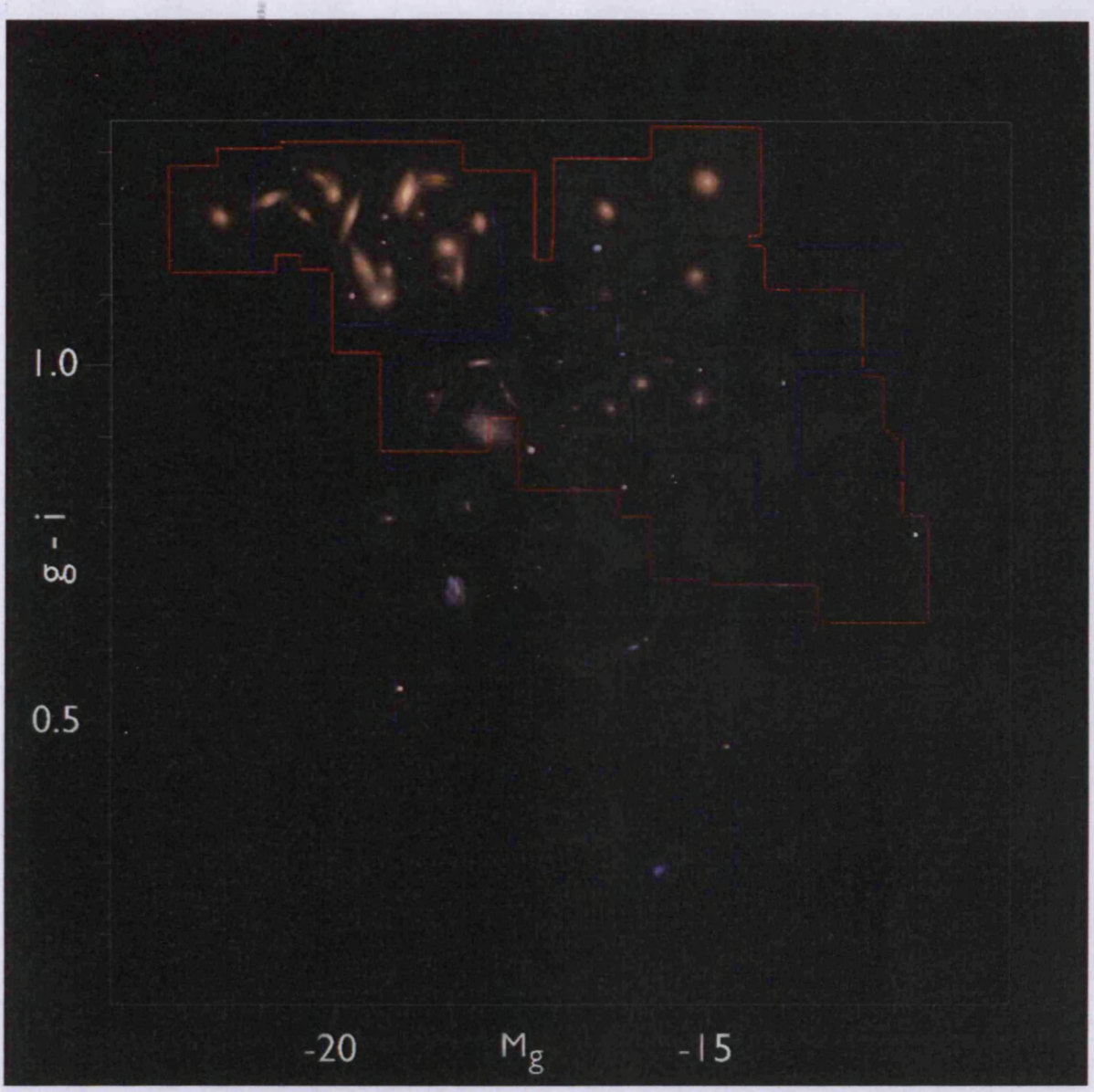


Figure 8.5: Colour-magnitude diagram for objects in VC1 with galaxies rendered as SDSS RGB images. HI detections are outlined in blue, non-detections in red. All images are the same fixed size regardless of actual galaxy size. A more conventional version of this plot can be found in chapter 5 figure 14.

evolution occurs. Perhaps the reverse is also true - gas loss may need to be very substantial before changes become obvious. A more quantitative analysis is necessary, but must be left to future research.

As mentioned in chapter 1, while individual objects may show signs of morphological evolution, understanding whether the process is widespread is more difficult. As discussed, the HI-detected S0s all lie in the transition region of the colour-magnitude diagram. A combination of UV and H-band data, following Hughes & Cortese 2009, would separate the red and blue sequences far more clearly, allowing, for example, analysis of whether transition objects are the most deficient.

The evidence presented in this thesis is simply not sufficient to say whether all S0s are faded spirals - there are, after all, only 4 detected S0s in this sample. As noted in chapter 1, Boselli & Gavazzi 2006 state that only faint, disc-dominated S0s may have evolved from spirals via gas loss, as the different bulge-to-disc ratios are difficult to explain by ram pressure stripping. However, simulations have shown that the stellar disc can be strongly affected by the gas removal (Kapferer et al. 2009, Smith 2009) but have yet to show whether the observed bulge-to-disc ratios can be produced by this process. Additionally, the high detection rate of the S0s of at least 20% suggests

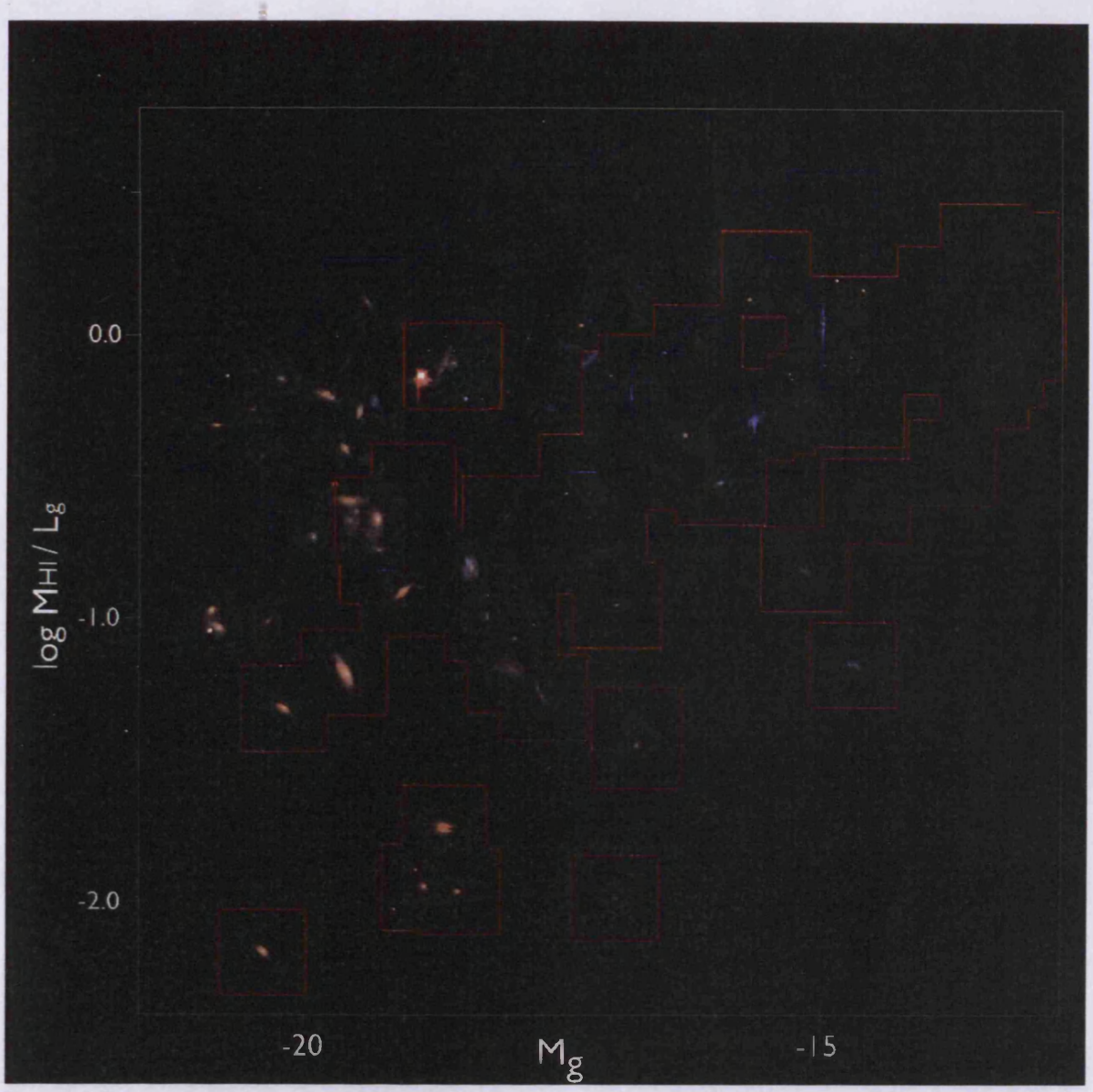


Figure 8.6: Variation of M_{HI}/L_g ratio as a function of absolute magnitude, for objects in VC1 with galaxies rendered as SDSS RGB images. Virgo detections are outlined in red, background detections in blue. All images are the same fixed size regardless of actual galaxy size. A more conventional version of this plot can be found in chapter 5 figure 16.

that morphological evolution is at least not an uncommon process for these objects.

In chapter 1 it was mentioned that some objects which are seemingly in a transition between morphological types look "odd and forced - like a dog wearing shoes", to quote Dressler 2004, referring to NGC 4650A. Some of the objects described in Cortese & Hughes 2009 surely also qualify for such a description (see figure 8.7). These objects have been suggested as possibly early-types that have been rejuvenated by gas capture. The S0s detected by AGES, suggested to indicate evolution in the other direction, do not at all resemble these peculiar systems, as shown by figure 5.9.

Nor could the S0s detected by AGES be described as blue sequence objects, as those in Kannappan, Guie & Baker 2009 (see figure 1.5) clearly are. Rather, as discussed, they appear to be the result of spirals having lost most of their gas. Despite the objections raised in Boselli & Gavazzi 2006, Boselli et al. 2006 find that S0s can be formed by ram pressure stripping of spirals, though other effects must also be at work (ram pressure cannot account for gas deficiency at the cluster periphery, for example). They predict that the colour gradients of galaxies transforming in this way should be reversed with respect to normal galaxies - gas is removed from the outside-in, so the galaxy

(a) Hubble Heritage image of NGC 4650A

(b) SDSS RGB image of NGC 5701

Figure 8.7: Objects possibly evolving from early to late types, as cited by Dressler 2004 and Hughes & Cortese 2009.

remains star forming for longer in the center. A future analysis of the surface brightness profiles could thus reveal if the S0s detected by AGES have been transformed in this way.

It has been mentioned in chapters 5 and 6 that there are some galaxies detected with negligible deficiencies, while some are highly deficient. The interpretation of this is that the non-deficient objects must be recent arrivals, since many of these are dwarfs, and simulations show that dwarfs will be entirely stripped of hydrogen on their first orbit (e.g. Smith 2009). Understanding if this view is correct has important consequences. If the environment is really an important factor in galaxy evolution, then dwarf galaxies should be more strongly affected than giants (Bosseli & Gavazzi 2006).

Observationally it is clear that the environmental effects can remove virtually all of the gas in giant spirals, with some galaxies in VC1 as deficient as 1.8 (even VCC 222, with an absolute g-magnitude of -20.4, has a deficiency of 1.6). If the cluster can have this dramatic an effect on the giant spirals, the simplest explanation for the presence of non-deficient dwarfs is that they are recent arrivals. Furthermore, some dwarf galaxies (such as object 30 in VC2) do seem to be deficient, so the cluster is affecting both dwarfs and giants. Conversely, some giants are non-deficient - for instance, VCC 975 is spiral with an absolute g-magnitude of -18.5 detected in VC1 with an HI deficiency of 0.1. The most obvious picture is one of a cluster still assembling, with some galaxies having yet to experience the full force of the cluster environment, while others have only the last vestiges of their original gas content remaining.

A caveat is that the HI deficiency is not well-calibrated for dwarf galaxies, so their HI deficiencies estimates may be wrong. Numerical simulations and analytic modelling suggest that dwarfs are the most vulnerable to gas stripping (see chapter 1 and Boselli & Gavazzi 2006), but as described in chapter 1 the results of simulations can be highly model dependent. Only observations can ultimately determine whether gas can survive in cluster dwarfs or not. If, for instance, their orbits never take them close to the cluster core they would be far less susceptible to the stripping mechanisms described in chapter 1 (if, more strangely, they manage to retain their gas despite passage through the cluster center then something is very wrong with the theoretical models).

As mentioned in chapter 6, Lee et al. 2003 present a more complex method to determine HI deficiency in late-type dwarf galaxies by measurements of the oxygen abundance (their definition of "dwarf" is, however, somewhat liberal - their faintest object has M_B = -15.3, their brightest, M_B = -19.0). Their sample is small (11 galaxies) but they find that some are highly deficient (deficiency $\simeq$ 1.5), with 4 being non-deficient (deficiency $<$ 0.4). Using the conventional method of estimating deficiency, the late-type objects in chapters 5 and 6 have a similar range of estimated HI deficiency.

It is also significant that there is a near-total lack of late-types which are actually undetected in HI, though some late-type objects are found with very high HI deficiencies. Morphological evolution from late to early-types could explain this - given the high deficiencies of some detections, it seems unlikely that no object in the cluster has

yet lost all of its gas - particularly given simulations which show dwarfs lose all their gas on a single orbit. It was also shown in chapter 1 that, on the large scale, HI deficient late-types better trace the structure of the clouds comprising the cluster (which are also more clearly seen in the distribution of early-types) than the non-deficient objects. Unlike Dressler's view expounded in chapter 1, the conclusion here is that environment does play a role in contemporary galaxy evolution, and gas loss is a significant part of this process.

The AGES results certainly do not explain the origin of all S0s by morphological evolution or all dEs by recent field accretion. Indeed the latter seems highly unlikely given the very low detection rate of dEs in HI, and the question of why these objects are not late-type remains unanswered. This is particularly surprising given that virtually all of the new detections in the cluster are clearly very blue late-type dwarfs (see figure 6.12). Likewise high HI deficiency does not always equate with an earlier-type morphology, so the formation of S0s may be considerably more complicated than simple gas loss.

Besides the presence of HI, there are other reasons why dwarf ellipticals might be evolved from later-type objects - even if the particular examples cited here have not. Given that over 1,100 objects are listed in the VCC as type -3 to -1 (dS0 - dE), so the fact that 3 of them may be gas rich (a small fraction even considering the total number present in the areas surveyed by AGES) does not necessarily indicate a great deal about the overall evolution of the population. Furthermore, it has already been described in chapter 1 that some simulations (Smith et al. 2009) suggest that morphological transformation can result from gas stripping, and some dwarf ellipticals show signs of spiral structure (Lisker et al. 2006). The issue is highly complex, and a detailed analysis is given in Boselli et al. 2008 (for instance, their modelling includes not only HI content, but also stellar populations, UV colours, metallicity and morphological properties). They conclude that ram pressure stripping can cause such morphological evolution.

In summary the S0s detected by AGES are evidence of morphological evolution of spiral galaxies, a conclusion backed by previous modelling. Future analysis of surface brightness profiles may be able to show in more detail the mechanism by which evolution occurs. The detected dEs do not appear to show morphological evolution, but cannot rule out that this occurs in the general population of dEs.

8.4 The Fate of the Stripped Gas

There is no doubt that gas in cluster galaxies is being depleted. There is also no doubt that in some cases the gas is removed into the intracluster medium and survives as long, neutral, detectable streams. What is still uncertain is whether these streams are highly unusual features or if they are very common, with observations simply lacking sufficient sensitivity and resolution to detect the majority of such features. The results of chapter 7 argue for the former, though with many caveats.

There are three key issues here. The first is whether gas is ordinarily removed into long streams or not. The second is whether the gas normally remains neutral or is rapidly ionised. The third is whether the gas, if it remains neutral and in a stream or other contiguous feature, is sufficiently dense as to be readily detectable by an HI survey. In short, do the streams which can be detected actually exist ?

The number of very extended HI features in the Virgo Cluster is not large, and no survey of any great sensitivity has yet covered its entirety. The current census of such features (those spanning more than 3.5' and thus in principle detectable by AGES) includes VIRGOHI21, the 500 Kpc stream of Koopmann et al. 2008, the complex of 5 clouds described in Kent et. al. 2009 (discovered by Kent et al. 2007), 3 contiguous clouds described in Tripp 2008 and Briggs & Barnes 2006, three galaxies in Chung et al. 2007 (VCC 979, the VCC 491/465 pair, and VCC 1987), and the HI associated with VCC 2066/2062 (see figure 8.8). Over the 140 square degrees of the Virgo Cluster, this is an average of just 0.1 features per square degree. With 15 square degrees surveyed by AGES, the lack of a direct detection is consistent with previous results.

This is further emphasised by ALFALFA, being the most sensitive large-area survey to date, and which has covered most of the cluster, but has not found extended HI to be common. It is not at all surprising that AGES makes no new such discoveries having only surveyed $\sim$10% of the cluster. So the first issue appears to be resolved : extended HI features are rare at current sensitivity levels. Nevertheless, gas is being depleted within cluster galaxies, and unless it is all consumed in star formation, some of it must be removed into the intracluster medium. The possibility remains that extended HI is more common at low column densities, in analogy to low surface brightness stellar

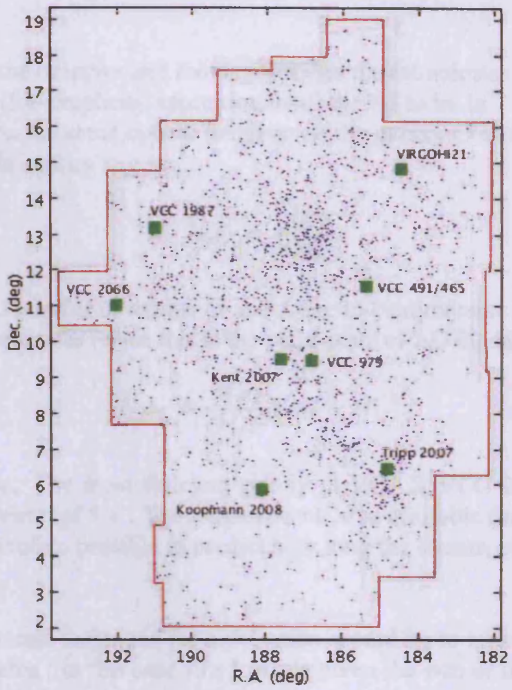


Figure 8.8: Map of the Virgo Cluster showing VCC galaxies as blue points, with green squares highlighting the approximate centers of known extended HI features spanning more than 3.5'.

features of Mihos et. al. 2005.

The second issue - whether the HI is rapidly ionised or not - is far more complex. The clouds described above demonstrate that extended HI can exist for some period in some cases. Their survival will depend on their intrinsic features (volume density, self-gravity and velocity dispersion) and their environment (interactions with other galaxies and hot X-ray gas). Vollmer et. al. 2001 try to account for these factors, and find that only about 10% of stripped gas remains atomic, and then only if of sufficient density. Their modelling is primarily concerned with the survival of small ($\sim$10 pc) spherical clouds and likely invalid when considering much larger, asymmetric structures.

HI can at least exist in low column density streams below the threshold of star formation - 2.5×10^{19} atoms cm^{-2} in the case of VIRGOHI21 (Minchin et al. 2007), and 2×10^{18} atoms cm^{-2} in clouds around M33 (Grossi et. al. 2008). Minchin 2001 found evidence for a column density threshold of HI within galaxies of about 2×10^{20} atoms cm^{-2}, though of course this does not preclude the existence of much lower densities caused by gas stripping. VIRGOHI21 is also evidence that very long streams can be created from relatively small amounts of gas, only a few % of the mass in NGC 4254 (assuming that galaxy to be the source of the gas). So even galaxies of relatively low deficiency might have associated extended HI features.

Through spatial averaging, the column density sensitivity limits of a few times 10^{17} atoms cm^{-2} quoted in chapter 7 mean the technique is exploring a new level of column density sensitivity. ALFALFA quote a similar column density limit (Grossi et al. 2008) without using any stacking techniques, but this is an absolute 1σ limit for a mass of gas spanning a single channel - in practise this is wholly undetectable. The limits quoted in chapter 7 are intended as real-world detection limits (if approximate), which no survey has yet reached directly.

Such high sensitivity comes at a suitably high price. Low column density features are only detectable provided those features are sufficiently large, uniform (continuous over their spatial extent), and (for a given gas mass) without much spread in velocity. The assumption of uniformity does not seem unreasonable given the known HI clouds and streams[6]. The velocity spread is more variable, with the stream in Koopmann et. al. 2008 only spanning about 100 km/s, whereas the complete VIRGOHI21 complex spans 500 km/s. However, given the estimates of HI deficiency, it is possible to constrain the kinds of clouds that could be produced by such gas loss that would be

[6]Although they are by no means perfectly uniform or contiguous, they are very far from being the HI "fluff" described in chapter 7 that would render them completely undetectable.

detectable by stacking.

Suppose a galaxy is edge-on to the observer and moving through the intracluster medium only across the plane of the sky relative to the observer (for simplicity approximate all the HI to be in 1 channel). Assume for simplicity that gas is stripped into a column the same optical diameter of the galaxy O along a length L of column density N_{HI}, with total mass M_{HI}. This implies :

$$N_{HI} = \frac{M_{HI}}{OL} \tag{8.1}$$

Where M_{HI} is actually the total number of atoms; O and L are in centimeters. It is more useful to have M_{HI} in solar masses and O and L in arcminutes (since this is the pixel scale of AGES data). The equation then becomes :

$$N_{HI} = \frac{M_{HI}}{OLd^2} 1.47 \times 10^{15} \tag{8.2}$$

Where d is the distance in Mpc. The most deficient galaxy in VC1 is VCC 222[7], with a missing HI mass of 5.4×10^9 $M_\odot$ and an optical diameter of 5.1'. The longest blank area available for stacking in VC1 had a length of 73'. Using equation 8.2, it is therefore possible to predict how long the stream of HI would need to be in order to escape detection by stacking.

As has been discussed, the optimum technique for a detection would be to average only those pixels containing the stream within the available area : in this case 73×5 pixels given the size of the area and the source producing the stream. Equation 3 of chapter 7 predicts a column density sensitivity of 7.4×10^{17} atoms cm^{-2} for this size area. Using equation 8.2, to reach this column density the stream from VCC 222 would have a length of 37 Mpc, a structure far larger than the Virgo Cluster itself. Equally, if for some reason the HI was distributed in a thin (i.e. again 1 channel velocity width) disc it would have to have a radius of 540 Kpc - about 1.8 degrees. Spatial averaging is thus able to detect, in principle, far more diffuse structures than can actually be present.

To be absolutely clear, it is important to note that this is not a minimum size limit the HI feature must have in order to be detectable by stacking. Stacking can detect features equivalent to this column density, but of course could detect smaller or denser features. While the HI may be in a stream many megaparsecs in length, stacking need only be performed on a very small fraction of it to detect the HI.

The velocity width of the HI is of course very significant, as shown in equation 7 of chapter 7. Assume the same stream is produced as before, but this time let its velocity width be 100 km/s - 10 channels (strictly speaking resolution elements). A length of 73' available for stacking will only enclose about 1% of the whole stream, so the total M_{HI} being stacked will be 5.4×10^7 $M_\odot$. Stacking an area of 73×5 pixels, equation 7 predicts a S/N of 2.0, which is undetectable.

This is perhaps self-evident - since the column density sensitivity is calculated as a limit, anything that would reduce this will necessarily render the gas undetectable. Consider instead a more physical example. The median amount of missing HI from each galaxy in VC1 is about 5×10^8 $M_\odot$. Suppose this is all lost at an even rate on a single pass of length 1.5 Mpc, roughly the projected length from the center of VC1 to the edge in R.A. Assume the velocity of the galaxy changes by 1,000 km/s, approximately the velocity width of each sub-cluster (see chapter 5). Each channel will thus contain 5×10^6 $M_\odot$ of HI.

It is not really suitable to use equation 7 of chapter 7 in this case as it assumes a uniform cloud geometry. Given the discrete nature of pixels and channels, observations would show the stream tracing a staircase pattern in velocity space - see figure 8.9. This limits the number of pixels which can be averaged. Since the galaxy has moved through 100 channels, the stream will be in 100 segments each of mass 5×10^6 $M_\odot$. Since there are about 300 pixels from one edge of VC1 to the center, each segment will only span 3×5 pixels, assuming the galaxy to still be of 5' diameter. Spatially averaging this few number of pixels will barely increase sensitivity at all.

However, since the HI is distributed over such a narrow velocity range in each segment, each would have a S/N of 14 and thus be easily detectable by eye. More realistically, the HI in each segment will span a velocity range comparable to that of the galaxy from which it is stripped. This only need exceed 30 km/s before the S/N drops

[7]Actually it's VCC 1555 but the optical center is slightly outside VC1, so some HI may have been missed, and also close to face-on to the observer. VCC 222 is edge-on and entirely inside the cube.

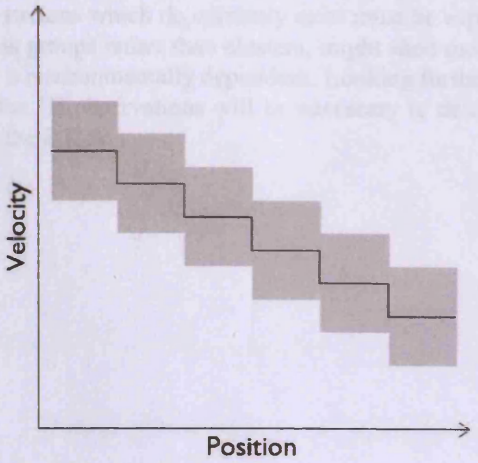


Figure 8.9: Schematic to indicate how a stream would appear in position-velocity space in a data cube. The solid line shows the path of a galaxy losing hydrogen. The grey area shows the appearance of a stream if its velocity width is 3 times the channel width. It is only possible to spatially average (in any useful way) over individual segments, so although the total mass may be large it may still not be possible to detect the stream.

below 5, and above 150 it will be less than 1. Factors of a few (i.e. less than 10) in terms of stream length, gas mass and galaxy width can greatly affect both whether a stream is detectable at all as well as whether spatial averaging will offer any benefit.

Imposing very robust constraints requires much further work. Although the cluster has a projected span ~3 Mpc at 17 Mpc distance, very massive galaxies may require multiple passes to completely remove their gas (Smith 2009). A linear 37 Mpc stream as described above as may seem unlikely, but this is little more than 4 circular orbits at the edge of the cluster. Additionally, the velocity of a galaxy through the intracluster medium will vary along its orbit, giving a varying density of the stripped gas. Variations in the intracluster medium will cause the same effect, but also cause varying amounts of ionisation. Tidal disruption of streams by other galaxies is also an important factor that is difficult to account for analytically. At best, any analytic model can certainly be no more than a ballpark estimate of whether low density streams should be detectable or not.

Assessing whether the low-density streams formed by gas removal are actually detectable is a formidable challenge, likely requiring numerical simulations. Vollmer et. al. 2001 were able to analytically estimate that only 10% of the stripped gas remains neutral, but this was for the specific case of a circular cloud of 10 pc radius, very different to the streams that are actually detected. While the results of chapter 7 do rule out streams of low velocity width, these are likely very rare, and they cannot at all constrain streams of high velocity width.

An automatic program is described in chapter 7 which averages many thousands of different areas of different sizes, shapes and positions within selected areas of the data cubes. This is no doubt better than manually averaging a few targeted areas, but the same reasons as above still prevent a guaranteed detection of any extended HI. Since these areas are necessarily free of any detected galaxies, this also makes it less likely that there are any HI streams present within these areas. Although the automatic program does find detections, it detects similar numbers of similar clouds (i.e. size and S/N) at all redshifts, with no relation to galaxy density. They are likely only statistical artifacts due to the huge number of areas and channels searched.

The data reduction techniques have been shown not to prevent the detection of extended features, but the HVCs described in chapter 6 prove this conclusively - as does the detection of the HI complex around VCC 2066/2062. Methods of examining the distribution of the flux have also been described - i.e. summing the flux in every pixel in a volume devoid of obvious galaxies and looking for a net positive flux or a deviation from Gaussian distribution. The idea is to detect HI below the ordinary sensitivity limit, albeit at the expense of position and velocity information. However, the errors intrinsic to this technique appear prohibitive. Bright sources can easily be seen by examining the flux distribution, but even small deviations from Gaussianity can overwhelm faint sources.

Despite the many complications, if Vollmer et. al. are correct it seems that gas stripped in the cluster does not

remain neutral for long. The elephant in the room is the hot X-ray gas - is this where all the stripped gas goes ? If so then the survival of those HI streams which do currently exist must be explained. In the short term, studies of less hostile environments, such as groups rather than clusters, might shed more light on whether the creation and survival of extended HI features is environmentally dependent. Looking further ahead, as with dark galaxies much more sensitive and high-resolution HI observations will be necessary to detect very low-density streams, which again is a capability afforded by the S.K.A.

Bibliography

Agertz, O., Moore, B., Stadel, J., et al. 2007, MNRAS, 380, 963
Andreon, S. 1998, ApJ, 501, 533
Auld, R., Minchin, R. F., Davies, J. I., et al. 2006, MNRAS, 371, 1617
Barnes, D. G., Staveley-Smith, L., de Blok, W. J. G., et al. 2001, MNRAS, 322, 486
Bekki, K., Couch, W. J., & Shioya, Y. 2002, ApJ, 577, 651
Bell, E. F., Zheng, X. Z., Papovich, C., et al. 2007, ApJ, 663, 834
Biermann, P., Clarke, J. N., & Fricke, K. J. 1979, A&A, 75, 7
Binggeli, B., Sandage, A., & Tammann, G. A. 1985, AJ, 90, 1681
Blitz, L., Spergel, D. N., Teuben, P. J., Hartmann, D., & Burton, W. B. 1999, ApJ, 514, 818
Boselli, A., Boissier, S., Cortese, L., & Gavazzi, G. 2008, ApJ, 674, 742
Boselli, A., Boissier, S., Cortese, L., et al. 2006, ApJ, 651, 811
Boselli, A. & Gavazzi, G. 2006, PASP, 118, 517
Bottema, R. 2003, MNRAS, 344, 358
Bournaud, F. 2010, Advances in Astronomy, 2010
Briggs, F. H. & Barnes, D. G. 2006, ApJ, 640, L127
Broeils, A. H. & Rhee, M. 1997, A&A, 324, 877
Brosch, N., Kniazev, A. Y., Moiseev, A., & Pustilnik, S. A. 2010, MNRAS, 401, 2067
Catinella, B., Haynes, M. P., Giovanelli, R., Gardner, J. P., & Connolly, A. J. 2008, ApJ, 685, L13
Cannon, J. M., Salzer, J. J., & Rosenberg, J. L. 2009, ApJ, 696, 2104
Cayatte, V., Kotanyi, C., Balkowski, C., & van Gorkom, J. H. 1994, AJ, 107, 1003
Cayatte, V., van Gorkom, J. H., Balkowski, C., & Kotanyi, C. 1990, AJ, 100, 604
Chung, A., van Gorkom, J. H., Kenney, J. D. P., & Vollmer, B. 2007, ApJ, 659, L115
Conselice, C. J., Gallagher, III, J. S., & Wyse, R. F. G. 2001, ApJ, 559, 791
Conselice, C. J., O'Neil, K., Gallagher, J. S., & Wyse, R. F. G. 2003, ApJ, 591, 167
Cortese, L., Bendo, G. J., Isaak, K. G., Davies, J. I., & Kent, B. R. 2010, MNRAS, 403, L26
Cortese, L. & Hughes, T. M. 2009, MNRAS, 400, 1225
Cortese, L., Minchin, R. F., Auld, R. R., et al. 2008, MNRAS, 383, 1519
Cowie, L. L. & Songaila, A. 1977, Nature, 266, 501
Davies, J., Minchin, R., Sabatini, S., et al. 2004, MNRAS, 349, 922
Davies, J. I., de Blok, W. J. G., Smith, R. M., et al. 2001, MNRAS, 328, 1151
Davies, J. I., Disney, M. J., Minchin, R. F., Auld, R., & Smith, R. 2006, MNRAS, 368, 1479
Davies, J. I., Roberts, S., & Sabatini, S. 2005, MNRAS, 356, 794
de Vaucouleurs, G. 1961, ApJS, 6, 213
di Serego Alighieri, S., Gavazzi, G., Giovanardi, C., et al. 2007, A&A, 474, 851
Disney, M. 2008, in IAU Symposium, Vol. 244, IAU Symposium, ed. J. Davies & M. Disney, 1-6
Doyle, M. T., Drinkwater, M. J., Rohde, D. J., et al. 2005, MNRAS, 361, 34
Dressler, A. 1980, ApJ, 236, 351
Dressler, A. 2004, in Clusters of Galaxies: Probes of Cosmological Structure and Galaxy Evolution, ed. J. S. Mulchaey, A. Dressler, & A. Oemler, 206
Dressler, A., Smail, I., Poggianti, B. M., et al. 1999, ApJS, 122, 51
Duc, P. & Bournaud, F. 2008, ApJ, 673, 787
Duc, P., Braine, J., Lisenfeld, U., Brinks, E., & Boquien, M. 2007, A&A, 475, 187
Elmegreen, D. M., Elmegreen, B. G., Frogel, J. A., et al. 2002, AJ, 124, 777
Ferrarese, L., Côté, P., Jordán, A., et al. 2006, ApJS, 164, 334
Freudling, W., Catinella, B., Calabretta, M., et al. 2008, in American Institute of Physics Conference Series, Vol. 1035, The Evolution of Galaxies Through the Neutral Hydrogen Window, ed. R. Minchin & E. Momjian, 242-245
Gavazzi, G., Boselli, A., Donati, A., Franzetti, P., & Scodeggio, M. 2003, A&A, 400, 451

Gavazzi, G., Boselli, A., Scodeggio, M., Pierini, D., & Belsole, E. 1999, MNRAS, 304, 595
Gavazzi, G., Giovanelli, R., Haynes, M. P., et al. 2008, A&A, 482, 43
Gibson, S. J., Douglas, K. A., Heiles, C., et al. 2008, in American Institute of Physics Conference Series, Vol. 1035, The Evolution of Galaxies Through the Neutral Hydrogen Window, ed. R. Minchin & E. Momjian, 249-251
Giovanelli, R. 2007, Nuovo Cimento B Serie, 122, 1097
Giovanelli, R. & Haynes, M. P. 1985, ApJ, 292, 404
Giovanelli, R., Haynes, M. P., Kent, B. R., & Adams, E. A. K. 2010, ApJ, 708, L22
Giovanelli, R., Haynes, M. P., Kent, B. R., et al. 2005, AJ, 130, 2598
Giovanelli, R., Haynes, M. P., Kent, B. R., et al. 2007, AJ, 133, 2569
Grcevich, J. & Putman, M. E. 2009, ApJ, 696, 385
Grossi, M., di Serego Alighieri, S., Giovanardi, C., et al. 2009, A&A, 498, 407
Grossi, M., Giovanardi, C., Corbelli, E., et al. 2008, A&A, 487, 161
Haynes, M. P. 1981, AJ, 86, 1126
Haynes, M. P. 2007, Nuovo Cimento B Serie, 122, 1109
Haynes, M. P. & Giovanelli, R. 1984, AJ, 89, 758
Haynes, M. P., Giovanelli, R., & Kent, B. R. 2007, ApJ, 665, L19
Henning, P. A., Springob, C. M., Day, F., et al. 2008, in American Institute of Physics Conference Series, Vol. 1035, The Evolution of Galaxies Through the Neutral Hydrogen Window, ed. R. Minchin & E. Momjian, 246-248
Hopkins, P. F., Somerville, R. S., Cox, T. J., et al. 2009, MNRAS, 397, 802
Irwin, M. & Lewis, J. 2001, New Astronomy Review, 45, 105
Kannappan, S. J., Guie, J. M., & Baker, A. J. 2009, AJ, 138, 579
Kantharia, N. G., Rao, A. P., & Sirothia, S. K. 2008, MNRAS, 383, 173
Kapferer, W., Sluka, C., Schindler, S., Ferrari, C., & Ziegler, B. 2009, A&A, 499, 87
Karachentsev, I., Musella, I., & Grimaldi, A. 1996, A&A, 310, 722
Kenney, J. D. P., Tal, T., Crowl, H. H., Feldmeier, J., & Jacoby, G. H. 2008, ApJ, 687, L69
Kennicutt, Jr., R. C. 1989, ApJ, 344, 685
Kent, B. R., Giovanelli, R., Haynes, M. P., et al. 2008, AJ, 136, 713
Kent, B. R., Giovanelli, R., Haynes, M. P., et al. 2007, ApJ, 665, L15
Kent, B. R., Spekkens, K., Giovanelli, R., et al. 2009, ApJ, 691, 1595
Khochfar, S. & Burkert, A. 2003, ApJ, 597, L117
Kilborn, V. A. 2002, in Astronomical Society of the Pacific Conference Series, Vol. 276, Seeing Through the Dust: The Detection of HI and the Exploration of the ISM in Galaxies, ed. A. R. Taylor, T. L. Landecker, & A. G. Willis, 80
Klypin, A., Kravtsov, A. V., Valenzuela, O., & Prada, F. 1999, ApJ, 522, 82
Knapp, G. R., Kerr, F. J., & Williams, B. A. 1978, ApJ, 222, 800
Kniazev, A. Y., Grebel, E. K., Pustilnik, S. A., et al. 2004, AJ, 127, 704
Koopmann, R. A., Giovanelli, R., Haynes, M. P., et al. 2008, ApJ, 682, L85
Koposov, S. E., Yoo, J., Rix, H., et al. 2009, ApJ, 696, 2179
Kovac, K., Oosterloo, T. A., & van der Hulst, J. M. 2005, in IAU Colloq. 198: Near-fields cosmology with dwarf elliptical galaxies, ed. H. Jerjen & B. Binggeli, 351-354
Kroupa, P., Theis, C., & Boily, C. M. 2005, A&A, 431, 517
Kuhlen, M., Diemand, J., Madau, P., & Zemp, M. 2008, Journal of Physics Conference Series, 125, 012008
Larson, R. B., Tinsley, B. M., & Caldwell, C. N. 1980, ApJ, 237, 692
Lee, H., McCall, M. L., & Richer, M. G. 2003, AJ, 125, 2975
Leroy, A. K., Walter, F., Brinks, E., et al. 2008, AJ, 136, 2782
Lisker, T., Grebel, E. K., & Binggeli, B. 2006, AJ, 132, 497
Mateo, M. L. 1998, ARA&A, 36, 435
Mei, S., Blakeslee, J. P., Côté, P., et al. 2007, ApJ, 655, 144
Mihos, J. C., Harding, P., Feldmeier, J., & Morrison, H. 2005, ApJ, 631, L41
Milgrom, M. 1983, ApJ, 270, 365
Minchin, R., Davies, J., Disney, M., et al. 2005, ApJ, 622, L21
Minchin, R., Davies, J., Disney, M., et al. 2007, ApJ, 670, 1056
Minchin, R. F. 2001, PhD thesis, Department of Physics and Astronomy, Cardiff University, 5 The Parade, Cardiff, Glamorgan, CF24 3YB, Wales, UK
Moore, B., Ghigna, S., Governato, F., et al. 1999, ApJ, 524, L19
Moore, B., Katz, N., Lake, G., Dressler, A., & Oemler, A. 1996, Nature, 379, 613
Morgan, H. R. & Oort, J. H. 1951, Bull. Astron. Inst. Netherlands, 11, 379
Moster, B. P., Macciò, A. V., Somerville, R. S., Johansson, P. H., & Naab, T. 2010, MNRAS, 403, 1009

Nichols, M. & Bland-Hawthorn, J. 2009, ApJ, 707, 1642
Nulsen, P. E. J. 1982, MNRAS, 198, 1007
Putman, M. E., de Heij, V., Staveley-Smith, L., et al. 2002, AJ, 123, 873
Quinn, P. J. 1984, ApJ, 279, 596
Rosenberg, J. L. & Schneider, S. E. 2002, ApJ, 567, 247
Roškar, R., Debattista, V. P., Quinn, T. R., Stinson, G. S., &Wadsley, J. 2008, ApJ, 684, L79
Sabatini, S., Davies, J., Scaramella, R., et al. 2003, MNRAS, 341, 981
Sabatini, S., Davies, J., van Driel, W., et al. 2005, MNRAS, 357, 819
Saintonge, A. 2007, AJ, 133, 2087
Sarazin, C. L. 1988, X-ray emission from clusters of galaxies, ed. Sarazin, C. L.
Schaye, J. 2004, ApJ, 609, 667
Schneider, S. E., Stage, M. D., Auld, R., & Cortese, L. 2008, in American Institute of Physics Conference Series, Vol. 1035, The Evolution of Galaxies Through the Neutral Hydrogen Window, ed. R. Minchin & E. Momjian, 17-23
Silk, J. 2006, Annalen der Physik, 15, 75
Simon, J. D., Blitz, L., Cole, A. A., Weinberg, M. D., & Cohen, M. 2006, ApJ, 640, 270
Simon, J. D. & Geha, M. 2007, ApJ, 670, 313
Smith, R. 2009, PhD thesis, Department of Physics and Astronomy, Cardiff University 5 The Parade, Cardiff, Glamorgan, CF24 3YB, Wales, UK
Solanes, J. M., Giovanelli, R., & Haynes, M. P. 1996, ApJ, 461, 609
Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629
Stewart, K. R., Bullock, J. S., Barton, E. J., & Wechsler, R. H. 2009, ApJ, 702, 1005
Thilker, D. A., Bianchi, L., Meurer, G., et al. 2007, ApJS, 173, 538
Tollerud, E. J., Bullock, J. S., Strigari, L. E., & Willman, B. 2008, ApJ, 688, 277
Tonnesen, S., Bryan, G. L., & van Gorkom, J. H. 2007, ApJ, 671, 1434
Toomre, A. & Toomre, J. 1972, ApJ, 178, 623
Tripp, T. M. 2008, in IAU Symposium, Vol. 244, IAU Symposium, ed. J. Davies & M. Disney, 63-72
Tully, R. B. 2008, in IAU Symposium, Vol. 244, IAU Symposium, ed. J. Davies & M. Disney, 146-151
Tully, R. B. & Fisher, J. R. 1987, Nearby galaxies Atlas, ed. Tully, R. B. & Fisher, J. R.
van der Wel, A., Rix, H., Holden, B. P., Bell, E. F., & Robaina, A. R. 2009, ApJ, 706, L120
van Driel, W. & van Woerden, H. 1989, A&A, 225, 317
Verheijen, M., van Gorkom, J. H., Szomoru, A., et al. 2007, ApJ, 668, L9
Vollmer, B., Cayatte, V., Balkowski, C., & Duschl, W. J. 2001, ApJ, 561, 708
Wakker, B. P. & van Woerden, H. 1991, A&A, 250, 509
Walter, F., Brinks, E., de Blok, W. J. G., et al. 2008, AJ, 136, 2563
Westmeier, T., Brüns, C., & Kerp, J. 2008, MNRAS, 390, 1691
Woods, D. F., Geller, M. J., & Barton, E. J. 2006, AJ, 132, 197
Zwaan, M. A., Staveley-Smith, L., Koribalski, B. S., et al. 2003, AJ, 125, 2842
Zheng, X. Z., Bell, E. F., Rix, H., et al. 2006, ApJ, 640, 784
Zibetti, S., White, S. D. M., Schneider, D. P., & Brinkmann, J. 2005, MNRAS, 358, 949