HI Properties of Massive Galaxies from Stacking. Quenching Mechanisms.

Silvia Fabello



München 2012

HI Properties of Massive Galaxies from Stacking. Quenching Mechanisms.

Silvia Fabello

Dissertation
an der Fakultät der Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Silvia Fabello
aus Mailand

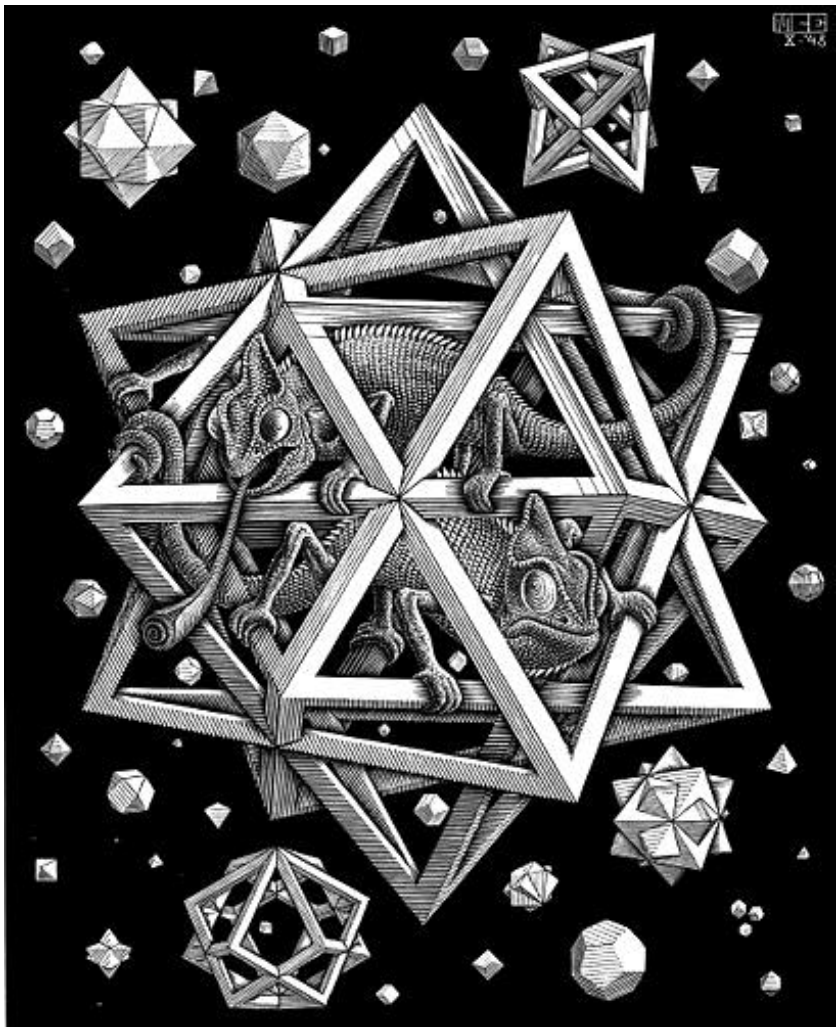
München, den 17. Februar 2012

Erstgutachter: Simon White

Zweitgutachter: Ralf Bender

Tag der mündlichen Prüfung: 18. April 2012

Astronomy? Impossible to understand
and madness to investigate.
Sophocles ($\sim$ 420 BC)



Stars (1948) – Escher, M.C.

Contents

Summary **xv**

Zusammenfassung **xvii**

1 Introduction **1**
1.1 Galaxies at $z \sim 0$ . . . 1
1.2 Neutral hydrogen and star formation . . . 6
1.3 HI properties from observations . . . 8
1.4 This work . . . 17
1.4.1 Thesis outline . . . 18

2 The sample **21**
2.1 Selection criteria . . . 21
2.2 Sample properties . . . 22
2.3 ALFALFA HI data and parameters . . . 25
2.4 SDSS data and optical parameters . . . 27
2.5 GALEX data and NUV parameters . . . 31

3 Stacking of HI spectra **35**
3.1 ALFALFA HI data . . . 37
3.2 ALFALFA stacking tool . . . 38
3.2.1 Creating a catalogue of HI spectra . . . 38

3.2.2 The stacking method . . . 41
3.2.3 Evaluating HI gas fractions . . . 44
3.2.4 Errors . . . 46
3.3 HI scaling relations . . . 47
3.4 Conclusions . . . 51

4 Morphological quenching 53
4.1 Introduction . . . 53
4.2 Bulge-Dominated sample selection . . . 55
4.3 HI study of a complete sample of early-type galaxies . . . 58
4.4 A test of the morphological quenching scenario . . . 60
4.5 Conclusions . . . 63

5 AGN feedback on the gas content of massive galaxies. 69
5.1 Introduction . . . 69
5.2 AGN and control samples . . . 71
5.2.1 The AGN sample . . . 71
5.2.2 The control sample . . . 74
5.3 The HI gas fractions of AGN hosts . . . 75
5.3.1 Summary of the HI analysis . . . 78
5.4 The H_2 content of AGN hosts . . . 80
5.5 Conclusions . . . 87

6 Environmental effects on HI and star formation properties of massive galaxies 91
6.1 Introduction . . . 91
6.2 Our environmental tracer . . . 95
6.3 Specific star formation rate scaling relations . . . 97
6.4 HI scaling relations . . . 101
6.4.1 Comparison with previous work . . . 102
6.5 Relative environmental effects . . . 103
6.5.1 Comparison with models . . . 105
6.6 Conclusions . . . 111

7 Summary and future prospects 115
7.1 Main results . . . 115
7.2 Outlook . . . 117

7.2.1 Future prospects 118

A Further Stacking tests **121**

A.1 Stacking of simulated HI spectra 121

A.1.1 Model of an HI spectrum 121

A.1.2 Catalogue of non detections 125

A.1.3 Stacking of spectra 127

A.2 Comparison of two different stacking procedures 130

B Correction for Arecibo beam confusion **135**

Bibliography **139**

Acknowledgments **149**

List of Figures

1.1 Observed galaxy bimodality 3
1.2 Example of an HI spectrum 9
1.3 Comparison of ALFALFA and GASS surveys 13
1.4 Images of VIVA galaxies in Virgo 16

2.1 Sky distribution of the sample. 23
2.2 Sample properties. 24
2.3 Schematic representation of an ALFALFA 3-D data-cube. 25
2.4 CMD from optical colours 33

3.1 *Rms* of stacked spectra vs number of co-added objects. 42
3.2 Examples of stacked spectra for five bins of $M_\star$. 45
3.3 HI gas fraction scaling relations for *sample A*. 49

4.1 Properties of the bulge-dominated (B-D) sample. 56
4.2 Examples of B-D galaxies. 57
4.3 HI gas fraction scaling relations for B-D galaxies. 59
4.4 HI gas fraction variation in the $\mu_\star$- (NUV$-r$) plane (B-D sample). 61
4.5 Test for the morphological quenching 62
4.6 *Disk* HI gas fractions for both B-D and *sample A* galaxies. 65

5.1 BPT diagram for galaxies in our sample. 72
5.2 $\mu_\star$-(NUV$-r$) plane for quiescent and active galaxies. 74

5.3 HI gas fraction as a function of L[OIII] and L[OIII]/σ^4. 76
5.4 HI gas fraction as a function of L[OIII]/σ^4 for the blue/red population. . . 79
5.5 HI gas fraction as a function of L[OIII]/σ^4 for the young/old population. . 80
5.6 BPT diagram for COLD GASS galaxies 81
5.7 NUV$-r$ distribution for COLD GASS objects 82
5.8 Cold gas and sSFR as a function of L[OIII]/σ^4 (COLD GASS) 84
5.9 M_{H_2}/M_{HI} for AGN hosts and control galaxies (one case). 85
5.10 Comparison of gas properties in AGN hosts and control galaxies 86

6.1 Local density and $M_\star$ properties of *sample A*. 96
6.2 Sky distribution of two groups in our sample. 97
6.3 sSFR distribution for the different density bins considered. 98
6.4 $M_{HI}/M_\star$ and sSFR vs $M_\star$ scaling relations for the different density bins . . 100
6.5 Comparison with literature results 102
6.6 Environmental effects on different properties of galaxy (*sample A*) 104
6.7 Local density and $M_\star$ properties of the mock catalogue 107
6.8 Environmental effects on different properties of galaxy (mock data) 108
6.9 Comparison of *sample A* and mock data: sSFR 109
6.10 Comparison of quenched fractions for mock data and *sample A* 110

A.1 Assumed HI surface density and kinematics. 122
A.2 Simulated HI spectra for different inclinations. 126
A.3 Examples of simulated non detections. 127
A.4 Simulated stacked spectrum. 129
A.5 (Simulated) *Rms* of stacked spectra vs number of co-added objects. 129
A.6 Comparison of stacking methods: *rms* 131
A.7 Comparison of stacking methods: ALFALFA detections 131
A.8 Comparison of stacking methods: HI scaling relations. 133

B.1 Example of beam confusion 136

List of Tables

3.1 Average HI scaling relations for *sample A*. 50
3.2 Fits to HI scaling relations 50

4.1 Average HI scaling relations for *B-D sample* and *sample A*. 67
4.2 Average HI scaling relations for additional *B-D samples*. 68

Summary

Galaxies have been found to divide into two families: one dominated by late-type, star forming, blue objects, which are rich in cold gas and have a low stellar mass surface density ($\mu_\star$); the other is made of early-type, red and passive galaxies with higher $\mu_\star$ and on average low gas content. The physical mechanisms responsible for the galaxy transition between the active and passive regime are still debated. In the high mass range, mechanisms proposed to quench the star formation (SF) through cold gas heating or depletion are not efficient enough to reproduce the correct red sequence of passive systems, when implemented in models of galaxy evolution. Input for a better understanding of the physics of quenching mechanisms, and of their relative importance and efficiency, can come from a comparison of the cold atomic neutral hydrogen (HI) content and SF for a statistically significant sample of massive systems where quenching is at work. However, existing surveys do not sample this high mass, gas poor regime well enough. In this work, we study the HI properties of a volume-limited sample of $\sim$5000 nearby galaxies with stellar mass $M_* > 10^{10}\,M_\odot$, selected from the state-of-the-art blind HI survey ALFALFA to have optical and ultraviolet data so that star formation and galaxy properties can be derived. As ALFALFA does not sample with sufficient sensitivity the high mass, gas poorest range, we developed a software tool to co-add its data, in order to obtain average gas properties of galaxy classes which individually may be largely undetected.

Using this technique, we study three types of quenching processes, namely the presence of a bulge component, feedback from an active galactic nucleus (AGN), and environmental mechanisms acting on the interstellar medium. Simulations of early-type galaxies with non star-forming HI disks have suggested that the presence of a bulge can stabilize the gas, thus

preventing star formation, but on average we do not observe this. We find that, once $\mu_\star$ and NUV$-r$ colours are fixed, the HI gas fraction in massive bulge- and disk-dominated galaxies is the same *(Chapter 4)*. A similar negative result is obtained if we compare $M_{HI}/M_\star$ of AGN hosts and control galaxies, despite simulations that invoke feedback from AGN to heat or deplete cold gas in massive systems. The relation we observe between the cold gas content and the accretion rate in the red population actually points towards a co-evolution of SF and AGN activity, both driven by the amount of gas available *(Chapter 5)*. The last class of quenching mechanisms studied in this work includes environmental processes, which are known to affect the SF properties of galaxies and, at least in rich clusters, their cold gas content. For the first time, though, we study the effect of the environment on the HI content as a continuous function of local density, comparing it with global and inner specific star formation rate. The gradual increase in the suppression of SF from the inner to the outer regions that we observe, and the even stronger HI deficiency as a function of increasing local density, can be explained by a mechanism acting on the disk from the outside-in, like ram-pressure stripping of the HI. A comparison with mock catalogs from models, which include only removal of the hot gas, shows how models underestimate environmental effects, especially on the cold gas component of galaxies *(Chapter 6)*.

We therefore suggest that, in order to improve our understanding of the galaxy bimodality in the local Universe, observations and models should particularly focus on environmental mechanisms acting on the cold interstellar medium. These processes are efficient over a broader range of local densities than previously thought, and could solve parts of the puzzle in the formation of massive and passive systems.

Zusammenfassung

Bekanntermaßen teilen sich Galaxien in zwei Familien auf. Die erste wird von blauen Objekten mit geringer stellarer Oberflächenmassendichte ($\mu_\star$) und einem hohem Anteil an kaltem Gas dominiert. Diese Objekte befinden sich in der späten Entwicklungsphase und bilden aktiv neue Sterne. Die zweite Gruppe setzt sich aus jungen, roten und im Allgemeinen passiven Galaxien zusammen, die höhere $\mu_\star$ aufweisen und im Mittel gasärmer sind. Die genauen physikalischen Mechanismen, die für den Übergang von Galaxien von der aktiven Zone hin zur Passiven verantwortlich sind, stehen immer noch zur Diskussion. Alle Prozesse, die im Bereich hoher Massen die Ausbildung neuer Sterne durch Aufheizen oder den Abbau von kaltem Gas dämpfen sollen, scheinen in Galaxie-Entwicklungsmodellen nicht effizient genug zu sein um die beobachtete Sequenz von passiven, roten Systemen reproduzieren zu können. Diese Dämpfungsmechanismen können durch den Vergleich des Anteils an kaltem, atomaren Wasserstoff (HI) mit der Sternentstehung (SF) in einer statistisch aussagekräftigen Auswahl an massereichen Systemen untersucht werden und hinsichtlich der beteiligten physikalischen Prozesse und ihrer Relevanz und Effizienz besser verstanden werden. Leider erfassen die existierenden Galaxien-Durchmusterungsprogramme das Regime der massereichen, gasarmen Systeme nicht ausreichend genau. In dieser Arbeit werden die HI-Eigenschaften einer volumenbeschränkten Auswahl von 5000 nahen Galaxien, deren Masse $M_* > 10^{10}\,M_\odot$ erfüllt, studiert. Um wichtige Eigenschaften und die Sternentstehungsrate von Galaxien bestimmen zu können, werden die optischen und ultravioletten Beobachtungsdaten der Objekte aus der ALFALFA Studie gewählt, die den modernsten Stand der Technik in blinden Galaxien-Durchmusterungen darstellt. Da auch die ALFALFA Beobachtungen den Bereich der massereichen, gasarmen Objekte nicht mit ausreichender

Sensitivität erfassen, wurde im Rahmen dieser Arbeit eine Software entwickelt, die die Beobachtungsdaten statistisch „aufschichtet“ (co-add, stack) um Mittelwerte für die Eigenschaften von Galaxienklassen zu bestimmen, deren HI-Anteil einzeln weitgehend nicht aufgespürt werden könnte.

Mit Hilfe dieser entwickelten Technik werden drei Typen von Dämpfungsmechanismen studiert, nämlich das Vorhandensein eines Zentralbereichs (Bulge), die Rückkopplung von aktiven galaktischen Kernen (AGN) und Umgebungsprozesse, die auf das interstellare Medium wirken. Simulationen von Galaxien, die sich in ihren frühen Entwicklungsstadien befinden und keine Scheiben als Sternentstehungsgebiete besitzen, deuten auf einen stabilisierenden Effekt des Zentralbereichs auf den Gasanteil hin. Im Mittel wurde dieser Effekt in dieser Studie jedoch nicht festgestellt. Insbesondere ist der Gasanteil in Scheibengalaxien und solchen, die vom Zentralbereich geprägt sind, gleich, sobald man sich auf $\mu_\star$ und die NUV$-r$ Farbe festlegt *(Kapitel 4)*. Ein ähnliches Resultat ergibt der Vergleich von $\mathrm{M_{HI}/M_\star}$ in AGN und Kontrollgalaxien, obwohl in Simulationen die Rückkopplung von AGN verwendet wird, um kaltes Gas in massereichen Galaxien aufzuheizen oder abzubauen. Die beobachteten Beziehungen zwischen dem kalten Gasanteil und der Akkretionsrate deuten auf eine gemeinsame Entwicklung der Sternentstehungsrate und der Aktivität von AGN hin, die von der vorhandenen Menge an Gas getrieben wird *(Kapitel 5)*. Als letzte Klasse von Dämpfungsmechanismen werden die Umgebungprozesse untersucht, deren Einfluss auf die Sternentstehungseigenschaften von Galaxien und, im Falle von gasreichen Galaxienhaufen, auch auf den Gasanteil bereits bekannt ist. Zum ersten Mal wird hier der Einfluss der Umgebung auf den HI-Anteil in Abhängigkeit von der lokalen Dichte studiert und mit der globalen und inneren spezifischen Sternentstehungsrate verglichen. Die beobachtete, allmählich steigende Unterdrückung der Sternentstehung von inneren Regionen nach außen sowie der noch stärkere HI-Mangel mit steigender lokaler Dichte kann durch einen nach innen gerichteten Mechanismus erklärt werden, wie zum Beispiel das Abziehen von HI durch den Staudruck. Ein Vergleich mit Modellgalaxien, die nur den Abbau von heißem Gas berücksichtigen, legt nahe, dass die Umwelteinflüsse in diesen Modellen unterschätzt werden, vor allem im Hinblick auf die kalte Gaskomponente *(Kapitel 6)*.

Diese Studien stellen die Bedeutung von Umwelteinflüssen, die auf das kalte interstellare Medium wirken, deutlich hervor. Im Hinblick auf das Verständnis der Bimodalität von Galaxien im lokalen Universum wird angeraten, dass zukünftige Beobachtungsprogramme und Modellierungsansätze sich speziell auf diese Einflüsse konzentrieren. Diese Prozesse arbeiten effizient über einen größeren Bereich von lokalen Dichten als bisher vermutet und könnten Teile des Rätsels über die Entstehung von massereichen und passiven Systemen lösen.

1

Introduction

Hydrogen is the most abundant element in the Universe and the primordial building block of the baryonic structures in it. At present, at redshift $z = 0$, the vast majority of baryons exists in the form of hot intergalactic gas, while the fraction of matter as cold gas is tiny ($\sim 0.3\%$ of the mass/energy budget of the Universe). Of this, more than half is in the form of neutral hydrogen. While this might appear a negligible constituent of the Universe, nevertheless neutral hydrogen in both its molecular (H_2) and atomic (HI) phase is important in many astrophysical processes, and it is a reliable indicator of the star forming potential for galaxies. In addition, HI is a tracer of galaxy dynamics and it is sensitive to tidal interactions, mergers and other environmental effects. In this work, we use atomic hydrogen to study some open issues in the observed properties of massive galaxies in the local Universe. In this first chapter, we present a brief overview of galaxy properties at $z \sim 0$, and their connection with the current picture of galaxy evolution (Section 1.1). We discuss the role of the atomic hydrogen component in galaxies (Section 1.2), and present a review of the current state of HI observations, with a particular focus on analyses related to our work (Section 1.3). In Section 1.4, we discuss how we plan to address some of the open issues in the evolution of massive systems, and why our work improves our understanding of galaxy evolution. Section 1.4.1 draws an outline of this thesis.

1.1 Galaxies at $z \sim 0$

According to the standard cosmological model, the formation of structures in the Universe originates from quantum perturbations in the matter density expanded to cosmological scales by inflation. Dark matter fluctuations grow via gravitational instabilities to form

dark matter halos, and galaxies are believed to form within these density peaks once they are sufficiently massive for the baryons to be able to cool and condense. In the framework of the dark matter halo merger history (White & Frenk 1991), galaxy evolution can then be thought of, at a simplistic level, as a three phases process, each directly connected with gas: accretion of gas from the intergalactic medium (Kereš et al. 2005; Dekel et al. 2009), followed by the processing of gas in the galaxy through star formation, and, finally, expulsion of enriched gas and energy back in the intergalactic medium (Dave 2006). Numerical simulations of large-scale structures have to date produced realistic maps of the dark matter distribution (Springel et al. 2005; Boylan-Kolchin et al. 2009). More difficult is for models to reproduce smaller scale observations such as individual galaxies, for which detailed baryonic physics is essential.

From an observational point of view, in the last decade significant improvement in understanding galaxy evolution has been possible thanks to the availability of overlapping large surveys at different wavelengths, which provide empirical diagnostics of the physical processes at work. The Sloan Digital Sky Survey (SDSS; York et al. 2000), with 5 band optical photometry and spectroscopy, has produced in addition to magnitudes and colours accurate structural parameters such as stellar mass, size, surface brightness and stellar surface density, and active galactic nuclei (AGN) content, metallicity and velocity dispersion for $\sim$ one million galaxies. The Two-Micron All Sky Survey (2MASS Skrutskie et al. 2006) scanned the entire sky in three near-infrared bands and detected a similar number of galaxies. The structural parameters that can be extracted from its data are similar to the ones obtained from SDSS photometry, such as galaxy mass, size, morphology, but at wavelengths that are sensitive to the older stellar populations and not obscured by dust. In the UV, the Galaxy Evolution Explorer (GALEX; Martin et al. 2005) observed over 40,000 square degrees of the sky in two imaging and spectroscopic bands, providing, in combination with SDSS, star formation rates and dust attenuation measurements. In principle, with information about all components in galaxies, we can *quantify* in detail the relations between global properties of galaxies and trends with the environment in which they are located. These correlations are the observational tool to understand the internal physics of these systems, to break the degeneracies in the several physical mechanisms proposed to be at work. In order for these analyses to be effective, samples have to be: *(i)* unbiased and spanning the broadest possible range of galaxy properties; *(ii)* homogeneous; *(iii)* large enough to allow cross-correlations among different properties.

In this respect, the surveys mentioned above have played a significant role in boosting our knowledge of galaxies in the local Universe, thanks to their large samples spanning a

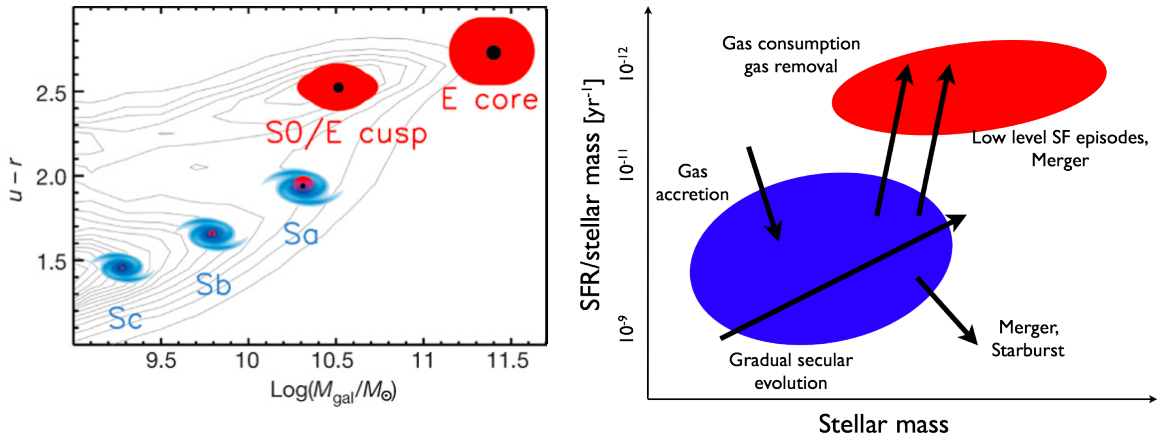


Figure 1.1: Representations of galaxy bimodality. *Left:* galaxy distribution contours on a total stellar mass - colour diagram, from SDSS data. On top of the contours, cartoons of galaxy types are shown, to illustrate the transition between blue disks and red bulges/elliptical galaxies. Black dots represent black holes. Credits: Cattaneo et al. (2009). *Right:* Schematic representation of the processes that could be responsible for the observed galaxy bimodality. Adapted from Schiminovich et al. (2007).

broad range in galaxy parameter space. One of the main outcomes of the analyses performed on their data has been the proper characterisation of the relations between galaxy properties. While galaxies have long been known to follow ordered sequences in many properties, which change from spirals to ellipticals, SDSS analyses actually revealed that the transition between these is not smooth. In a plane identified by specific star formation rate (or colour) and stellar mass (or luminosity), galaxies occupy either a blue cloud dominated by star-forming, blue, late-type morphologies, or a tighter sequence of red and passive, high stellar mass surface density spheroidals. This bimodality is presented in Figure 1.1, left panel (Cattaneo et al. 2009). The gray lines indicate the galaxy distribution contours on a total stellar mass-colour diagram, from SDSS data. On top of the contours, cartoons of galaxy types are shown, to represent the change from blue disks to red bulges/elliptical galaxies. Black dots represent black holes, which are almost ubiquitous in massive galaxies and whose mass increases with the host mass (Kormendy 2004). The transition between the two families occurs at a stellar mass of about $3 \times 10^{10} M_{\odot}$ (Kauffmann et al. 2003b; Baldry et al. 2004), but it is associated also with transitions in other structural parameters, such as the stellar mass surface density ($\mu_{\star}$). At Log $\mu_{\star}[M_{\odot} kpc^{-2}] \sim 8.5$, a bimodality has been identified between disk-dominated, late-type galaxies and bulge-dominated, early-

type objects (Kauffmann et al. 2006) and in the UV-based specific star formation rates (Schiminovich et al. 2007). With respect to the cold gas content, the current picture is that blue-cloud, low $\mu_\star$ galaxies are gas-rich while red-sequence ones or galaxies with high $\mu_\star$ are on average gas-poor (Roberts & Haynes 1994; Catinella et al. 2010; Saintonge et al. 2011a). To date, though, our understanding of the statistical properties of gas in galaxies lags behind the optical one. This gap arises because gas is more difficult to detect, so that at present we have data for the global HI content of a few 10^4 galaxies, two orders of magnitude less than the size of the galaxy sample observed by SDSS. If we consider resolved HI data, samples are even smaller (few hundred objects). In addition to the smaller numbers, which do not allow similar cross-correlation studies as performed with SDSS or GALEX, an additional limits comes from the fact that the bulk of available HI data assembled by large area sky surveys is biased toward the most gas-rich objects.

Observations in the local Universe, like the ones presented above, are used as constraints for models, whose prescriptions have to reproduce observed galaxies. To date, the most promising approach to understand the formation of galaxies and clusters within the dark matter gravitational framework are Semi-Analytic Models (SAMs; White & Rees 1978; Cole 1991; White & Frenk 1991; Kauffmann et al. 1993). SAMs are sets of analytic prescriptions for baryonic physics, implemented on top of the dark matter halo merger history. In SAMs, gas evolution is regulated by cooling, heating, star formation, feedback from supernovae and AGN. Current models can reproduce the observed bimodality, but still have difficulties in creating a realistic sequence of massive, red and passive galaxies despite the negative feedback prescriptions implemented. Modeled massive galaxies at low redshifts are in fact too active, so that some additional physical process, or a refinement of the implemented ones, is required to quench their star formation and keep them passive (e.g. Croton et al. 2006; Cattaneo et al. 2007). In Figure 1.1, right panel, we show again a representation of the bimodality, adapted from Schiminovich et al. (2007), on top of which we have plotted a scheme of the main mechanisms suggested to be at work in galaxies. The gradual secular evolution causes galaxies to be more massive and less star forming as they consume their gas into stars; as do mergers, which can quickly turn galaxies red after an eventual starburst phase. In addition to these, mechanisms proposed to solve the overabundance of star formation in massive systems by quenching their star formation should either prevent gas cooling or deplete the already existing interstellar medium (ISM) reservoir. They can be split into two main classes, as follows.

i) The first class of quenching mechanisms includes all processes that act on central

galaxies, i.e. galaxies lying in the center of their dark matter halo, over which any inflow of gas is likely to be directed. The main candidate for this class is negative feedback from accretion of matter onto supermassive black holes (SMBH). The fact that SMBH are almost ubiquitous in massive galaxies together with observations that the mass of the BH correlates with the bulge mass (Kormendy 2004) led to the idea that star formation and BH evolution are somehow related. Two main modes of feedback from AGN are today considered, acting either through radiative (e.g. Kauffmann & Haehnelt 2000; Hopkins et al. 2006; Tremonti et al. 2007) or mechanical (Croton et al. 2006; Fabian et al. 2006) feedback. The radiative feedback action should occur mostly at high redshift (z$\sim$2) when quasars, the most energetic and brightest AGN, reach their peak activity. Massive, large-scale outflows have actually been detected in the hosts of local quasars (e.g. Fischer et al. 2010; Sturm et al. 2011), but whether the feedback is efficient on statistical scales also from lower luminosity AGN, more common in the local Universe, is still unclear. There is observational evidence that star formation and AGN presence correlate in the local Universe (e.g. Kauffmann et al. 2003a; Schawinski et al. 2007) and at intermediate redshifts (e.g. Nandra et al. 2007; Silverman et al. 2008), but a systematic analysis of the cold gas component is currently missing.

Another possibility that has recently been suggested to halt the star formation in early-type galaxies is that the presence of a bulge itself is enough to stabilize a gas disk against fragmentation, thus preventing further star formation (Martig et al. 2009). This "morphological quenching" would not require any gas consumption, removal or suppression of gas supply, and would explain why early-type systems do not undergo a second phase of star formation.

ii) The second class of quenching mechanisms includes environmental interactions, which can halt the star formation in galaxies once they enter a dense region. In particular, models assume a shutoff of hot gas cooling in galaxies once they fall into a larger halo (e.g. Guo et al. 2011). This mechanism, cutting the primary fuel for star formation, produces a population of clustered, red-sequence galaxies that seems to be in agreement with SDSS data (e.g. Baldry et al. 2006; Weinmann et al. 2009). Additional stronger environmental mechanisms, such as tidal interactions between galaxies, or hydrodynamic interactions between the ISM and the dense intracluster medium, should be even more efficient in quenching the star formation in dense environments (Boselli & Gavazzi 2006).

Although the relative importance of these quenching mechanisms is still debated, it

is clear that every process that halts star formation in galaxies must also affect their gas content. Therefore, quantifying the gas properties of massive systems in the process of being quenched and comparing them with other galaxy properties is of primary importance for constraining the action and efficiency of quenching mechanisms. To clarify why the study the *atomic* phase of hydrogen is particularly important to address the quenching issue, let us briefly outline the hydrogen role in galaxies and star formation[1] (§ 1.2). In the following section (§ 1.3), we are going to review the main HI properties of nearby galaxies deduced from observations.

1.2 Neutral hydrogen and star formation

The "life-cycle" of hydrogen starts in the intergalactic medium (IGM) where the gas is mostly ionized and highly inhomogeneous, as probed by studies of hydrogen absorption features in the UV spectrum of light coming from distant quasars (e.g. Kanekar & Briggs 2004, for a review). From the IGM, hydrogen can be accreted onto dark matter haloes, where it can settle as interstellar medium of the central galaxy (White & Frenk 1991), or be shock-heated to the virial temperature into diffuse hot gas in the halo, which eventually cools onto the galaxy as well (Kereš et al. 2005). In the ISM, hydrogen spans a large range of densities and temperatures, and plays, in its different phases, a direct role in the process of star formation. The cold, dense ISM component consists of neutral hydrogen; the majority is in the atomic phase ($n \simeq 100\,\mathrm{cm}^{-3}$ and $\mathrm{T}\simeq 100\,\mathrm{K}$), which can condense into molecular one ($n \simeq 10^4\,\mathrm{cm}^{-3}$ and $\mathrm{T}\simeq 10\,\mathrm{K}$). Molecular hydrogen forms in the environment of the HI disk where the surface density is high and dust is available: two atoms of hydrogen collide with a dust grain, which allows the formation of an H_2 molecule, acting as a catalyst by absorbing the excess of energy. Subsequently, when the Jeans density necessary for a cloud of material to become gravitationally unstable is exceeded (Jeans 1902), the molecular cloud starts to collapse. If the energy released is effectively dissipated through e.g. dissociation of H_2 and ionization of H, the collapse will continue until a star (or a group of stars) forms. These newly born stars, in turn, influence the remaining gas. The most massive, hot stars emit UV radiation which ionizes the hydrogen, creating regions in photo-ionization equilibrium (HII regions) which can be observed through lines of the hydrogen Balmer series (primarily trough Hα emission at 6562.8 Å). Stars can return enriched material to the circumgalactic medium, outside the galaxy but inside its host halo, through mass outflows or at the end of their evolutionary cycle through

[1] A recent, detailed review about the theory of star formation is presented in McKee & Ostriker (2007).

supernova explosions. In addition, processes like feedback from a supermassive black hole or environmental interactions can heat the cold gas or expel material from the galaxy itself.

The hydrogen phase directly connected with star formation, on sub-kpc scales characteristic of gas clouds, is actually the molecular one. A step before this, though, *atomic hydrogen* constitutes the primary supply for the potential growth of galaxies. If this is removed, for example by the mechanisms listed in the previous paragraph, new star formation cannot proceed. Observationally, star formation is globally quenched in galaxies that are HI deficient due to environmental effects, a result that can be interpreted in terms of HI feeding SF (Boselli et al. 2001). On global scales, a relation between atomic hydrogen and star formation has been measured (Kennicutt 1998; Schuster et al. 2007). Such a relation may not be representative of the detailed physics involved in the formation of individual stars, but in order to study statistical properties of galaxies, and the evolution of a galaxy population over cosmological timescales, such coarse characterisation of the star formation - hydrogen connection is adequate. The amount of atomic hydrogen in galaxies is a sensitive tracer of the evolution of these systems, as it is set by the balance of inflows and outflows, and it is modulated by both internal processes and interactions with the surroundings.

To study the effects of quenching mechanisms, therefore, information about the atomic hydrogen combined with optical and UV data, to trace star formation and galaxy properties, would be already effective. With homogeneous data for a large sample of galaxies spanning the whole population range down to the gas-poor regime, one can ask which process is the dominant one in halting the star formation, and in which objects. Additionally, one can test the physics of the mechanisms at work in real galaxies and compare their observational signatures with expected effects. Finally, it is possible to constrain the statistical influence of quenching processes in shaping galaxy properties. Due to the lack of HI data, only limited attempts on small datasets have to date been performed. The atomic hydrogen content in galaxies, as we already discussed, is currently not properly sampled especially in the high mass range where galaxies are gas poorer, although it is one of the missing ingredients for a complete picture of the physics of quenching mechanisms at work in massive systems.

1.3 HI properties from observations

Neutral atomic hydrogen in its ground state emits at 21 cm, in the radio range, through an hyperfine magnetic dipole transition. The electron can have its magnetic moment either parallel to that of the proton, or anti–parallel. The parallel state, populated by collisions, has a slightly higher energy, so that a transition to the anti–parallel state produces radiation at a wavelength of 21 cm (a frequency of 1420.40575177 MHz). The lifetime of the upper state is only about 11 million years, too short to be observed in a laboratory but measurable by radio telescopes given the large numbers of atomic hydrogen atoms in the interstellar medium. To date, however, HI emission has been detected only out to $z \sim 0.2$ (Verheijen et al. 2007; Catinella et al. 2008), both because the flux quickly decreases with distance and its detection with current facilities would require unfeasible integration times, and because at higher redshift the 21 cm emission line shifts in a frequency band affected by man-made interference, making observations challenging.

Global properties

Such radiation, theoretically predicted by van de Hulst in 1944, was first detected a few years later by Ewen, who observed the signal coming from the spiral arms of our own Galaxy (Ewen & Purcell 1951). Two years later, the first extragalactic HI emission from the Magellanic clouds, the two galaxies nearest to the Milky Way, was detected by Kerr & Hindman (1953). After that, it took nearly two decades of pioneering work, largely by Roberts, to increase the number of galaxies detected to more than one hundred (Roberts 1975). During this period, almost all extragalactic HI observations were performed with single-dish telescopes, which have spatial resolutions ranging from few to tens arcminutes, so that for the majority of the galaxies only the total emission, spatially integrated over the whole object could be determined. The HI data resulting from a single-dish observation of a galaxy with apparent size smaller than the telescope beam are spectra, i.e. measures of the flux as a function of the frequencies of emission, Doppler-shifted by the radial velocities of the emitting atoms in the galaxy. An example of a typical HI spectrum of a galaxy is shown in Figure 1.2. The characteristic double horn profile is the result of the convolution of the spatial HI distribution with the underlying velocity field and with the beam response. From the HI spectrum, fundamental global properties can be derived. The redshift (or recessional velocity) of the emitting system can be estimated from the midpoint of the profile (flagged as red dashed line in the figure); the maximum rotational velocity is extracted from the observed width of the HI profile (as flagged by the blue lines), once

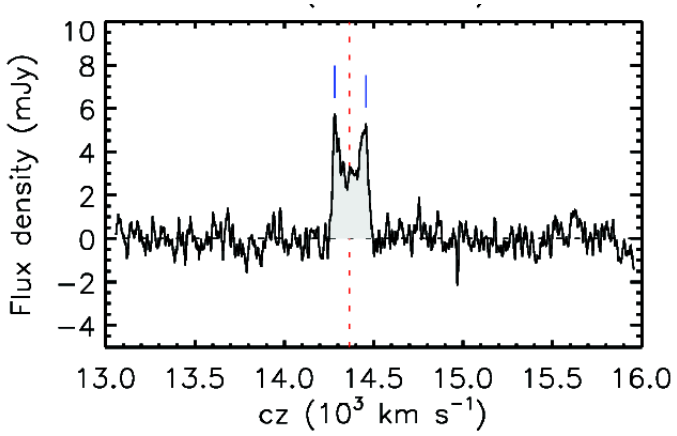


Figure 1.2: A typical HI line profile for a galaxy. The red line marks the midpoint of the emission.

corrected for inclination effects; finally, the total HI flux integrated under the line profile (gray area in the figure) is proportional to the total HI mass.

The total HI content, which is the main quantity that we use in this work, is particularly interesting because of its correlation with other global galaxy properties, especially the total stellar content. Earlier studies already pointed out a correlation between the hydrogen mass and properties like galaxy colour, total stellar mass (M_{tot}), and luminosity (Roberts 1975). In addition, a clear dependence of the relative gas content M_{HI}/M_{tot} on morphological types was measured for irregular and spiral galaxies. Roberts (1975) and later Roberts & Haynes (1994) compared integrated values to show how the gas content in galaxies decreases monotonically with type: dwarf blue galaxies can have the largest fraction of their mass in the form of atomic hydrogen, typical spirals have on average ∼10-20% HI, while for earlier morphological types the gas fraction decreases down to few percent, to be almost zero in ellipticals. These observations led to speculation on the possible evolutionary paths and on possible gas consumption time-scales, issues that are still partially open nowadays. A second, fundamental result of the studies of HI properties was the discovery of an empirical correlation between the hydrogen line width, a distance-independent observable, and the absolute magnitude (or diameter) in spiral galaxies (Tully & Fisher 1977). This relation, one of the tightest measured in spirals, turned out to be a powerful tool to study the large-scale galaxy distribution. If one measures the rotational velocity of a galaxy, using the Tully-Fisher (TF) relation it is possible to estimate its intrinsic brightness; by comparing it with the apparent magnitude it is then possible to estimate the galaxy distance, independently of redshift. In the 1980s and 1990s, the TF relation was exploited both for the measurement of the Hubble constant and for characterising galaxy

peculiar velocities, which arise from gravitational perturbations due to inhomogeneities in the density field. In addition, the HI line was used for accurate redshift estimates, which showed the filamentary large scale structure of the Universe.
In disk galaxies, another empirical scaling relation adds to the observed dependence of rotational velocity on luminosity, providing further insight into these objects. Hewitt et al. (1983) observed a trend between the total HI content and the galaxy disk size. This property was used by Haynes & Giovanelli (1984) to calibrate a relation that allowed them to estimate the "normal" HI content of a galaxy given its size and morphological type, by comparison with a reference sample of isolated galaxies. By using this reference HI content, they introduced the "HI deficiency" parameter defined as the difference between the observed and expected gas content, and proved that galaxies located in a dense environment are typically depleted of their HI (Giovanelli & Haynes 1985). Already in 1972, Gunn & Gott had proposed that stripping of the ISM of a galaxy moving in a dense intracluster medium would have an important effect on the evolution of galaxies in clusters. The relation between the atomic gas content and the environment, likely to be caused by stripping, has been an open question in the field of HI astronomy since then. In particular, the effect of environments less dense than clusters on galaxy properties is poorly constrained. The Arecibo Galaxy Environment Survey (AGES; Auld et al. 2006) is currently blindly observing 200 square degrees of the sky in 13 regions, chosen to sample a range of galaxy environments from voids to clusters. When completed, it will allow us the study of environmental impact on HI content and morphology on much larger samples. Preliminary results indicate that the faint-end slope of the HI Mass Function (HIMF) is strongly influenced by the environment: the fraction of small HI-galaxies decreases in less dense environments. We are going to further discuss environmental results in the next part of the Section, and we will address open problems in Chapter 6.

Statistical studies of cluster galaxies often focused on the effects of the removal of atomic gas on galaxy star formation, as compared with field galaxies. Star formation is quenched in galaxies that are HI deficient due to environmental effects (Kennicutt 1983; Gavazzi et al. 2002; Koopmann & Kenney 2004; Gavazzi et al. 2008). The first relation between atomic gas and star formation actually dates back to 1959, when Schmidt observed it on data from the Milky Way. In the following 50 years, many studies have looked into the properties and validity of the Schmidt law, a milestone being the work by Kennicutt (1998). Kennicutt used a sample of mostly spirals with Hα (as tracer of star formation), HI and (indirect) H_2 measurements, and formulated a "law" connecting the star formation rate (SFR) surface density and the surface density of the cold gas: $\Sigma_{SFR} = A\Sigma_{gas}^{n}$, which holds over more than 4 orders of magnitude in Σ. This Kennicutt-Schmidt relation implies

that the efficiency with which stars form from gas in galaxies is constant, although not linear ($n \neq 1$). Kennicutt in fact measured a logarithmic slope of the relation between 1.3 to 2.5, for normal galaxies, and a much tighter relation with $n \sim 1.4$ including starburst galaxies, where molecular surface densities are particularly high. Finally, Kennicutt (1998) measured a stronger correlation of total star formation rates with the total gas content (HI+H_2), but which of the two hydrogen phases is the dominant depends on the galaxies, and on the galactic scale considered. To resolve these open issues, spatially resolved data are needed, like the ones we review in the second part of the section. One of the most recent studies based on integrated properties (Bothwell et al. 2009) tried to characterise the nature of star formation and ISM dependence in the local Universe. A new result of their analysis was the observation that at the high-mass end of the galaxy population the distribution of (low) specific star formation rates matches the (gas-poor) HI distribution. This supports the idea that the transition between the red and blue sequences in massive galaxies is connected with a lack of available atomic gas supply.

In the last two decades, multi-beam receivers installed on single-dish telescopes allowed astronomers to undertake the first HI blind, large area surveys. The first wide-area HI survey was the HI Parkes All Sky Survey (HIPASS; Meyer et al. 2004), which scanned 70% of the sky identifying ∼5000 sources, thus increasing the number of detections available at that time by two orders of magnitudes. Thanks to this new blind approach, the study of gas-rich galaxies over the widest (for the time) range of luminosity and environments became possible. It provided, among many results, the first HI mass function of the local Universe (Zwaan et al. 2003); it demonstrated that most of the HI is associated to galaxies, and led to some interesting discoveries such as the Leading Arm of the Magellanic Stream (Putman et al. 1998). The successor of HIPASS and state-of-the-art HI large survey is the on-going Arecibo Legacy Fast ALFA survey (ALFALFA; Giovanelli et al. 2005), whose data we use in this work. ALFALFA is a blind survey of 7000 deg^2 of the sky out to redshift 0.06, performed at Arecibo with 8 times better sensitivity and 4 times better angular resolution than its precursor. ALFALFA is expected to detect more than 30,000 galaxies, presenting the HI view of a cosmologically significant volume in the local Universe. More than 2/3 of its sources have never been previously observed in the 21 cm line because selection criteria for past HI observations were based on optical parameters and missed the majority of HI rich systems. One of ALFALFA primary goals is the robust determination of the HIMF in the local Universe (Martin et al. 2010). Additional results include the proof that stable HI disks without an optical counterpart do not exist (Haynes 2007); typically, HI systems that are optically dark are tidal debris. ALFALFA has sparked a variety of papers on specific

morphological classes and environments, particularly connected with the Virgo cluster. The survey has detected a number of large (>250 kpc) HI streams on the outskirts of the cluster, which could represent tidal or interaction debris (Kent 2008; Koopmann et al. 2008). The HI mass function in Virgo has been derived, showing significant shortfall of high HI mass galaxies caused by the known HI deficiency of cluster spirals (Kent et al. 2008). Star formation and gas content in the Virgo cluster region have been extensively studied (e.g. Gavazzi et al. 2008; Grossi et al. 2009), although with a predominant focus on spiral galaxies, which are easier detected. Interestingly for our work, an analysis of more HI–rich than average massive galaxies has revealed how they have stronger colour gradients than corresponding normal galaxies with similar structural properties. Observations of massive galaxies with outer regions bluer than the inside (and higher specific star formation rates) support the idea of an outside-in disk formation: galaxies have recently accreted new gas in the outer regions, and it is currently forming stars (Wang et al. 2011).

ALFALFA data constitute the largest HI sample available to date, but is still not ideal for understanding the quenching mechanisms that act on massive galaxies and that we mentioned in Section 1.1. The ALFALFA sensitivity, in fact, because of the short integration time (∼1 minute), is not sufficient to detect a large fraction of the massive, gas-poorer systems over a cosmologically significant volume. In Figure 1.3, ALFALFA detections for galaxies with $M_\star > 10^{10} M_\odot$ and redshifts $0.025 < z < 0.05$ are plotted as black dots, as a function of both structural parameters ($M_\star$ and $\mu_\star$) and NUV$-r$ colour. Note that at these redshifts (which is not true for very nearby objects) the detected gas fractions are always high, and detections are biased toward blue galaxies (NUV$-r$<4). An on-going survey that has been specifically designed to study the transition of massive galaxies between the blue cloud and the red sequence (see again Figure 1.1) is instead the GALEX Arecibo SDSS survey (GASS; Catinella et al. 2010). GASS is a targeted survey of ∼1000 massive ($M_\star > 10^{10} M_\odot$) galaxies in the local Universe ($0.025 < z < 0.05$), which are observed at Arecibo down to a very low gas fraction limit of few percent. In Figure 1.3, GASS detections are overplotted as red dots; red arrows are non detections in the survey. Note how, at fixed stellar mass (left panel), GASS reaches much lower gas fractions than ALFALFA. Thanks to its selection criteria, based only on a cut in stellar mass and redshift, and on the intersection with SDSS and GALEX, GASS is assembling the first statistically significant sample of massive galaxies in the transition region, with homogeneously measured stellar masses, SFRs, gas properties, and other galaxy parameters. This unique sample, whose observations are still on-going, has already produced several new results. First, GASS has quantified gas fraction scaling relations as a function of $M_\star$, $\mu_\star$ and NUV$-r$ colour for massive galaxies. It showed that the last two parameters actually drive the

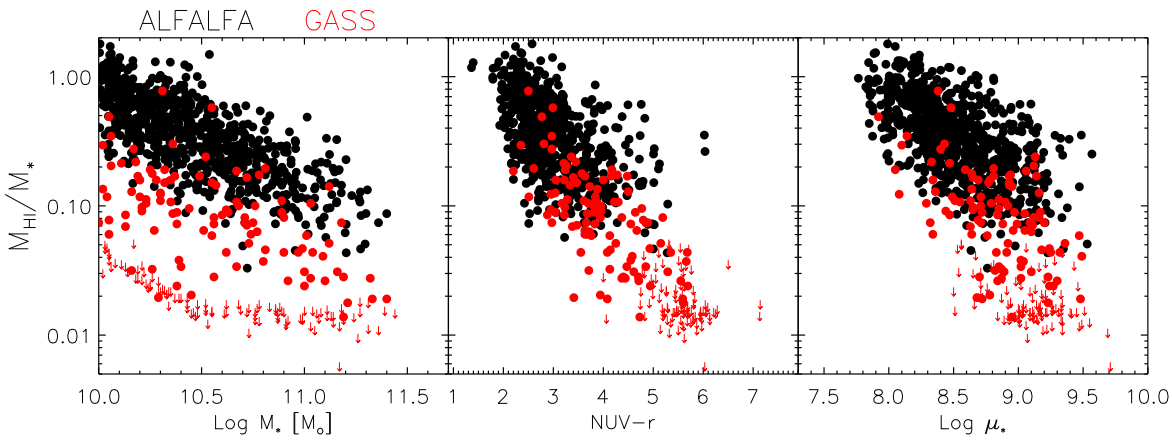


Figure 1.3: Comparison of ALFALFA (black dots) and GASS (red dots) results, in the massive range observed by GASS. Red arrows are GASS non-detections. From left to right: HI gas fraction as a function of stellar mass, NUV$-r$ colour and stellar mass surface density.

strongest correlation of $M_{HI}/M_\star$ in these systems, and that by using $\mu_\star$ and NUV$-r$ it is possible to define an HI-gas fraction plane, which allows us to identify candidates for systems in the process of transiting between the two families (Catinella et al. 2010). GASS has also shown that nearly 40% of the HI in the local Universe is found in galaxies more massive than $M_\star = 10^{10} M_\odot$, where a corresponding ∼50% of the star formation rate is measured (Schiminovich et al. 2010). The little variation observed in star formation efficiency (SFR/M_{HI}) with stellar mass, $\mu_\star$, NUV$-r$ colour suggests that external processes or feedback mechanisms, rather than internal structure, play a role in regulating the star formation in massive galaxies. Finally, GASS identified a threshold for stellar mass surface density $\mu_\star \simeq 10^{8.5}$ $M_\odot kpc^{-2}$, above which the detection fraction declines strongly (Figure 1.3, right panel). This is the same threshold observed in the bimodality of specific SFR (Schiminovich et al. 2007) and morphology (Kauffmann et al. 2006).

Resolved studies

Parallel to the analyses just presented, the successes produced by the study of the HI line triggered a steady improvement in radio facilities. Starting from the seventies, the use of interferometric technique developed, leading to an increase in the spatial resolving power. The spatial resolution of an interferometer is in fact equivalent to the one of a single antenna with a diameter equal to the array size. As a reference example, the largest single-dish telescope currently existing is Arecibo, which, with its 305m diameter, has a resolution of 3.5 arcminutes at 21 cm. The Very Large Array (VLA), the most sensitive

interferometer currently working, has 27 radio antennas of 25m in diameter, which in their more extended configuration are equivalent to a 36km diameter antenna. This corresponds to a spatial resolution of ∼1.4 arcseconds at 21 cm. Note, however, that the sensitivity of a telescope is given by the total collecting area, which does not change by separating the antennas: the VLA sensitivity corresponds "only" to the one of a 130m dish. As a result, the samples of galaxies with resolved HI maps are small, including only a few hundreds objects against the few thousands observed with single dishes. However, high spatial resolution HI observations of selected targets are extremely valuable, as they allow us to study in detail the physics and dynamics of the interstellar medium in nearby galaxies.

The first studies with interferometric technique focused on characterising the resolved distribution of HI in many more objects than the few, nearest systems that had been mapped with single-dish instruments. These studies confirmed that the distribution of atomic hydrogen is in most cases much more extended than the stars, and revealed that in disk galaxies the HI is distributed in a rotating disk, which mimics the stellar one. In addition, the HI distribution often exhibits a central depression (e.g. Hewitt et al. 1983), which likely corresponds to denser regions where the molecular hydrogen is the dominant constituent of the interstellar medium. An early, important finding of spatially resolved studies, and a milestone in the establishment of dark matter theories, was the discovery of some "dark" material in galaxies through the study of the rotation curves of spiral galaxies. A rotation curve is the plot of the tangential velocity in the plane of a rotating disk as a function of the distance from the galactic center. HI rotation curves were found not to decrease as expected from Newtonian mechanics, but to be actually flat out to larger radii (Roberts & Rots 1973), as if a substantial amount of matter that is not emitting light was distributed far from the center of the galaxy. This trend had not been noticed before (for example from rotation curves built using Hα emission line) because the HI is an outer tracer of the dynamical mass, as it extends far beyond the optical disk.

In the following years, the focus of spatially resolved studies has been mainly directed towards the understanding of the connection between neutral gas and star formation, and how this is affected by the environment. To investigate the physics of star formation in detail, observations of both HI and H_2 with high spatial resolution are needed, and this approach has been followed by many. Kennicutt et al. (2007), among the most recent results, studied the local star formation law in the nearby galaxy M51, with a spatial resolution of 0.5-2 kpc. On these small scales, they measured no correlation of SFR with Σ_{HI} alone, as the atomic component had almost constant column density, while the correlation with molecular gas only was similar to the one obtained considering the total gas. The lack of correlation between SFR surface density and HI surface density on local scales was

in contrast with the relatively strong dependence observed on global scales in disks (e.g. Buat et al. 1989; Kennicutt 1989, 1998). The difference probably arose from the different molecular fractions characteristic of the samples (H_2 constitutes more than 90% of the cold gas in M51, while HI constitutes on average 50% of the cold gas in the disks of the spirals studied by Kennicutt). As Kennicutt et al. discuss in their work, however, it may be the case that the global correlation breaks down on subkiloparsec scales. One of the largest surveys performed to study this issue, The HI Nearby Galaxy Survey (THINGS; Walter et al. 2008), observed the HI content of 34 galaxies drawn from a multiwavelength sample to span Hubble types from dwarfs to spirals, and a wide range of star formation rates, absolute luminosities and metallicities. Thanks to the high sensitivity of the survey, the authors could investigate in detail star formation properties and the structure of the ISM. First, they found evidence of the effect of feedback from supernovae on the galaxy ISM: out of 20 galaxies, they observed more than 1000 HI shells or holes in the HI distribution of gas-rich systems. The estimate of the star formation rate from these features correlates with values obtained by other SF tracers, thus confirming the feedback scenario (Bagetakos et al. 2007). Second, the multiwavelength sample allowed the authors a comparison of the Schmidt-Kennicutt law using the different neutral hydrogen components HI, H_2 and HI+H_2 on kpc scales (Bigiel et al. 2008). The result was that locally the gas – star formation connection depends on environmental quantities such as the pressure of the ISM. Where the gas is predominantly molecular, the Schmidt-Kennicutt relation depends only on this phase and is compatible with a linear relation. Instead, where HI dominates, like in the outer galaxy disks, there is a clear relationship between star formation and HI, although star formation is less efficient. Many additional studies had already investigated, and have investigated after, the star formation process with an obvious focus on H_2, but they are beyond the scope of this review.

A last important (for our work) branch of spatially resolved HI analyses is the study of how the evolution of galaxies can be influenced by the environment in which objects are located. The occurrence of HI deficient galaxies in clusters has been known for a long a time (Giovanelli & Haynes 1985). Recently, an HI imaging survey of galaxies in the Virgo cluster performed at the VLA, the VLA Imaging of Virgo in Atomic gas (VIVA; Chung et al. 2009), mapped 53 late-type Virgo cluster galaxies to actually study the various physical mechanisms that are expected to drive galaxy evolution in a dense environment. In Figure 1.4 (left panel), a composite image of the HI data for the individual galaxies (blue) is shown, overlaid on the ROSAT X-ray image (orange) tracing the hot ICM (by Böhringer et al. 1994). X-rays trace the hot ICM. On the right, four examples of individual galaxies are reported with the HI contours from VIVA overplotted on SDSS composite images. The

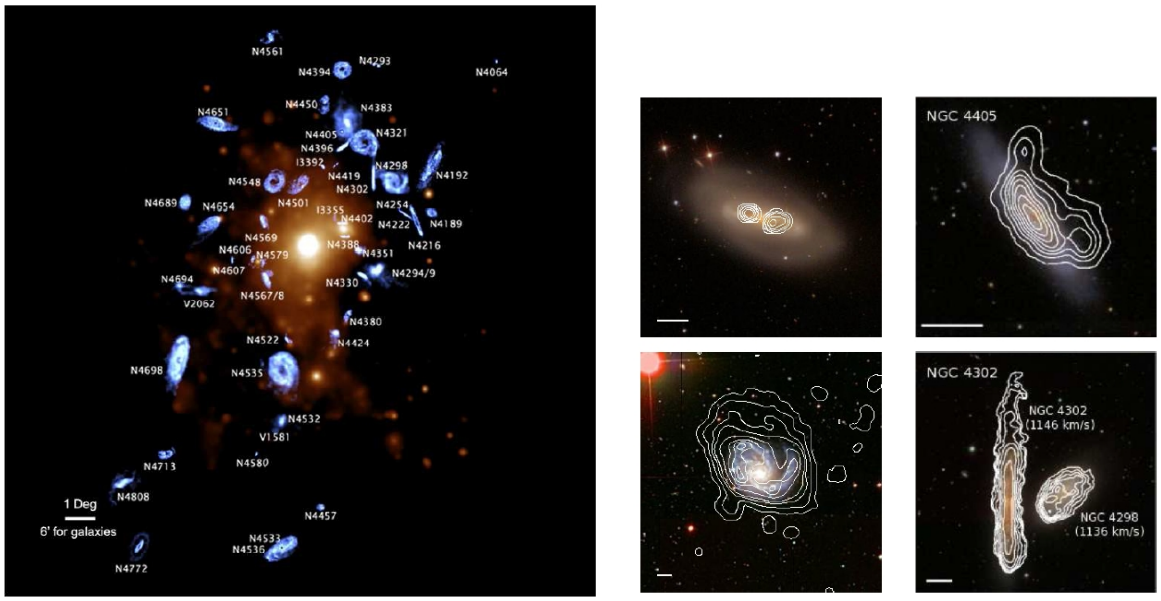


Figure 1.4: VIVA galaxies, from Chung et al. (2009). *Left:* a composite image of the HI data of the individual galaxies (in blue) overlaid on the ROSAT X-ray data (orange) by Böhringer et al. (1994). The galaxies are located at the proper position in the cluster but each HI image is magnified by a factor 10. *Right:* four examples of galaxies in the Virgo cluster. The HI contours from VIVA are overplotted on a SDSS composite image.

survey showed that in the denser regions of the cluster HI deficient galaxies with truncated HI profiles (as Figure 1.4, top row, left) are often associated with truncated star formation profiles. Together with examples of HI disturbed morphologies (top row, right) and one-side tails (bottom row), data pointed towards the mechanism of ram-pressure stripping, supported in addition by comparison with hydrodynamic simulations. VIVA also revealed a population of gas-rich galaxies at the edge of the cluster with long HI tails, suggesting that even far from the central denser regions environment has an impact on the gas content and consequently on the evolution of individual galaxies. Stripping has been observed acting on the very outer parts of disks where HI has low densities, so that the idea of a ram-pressure event far from the cluster center was not surprising.

While the nearby Virgo cluster is well studied, the effect of smaller environments on the atomic gas in galaxies is on the contrary still poorly constrained. A survey of 6 loose groups analogous to the Local Group (Pisano et al. 2011) measured a flatter HIMF than the equivalent field HIMF, consistent with the idea that the fraction of small HI galaxies increases with local density. One of the few other studies of HI in groups showed how this material can trace the encounters between galaxies, which are expected to be frequent when

the velocity dispersion is of the same order as the internal velocities of the galaxies in the group. Freeland et al. (2009) found that the majority of galaxies in spiral-rich groups are interacting, and in some cases multiple galaxies are observed inside a common HI envelope. In elliptical-dominated groups instead, likely at the end of the evolutionary sequence, the observed HI tidal debris are likely to be the remnants of the interactions.

In this section, we have presented a partial review of HI observations in nearby galaxies, and what we have learned from them. From the first detections to current studies, atomic hydrogen has proven to be a fundamental component to study dynamics, physics and evolution of galaxies. Unfortunately, few HI studies have specifically focused on the massive, gas-poorer range and on the mechanisms halting star formation in these systems. In Section 1.1, we have summarized the two main classes of quenching mechanisms expected to be at work in the high mass range. One includes environmental interactions, which are vastly studied in HI but almost only with respect to the dichotomy field/cluster and a particular focus on spiral galaxies. The other class comprises all kinds of processes that halt cooling of the hot gas or heat/removal the cold gas thus preventing star formation in central galaxies. Although the GASS survey represents a first step to gain insight into these physical processes, it still fails to detect a large fraction of red galaxies with high stellar surface densities (Figure 1.3, second and third panels). In addition, the GASS sample, although large, is not enough for the study of simultaneous cross-correlations among galaxy properties, which would allow us to identify second order dependencies. ALFALFA, on the contrary, although not designed for this analysis, produces a plethora of data and overlaps with large multiwavelength surveys such as SDSS and GALEX. Because of its characteristics, its sample is ideal to study massive galaxies over the broadest parameter space, and to study the role of the cold gas in the quenching mechanisms in the local Universe. Even if the ALFALFA sensitivity is not sufficient to obtain a complete census of the gas-poorer objects, its sample size allow us a statistical analysis through co-adding of non-detected data.

1.4 This work

In this thesis, we study the HI content of a complete, multiwavelength and statistically significant sample of massive galaxies, asking what is the relation between the amount of cold gas and other galaxy properties in the regime where current data are not sensitive enough, and how do quenching mechanisms affect this. With a complete and statistically

significant multiwavelength sample like the one we build, we can study if the HI scaling relations that we measure for the gas-rich objects hold for the gas-poorer as well. Once we clarify this, we can investigate whether there is a preferential mechanism depleting their gas content and quenching their star formation, turning them red and passive. Our work is statistical in nature, as it measures only average properties for bins of galaxies, but because of this characteristic it can point out the "weight" that the different mechanisms have in shaping the observed galaxy bimodality in the local Universe. In the framework of current models of galaxy formation, such an analysis is complementary to detailed studies of the processes, and equally fundamental. The characterisation of the average relations between galaxy properties from SDSS has shown the bimodality in several aspects of the galaxy population, and pointed out the "dominant" galaxy properties. These results are used as constraints in semi-analytics models of galaxy formations, but model outcomes are not yet entirely realistic because the input prescriptions are still incomplete. One of the missing components is the gas, especially information on its behaviour in the massive, gas poorer range. With our work, we are going to provide a statistical description of the atomic hydrogen content in massive systems, interpreting our results in light of the mechanisms proposed to quench the star formation.

To achieve this goal, we use data from ALFALFA, the state-of-the-art HI survey. Current blind HI surveys are too shallow to detect these systems, but if positions and redshifts of target galaxies are known, it is possible by co-adding signals from many objects to decrease the background noise. With this stacking technique, one can constrain average gas properties of galaxy classes which, individually, may be largely undetected but globally can yield a signal. We have therefore developed a software tool to co-add ALFALFA data, which we applied to a volume-limited sample of ~ 5000 galaxies selected to have stellar masses greater than $10^{10}\,\mathrm{M}_{\odot}$ and redshifts in the range $0.025 < z < 0.05$, and to lie in the intersection of the SDSS spectroscopic and the GALEX ultraviolet surveys. With these requirements, precise redshifts and estimators of different galaxy components are available. The chosen high mass range, probing both star forming and passive galaxies over a broad range of other galaxy properties (such as morphology, $\mu_{\star}$, AGN content), is suitable for the study of some of the mechanisms proposed to be responsible for the quenching of the star formation, and probably the evolution between the sequences.

1.4.1 Thesis outline

In the first part of this work we present the sample and the stacking tool. In particular, in Chapter 2 we describe the sample selection criteria and the main parameters that we have

derived from optical, ultraviolet and HI data and that we use for the analysis. In Chapter 3, we present the ALFALFA survey and data in detail, and introduce the stacking tool. We then describe a first application of the stacking, which also serves as a test for the developed procedure, in which we measure the HI scaling relations as a function of galaxy properties for our sample.

In the second part of the thesis, we address the scientific questions. In Chapter 4 we constrain the average HI content of a large, unbiased sample of massive early-type galaxies, and test a specific proposed model of morphological quenching. Chapter 5 addresses the question whether we see observational signatures of the feedback from an active nucleus on the cold gas properties of massive galaxies. We use both HI data from ALFALFA, and molecular hydrogen data from an on-going survey presented in the chapter, to probe the feedback at different galactic scales. Finally, in Chapter 6, we use ALFALFA data to measure for the first time the HI gas fraction as a continuous function of local density. The goal is to compare the effects of the environment at different densities on global and nuclear specific SFR and HI content, to gain insight into the mechanisms at work, using in addition comparisons with mock catalogs extracted from SAMs. The last part, Chapter 7, presents a summary of the work, with a discussion of its implications and future prospects.

2

The sample

This work is based on a complete sample of massive nearby galaxies, for which we collect HI, optical and ultraviolet (UV) data. Such a multiwavelength data-set is crucial for the analysis, as it allows us to estimate different structural and physical parameters for our targets like stellar mass, mass surface brightness, size, colour, star formation. The detailed knowledge of these galaxy properties is fundamental to study how the cold gas is related to, and how it evolves with, its host.

In this chapter, we first summarize the selection criteria that define our parent sample, which we call *sample A* (§2.1), and shortly outline its main properties (§2.2). Later, we present the multiwavelength data and the main parameters that we derive from them. We describe the HI data in Section 2.3, the optical data in Section 2.4, and finally the ultraviolet data in Section 2.5.

2.1 Selection criteria

The sample used in this work is based on the spectroscopic Sloan Digital Sky Survey (SDSS; York et al. 2000). From this data-set, we extract galaxies that fulfill the following requirements:

- their stellar masses lie in the range $10^{10} < \mathrm{M}_\star\ [M_\odot] < 10^{11.5}$;
- they are located in the intersection of the sky regions observed by the 6-th data release of SDSS (Adelman-McCarthy et al. 2008, hereafter DR6), the projected GALEX Medium Imaging Survey (MIS; Martin et al. 2005) and ALFALFA;
- they have redshifts in the range $0.025 < z < 0.05$;

The requirement on the stellar mass range is imposed by the scientific goal of this work (see Introduction).

The multiwavelength coverage of the sample is ensured by the second criterion. In particular, we select targets from the SDSS *spectroscopic* survey because it provides us with very accurate redshift measurements (the average uncertainty in the SDSS spectroscopic redshifts is 0.0002), which are essential if we wish to recover an accurate estimate of the mean HI content of a population of galaxies using stacking techniques.

Finally, the redshift range is determined by considerations on the characteristics and the completeness of the survey. The lower limit is set at $z = 0.025$ because: *i)* at smaller redshifts the SDSS spectroscopic targeting completeness falls below 50% for our stellar mass range. At $z = 0.025$ a stellar mass of few times $10^{10}\,M_{\odot}$ corresponds to $r \sim$13; the SDSS primary spectroscopic sample, which targets all galaxies with $r <$17.77, has completeness greater than 80% for $r >$14.5. *ii)* In addition, a single pointing on galaxies closer than z<0.025 may underestimate the HI flux if the gas disk size is larger than the Arecibo beam (the beam is 3.5 arcminutes in diameter, which corresponds to physical scales of 0.1-0.2 Mpc at the redshifts considered). The upper redshift limit is primarily set by ALFALFA, which observes up to redshift $z \sim$0.06. We further restrict ourselves to 0.05 in order to avoid a range where known radio frequency interference is a problem.

We remark that our selection criteria are the same used to extract the "parent sample" of the GASS survey (Catinella et al. 2010).

2.2 Sample properties

Our main limit on the data collection is the availability of HI data, whose acquisition is still on-going. The already available ALFALFA 40% data set (Martin et al. 2010; Haynes et al. 2011) includes the following SDSS sky regions: $7.5\,\mathrm{h} < \alpha_{2000} < 16.5\,\mathrm{h}$, $+4^{\circ} < \delta_{2000} < +16^{\circ}$ and $+24^{\circ} < \delta_{2000} < +28^{\circ}$, and $22\,\mathrm{h} < \alpha_{2000} < 3\,\mathrm{h}$, $+14^{\circ} < \delta_{2000} < +16^{\circ}$. Within this area, our selection criteria extract a total of 5350 objects, but we discard 11% of them because they have poor quality HI data (see section 3.2.1, c). Our final sample, hereafter identified as *sample A*, is therefore composed of 4726 objects. Their distribution in the sky is shown in Figure 2.1, where we also illustrate the potential sample we can build when ALFALFA is completed. In the plot, blue dots represent galaxies in *sample A*, the gray dots in the background are the targets that match our selection criteria but for which we still do not have ALFALFA data. As a note, the two vertical gaps correspond to the Galactic Plane and the horizontal one to a strip not covered by the spectroscopic SDSS DR6.

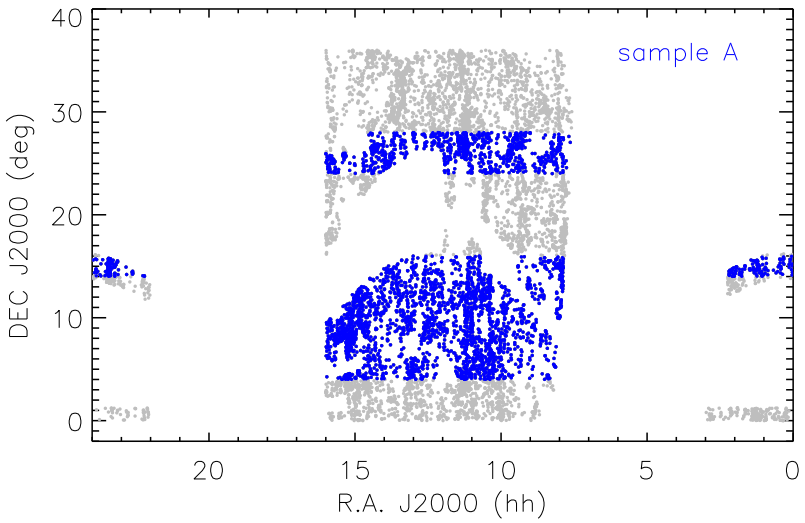


Figure 2.1: The sky distribution of galaxies meeting our selection criteria (cuts in $M_\star$ and z, and location in the intersection of the projected ALFALFA, SDSS DR6 and GALEX MIS surveys) is shown as gray dots. In blue we overplot *sample A*, i.e. galaxies for which we already have ALFALFA data.

Even if we do not yet have access to the all potential ALFALFA data, *sample A* includes already enough statistics to allow us for analysis over a broad range of galaxy properties. In Figure 2.2, we present the sample characteristics: the first three panels display as black histograms the stellar mass $M_\star$ (a), redshift z (b) and colour (c) distributions of the galaxies in *sample A* (we use NUV$-r$ colours, described in Section §2.5); panel (d) shows the colour versus absolute r-band magnitude (M_r) for our galaxies, as gray dots. It is evident that *sample A*, which is complete in stellar mass and redshift over the ranges considered, spans the whole interval of colours of the normal galaxy population, from the very blue, gas-rich objects to the red, gas poorer ones.

Out of our targets, only 23% of the galaxies have a reliable ALFALFA detection (i.e. objects corresponding to ALFALFA detection codes 1 or 2[1]). If we overplot the individual detections in all the panels (red histograms and dots) we note that the detection rate is even across the stellar mass range (∼18%), but the detections are clearly biased toward

[1] As discussed in Giovanelli et al. (2005), ALFALFA HI-line detections are coded into three different categories. Code 1 detections have a signal-to-noise ratio greater than 6.5 and are reliable at greater than 95% confidence. Code 2 detections, referred to as "priors", have a lower signal-to-noise ratio, between 4.5 and 6.5 but an optical counterpart at the same, known redshift. Their reliability is estimated to be greater than 85%. Code 9 (which we do not consider) refers to objects assumed to be high velocity clouds.

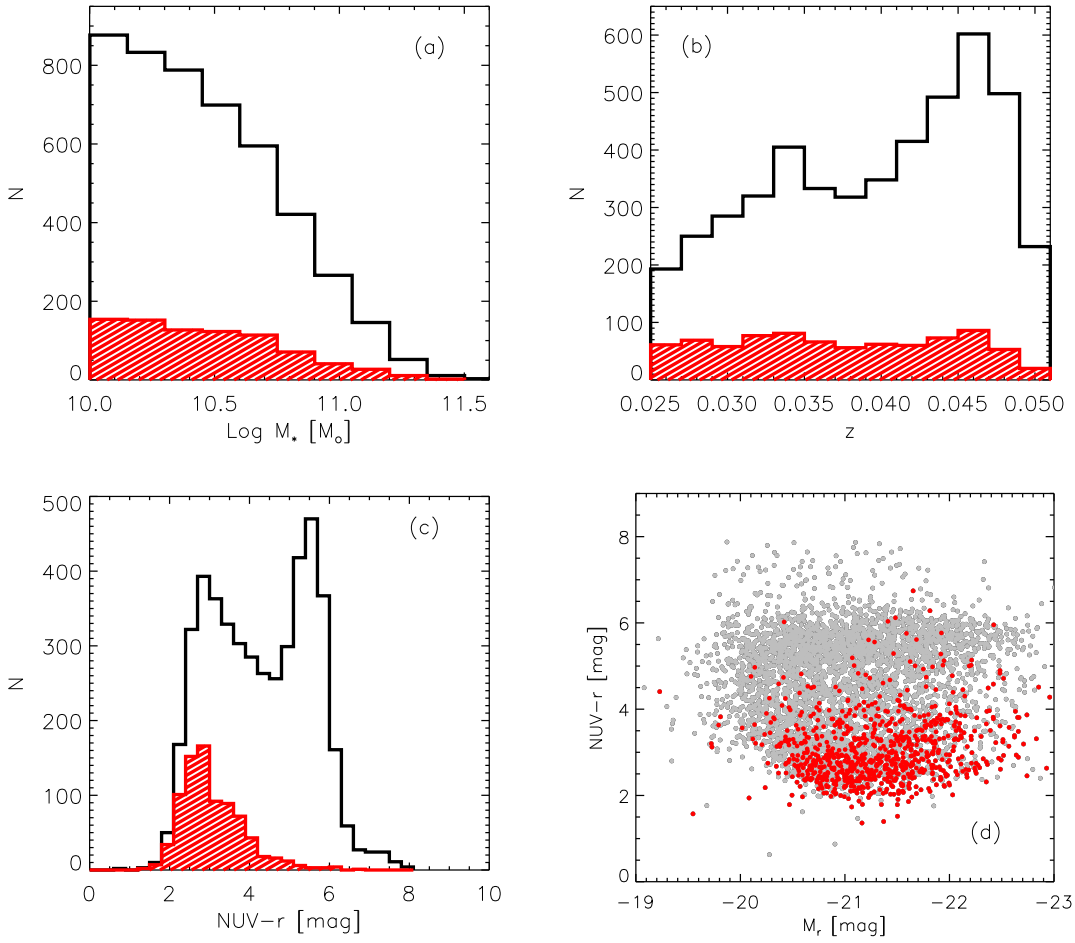


Figure 2.2: Distributions of stellar masses (a), redshifts (b) and NUV-r colours corrected for Galactic extinction (c) for galaxies in *sample A* (black solid histograms). Red, dashed histograms indicate the distributions for the sub-sample of galaxies with ALFALFA detections. Panel (d) shows the colour-magnitude diagram for *sample A* galaxies (gray dots) and for the sub-sample with ALFALFA detections (red dots).

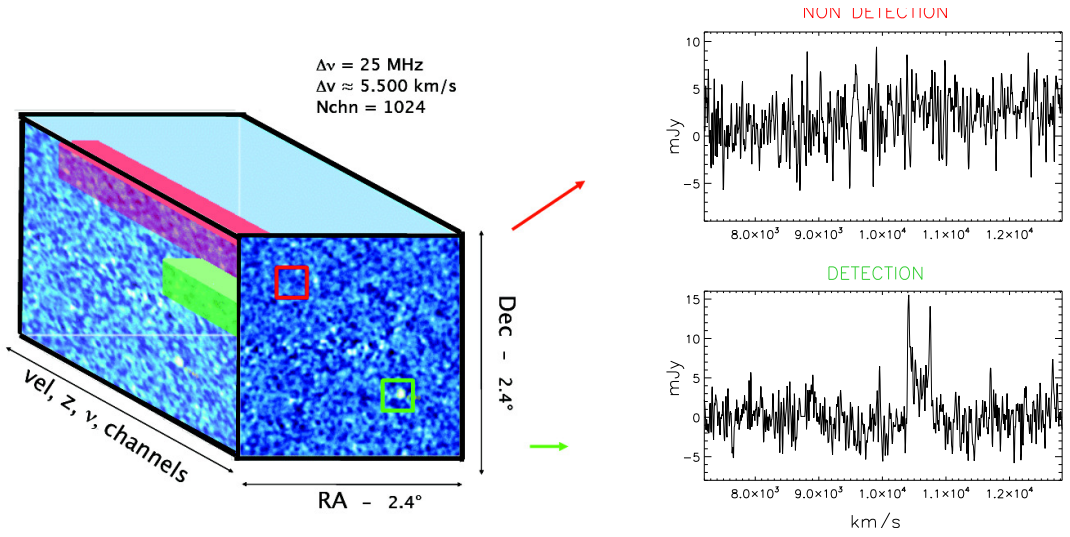


Figure 2.3: A schematic representation of the fully processed ALFALFA 3-D data-cube. See the text for further detail. For each target in *sample A*, we extract a spectrum at a given position of the sky, over the velocity range of the data-cube that contains the source. Two examples of extracted spectra are shown on the right, illustrating an HI detection (green, bottom) and an HI non-detection (red, top).

blue objects (with NUV$-r \lesssim 4$), which are on average gas richer and therefore above the ALFALFA sensitivity limit.

2.3 ALFALFA HI data and parameters

The on-going ALFALFA survey, which we already introduced in Section 1.3, is a blind HI survey of the extragalactic sky observable with the Arecibo telescope. It will scan 7000 square degrees of the declination range between 0 and 36 degrees, avoiding the Galactic Plane (the two vertical "holes" in R.A. in Figure 2.1). It is expected to detect more than 30,000 extragalactic HI sources over the velocity interval $v[\mathrm{km\,s^{-1}}] \simeq$ [-2500;18000], which corresponds to the local universe out to z$\sim$0.06. ALFALFA has an angular resolution of 3.5 arcminutes, the FWHM of the telescope beam at the frequencies observed. This resolution is significantly better than the one of previous blind surveys: HIPASS, for example, has an angular resolution of 15.5 arcminutes, 5 times lower. ALFALFA velocity resolution, after smoothing, is of $\sim$ 11 km s^{-1}.

The survey scans the sky using a 2-pass drift scanning technique (Giovanelli et al. 2005), which results in an effective integration time per beam solid angle of $\sim$48 seconds.

The 2-D scan data covering a given region of the sky are combined to form 3-D cubes of dimension $2.4° \times 2.4°$ and 5500 km s^{-1} in velocity "depth" (so that four cubes are needed for each sky region in order to cover the whole ALFALFA frequency range). In Figure 2.3 we have drawn a schematic representation of one such data-cube, and reported two examples of ALFALFA spectra that can be extracted. The final average noise level of a reduced spectrum is 2.2 mJy. As a reference, ALFALFA sensitivity would allow us to detect an HI mass of few 10^8 M$_\odot$ at the redshift of 0.025, assuming a signal-to-noise of 6 and a width of the HI line of 300 km s^{-1} (average for our massive galaxies).

More technical details about ALFALFA data and data reduction are presented in the chapter dedicated to the stacking tool (§3.1).

Since our targets are extracted from the SDSS spectroscopic survey, we have precise information about their positions and redshifts. We extract from the ALFALFA data-cubes a spectrum for each galaxy in the sample, independently of the fact that the signal is individually detected or not. The green and red regions in Figure 2.3 represent examples of these two cases respectively, with the corresponding extracted spectra displayed on the right. Once we have a catalogue of spectra we co-add them in bins of a given property, and if we recover a stacked signal we measure it. The parameters that we measure, or derive, from a stacked spectrum are: the HI flux, the HI mass and the HI gas fraction, as described below. Unlike for an individual detection, we do not measure the width of the HI stacked emission. For an individual line, the width is a measure of the (observed) rotational velocity of the emitting object, but in the stacking case the shape of the signal is the convolution of the individual ones co-added. Since we do not use this information in the work, we do not record the widths of the stacked profiles.

HI flux The total flux S_{int} is the integral of the signal detected, in Jy km s^{-1}. When we detect a signal, we simply flag by visual inspection the edges of the emission, and integrate inside this region. In particular, we consider detections the spectra for which the signal-to-noise ratio (evaluated following Saintonge 2007, and eq. 3.3 in the next Chapter) is greater than 6.5. When this is not the case, we evaluate an upper limit assuming a 5σ signal with a width of 300 km s^{-1}, smoothing the spectrum to 150 km s^{-1}. (The width of 300 km s^{-1} is chosen because it corresponds to the peak of the distribution of velocity widths for the detected galaxies in our sample.)

HI mass It is computed using the formula (Roberts 1963):

$$\frac{\mathrm{M_{HI}}}{\mathrm{M_\odot}} = \frac{2.356 \times 10^5}{1+z} \left(\frac{D_L(z)}{\mathrm{Mpc}}\right)^2 \left(\frac{S_{int}}{\mathrm{Jy\ km\ s^{-1}}}\right), \tag{2.1}$$

where S_{int} is the integrated HI line flux, z the redshift and $D_L(z)$ the luminosity distance.

HI gas fraction It is simply defined as $M_{HI}/M_\star$.

2.4 SDSS data and optical parameters

SDSS (for the survey specification see York et al. 2000) is an optical survey, which scanned the sky in the five bands: *u, g, r, i, z*. The targets for the spectroscopic survey are selected from the photometric catalogue of galaxies with $r \leq 17.77$ mag and r-band half-light surface brightness $\mu_{50} \leq 24.5$ mag arcseconds^{-2} (Strauss et al. 2002). Although our sample is extracted from the 6-th data release, we use the most recent DR7 (Abazajian et al. 2009) to obtain the parameters we are interested in. These are directly obtained from Structured Query Language (SQL) queries to the SDSS DR7 database server, or from the MPA-JHU SDSS DR7 release of spectrum measurements[2].

In the following we summarize the main parameters that, in addition to the redshifts, we use in this work. We start from the photometric ones, first presenting the parameters that we extract directly from the catalogues, such as magnitudes and stellar mass, then describing the derived ones as stellar mass surface density, concentration index and inclination.

Magnitudes. We use the SDSS *model magnitudes* as obtained from the database, which we only correct for the tabulated corresponding Galactic extinction factor. According to SDSS documentation[3], the reduction pipeline fits two models to the two-dimensional image of each object in each band:

- A pure de Vaucouleurs profile:

$$\mathrm{I(r)} = \mathrm{I_0} \exp\{-7.67[(r/r_e)^{1/4}]\}, \tag{2.2}$$

  truncated beyond $7r_e$ to smoothly go to zero at $8r_e$, and with arbitrary axis ratio and position angle;

- A pure exponential profile:

$$\mathrm{I(r)} = \mathrm{I_0} \exp(-1.68\, r/r_e), \tag{2.3}$$

[2]See http://www.mpa-garching.mpg.de/SDSS/DR7/
[3]See http://www.sdss.org/dr7/algorithms/

truncated beyond $3r_e$ to smoothly go to zero at $4r_e$, with arbitrary axis ratio and position angle.

The models are convolved with a double-Gaussian fit to the point spread function (PSF). In the database are also reported errors on the magnitudes.

We use model magnitudes for extended sources because they are a good proxy for Petrosian magnitudes but have smaller errors (the mean uncertainty for the model magnitudes is ∼0.02%, in contrast with the 0.1% of the Petrosian ones). The Petrosian magnitudes would be better measurements for galaxies, because they measure galaxy fluxes within a circular aperture whose radius is defined by the shape of the azimuthally-averaged light profile. This definition, introduced by Petrosian (1976), is not biased because measures a constant fraction of the total light, independent of the position and distance of the objects.

Stellar mass. Stellar masses, $M_\star$, are derived from SDSS photometry using the spectral energy distribution (SED) fitting technique described in Salim (2007) with a Chabrier 2003 initial mass function. A variety of model SEDs from the Bruzual & Charlot (2003) library are fitted to each galaxy, building a probability distribution for its stellar mass. The stellar mass assigned to a galaxy is therefore the mean of this distribution, while the measurement error is estimated from its width. Over our stellar mass range, these values are believed to be accurate to better than 30%, significantly smaller than the uncertainty on other derived physical parameters such as e.g., specific star formation rates. This accuracy is more than sufficient for this study.

Stellar mass surface density. The stellar mass surface density, $\mu_\star$, is defined as the stellar mass inside an effective radius. We measure it as $\mu_\star = M_\star/(2\pi R^2_{50,z})$, where $R_{50,z}$ is the Petrosian radius (radius inside which the Petrosian magnitude is measured) containing 50% of the flux in z-band in units of kpc. Mass surface densities trace the mass concentration, which appears to decrease with morphological class from ellipticals to spirals (as first noticed by Roberts & Haynes 1994).

Concentration index. The concentration index, C, is evaluated as $C{=}R_{90}/R_{50}$, where R_{90} and R_{50} are the radii enclosing 90% and 50% of the r-band Petrosian flux. This parameter is a good indicator of the bulge-to-total ratio (B/T) of the light distribution, as shown by (Gadotti 2009) with full bulge/(bar)/disk decomposition analysis on the 2-dimensional SDSS imaging. Because of its correlation with the bulge-to-

total light, it can be used to discriminate galaxies according to their morphology in disk- and bulge-dominated (as we do in Chapter 4). As a reference, $C \geq 2.6$, which is the threshold we use to select bulge-dominated galaxies, corresponds to B/T$\geq$0.3 (see also Fig.1 in Weinmann et al. 2009).

Inclination. The inclination i to the line-of-sight is evaluated according to: $\cos i = b/a$, where a and b are the semi-major and semi-minor axes from the r-band exponential fit, respectively (Eq. 2.3). b/a for the exponential fit is tabulated as *expAB_R* in the SDSS database. Inclinations are used only as an input of the stacking tool, as explained in Section 3.2.1.

In addition to the described photometric parameters, we extract some spectroscopic ones, namely: fluxes of several emission lines, central velocity dispersion, the $D_n(4000)$ parameter and fibre (specific) star formation rates.

Emission lines. We extract the following emission line fluxes from the MPA-JHU catalogue: Hα, Hβ, [NII], [OIII] (in units of $10^{-17}\,\mathrm{erg\,s^{-1}cm^{-2}}$), which we use to classify the nuclear activity of our galaxies in Chapter 5. The fluxes are measured from a Gaussian fit to the SDSS spectra, after the continuum has been subtracted.

We have to correct the fluxes for dust extinction, which attenuates the observed flux F_{obs}:

$$F_{\lambda;obs} = F_\lambda \mathrm{e}^{-\tau_\lambda}, \tag{2.4}$$

where F_λ is the emitted flux at a given wavelength λ and τ_λ is the corresponding optical depth. To estimate the attenuation correction we adopt the prescription introduced by Wild et al. (2007):

$$\frac{\tau_\lambda}{\tau_\mathrm{v}} = (1-\mu)\left(\frac{\lambda}{5500\text{Å}}\right)^{-1.3} + \mu\left(\frac{\lambda}{5500\text{Å}}\right)^{-0.7} \tag{2.5}$$

where τ_V is the total effective optical depth in the V band and μ is the fraction of it caused by the ambient interstellar medium (ISM). This dust prescription tries to account for the attenuation of the emission lines within their own birth clouds (first term; the power law index is obtained from data of 3 local galaxies) and the attenuation caused by the ISM (second term; the index is given by a fit to the ultraviolet to infrared continua of starburst galaxies). Following Wild et al. (2007), we set μ= 0.3.

Finally, we have to assume an intrinsic Balmer ratio $F_{H\alpha}/F_{H\beta}$, from which we can extract τ_V inverting Equation 2.4. We assume the value of 2.87, estimated for an idealized HII region (Osterbrock 1989, Case B nebula).

Central velocity dispersion. We acquire the velocity dispersion σ as the *vdisp* parameter in the SDSS database.

In a galaxy, σ is the result of the convolution of the individual stellar spectra, Doppler shifted by the motion. In the SDSS pipeline, σ is evaluated using a "direct fitting" method to the galaxy spectrum[4]. The method consists in comparing the spectrum of the analysed galaxy with a fiducial spectral template, which can be either the convolution of different stellar type spectra, or an high S/N spectrum of a galaxy with known velocity dispersion. Dispersions are measured from the 3-arcseconds fibre spectrum, therefore they are characteristic of the inner regions. According to the SDSS documentation, only values between 70 and 420 $\mathrm{km\,s^{-1}}$ are reliable. Velocity dispersions smaller than 70 km $\mathrm{s^{-1}}$ are not reliable because of the resolution of the SDSS spectra ($\lambda/\delta\lambda \sim$1800). The maximum σ to which the spectra are convolved is 420 km $\mathrm{s^{-1}}$, which renders larger velocity dispersions not reliable.

We correct *vdisp* for aperture effects, which can be significant for the farther objects. Following Graves et al. (2009a) and references therein, we scale the fibre velocity dispersion to that at 1/8 of the effective radius: $\sigma_{1/8} = vdisp \cdot \left(r_{fib}/\frac{1}{8}r_\circ\right)^{0.04}$, where $r_{fib} = 1''.5$ and $r_\circ$ is the circular galaxy radius, defined as $r_\circ = R_e\sqrt{b/a}$. R_e is the r-band de Vaucouleurs radius (tabulated as *devR_r*), and a, b are now the major and minor axis from the de Vaucouleurs fit (b/a is tabulated as *devab_r*). In general, this correction is small, $\sim 5\%$. (Note that for galaxies included in *sample A*, R_e has a mean value of $8.2''$.)

$\mathbf{D_n(4000)}$. The break strength at 4000 Å (the strength of the absorption feature bluewards 4000 Å in a spectrum, hereafter $D_n(4000)$) is defined as the ratio of the flux in the red part of the continuum (e.g., from Balogh et al. 1999, 3850–3950Å) to that in the blue continuum (e.g., 4000–4100Å). In a composite stellar system, the strength of this discontinuity is a function of the continuum emission itself and the strength of the absorption features, which depend on temperature and metallicity of the stars. As discussed in Kauffmann et al. (2003b), $D_n(4000)$ can serve as a measure of the age of the stellar population of galaxies when calibrated using stellar population synthesis models. There is however a secondary dependence on metallicity, which is stronger for older stellar populations.

$D_n(4000)$ is measured from the SDSS 3 arcseconds-diameter fibre spectrum. The galaxies in our sample are located at low redshifts ($0.025 < z < 0.05$), and in 72%

[4]See http://www.sdss.org/DR6/algorithms/veldisp.html for more discussion

of our sample the 3 arcseconds fibre covers less than 20% of the Petrosian radius enclosing 90% of the total light of the galaxy. The 4000Å break values should thus be regarded as a measure of stellar population age for the inner regions of the galaxies.

Fibre specific star formation rate. The specific star formation rate is defined as the star formation rate (SFR), per unit of stellar mass: SFR/$M_\star$ [yr^{-1}]. To estimate it, we acquire both star formation rates and stellar masses measured inside the SDSS fibre, from the MPA-JHU catalogue. In particular, the fibre stellar masses are obtained from fits to photometry inside the fibre with the method already explained for the total masses. The star formation rates, instead, are evaluated from the spectra following Brinchmann et al. (2004), using the models of Charlot & Longhetti (2001) for computing the line and continuum emission from galaxies optimized for integrated spectra. The method differs for the different classes of galaxies. Briefly, *(i)* the SFRs of pure star forming objects are derived directly from the spectrum emission lines. A grid of models with different parameters (among which metallicity Z and the total dust attenuation in the V-band) is generated; the likelihood of each model given the data is calculated; the SFR is then the best-fitting value. *(ii)* For objects whose signal is contaminated by AGN emission, or for objects with low S/N of the emission lines, the SFRs are derived indirectly from the break strength at 4000 Å. The likelihood distribution for SFR/$M_\star$ as a function of $D_n(4000)$ is built using star-forming galaxies only. The method applied to the MPA-JHU catalogue differs here with respect to the one of Brinchmann et al. (2004): for AGN or composite galaxies where the S/N of Hα and Hβ is greater than 3, the probability distribution function is estimated using only star-forming galaxies with similar Hα/Hβ. This step removes trends with dust attenuation, and is possible on the new data because of the improved statistics.

2.5 GALEX data and NUV parameters

The Galaxy Evolution Explorer (GALEX) is an orbiting space telescope observing galaxies in ultraviolet light with two imaging and two spectroscopic channels in far ultraviolet (FUV – extending from 1350Å to 1750Å) and near ultraviolet (NUV – from 1750Å to 2750Å).

The GALEX mission is performing several surveys, each with different characteristics (Martin et al. 2005). The larger one is the All-sky Imaging Survey (AIS), with average exposure times of 100s over 40,000 square degrees of the sky. The intermediate one, or Medium Imaging Survey (MIS), is characterised by longer exposures (average 1.5ks) of 1000 square degrees of the sky, with extensive overlap with the SDSS. Finally, the Deep

Imaging Survey (DIS) aims to integrate 30ks but over a limited sky region of 80 square degrees.

Our sample is selected from the projected MIS survey, whose integration times allow to reach NUV sensitivities of $m_{AB} \simeq 23$. With such low magnitudes, we can probe the full colour range and specific star formation rates of the normal galaxy population. Unfortunately the MIS is still on-going, so we do not yet have data for 70% of the objects. For these galaxies we use the AIS, which reaches depth of $m_{AB} \simeq$20-21.

The UV parameters we use are extracted from the GALEX UV photometry and reprocessed by Wang et al. (2011). As explained there in more detail, the process includes smoothing the GALEX and SDSS images to a common point spread function (PSF). The SDSS r-band images are convolved to the resolution of the UV imaging, which is lower, before a fitting routine is used to calculate magnitudes in consistent apertures, therefore ensuring that measurements in different bands represent similar physical regions of the galaxies. A description of colour and star formation rate measurements is presented below.

NUV-r colour. The colours we use are derived from the magnitudes measured on the GALEX NUV and SDSS r-band images. NUV and r magnitudes are corrected for Galactic extinction following the prescription of Wyder et al. (2007). They adopt $A(\lambda)/E(B\text{-}V)=2.751$ for the SDSS r-band and $A(\lambda)/E(B\text{-}V)=8.2$ for GALEX NUV. With these assumptions, the corrections for NUV$-r$ can be estimated as: $A_{NUV}\text{-}A_r=1.9807A_r$, where A_r denotes the extinction obtained from the SDSS database (listed there as ext_r). Corrections for internal dust-attenuation are very uncertain for galaxies outside the blue sequence, especially in absence of far-infrared data. As a consequence, we do not only apply these corrections to the colour estimates.

NUV$-r$ colours can be considered a proxy for the global specific star formation rate (i.e. star formation rate per unit stellar mass), since they measure the ratio (as logarithmic difference of magnitudes) between the emission coming from young and old stars. The UV band is in fact dominated by B stars and traces the star formation over the last $\sim$ 100 Myrs, while the r-band, dominated by old stellar populations, measures the total light coming from the galaxy. Moreover, NUV$-r$ colours span a wider dynamic range than other optical derived ones, thus the bimodality of the galaxy population is better displayed. This is clearly visible if we compare the colour – magnitude diagram obtained with NUV magnitudes, reproduced here for convenience (Figure 2.4, right panel), with one built using only optical colours from SDSS. In Figure 2.4 (left panel) we show the diagram using $u-r$ colours, corrected for Galactic extintion. Not only the dynamic range spanned is smaller, but the bimodal trend is

also less defined.

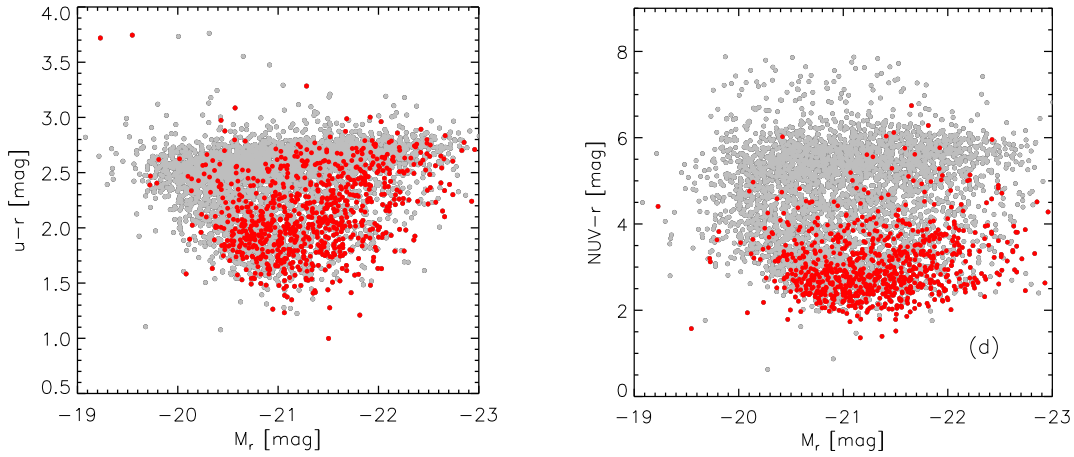


Figure 2.4: NUV-r colour – magnitude diagram (right) compared to the $u-r$ colour – magnitude diagram (left) for *sample A*. The optical colours have been derived from the SDSS magnitude corrected for Galactic extinction. Red dots are the sub-sample with ALFALFA detections.

Star formation rates. The star formation rates are obtained from spectral energy distributions (SED) fitting technique to the five optical and two UV bands, from Wang et al. (2011). A more thorough discussion of this procedure is presented there and in Saintonge et al. (2011b). Here we only summarize it briefly. They generate a population of model galaxies from Bruzual & Charlot (2003). These population synthesis models span a range of ages, exponentially declining star formation histories, metallicities and dust attenuations. The models are fitted to the galaxy's SED, and for each galaxy the SFR and dust attenuation are the χ^2-weighted average of the model parameters. This approach is similar to Salim (2007), but for the empirically-calibrated dust extinction coefficient (A_v). A_v is a critical parameter to obtain accurate SFRs, and a direct estimate of it would be possible only knowing a galaxy SED from UV trough far infrared. Therefore they estimate A_v from the SDSS/GALEX data using either the Balmer decrement or a combination of $D_n(4000)$ and NUV$-r$ (or both) and calibrate it on the Johnson et al. (2007) sample. They adopt different prescriptions for A_v based on the SDSS data for different colour intervals, because they find that different combinations of the SDSS parameters minimize best the scatter with the direct measures of A_v in different NUV$-r$ ranges.

3

Stacking of HI spectra

Based on
S. Fabello, et al.
MNRAS, 2011, 411, 993

Stacking has recently become a common tool to constrain the statistical properties of a population of objects that lack individual detections in a survey. If one has an ensemble of objects that are too faint to yield individual detections but knows their positions (and redshifts if one works with spectra), co-adding their data allows to decrease the noise and recover the total signal of the ensemble. This technique has already been applied to a wide variety of different astrophysical data. Examples include stacking of maps of the 20 centimeter emission of optically selected galaxies, as in Hodge et al. (2009); they selected faint radio active galactic nuclei in luminous red galaxies, the vast majority of which is radio-quiet and would not be detected even in deep surveys, and were able to constrain their typical flux densities. A similar analysis was carried out by Shen et al. (2006) to study soft X-ray properties of high redshift quasars selected from SDSS. Other examples include studies of the star formation in high redshift Lyman Break Galaxies (Carilli et al. 2008), and studies of the spatial distribution and the colour of the intracluster light using stacked optical images (e.g., Zibetti et al. 2005).

Stacking can of course be applied to HI data as well, and in the past decade has been used by several authors to explore the properties of the cold gas beyond $z \sim 0$, at redshifts that are currently not well probed by existing facilities. The approach was similar in all the works: HI interferometric data (to have the angular resolution necessary to resolve the individual galaxies at those distances) have been collected for regions of the sky

already covered by optical surveys so that redshifts were available. Spectra have then been extracted for the individual galaxies at their known positions, and stacked. In one of the first works, for example, Chengalur et al. (2001) observed a cluster at z∼0.06 and extracted spectra for about 150 known galaxies with available redshifts. The stacked spectrum had a low signal-to-noise, nevertheless yielded a signal almost an order of magnitude lower than what would have been possible for an individual source. More recently, Lah et al. (2007) used the stacking to measure the HI content of 121 star-forming galaxies in the field, while Lah et al. (2009) attempted to constrain, with the same technique, the gas content of the furthest cluster observed in HI, at z∼0.37. They stacked 324 galaxies but obtained a marginal detection, because their data were not deep enough. As last example, Verheijen et al. (2007) observed in a pilot study two clusters at z∼0.2, integrating ∼240 hours on one and ∼180 hours on the other. With such deep data they were able to detect 42 galaxies, and used in addition the stacking to exploit the non-detections.

In conclusion, the experience from previous works has shown that stacking can be a powerful technique, but requires sensitive data and large samples. To have these requirements fulfilled at higher redshifts we have to wait for the new-generation HI surveys, but in the local universe large HI observational campaigns are already on-going. ALFALFA, for example, is blindly scanning a large volume on the sky of ~ 7000 deg ^{2} and is expected to detect more than 30,000 extragalactic HI sources out to z∼0.06. It will produce a plethora of data, both detections and non-detections, so that it will be possible to extract spectra for any desired galaxy, granted that we know its position and redshift. The stacking of ALFALFA data hence allows to probe gas-poorer regimes where we lack of individual detections, and it becomes useful and necessary to complement any study of HI properties of nearby galaxies.

We have developed a stacking tool to exploit the data from ALFALFA, but this tool can in future be adapted for any of the new generation HI surveys, to broaden the range of possible HI science also at higher redshifts. Due to the importance of the technique for this work, we dedicate this chapter to a detailed description of the stacking tool. We first summarize the relevant properties of the ALFALFA data and data reduction in §3.1, then describe step by step the stacking process (§3.2). The last part of the chapter (§3.3) discusses a first application, which also serves as a test for the developed procedure. In this application, the HI scaling relations for our sample are determined as a function of galaxy properties, and compared to the results of Catinella et al. (2010).

3.1 ALFALFA HI data

We already introduced the ALFALFA survey in the previous chapters (§ 1.3 and § 2.3). Here we describe the structure of the ALFALFA data and the reduction pipeline, focusing on the details that are relevant to the stacking procedure.

ALFALFA is scanning the high galactic latitude sky using a simple "minimum-intrusion" drift scanning technique (Giovanelli et al. 2005), which exploits the new seven-horn Arecibo L-band (the radio band including 21 centimeters) feed array. Without this receiver, called ALFA and installed on 2004, the survey would not be feasible: with a single pixel the same region of the sky would have been scanned seven time slower, so that up to 50 years would be needed to complete ALFALFA. For each of the seven beams, data are recorded separately for the two orthogonal polarizations. This not only provides two independent samples of the HI signal (as the 21 cm emission is not polarized), but also allows to discard in a second time spurious signals that appear only in one of the two samplings (as interference may be polarized). The 2-D drift scan data from both polarizations and all beams covering a given portion of the sky are combined to form 3-D cubes of dimension $2.4^\circ \times 2.4^\circ$ on the sky and 5500 $\mathrm{km\,s^{-1}}$ in velocity "depth". We refer the reader again to Figure 2.3, where a schematic representation of one such data-cube is shown. The raw spectral resolution of the ALFALFA data, before smoothing, is $\sim$5.5 $\mathrm{km\,s^{-1}}$ and the angular resolution is $\sim 3.3' \times 3.8'$ (corresponding to the FWHM of each ALFA beam). The data-cubes or "grids" are constructed from the drift scan data so that each spatial pixel is $1'$ on a side. The next step in the reduction process is the calibration of the 2-D drift scans; a noise signal from a diode of known temperature is injected in the receiver, so that by comparison with the data it is possible to calibrate the observed flux. The positions and flux scales of the final grids are improved using fits to the many radio continuum sources they contain (Kent et al. 2008).

The ALFALFA processing scheme (Giovanelli et al. 2005) retains all individually recorded spectra for each ALFA beam and each polarization separately. At this stage no filtering is performed to discard bad channels. However, in order to deal with poor quality data, the 2-D data for each beam and polarization are visually inspected to flag bad records and frequency channels. Data quality, in fact, may be limited by occasional instrumental problems or incomplete sky coverage, or channels may be flagged as bad because they are contaminated by interference, also known as RFI (Radio Frequency Interference). RFI consists of man-made signals whose strength is often larger than any extragalactic source. Radars and GPS are only the most common examples of devices that can produce it in the frequency interval we observe, but any mobile communication device, ground-based

and satellite-based navigation systems is a source of interference (See also, e.g. Burke & F. 2011).

The information about possible bad data are used during the 3-D grid construction. In order to account for possible effects that degrade the signal, each 3-D pixel of the data-cube is assigned a quality weight w ranging from 0 (unusable data) to 20 (good data), which is computed according to the availability and quality of data contributing to each pixel. The retention of the full data-set and the construction of the accompanying "weights map" allows us to judge whether or not adequate data exist for each target, so that it can be meaningfully included in the stacking process.

3.2 ALFALFA stacking tool

The stacking process consists of several different steps, described in detail in the followings. We extract a spectrum for each galaxy in our sample and evaluate its quality, creating a catalogue of reliable spectra, which we use for the analysis. We proceed to co-add the spectra of the desired targets and finally we measure the signal of the stacked objects.

3.2.1 Creating a catalogue of HI spectra

All our targets are selected from the SDSS spectroscopic survey, so we accurately know both their positions on the sky and their redshifts. To extract a spectrum for a given galaxy, we first select the ALFALFA data-cube that contains the target and then: *(a)* integrate over the sky to extract the raw spectrum; *(b)* evaluate the noise level of the spectrum to establish its quality; *(c)* perform a final quality check. Here we discuss these steps.

a. Spectrum extraction

The signal from each target is integrated over the whole frequency range of the corresponding data-cube and over a sky region centered on the object position. Because noise increases with the square root of the integration area, integrating over too large a region lowers the quality of the spectrum without increasing the signal. Our targets are always smaller than the ALFA beam (the mean R_{90} for *sample A* is $10''$), so we simply integrate over a sky region of $4'\times4'$. In Figure 2.3, we illustrate examples of spectra extracted at two different positions of the sky inside the same data-cube. The coloured regions indicate where the spectra would be evaluated (but their size is not scaled).

The HI spectrum is an histogram of flux density S as a function of velocity. For each velocity channel v, the corresponding flux density S_v is obtained by integrating the signal

$s_v(x, y)$ over the spatial pixels centered at the target galaxy position, as observed by a radio telescope of beam response pattern B:

$$S_v[\mathrm{Jy}] = \frac{\Sigma_x \Sigma_y s_v(x, y)}{\Sigma_x \Sigma_y B(x, y)},$$

where x, y are the sky coordinates (the two polarizations are kept separated). The expression above means that the spatially integrated profile is obtained by summing the signals over all the spatial pixels of interest and dividing by the sum of the normalized beam $B(x, y)$ over the same pixels (for a detailed discussion see Shostak & Allen 1980). The ALFALFA beam pattern can be approximated by:

$$B(x, y) = \exp\left(-\frac{1}{2}\left(\frac{x}{\sigma_x}\right)^2 - \frac{1}{2}\left(\frac{y}{\sigma_y}\right)^2\right),$$

with $\sigma_x = (2\sqrt{2\ln 2})^{-1} \times 3.3'$, and $\sigma_y = (2\sqrt{2\ln 2})^{-1} \times 3.8'$ (Giovanelli et al. 2005).

We note that we discard a spectrum if more than 40% of the pixels of the integrated region have a quality weight w less than 10. We also keep track of the three strongest continuum sources in an area covering $40' \times 40'$ around each source, since they can affect our spectra by creating standing waves. Standing waves are periodic fluctuations in the background, which occur when radiation from a source located near the target is multiply reflected and scattered by the telescope structure before reaching the receiver. These multiple reflections, which happen in telescopes with blocked aperture like Arecibo, can degrade the data and increase the noise. In the following we do not discard blindly any spectrum because of standing waves, but we record the strongest continuum sources for a possible subsequent inspection. Moreover, this information would be useful if an automatic quality threshold had to be implemented.

b. *Rms* evaluation

For each spectrum we need to measure the noise level, which is an indicator of the quality of the spectrum and which will later be used as a weighting factor when we stack spectra. Ideally the signal is zero in any channel outside the galaxy emission region; in practice there are fluctuations of the signal measured in the different channels. As noise estimate we use the standard deviation of the measured baseline from the expected zero line, hereafter identified as *rms* (root mean square). The *rms* has to be evaluated in regions of the spectrum where there is no emission from the target galaxy and where there are no spurious signals, such as HI emission from companion galaxies. In Figure 3.2, for example, some of these features are visible in the first two rows in the velocity range [-1500;-1000] km s^{-1}.

These strong RFI "absorption features" must not be included in the *rms* evaluation; since they are far from the galaxy signal we anyway do not need to discard the spectrum.

In order to define the spectral region that might contain galaxy emission, we estimate its expected observed HI width. The measured width of the HI line depends on the rotational velocity of the galaxy, which can be estimated from its luminosity using the well-known Tully-Fisher relation (Tully & Fisher 1977), and on the galaxy inclination since we actually measure the velocity component along the line-of-sight. For each target we estimate the expected observed velocity $w_{TF;o}$ from the SDSS data, following Giovanelli et al. (1997a). We use the i-band magnitude, k-corrected and corrected for Galactic and internal extinction (according to equations 11 and 12 in Giovanelli et al. 1997b) to estimate w_{TF}, and then project this value on the line-of-sight using the estimated inclination in order to obtain $w_{TF;o}$. We are aware that the Tully-Fisher relation does not hold for all morphological types and environments, but we think this is not a major issue since these velocities are only used to estimate the region of the spectrum that should contain a significant signal from the galaxy. In any case, the mean observed velocity that we estimate, of about $300\,\mathrm{km\ s^{-1}}$, is compatible with the one measured by the GASS survey for the same stellar mass range (Catinella et al. 2010).

Once we know where the signal from the target should be located, we apply a robust linear fit to the baseline excluding these channels and regions of the spectrum with a quality factor $w < 10$ (half the maximum value). This step allows us to eliminate possible gradients in the background, as for example seen in the top right panel in Figure 2.3, where the spectrum has a tilted baseline. We do not try to eliminate possible residual wavy baselines at this stage, as we will perform a manual reduction of the final stacked spectrum. We underline that the ALFALFA baselines are already almost flat and all with similar noises, and both characteristics are essential for the stacking. The quality of the baselines and their homogeneity is confirmed by the *rms* that we measure over the same regions of the spectrum just fitted. The average *rms* for the whole sample is 3.6 ± 0.5 mJy for each polarization. After averaging the two polarizations, a *rms* of $\sim$2.5 mJy is obtained, which is comparable to the average *rms* of 2.2 mJy evaluated for published, individually reduced ALFALFA spectra of detections.

c. Final quality check

After we have extracted the spectra, we visually inspect each of them. We check the extraction process, and we discard spectra with bad baselines caused by continuum sources and those with possible spurious signals close to the galaxy. For example, if one polarization has a significantly stronger signal than the other, we discard the spectrum because it is

contaminated by RFI, as the 21 cm emission is not polarized. These cuts eliminate 624 objects (out of 5350) from the initial sample.

3.2.2 The stacking method

We want to co-add the signals from N different sources located at different redshifts, i.e. at different positions in the bandwidth. First we shift each spectrum to the galaxy rest frequency, so each spectrum is centered at zero velocity. We stack together the spectra S_i (i=1,..N) using their *rms* as a weight, a standard approach in stacking analysis. Thus the final spectrum S_{stack} is determined by:

$$S_{stack} = \frac{\Sigma_{i=0}^{N} S_i \cdot w_i}{\Sigma_{i=0}^{N} w_i} \tag{3.1}$$

with the weighting function:

$$w_i = \frac{1}{rms_i^2} \ . \tag{3.2}$$

The stacking of the spectra is done separately for each polarization. Note that ALFALFA is a blind survey, which scans the sky with fixed integration time, so the *rms* are similar for most spectra. If the noise of the input targets is purely Gaussian, the *rms* of the stacked spectrum is expected to decrease as $1/\sqrt{N}$, where N is the number of objects co-added. In reality, there are likely to be systematic components in addition to the Gaussian noise, for example standing waves (we discarded spectra dominated by standing waves, but a weak residual signal could still remain). Because of these additional noise sources, we expect the *rms* to approach an upper limit as the number of co-added spectra becomes very large and the non Gaussian noise dominates. In Figure 3.1 (left panel), we show how the measured *rms* of the stacked spectra decreases as a function of the number of co-added objects. We stacked increasing numbers of randomly selected galaxies from *sample A*, and for each stack we evaluated an *rms* as described above. These measurements are shown as dots, while the dashed line is the expected average *rms* value assuming purely Gaussian noise, i.e. $2.5/\sqrt{N}$ mJy. We see that the *rms* decreases with the increasing number of spectra as $(\sqrt{N})^{-1}$, and then flattens as conjectured when the non Gaussian noise becomes dominant. For our data, it ceases to decrease significantly as N approaches values of about 300. The same trend of *rms* versus number of stacked objects is displayed by our simulation (see also Appendix A), where we can reproduce the flattening of the relation if we add a sinusoidal signal to the noise mimicking the non-Gaussian component (e.g. standing waves). This trend is illustrated in Figure 3.1 (right panel). Black points show the *rms* decrease if

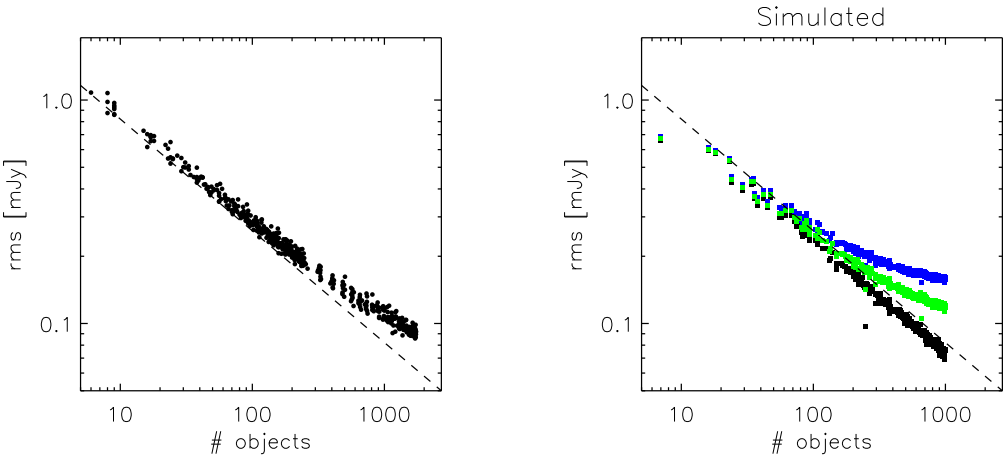


Figure 3.1: Dependence of the *rms* of the stacked spectra as a function of the number of objects co-added. The dashed line is the expected $1/\sqrt{N}$ dependence. *Left:* results obtained stacking real data: the flattening of the relation around $N \sim 300$ arises because the non-Gaussian noise becomes dominant. *Right:* black points are obtained stacking simulated spectra with only Gaussian noise. The coloured points represent the case in which we include an additional sinusoidal component to the noise: for a stronger component (blue) the flattening starts at smaller N.

we consider only a Gaussian noise component, in which case the noise does not deviate from the expected relation. An increase of the non-Gaussianity, instead, translates into a flattening for smaller galaxy ensembles: both green and blue points are obtained by adding a sinusoidal component to the noise, larger for the blue case.

After stacking, we manually process each co-added spectrum. The reduction includes the following steps. *i)* We average the two polarizations. *ii)* A default Hanning smoothing[1] is performed on the spectrum so that the final resolution is ~ 11 km s^{-1}. We may apply a boxcar smoothing to decrease the noise if the signal is marginal and the signal-to-noise is low. In any case we never average over more than nine channels (corresponding to $\Delta v \sim 50\,\mathrm{km\,s^{-1}}$). *iii)* Finally, the baseline is subtracted. We note that the baseline of a stacked spectrum is already almost flat (we have already automatically fitted a line to the individual baselines, which accounts only for possible gradients; in addition, different possible noise features tend to cancel out when we average many spectra). To eliminate possible residual features, we fit at this point a polynomial of maximum degree 2 to the baseline.

[1]Hanning smoothing consists in smoothing three channels together using the algorithm: $x(i) = 0.25 \cdot x(i-1) + 0.5 \cdot x(i) + 0.25 \cdot x(i+1)$.

Figure 3.2 shows some examples of reduced stacked spectra, obtained by stacking *sample A* galaxies in five different stellar mass intervals. The vertical axis in each panel shows the flux density in mJy, while the horizontal one shows the velocity in $\mathrm{km\,s^{-1}}$. Since we shifted each object to its rest-frequency, the stacked signal is centered at zero velocity. For each stacked spectrum, the mass range and the number of objects stacked are reported, as well as the signal-to-noise ratio of the detected HI signal. As explained above, when the signal-to-noise is low, we smooth the spectrum over up to nine channels as done, for example, for the last spectrum in Figure 3.2. In some stacked spectra there are spikes/holes caused by poor quality data, but the individual spectra containing the bad data were not discarded because the bad pixels are located far away from the central regions of interest. Examples of bad regions occur in the first two rows in the velocity range v = [-1500; -1000] $\mathrm{km\,s^{-1}}$. In the left column of the Figure, we show spectra obtained by stacking ALFALFA detections and non-detections together, which is the approach used in this work. In the right column instead, we plot the spectra obtained by stacking only the non-detections, to demonstrate that the stacked signal is not dominated by few individual detections: the process recovers a signal even if no individual galaxy is detected (dotted lines in Fig. 3.2 indicate the edges of the emission). Notice that the *rms* decreases with increasing number of co-added objects. The width of the profile is smaller for less massive objects, as expected from the Tully-Fisher relation since they have on average lower circular velocities.

If we recover a signal in the stacked spectrum, we flag its edges by visual inspection, and measure it by integrating the spectrum inside the flagged region (see dotted lines in Figure 3.2). We expect the signal to be symmetric around zero velocity; even if in some cases the SDSS redshift is slightly off-set with respect to the HI emission, the discrepancy will be random and will cancel out when co-adding multiple spectra (SDSS redshift uncertainty is 0.0002, which corresponds to $\sim 60\,\mathrm{km\,s^{-1}}$). This consideration may in addition help the user to identify the edges of the signal in case of marginal detection: they should be symmetric around zero. Finally we evaluate a signal-to-noise ratio following Saintonge (2007). The S/N is defined as the ratio between the mean flux across the emission line and the *rms* if the spectrum was smoothed to half the velocity width of the signal:

$$\mathrm{S/N} = \frac{\mathrm{F/W}}{rms}\left(\frac{\mathrm{W}/2}{10\,\mathrm{km\,s^{-1}}}\right)^{1/2}. \tag{3.3}$$

F is the integrated flux, 10 km $\mathrm{s^{-1}}$ corresponds to the spectral resolution of the ALFALFA data, and W is the width of the signal in km $\mathrm{s^{-1}}$. If the width is larger than 400 km $\mathrm{s^{-1}}$, we assume W= 400 km $\mathrm{s^{-1}}$ anyway, since for larger values the typical baseline fluctuations start to be of the same width as the galaxy profiles.

We define the spectrum an HI detection if the S/N is greater than 6.5. If there is no detection we evaluate an upper limit, assuming a 5 σ signal with a width of 300 $\mathrm{km\,s^{-1}}$, smoothing the spectrum to 150 $\mathrm{km\,s^{-1}}$. (The width of 300 $\mathrm{km\,s^{-1}}$ is chosen because it corresponds to the peak of the distribution of velocity widths for the detected galaxies in our sample.)

3.2.3 Evaluating HI gas fractions

Our aim in stacking spectra is to characterise the average HI content of a given sample of galaxies. This requires converting our recovered signal into an average HI mass and then into an HI gas fraction. For a single object, these two quantities are well defined. The HI mass can be estimated using the formula (Roberts 1963):

$$\frac{\mathrm{M_{HI}}}{\mathrm{M_\odot}} = \frac{2.356 \times 10^5}{1+z} \left(\frac{D_L(z)}{\mathrm{Mpc}}\right)^2 \left(\frac{S_{int}}{\mathrm{Jy\ km\ s^{-1}}}\right), \tag{3.4}$$

where S_{int} is the integrated HI line flux, z is the redshift and $D_L(z)$ is the luminosity distance. A correction for HI self-absorption is not applied to the HI mass, as it is likely to be negligible (Haynes & Giovanelli 1984, Appendix B). The HI gas fraction is simply defined as $\mathrm{M_{HI}/M_\star}$.

For N individual detections (each with a flux/gas mass measurement), the *average* value of their gas fractions is obtained as the weighted mean value:

$$\langle \frac{M_{HI}}{M_*} \rangle \quad = \quad \left(\Sigma_{i=0}^{N} \frac{M_{HI;i}}{M_{*;i}} \cdot w_i \right) / \left(\Sigma_{i=0}^{N} w_i \right) \tag{3.5}$$

where the w_i are defined as in equation 3.2 so that good quality spectra contribute more to the mean. If we stacked fluxes, we would measure the mean flux of the ensemble, which we could convert into a gas fraction assuming mean values for both the redshifts and the stellar masses. Our targets span a significant relative range in redshift ($z_{max} = 2 \cdot z_{min}$), so given the same HI mass an object at the lowest redshift limit contributes 4 times more signal than one at the upper redshift limit. Because we are stacking mainly non-detected spectra, we do not know how much each galaxy contributes to the total signal. Moreover, both the mean and median values of redshift and stellar mass may not be representative of the underlying population. In order to weight each spectrum in a consistent manner, we choose to stack "gas-fraction" spectra S'_i, where the signal that we co-add is no longer the flux $S\,[\mathrm{Jy}]$, but a gas fraction itself. To achieve this, we weight the spectrum S_i of each galaxy by its distance and stellar mass, so that the units become:

$$S'_i\,[\mathrm{Jy\,Mpc^2\,M_\odot^{-1}}] \;=\; \frac{S_i \cdot D_L^2(z_i)}{M_{*;i}} \tag{3.6}$$

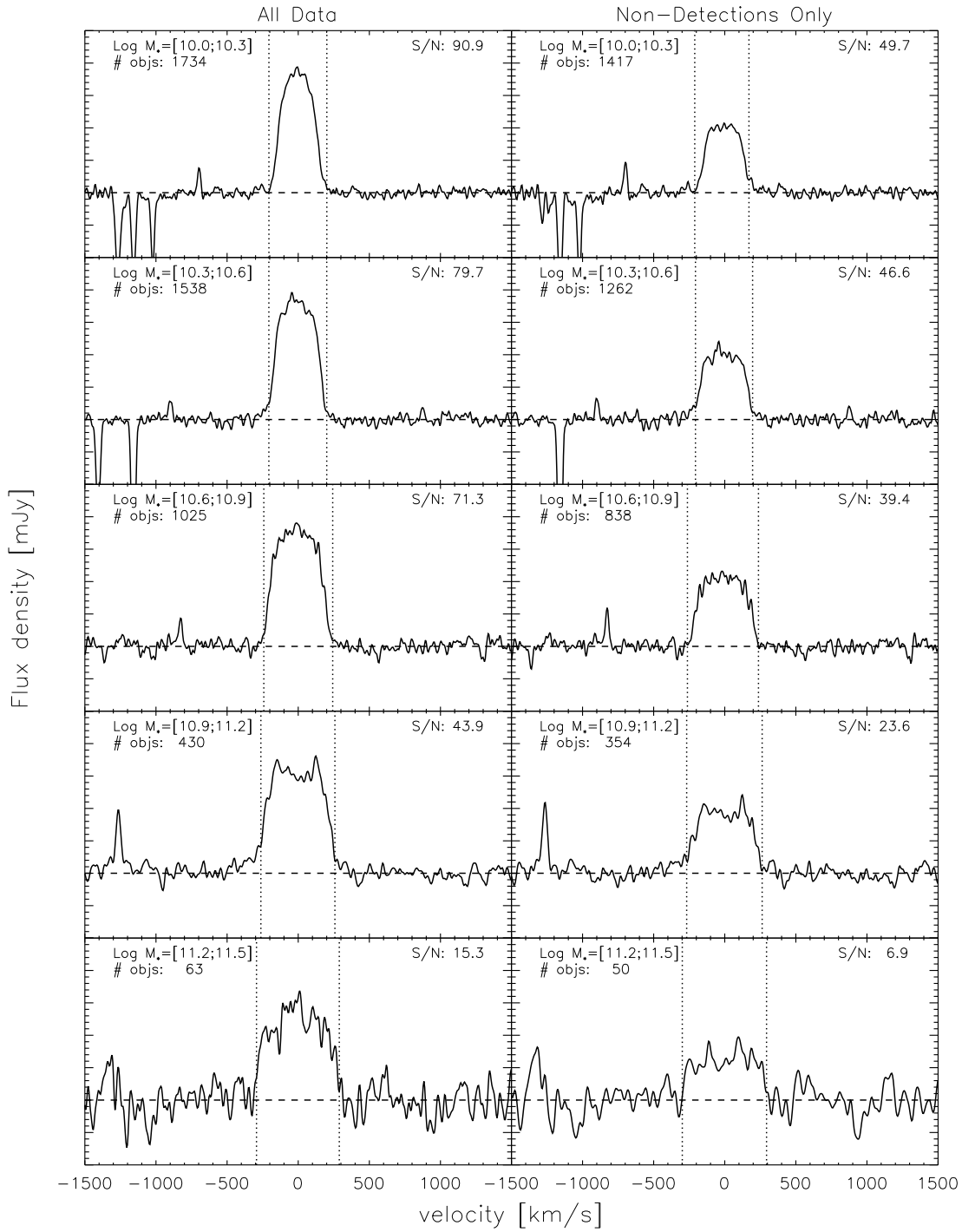


Figure 3.2: Examples of stacked spectra for 5 stellar mass bins. The x-axis is velocity in $\mathrm{km\,s^{-1}}$, the y-axis is HI flux density in mJy. Since we shifted each object to its rest frequency, the signal is centered at zero velocity. For each spectrum, the mass range and the number of stacked objects are reported, as well as its S/N ratio. Dotted lines show the boundaries of the signal, inside which we integrate the flux. Left column: stacked spectra using all galaxies. Right column: stacked spectra using only those galaxies that were not detected by ALFALFA. As expected, the signal is systematically lower for the right column.

The stacking is then performed according to equation 3.1, where we co-add the new "gas-fraction" spectra weighted by their noise. The resulting signal is the average HI gas fraction of the individual galaxies used to build the stack. In Appendix A, we discuss an alternative method to evaluate average gas fractions from stacked spectra, and present comparisons between the two.

3.2.4 Errors

When we consider the average gas fractions measured from stacking, we should take into account different possible sources of error. *i)* In the gas fraction formula, redshifts and stellar masses are used. Uncertainties on the stellar masses are less then 30%, and uncertainties on the SDSS spectroscopic redshifts are considered negligible (see Section 2.4). *ii)* The stacked signal is obtained by integrating the spectrum over a region flagged by the user. Errors on these measurements are significant when the signal-to-noise is low and the signal edges uncertain. We consider detections only stacked spectra for which the S/N is greater than 6.5, which results in errors on the integrated signals smaller than 20%. *iii)* The dominant source of uncertainty in the stacking procedure is therefore the dispersion on the average gas fractions measured. A limit in the stacking analysis, though, is the lack of information on the underlying distribution of the individual objects studied, since we consider (mainly) non detections. A common procedure used to estimate the error of a parameter (in our case the mean gas fraction) extracted from an ensemble of data is the *bootstrap* technique (Efron 1979). As summarized in Jenkins (2003), the bootstrap consists in drawing a random sample with repetition from the main pool of data, calculate the mean, and repeat the process as many times as possible. Under the assumption that the data are independent, the distributions of the means is indicative of the uncertainty on the one measured for the whole sample. It is also an indicator of how much possible outliers weight in the measure.

We apply the bootstrap to the stacking, with a difference. The standard method would require us to extract samples as large as the original, with repetitions. However we do not want to co-add the same galaxies, since the signal-to-noise would not increase in contrast with what we want to achieve with stacking. We extract 200 sub-samples, each 80% the main sample in size and without repetitions. The errors we assume are the dispersion on the repeated mean measures.

3.3 A first stacking application: HI gas fraction scaling relations for massive galaxies

To prove the reliability of the results that can be obtained with a stacking analysis, we performed some tests using both simulations and real data. These tests are described in detail in Appendix A. Here we present an application of the stacking, used to characterise how the average HI gas fraction depends on a variety of different properties, for our sample of galaxies with stellar masses greater than $10^{10}\,\mathrm{M}_{\odot}$.

Galaxy scaling relations are an important tool from both an observational point of view, because they provide statistic descriptions of galaxy properties, and from a theoretical one, as they constitute boundary conditions for models and simulations. In a recent work, Catinella et al. (2010) studied HI gas fraction scaling relations as a function of stellar mass, stellar mass density ($\mu_{\star}$), colour and concentration index (C) for a representative sample of $\sim$ 200 galaxies with $M_{*} > 10^{10} M_{\odot}$ from the GASS survey. Here we study the same trends for our *sample A* galaxies, and compare the results with those of Catinella et al. (2010). In the GASS survey, HI observations reach a much deeper flux limit compared to ALFALFA, so average HI gas fractions can be derived using individual HI detections rather than by stacking. It is important to check whether our stacking method gives answers that are consistent with these published measurements.

In Figure 3.3 we present our results for *sample A*. In each panel our estimates of the HI gas fraction derived from the stacked spectra are plotted as red circles; at the bottom of each panel, we record the number of galaxies included in each stack, as well as the percentage of galaxies with ALFALFA detections (inside the parenthesis). Errors are evaluated using the bootstrap technique as explained in Section 3.2.4. In cases where the stack includes more than a few hundred galaxies the errors on the mean HI fraction are negligible, as the bootstrap is a measure of the "reliability" of the estimated parameter. The green solid lines show the mean values of $\mathrm{M_{HI}/M_{\star}}$ measured by Catinella et al. (2010, see Figure 9 in their paper), where the non-detections are set to their upper limits. The dashed lines indicate the 1σ confidence interval estimated with a bootstrap method. Finally, the gray dots show the galaxies in *sample A* that are detected by ALFALFA; as expected, ALFALFA detections are offset toward the gas rich range. We will not show all the stacked spectra that yield our measurements; however, the spectra reported in Figure 3.2 (left column) are the ones from which we evaluated the points in Figure 3.3 (a). We note that we are able to measure an HI gas fraction in every bin, even in the very reddest NUV$-r$ colour bins where

the ALFALFA detection rates are close to zero. The fact that the errors remain relatively small for these bins indicates that the stacked spectra are not dominated by signal coming from a small fraction of the galaxies. If that was the case, in fact, bootstrapped repetitions including the outliers would have increased the errors.

We find that the HI gas fraction is a decreasing function of stellar mass, stellar mass surface density, concentration index and colour. The correlations between gas fraction and both $M_\star$ and concentration index (Figure 3.3, first row) are weak: the gas fraction values span ~ 1 dex across the whole range of the two parameters. Panel (b) seems to suggest a transition in the gas content of galaxies that are bulge dominated ($C \gtrsim 2.5$), but we will investigate this further in the next chapter. The average value of $M_{HI}/M_\star$ is instead strongly correlated with NUV$-r$ colour and stellar mass surface density, properties for which the gas fraction drops by up to a factor of 25 over the x range. We report the fits to these relations in Table 3.2 and the individual average points in Table 3.1, but here we just list the total change in $\mathrm{Log}\, M_{HI}/M_\star$ for each relationship, which allows the comparison among the different properties:

$$\begin{aligned} M_* &: \quad \Delta \log M_{\rm HI}/M_* = 0.20 \\ C &: \quad \Delta \log M_{\rm HI}/M_* = 0.24 \\ \mu_* &: \quad \Delta \log M_{\rm HI}/M_* = 0.45 \\ \mathrm{NUV}-r &: \quad \Delta \log M_{\rm HI}/M_* = 0.52 \end{aligned}$$

Our average HI gas fractions derived from the stacked spectra are in excellent agreement with the results reported in Catinella et al. (2010). The point-by-point agreement is generally within 1σ, with the possible exception of the lowest stellar mass and stellar mass surface density bins, where we find gas fractions which are systematically lower than those reported in Catinella et al. (2010) paper. We caution that low mass galaxies were somewhat under-represented in the first GASS data release. On the contrary, our error bars only indicate statistical errors and do not account for other effects, such as e.g. cosmic variance of high mass galaxies, so we do not think that the discrepancy is significant.

Before proceeding with the analysis, we point out once again that, although stacking recovers the average trends of the population extremely well, it does not provide any information on the underlying scatter. This can only be studied with individual gas fraction measurements.

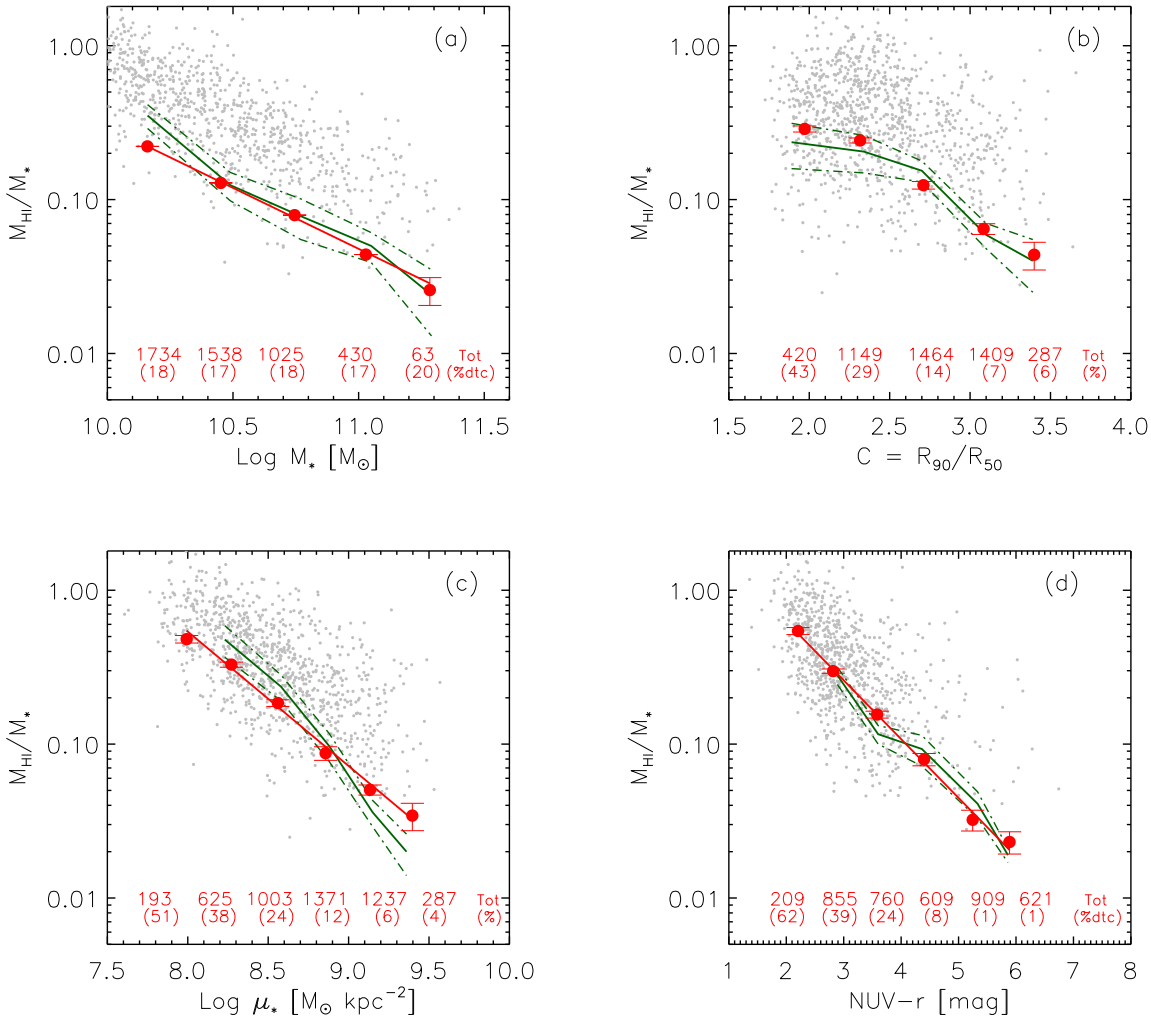


Figure 3.3: Average gas fractions of *sample A* galaxies derived from stacking (red circles) are plotted as a function of stellar mass $M_\star$ (a), concentration index C (b), stellar mass density $\mu_\star$ (c) and NUV$-r$ colour (d). The points are plotted at the mean value of each bin. The red lines show a linear fit to the red points, which are compared with the average HI gas fractions from Catinella et al. (2010) (green lines; green dashed lines show the 1σ uncertainties). Gray dots represent galaxies in *sample A* with ALFALFA detections. The numbers at the bottom of the panels indicate the total numbers of objects co-added in each bin, and the percentage of them directly detected by ALFALFA (inside the parenthesis). The values plotted are reported in Table 3.1 and the fits in Table 3.2.

x	$\langle x \rangle$	$\langle \mathrm{M_{HI}/M_\star} \rangle$	N		x	$\langle x \rangle$	$\langle \mathrm{M_{HI}/M_\star} \rangle$	N
Log $\mathrm{M_\star}$	10.16	0.222±0.008	1734		R_{90}/R_{50}	1.97	0.288±0.013	420
	10.45	0.128±0.005	1538			2.32	0.242±0.008	1149
	10.74	0.079±0.004	1025			2.71	0.124±0.006	1464
	11.03	0.044±0.004	430			3.08	0.064±0.004	1409
	11.28	0.026±0.005	63			3.40	0.044±0.007	287
Log $\mu_\star$	7.99	0.481±0.025	193		NUV$-r$	2.20	0.543±0.024	209
	8.27	0.328±0.011	625			2.82	0.298±0.009	855
	8.56	0.185±0.008	1003			3.58	0.156±0.007	760
	8.86	0.087±0.008	1371			4.40	0.079±0.006	609
	9.13	0.050±0.003	1237			5.24	0.032±0.004	909
	9.40	0.034±0.006	287			5.89	0.023±0.004	621

Table 3.1: Average HI scaling relations for *sample A*, as in Figure 3.3.

$$\mathrm{Log}\langle \mathrm{M_{HI}/M_\star} \rangle = (\text{-}0.79\pm0.03)\times\mathrm{Log\,M_\star}+(7.39\pm0.33)$$
$$\mathrm{Log}\langle \mathrm{M_{HI}/M_\star} \rangle = (\text{-}0.87\pm0.03)\times\mathrm{Log}\,\mu_\star+(6.71\pm0.22)$$
$$\mathrm{Log}\langle \mathrm{M_{HI}/M_\star} \rangle = (\text{-}0.38\pm0.01)\times(\mathrm{NUV}-r)+(0.56\pm0.04)$$

Table 3.2: Linear fits to the HI scaling relations in Figure 3.3.

3.4 Conclusions

Stacking has proven to be a powerful technique to exploit surveys that have a large amount of non-detected data. In the HI case, the on-going ALFALFA survey is the current best data-set to which stacking could be applied, as it fulfills both the requirements of having good quality data and a large sample.

We have developed a stacking technique for ALFALFA survey, which allows us to measure average gas fractions of samples of galaxies that are individually non detected. We have presented the tool step-by-step, and showed a first application of it. We measured average HI scaling relations for massive galaxies and compared our results with the ones of Catinella et al. (2010), obtained for a different sample with individual detections available, proving that both analyses lead to compatible results.

In the rest of this thesis we use the stacking to address some open issues in the cold gas properties of massive galaxies, but of course the range of possible science cases that the stacking can address is much broader. The natural extension of this tool would be applying the stacking to the higher redshift data that in the next decade will be acquired by the new generation HI surveys programmed with the Square-Kilometer-Array precursor telescopes.

4

Morphological quenching

Based on
S. Fabello, et al.
MNRAS, 2011, 411, 993

4.1 Introduction

The cold gas content of a galaxy is known to vary strongly with colour and star formation rate. The connection between gas content and galaxy morphological type remains on the contrary unclear. Whereas star-forming spiral galaxies almost always contain atomic hydrogen, the HI content of early-type galaxies is considerably more difficult to predict. Some Es and S0s have neutral atomic hydrogen gas content similar to those of Sb-Sc type spirals, while others contain several orders of magnitude less HI (Roberts & Haynes 1994). It has been speculated that these variations may be an indication of the external origin of the HI in ellipticals (Knapp et al. 1985), but direct proof of this conjecture is still lacking.

Because the HI is on average more difficult to detect in early-type galaxies, the samples discussed in the literature have generally been quite small. Some of the largest systematic studies of gas in early-type galaxies were carried out in the 1980's. Knapp et al. (1985) analysed a sample of 152 nearby elliptical galaxies, of which 23 were detected in the HI line. These authors studied the distribution of the quantity M_{HI}/L_B for these systems. In contrast to spiral galaxies, where the distribution $N(M_{HI}/L_B)$ has a well-defined mean value and a small dispersion, M_{HI}/L_B spans a wide range in ellipticals. Wardle & Knapp (1986) extended this work to S0 galaxies and found that the HI detection rate was twice as high

compared to ellipticals. Some years later, Bregman et al. (1992) carried out a study of the interstellar components of 467 early-type galaxies in the Revised Shapley Ames Catalogue and again reported a trend of increasing neutral gas content from E to Sa. These authors suggested that the cold gas in early-types is associated mainly with disks and not with the bulge components of these galaxies. However, more recent studies that have mapped HI in nearby early-type galaxies (Morganti et al. 2006) have concluded that the HI can be organized in a variety of different configurations, e.g. in regular disks, in clouds, in rings, or even in tidal tail-like structures. Such a variety of structural distributions is likely to be connected with the environment in which the ETG are found. Preliminary observations of ETG (E+S0) galaxies from the ATLAS 3D survey (Cappellari et al. 2011) have shown that the HI distribution of early-types in the Virgo cluster is strongly irregular and becomes more settled, approaching a rotating disk, when objects are located in less and less dense groups (Serra et al. 2010). When completed, the ATLAS 3D sample of 170 ETG mapped in the HI line will be the first complete large sample of early-types with resolved HI data. Similar results on the dependence of early-types' HI content on environment have also been found by Oosterloo et al. (2010) on a sample of 33 nearby objects.

One major problem that has plagued our understanding of cold gas in early-type galaxies is that the available HI data have been inhomogeneous. As already extensively discussed, large blind HI surveys such as ALFALFA offer uniform coverage over large regions of the sky and allow one to construct complete, unbiased samples of HI-selected galaxies. However, these surveys are shallow and do not in general detect gas-poor early-type galaxies. This limitation has been pointed out in recent works by di Serego Alighieri et al. (2007) and Grossi et al. (2009), which used ALFALFA data to study an unbiased sample of early-type galaxies in the Virgo cluster region. They were able to compare the HI content of early-type galaxies drawn from field and group environments, but their average detection rates were much smaller than earlier studies based on incomplete and inhomogeneous data.

The Hubble morphological classification scheme is based on the optical appearance of a galaxy which in turn depends both on structural properties such as its bulge-to-disk ratio, and on star formation rate. If we wish to understand the physical processes that regulate the neutral gas content of galaxies, it is preferable to analyse the effects of star formation and galaxy structure *separately*. Recently, Helmboldt (2007) studied 30 E and S0 galaxies with signs of recent or ongoing star formation and concluded that such systems are more gas rich than E/S0 galaxies with old stellar populations. The availability of sizes, surface brightnesses and parameters measuring the concentration of the light for samples of millions of galaxies (made possible due to recent advances in large scale CCD surveys

and automatic image processing techniques) enabled a new approach in understanding the interplay between stars and gas in early-type systems.

Here, we use ALFALFA survey data to constrain the average HI gas fractions of a large, unbiased sample of massive early-type galaxies. We present the sample and the selection criteria adopted to extract it in Section 4.2. We study how the HI content depends on parameters such as stellar mass, stellar mass surface density, concentration index, central velocity dispersion, and on UV/optical colour in § 4.3; there we also compare early-types' gas content with the global scaling relations obtained without applying any morphological cut. Finally, we test a specific proposed model of morphological quenching in § 4.4.

4.2 Bulge-Dominated sample selection

We have chosen to define "early-type" galaxies purely in terms of their structural properties, without regard to their stellar populations or star formation rates. We note that our definition is in contrast to others in the literature, which have excluded galaxies with emission lines (e.g., Bernardi et al. 2003; Graves et al. 2009a). Our goal is to explore the extent to which the presence or absence of a significant bulge component influences the HI content of a galaxy, so we do not wish to bias our conclusions by selecting against bulge-dominated galaxies with emission lines.

Starting from *sample A* (see Section 2.1), we extracted a subset of early-type, bulge-dominated objects (hereafter *B-D sample*) with the following properties:

- concentration index $C{=}R_{90}/R_{50} \geq 2.6$;
- the likelihood that the light profile is fitted by a de Vaucouleurs model is greater than it is by an exponential one (*lnldev_r* and *lnlexp_r* parameters from SDSS);
- inclination less than 70° (this cut rejects the most inclined systems).

We remind that the C parameter has been shown to be an excellent indicator of the bulge-to-total ratio (B/T) derived from bulge/bar/disk decomposition analysis on 2-dimensional SDSS images (Gadotti 2009). A cut at $C \geq 2.6$ restricts the sample to galaxies with $B/T \geq 0.3$ (see Fig.1 of Weinmann et al. 2009). For this analysis we choose our default cut at $C \geq 2.6$ so to obtain a statistically significant sample, but we also experiment with more stringent cuts at larger values of C.

Application of our default cut leads to a final *B-D sample* consisting of 1833 objects, whose properties are shown in Figure 4.1. The solid histograms represent the entire sample; the red dashed histograms the sub-sample detected by ALFALFA. The stellar mass,

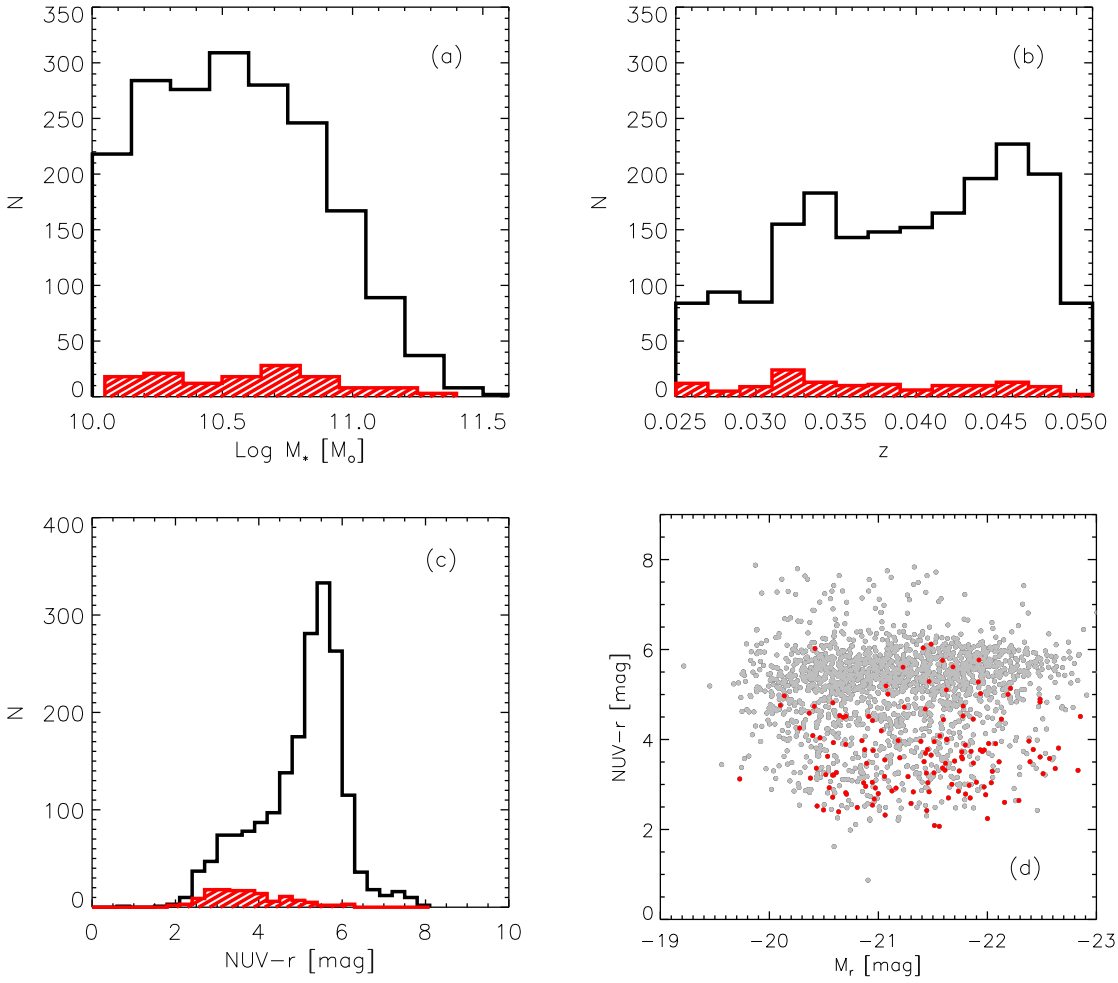


Figure 4.1: Distributions of (a) stellar mass, (b) redshift and (c) NUV-r colour (corrected for Galactic extinction only) for galaxies in the *B-D sample* (for a comparison with the whole *sample A* see Figure 2.2). The black solid histograms represent the whole *B-D sample*, while the red, dashed histograms show the distributions for the sub-sample of galaxies with ALFALFA detections. Panel (d) shows the colour-magnitude diagram of *B-D sample* galaxies (gray dots) and the sub-sample with ALFALFA detections (red dots).

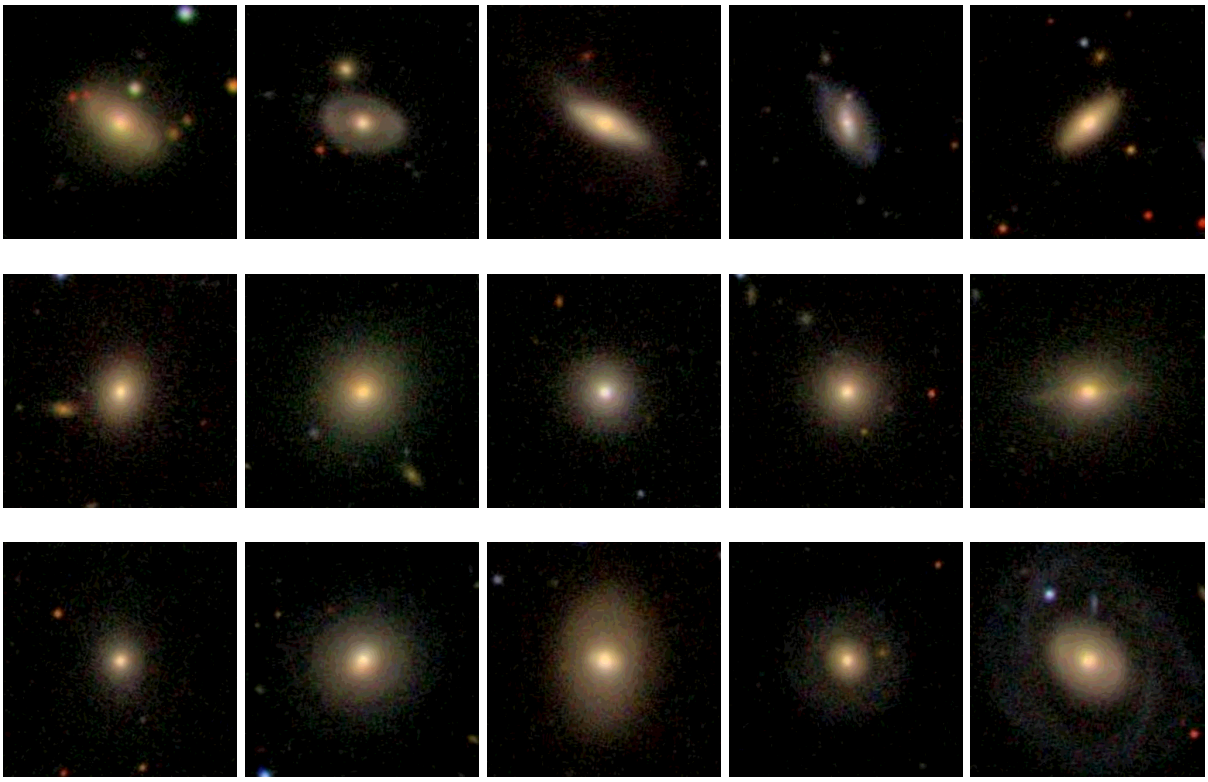


Figure 4.2: SDSS postage images of galaxies randomly selected from the *B-D sample* and ordered by increasing C. From left to right, top to bottom C increases from 2.6 to 3.8. The images are 1 arcminute square in size.

redshift and the NUV-r colour distributions are shown in panels (a), (b) and (c), respectively. A pure morphological cut results in selecting against blue, gas rich galaxies. The average ALFALFA detection rate decreases significantly, so that now only $\sim 9\%$ of the bulge-dominated galaxies are detected (unlike the $\sim$23% of *sample A*). Finally, the colour-magnitude diagram is shown in panel (d). Most of the early-type targets lie on a well-defined red sequence, but we note that some objects do scatter bluewards of this locus. Such objects may be star-forming, transitional or Seyfert galaxies (Schawinski et al. 2007). In addition, by selecting targets based on concentration index and inclination we may also include galaxies with some disk component, as clear by visual inspection of our targets. In Figure 4.2 we present 1×1 arcminute SDSS postage stamps of a randomly selected set of galaxies from our *B-D sample*, which have been ordered by increasing values of C.

4.3 HI study of a complete sample of early-type galaxies

In this section, we ask whether bulge-dominated galaxies with $M_* > 10^{10} M_\odot$ lie on the same HI scaling relations as the general population of galaxies with similar stellar masses. We have shown in the previous chapter (§ 3.3) the global HI scaling relations for all the targets, underlying how they are important for describing galaxy properties and constraining evolutionary models. Our goal is now to explicitly determine whether the presence of the bulge plays a role in regulating the rate at which gas is coverted into stars, for example by stabilizing the HI disk and preventing its collapse (e.g., Martig et al. 2009). In order to test the role of the bulge in a clean way, we must account for the fact that the physical properties of galaxies are strongly correlated. If one selects a sub-sample of bulge-dominated galaxies from the parent *sample A*, one will automatically select a sample of galaxies with higher stellar masses, higher stellar mass surface densities, and redder colours. It is therefore important to understand whether or not bulge-dominated galaxies differ in HI content from the parent sample at *fixed* values of these parameters.

Our results are presented in Figure 4.3, where we show the HI scaling relations as a function of colour and structural parameters. In all the plots, blue circles are the average gas fractions obtained from stacking of *B-D sample* targets. The red lines are the fits to the mean HI gas fraction relations obtained for *sample A*; in the case of σ, not considered in the previous chapter, we report the actual data points. Errors are evaluated using the *bootstrap method* (§ 3.2.4). Finally, triangles indicate the upper limits evaluated in the case of a stack that yields a non-detection.

The average gas content of bulge-dominated objects decreases weakly with increasing values of $M_\star$ and central velocity dispersions, and more strongly with increasing $\mu_\star$ and colours. Our main result is that the average HI gas fractions of B-D galaxies are significantly lower (by approximately a factor of 2) than those of the parent sample at a given value of stellar mass. A similar, but weaker reduction in the average HI gas fraction is seen for the *B-D sample* when it is plotted as a function of the central velocity dispersion (corrected for aperture effects, see § 2.4) of the bulge. However, the relation between gas fraction and stellar mass surface density and NUV$-r$ colour appears to be insensitive to the morphological cut[1]. In their paper, Catinella et al. (2010) showed that a linear combination of stellar mass surface density and NUV$-r$ colour provided an excellent way to "predict" the HI content of star-forming galaxies more massive than $10^{10} M_\odot$. Here we show that this conclusion holds true independent of the bulge-to-disk ratio of the galaxy.

[1]For completeness, we report all the values of the points in the plots, for the different cuts and as a function of the different parameters, in Table 4.1.

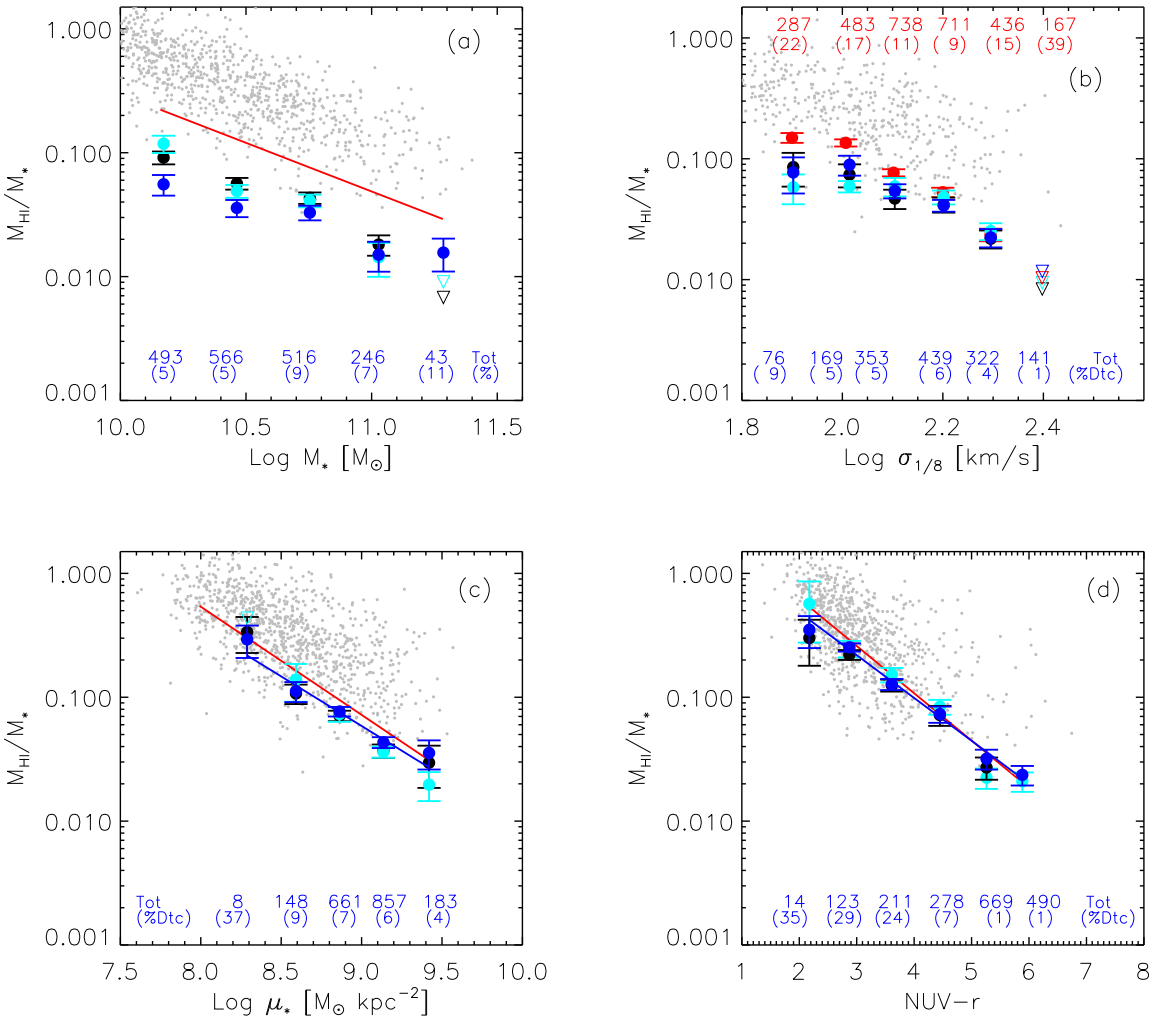


Figure 4.3: The dependence of the average HI gas fraction on stellar mass $M_\star$ (a), central velocity dispersion σ (b), stellar mass surface density $\mu_\star$ (c) and NUV$-r$ colour (d) for the *B-D sample* (blue symbols). The relations found for *sample A* are shown in red for comparison (the fit for the $M_\star$ relation, the actual points for σ). Upside-down triangles indicate upper limits in the case of a non-detection. The numbers written in the panels indicate the numbers of objects co-added in each bin (Tot), and the percentage of them directly detected by ALFALFA inside the parenthesis (%Det). Gray dots show *sample A* galaxies with ALFALFA detections. We have also applied more stringent cuts to the *B-D sample*, as explained in the text: cyan circles represent a sample with C>3 (52% of the original ETG sample); black circles are for a sample with b/a >0.6 (or inclination lower than 55° - 68% of the original sample). The values plotted for each *ETG sample* are reported in Table 4.1

We have tested that our result is robust if we define the *B-D sample* using more stringent cuts on concentration index ($C > 3$; cyan dots in Figure 4.3) or on the axis ratio of the galaxy ($b/a > 0.6$, which implies inclination lower than $\sim$55°; black circles). These cuts reduce the early-type galaxies by 50% and 30% respectively. Nevertheless, results shown in Figure 4.3 demonstrate that the average HI gas fractions of these systems still lie on the same relations when plotted as a function of stellar mass surface density and NUV-r colour.

Last, we measure how the average gas fractions of galaxies in the *B-D sample* vary as a function of position in the two-dimensional plane of colour versus stellar mass density μ_*, since these are the two main parameters driving the gas content in massive galaxies. In Figure 4.4 we plot the B-D targets on the μ_*-(NUV$-r$) plane. Bulge-dominated galaxies are mainly found on the red sequence, but there is a minority population with bluer colours. We adaptively bin the sample in two dimensions by recursively dividing the plane into axis-aligned rectangles. We stop dividing a region when a further split would lower the S/N below the detection threshold of 6.5. Figure 4.4 (bottom panel) shows the final binning used. In each bin, the measured gas fraction (expressed as a percentage of the stellar mass) is reported. In Figure 4.4 (top panel) we colour-code the (NUV$-r$)$-\mu_*$ plane according to gas fraction. The HI content decreases going from left to right (towards increasing stellar mass surface density) and from bottom to top (towards redder colours). The most significant variation is clearly along the colour direction.

4.4 A test of the morphological quenching scenario

The idea that galaxy disks are more resistant to the formation of bars, spiral density waves and other instabilities if they are embedded within a dynamically hot halo or bulge, has its origins in early work by Ostriker & Peebles (1973). Recently, Martig et al. (2009) have proposed this so-called "morphological quenching" mechanism as a way of explaining why present-day bulge-dominated galaxies on the red sequence cease growing in stellar mass in environments where they continue to accrete gas. In their picture, a disk with similar gas content will be much less efficient at forming stars if it is embedded in a galaxy with a significant bulge component. As stated in the abstract of their paper, "our mechanism automatically links the colour of the galaxy with its morphology, and does not require gas consumption, removal or termination of the gas supply."

To test whether the "morphological quenching" process is truly important in maintaining the low observed rates of star formation in red sequence galaxies, we have performed the

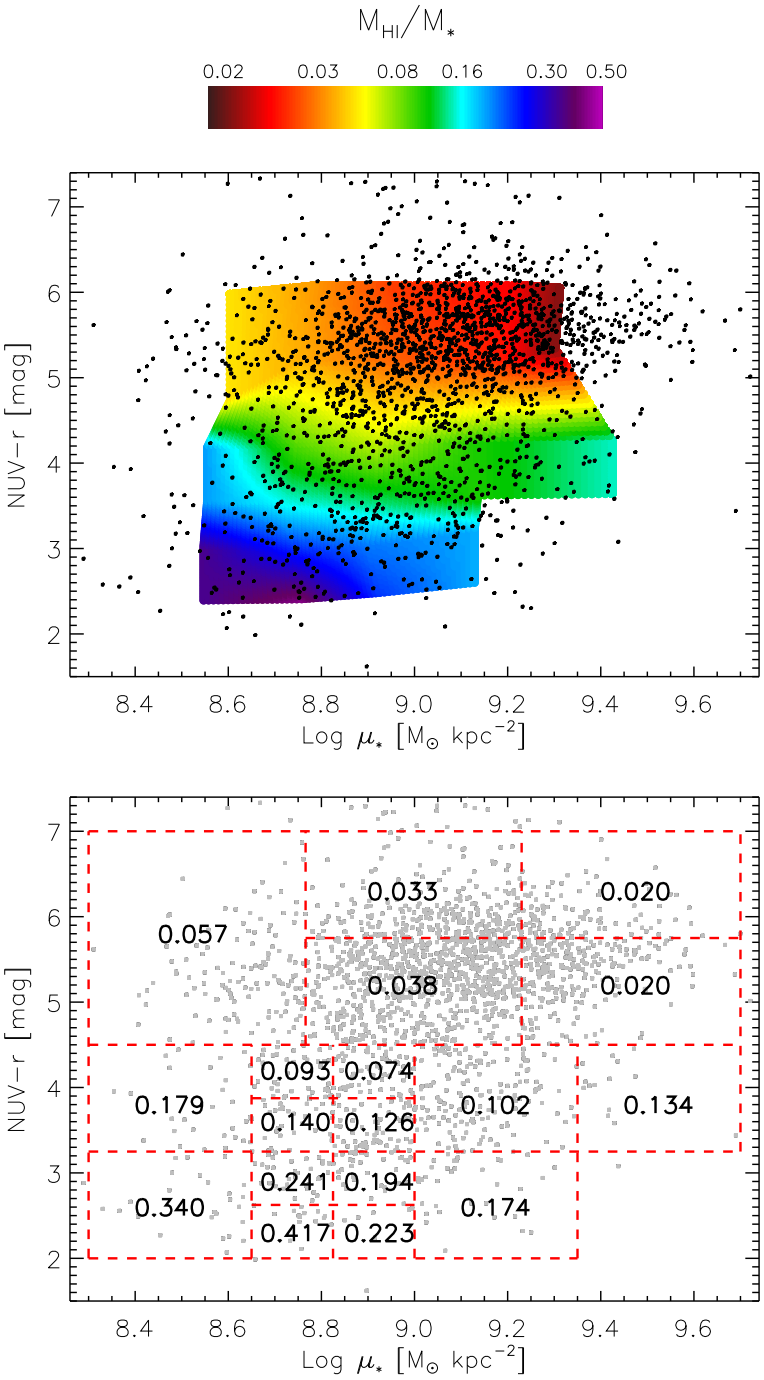


Figure 4.4: Average HI gas fraction dependence in the 2-dimensional plane of stellar mass surface density $\mu_\star$ and NUV$-r$ colour for galaxies in the *B-D sample*. In the top panel the dots show individual objects, while the colours show the (interpolated) gas fractions measured with the stacking as a function of position in the plane (the colour scale key is included above the plot). The bottom panel shows the adopted binning. For each bin, the gas fraction measured from the stacked spectrum is noted.

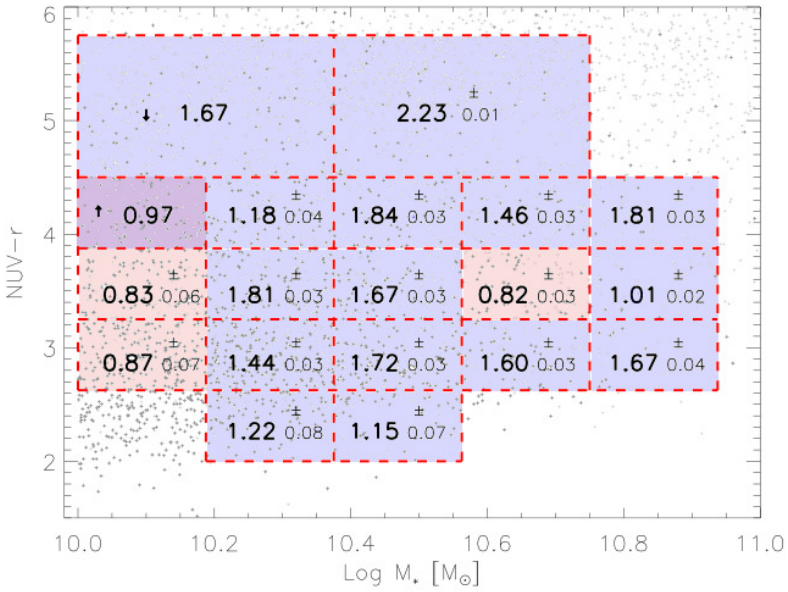


Figure 4.5: Morphological quenching test. For each bin of NUV$-r$ and $M_\star$ the ratio between the gas fractions of the disk-dominated galaxies with $C < 2.6$ f(HI)$_{DD}$ and the bulge-dominated galaxies with $C > 2.6$ f(HI)$_{BD}$ is reported. The panel shows the adopted binning with the values measured for each bin. The arrows indicate upper/lower limit if one of the stacked spectra yielded a non-detection. Gray dots show how the targets (LTGs+ETGs) distribute along the plane. For clarity, we have coloured in blue the bins in which that is true, and in red the bins in which bulge-dominated ones are gas richer.

following experiment. We have binned galaxies with $C < 2.6$ and $C > 2.6$ in the two-dimensional plane of NUV-r colour versus stellar mass (note that we use the same bin boundaries for both samples). We stack the HI spectra of the galaxies in each bin and calculate the average HI gas fraction. In Figure 4.5, we report for each bin the ratio between the gas fraction of the disk-dominated (DD) objects and the bulge-dominated (BD) ones, i.e.:

$$r = \left(\frac{M_{HI}}{M_*}\right)_{DD} / \left(\frac{M_{HI}}{M_*}\right)_{BD}$$

If the morphological quenching scenario is correct, at fixed stellar mass we would expect to find higher average HI gas fractions for bulge-dominated galaxies on the red sequence than for disk-dominated galaxies with similar colours.

The results in Figure 4.5 show that in general the *opposite* is true. Gas fractions are (almost) always higher for the disk-dominated galaxies than for bulge-dominated ones. For clarity, we have coloured in blue the bins for which this holds, and in red the bins in which

bulge-dominated galaxies are gas richer. We note that in the three red bins (the violet one is actually a lower limit for which the ratio is compatible with 1) the difference in the gas content is anyway small: disk-dominated galaxies contain at least 80% of the gas found in B-D with similar properties. The gas fraction differences do appear to be largest for red sequence galaxies with NUV-$r > 5$, but the sign of the difference contradicts the predictions of Martig et al. (2009).

It is important to check that our result is not simply due to extinction effects. Some of the reddest, gas-rich disk-dominated objects may actually be heavily obscured systems which will move blueward when dust corrections are applied. Following Schiminovich et al. (2010), we have applied dust corrections to the NUV$-r$ colours of the galaxies in our sample with $D_n(4000)<1.7$ (see § 2.4). In order to make up for the loss of red late-type galaxies, we applied a stronger cut in concentration index ($C>3$) to define the early-type sample. The ratio of gas fraction between disk-dominated and bulge-dominated objects decreases, but we still find that the HI gas fraction of the disk-dominated galaxies *never* falls below that of the bulge-dominated ones. In particular, if we divide the red sequence objects with $4.5 < \mathrm{NUV}-r < 7$ in two bins of stellar mass, we find a value of r=1.03 for the less massive galaxies, and r=1.71 for the more massive ones.

We note that our results are in agreement with those presented in Schiminovich et al. (2010). In their paper, Schiminovich et al. use a volume-limited sample of 200 galaxies from the GASS survey to explore the global scaling relations associated with the ratio $\mathrm{SFR}/\mathrm{M_{HI}}$, which they call the *HI-based star formation efficiency*. They found that the average value of this star formation efficiency has little variation with any galaxy parameter, including the concentration index.

4.5 Conclusions

We have extracted a sub-sample of 1833 "early-type" galaxies from our *sample A*, selected only by morphology to have inclinations less than 70°, concentration indices $C>2.6$, and with light profiles which are well fitted by a De Vaucouleurs model. We have stacked their spectra in bins of stellar mass, stellar surface mass density, central velocity dispersion and NUV$-r$ colour, and we have estimated the average HI gas fractions of the galaxies in each bin.

Our main result is that the HI content of both early-type and late-type galaxies correlate primarily with NUV$-r$ colour and stellar mass surface density (see also § 3.3). The relation between average HI gas fraction and these two parameters is independent of C, and hence

of the bulge-to-disk ratio of the galaxy. At fixed stellar mass, early-type galaxies do have lower average HI fractions than late-type galaxies, but this effect does not arise as a direct consequence of the presence of the bulge and we discuss possible implications below.

We have also tested whether the HI gas content of bulge-dominated galaxies differs from that of late-types at fixed values of NUV$-r$ and $\mu_\star$. We find no evidence that red-sequence galaxies with a significant bulge component are less efficient at turning their available gas reservoirs into stars. This result seems in contradiction with the "morphological quenching" scenario proposed by Martig et al. (2009).

We can now discuss in more detail possible implications of this work.

1. The HI content of a galaxy is independent of its bulge-to-disk ratio. This can be understood if the following two conditions are satisfied: 1) The HI gas in early-type galaxies is always associated with disks, or with material that is in the process of settling into disks. 2) The formation of the galactic disk is decoupled from the formation of the bulge. One way that condition (2) could be satisfied is if the disk is formed from gas that accretes *well after the bulge formation event*, so that the amount and the configuration of the accreted material is not in any way related to properties of the bulge, such as its mass or its velocity dispersion.

At fixed *total stellar mass*, we found that galaxies with larger C have lower gas fractions than galaxies with smaller C (Figure 4.3, panel a). If the gas is associated mainly with the disk, then the ratio $\mathrm{M_{HI}}/M_{*;disk}$ should *not* depend on C. We have fit the Weinmann et al. (2009) relation between B/T and C. Our bi-linear fit result is $C = 1.92 + 2.28 \cdot B/T$[2]. For each galaxy in our sample, we compute $M_{*;disk}$. In Figure 4.6, we plot $\mathrm{M_{HI}}/M_{*;disk}$ as a function of $\log M_*$ for both *B-D sample* and *sample A* galaxies. As can be seen, the difference in gas fraction between the two samples is greatly reduced when the HI mass is divided by the disk stellar mass. For reference, the red and blue lines on the plot show $\mathrm{M_{HI}}/M_{*;total}$, as in Figure 4.3 (panel a).

We note that our result that the HI fraction of early-type galaxies depends on the size of the galaxy, but not on its bulge properties, is opposite to what is found for bulge stellar populations. In a recent paper, Graves et al. (2009b) carried out an analysis of 16,000 nearby quiescent galaxies with 3-arcsecond aperture spectra from the Sloan Digital Sky Survey. Their paper demonstrates convincingly that mean stellar age, [Fe/H], [Mg/H], and [Mg/Fe] scale strongly with the velocity dispersion of the bulge, but there is no dependence

[2]We correct an error in the published Fabello et al. (2011), where we mistyped the constant.

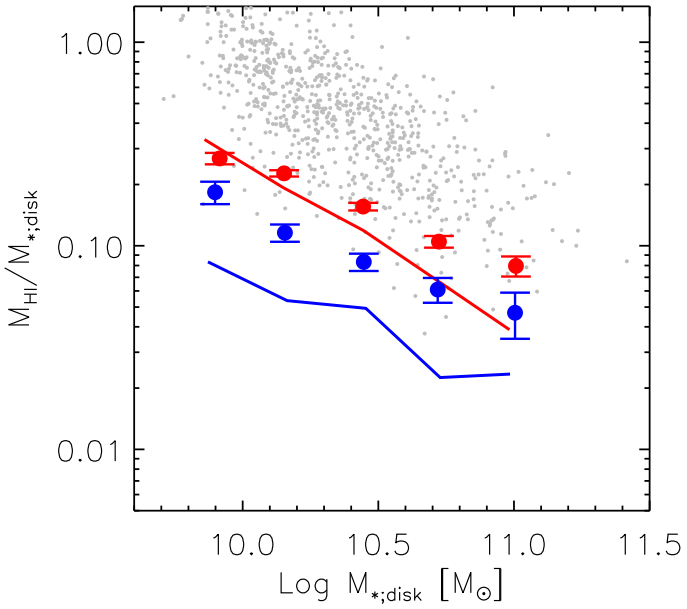


Figure 4.6: Average *disk* gas fractions of *sample A* (red circles) and *B-D sample* galaxies (blue circles) as a function of total stellar mass $M_\star$. The lines show the results obtained measuring the total gas fraction, as in Figure 4.3 (panel a). Gray dots show galaxies with ALFALFA detections from *sample A*.

on R_e at a fixed value of σ. We thus infer that the star formation *history* of the central bulge does not depend on the size of the galaxy. Our own results show that the present-day gas content of early-types does depend on size. This again argues for bulge and disk formation processes that are decoupled.

2. Galaxies with significant bulge component are not less efficient in turning their available gas into stars. One possible explanation of this result is that the rate of gas consumption in galactic disks is primarily regulated by externally-driven rather than internally-generated instabilities. In recent work, Chakrabarti & Blitz (2009) present an analysis of the observed perturbations of the HI disk of the Milky Way and infer the existence of a dark sub-halo which tidally interacted with the Milky Way disk. In addition to dark sub-halos, luminous satellite galaxies are observed to interact with galactic disks. Finally, the dark matter environment at the centers of present day halos is neither static nor in equilibrium. Gao & White (2006) used the Millennium Simulation to study asymmetries in dark matter halo cores. They found that 20 per cent of cluster haloes have density center

separated from barycenter by more than 20 percent of the virial radius, while only 7 percent of Milky Way haloes have such large asymmetries. Because early-type galaxies reside in more massive dark matter halos than late-type galaxies (Mandelbaum et al. 2006), their disks may be subject to considerably larger externally-driven perturbations. All these effects may counteract the stabilizing effect of the bulge.

Further tests of the proposed external origin of the gas in early-type galaxies will come from more detailed analysis of its spatial distribution and kinematics. By studying the relation between gas and stellar angular momentum, one can hope to gain further understanding of how the gas was accreted. Studies of this nature are planned as part of the next generation of integral-field spectroscopy studies of nearby early-type galaxies, such as the ATLAS 3D survey.

x	$\langle x \rangle$	$\langle M_{HI}/M_\star \rangle$ *Sample A*	N	$\langle M_{HI}/M_\star \rangle$ *ETG sample*	N
Log $M_\star$	10.16	0.222±0.008	1734	0.0556±0.009	493
	10.45	0.128±0.005	1538	0.036±0.005	566
	10.74	0.079±0.004	1025	0.033±0.004	516
	11.03	0.044±0.004	430	0.015±0.004	246
	11.28	0.026±0.005	63	0.016±0.004	43
R_{90}/R_{50}	1.97	0.288±0.013	420	-	-
	2.32	0.242±0.008	1149	-	-
	2.71	0.124±0.006	1464	-	-
	3.08	0.064±0.004	1409	-	-
	3.40	0.044±0.007	287	-	-
Log $\mu_\star$	7.99	0.481±0.025	193	-	-
	8.27	0.328±0.011	625	0.293±0.082	8
	8.56	0.185±0.008	1003	0.112±0.018	148
	8.86	0.087±0.008	1371	0.077±0.006	661
	9.13	0.050±0.003	1237	0.043±0.004	857
	9.40	0.034±0.006	287	0.035±0.008	183
NUV$-r$	2.20	0.543±0.024	209	0.351±0.091	14
	2.82	0.298±0.009	855	0.253±0.019	123
	3.58	0.156±0.007	760	0.127±0.011	211
	4.40	0.079±0.006	609	0.073±0.009	278
	5.24	0.032±0.004	909	0.032±0.005	669
	5.89	0.023±0.004	621	0.024±0.004	490
Log σ	1.90	0.149±0.014	287	0.077±0.020	76
	2.01	0.135±0.009	483	0.089±0.014	169
	2.10	0.077±0.005	738	0.054±0.007	353
	2.20	0.053±0.005	711	0.041±0.004	439
	2.29	0.023±0.003	436	0.022±0.003	322
	2.40	0.010±0.002	167	0.012±0.002	141

Table 4.1: Average gas fractions for the samples shown in Figures 4.3: *sample A* (red circles in the Figure), and the main *B-D sample* (blue circles).

x	$\langle x \rangle$	$\langle M_{HI}/M_\star \rangle$ *ETG sample 1*	N	$\langle M_{HI}/M_\star \rangle$ *ETG sample 2*	N
Log $M_\star$	10.16	0.091±0.011	311	0.118±0.019	139
	10.45	0.056±0.006	389	0.049±0.006	294
	10.74	0.043±0.005	351	0.041±0.005	328
	11.03	0.018±0.003	183	0.014±0.004	178
	11.28	0.007*	32	0.009*	25
Log $\mu_\star$	7.99	-	-	-	-
	8.27	0.336±0.109	5	0.435±0.154	3
	8.56	0.107±0.019	120	0.139±0.047	41
	8.86	0.071±0.007	461	0.073±0.010	258
	9.13	0.037±0.004	591	0.036±0.004	535
	9.40	0.029±0.011	89	0.020±0.005	124
NUV$-r$	2.20	0.301±0.121	10	0.569±0.293	3
	2.82	0.220±0.020	77	0.247±0.038	29
	3.58	0.125±0.014	127	0.152±0.020	74
	4.40	0.072±0.013	180	0.084±0.011	118
	5.24	0.027±0.005	468	0.022±0.004	387
	5.89	0.014±0.004	349	0.021±0.004	315
Log σ	1.90	0.085±0.026	50	0.058±0.016	99
	2.01	0.074±0.016	114	0.059±0.006	302
	2.10	0.047±0.008	234	0.059±0.010	136
	2.20	0.042±0.006	287	0.048±0.006	252
	2.29	0.022±0.004	222	0.025±0.004	246
	2.40	0.008*	112	0.009*	123

Table 4.2: Average gas fractions for the additional *B-D samples* shown in Figures 4.3: *ETG sample 1* defined with more stringent cuts on inclination (black circles in the Figure) and *ETG sample 2* obtained with the cut on C (cyan circles). Asterisks indicate upper limits.

5

AGN feedback on the gas content of massive galaxies.

Based on
S. Fabello, et al.
MNRAS, 2011, 416, 1739

5.1 Introduction

Observations have shown that most low redshift massive galaxies contain a black hole (BH) in their bulges (Kormendy 2004) and that in some cases, the black hole may be actively accreting. The mass of the black hole is strongly coupled with the mass and velocity dispersion of the bulge of its host galaxy (Kormendy & Richstone 1995; Magorrian et al. 1998; Gebhardt et al. 2000; Ferrarese & Merritt 2000; Tremaine et al. 2002). Whether or not energy output by accreting black holes can influence the evolution and properties of their host galaxies is still a matter of debate. In some theoretical models of galaxy evolution, (radio) AGN are assumed to heat the external gas reservoir in massive dark matter halos (eg. Bower et al. 2006; Cattaneo et al. 2006; Croton et al. 2006; De Lucia & Blaizot 2007; Somerville et al. 2008), thereby quenching the growth of the most massive galaxies. It is believed that this process may solve the over-cooling problem in simulated galaxy clusters (eg. Churazov et al. 2001; Reynolds et al. 2001; Quilis et al. 2001; Ruszkowski & Begelman 2002; Sijacki & Springel 2006; Teyssier et al. 2011). Another favored evolutionary scenario is that the energy output by the AGN actually drives gas out of the galaxy itself, depleting its interstellar medium and therefore shutting down star formation (e.g. Di Matteo et al.

2005; Hopkins et al. 2006).

Observationally, the situation is still very unclear. In nearby galaxies (e.g. Kauffmann et al. 2003a; Heckman et al. 2004; Schawinski et al. 2007) and galaxies at intermediate redshifts (e.g. Nandra et al. 2007; Silverman et al. 2008), the host galaxies of AGN are found to occupy the "green" part of the colour - magnitude diagram, i.e. their stellar populations are intermediate in age between the blue, actively star-forming spiral galaxies, and the gas-poor, quiescent early-type population. One interpretation of these results is that there is an evolutionary sequence driven by nuclear activity: i.e. AGN are triggered in galaxies on the blue sequence, gas in these systems is heated, and galaxies move onto the red sequence as star formation shuts down (Schawinski et al. 2009b). However, this interpretation is certainly not unique. At a fixed black hole mass, AGN tend to be found preferentially in galaxies with younger stellar populations (and hence more gas). The presence of a gaseous fuel supply may simply be a necessary condition for both star formation and accretion onto the black hole (Kauffmann et al. 2003a; Kauffmann & Heckman 2009). A possible intermediate solution has been proposed by Diamond-Stanic & Rieke (2012), who compared the star formation of AGN hosts on different scales versus their nuclear activity. They found a strong correlation between BH accretion rate and star formation in the inner regions, but only a weak trend with the global star formation. They therefore argued for a connection between the BH and the gas only on scales smaller than $\sim$100 pc.

One way to ascertain whether AGN are really responsible for terminating the growth of massive galaxies, is to establish a direct link between the presence of an active nucleus and the *gas content* of such galaxies. There have only been a few *systematic* studies of the cold gas content, both neutral and molecular, of the host galaxies of moderately large samples of AGN. Ho et al. (2008) conducted a survey of neutral gas in a representative sample of 154 moderate- to high-luminosity nearby Type I AGN. Their results appear to challenge the feedback scenario because AGN hosts are found to be at least as gas rich as quiescent objects of the same morphological type. Hughes & Cortese (2009) carried out a study of HI in a volume-limited sample of galaxies in different environments around the Virgo cluster, trying to disentangle between the effect of AGN and environment on star formation. In their sample, the incidence of AGN peaks in the "green valley", but galaxies with quenched star formation are found to be mainly HI-deficient galaxies located in high-density environments. No direct connection between AGN activity and quenching of star formation was observed, at odds with the previous cited studies. Observations of the molecular gas (H_2) may give better insight, as this gas component is typically more centrally distributed than the atomic one. So far there are unfortunately very limited studies. Schawinski et al. (2009a) observed the total molecular content of a small sample

of early-types with the aim of directly observing the feedback from the BH as the transition mechanisms from star forming to passive systems. They have a limited sample, but argue for a possible quenching effect since the active galaxies in their sample are gas poorer than the star forming ones, and their depletion time too rapid to be due to pure star formation.

Aiming to establish a direct connection between AGN activity and gas content, in this chapter we measure the average HI content of a sample of ∼2000 AGN host galaxies, using the stacking. We build control samples and we look for differences in HI mass fraction ($M_{HI}/M_\star$) between AGN and control galaxies. In addition, we use the data from the CO Legacy Database for GASS (COLD GASS, Saintonge et al. 2011a) survey of molecular gas, which has selection criteria analogous to ours, to repeat the analysis for the H_2 content.

In Section 2 we describe the HI sample selection and the data we use. In Section 3 we compare the gas content of AGN and inactive counterparts in bins of nuclear properties; in Section 4 we study the molecular gas content. A discussion of the results is presented in Section 5.

5.2 AGN and control samples

We have shown in the previous chapter that the HI content of massive galaxies is, to first order, most tightly correlated with optical/UV (NUV$-r$) colour and stellar mass surface density. For an unbiased comparison of the cold gas in AGN hosts and quiescent galaxies we must match the two classes in these two properties, to remove first order dependencies.

Starting from *sample A* (§ 2.1) we extract the AGN (§ 5.2.1) and a control sample of galaxies where there is no evidence for accretion onto a central black hole (§ 5.2.2). We present the selection criteria applied, and the additional parameters we use in the analysis but that have not yet been described in Sections 2.4 and 2.5.

5.2.1 The AGN sample

We identify the AGN in *sample A* using the diagnostic diagram first defined by Baldwin et al. (1981, hereafter BPT). They showed that it is possible to distinguish type 2 AGN (in which the emission is obscured by a dusty circumnuclear medium, see the unified model by Antonucci 1993) from normal star-forming galaxies using the intensity of the ratios of nebular emission lines. Since the different classes of objects are dominated by different excitation mechanisms, the relative line intensities are different. In particular, we consider the ratio [OIII]/Hβ versus [NII]/Hα to identify AGN, applying a cut in signal-to-noise ≥ 3 in *all* four line measurements in order to have reliable estimates. The fluxes

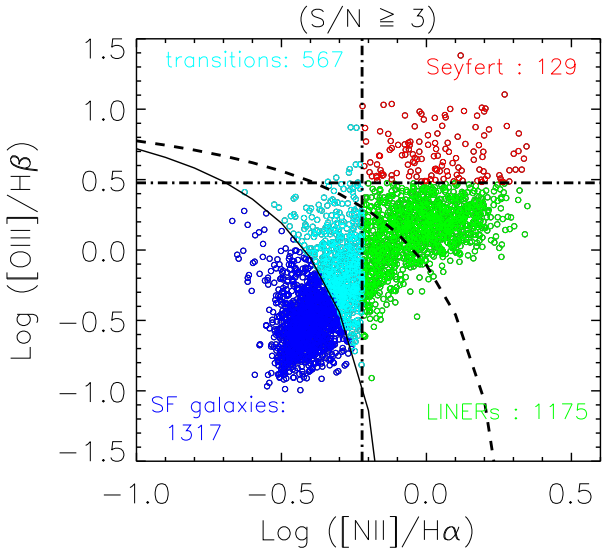


Figure 5.1: BPT diagram for galaxies in our sample with S/N≥3 for the four emission lines Hα, Hβ, [OIII] and [NII]. The solid curve we use to demarcate the boundary between AGN and "normal" star-forming (SF) galaxies is from Kauffmann et al. (2003a); the dashed line shows the demarcation boundary for "pure" AGN from Kewley et al. (2001). Objects are colour-coded according to their nuclear properties as labeled in the diagram.

are measured inside the SDSS 3-arcseconds fibre, which covers the inner regions of our galaxies, therefore we can safely state that the AGN properties derived from them are not significantly contaminated by outer star forming regions (see Kewley et al. 2006).

The locus of *sample A* galaxies that pass our signal-to-noise cut on the BPT diagram is shown in Figure 5.1. We adopt the Kauffmann et al. (2003a) prescription to separate AGN and star forming objects, shown as solid line in the Figure, so that active galaxies lie in the plane region identified by:

$$\log([\mathrm{OIII}]/\mathrm{H}\beta) \geq 0.61/\{\log([\mathrm{NII}]/\mathrm{H}\alpha) - 0.05\} + 1.3.$$

We note that Kewley et al. (2001) suggested a more stringent cut to separate galaxies where the nuclear signal is almost completely excited by emission from the gas accreting onto the black holes, rather than from star-forming regions. With this cut (shown as a dashed line in Figure 5.1), we obtain a subsample of 912 "pure" AGN, which is too small for our stacking analysis. We then consider the demarcation line suggested by Kauffmann et al. (2003a), which allows us to define a final sample of 1871 active galaxies (hereafter *AGN sample*).

We can further subdivide the AGN targets into Seyferts, LINERs (which make up the

bulk of our sample) or "transition", as indicated by the different colours in Figure 5.1. As a quick review (but see also Ho et al. 1997; Kauffmann et al. 2003a, and references therein), these are the classes in which type 2 AGN are commonly divided. LINERs (Heckman 1980b) display high values of the ratio of lower ionization lines [NII]/Hα and are characterised by lower nuclear luminosities than Seyfert galaxies. The class of "transition" objects has been introduced to separate galaxies with nuclear emission-line properties that are intermediate between normal star-forming galaxies and Seyferts and LINERs; they are actually likely to be normal AGN whose emission is contaminated by HII regions close to the nucleus (Ho et al. 1993).

Finally, we plot the AGN objects in the NUV$-r$ colour - stellar mass surface density plane, which are the two parameters we use to match AGN and quiescent targets. In the left panel of Figure 5.2, black dots represent the *AGN sample*, magenta dots the "pure" AGN defined using the Kewley et al. (2001) cut (dashed line in Figure 5.1). We note that "pure" AGN span the same range of parameter space as the larger *AGN sample* adopted in this work; the inclusion of transition objects does not bias our study toward bluer, star forming galaxies.

Additional parameters

Once we have identified the active galaxies in our sample, we need a tracer of their nuclear activity. As discussed in Kauffmann et al. (2003a), the luminosity of the [OIII]λ5007 line should be a reasonably reliable tracer of black hole activity once corrected for dust extinction. Because it is a high excitation line, it is less contaminated by emission from HII regions than other lines such as Hα. We correct the line fluxes for dust extinction using the Balmer decrement, as explained in § 2.4, and then convert the [OIII] fluxes into luminosities.

From the luminosity, we can estimate the Eddington ratio for the AGN, which is defined as the ratio between the bolometric luminosity and the Eddington luminosity: $\mathrm{L}_{EDD} = 1.26 \cdot 10^{38}\,\mathrm{M}_{BH}/\mathrm{M}_{\odot}\,erg\,s^{-1}$. The Eddington luminosity is the limit at which the gravitational attraction balances the radiation pressure of the infalling material. We actually estimate a proxy for the Eddington ratio, using: :

$$\frac{\mathrm{L}_{bol}}{\mathrm{L}_{EDD}} \propto \frac{\mathrm{L}_{bol}}{\mathrm{M}_{BH}} \propto \frac{\mathrm{L[OIII]}}{\sigma^4}$$

where L_{bol} is the bolometric luminosity, and M_{BH} the central black hole mass. The Eddington luminosity L_{EDD} scales with M_{BH}, which we estimate using the stellar velocity dispersion σ measured from the fibre spectrum: $\mathrm{M}_{BH} \propto \sigma^4$ (Tremaine et al. 2002, and

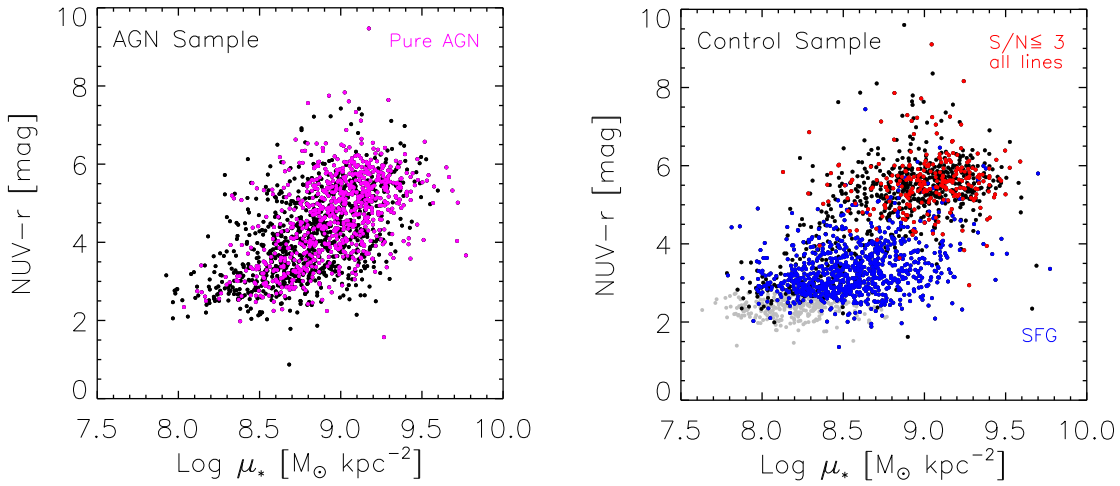


Figure 5.2: Galaxies in our samples are plotted in the (NUV−r)-$\mu_\star$ plane. *Left*: AGN targets. Black points show the whole *AGN sample*, but we overplot as magenta points galaxies that lie above the stronger cut defined by Kewley et al. (2001). *Right*: control galaxies, including "inactive" galaxies with S/N([OIII])< 3 (red dots are galaxies for which the S/N in all the four lines is smaller than 3) plus star-forming objects (SFG; blue dots). Gray dots are the galaxies that are discarded when matching to the AGN sample (i.e. mainly very blue, low stellar surface density objects). Black and coloured dots show the final *control sample* extracted.

references therein). L_{bol} scales linearly with L[OIII] (see Heckman et al. 2004), so to first order the ratio L[OIII]/σ^4 can be used as a proxy for the accretion rate. For a precise estimate of the Eddington parameter we refer the reader to Heckman et al. (2004) and Kauffmann & Heckman (2009).

5.2.2 The control sample

Our control sample is drawn from a combined sample of "inactive" galaxies and star-forming systems. Inactive galaxies are those whose spectrum exhibits a S/N of the [OIII] line smaller than 3. We exclude 474 galaxies where the S/N of the [OIII] line is greater than 3 but other lines lie below this S/N threshold, because these galaxies cannot be accurately classified. By applying a cut in the [OIII] line only, our control sample may be contaminated by weak LINERs. The contamination is anyway negligible: for our targets, less than 5% of the galaxies with S/N([OIII]) < 3 have Log([NII]/Hα) > Log(0.6) (vertical dash dotted line in Figure 5.1).

Inactive galaxies, together with galaxies classified as star-forming on the BPT diagram,

constitute our pool of 2377 galaxies for which the nuclear emission is not dominated by an AGN. The colour-$\mu_\star$ plane for those galaxies is shown in the right panel of Figure 5.2. Two distinct populations are clear in this diagram: a red sequence of galaxies with weak or absent emission lines, and a blue cloud consisting of star-forming galaxies which extends until lower $\mu_\star$ with respect to the active sample.

We stack AGN in bins of the nuclear properties, but for each bin we want to compare the average HI mass fraction with that obtained for control galaxies matched in stellar mass surface density $\mu_\star$ and NUV$-r$ colour. For each AGN we then select a galaxy from the control pool by searching for its closest neighbour in the plane defined by these two parameters. We match objects in order of decreasing $\mu_\star$, and we do not allow a galaxy to enter the control sample more than once. We do not have enough galaxies to build larger control samples, but with this method we are able to reproduce the original $\mu_\star$ and NUV$-r$ distributions of the *AGN sample*. The gray dots in Figure 5.2 (right panel) are the objects discarded by this method because they do not match any AGN. The black and coloured points show the final *control sample*. In particular, blue dots are star forming galaxies, while galaxies for which none of the four emission lines Hα, Hβ, [OIII] and [NII] are detected ($S/N < 3$) are overplotted as red dots.

5.3 The HI gas fractions of AGN hosts

The goal of this section is to establish whether or not the HI content of active galaxies is influenced by the BH properties and, if it is, by how much with respect to the quiescent corresponding targets.

We stack our *AGN sample* in bins of [OIII] line luminosity and in bins of black hole accretion rate proxy L[OIII]$/\sigma^4$. In Figure 5.3, we plot the average gas fraction $M_{HI}/M_\star$ as a function of both parameters, for both the AGN sample (red) and for the control galaxies (black). Errors are evaluated using the bootstrap method (§ 3.2.4).

We first focus on the left plot, where we study whether there is a correlation between gas content and nuclear activity (as traced by L[OIII]). As can be seen, the average HI mass fraction increases weakly as a function of [OIII], from about 9% for the weakest AGN to ∼14% for the most luminous systems. This nearly flat trend reflects the fact that AGN of different [OIII] luminosities are located in roughly the same region of the $\mu_\star$-(NUV$-r$) plane. If we scale the nuclear activity to the size of the black hole, though, the dependence of HI mass fraction becomes considerably stronger. This is clear in the right panel, where we consider the black hole accretion rate. Objects with the lowest measurable accretion

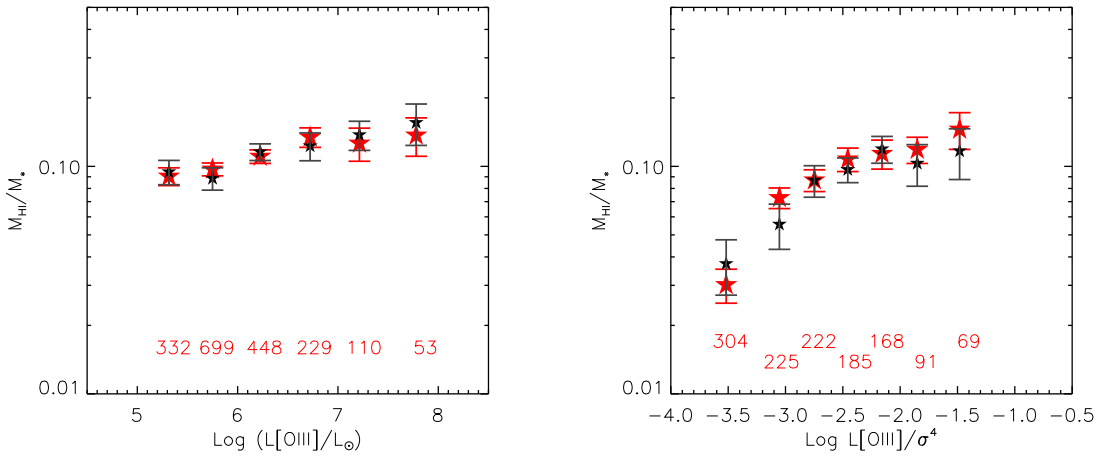


Figure 5.3: HI gas fraction as a function of L[OIII] (*left panel*) and L[OIII]/σ^4 (*right panel*). Filled red stars are the values obtained for the AGN hosts, while the black stars are the values obtained for the corresponding control galaxies. The number of galaxies included in each stack is reported at the bottom of the plot at the position of each bin.

rates are very gas poor ($M_{HI}/M_\star \sim 3\%$), and the average gas fraction increases as a function of L[OIII]/σ^4 up to a maximum value of $\sim$14%. We likewise find that galaxies with higher black hole accretion rates tend to be bluer and have lower values of $\mu_\star$. We note that the trend in HI mass fraction as a function of black hole accretion rate is considerably weaker than as a function of NUV$-r$ colour of the host galaxy. In Section 3.3, we found in fact that the average HI mass fractions change from $\sim$55% for the bluest objects to $\sim$2% for the reddest ones.

The next step is to assess whether AGN differ from control galaxies in terms of mean HI gas mass fraction. We do not have a measurement of nuclear activity for quiescent objects, of course, but for each bin of AGN we extract control galaxies as previously explained. Results for the control sample are plotted as black stars in Figure 5.3, at the x-axis value of the corresponding AGN bin. As evident, there is no difference in the atomic gas content of the two samples, at any value of luminosity or accretion rate.

Kauffmann & Heckman (2009) showed that there appear to be two distinct regimes of black hole growth in galaxies in the local Universe. The first is associated with galaxies with significant ongoing star formation; in this regime the distribution of accretion rates shows little dependence on the central stellar population of the galaxy. The second regime is associated with "passive" galaxies, and is characterised by a power-law distribution of

accretion rates. There, the accretion rate does depend strongly on the age of the central stellar population in the galaxy.

We now take a look at HI gas mass fraction trends for blue and red galaxy subpopulations. We define red sequence galaxies to have NUV$-r \geq 4.5$, and blue sequence galaxies to have NUV$-r < 4.5$. This cut is based on the split in the bimodal colour distribution of the control sample, visible also in Figure 5.2 (right panel). Results are shown in Figure 5.4 (left panel): blue stars are blue cloud AGN, black stars are the corresponding control galaxies; red triangles represent the red sequence AGN, while black triangles are control galaxies. Symbols are empty when the measure is a non-detection (upper limit). The average HI gas fraction remains constant as a function of the accretion rate parameter L[OIII]$/\sigma^4$ for the blue population. The increase in HI gas fraction as a function of L[OIII]$/\sigma^4$ seen in Figure 5.3 is in fact driven by galaxies on the red sequence. Note that we had to increase the bin size for these objects in order to recover sufficient signal in our stacks. Our control samples of red and blue galaxies exhibit the same trends with L[OIII]$/\sigma^4$ as the AGN, but the gas mass fractions are slightly higher for the blue control objects, and lower for the red control objects (in fact, we do not detect HI in any of the red control galaxy stacks).

The NUV$-r$ colours we use are not corrected for dust extinction. It is conceivable that AGN may be found in galaxies with more dust than average, so we should ascertain whether using uncorrected colours to create matched control samples will bias our results. Reliable estimates of the specific star formation rates of our galaxies may be obtained by fitting spectral energy distribution models to the 7-band GALEX and SDSS data, and calibrating the attenuation as function of NUV$-r$ colour using a reference sample with direct measurements of the UV through far-IR spectral energy distribution. This methodology is explained in detail in § 2.5. We repeat our stacking exercise, this time matching AGN hosts to control galaxies using $\mu_\star$ and specific star formation rate (sSFR) rather than $\mu_\star$ and NUV-r colour. The new results are shown in the right panel of Figure 5.4. Red symbols denote objects with low specific star formation rates (Log SFR/M$_\star$ [yr^{-1}] < -11.0), while blue symbols are for more strongly star-forming galaxies (Log SFR/M$_\star$ [yr^{-1}] > -11.0). Once again, this division is based on the bimodal distribution of sSFR in our sample. We see the same trends as before, with no significant difference in HI content between the blue AGN hosts and control galaxies. The same conclusion holds for red AGN hosts, although we note that the two highest accretion rate bins yield non-detections. This is because many of these galaxies shift over to the blue sequence after dust corrections are made.

We have already remarked that the amount of accretion onto the central supermassive

black hole in a galaxy is more likely to be sensitive to the properties of the interstellar medium in the central parts of the galaxy rather than on global scales. We now investigate what happens if we split our AGN sample into "red" and "blue" subsamples using $D_n(4000)$, which is measured within the 3 arcsecond diameter SDSS fibre aperture, and remeasure gas fractions for these two new samples, plus corresponding control galaxies. In this case, blue galaxies have $D_n(4000)<1.6$ and red galaxies have $D_n(4000)>1.7$ (this roughly separates the peaks of the bimodal distribution observed for 4000Å break strengths). In Figure 5.5 we show our results. AGN hosts exhibit no significant differences in their gas content with respect to non-AGN in the same $D_n(4000)$ range matched by global colour and $\mu_\star$. Also the differences in the red population seen in Figure 5.4 have now vanished, given that now we detect the same signal for both AGN and control galaxies. We note that the two bins of $D_n(4000)$ clearly separate in terms of the trend in gas fraction as a function of the Eddington parameter L[OIII]/σ^4, consistent with the suggestion of Kauffmann & Heckman (2009) that accretion onto the black hole is supply-limited in the red galaxy population, but not in the blue one.

5.3.1 Summary of the HI analysis

We briefly summarize the results obtained in this first part, before proceeding with the analysis. All results will be discussed in the last section of the Chapter.

We have used the stacking to measure the average HI content of a sample of 1871 AGN host galaxies extracted from *sample A*, and compared them with a control sample of non-AGN matched to the AGN hosts in the colour and $\mu_\star$ properties. We found no hint of quenching feedback acting on the global HI content of AGN at all values of black hole accretion rate probed by the galaxies in our sample. This indicates that AGN do not influence the large-scale gaseous properties of galaxies in the local Universe.

In addition, we have studied the variation of HI mass fraction with black hole accretion rate in the blue and red galaxy populations. In the blue population, the HI gas fraction is independent of BH accretion rate, indicating that accretion is not sensitive to the properties of the interstellar medium of the galaxy on large scales. However, in the red population BH accretion rate and gas fraction do correlate. We speculate that the fuel supply in the red AGN population could be a mixture of mass loss from stars and gas present in disks, but a more detailed discussion is presented later.

There are, however, two possible caveats about our results. First, we are not probing very high accretion regimes where a quenching effect from the BH may be significant.

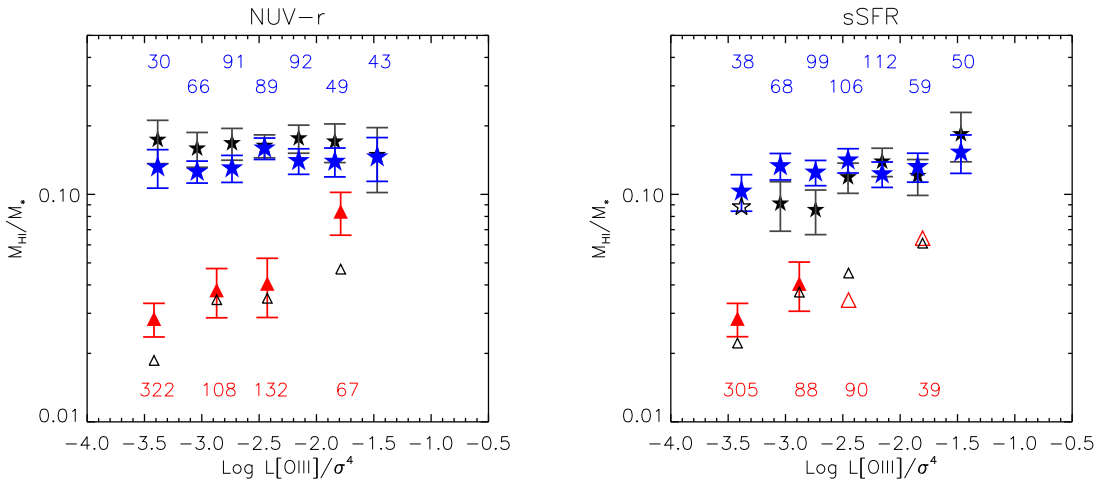


Figure 5.4: HI gas fraction as a function of L[OIII]/σ^4 for galaxies and AGN hosts, split into two bins of star formation properties. Coloured symbols show results for galaxies with active galactic nuclei, while black symbols show results obtained for the corresponding control samples. The empty symbols denote measurements that are upper limits. The number of objects stacked in each bin is reported. *Left panel*: the trend with gas fraction is shown separately for blue galaxies (defined as NUV−r<4.5), and red sequence ones (NUV−r>4.5). *Right panel*: galaxies are matched and divided into two bins of specific star formation rate; blue symbols are for star forming targets (Log SFR/$M_\star$ > -11.0), red for quiescent ones (Log SFR/$M_\star$ < -11.0).

Second, our HI measurements are integrated over the whole galaxy, but a feedback effect is likely to be stronger in the central regions. For more conclusive insights we need to compare the gas located closer to the nucleus. Since the H_2 distribution is generally more concentrated than the HI and it is reasonably affected first, we compare the molecular gas content of AGN and inactive galaxies in the next section.

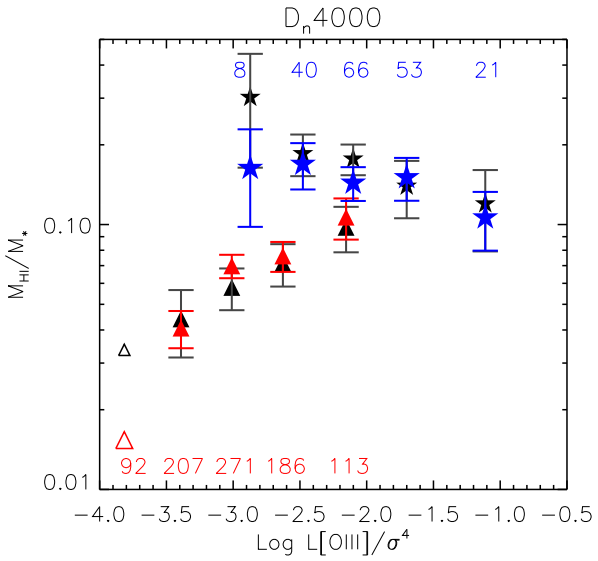


Figure 5.5: HI gas fraction as a function of L[OIII]/σ^4, for galaxies divided into two subsamples according to 4000Å break strength: red triangles are galaxies with $D_n(4000)$>1.7 and blue stars with $D_n(4000)$<1.6. Symbols are as in Figure 5.4.

5.4 The H_2 content of AGN hosts

We do not have H_2 measurements available for our *sample A*, therefore we use data from the on-going COLD GASS survey. COLD GASS observes the CO(1→0) line of a sample of ∼ 350 objects randomly extracted from the GASS sample. The survey is carried out at the IRAM 30m telescope, whose beam has a FWHM of 22 arcseconds at the wavelength observed. In most of the cases only a single pointing is necessary to cover the galaxy; for targets with optical diameter larger than the beam, either an aperture correction is applied or and additional pointing is performed (for further details see Saintonge et al. 2011a, § 4.4 there).

COLD GASS has targeted 291 galaxies to date (DR2, Saintonge et al. in preparation). Of these, 162 are detected in the CO(1→0) line and the others are observed until an upper limit of $M_{H_2}/M_\star$∼1.5% is reached. For the COLD GASS sample H_2 and HI (from GASS) total masses are available, and we estimate from GALEX and SDSS the needed parameters following the same procedures described in Chapter 2. The sample is of course different from *sample A*, but the selection criteria for the survey are the same adopted in this work (see details in § 2.1), so that we can complement the HI analysis without bias over the same high mass range. Unfortunately, the statistics for COLD GASS is limited, so we cannot perform a complete analysis as in the HI case. We will not use the stacking technique

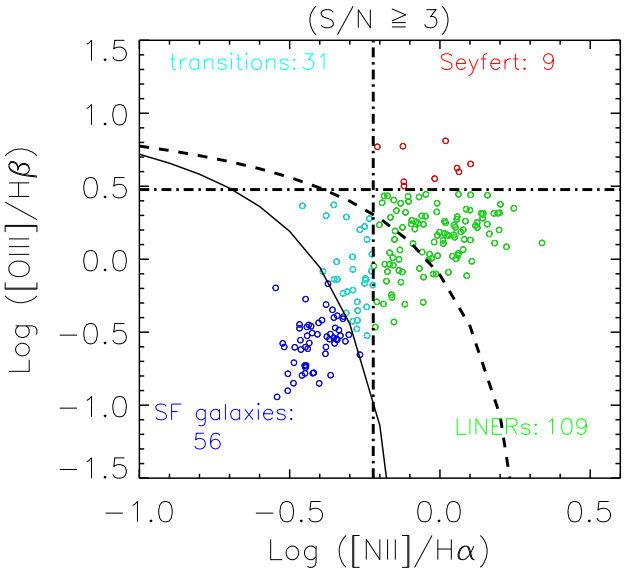


Figure 5.6: BPT diagram for COLD GASS galaxies, detected in the four emission lines used. Galaxies are colour-coded according to their nuclear properties. We identify as AGN targets above the solid line from Kauffmann et al. (2003). The symbols and lines are the same as in Figure 5.1.

either, since even a spectrum obtained by summing all the H_2 non detections does not yield a clear signal. Our approach in this section is to first study scaling relations as a function of nuclear activity using all the AGN hosts, and second to create control samples of inactive galaxies for a direct comparison of their H_2 content. Unlike for the HI analysis, we match AGN hosts and inactive galaxies only by their NUV$-r$ colour. Saintonge et al. (2011a) have shown that this is the parameter that exhibits the strongest correlation with molecular gas fractions for COLD GASS galaxies. The molecular gas content depends only weakly on stellar mass or $\mu_\star$.

As already noted, the COLD GASS measurements, like the ALFALFA ones, are obtained integrating over the whole galaxy. We therefore do not really resolve the central regions of our targets. However, Schawinski et al. (2009a) used the same telescope to measure integrated molecular masses of galaxies at slightly higher redshifts ($0.05< z <0.10$, for which the beam corresponds to larger physical sizes) and found a possible feedback signature; if our AGN are gas depleted we should be able to see the effect.

To extract the AGN sample and the inactive galaxies, we use the same BPT diagnostic applied to *sample A* (§ 5.2.1). We obtain a sample of 149 AGN galaxies, 118 of which are

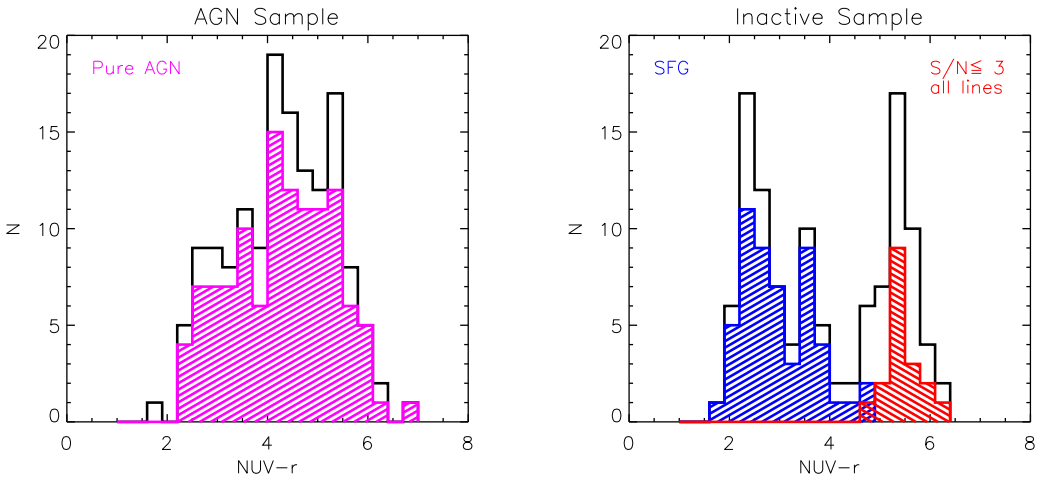


Figure 5.7: NUV$-r$ colour distribution for COLD GASS objects. *Left*: AGN targets. The black histogram shows the whole *AGN sample*, but we overplot as magenta galaxies that lie above the stronger cut defined by Kewley et al. (2001). *Right*: control galaxies. They include "inactive" galaxies with S/N([OIII])<3 (the red histogram displays the colours of galaxies for which the S/N in all the four lines is smaller than 3) plus the pure star-forming objects (blue histogram).

identified as "pure" AGN. The BPT diagram for the COLD GASS sample with S/N>3 in all the four lines needed is shown in Figure 5.6, where we colour code the targets according to their nuclear properties. As in the previous sections, star-forming galaxies plus objects non detected in the [OIII] line make up our sample of 115 inactive targets.

Figure 5.7 displays the NUV$-r$ histograms for the two samples: AGN (left) and inactive galaxies (right). We notice that the AGN tend to populate the whole NUV$-r$ range, while inactive galaxies are clearly separated in a blue and a red population.

Gas content and SF vs BH accretion rate

By stacking HI data we found that there are two regimes of accretion rate onto the central black hole: blue galaxies (both global and central colours) have a higher HI gas fraction which does not depend on the accretion rate, while red objects display a trend between the two properties. Other analyses, performed on samples lacking gas data (eg. Diamond-Stanic & Rieke 2011), claim that the accretion rate strongly correlates with nuclear star formation but only weakly with the global star formation.

If we assume that the black hole accretion rate correlates with the gas in the central regions, then we expect to reproduce both the afore-mentioned results with our sample,

for the following arguments. *i)* If we measure a dependence of the H_2 gas fraction on the BH activity we also expect a correlation of L_{bol}/L_{EDD} with SF (in the central regions), because star formation and H_2 content are strictly connected. *ii)* On larger scales and for gas rich objects, these correlations must instead get diluted, in particular in the galaxy regions where the atomic gas phase dominates. This may furthermore explain why in blue objects (therefore on average gas richer) we do not measure any dependence of the neutral gas on the accretion rate.

In Figure 5.8, we use the COLD GASS sample with individual H_2 and HI data to test these hypotheses (*i* and *ii*). In the top row, we show the H_2 (left) and HI (right) gas fractions as function of the Eddington accretion rate proxy (evaluated as L[OIII]/σ^4) for the active galaxies. Objects are colour-coded according to their NUV$-r$ colour to separate the red and blue populations. Arrows indicate the upper limits evaluated in the case of non detection.

The accretion rate clearly displays a tighter correlation with the molecular gas content (left panel) than with the atomic one (right). The correlation coefficient[1] measured considering detections only is $r = 0.55$ in the H_2 case, and $r = 0.21$ for the HI. H_2 non detections correspond on average to galaxies with very low accretion rate, while actively accreting BH are associated with H_2 rich systems. On the contrary the average HI gas fraction for the blue population, which is almost entirely detected, varies significantly at fixed accretion rate, but its mean value of $\sim$10% is consistent with the average value found with the stacking in Section 5.3. We cannot measure the average gas fraction of the red objects, as GASS does not detect half of them. However, the fact that the non detections all have low accretion rates suggests a combined decrease of HI gas content and accretion in red objects.

The stronger correlation of accretion rate with H_2 than with HI is reflected in the star formation properties. In the bottom row of Figure 5.8 we compare the specific SFR measured inside the SDSS fibre (left panel) and global star formation rate (right) as a function of accretion rate. The accretion onto the BH exhibits a better correlation with the central star formation ($r = 0.60$), while a much larger scatter is visible as a function of the global one ($r = 0.46$) .

A correlation indeed exists between the H_2 gas content in the central regions of a galaxy and the accretion activity onto its BH. Nevertheless, without a comparison with control galaxies we cannot univocally determine whether the BH is quenching the gas or

[1]Measured as the ratio of the covariance of the sample populations to the product of their standard deviations.

we are observing a natural co-evolution toward the ending phase (low gas supply, low star formation, low accretion) of both the AGN and the host.

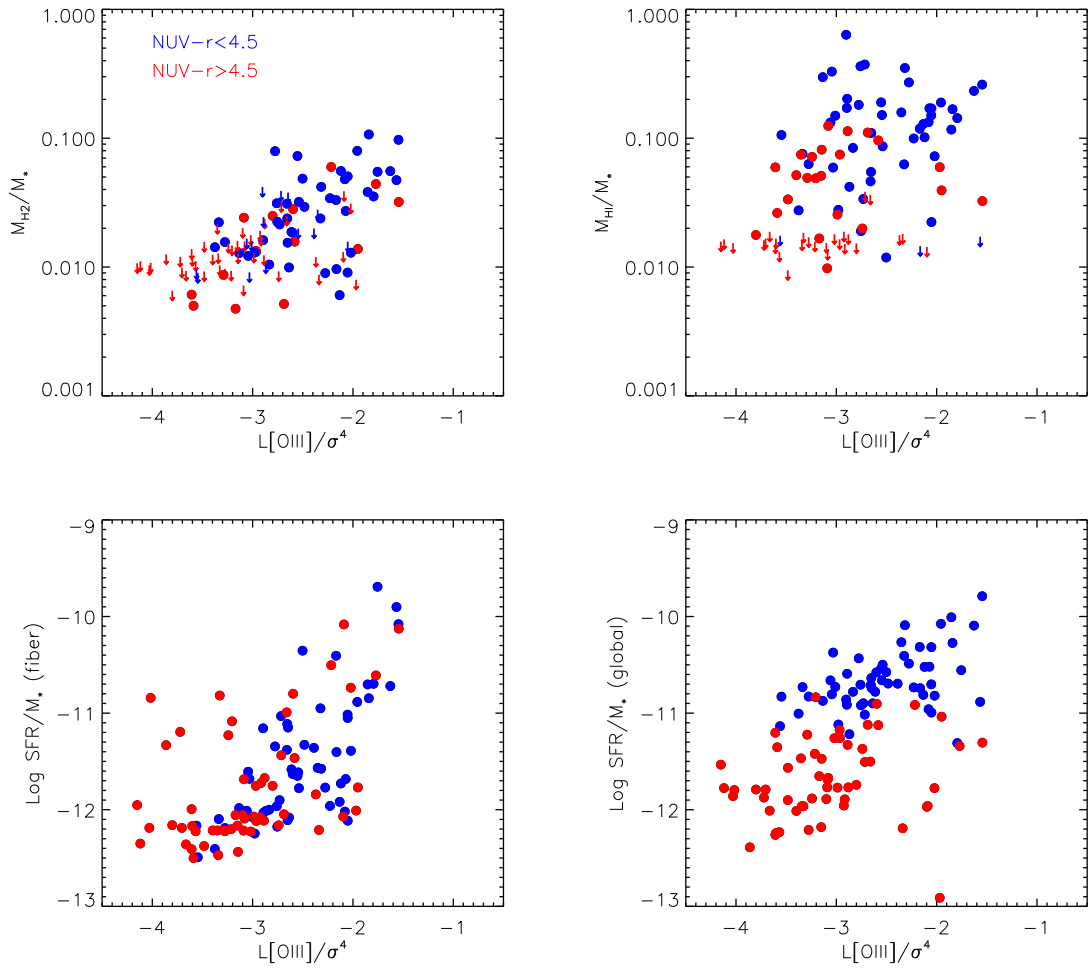


Figure 5.8: Gas and star formation properties versus BH accretion rate, plotted separately for the blue and red populations. The NUV−r colour bins used are indicated in the upper left corner. Arrows indicate upper limits. *Top row:* $M_{H2}/M_\star$ (left) and $M_{HI}/M_\star$ (right) as a function of L[OIII]/σ^4. *Bottom row:* fibre specific SFR (left) and global specific SFR (right) versus accretion rate.

H_2 in AGN-hosts and control galaxies

To check an eventual quenching effect of the active BH on the H_2 content of its host, we need to compare the gas content of active and inactive galaxies at *fixed* values of NUV−r

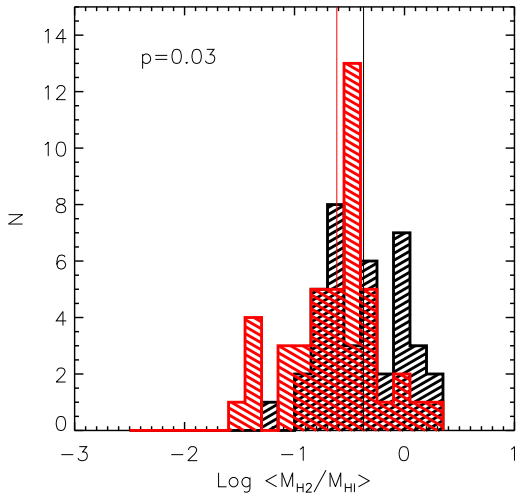


Figure 5.9: Comparison of the distributions of M_{H_2}/M_{HI} for one of the couple of sub-samples of AGN (red histogram) and control targets (black histogram) matched in NUV$-r$. The mean values of the two distributions are flagged; the probability that the two sub-samples are drawn from the same distribution is reported in the upper left corner.

colour. We do not have enough statistics to extract a control inactive sample that matches the properties of the AGN one, as done in Section 5.2.2. We mainly lack "green" inactive galaxies, as clearly visible in Figure 5.7. We can only extract sub-samples of both active and inactive galaxies so that their *final* distributions of NUV$-r$ are the same, even if they do not resemble the original ones; i.e., their average H_2 content would not be representative of the whole population, but the comparison between AGN and inactive objects would be meaningful.

We extract 5 couples of sub-samples matched in the NUV$-r$ properties. If we extract the largest possible sub-populations that match, the average sub-sample sizes are: 95 AGN and 80 inactive galaxies. We cannot repeat the extraction process too many times, because we would start extracting the same targets since the initial samples are small.

For each couple of sub-samples we compare the distributions of the gas properties of AGN and inactive galaxies, to asses whether they are similar or not. In Figure 5.9, for example, we show the distribution of the molecular-to-atomic ratio for a sub-sample of AGN (red histogram) and the correspondent control galaxies (black histogram), considering H_2 and HI detections only. The mean values of the distributions are flagged. In the upper left corner we report the probability, estimated with a Kolmogorov-Smirnov test, that the two distributions are drawn from the same one. The shape of the distributions are similar, but

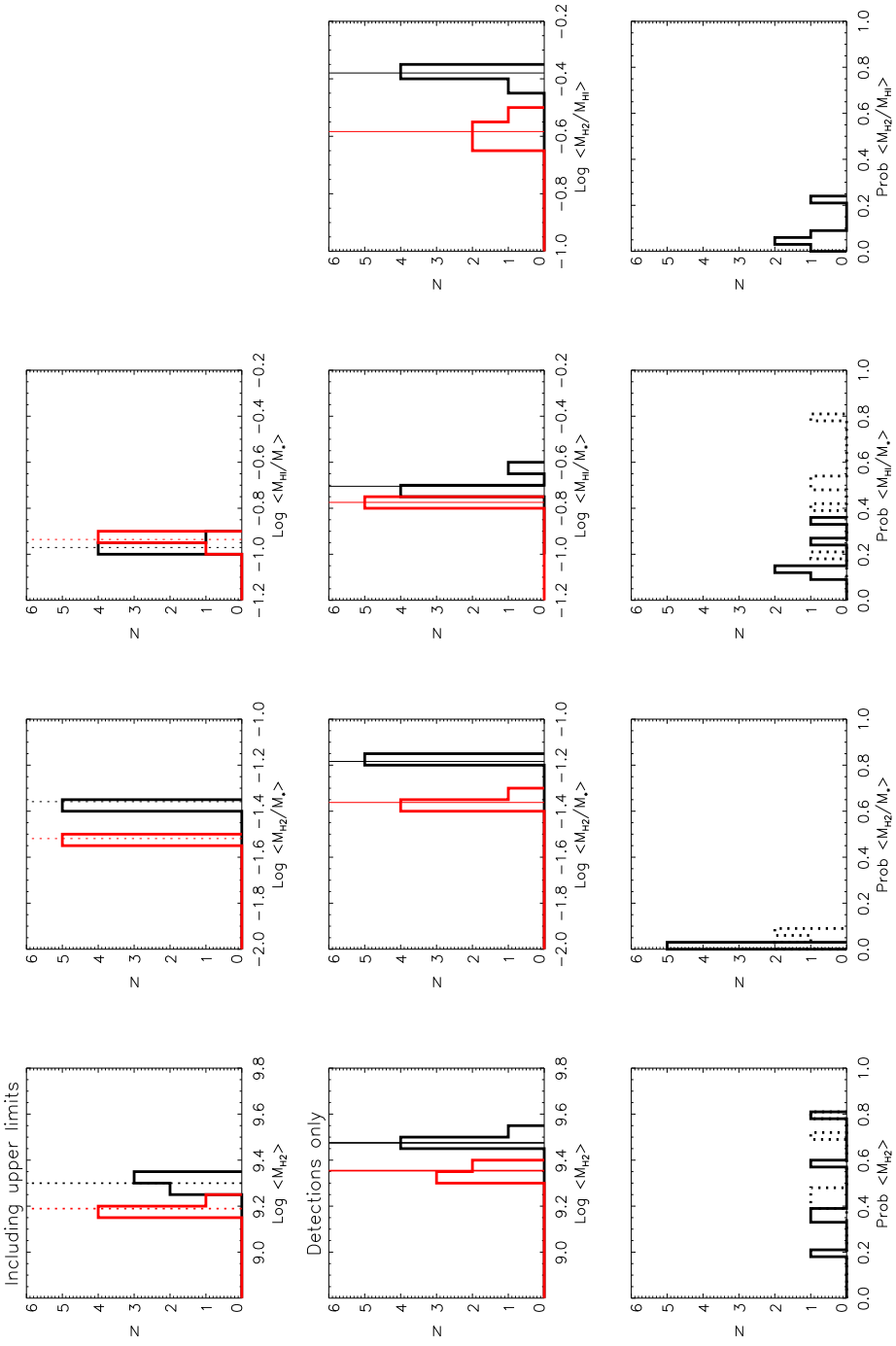


Figure 5.10: Comparison of gas properties of AGN (red histograms) and control targets (black histograms) matched in NUV$-r$. The first row includes non-detections as upper limits; the middle row considers only detections. The mean values of the two distributions are flagged in each panel; solid lines correspond to the detections only, dotted lines to detections plus upper limits. The bottom row is the probability that the two sub-samples are drawn from the same distribution. *From left to right:* average values of M_{H_2}, $M_{H_2}/M_\star$, $M_{HI}/M_\star$ and M_{H_2}/M_{HI}.

AGN are slightly offset toward lower molecular gas contents.

With the same approach, we compare AGN and inactive galaxies in different gas properties. We summarize our results in Figure 5.10 where we compare, from left to right respectively, the average values of M_{H_2}, $M_{H_2}/M_\star$, $M_{HI}/M_\star$ and M_{H_2}/M_{HI}. The bottom row in the Figure shows, for each property and each couple of sub-samples, the probability that the distributions of the considered property are the same for AGN and control galaxies. Solid lines are obtained using detections only, dotted lines using detections plus upper limits. In the first and second rows, we report the distributions of the mean values of each property for each sub-sample, considering only detections (middle row) or including upper limits as well (top row). AGN are represented by red histograms, control targets by black ones. If we visually inspect the different panels, we notice that given the same distribution of NUV$-r$ colours, AGN *tend* to have smaller M_{H2} (first column, this is the less evident result), smaller H_2 gas fraction (second column) and smaller H_2/HI ratios (fourth one). In the case of molecular-to-atomic ratio, we use the detections only. It is finally encouraging to see that the HI gas fractions (third column) are instead similar, in particular when we include also the non detections consistently with the outcome of the stacking analysis.

We are aware that because of the very limited statistics this part of the analysis returns only tentative results. Nevertheless, the plots in Figure 5.10 suggest that direct depletion of the H_2 content by the AGN, though not strong, is in place.

5.5 Conclusions

In this chapter we have looked at the possible interplay between nuclear activity and hydrogen content, both atomic and molecular, in massive nearby galaxies. For the HI case we have selected a sample of AGN and a control sample of inactive galaxies matched by $\mu_\star$ and NUV$-r$ starting from *sample A*, and we have performed a stacking analysis. For the study of the molecular gas, we have used a small sample of massive galaxies extracted from the COLD GASS survey and looked for correlations using individual measurements (as opposed to stacking).

Our analysis focused mainly on two aspects, which we summarize and discuss here: *i)* we looked at interplays between gas and BH activity in AGN hosts; *ii)* we tried to assess whether the gas content of active galaxies differs from that of the control galaxies.

***i)* Gas content versus BH accretion rate.**

A first interesting result of our analysis is the scaling of the mean HI gas fraction of AGN host galaxies with accretion rate onto the black hole, as traced by L[OIII]/σ^4. In the blue galaxy population, the accretion rate onto the black hole and the HI gas fraction of the galaxy are independent. However, in the red galaxy population, these two quantities do track each other. Kauffmann & Heckman (2009) hypothesized that in blue galaxies, the black hole regulates its own growth at a rate which does not further depend on the properties of the interstellar medium. In red galaxies with old stellar populations, they deduced that the decrease in the accretion rate onto black holes was consistent with population synthesis model predictions of the decline in stellar mass loss rates as a function of mean stellar age.

Because we now have measured HI mass fractions for the red population, we can check this in more detail. The average global (NUV$-r$)/($g-r$) colours of the host galaxies vary from 4.56/0.74 for the stack with the highest accretion rate in Figure 5.5 to 5.54/0.80 for the stack with the lowest accretion rate. As described in Kauffmann & Heckman (2009) we use population synthesis models (e.g., Bruzual & Charlot 2003; Maraston 2005) to calculate mass loss rates as a function of colour, assuming simple exponential declining star formation histories for our galaxies. If the HI content of red sequence AGN came only from mass loss, we would expect a 0.3 dex change in the average $\log(M_{HI}/M_\star)$ between AGN stacks with the lowest and highest accretion rates in our sample. Figure 5.5 shows a 0.5 dex change in this quantity, which is somewhat larger than predicted. However, we note that the average stellar mass surface densities of the host galaxies in the stacks decrease by a factor of 3 between AGN with the highest and lowest accretion rates (likewise the concentration index decreases from 3.15 to 2.79), so structural properties are not constant across the population. This may imply that the fuel supply in the red AGN population is in fact a mixture of mass loss from stars and gas present in disks.

We argued that the HI results here presented are consistent with previous works, which measured strong dependence of the *central* star formation on the accretion onto the black hole (Diamond-Stanic & Rieke 2012), and found hints of molecular gas quenching in AGN hosts (Schawinski et al. 2009a). The molecular gas is generally more centrally concentrated that the atomic one and may be more sensitive to feedback processes occurring in the bulge of the galaxy. The weak (or no) correlations we measured for the HI are the result of looking at too large scales, especially in gas rich (blue) objects where the neutral gas is expected to extend up to large radii.

Consistently with this picture, we measured a clear correlation between the H_2 gas fraction and the Eddington ratio ($r = 0.55$ with respect to a value of $r = 0.21$ measure for

the HI). This connection is reflected in the correlation between accretion rate and central star formation ($r = 0.60$), which is stronger than the one obtained considering global star formation ($r = 0.46$).

***ii)* Quenching of the gas by feedback?**

We do not find any significant difference between the HI gas content of AGN hosts and that of control galaxies. This conclusion holds at all values of the Eddington parameter proxy L[OIII]/σ^4 probed by the galaxies in our sample. If converted into a proper Eddington ratio, our sample would cover the range $\mathrm{Log}\,(\mathrm{L}/\mathrm{L}_{Edd}) \simeq [-4; -2]$. We are studying the relatively low accretion rate regime, and that could be a reason why we do not see any sign of AGN feedback. We would need data over a larger volume to stack a significant number of AGN with very high $\mathrm{L}/\mathrm{L}_{Edd}$. Still, up to the highest accretion rates we have studied, the gas fraction shows no hint of correlation with either L[OIII]/σ^4 (blue population) or AGN presence. These results are also in agreement with Ho et al. (2008). As discussed in their paper, the lack of any difference may be due to the fact that the HI traces gas on very large scales, but accreting black holes only heat gas close to the centers of the galaxies.

In some simulations of galaxy formation, AGN feedback is an extremely violent process which can drive out much of the interstellar medium from galaxies, truncating ongoing star formation and reddening their stellar populations (Di Matteo et al. 2005). In this scenario, AGN hosts would be expected to be gas deficient. Catastrophic global feedback processes are clearly not occurring in the majority of AGN in the local Universe. To ascertain whether AGN feedback is at least important in the centers of galaxies, one would ideally need to resolve the gas on the sub-kpc scales close to the nucleus, for a complete sample of AGN hosts and a well defined control sample of inactive galaxies. Such analysis on statistically significant samples will be easily performed with new upcoming instruments. The Atacama Large Millimeter Array (ALMA[2]), for example, will improve our knowledge of the cold gas down to small physical scales, thanks to its high resolution. Waiting for this new breakdown, we tried to measure clear quenching effects of the BH on the molecular gas content of COLD GASS galaxies, comparing them with control targets. Despite poor statistics, we measure promising signatures of AGN hosts being H_2 poorer than their non active counterparts.

[2]http://www.almaobservatory.org/

6

Environmental effects on HI and star formation properties of massive galaxies

6.1 Introduction

The first hint that properties of galaxies depend on the environment in which they are located dates back to the early thirties. Hubble himself first noticed how "elliptical nebulae and early spirals are relatively much more numerous than other morphological types in the dense Virgo cluster" (Hubble & Humason 1931). This observed morphology – local density relation has been the main driver of environmental studies ever since.

The first systematic studies of the relation between galaxies and their evolution in a dense environment started around the 80s with the work of Oemler (1974) and Dressler (1980). They measured how the fraction of elliptical and S0 galaxies increases with local density, with a corresponding decrease in the fraction of spirals. In the following decade, the main debate has been whether the observed morphology – density relation is a consequence of initial conditions, or is created by processes acting in dense clusters. Unfortunately, samples have been limited, so that different galaxy properties could not be studied at the same time. The dominant approach has been to focus on the differences between field/cluster galaxies, both in the optical and at radio frequencies. After the discovery of the morphology – density relation, the radio community focused on comparing the HI content of spiral galaxies located in clusters with that of their field counterparts (Haynes et al. 1984). HI in earlier morphological types was generally not detected because of the limiting sensitivity of the available radio telescopes. The observations led to the conclusion

that disks in very dense environments exhibit a deficiency of their HI content (Giovanelli & Haynes 1985), which decreases strongly toward the cluster cores (Gavazzi 1989). The trend in HI depletion matched the predicted effect of *ram pressure stripping*, a mechanism first introduced by Gunn & Gott (1972). The cold gas stripping rate in galaxies moving through a dense intracluster medium (ICM) is proportional to the density of the ICM and the velocity with which the galaxy moves through it, i.e. $\propto \rho_{ICM} v_{glx}^2$. Stripping is efficient in removing the gas when the ram pressure exceeds the gravitational force binding the ISM to the galaxy. This implies that a strong effect is expected in the dense, central regions of clusters. Interestingly, observations of HI deficient galaxies that were still star forming showed normal molecular gas in their central regions (Kenney & Young 1989). A second mechanism explaining why cluster galaxies are gas-deficient that has been proposed is *starvation*. Larson et al. (1980) suggested that the diffuse hot gas component which surrounds a galaxy is easily stripped by the ICM, thus preventing new material from cooling onto the disk and forming new stars. After a few Gyrs, the galaxy will then convert all its HI into molecular gas, consume its reservoir, "starve" and stop forming stars.

Hydrodynamic mechanisms such ram-pressure stripping could not explain an eventual morphological transformation of spirals into early-type objects; more dramatic events like mergers between galaxies are required for such changes. Investigations were carried out of the dependencies of star formation on environment at fixed morphological type. Most early studies were restricted to a field/cluster comparison and results were controversial. Some analyses found that galaxies in clusters exhibit reduced star formation (e.g. Kennicutt 1983; Dressler et al. 1985), while others found enhanced activity (e.g. Kennicutt et al. 1984; Gavazzi & Jaffe 1985). It was still not possible to understand whether the observed changes were internally driven or caused by the external environment. In the late 90s it became clear that only with a careful parametrization of environmental density and morphology the question could be properly addressed. Hashimoto et al. (1998) used a sample of 15,749 galaxies from the Las Campanas Redshift Survey to study the effects of environment on star formation using a three dimensional local density estimator. After removing the effect of the morphology - density relation using the concentration index (as objective morphological tracer), they proved that star formation rate of galaxies with a given morphology is a continuous, decreasing function of increasing local density.

From 2003 till now, large optical surveys have provided complete samples of galaxies large enough to allow galaxies to be binned according to several properties at once. The first strong result of these investigations has been that galaxy evolution is to first order set

by "nature" because the main properties of galaxies are determined by their stellar mass, and only weakly depend on environment (Kauffmann et al. 2003b; Tanaka et al. 2004). Once stellar mass is fixed, an environmental "nurture" effect does modify their properties. A second important result is that the correlation between colour/star formation and environment is actually stronger than the one between morphology and environment (Ball et al. 2008; Bamford et al. 2009; Skibba et al. 2009). Dense environments are dominated by an increasing fraction of red and quiescent objects (Bower & Balogh 2004; Balogh et al. 2004, 2009).

Disentangling different physical processes, instead, has been more difficult. Pure optical data are not sufficient to investigate the physics behind the observations, because they do not provide complete information about different structural and physical parameters. Models of galaxy evolution are currently the most successful tool we have to try to improve our understanding of the data. Semianalytical models (SAM) have been significantly improved with the aid of SDSS data, which have provided better and more detailed constraints on physics in the models. One of the most recent semi-analytic models (e.g. Guo et al. 2011) includes only a starvation-like environmental effect in their recipes, which acts on galaxies once they enter the virial radius of a bigger halo and became "satellites". Strictly speaking, starvation is not a physical process, but the global effect of a variety of mechanisms that remove the hot gas in galaxies. In Guo's SAMs, starvation is produced through both tidal and ram-pressure stripping of the hot component. To avoid confusion, here and throughout the chapter ram-pressure refers only to the stripping of the cold gas, and starvation to any mechanism removing the hot one. Models predict that, at fixed stellar mass, satellites are redder than the corresponding centrals because their star formation is not sustained by fresh material. Once a similar division in satellites and centrals is applied to redshift surveys, observations and models are qualitatively consistent (van den Bosch et al. 2008; Weinmann et al. 2009; Kimm et al. 2009). Since the optical observations of the last decade pointed toward weak environmental mechanisms that were not strongly dependent on group mass and that extended out to large distances from the center of the halo (Bower & Balogh 2004; Balogh et al. 2004; Weinmann et al. 2006), and because these effects were reproduced by the models, starvation has come to be considered as the main driver of environmental evolution in galaxies. Ram-pressure stripping of the cold gas is currently considered unlikely to play a major role in shaping galaxies properties, at least in lower density environments. According to simulations, to observe a significant effect environments richer than Virgo are necessary (Quilis et al. 2000).

Although the HI is a sensitive tracer of interactions with the ICM, current data are

still limited compared to what available in the optical. Samples are small because HI observations are time consuming, and targeted environmental studies mainly focused on the field/cluster dichotomy. All observations have actually supported the stripping scenario, finding that the fraction of HI-deficient galaxies increases in very-rich environments (Solanes et al. 2001; Boselli & Gavazzi 2009), along with the frequent presence of signatures of disturbed morphologies (Bravo-Alfaro et al. 2000; Chung et al. 2009). Comparison between dynamical models or SPH simulations and data (Abadi et al. 1999; Vollmer 2009) has shown that ram pressure can in fact be responsible for the observed distortion of the HI disk in cluster galaxies. Observations of statistically significant samples are still lacking, however, so it has been difficult to quantify at what density or halo mass the effect sets in. Fortunately, the astronomical community is currently putting a lot of effort in making a step forward in our picture of the HI content of galaxies, similar to what SDSS has done for the optical observations. The planned, next generation radio telescopes and instruments will make large, complete and deep HI surveys feasible, so that the analysis of the cold gas depletion as a continuous function of the local density, and in bins of several properties, will be possible.

Current blind HI surveys like ALFALFA do map the whole density range from void to rich clusters, but because they are shallow, the galaxies with HI detections are biased against dense environments where objects are known to be gas poor. Nevertheless, with the large amount of data available from ALFALFA, we can use the stacking technique to perform the first systematic study of the cold gas content as a function of local density for a statistically significant sample. If we constrain the (relative) effect of local density on different galaxy properties, including HI and star formation, we can gain insight into the processes at work, and try to address the still open issue of the dominant mechanisms acting at different environmental densities.

In this chapter, we study the relation between HI and SF as a function of the environment. We start by presenting the local density tracer we use for our analysis (§ 6.2). To compare star formation and cold gas dependence on environment, we measure average scaling relations for both parameters as a function of $M_\star$, because it is crucial to disentangle the effect of stellar mass and environment. In section 6.3 we present star formation rate scaling relations, and in section 6.4 the HI ones. Finally, we constrain the *relative* effect of environment on galaxy properties. To gain insight into the processes at work, we compare our results with mock catalogues from semi-analytic models applied to the Millennium Simulation. (§ 6.5). Summary and discussion on future work are presented in Section 6.6.

6.2 Our environmental tracer

Different environmental tracers are suitable to study different scales, and may produce different results (Muldrew et al. 2012). The local galaxy density evaluated inside a fixed aperture, for example, is better suited to measure large scale environments. If one chooses a scale of the order of a typical group halo, there will be a direct correlation between the environment parameter and the size of the halo itself. Following recent work based on SDSS data (e.g. Kauffmann et al. 2004; Blanton & Berlind 2007; Thomas et al. 2010), we define a density parameter for each galaxy as the number of neighbours with Log $M_\star[M_\odot] \geq 9.5$ located inside a "cylindrical" aperture of 1 Mpc radius, and $\pm 500\,\mathrm{km\ s^{-1}}$ depth, centred on the target. The value of 1 Mpc as a preferred scale for measuring environment comes about, because it is somewhat larger than the typical virial radius of haloes hosting L_* galaxies in the local Universe. On smaller scales, the number of tracer galaxies becomes too small to estimate environment reliably. A strong correlation between star formation and environment on these scales was found observationally by Kauffmann et al. (2004) and many others.

We search for neighbours among galaxies in the MPA-JHU DR7 spectroscopic catalogue. We take into account the fact that two SDSS fibres cannot be closer than 55 arcseconds, so we might miss close companions. In order to correct for this "fibre collision" effect, we follow the statistical approach of Li et al. (2006), who measured the angular two-point correlation function for the SDSS spectroscopic sample $[w_z(\theta)]$ and for the parent photometric sample $[w_p(\theta)]$. The ratio

$$F(\theta) = \frac{1 + w_p(\theta)}{1 + w_z(\theta)} \tag{6.1}$$

can then be used to account for the effect of fibre collisions. We adopt the correlation functions from Li et al. (2006), and weigh each neighbour by $F(\theta)$, where θ is the angular separation from the main target. The fibre collision corrections are actually negligible ($\sim 3\%$). The cut in Log $M_\star[M_\odot] \geq 9.5$ we impose on the spectroscopic sample ensures that we estimate the density field consistently throughout the whole sample. We note, however, that we are not considering companions down to a fixed mass fraction limit with respect to the primary galaxy.

The distribution of the density parameter for *sample A* is shown in Figure 6.1, left panel (where we plot N+1 for convenience). Almost 50% of the sample has zero or one massive neighbour inside 1 Mpc, as shown by the blue region of the histogram. In this chapter, the colour legend is the following: blue= $(0 \leq N \leq 1)$; green= $(2 \leq N \leq 9)$; red= $(N \geq 10)$.

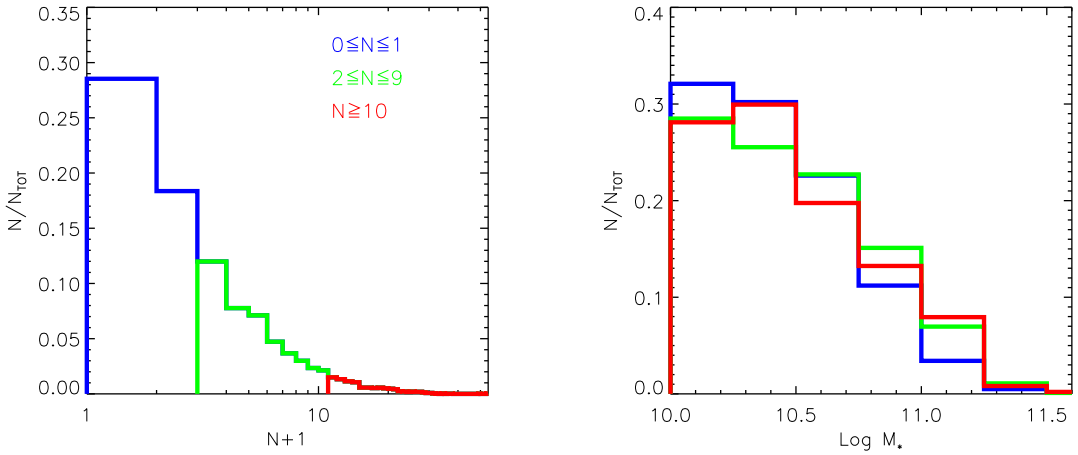


Figure 6.1: Local density and stellar mass properties for *sample A*. *Left:* normalized distribution of the density parameter, as the number of neighbours inside the chosen aperture N (+1 for convenience). Most of our targets are isolated, or in small groups. The colours represent the three density bins we will use throughout the chapter, as reported in the legend. *Right:* Normalized stellar mass distribution for each of the density bins considered.

The right panel of Figure 6.1 shows the normalised stellar mass distribution for each density bin.

Finally, in Figure 6.2 we plot two examples of rich environments in our sample. The first one (top row) is the richest bin in *sample A*, with 72 neighbours inside the aperture. On the left we plot the projection on the sky of galaxies in units of Mpc; the target galaxy is represented by the red dot. On the right, we show the three-dimensional distribution of the neighbours. This region, which is at the lower edge of our redshift range (z=0.0257), actually lies in the far outskirts of the Coma cluster. The mean redshift of the cluster has been estimated to be $cz = 6853 \pm 1082$ km s^{-1} (Colless & Dunn 1996), and the central target of our bin has $cz = 7860$ km s^{-1}. Unfortunately, most of the Coma cluster is outside our redshift range. The mean local density for *sample A* galaxies in the bin with $N \geq 10$ is 15 neighbours. We also show an example of a typical group, in this particular case with 17 neighbours, in the bottom row of Figure 6.2. The black circle around the main target indicates the physical size of the Arecibo beam at the redshift of the target (z =0.041).

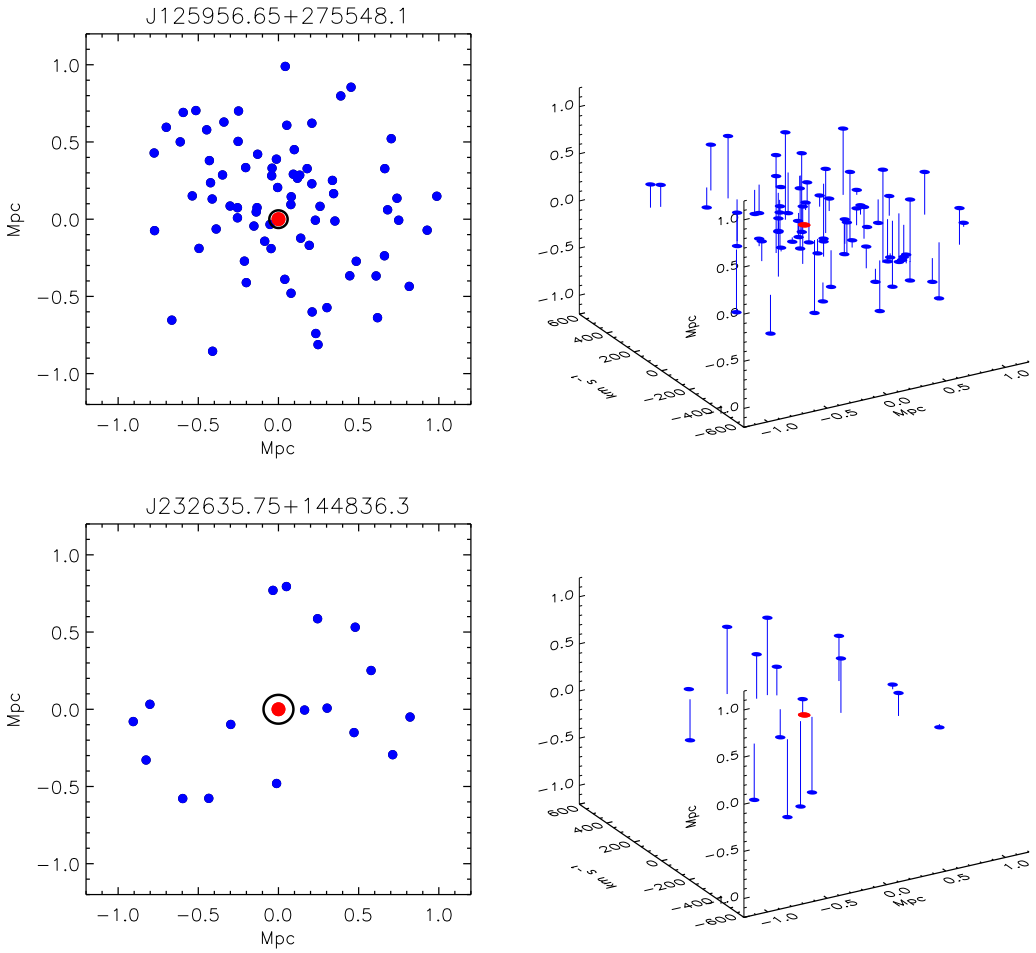


Figure 6.2: Sky distribution of two groups in our sample. *Left:* sky projection as a function of distance (in Mpc) from the central target. On the central galaxies (red dot, SDSS name on top), the size of the Arecibo beam is overplotted as a black circle. *Right:* three-dimensional view of the same group, with the redshift/velocity component added. The top row shows the richest environment in our sample, the bottom a typical one.

6.3 Specific star formation rate scaling relations

It is well known that the star-forming properties of galaxies are affected by the environment in which they are located. As reviewed in the introduction, colours are redder and specific star formation rates decrease with increasing environmental density. There is general consensus that this trend is mostly driven by an increasing fraction of red and passive galaxies. As a result, the mean specific star formation decreases when averaging over the whole galaxy population. Balogh et al. (2004) and Wetzel et al. (2012) measured no

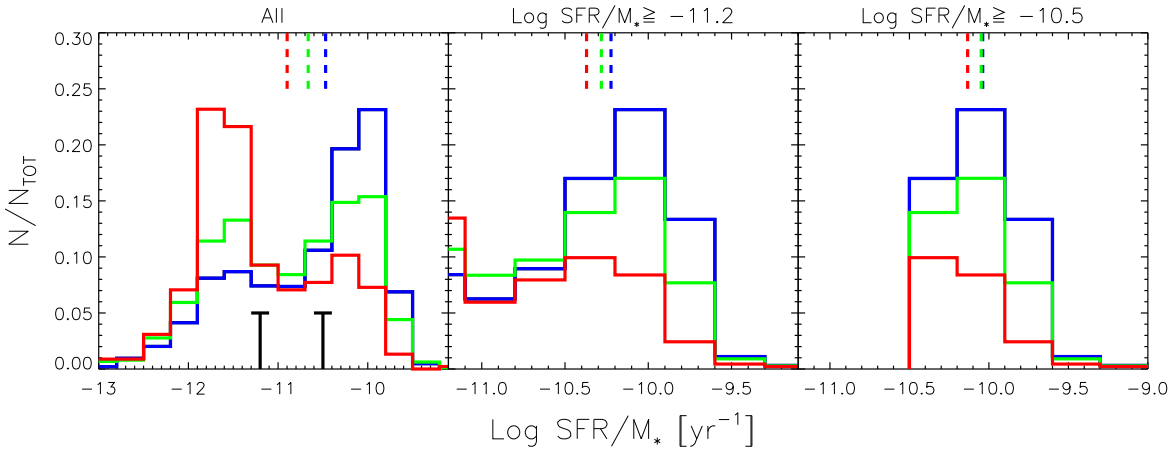


Figure 6.3: Global specific star formation rate distribution for the different density bins considered in this work. Colours as in Figure 6.1. The mean value of each distribution is indicated on top by a dashed line. *Left:* distributions for the whole galaxy population. The black lines flag Log SFR/$M_\star \geq -11.2$ and Log SFR/$M_\star \geq -10.5$, the cuts we apply in the second and third panel.

significant difference in the distribution of specific star formation rates when looking at *star-forming* galaxies in different environments. However, von der Linden et al. (2010) observed a (weak) shift toward lower values from the outskirts to the center of clusters. Understanding the star formation – density relation for active objects is important because this constrains environmental quenching time scales and mechanisms.

In our sample, we observe a decreasing star formation with increasing local density when we consider the whole population. This is visible in Figure 6.3 (left panel), which shows the normalized specific star formation rate distributions for three bins of local density (the colour coding is the same as in Figure 6.1). Each histogram exhibits a bimodal distribution, with an increasing fraction of quenched, red galaxies for higher density. The two black solid lines divide the specific star formation rate distribution into active blue galaxies (Log SFR/$M_\star \geq -10.5$), transitioning ones ($-11.2 \leq$ Log SFR/$M_\star \leq -10.5$) and passive red objects (Log SFR/$M_\star \leq -11.2$). The mean value of each histogram (measured as $\sum$SFR/$\sum M_\star$ and flagged by a dashed line) shifts toward lower specific star formation rates. The middle panel of Figure 6.3 focuses on galaxies with Log SFR/$M_\star \geq -11.2$, to include both transitioning and star forming galaxies. This cut is similar to the one applied by von der Linden et al. (2010). When we neglect the passive, already quenched galaxies, the shift in the average star formation properties becomes weaker but it is still evident. The change in mean value now measures 0.15 dex, compared to the 0.43 dex obtained for

the whole population. If we restrict our sample to the still actively star-forming galaxies (Log SFR/$M_\star \geq -10.5$, right plot of Figure 6.3), the average trend disappears for all the density bins except the last (red), where we can still observe a shift in the mean of 0.1 dex.

In the following, we will always analyse both the whole population, and the sub-sample with Log SFR/$M_\star \geq -11.2$ separately. Including all galaxies in the analysis will also include the effect of early quenching mechanisms which may have shut down the star formation before galaxies entered a dense environment (e.g., AGN feedback, mergers). The analysis of star-forming objects isolates the present-day effect of mechanisms that quench the star formation in clusters.

We will additionally disentangle the stellar mass dependence from the star formation quenching induced by environment. For each density bin, we measure the specific star formation rate (sSFR) as a function of $M_\star$. We do this both for the global specific star formation rate and for sSFR measured within the SDSS fibre aperture. The comparison between the two provides further insight into the environmental effects acting on our galaxies. An outside-in mechanism, for example, would result in the outer regions being quenched before the inner ones.

Results are presented in Figure 6.4, top row. On the left we show the global specific star formation rate, which is evaluated from SED fitting techniques using SDSS and GALEX global photometry (§2.5). On the right, the fibre sSFR acquired from the MPA DR7 database is shown (§2.4). Colours refer to the three density bins considered (see the legend in the bottom row). The vertical scale has been chosen to span the same 2.5 dex range in all the plots. The error in the mean is evaluated using bootstrap re-sampling.

We notice that the dependence of specific star formation rate on stellar mass is stronger than the one on local density. The global specific star formation rate (top left) of isolated objects decreases by $\sim$0.8 dex with increasing $M_\star$; a similar change occurs only if we move a low mass galaxy from the field to a very rich environment (red line). Massive galaxies exhibit a much weaker dependence on the number of neighbours: they are passive with low specific star formation rates when located in the field and show little change with increasing density. This result is in agreement with previous work (e.g. Kauffmann et al. 2004; Tanaka et al. 2004).

The specific star formation rate measured in the inner regions of our galaxies (Figure 6.4, top right panel) displays the weakest dependence on both $M_\star$ and local density.

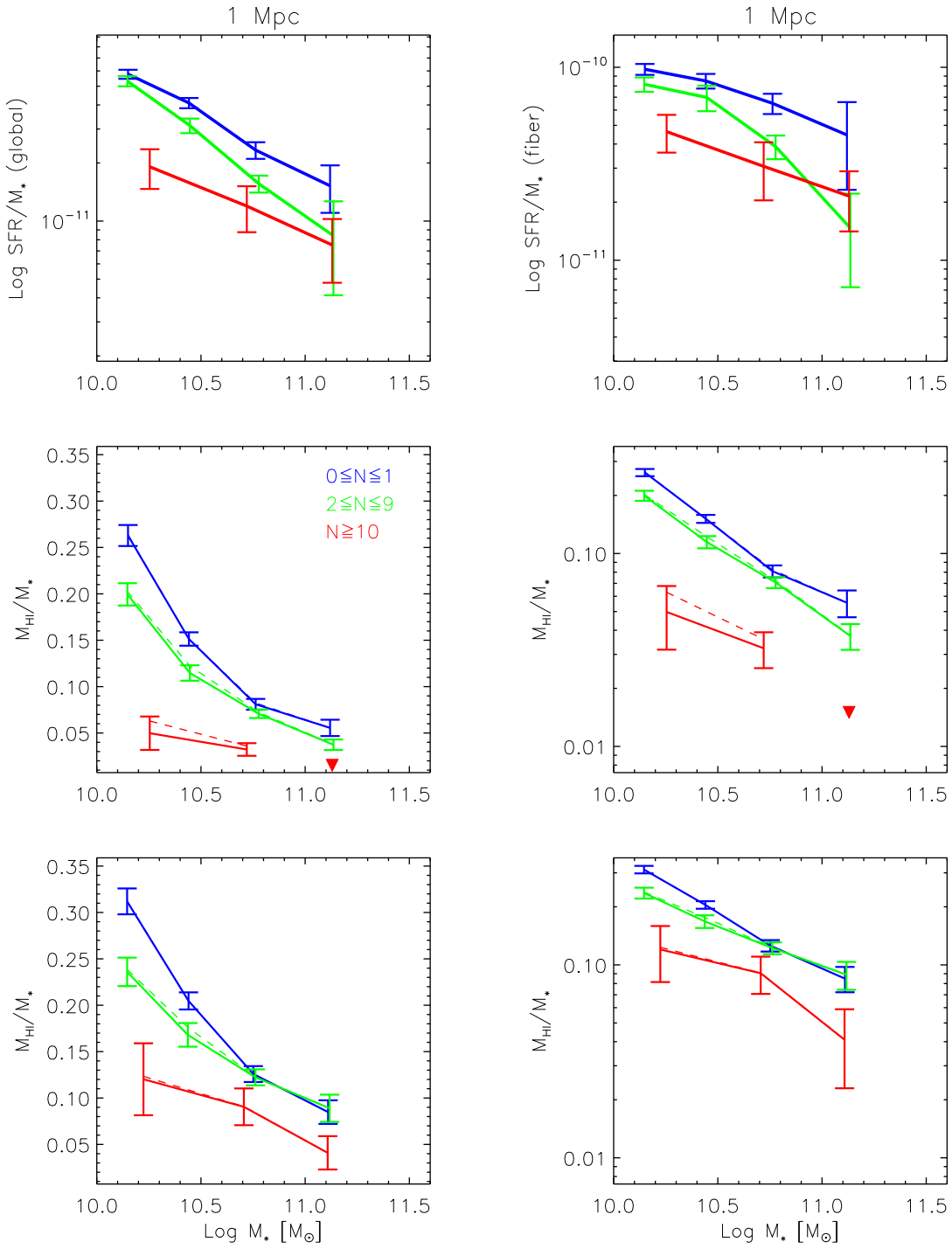


Figure 6.4: Scaling relations for the different density bins, as a function of $M_\star$. *Top row*: global (left) and fibre (right) specific star formation rate. *Middle row*: average HI gas fraction, in both logarithmic scale (right) and linear scale (left). Triangles indicate an upper limit/marginal detection. Dashed lines show the gas fractions as measured before the confusion correction is applied (see the text for details). Errors are evaluated using bootstrap re-sampling. *Bottom row:* same as the middle one, but for star forming galaxies only.

6.4 HI scaling relations

We measure how the average HI gas fraction evaluated from stacking scales with N as a function of environment. Our results are presented in the middle and bottom rows of Figure 6.4. The vertical range spans 2.5 dex as in the star formation plots, but the left panel have a linear scale which better highlights the behavior at small stellar masses. The different colours correspond to the different density bins; errors are evaluated with bootstrapping of 80% of the sample as explained in §3.2.4.

The first result is similar to that in the previous section; the HI content correlates most strongly with stellar mass. In the middle row, where we consider all galaxies, there is almost no difference between the density bins except the red one. The gas fraction decreases by about 0.75 dex across the stellar mass range, but by less than 0.2 dex in the first three density bins. A significant HI depletion is only measured for galaxies located in the densest environments ($N \geq 10$, red line).

The second result is that the effect of environment on gas content is stronger for lower mass galaxies. The linear plot (Figure 6.4, middle left) clearly shows how the HI gas fraction decreases for increasing density: $M_{HI}/M_{\star}$ drops on average from $\sim 27\%$ in isolated lower mass galaxies to $\sim 5\%$ for objects in the richest environments. This decrease is comparable to the change in HI mass fraction across the stellar mass range. The effect of environment is much smaller for more massive galaxies.

Note that at the median redshift of *sample A*, the size of the Arecibo telescope beam corresponds to physical scales of 0.15 Mpc and may include more galaxies than the targeted one. This is the case, for example, in the upper left panel of Figure 6.2. The black circle around the main target indicates the size of the beam and encloses one neighbouring companion. If the redshift/velocity difference between the target and the companion is small, the fluxes will blend together and this would result in an overestimate of the HI flux. For this reason, we apply a correction described in detail in Appendix B. In summary, for each galaxy that lies inside a region of the beam size $\pm 300\,\mathrm{km\ s^{-1}}$ around the main target, we estimate its gas fraction from photometry and subtract it from the HI mass of the target galaxy. We use the relation between $\mu_{\star}$, NUV$-r$ and HI defined by Catinella et al. (2010). In the bottom plots of Figure 6.4, the dashed lines indicate the gas fractions before correction for contamination, while the solid lines are the final corrected values. Note that the corrections are always smaller than a few percent, even in the highest density bins. Throughout the rest of the chapter, we plot only the corrected values.

Finally, in the bottom row of Figure 6.4 we show trends for the star forming galaxies

defined as galaxies with $SFR/M_\star \geq 10^{-11.2}$. The trends are weaker than the ones obtained for the whole sample, but still clearly present.

6.4.1 Comparison with previous work

Before proceeding with the analysis, we compare our gas fraction scaling relations with recent results from Cortese et al. (2011). These authors analysed a nearby sample, including members of the Virgo cluster, and studied gas fraction as a function of internal properties like $M_\star$, colour and $\mu_\star$ in a range of different environments. Their average trends are reported in Figure 6.5[1]. The different gray lines indicate: galaxies in any environment that have a normal HI content for their size and morphological type (light gray), field objects (gray) and galaxies belonging to the Virgo cluster (dark gray). If we compare our results with these relations, we find that the gas fraction of galaxies with $N \leq 10$ is comparable with the gas fraction of the "normal" objects in the Cortese et al. (2011) sample. Galaxies in our bin with $N \geq 10$ are as deficient in their HI content as the objects in Virgo. Note that Virgo is not a relaxed cluster, so it is composed by different substructures, which are still merging.

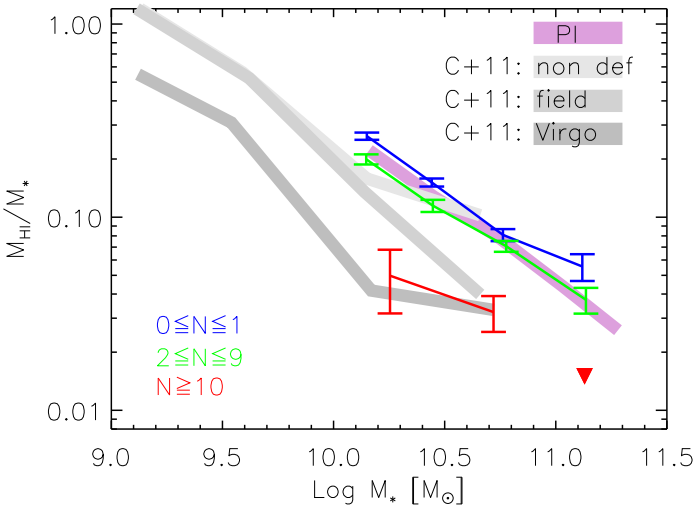


Figure 6.5: Comparison of our scaling relations with Cortese et al. 2011 (gray lines) and Figure 3.3 (purple one). See the text for detail.

[1]Here we consider the average of the individual gas fractions to compare with our our measures. In the published paper Cortese et al. (2011) consider the average of the logarithmic values.

6.5 Relative environmental effects

Qualitatively, as seen in the previous sections, the decrease in the gas content of galaxies in the richest environments is significantly stronger than the decrease in the specific star formation rate.

To properly compare the different quantities as a function of local density, we place the scaling relations for both HI and star formation in Figure 6.4 on the same scale. We divide our sample in three bins of stellar mass, and for each bin we evaluate $M_{HI}/M_\star$, global and fibre specific star formation rates. For each parameter and each mass bin, we fix the values measured for the lowest density bin ($0 \leq N \leq 1$) to 1, and scale the others to them. With this normalization, we compare the *relative* decrease of each quantity with density. In a pure starvation scenario, where there is no replenishment of the cold disk from the hot halo, we would expect the cold gas and the star formation to decrease at the same rate as a function of density. If stripping of the cold gas takes place, we expect to observe a stronger environmental dependence for the HI than for the star formation. The HI is in fact distributed at larger radii and thus has lower density than the molecular clouds where the stars are forming. Our results are presented in Figure 6.6, first row. We have actually divided galaxies into four density bins, by splitting the middle one into: $2 \leq N \leq 5$ and $6 \leq N \leq 9$. The orange lines represent the HI gas fractions, the blue ones the global specific star formation rates, and the green lines the fibre ones. We do not present results for the highest mass bin ($\mathrm{Log}\, M_\star$=[11.0; 11.5]), because statistics (∼250 objects) are too poor.

The results support an *outside-in mechanism acting on the disks* of lower mass galaxies (top left panel). The HI gas content declines more rapidly as a function of N than the star formation. Moreover, we see that the global specific star formation rate declines more rapidly than the fibre one (if we compare blue lines with green lines). When an object is in an environment with $\langle N \rangle \cong 16$, its HI content drops to almost 20% of the value measured in isolated galaxies, the global star formation to less than 40%, and the fibre one to half. This progression points toward a cold-gas-stripping mechanism. The HI is typically distributed in a disk which extends well outside the optical radius. If the external regions are stripped first, the star formation inside the disk is not immediately affected.
A similar gradually increasing effect is unfortunately not measured in the higher mass bin (upper right plot). We are probably in an environmental regime that is not dense enough to affect galaxies with $\mathrm{Log}\, M_\star \geq 10.5$, which have higher stellar (and probably) gas surface density. We would require larger numbers of neighbours than the ones sampled here to reach ICM densities able to strip the HI.

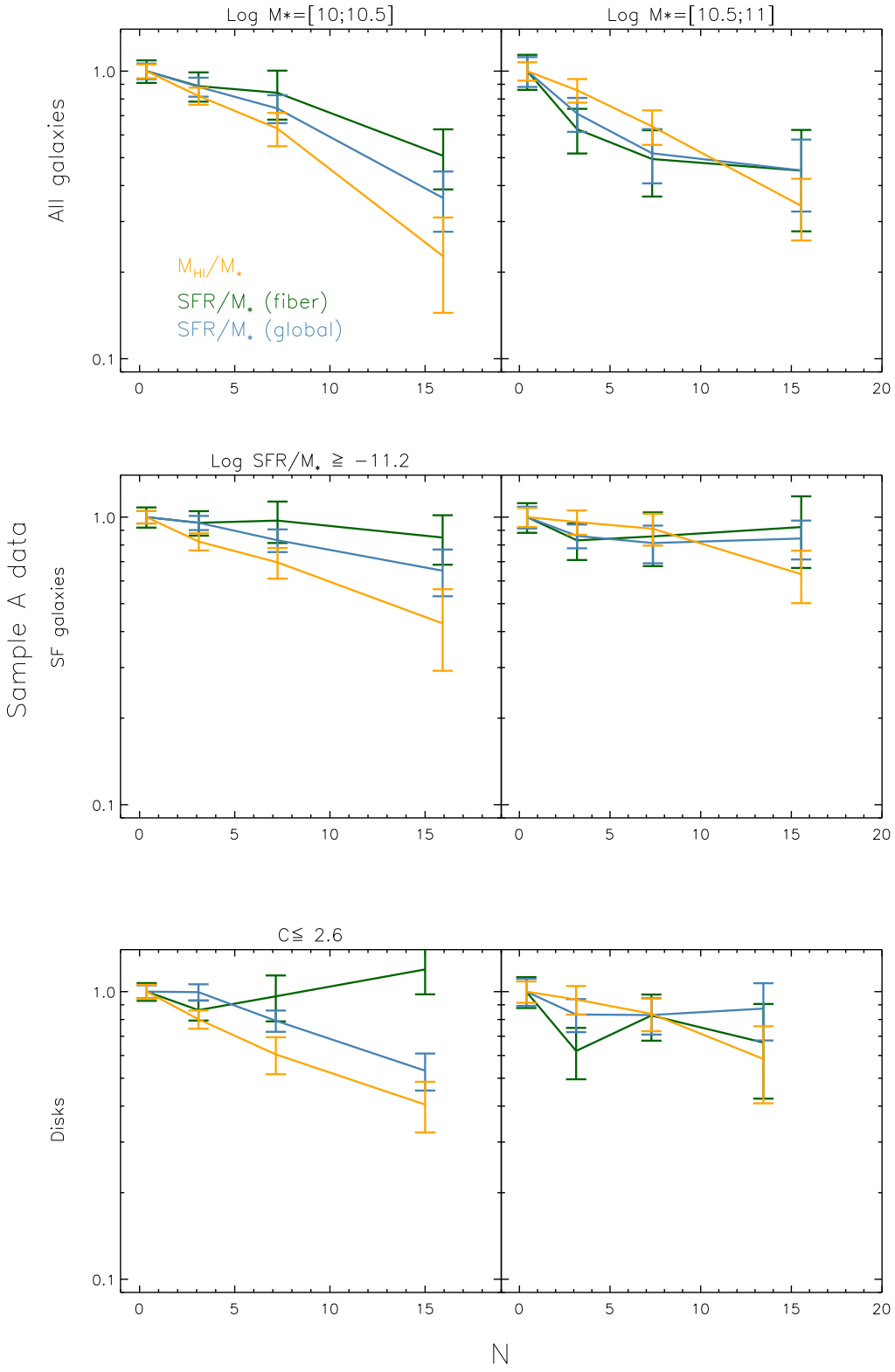


Figure 6.6: *Sample A data*: Comparison of HI gas fraction depletion (orange), global specific star formation rate (blue) and fibre specific star formation rate (green) quenching, as a function of local density for 2 bins of $M_\star$, as indicated on top. *First row*: all galaxies. *Second row*: star forming ones ($\mathrm{Log}\, SFR/M_* \geq -11.2$). *Third row*: disk-dominated galaxies ($C \leq 2.6$).

We can nevertheless test the stripping scenario as follows. An HI deficiency caused by stripping is likely to be more effective on disk-dominated, star forming galaxies. We investigate this in the second and third rows of Figure 6.6.
In the second row, we analyse only star forming galaxies with $\mathrm{SFR/M_\star} \geq 10^{-11.2}\,\mathrm{yr}^{-1}$. With this cut, we select objects of any morphology, but which have not yet been quenched. In this case, the general trends with environment are weaker than if we consider the whole population. Massive, star-forming galaxies (right panel) exhibit no correlation between star formation and HI content and the local density field. However, less massive objects do show an effect which mirrors the one just discussed for the whole population: the fibre specific star formation (green line) is insensitive to the local density, while the global one slowly decreases to ∼70% of the initial value and the HI to 50%.
The same picture appears if we focus on disk-dominated galaxies, selected to have concentration indexes smaller than 2.6, or bulge-to-total ratio smaller than 0.3[2]. The environmental quenching of the global SF, together with the depletion of the HI, is observed in galaxies with $\mathrm{Log\,M_\star} \leq 10.5$ (the specific star formation rate drops to 50% of its field value, HI to 40% of it), while no trend is measured for the fibre specific star formation rate within the errors. More massive systems do not show significant trend with the local density.

Finally we can compare the trends with N for lower mass galaxies selected by star formation (middle row, left) and "morphology" (bottom row, left), with the trends for the whole *sample A* (Figure 6.6, upper left). We find that the environmental dependence is mostly driven by disk-dominated galaxies, suggesting once again an effect which, like ram-pressure, primarily acts on the disk of our targets.

6.5.1 Comparison with models

In the introduction, we mentioned that ram-pressure stripping of the cold gas is currently not included in most semianalytic models of galaxy formation. Ideally, a successful model should be able to statistically reproduce galaxy properties at low redshift, including the differences in galaxy properties as a function of environment. We show now that current models actually fail to reproduce trends in the cold gas content in galaxies as a function of density.

In semianalytic models, gas evolution is regulated by a set of analytic prescriptions that

[2]We remind the reader that C is an indicator of the bulge-to-total ratio. See Section 4.2 for more details.

describe the physics of the baryons, such as heating, cooling, star formation, supernovae and AGN feedback. Because our understanding of the physical processes involving baryons is still incomplete, the models themselves are approximate.

Here, we use the SAMs from Guo et al. (2011), which contain an improved description of the environmental effects. The models are implemented on the Millennium II simulation (Boylan-Kolchin et al. 2009). In Guo's models, environmental processes are activated when a galaxy enters the virial radius of a more massive halo. Both tidal and ram-pressure stripping act on the *hot gas* distributed around the galaxy, which is then gradually stripped. This roughly reproduces the fraction of actively star forming galaxies as a function of projected distance from the cluster center, as shown in Figure 3 of Guo's paper. The stripping does not affect the cold gas component of galaxies. This assumption, as Guo et al. themselves discuss, is unrealistic for galaxies in the central regions of rich clusters. We now show that this implementation fails to reproduce the cold gas trends in less extreme environmental regimes.

From the results of Guo's models on the MII simulation at $z \sim 0.05$, we extract a mock catalogue of galaxies with $\mathrm{Log\,M_\star} = [10; 11.5]$. We then apply the same criteria used on our data to evaluate a local density parameter (N) for each target, counting the neighbours more massive than $\mathrm{M_\star} = 10^{9.5}\,\mathrm{M_\odot}$ inside a cylinder of 1 Mpc radius and depth $\pm 500\,\mathrm{km\,s^{-1}}$. In addition, we acquire from the catalogue the following quantities: stellar masses, cold gas masses (used as a proxy for HI), bulge masses, and star formation rates. For details on how these different properties are computed, we refer the reader to Guo et al. (2011). Here we only recall how gas content and star formation are regulated. The cold gas content of a galaxy can be supplied both by diffuse infall from the surroundings and by new material from accreted satellites, and can be depleted by SN heating or SF. The ISM is distributed in a disk, with size which scales as the product of the virial radius and the spin parameter of the corresponding host halo. The total star formation rate $\delta \mathrm{M_\star}/\delta t$ follows a simplified version of the Kennicutt-Schmidt law. It is proportional to the total mass of cold gas above a given threshold, at each timestep. This critical gas mass required for stars to form is set because star formation is believed to happen if the surface gas density exceeds a critical value given by the Toomre stability criterion. In the model, this stability criterion is a global rather than local one.

In Figure 6.7, we show the normalized local density parameter distribution and the stellar mass distribution for each density bin used in this section for our simulation catalogue. As expected, the mock data span a wider range of local densities than *sample A* data; for this comparison we restrict the analysis to $N \leq 30$. With this cut, the mean N of

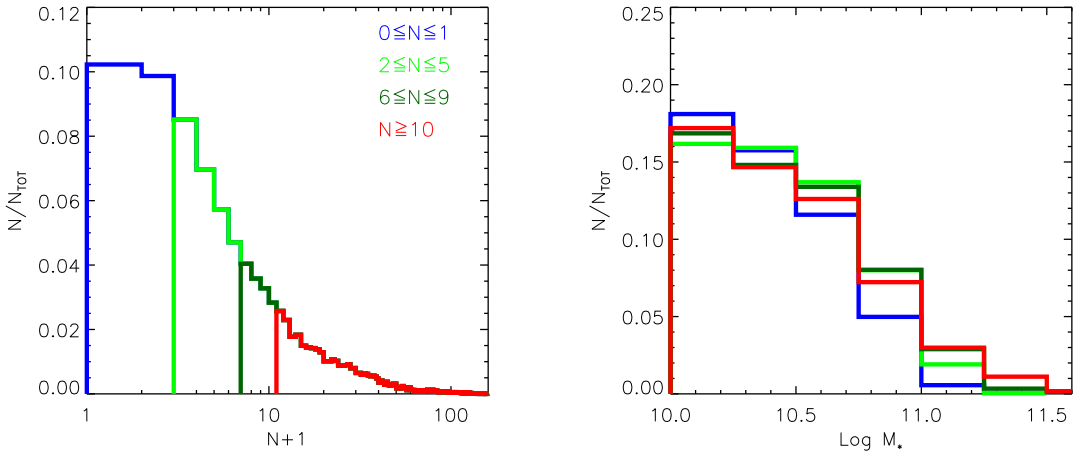


Figure 6.7: Local density and stellar mass properties for the mock catalogue. *Left:* distribution of the number density parameter N (+1 for convenience). In the mock data, we have objects located in higher density environments. *Right:* Normalized $M_\star$ distribution for each of the density bins considered.

the higher mock density bin results 17.1, similar to the mean value for *sample A* of 15.7.

In Figure 6.8 we reproduce the comparison plot between HI and star formation of Figure 6.6 using mock data. Orange lines represent the cold gas fractions, blue lines the global specific star formation rates (we do not have fibre ones). We overplot *sample A* trends from Figure 6.6 as dashed lines, to allow the reader a direct comparison.

In the upper row, we notice that mock data exhibit a similar dependence on environment in both intervals of stellar mass. This is expected because a starvation-like mechanism does not vary significantly with $M_\star$. The models reproduce the observed *average* star formation decrease with N, but the gas trends (in particular for lower mass galaxies) are too shallow: the cold gas drops to only 40-50% of the field values, as opposed to the 20-30% measured for the HI in the previous section.

We note that the apparently puzzling result that the star formation (blue line) is more significantly affected by environment than the cold gas (orange) is in fact a consequence of the way star formation is treated in the models. In galaxies where the cold gas mass has fallen below the threshold value, star formation shuts down[3]. Cold gas can linger just below the threshold. This is the reason why star formation is quenched faster than the cold gas is depleted. In Figure 6.9 we compare the specific star formation distributions of *sample A* galaxies (filled black histograms) with values from the mock catalogue (empty

[3]Excluding episodes of merger-induced starburts.

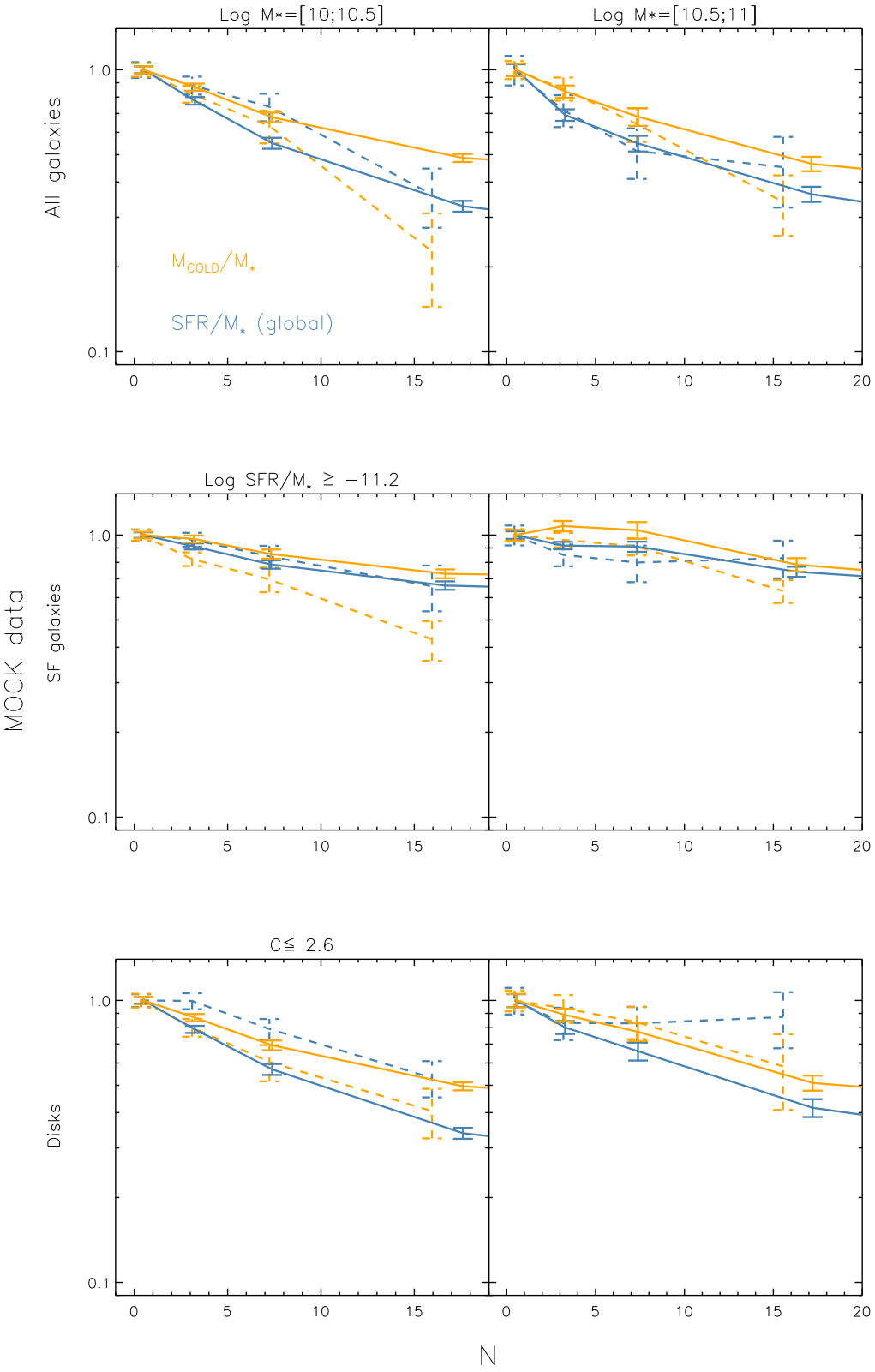


Figure 6.8: *Mock data from SAMs*: Comparison of HI gas fraction (orange) depletion, and global specific star formation rate (blue) quenching, as a function of local density for 2 bins of $M_\star$, as indicated on top. Colours and panels as in Figure 6.6. The dashed lines are the trends for *sample A* (Figure 6.6).

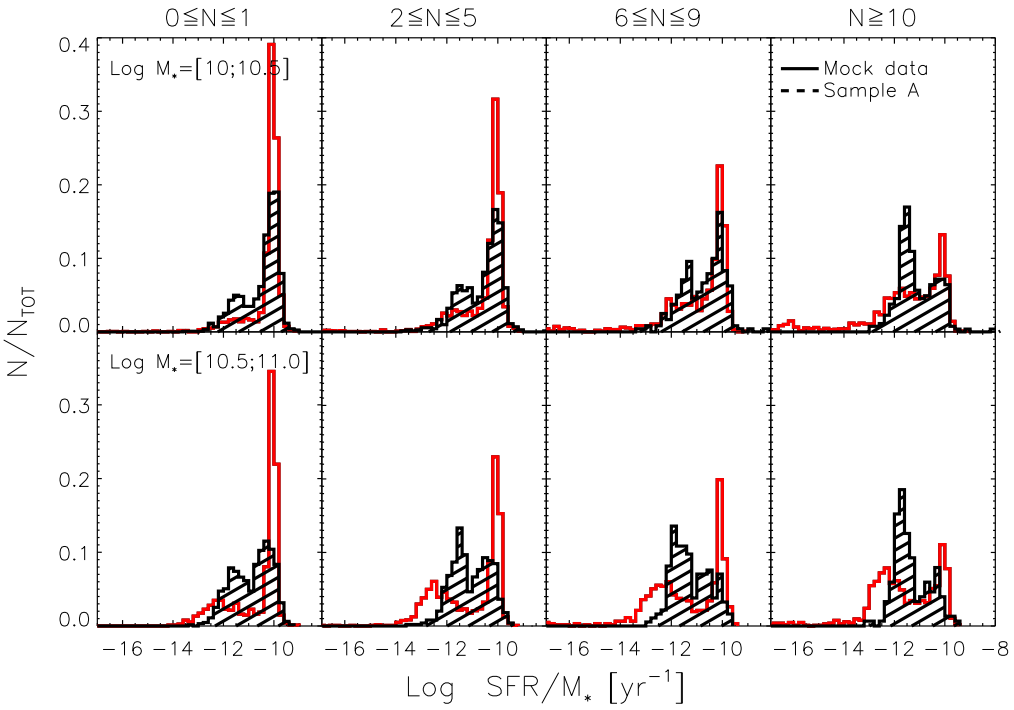


Figure 6.9: Comparison of specific star formation rates of our data (hatched histograms) and the mock ones (solid histograms). The two different rows are for the different stellar mass bins, while the bins refer to different densities (as reported on top).

red histograms) in the two $M_\star$ bins considered (top and bottom row), and for each density bins (as indicated on top). The relative behavior of real and mock data is similar, but mock data exhibit a tail of very low values down to $SFR/M_\star \sim 10^{-16} yr^{-1}$, which contributes in lowering the mean (note that observational estimates of star formation rates are not reliable in this regime; it is more meaningful to focus on star forming objects only). An opposite effect appears for the cold gas content, which is not properly depleted in the models. Most of mock galaxies (76%) have more than 10% of cold gas; in the same mass range, only 50% of the detections of the GASS representative sample are characterised by a high gas fraction (and the number decreases if we include the non-detections).

If we restrict the analysis to galaxies with $SFR/M_\star \geq 10^{-11.2}\,yr^{-1}$, for which the gas is above the star formation threshold (middle row in Figure 6.8), we see how the star formation and the cold gas are coupled by construction. The slow decrease with density is the result of the starvation mechanism: since the cold gas is not directly removed, only star formation consumes it. Without new replenishment of the cold fuel, star formation in satellites galaxies declines more rapidly than in isolated galaxies, so that the fraction of

quenched objects increases with N. The trend is weak, because we are selecting against the objects that have already been strongly affected. The model clearly does not reproduce the observed dependence of cold gas content on density in lower mass galaxies.

We can also select disk-dominated galaxies in the mock catalogue, using the cut in bulge-to-total ratio at 0.3. In this case, as seen in Figure 6.8, bottom row, the cut does not affect the global trends with environmental local density. Over the whole stellar mass range, the disk galaxies in our mock catalogue exhibit almost exactly the same trends as seen in the first row, as starvation does not preferentially act on disks. For this reason, the models once again fail to reproduce the observed trends for galaxies with $M_\star \leq 10^{10.5}\,M_\odot$, where gas is clearly depleted faster than the SF is quenched.

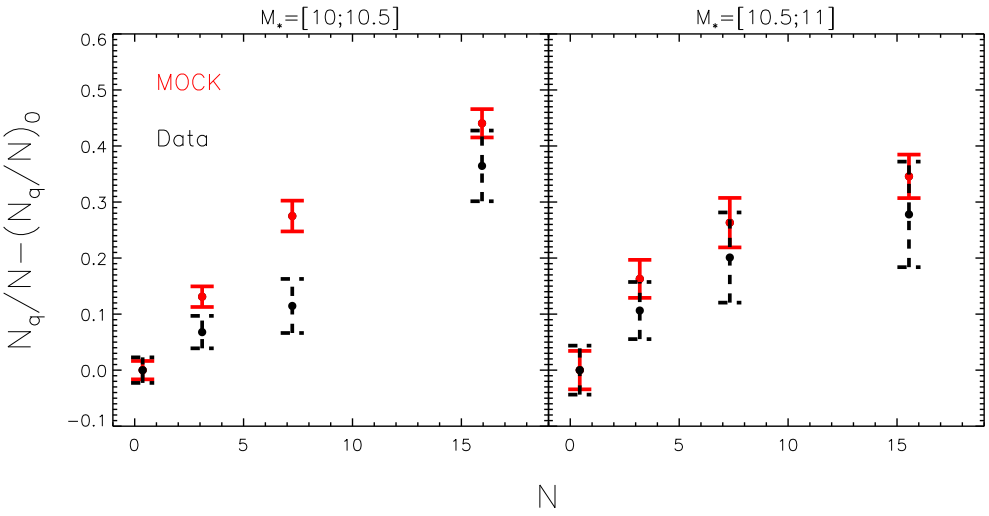


Figure 6.10: Comparison of the fraction of galaxies quenched by environmental effect (with respect to the isolated objects) for our data (black, hatched lines) and the mock catalogue (solid red ones). The two different panels refer to the different stellar mass bins, as reported on top.

A robust comparison can also be performed between the fraction of quenched objects with $SFR/M_\star \leq 10^{-11.2}\,yr^{-1}$. Although in each bin the fraction of quenched galaxies in the data is higher than in the models (Figure 6.9), models and observations agree reasonably well in terms of the trend with density. In Figure 6.10, we compare the quenched fraction N/N_q of each density bin to the one of isolated objects $(N/N_q)_0$, at fixed stellar mass, in both real (black) and mock data (red points). The values of N/N_q-$(N/N_q)_0$ are consistent, and, as expected, increase with increasing local density.

6.6 Conclusions

It has been suggested that the main interactions that shape the properties of galaxies in dense environments are the ones between the ISM (both cold and hot) of the galaxy and the ICM of its surrounding halo. In particular, both ram pressure stripping of the cold HI disk and starvation from removal of the hot gas have some observational support. Ram pressure stripping of HI gas is clearly observed in galaxies in the central regions of rich clusters; quenching of star formation on longer timescales and in lower environmental density is suggested by optical observations and supported by general agreement of star formation trends with environment in semianalytic models, which include only hot gas removal. Stripping has in fact never before been shown to be effective for the average galaxy in groups because observations of HI gas have focused only on the field/cluster dichotomy.

In this chapter, we have used a complete, volume limited sample of nearby galaxies with $M_\star > 10^{10} M_\odot$ for which we have both star formation data and coverage by the ALFALFA survey to compare how the average HI content and specific star formation rates depend on local density (once we remove the dependence on stellar mass). *For the first time*, we have been able to trace the near HI gas fraction as a continuous function of environment, filling the observational gap between field and rich clusters. Our results, part of which simply confirm previous findings, can be summarized as follows.

- The properties of galaxies (HI and star formation) depend to first order on their stellar mass. However, a residual dependence on environment as traced by the local density field is measured. This environmental quenching/ gas depletion is weaker for more massive galaxies, and for the star formation measured inside the SDSS fibre. The difference observed among parameters that trace different galactic scales (like the nuclear and global star formation, and the integrated HI) already points towards a mechanism that acts from the outside in.

- The HI gas fraction exhibits the strongest depletion for galaxies with $M_\star < 10^{10.5} M_\odot$ located in environments with more than 10 $M_\star \geq 10^{9.5} M_\odot$ members within a volume of 1 Mpc, ± 500 km s^{-1} in depth. These values may represent a threshold in halo mass, above which stronger environmental effects become dominant.

- When we compare the *relative* decrease of HI gas fraction and star formation on global and inner scales, we see clear signatures of gas stripping. Ram-pressure stripping of the cold gas in the disk from the external regions of galaxies with $10^{10} \leq M_\star[M_\odot] \leq$

$10^{10.5}$ can explain the gradual increase in the suppression of star formation from the inner to the outer regions, and the even stronger HI deficiency in these systems.

- If we focus on star-forming galaxies only, the trends of gas fraction with environment weaken but do not disappear. By discarding passive galaxies, we: *i)* discard systems which may have been quenched in the past by internal processes, and *ii)* focus on galaxies that are currently in the process of being transformed by the environment. This implies that (some of the) galaxies that are still star-forming have already been depleted of their cold gas reservoir.

- A pure starvation scenario cannot reproduce the observed trends of star formation and cold gas with environment. We used a mock catalogue extracted from Guo et al. (2011) SAMs applied onto the Millennium II simulation to show how current models underestimate the environmental effects, especially on the cold gas component of the galaxy.

The analysis presented in this chapter is only statistical in nature. We do not have resolved information on the spatial extent of the atomic gas for our targets, which would more clearly constrain the stripping efficiency. Nor can we estimate timescales for the stripping events. When stacking, it is of course not possible to distinguish between the effects of starvation or ram-pressure mechanisms in individual galaxies, and they may indeed both play a significant role in galaxy evolution.

Nonetheless, we have shown how on average depletion of the cold gas happens *before* star formation is quenched, and this is seen outside the central regions of rich clusters. Future models of galaxy formation should include ram-pressure effects on the cold gas. From the observational side, we still need to improve our analysis of the environmentally driven evolution, especially if it has to be applied as a constraint on the models. Some of the still open issues can be addressed soon. How does the local density parameter we used correlate with the group halo size? In the group catalogue for SDSS DR4 generated by Yang et al. (2007), for example, groups with more than 10 massive neighbours have an estimated halo mass greater than $10^{13}M_{\odot}$. An implementation of friend-of-friend group finder onto the latest SDSS data release, combined with the full ALFALFA data set (to be completed in the next few years), would allow us to extend this analysis to the HI properties of galaxies as a function of dark matter halo mass. In addition, with better statistics we could measure the decrease in HI for increasing group/clustercentric distance; this would put stronger constraints on ram-pressure stripping mechanisms.

Finally, not only better statistics, but also better resolution is necessary to gain additional insights into the physical processes acting in different environments. These issues will be hopefully solved in the next decade thanks to the upcoming HI surveys on the next generation of radio telescopes. The all-sky surveys planned at the Westerbork telescope and at ASKAP (WALLABY), will scan the sky in the 21 centimeter line with much better sensitivity and resolution.

7

Summary and future prospects

Massive and passive galaxies are a challenge to our current picture of galaxy evolution. Observations show that galaxy evolution produces a well defined bimodality in the colours of galaxies, and that in the local Universe the red sequence is dominated by spheroidal, massive systems with high stellar mass surface densities and on average low cold gas content. Despite the increasing observational constraints, theoretical models still have difficulties in reproducing the low star formation properties of massive galaxies. To date, two main classes of mechanisms have been proposed to explain the quenching of the star formation in these systems; however, in their current implementation, they are not sufficient to properly halt the activity in massive galaxies. The first class comprises environmental mechanisms, like removal of hot gas and the subsequent shut off of gas cooling in galaxies once they fall into a larger halo. The second class includes mechanisms that act in isolated galaxies by either preventing gas cooling onto a potentially star forming disk, or via direct depletion of the cold interstellar medium.

In this thesis, we studied the HI content of nearby massive galaxies to address some of the open issues in the quenching processes and to constrain their efficiency in shaping the observed bimodality. This final chapter summarizes the main results of this work (Section 7.1) and puts them into context with the current picture of galaxy evolution, presenting questions that remained open and how they can be addressed in the future (Section 7.2).

7.1 Main results

We selected a volume-limited sample of ~ 5000 galaxies to have stellar masses larger than $10^{10}\mathrm{M}_{\odot}$, redshifts in the range $0.025 < z < 0.05$, and to lie in the intersection of the

ALFALFA HI, SDSS optical, and GALEX ultraviolet surveys. These selection criteria provided us with a statistically significant, multiwavelength sample with homogeneous estimators for different galaxy components. To exploit data from ALFALFA, which detects only 20% of the massive systems at the redshifts considered, we developed a software tool to co-add or "stack" its data.

Morphological quenching – *Chapter 4*
Using stacking, we analysed the properties of 1833 bulge-dominated galaxies selected from our main sample to be a proxy for early-type objects. We confirmed that the HI gas fraction of both massive early-type and late-type galaxies correlates *primarily* with NUV-r colour and stellar mass surface density ($\mu_\star$). The main result of our analysis is that the relation between average HI gas fraction and the two parameters is *independent* of the bulge-to-disk ratio of the galaxy (§ 4.3). In addition, we found no evidence that red-sequence galaxies with a significant bulge component are less efficient than disk galaxies in turning their available gas reservoirs into stars, as suggested by some simulations that produce early-type galaxies with non star-forming HI disks (§ 4.4). We actually observed that the *disk* gas fraction ($M_{HI}/M_{\star;DISK}$) is independent of galaxy morphology at fixed stellar mass, in agreement with the idea that the cold gas is mainly associated with the disk itself (§ 4.5).

AGN feedback on the gas content of massive galaxies – *Chapter 5*
To investigate the AGN - atomic gas connection, we compared 1871 AGN host galaxies with a control sample of non-AGN objects, matched to the former in NUV$-r$ colour and $\mu_\star$ (since $M_{HI}/M_\star$ in massive galaxies correlates strongest with these two properties). We found no significant differences in HI content between AGN and control galaxies at all values of black hole accretion rate probed by the galaxies in our sample. This indicates that AGN do not influence the global gaseous properties of galaxies, at least in the local Universe. In addition, accretion itself is not sensitive to the global properties of the interstellar medium in blue, gas-rich galaxies; on the contrary, in the red population, accretion rate and gas fraction do decrease together (§ 5.3). The measured gas fractions for this population are not too different from the ones expected from a stellar mass loss origin, implying that the fuel supply in the red AGN population could be a mixture of mass loss from stars and residual gas in the disk (§ 5.5).

We could not yet unequivocally ascertain whether AGN feedback is important in the central regions of galaxies. Since the molecular hydrogen phase (H_2) is generally more centrally concentrated than the atomic one and may be more sensitive to feedback processes occurring in bulges, we examined trends in H_2 gas fractions for AGN hosts. A first analysis

on a small sample from the second data release of the COLD GASS survey (Saintonge et al 2011) shows encouraging gas depletion signatures (§ 5.4). A more detailed study with better statistics is needed.

Environmental effects on HI and star formation properties of massive galaxies – *Chapter 6*
For the first time, we studied the HI content of massive galaxies as a continuous function of local density, filling the observational gap currently existing between fields and clusters. By comparing the effects of environment at different density scales on global and inner specific SFR (§ 6.3) and HI content (§ 6.4), we showed clear signs of cold gas stripping in galaxies with $10 < \mathrm{Log}\, M_* [M_\odot] < 10.5$, outside clusters. Once the first order dependence of galaxy properties on stellar mass had been removed, in fact, we measured a gradual increase in the suppression of star formation from the inner to the outer regions, and an even stronger HI deficiency as a function of local density. All these effects can be explained by a mechanism acting on the disk from the outside-in, like the ram-pressure of the cold ISM (§ 6.5). Finally, a comparison with mock catalogs extracted from one of the most recent semi-analytic models (Guo et al. 2011), which includes only starvation through removal of the hot gas, showed how models underestimate the environmental effects, especially on the cold gas component of galaxies (§ 6.5).

7.2 Outlook

In our work, we investigated the connection between atomic hydrogen and star formation with the purpose of testing quenching mechanisms, using for the first time homogeneous data for a large sample not biased towards any morphology or star formation regime. As already underlined in the Introduction, this work is statistical in nature and does not aim at ruling out any of the suggested mechanisms acting in individual galaxies. Rather, we wanted to understand their "statistical efficiency" in producing the observed bimodality of massive galaxies in the local Universe. We proved that bulge-dominated galaxies are not less efficient in forming stars than disk-dominated ones: once a galaxy has a gas disk component, star formation will eventually take place in it. Therefore, in isolated galaxies there must be additional mechanisms at work which heat or deplete the cold gas, and/or prevent hot gas from cooling onto a disk. The main mechanism currently suggested is feedback from AGN, in both its radio (mechanical) and quasar (radiative) mode. Our results show no signature of negative feedback on the atomic reservoir of

massive galaxies, but rather suggest a co-evolution of AGN and star formation, both fed by the gas available. Observations of galaxy colours becoming redder when galaxies host an AGN (e.g. Kauffmann et al. 2003a; Heckman et al. 2004; Schawinski et al. 2007), which have been considered as supporting evidence of an AGN-driven galaxy evolution, could be explained by the following: when a galaxy starts to consume its atomic reservoir, there is no material for either accretion onto black hole or conversion into molecular gas and subsequent star formation. Of course, we cannot and we do not exclude that radiative feedback is efficient at higher redshifts, when AGN were more efficiently accreting. And we do not exclude that the radio-mode, for which we do not have a significant sample in our data, plays a major role. There is some evidence of atomic and molecular hydrogen outflows driven by radio jets at low and high redshifts (Nesvadba et al. 2008; Guillard et al. 2012), or quasar feedback through AGN winds at high redshifts (Feruglio et al. 2010), but to date they are based on a very limited number of galaxies so that their incidence is not constrained.

While we did not observe both bulges and AGN to have a significant influence on the global cold gas content of massive galaxies, we actually proved that the environment plays a major role in quenching their star formation. In fact, we measured signatures of atomic hydrogen stripping in intermediate mass galaxies located in environments less dense than clusters. Environmental effects on real galaxies are actually stronger than currently assumed in models: we clearly showed that models that include only hot gas removal are not able to reproduce observations. We therefore suggest that, in order to improve our understanding of galaxy evolution, models should implement additional environmental effects on the cold ISM of galaxies, which could solve part of the puzzle in the formation of massive and passive systems.

7.2.1 Future prospects

We showed how the study of the atomic gas phase in massive galaxies on statistically significant, unbiased samples is necessary and effective to properly constrain the dominant mechanisms creating the observed galaxy bimodality. The most promising mechanisms, which both models and observations should investigate, are environmental ones. From the observational side there are in fact some aspects that need to be better investigated. First, we need to extend the mass range considered in this work to lower mass galaxies: if we are observing signatures of cold gas stripping, we should measure an even stronger effect on less massive galaxies, which are on average gas richer and in which the gas is less bound. Second, we should increase the statistics for the more massive systems and expand the

local density range studied, in order to identify the threshold in both $M_\star$ and environment at which stripping of the cold gas starts/ceases to be effective. For this latter investigation, a larger volume is necessary. The paucity of massive systems (Log $M_\star > 10^{11}\,M_\odot$) in our sample is in fact a consequence of cosmic variance; the small number of large groups and the lack of clusters in our sample is also a consequence of the limited volume sampled. Finally, spatially resolved data showing the outside-in stripping effect on atomic gas which is suggested by our trends would represent the "smocking gun" evidence for our conclusions. To improve the analysis performed here, therefore, new data and samples should fulfill the following characteristics. *(i)* Observations should spatially resolve the gas distribution. *(ii)* Samples should include data about additional hydrogen phases, first of all the one directly coupled with star formation on short time scales, i.e. H_2. This will further clarify the connection ISM-star formation in massive galaxies, as HI constitutes only the first phase in the process leading to star formation in galaxies. *(iii)* Samples should cover larger volumes over broader redshift ranges. Quantifying how the relative importance of the studied processes and their efficiency evolves with redshift is key to understand the main drivers of galaxy evolution. Recent studies have started to assemble statistical samples (Tacconi et al. 2010, Förster Schreiber et al. 2011) to constrain the dependence of star formation on molecular gas content and how this dependence varies with environment at intermediate to high redshift ($1 < z < 3$), but samples are still small, biased toward the luminous end of the galaxy population and without HI data available.

In the next decade, new radio facilities are going to see the light, and fulfill the listed requirements. New telescopes and new receivers are under development and are specifically designed to achieve the characteristics required for the undertaking of major HI surveys, with better sensitivity and angular resolution, and larger redshift ranges probed. The exploitation of new technologies has already started on the existing facilities, like the VLA telescope which is currently being upgraded to the new Expanded VLA (EVLA[1]). It will be particularly useful to spatially resolve the HI in galaxies in environments from groups to clusters up to $z \sim 0.53$, and to map the intragroup medium. Similarly, the Westerbork Synthesis Radio Telescope (WRST) is going to be equipped with APERTIF (Verheijen et al. 2008), a new focal plane array which provides several beams on the sky increasing the survey speed of the telescope by a factor of 20. One of the major programme planned is an HI survey of the northern sky that overlaps with SDSS out to $z \sim 0.15$, to study the evolution of the gaseous content of $\sim 2 \times 10^5$ galaxies and its role in galaxy evolution. With such survey, we will be able to compare the efficiency of both local and

[1]http://www.aoc.nrao.edu/evla

global environments of galaxies in regulating their gas content. Its data could detect, thus confirm, stripping signatures on the atomic gas. We could study the role of HI when galaxies evolve from the blue cloud to the red sequence, and which is the gas content of massive, red galaxies at earlier stage of their evolution than the ones probed here. In the far future, the Square Kilometer Array (SKA[2]) will be an aperture synthesis radio telescope with a collecting area of up to one million square meters spread over at least 3000 km. SKA science cases are similar to the ones addressed by APERTIF, but the telescope will push observations out to z ∼3, performing HI line surveys of a billion galaxies. In the meanwhile, two of its pathfinders, namely MeerKAT (Booth et al. 2009) and the Australian Square Kilometre Array Pathfinder (ASKAP Johnston et al. 2009), are under construction to test new technologies.

From the molecular gas side, the Atacama Large Millimeter/Submillimeter Array (ALMA[3]) should see complete light in 2013, counting 66 antennas operating at wavelengths of 0.3 to 9.6 mm (thus suitable for indirect measurements of H_2). Although the telescope is not designed as a survey instrument, it will allow the study of the star formation - ISM - galaxy connection out to $z \sim 3$. In addition, ALMA resolution will be for example suitable to study the dynamics of the molecular gas in the few tens of parsec around accreting AGN, thus giving better insight into their physics.

To conclude, upcoming data will be of primary importance to expand the analysis of this thesis, because they will show how and when gas is converted into stars, and with which efficiency, up to the epoch when massive galaxies had the bulk of their star formation. At the same time, our stacking approach will be crucial for the exploitation of upcoming surveys. The sensitivity limits that we face now in the local Universe, in fact, will appear in deeper data at higher redshift, where the numbers of individual detections will be limited by flux dilution.

[2]http://www.skatelescope.org/
[3]http://www.almaobservatory.org/

A

Further Stacking tests

In this Appendix we present further tests that we performed before and after the development of the stacking tool in order to verify the reliability of the whole process.

In the first part (A.1), we describe a simple simulation which we completed before building the actual stacking tool. The goal was to mimic the same steps that we would then have to apply to the ALFALFA data, but implemented for a sample of model HI spectra for which we knew a priori both noises and signals. In the second part (A.2), we present the comparison between two different possible stacking procedures, to support the choice of the one used in this work. We co-add "gas-fraction" spectra, as explained in Section 3.2.3, but here we show that different approaches lead to compatible results.

A.1 Stacking of simulated HI spectra

In order to simulate the stacking of HI spectra, we first have to create a model of such a spectrum and generate a catalogue of them for a range of realistic galaxy properties. We describe here how we achieved this, by presenting: how we modeled the HI emission from simple assumptions on the cold gas distribution and kinematics (§A.1.1); how we generated a catalogue of realistic (non-detected) HI signals (§A.1.2); how we stacked them to recover the average signal we had in input (§A.1.3).

A.1.1 Model of an HI spectrum

The neutral gas in normal disk galaxies is usually distributed in a rotating disk. The HI emits in its rest frame at 21 cm, but the emission will be measured by an observer as

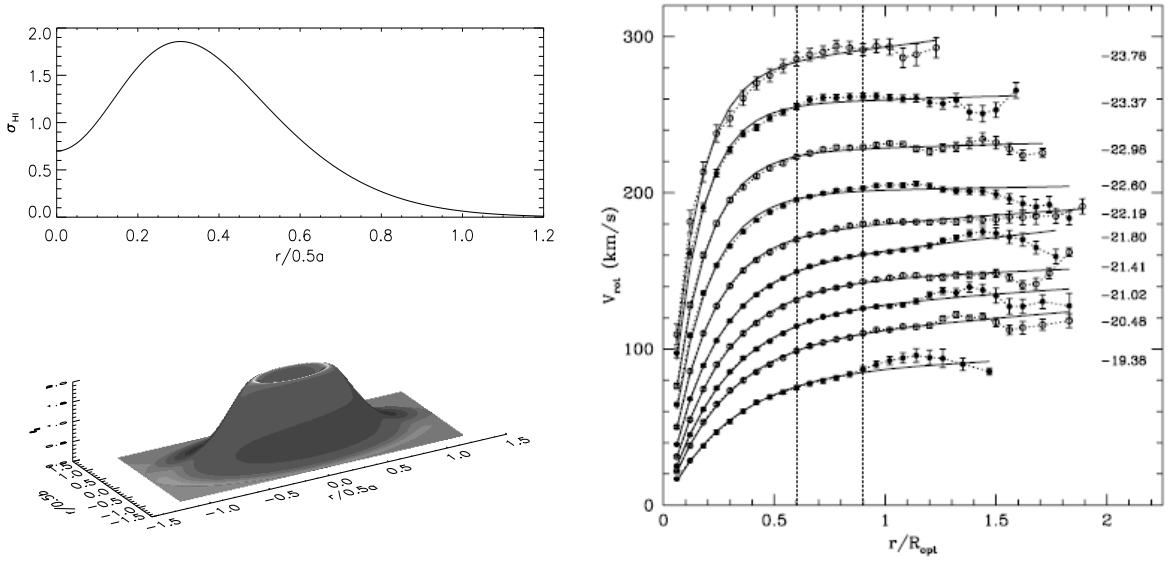


Figure A.1: *Left:* Model of the double Gaussian profile for the HI surface density distribution in galaxies, following Hewitt et al. (1983). The horizontal scales are normalized to the optical major and minor axis. *Right:* Figure 4 from Catinella et al. (2006). Template RCs for 10 bins of I-magnitude, parametrized as a function of optical radius. Each curve is labeled on the right by its mean I-band absolute magnitude.

Doppler shifted by both the galaxy recessional velocity and by the rotation of the galaxy itself.

An HI spectrum is then a measure of the HI flux intensity [Jy] per unit velocity at which the signal is observed [$\mathrm{km\,s^{-1}}$], over a given velocity range. To generate it, we need to model the HI distribution in a galaxy, assuming a surface density which gives us information about the quantity of emitting gas located at each position inside the galaxy, and a velocity field which tracks the movement of each of the HI "points". Below, we describe the different assumptions we used.

HI surface density. Observations show that in normal disk galaxies the HI is usually distributed in a disk which extends to larger radii than the optical one. In addition, its distribution often exhibits a central depression in the densest regions (e.g. Hewitt et al. 1983) where the hydrogen is likely to be in a molecular phase.

We start from the simplest assumption of an optically thin HI disk, and use a radial HI surface density profile $\sigma_{HI}(r)$ from Hewitt et al. (1983):

$$\sigma_{HI}(r) = 3\exp\left(-\frac{r^2}{R_0^2}\right) - 1.8\exp\left(-\frac{r^2}{0.23R_0^2}\right). \tag{A.1}$$

This distribution is the sum of two Gaussians, a wide positive one and a narrow negative one at the center to model the observed central depression. R_0 sets the scale of the HI distribution, while the profile parameters have been evaluated by Hewitt et al. (1983) from real data. They measured the HI fluxes along the major axis of 52 nearby galaxies, and fitted a convolution of the beam pattern with a double Gaussian profile to the observations. Since they found that the cold gas distribution is on average about 1.2 times larger than the optical blue diameter, we set the HI cutoff radius (R_0, where the density falls to zero) at the same distance.

The shape of the modeled distribution is shown in Figure A.1 (left column). The top panel is the analytic relation between $\sigma_{HI}(r)$ and r, normalized to half of the galaxy major axis $a/2$. On the bottom panel, we plot the 2-dimensional distribution, where now the two horizontal radii are scaled respectively to the major and minor axis of the galaxy.

HI kinematics. We use again a simple assumption: the HI is distributed in a regularly rotating disk. The complete analytic description of the observed 2-dimensional velocity field of a rotating disk as a function of the distance from the center of the galaxy (r) and the angle with the major axis (ϕ), would be:

$$v(r,\phi) = v_0 + v(r)\sin i\cos\phi + w(r,\phi)\sin i\sin\phi + z(r,\phi)\cos i, \tag{A.2}$$

where: v_0 is the systemic velocity of the object; $v(r)$ is the pure rotational velocity component; w is the radial velocity in the plane of the galaxy; z is the velocity out of the plane of the galaxy. All the velocity components of the disk need to be projected along the line-of-sight (through the factor $\cos i$, where i is the inclination) to obtain the observed field.

For the purpose of this exercise, we consider a galaxy in its rest frame ($v_0 = 0$) and neglect all terms but the pure rotational velocity $v(r)$. To model this component, we assume the template rotation curves (RCs, i.e. the measure of the rotational velocity as a function of radius) for disk galaxies derived by Catinella et al. (2006). In their paper, they created template rotational velocity curves by binning a sample of ~ 2200 local disk galaxies in 10 I-band luminosity intervals and computing the average RC in each bin. They fitted the template RCs with a Polyex model (Giovanelli & Haynes 2002), and obtained a model for $v(r)$ as function of r and I-band luminosity:

$$v_I(r) = v_{0;I} \cdot \left(1 - e^{-\frac{r}{r_I}}\right)\left(1 + \frac{\alpha_I r}{r_I}\right), \tag{A.3}$$

where $v_{0;I}$, r_I and α_I are the amplitude, the exponential scale of the inner region and the slope of outer part of the RC respectively, and depend on the I-band luminosity bin considered. The templates are shown in Figure A.1, right panel: in the plot, a rotation curve is displayed for each of the intervals considered. Catinella et al. (2006) results shows that the more massive systems have steeper initial velocity rises and flatter outer slopes than the smaller ones.

As explained, we considered only a pure rotational disk to model the velocity field of our objects. In order to obtain a more realistic description, though, we add at the end a dispersion component. HI disks are characterise by small motions in the direction perpendicular to the disk with $\sigma < 10\,\mathrm{km\ s^{-1}}$ (e.g. Schulman et al. 1997), so we choose to extract a σ from a randomly generated normal distribution of velocities with standard deviation of $6\,\mathrm{km\,s^{-1}}$ and mean $6\,\mathrm{km\,s^{-1}}$ (excluding negative values).

From the two assumptions above - analytic profiles for both the HI surface density distribution and the velocity field - we can simulate an observed HI spectrum. Since this is the measure of the HI flux intensity per unit velocity at which the signal is observed, we have to map each point of the HI distribution with its correspondent observed velocity, and then sum HI points with similar velocity which will emit at similar frequency.

We map the 2-dimensional HI distribution (modeled as in Figure A.1, left) with a 1000×1000 grid, and we assign a coordinate to each point $(r;\phi)$. For each grid bin, we then extract the corresponding $\sigma_{HI}(r)$ and the component of the velocity along the line of sight $v(r,\phi) = v(r)\sin i\cos\phi$. The spectrum is just the sum of the HI densities of the spatial grid bins that "emit" in a given velocity interval, for all the velocities. We use a velocity resolution of 1 km $\mathrm{s^{-1}}$ to extract the spectrum, and the smooth to 11 km $\mathrm{s^{-1}}$ to mimic ALFALFA observations.

We notice that the physical units of the HI flux are Jy, but we are summing densities. We should convert the surface density into a flux density, but for the purpose of this simulation, considering that the flux scales linearly with the density, we just use arbitrary units.

In Figure A.2 (top row) we show two examples of simulated HI spectra, for two different inclinations (and randomly generated input parameters; see the next Section). The two spectra shown are obtained with a higher grid resolution (6000×6000) only for illustrative purposes. When the object is (almost) face-on as in the right panel,

the emission profile is a Gaussian broaden by the velocity dispersion only, because there is (almost) no rotational velocity *observed*. If we had not added this velocity dispersion component, the line in the face-on case would be an unrealistic spike at velocity $0\,\mathrm{km\,s^{-1}}$. When the inclination is larger, as in the left panel, the observed line displays the characteristic double-horn profile caused by the rotation which Doppler shifts the emitted signal.

The last component necessary to model a realistic spectrum is a noise over the velocity channels. Real observations are of course characterised by a background noise which can be as strong as (or even stronger than) the galaxy emission itself. We consider only a Gaussian component of constant *rms* across the velocity range, which we randomly generate for each channel. In Figure A.2, second and third rows, the same examples of simulated spectra as before are displayed, but now a background noise has been added. The bottom row has been smoothed to 11 km $\mathrm{s^{-1}}$ resolution to mimic the ALFALFA velocity resolution.

A.1.2 Catalogue of non detections

We want to create a catalogue of many realistic spectra extracted from a range of galaxy properties, in particular with different HI signal intensities and widths. As explained in the previous section, the input parameters that we need to model our spectra are: major (a) and minor (b) axis, from which we set the scale of the HI distribution and estimate the inclination; an I-band magnitude (m_I) from which we extract the parameters ($v_{0;I}$, r_I, α_I) which characterise the rotation curves; an *rms* as noise measurement.

We define a range of possible values for the main parameters, namely: m_I in the range [-24;-18], a and b in the range [5;20] kpc, the *rms* in the interval [0.5;2] (set by considerations on the flux density scales that we typically measure). We use a random generator to extract a set of values for each spectrum we want to model. In addition, once we have an I-band magnitude, we can assign $v_{0;I}$, r_I, α_I. We use the parameters and their errors for each m_I bin from Catinella et al. (2006); we generate a normal distribution of possible values centered at the parameter and with standard deviation equal to its error, and extract one value per object.

With all the assumptions described, we generate a catalogue of 1500 spectra spanning several combinations of the parameters in input. Since we are interested in stacking non

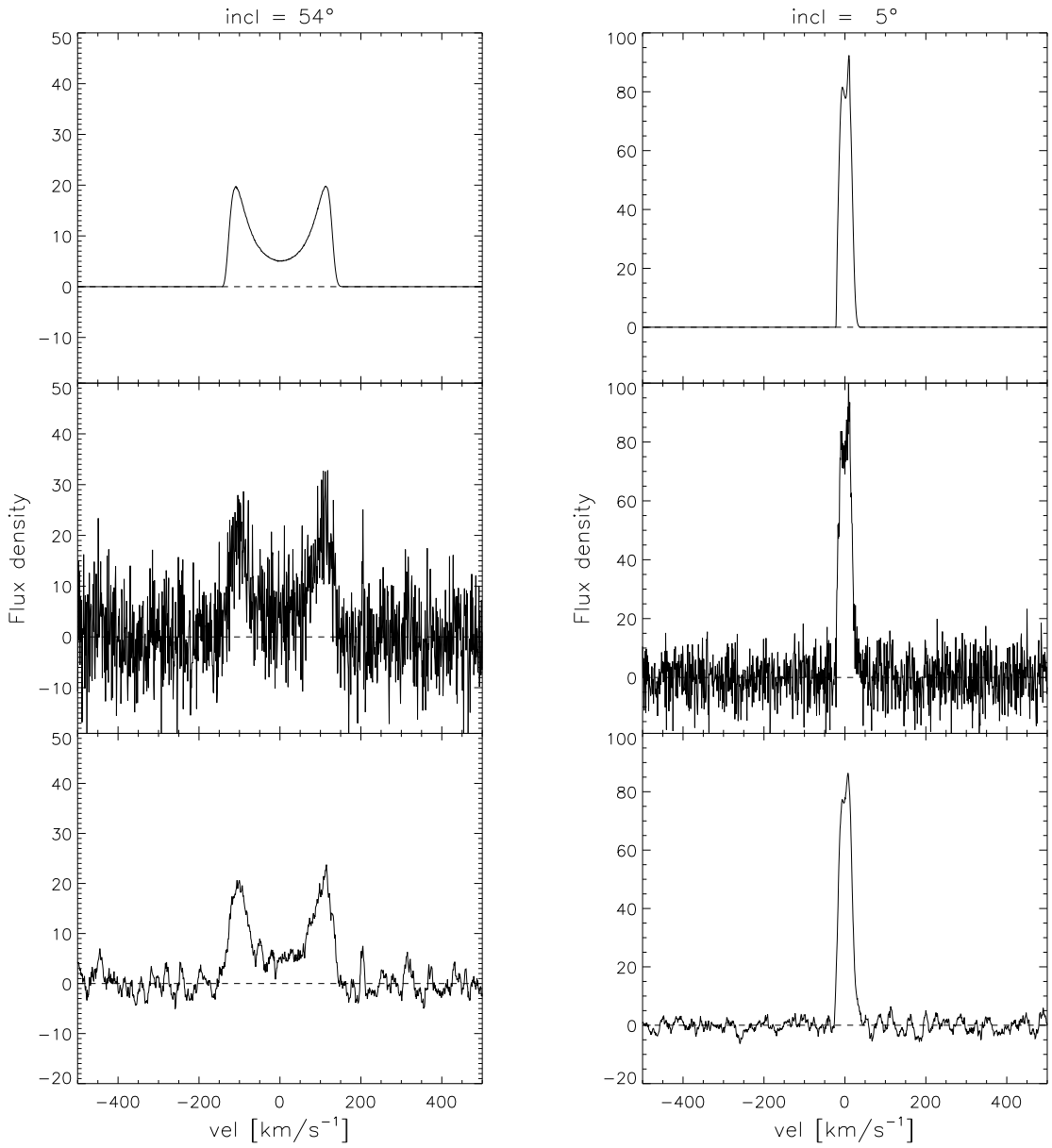


Figure A.2: Simulated HI spectra for two different inclinations (reported on top of each column). The first row shows the spectra without noise, the second and third with Gaussian white noise added. In particular, the bottom row has been smoothed to 11 km s^{-1} to mimic the ALFALFA resolution.

detection-like spectra, defined as spectra in which the noise is larger than the signal beneath, we select at the end the subset of 90 non detected like spectra, some of which are shown in Figure A.3. In some cases the detection is actually marginal, as in the middle panel of the bottom row where it is possible to guess the profile of the HI emission.

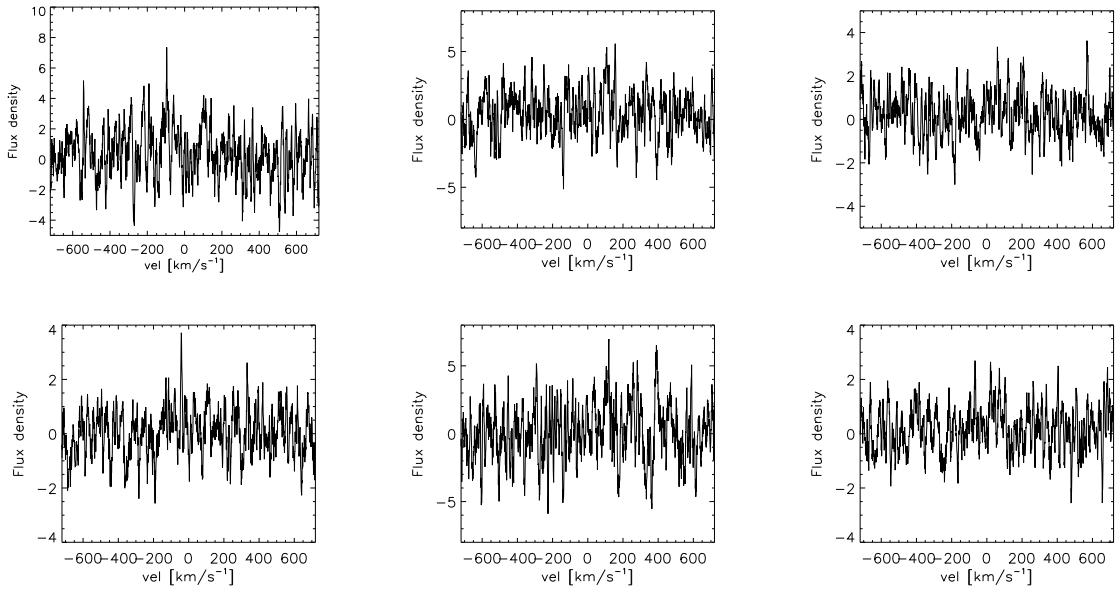


Figure A.3: Some examples of simulated non detections.

A.1.3 Stacking of spectra

We have created 90 non detection like spectra $S_{v;i}$, i.e. for which the noise is higher then the signal itself. For each of them we know the input rms_i and the input total signal ($S_{int;i}$), which is the integral of the HI distribution (eq. A.1).

We use the standard stacking procedure of co-adding the spectra $S_{v;i}$ using their *rms* as a weight, so that worse data (higher noise) contribute less to the mean. The final stacked spectrum S_v is obtained according to:

$$S_v = \sum_{i=1}^{90} \left(\frac{S_{v;i}}{rms_i^2} \right) / \sum_{i=1}^{90} \left(\frac{1}{rms_i^2} \right) \tag{A.4}$$

In Figure A.4 we show the stacked spectrum. The noise has decreased as expected, and the signal is clearly visible with a high signal-to-noise ratio. The shape of the profile is now given by the convolution of the individual lines, so no longer have a double-horn profile.

We manually flag the edges of the stacked emission, as we would do for a real spectrum, and mark the edges as dotted lines in Figure A.4. When we flag the edges, we may be missing some signal at the borders of the stacked line where only the wider profiles contributed. This is a limit of the stacking process, but if the signal-to-noise is good the fraction of lost signal is marginal.

We integrate the stacked signal inside the flagged region, measure an error using the bootstrap technique (measuring $S_{int;\mathrm{stack}} = 26.5 \pm 3.1$), and evaluate the *rms* over the remaining channels (rms_{stack}=0.07). The advantage of the simulation is that we can compare these results with the average values, since we know the individual inputs by construction. First, the recovered value from the stacking is in excellent agreement with the mean of the individual input signals ($\langle S_{int} \rangle = 26.7 \pm 1.9$). Second, the *rms* decreases as expected: we measure a final *rms* of 0.07 with respect to theoretical value of $\langle rms_i \rangle / \sqrt{90} = 0.08$.

The last check that we performed over the model data is a more detailed study of how the measured *rms* scales with the number of objects co-added, and the type of noise in input. We have already discussed this test in Section 3.2.1, so we just summarize it here for completeness and report the plot (Figure A.5). Using both detected and non-detected model spectra, we stack increasing numbers of randomly selected ones at which: we have first added only Gaussian white noise (black dots), and subsequently a further sinusoidal component (blue and green dots). The dashed line in the Figure is the expected average *rms* value assuming purely Gaussian noise. There is perfect agreement between the trends for the samples with only Gaussian component, while we observe a flattening in the examples including the additional sinusoidal noise. In particular, the flattening of the relation happens for smaller ensembles if the non-Gaussian component is larger, as in the blue case.

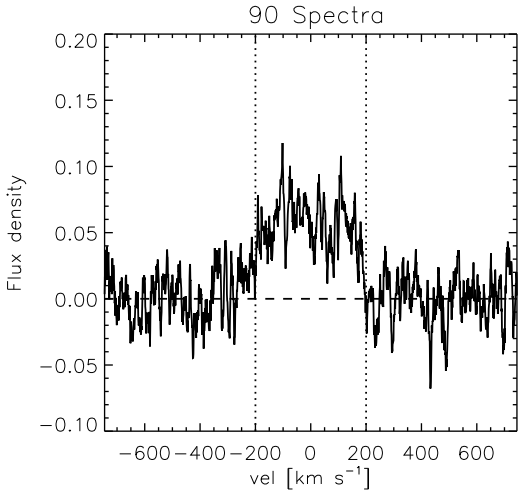


Figure A.4: Spectrum obtained stacking 90 non detections like spectra. The dotted vertical lines mark the region inside which we integrated the signal.

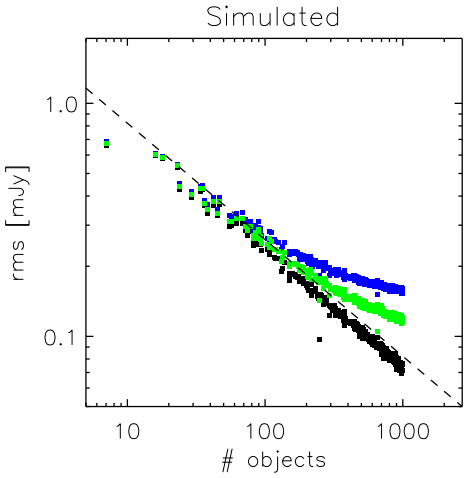


Figure A.5: Dependence of the *rms* of the stacked spectra as a function of the number of objects co-added. The dashed line is the expected $1/\sqrt{N}$ dependence. Black points are obtained considering only Gaussian noise; coloured points are obtained adding an additional sinusoidal component to the noise (stronger for the blues ones).

A.2 Comparison of two different stacking procedures

The physical quantity that we want to measure from the stacking is a gas fraction, i.e. the ratio between the HI mass and the stellar mass. For our analysis we decided to stack directly spectra converted into "gas fractions", so that it is not necessary to estimate average values for both $M_\star$ and z, whose distributions may not be well represented by means or medians (see §3.2.3 for further discussion). In this section we compare this method with an alternative approach for deriving average HI gas fractions from stacked spectra, and show that the results are consistent.

We start presenting the two methods, and then we move to the comparison.

1) Stacking "gas fractions". This approach, better described in section 3.2.3, is the one adopted throughout this work. Briefly, we weight each spectrum by the corresponding ${M_\star}^{-1}$, $(1+z)^{-1}$ and $D_L^2(z)$, so that the resulting stacked signal is already the average gas fraction of the ensemble.

2) Stacking HI fluxes. A different approach consists in co-adding the HI spectra S_i in their original form, so that the signal measured form the resulting spectrum is an HI flux [Jy].

To derive the average HI gas fraction, which can be estimated from equation 3.4, we have to assume average values of z and $M_\star$. We note that our redshift range is compact enough so that: $D_L^2(\langle z\rangle) \simeq \langle D_L(z)\rangle^2 \simeq \langle D_L^2(z)\rangle$, and we can just evaluate the distance at the average redshift.

The stacked gas fraction would then be obtained as:

$$\langle\frac{M_{HI}}{M_*}\rangle = \frac{2.356\times10^5}{1+\langle z\rangle}\frac{D_L^2(\langle z\rangle)}{\langle M_*\rangle}\frac{\Sigma_{i=0}^N S_i\cdot w_i}{\Sigma_{i=0}^N w_i}, \tag{A.5}$$

where $\langle z\rangle$ and $\langle M_\star\rangle$ are the mean values, and w_i the inverse of the squared *rms*.

The main limitation is that this method may produce biased results if galaxies in the bin span a significant range in redshift or stellar mass. In addition, by stacking mainly non detected spectra we do not know how much each galaxy contributes to the total signal.

This approach will work best if the bins are small. To apply it, we choose to split each ensemble we want to stack into three sub-bins in redshift: [0.025;0.033], [0.033;0.042], [0.417;0.050]. For each of these sub-bins, we evaluate the mean stellar mass $\langle M_\star\rangle$, and split the sample into two further bins according to: $M_\star \gtrless \langle M_\star\rangle$. We stack the

spectra in each of the 6 sub-bins, measure a flux and evaluate a gas fraction using both the average and median values of z and $M_\star$. Then we evaluate the final gas fraction by averaging the six values obtained (weighted by the number of objects co-added).

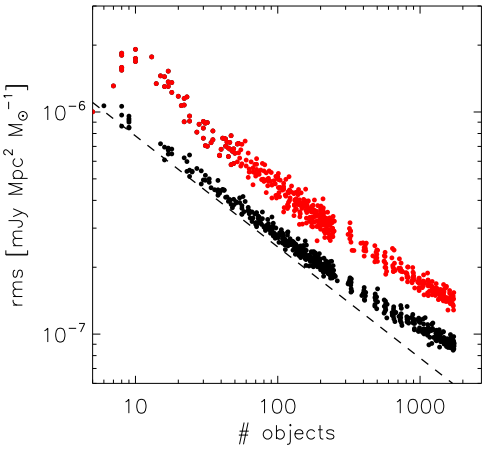


Figure A.6: Dependence of the *rms* of the stacked spectra on the number of objects stacked, for the two different methods. The red dots are obtained by stacking "gas fractions". The black ones by stacking HI fluxes. The dashed line is the expected $1/\sqrt{N}$ dependence (note that the black points have been multiplied by the average squared distance and divided by the average $M_\star$ in order to have the same units as the red points [mJy Mpc^2 M$_\odot^{-1}$], and then offset for display purpose).

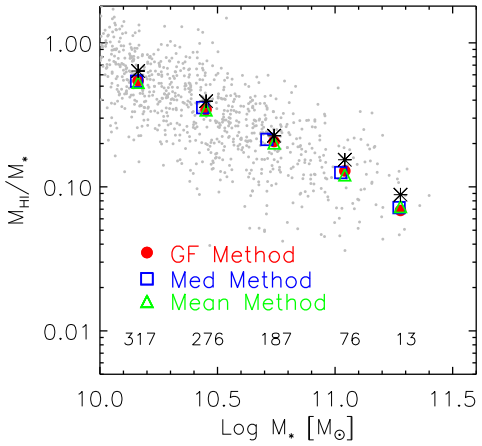


Figure A.7: Comparison of average gas fractions obtained with the two different stacking methods applied to galaxies with an ALFALFA detection. Red circles are obtained using method 1; green triangles/blue squares are from method 2 (assuming respectively means or medians in equation A.5). The stars are the means obtained from averaging the catalogued ALFALFA values. Gray dots show the ALFALFA detections. The numbers of objects co-added in each bin is reported.

As first test, we verify that the *rms* of the co-added spectra decreases as $1/\sqrt{N}$ in both cases; when method (1) is applied, i.e. by multiplying each flux by distance and stellar mass, we are in fact rescaling the noise in the spectra in a sensible way. Results are shown in Figure A.6: red circles are obtained by stacking "gas fractions", black circles are obtained by stacking fluxes. The black points have been multiplied by the average squared distance and divided by the average $M_\star$ in order to have the same units as the red points [mJy Mpc^2 M$_\odot^{-1}$], and slightly offset for display purposes. We caution that the *rms*

obtained with method (1) are systematically higher (by less than ∼20%) because they are obtained by co-adding spectra with inhomogeneous noises. Since the different is small, and the expected $1/\sqrt{N}$ dependence is still recovered, method (1) is convenient.

We then compare the results of both stacking methods against the "real" averages from individual data. We select all the galaxies in *sample A* that are detected by ALFALFA (detection codes 1+2) and divide them into five stellar mass bins. For each bin, we can then estimate an average gas fraction in four different ways, which are shown in Figure A.7: *i)* we evaluate the mean of the individual detections, using the catalogued ALFALFA measurements (black stars); *ii)* we stack "gas-fractions" (method 1) and measure the resulting signal (red circles); *iii)* we use method (2), assuming mean values for z and $M_\star$ (green triangles); *iv)* we use method (2), assuming median values for z and $M_\star$ (blue squares). Gray dots show the individual ALFALFA detections. If we compare all the results we find a good agreement, as shown in Figure A.7. First, the signal recovered from the stacking method used in this thesis is consistent with the mean value of the individual detections for each bin. Second, the two different stacking methods yield results which are also consistent with each other. Small differences of around ~ 10 % in M_{HI}/M_* do occur in the two largest mass bins that contain the fewest objects.

As a final, additional test, we compute the same HI scaling relations we studied in Figure 3.3 using all *sample A*, with both method (1) and (2). In Figure A.8 we show the results, where symbols are the same as described above. Once again, we obtain good consistency between the two methods.

Even if the two approaches lead to equivalent results, the second requires an accurate estimate of the average values of z and $M_\star$, which implies considering small sub-samples of these two properties. Since we do not want to sacrifice statistics in our stacking analysis by splitting a given ensemble into further sub-bins, we choose to use the "gas-fraction" method.

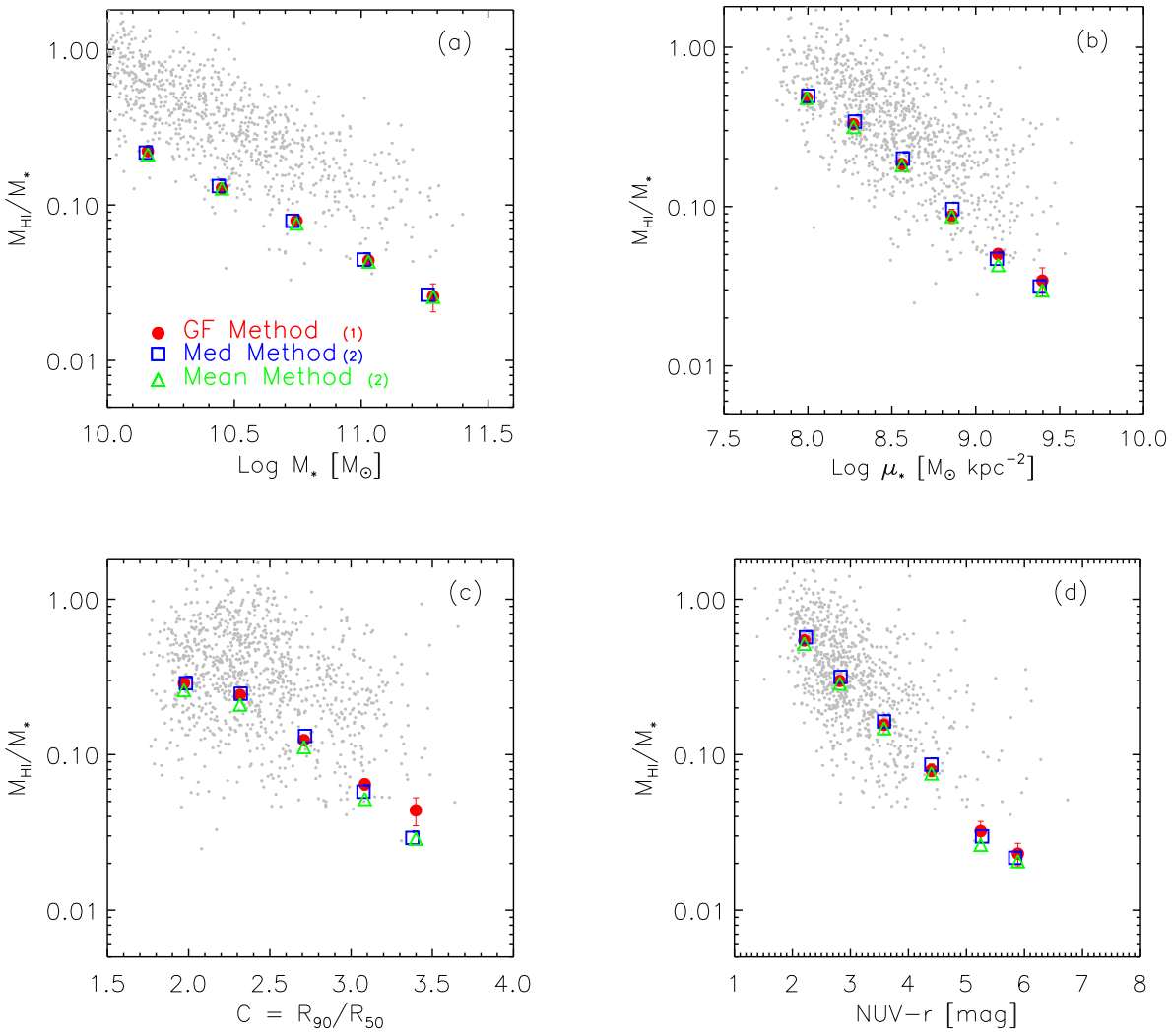


Figure A.8: Comparison of average gas fractions obtained with the two different stacking methods, applied to all *sample A* galaxies. Symbols and colours are the same as Figure A.7. Errors on method (2) points are evaluated with bootstrap.

B

Correction for Arecibo beam confusion

The Arecibo telescope, single dish of 305m in diameter, has a spatial resolution of ∼3.5 arcminutes at 21 cm (the beam FWHM). Such angular size corresponds to a physical scale of 0.15 Mpc at the mean redshift considered in this work (i.e. $z \simeq 0.037$). Therefore, one concern in doing an analysis using Arecibo data is the possible confusion of signals coming from different galaxies within the beam at similar redshift. An example of this case is shown in Figure B.1, where two companions are within the beam in addition to the (central) main target. On the left, the SDSS image is shown, where the yellow circle indicates the Arecibo beam size. On the top right, the resulting spectrum obtained with Arecibo is shown. The vertical lines flag the expected central velocities of the three objects as evaluated from SDSS redshifts. The presented example, from GASS data, has been chosen to illustrate the problem but actually does not represent exactly the kind of confusion that we may have in stacking data. In fact, as we remind the reader, we visually inspected each spectrum (§ 3.2.1), and we discarded the ones with a strong signal close to the galaxy but not centered at the expected redshift (in the example, the pink one). Therefore, confusion in our analysis mainly arises from the stacking of non-detected emissions, or if companion and main target have almost exactly the same redshift (as for the black and green profiles, although this example spectrum would have already been discarded).

In Chapter 4 and 5 we did not apply any correction because the contribution to the measured HI emission from the stacking was not significant. For the environmental analysis, though, the confusion becomes larger, especially in the high density bins where galaxies are more clustered. In order to identify confused objects, we search inside the MPA-JHU spectroscopic sample of galaxies with $M_\star > 3 \times 10^9 M_\odot$, for which the projected distance from the target is smaller than the beam FWHM, and the velocity separation is smaller

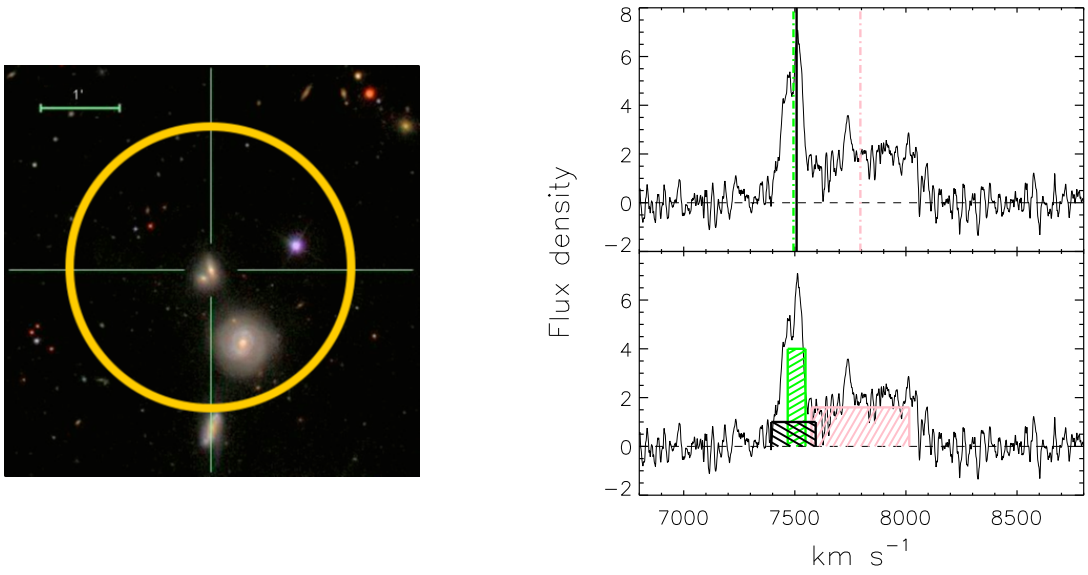


Figure B.1: (Exceptional) example of possible signal confusion inside the Arecibo beam. *Left:* SDSS image of the galaxy GASS 49727 and its companions within the beam. The yellow circle indicates the 3.5 arcminutes Arecibo beam size. *Right:* the spectrum obtained with Arecibo. On top, the main target (black solid line) and the two companions (coloured dotted ones) central velocities are flagged, as estimated from SDSS spectroscopy. On bottom, the shadowed regions show how we would model the HI signals, as described in the text.

than 300 km s^{-1}. This value is representative for the *observed*, i.e. not corrected for inclination, velocities in massive galaxies. If the velocity difference is larger, the signals will not overlap. Companions with strong emission at large velocity separation can eventually increase the noise in the stacked baseline, but do not affect the measured gas content. ∼20 % of *sample A* targets have at least one confusing companion, according to these criteria.

In order to account for contamination from the close companions, we proceed according to the following steps.

1) We estimate the expected gas content of each companion, using the relation between colour, stellar mass surface density and HI gas fraction found by Zhang et al. (2009):

$$\mathrm{Log}\left(\frac{\mathrm{M_{HI}}}{\mathrm{M_*}}\right) = -1.73238 \cdot (g-r) + 0.215182 \cdot \mu_i - 4.08451, \qquad \text{(B.1)}$$

where μ_i is the surface brightness in the i-band, and g and r are SDSS magnitudes corrected for Galactic extinction. We do not use here the relation from Catinella

et al. (2010) because *(i)* we consider also lower stellar masses than they do, and *(ii)* we do not have NUV data for all companions.

2) We estimate the actual signal contaminating the stacked spectrum. First, we calculate the fraction of it the companion emission observed (f_1), considering the projected distance between the target and the companion. The beam profile can be approximated with a 2D Gaussian with $\sigma_x = (2\sqrt{2 \cdot \ln 2})^{-1} \times 3.3'$ and $\sigma_y = (2\sqrt{2 \cdot \ln 2})^{-1} \times 3.8'$, so that its response decreases at the edges. The bigger companion in Figure B.1, for example, lies at a projected distance of $x \simeq 0.4$ arcminutes, and $y \simeq 1$ arcminutes. Therefore, $f_1 = \exp[-0.5 \cdot (x/\sigma_x)^2 - 0.5 \cdot (y/\sigma_y)^2]$=0.8 of its flux would be recorded. Second, only part of the companion signal would actually overlap with the main target in the velocity space. To estimate the fraction of contaminating emission (f_2), we calculate the expected observed widths of both the main target (w_t) and the companion (w_c), using a Tully-Fisher relation as in § 3.2, and assume a box-shape profile. The fraction of the contaminating flux is given by the velocity overlap (Δw) between w_t and w_c: $f_2 = \Delta w / w_c$. As an example, in Figure B.1 (bottom right) the dashed regions represent how we would have modeled the three signals contributing to the spectrum. In the example, the flux from the green companion contributes entirely to the measured signal ($f_2 = 1.$), as it entirely overlaps with the main target emission (black region); on the contrary, the pink companion contributes only for a very small fraction (overlap of $f_2 = 0.06$).

3) We check that the estimated flux for companions with $\Delta v > 50\,\mathrm{km\ s^{-1}}$ would be not detected above the average ALFALFA noise level of 2.2 mJy. If it were detected and not centered, we would have discarded the spectrum. In the opposite case, we assume that the correction is overestimating the HI content of the neighbour and we then set it to the actual ALFALFA upper limit.

4) Finally, we subtract the contributions from all the confused companions to the HI mass measured on the stacked spectrum, as follows.

   Because of the weight we apply to the spectra of the individual galaxies (i), a gas fraction measured from the stacked spectrum is (§ 3.2.2):

$$\frac{\mathrm{M_{HI}}}{\mathrm{M_*}} = \frac{2.356 \times 10^5}{\Sigma_i \mathrm{w}_i} \Sigma_i \frac{\mathrm{D_L^2}(z_i)}{(1+z_i)} \frac{S_i}{\mathrm{M_{*;i}}} \mathrm{w}_i. \tag{B.2}$$

   The total integrated signal S_i is actually the real emission from the main target (t) plus the ones from the confused objects (c), weighted for the two factors described

in (2):

$$S_i \;=\; S_t + \Sigma_c \, f_{1;c} f_{2;c} S_c. \tag{B.3}$$

We can rewrite B.2 as

$$\frac{\mathrm{M_{HI}}}{\mathrm{M_*}} = \left(\frac{\mathrm{M_{HI}}}{\mathrm{M_*}}\right)_t + \frac{2.356 \times 10^5}{\Sigma_i \mathrm{w}_i} \Sigma_i \frac{\mathrm{D_L^2}(z_i)}{(1+z_i)} \frac{\Sigma_c \, f_{1;c} f_{2;c} S_c}{\mathrm{M_{*;i}}} \mathrm{w}_i. \tag{B.4}$$

And finally, if we substitute the estimated gas fractions from photometry (gf_c):

$$\left(\frac{\mathrm{M_{HI}}}{\mathrm{M_*}}\right)_t = \frac{\mathrm{M_{HI}}}{\mathrm{M_*}} - \frac{1}{\Sigma_i \mathrm{w}_i} \Sigma_i \left[\frac{\mathrm{D_L^2}(z_i)}{(1+z_i)} \frac{\mathrm{w}_i}{\mathrm{M_{*;i}}} \Sigma_c \left(\frac{f_{1;c} f_{2;c} \cdot \mathrm{gf}_c \cdot \mathrm{M_{*;c}} \cdot (1+z_c)}{\mathrm{D_L^2}(z_c)} \right) \right]$$

With the procedure presented here, we correct the gas fraction values measured with the stacking in Chapter 6. As already mentioned there, corrections for contamination are always small. Even in the bins of higher local densities, the correction factor is smaller than few percent.

Bibliography

Abadi, M. G., Moore, B., & Bower, R. G. 1999, MNRAS, 308, 947

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543

Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2008, ApJS, 175, 297

Antonucci, R. 1993, ARA&A, 31, 473

Auld, R., Minchin, R. F., Davies, J. I., et al. 2006, MNRAS, 371, 1617

Bagetakos, I., Brinks, E., Walter, F., & de Blok, E. 2007, in IAU Symposium, Vol. 237, IAU Symposium, ed. B. G. Elmegreen & J. Palous, 393–393

Baldry, I. K., Balogh, M. L., Bower, R., Glazebrook, K., & Nichol, R. C. 2004, in American Institute of Physics Conference Series, Vol. 743, The New Cosmology: Conference on Strings and Cosmology, ed. R. E. Allen, D. V. Nanopoulos, & C. N. Pope, 106–119

Baldry, I. K., Balogh, M. L., Bower, R. G., et al. 2006, MNRAS, 373, 469

Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5

Ball, N. M., Loveday, J., & Brunner, R. J. 2008, MNRAS, 383, 907

Balogh, M., Eke, V., Miller, C., et al. 2004, MNRAS, 348, 1355

Balogh, M. L., McGee, S. L., Wilman, D., et al. 2009, MNRAS, 398, 754

Balogh, M. L., Morris, S. L., Yee, H. K. C., Carlberg, R. G., & Ellingson, E. 1999, ApJ, 527, 54

Bamford, S. P., Nichol, R. C., Baldry, I. K., et al. 2009, MNRAS, 393, 1324

Bernardi, M., Sheth, R. K., Annis, J., et al. 2003, AJ, 125, 1817

Bigiel, F., Leroy, A., Walter, F., et al. 2008, AJ, 136, 2846

Blanton, M. R. & Berlind, A. A. 2007, ApJ, 664, 791

Booth, R. S., de Blok, W. J. G., Jonas, J. L., & Fanaroff, B. 2009, ArXiv e-prints

Boselli, A. & Gavazzi, G. 2006, PASP, 118, 517

Boselli, A. & Gavazzi, G. 2009, A&A, 508, 201

Boselli, A., Gavazzi, G., Donas, J., & Scodeggio, M. 2001, AJ, 121, 753

Bothwell, M. S., Kennicutt, R. C., & Lee, J. C. 2009, MNRAS, 400, 154

Bower, R. G. & Balogh, M. L. 2004, Clusters of Galaxies: Probes of Cosmological Structure and Galaxy Evolution, 325

Bower, R. G., Benson, A. J., Malbon, R., et al. 2006, MNRAS, 370, 645

Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., & Lemson, G. 2009, MNRAS, 398, 1150

Bravo-Alfaro, H., Cayatte, V., van Gorkom, J. H., & Balkowski, C. 2000, AJ, 119, 580

Bregman, J. N., Hogg, D. E., & Roberts, M. S. 1992, ApJ, 387, 484

Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151

Bruzual, G. & Charlot, S. 2003, MNRAS, 344, 1000

Buat, V., Deharveng, J. M., & Donas, J. 1989, A&A, 223, 42

Burke, B. F. & F., G.-S. 2011, An introduction to Radio Astronomy, ed. Cambridge University Press

Cappellari, M., Emsellem, E., Krajnović, D., et al. 2011, MNRAS, 413, 813

Carilli, C. L., Lee, N., Capak, P., et al. 2008, ApJ, 689, 883

Catinella, B., Giovanelli, R., & Haynes, M. P. 2006, ApJ, 640, 751

Catinella, B., Haynes, M. P., Giovanelli, R., Gardner, J. P., & Connolly, A. J. 2008, ApJL, 685, L13

Catinella, B., Schiminovich, D., Kauffmann, G., et al. 2010, MNRAS, 403, 683

Cattaneo, A., Blaizot, J., Weinberg, D. H., et al. 2007, MNRAS, 377, 63

Cattaneo, A., Dekel, A., Devriendt, J., Guiderdoni, B., & Blaizot, J. 2006, MNRAS, 370, 1651

Cattaneo, A., Faber, S. M., Binney, J., et al. 2009, Nature, 460, 213

Chakrabarti, S. & Blitz, L. 2009, MNRAS, 399, L118

Charlot, S. & Longhetti, M. 2001, MNRAS, 323, 887

Chengalur, J. N., Braun, R., & Wieringa, M. 2001, A&A, 372, 768

Chung, A., van Gorkom, J. H., Kenney, J. D. P., Crowl, H., & Vollmer, B. 2009, AJ, 138, 1741

Churazov, E., Brüggen, M., Kaiser, C. R., Böhringer, H., & Forman, W. 2001, ApJ, 554, 261

Cole, S. 1991, ApJ, 367, 45

Cortese, L., Catinella, B., Boissier, S., Boselli, A., & Heinis, S. 2011, MNRAS, 415, 1797

Croton, D. J., Springel, V., White, S. D. M., et al. 2006, MNRAS, 365, 11

Dave, R. 2006, in The Fabulous Destiny of Galaxies: Bridging Past and Present, ed. V. Le Brun, A. Mazure, S. Arnouts, & D. Burgarella, 219

De Lucia, G. & Blaizot, J. 2007, MNRAS, 375, 2

Dekel, A., Sari, R., & Ceverino, D. 2009, ApJ, 703, 785

Di Matteo, T., Springel, V., & Hernquist, L. 2005, Nature, 433, 604

di Serego Alighieri, S., Gavazzi, G., Giovanardi, C., et al. 2007, A&A, 474, 851

Diamond-Stanic, A. M. & Rieke, G. H. 2012, ApJ, 746, 168

Dressler, A. 1980, ApJ, 236, 351

Dressler, A., Thompson, I. B., & Shectman, S. A. 1985, ApJ, 288, 481

Efron, B. 1979, Ann. Statist., 7, 1

Ewen, H. I. & Purcell, E. M. 1951, Nature, 168, 356

Fabello, S., Catinella, B., Giovanelli, R., et al. 2011, MNRAS, 411, 993

Fabian, A. C., Celotti, A., & Erlund, M. C. 2006, MNRAS, 373, L16

Ferrarese, L. & Merritt, D. 2000, ApJL, 539, L9

Feruglio, C., Maiolino, R., Piconcelli, E., et al. 2010, A&A, 518, L155

Fischer, T. C., Crenshaw, D. M., Kraemer, S. B., Schmitt, H. R., & Trippe, M. L. 2010, AJ, 140, 577

Freeland, E., Stilp, A., & Wilcots, E. 2009, AJ, 138, 295

Gadotti, D. A. 2009, MNRAS, 393, 1531

Gao, L. & White, S. D. M. 2006, MNRAS, 373, 65

Gavazzi, G. 1989, ApJ, 346, 59

Gavazzi, G., Boselli, A., Pedotti, P., Gallazzi, A., & Carrasco, L. 2002, A&A, 396, 449

Gavazzi, G., Giovanelli, R., Haynes, M. P., et al. 2008, A&A, 482, 43

Gavazzi, G. & Jaffe, W. 1985, ApJL, 294, L89

Gebhardt, K., Kormendy, J., Ho, L. C., et al. 2000, ApJL, 543, L5

Giovanelli, R. & Haynes, M. P. 1985, ApJ, 292, 404

Giovanelli, R. & Haynes, M. P. 2002, ApJl, 571, L107

Giovanelli, R., Haynes, M. P., Herter, T., et al. 1997a, AJ, 113, 53

Giovanelli, R., Haynes, M. P., Herter, T., et al. 1997b, AJ, 113, 22

Giovanelli, R., Haynes, M. P., Kent, B. R., et al. 2005, AJ, 130, 2598

Graves, G. J., Faber, S. M., & Schiavon, R. P. 2009a, ApJ, 693, 486

Graves, G. J., Faber, S. M., & Schiavon, R. P. 2009b, ApJ, 698, 1590

Grossi, M., di Serego Alighieri, S., Giovanardi, C., et al. 2009, A&A, 498, 407

Guillard, P., Ogle, P. M., Emonts, B. H. C., et al. 2012, ApJ, 747, 95

Gunn, J. E. & Gott, III, J. R. 1972, ApJ, 176, 1

Guo, Q., White, S., Boylan-Kolchin, M., et al. 2011, MNRAS, 413, 101

Hashimoto, Y., Oemler, Jr., A., Lin, H., & Tucker, D. L. 1998, ApJ, 499, 589

Haynes, M. P. 2007, Nuovo Cimento B Serie, 122, 1109

Haynes, M. P. & Giovanelli, R. 1984, AJ, 89, 758

Haynes, M. P., Giovanelli, R., & Chincarini, G. L. 1984, ARA&A, 22, 445

Haynes, M. P., Giovanelli, R., Martin, A. M., et al. 2011, AJ, 142, 170

Heckman, T. M., Kauffmann, G., Brinchmann, J., et al. 2004, ApJ, 613, 109

Helmboldt, J. F. 2007, MNRAS, 379, 1227

Hewitt, J. N., Haynes, M. P., & Giovanelli, R. 1983, AJ, 88, 272

Ho, L. C., Darling, J., & Greene, J. E. 2008, ApJ, 681, 128

Ho, L. C., Filippenko, A. V., & Sargent, W. L. W. 1993, ApJ, 417, 63

Ho, L. C., Filippenko, A. V., & Sargent, W. L. W. 1997, ApJS, 112, 315

Hodge, J. A., Zeimann, G. R., Becker, R. H., & White, R. L. 2009, AJ, 138, 900

Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2006, ApJS, 163, 1

Hubble, E. & Humason, M. L. 1931, ApJ, 74, 43

Hughes, T. M. & Cortese, L. 2009, MNRAS, 396, L41

Jeans, J. H. 1902, Royal Society of London Philosophical Transactions Series A, 199, 1

Jenkins, C. 2003, Practical statistics for astronomers (Cambridge University Press)

Johnson, B. D., Schiminovich, D., Seibert, M., et al. 2007, ApJS, 173, 392

Johnston, S., Feain, I. J., & Gupta, N. 2009, in Astronomical Society of the Pacific Conference Series, Vol. 407, The Low-Frequency Radio Universe, ed. D. J. Saikia, D. A. Green, Y. Gupta, & T. Venturi, 446

Kanekar, N. & Briggs, F. H. 2004, NewAR, 48, 1259

Kauffmann, G. & Haehnelt, M. 2000, MNRAS, 311, 576

Kauffmann, G. & Heckman, T. M. 2009, MNRAS, 397, 135

Kauffmann, G., Heckman, T. M., De Lucia, G., et al. 2006, MNRAS, 367, 1394

Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003a, MNRAS, 346, 1055

Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003b, MNRAS, 341, 54

Kauffmann, G., White, S. D. M., & Guiderdoni, B. 1993, MNRAS, 264, 201

Kauffmann, G., White, S. D. M., Heckman, T. M., et al. 2004, MNRAS, 353, 713

Kenney, J. D. P. & Young, J. S. 1989, ApJ, 344, 171

Kennicutt, Jr., R. C. 1983, ApJ, 272, 54

Kennicutt, Jr., R. C. 1989, ApJ, 344, 685

Kennicutt, Jr., R. C. 1998, ApJ, 498, 541

Kennicutt, Jr., R. C., Bothun, G. D., & Schommer, R. A. 1984, AJ, 89, 1279

Kennicutt, Jr., R. C., Calzetti, D., Walter, F., et al. 2007, ApJ, 671, 333

Kent, B. R. 2008, in American Institute of Physics Conference Series, Vol. 1035, The Evolution of Galaxies Through the Neutral Hydrogen Window, ed. R. Minchin & E. Momjian, 225–228

Kent, B. R., Giovanelli, R., Haynes, M. P., et al. 2008, AJ, 136, 713

Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, MNRAS, 363, 2

Kerr, F. J. & Hindman, J. V. 1953, AJ, 58, 218

Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A., & Trevena, J. 2001, ApJ, 556, 121

Kewley, L. J., Groves, B., Kauffmann, G., & Heckman, T. 2006, MNRAS, 372, 961

Kimm, T., Somerville, R. S., Yi, S. K., et al. 2009, MNRAS, 394, 1131

Knapp, G. R., Turner, E. L., & Cunniffe, P. E. 1985, AJ, 90, 454

Koopmann, R. A., Giovanelli, R., Haynes, M. P., et al. 2008, ApJL, 682, L85

Koopmann, R. A. & Kenney, J. D. P. 2004, ApJ, 613, 866

Kormendy, J. 2004, Coevolution of Black Holes and Galaxies, 1

Kormendy, J. & Richstone, D. 1995, ARA&A, 33, 581

Lah, P., Chengalur, J. N., Briggs, F. H., et al. 2007, MNRAS, 376, 1357

Lah, P., Pracy, M. B., Chengalur, J. N., et al. 2009, MNRAS, 399, 1447

Larson, R. B., Tinsley, B. M., & Caldwell, C. N. 1980, ApJ, 237, 692

Li, C., Kauffmann, G., Jing, Y. P., et al. 2006, MNRAS, 368, 21

Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, AJ, 115, 2285

Mandelbaum, R., Seljak, U., Kauffmann, G., Hirata, C. M., & Brinkmann, J. 2006, MNRAS, 368, 715

Maraston, C. 2005, MNRAS, 362, 799

Martig, M., Bournaud, F., Teyssier, R., & Dekel, A. 2009, ApJ, 707, 250

Martin, A. M., Papastergis, E., Giovanelli, R., et al. 2010, ApJ, 723, 1359

Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, ApJL, 619, L1

McKee, C. F. & Ostriker, E. C. 2007, ARA&A, 45, 565

Meyer, M. J., Zwaan, M. A., Webster, R. L., et al. 2004, MNRAS, 350, 1195

Morganti, R., de Zeeuw, P. T., Oosterloo, T. A., et al. 2006, MNRAS, 371, 157

Muldrew, S. I., Croton, D. J., Skibba, R. A., et al. 2012, MNRAS, 419, 2670

Nandra, K., Georgakakis, A., Willmer, C. N. A., et al. 2007, ApJL, 660, L11

Nesvadba, N. P. H., Lehnert, M. D., De Breuck, C., Gilbert, A. M., & van Breugel, W. 2008, A&A, 491, 407

Oemler, Jr., A. 1974, ApJ, 194, 1

Oosterloo, T., Morganti, R., Crocker, A., et al. 2010, MNRAS, 409, 500

Ostriker, J. P. & Peebles, P. J. E. 1973, ApJ, 186, 467

Pisano, D. J., Barnes, D. G., Staveley-Smith, L., et al. 2011, ApJS, 197, 28

Putman, M. E., Gibson, B. K., Staveley-Smith, L., et al. 1998, Nature, 394, 752

Quilis, V., Bower, R. G., & Balogh, M. L. 2001, MNRAS, 328, 1091

Quilis, V., Moore, B., & Bower, R. 2000, Science, 288, 1617

Reynolds, C. S., Heinz, S., & Begelman, M. C. 2001, ApJL, 549, L179

Roberts, M. 1975, Sandage, A. and Sandage, M. and Kristian, J. - Radio Observations of Neutral Hydrogen in Galaxies, ed. University of Chicago Press

Roberts, M. S. 1963, ARA&A, 1, 149

Roberts, M. S. & Haynes, M. P. 1994, ARA&A, 32, 115

Roberts, M. S. & Rots, A. H. 1973, A&A, 26, 483

Ruszkowski, M. & Begelman, M. C. 2002, ApJ, 581, 223

Saintonge, A. 2007, AJ, 133, 2087

Saintonge, A., Kauffmann, G., Kramer, C., et al. 2011a, MNRAS, 415, 32

Saintonge, A., Kauffmann, G., Wang, J., et al. 2011b, MNRAS, 415, 61

Salim, S. e. a. 2007, ApJS, 173, 267

Schawinski, K., Lintott, C. J., Thomas, D., et al. 2009a, ApJ, 690, 1672

Schawinski, K., Thomas, D., Sarzi, M., et al. 2007, MNRAS, 382, 1415

Schawinski, K., Virani, S., Simmons, B., et al. 2009b, ApJL, 692, L19

Schiminovich, D., Catinella, B., Kauffmann, G., et al. 2010, MNRAS, 408, 919

Schiminovich, D., Wyder, T. K., Martin, D. C., et al. 2007, ApJS, 173, 315

Schuster, K. F., Kramer, C., Hitschfeld, M., Garcia-Burillo, S., & Mookerjea, B. 2007, A&A, 461, 143

Serra, P., Morganti, R., Oosterloo, T. A., et al. 2010, in Astronomical Society of the Pacific Conference Series, Vol. 421, Galaxies in Isolation: Exploring Nature Versus Nurture, ed. L. Verdes-Montenegro, A. Del Olmo, & J. Sulentic, 49

Shen, S., White, S. D. M., Mo, H. J., et al. 2006, MNRAS, 369, 1639

Shostak, G. S. & Allen, R. J. 1980, A&A, 81, 167

Sijacki, D. & Springel, V. 2006, MNRAS, 366, 397

Silverman, J. D., Mainieri, V., Lehmer, B. D., et al. 2008, ApJ, 675, 1025

Skibba, R. A., Bamford, S. P., Nichol, R. C., et al. 2009, MNRAS, 399, 966

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163

Solanes, J. M., Manrique, A., García-Gómez, C., et al. 2001, ApJ, 548, 97

Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E., & Hernquist, L. 2008, MNRAS, 391, 481

Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629

Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, AJ, 124, 1810

Sturm, E., González-Alfonso, E., Veilleux, S., et al. 2011, ApJL, 733, L16

Tanaka, M., Goto, T., Okamura, S., Shimasaku, K., & Brinkmann, J. 2004, AJ, 128, 2677

Teyssier, R., Moore, B., Martizzi, D., Dubois, Y., & Mayer, L. 2011, MNRAS, 414, 195

Thomas, D., Maraston, C., Schawinski, K., Sarzi, M., & Silk, J. 2010, MNRAS, 404, 1775

Tremaine, S., Gebhardt, K., Bender, R., et al. 2002, ApJ, 574, 740

Tremonti, C. A., Moustakas, J., & Diamond-Stanic, A. M. 2007, ApJL, 663, L77

Tully, R. B. & Fisher, J. R. 1977, A&A, 54, 661

van den Bosch, F. C., Aquino, D., Yang, X., et al. 2008, MNRAS, 387, 79

Verheijen, M., van Gorkom, J. H., Szomoru, A., et al. 2007, ApJL, 668, L9

Verheijen, M. A. W., Oosterloo, T. A., van Cappellen, W. A., 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, 265–271

Vollmer, B. 2009, A&A, 502, 427

von der Linden, A., Wild, V., Kauffmann, G., White, S. D. M., & Weinmann, S. 2010, MNRAS, 404, 1231

Walter, F., Brinks, E., de Blok, W. J. G., et al. 2008, AJ, 136, 2563

Wang, J., Kauffmann, G., Overzier, R., et al. 2011, MNRAS, 412, 1081

Wardle, M. & Knapp, G. R. 1986, AJ, 91, 23

Weinmann, S. M., Kauffmann, G., van den Bosch, F. C., et al. 2009, MNRAS, 394, 1213

Weinmann, S. M., van den Bosch, F. C., Yang, X., & Mo, H. J. 2006, MNRAS, 366, 2

Wetzel, A. R., Tinker, J. L., & Conroy, C. 2012, MNRAS, 3128

White, S. D. M. & Frenk, C. S. 1991, ApJ, 379, 52

White, S. D. M. & Rees, M. J. 1978, MNRAS, 183, 341

Wild, V., Kauffmann, G., Heckman, T., et al. 2007, MNRAS, 381, 543

Wyder, T. K., Martin, D. C., Schiminovich, D., et al. 2007, ApJS, 173, 293

Yang, X., Mo, H. J., van den Bosch, F. C., et al. 2007, ApJ, 671, 153

York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579

Zhang, W., Li, C., Kauffmann, G., et al. 2009, MNRAS, 397, 1243

Zibetti, S., White, S. D. M., Schneider, D. P., & Brinkmann, J. 2005, MNRAS, 358, 949

Zwaan, M. A., Staveley-Smith, L., Koribalski, B. S., et al. 2003, AJ, 125, 2842

Acknowledgments

I am grateful to my supervisors, Guinevere Kauffmann and Barbara Catinella, for giving me the opportunity of working on this exciting project. Without their guidance and patience, you would not be reading this thesis. I thank Guinevere for her constant motivation and shared knowledge. I consider her a model for having being always helpful and willing to be involved in any step, while being the leader of a successful, big group. I owe even more, if possible, to Barbara Catinella for her experienced guidance into the HI world, I could almost say "hand in hand". I thank her for her honest encouragement and shared enthusiasm shown in every step of this work. And for having reminded me to "tirare fuori la grinta" when I forgot: you spurring me on has been very much appreciated (and effective)! I am glad to have you now as a friend.

Guinevere and Barbara, I would like to thank you both for your understanding when I took the decision to change direction in my career: without your support it would have been a little more difficult to undertake the next step in my working life.

This thesis would not have been possible without Riccardo Giovanelli and Martha Haynes, and not only because they gave me access to the remarkable ALFALFA data. Thanks for your help, suggestions and warm hospitality: that was particularly welcoming, in February, under one meter of snow!

I want to thank David Schiminovich and the whole GASS team for this awesome experience that brought me to the largest radio telescope of the world. It was a pleasure to get to know all of you guys!

I really enjoy interacting with Cheng Li and Luca Cortese, whom I warmly thank for

the useful insights that helped to improve the work a lot.

I am grateful to the many members of the ALFALFA team who have contributed to the acquisition and processing of the ALFALFA dataset over the last six years; and to the telescope operators and people I met during my awesome trip to Arecibo and Pico Veleta.

I owe my gratitude to the colleagues at MPA to whom I asked all sorts of questions, and my love to all the friends who shared the PhD experience with me, for better or for worse. Francesca and Alessia, Rob and Irina, and many more that I should mention: you have been my strength in Munich, joining me in these years of discovery - discoveries about galaxies but also about myself, who arrived as a scared student and leaves (≈)ready to conquer her future. During this time, I have learned the precious lesson of how far I can go counting on myself only; and yet how much more valuable time is when spent with wonderful people like you!

A special thanks goes to Michele, to whom everything just said apply despite the ocean between our friendship.

Finally, last because most important, huge thanks and hugs and love to my parents and little sister. Without our Skype dinners and your unreserved and overwhelming support, always and anyhow, it would have been so, so much harder - and a bit less meaningful. I am so proud of you!

Uli, a part of this work is also yours, as you kept me sane in these last months. A.i., but it's not adequate.

Now I am ready for another, new start.

"Will you succeed? Yes, you will indeed. ($98\frac{3}{4}$% guaranteed.)"[1]

[1] *Oh, the Places You'll Go!*, Dr. Seuss

List of publications

Refereed publications

[11] Catinella, B.; Schiminovich, D.; Kauffmann, G.; **Fabello, S.**; Hummels, C.; Lemonias, J.; Moran, S. M.; Wu, R.; Cooper, A. P.; Wang, J. *The GALEX Arecibo SDSS Survey. VI. Second Data Release and Updated Gas Fraction Scaling Relations.* A&A, in press.

[10] Li, C.; Kauffmann, G.; Fu, J.; Wang, J.; Catinella, B.; **Fabello, S.**; Schiminovich, D.; Zhang, W. *The clustering of galaxies as a function of their photometrically-estimated atomic gas content.* MNRAS, in press.

[9] Catinella, B.; Kauffmann, G.; Schiminovich, D.; Lemonias, J.; Scannapieco, C.; Wang, J.; **Fabello, S.**; Hummels, C.; Moran, S. M.; Wu, R.; Cooper, A. P.; Giovanelli, R.; Haynes, M. P.; Heckman, T. M.; Saintonge, A. *The GALEX Arecibo SDSS Survey - IV. Baryonic Mass-Velocity-Size Relations of Massive galaxies.* 2012, MNRAS, 420, 1959

[8] Moran, S. M.; Heckman, T. M.; Kauffmann, G.; Davé, R.; Catinella, B.; Brinchmann, J.; Wang, J.; Schiminovich, D.; Saintonge, A.; Graciá-Carpio, J.; Tacconi, L.; Giovanelli, R.; Haynes, M. P.; **Fabello, S.**; Hummels, C.; Lemonias, J.; Wu, R. *The GALEX Arecibo SDSS Survey V: The Relation between the HI Content of Galaxies and Metal Enrichment at their Outskirts.* 2012, ApJ, 745, 66

[7] **Fabello, S.**; Kauffmann, G.; Catinella, B.; Giovanelli, R.; Haynes, M. P.; Heckman, T. M.; Schiminovich, D. *ALFALFA HI Data Stacking II. HI content of the host galaxies of AGN.* 2011, MNRAS, 416, 1739

[6] Saintonge, A.; Kauffmann, G.; Wang, J.; Kramer, C.; Tacconi, L. J.; Buchbender, C.; Catinella, B.; Graciá-Carpio, J.; Cortese, L.; **Fabello, S.**; Fu, J.; Genzel, R.; Giovanelli, R.; Guo, Q.; Haynes, M. P.; Heckman, T. M.; Krumholz, M. R.; Lemonias, J.; Li, C.; Moran, S.; Rodriguez-Fernandez, N.; Schiminovich, D.; Schuster, K.; Sievers, A. *COLD GASS, an IRAM legacy survey of molecular gas in massive galaxies - II. The non-universality of the molecular gas depletion time-scale.* 2011, MNRAS, 415, 61

[5] Saintonge, A.; Kauffmann, G.; Kramer, C; Tacconi, L. J.; Buchbender, C.; Catinella, B.; **Fabello, S.**; Graciá-Carpio, J.; Wang, J.; Cortese, L.; Fu, J.; Genzel, R.; Giovanelli, R.; Guo, Q.; Haynes, M. P.; Heckman, T. M.; Krumholz, M. R.; Lemonias, J.; Li, C.; Moran, S.; Rodriguez-Fernandez, N.; Schiminovich, D.; Schuster, K.; Sievers, A. *COLD GASS, an IRAM legacy survey of molecular gas in massive galaxies - I. Relations between H2, HI, stellar content and structural properties.* 2011, MNRAS, 415, 32

[4] **Fabello, S.**; Catinella, B.; Giovanelli, R.; Kauffmann, G.; Haynes, M. P.; Heckman, T. M.; Schiminovich, D. *ALFALFA HI Data Stacking I. Does the Bulge Quench Ongoing Star Formation in Early-Type Galaxies?* 2011, MNRAS, 411, 993

[3] Schiminovich, D.; Catinella, B.; Kauffmann, G.; **Fabello, S.**; Wang, J.; Hummels, C.; Lemonias, J.; Moran, S. M.; Wu, R.; Giovanelli, R.; Haynes, M. P.; Heckman, T. M.; Basu-Zych, A. R.; Blanton, M. R.; Brinchmann, J.; Budavári, T.; Gonçalves, T.; Johnson, B. D.; Kennicutt, R. C.; Madore, B. F.; Martin, C. D.; Rich, M. R.; Tacconi, L. J.; Thilker, D. A.; Wild, V.; Wyder, T. K. *The GALEX Arecibo SDSS Survey - II. The star formation efficiency of massive galaxies.* 2010, MNRAS, 408, 919

[2] Catinella, B.; Schiminovich, D.; Kauffmann, G.; **Fabello, S.**; Wang, J.; Hummels, C.; Lemonias, J.; Moran, S. M.; Wu, R.; Giovanelli, R.; Haynes, M. P.; Heckman, T. M.; Basu-Zych, A. R.; Blanton, M. R.; Brinchmann, J.; Budavári, T.; Gonçalves, T.; Johnson, B. D.; Kennicutt, R. C.; Madore, B. F.; Martin, C. D.; Rich, M. R.; Tacconi, L. J.; Thilker, D. A.; Wild, V.; Wyder, T. K. *The GALEX Arecibo SDSS*

Survey - I. Gas fraction scaling relations of massive galaxies and first data release. 2010, MNRAS, 403, 683

[1] Gavazzi, G.; Giovanelli, R.; Haynes, M. P.; **Fabello, S.**; Fumagalli, M.; Kent, B. R.; Koopmann, R. A.; Brosch, N.; Hoffman, G. L.; Salzer, J. J.; Boselli, A. *HI content and other structural properties of galaxies in the Virgo cluster from the Arecibo Legacy Fast ALFA Survey.* 2008, A&A, 482, 43

Non Refereed Publications

[2] **Fabello, S.**; Kauffmann, G.; Catinella, B.; Li, C,: Giovanelli, R.; Haynes, M. P. *ALFALFA HI Data Stacking III. Comparison of environmental trends in H*I *gas mass fraction and specific star formation rate.* MNRAS, submitted.

[1] **Fabello, S.**; Catinella, B.; Kauffmann, G.; Giovanelli, R.; Haynes, M. P.; Schiminovich, D. *Exploiting HI surveys with stacking.* Proceedings of the ISKAF2010 Science Meeting. June 10 -14 2010. Assen, the Netherlands. Published online.