Gas flows in interacting galaxies: a multiwavelength study

by

Jillian Marie Scudder
B.A., Macalester College, 2009

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Physics & Astronomy

Gas flows in interacting galaxies: a multiwavelength study

by

Jillian Marie Scudder
B.A., Macalester College, 2009

Supervisory Committee

Dr. Sara L. Ellison, Supervisor
(Department of Physics & Astronomy)

Dr. Chris Pritchet, Departmental Member
(Department of Physics & Astronomy)

Dr. Lisa Rosenberg, Outside Member
(Department of Chemistry)

Supervisory Committee

Dr. Sara L. Ellison, Supervisor
(Department of Physics & Astronomy)

Dr. Chris Pritchet, Departmental Member
(Department of Physics & Astronomy)

Dr. Lisa Rosenberg, Outside Member
(Department of Chemistry)

ABSTRACT

A galaxy's evolution is quite sensitive to the impact of external influences. In this thesis, the impact of external environment from both large and small scale effects is investigated, along with a study of how the HI gas fraction of a galaxy can modulate a galaxy's response to perturbations by galaxy–galaxy interactions. This thesis makes use of the statistical power of the Sloan Digital Sky Survey Data Release 7 (SDSS DR7) to assemble a large spectroscopic sample of galaxies, select samples of interest, and select control samples of galaxies matched to each galaxy within the sample of interest in mass, redshift, and (if applicable) local density. It is possible to trace a galaxy's internal gas motions which mark its disturbance by using the metrics of star formation rate (SFR) and gas-phase metallicity.

To investigate the influence of large scale environment, a sample of star forming galaxies in a locally dense environment, but relatively isolated from larger scale structure, is constructed. This sample is further divided into groups which are truly isolated from any large scale structure (no cluster potential within 1 Mpc), and those which, in spite of their relative local isolation, are embedded within a larger cluster structure (within 1 Mpc of a cluster). As the local galaxy density is identical between isolated and embedded group structures, a fair comparison between the star forming properties of the galaxies within those group structures can be made. Star forming

galaxies whose groups are embedded within a larger structure are found to show statistically lower SFRs than those galaxies whose groups are truly isolated from any larger cluster potential.

The impact of local galaxy–galaxy interactions is subsequently considered. Using a sample of star-forming galaxies in pairs from the SDSS DR7, the enhancement in SFRs and the suppression of metallicities is traced as a function of projected separation (r_p). The metallicity dilution as a function of r_p is presented for the first time. Galaxies in pairs are found to have SFRs and metallicity values which are offset from a carefully selected control sample to separations of at least 80 kpc h^{-1}. Using a suite of simulations developed for the purposes of comparison with these observational results, a new interpretive framework is developed for enhancements as a function of r_p.

To investigate the role that gas fraction plays in moderating the strength of interaction triggered starbursts, new data is obtained from the Jansky Very Large Array (VLA). The VLA data supplements the existing SDSS data with HI gas masses for a subsample of resolvable galaxy pairs at small r_p(in kpc h^{-1}). HI masses are obtained and gas fractions are calculated for a sample of 34 paired galaxies. A positive correlation is detected at $> 2\sigma$ between the gas fraction of a galaxy and the SFR enhancement of that galaxy.

The work presented in this thesis has expanded the understanding of physical variables, both internal and external, which can change the star forming properties of a galaxy through an examination of tracers of internal gas flows in those galaxies.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

My thanks go out to:

Mum, Dad, & Matthew for being there for me, and for supporting me through another long scholastic journey.

Keaton, for his constant patience, support and perspective.

Sara, for helping me to achieve more, and for her encouragement and patience.

My fellow grad students, for keeping me company in both the high and the low moments, and making the lows a bit more bearable.

I could not have done it alone.

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I
I took the one less traveled by,
And that has made all the difference.
–Robert Frost
The Road Not Taken

DEDICATION

For Dr. Harvey Israel Scudder,
the first of three generations of Dr. Scudders.

Chapter 1

Introduction

1.1 Galaxies

At their most fundamental level, galaxies are massive, gravitationally bound collections of stars, gas, and dust, embedded within a larger halo of dark matter. As galaxies appear in a wide variety of shapes and sizes, there are a correspondingly large number of ways to describe them. A galaxy's colour can be described, its chemical composition can be analyzed, the number of stars it is currently forming can be deduced, the mass of the luminous portion of the galaxy can be determined, and there are both visual and algorithmic ways of precisely determining a galaxy's shape, among many other parameters. These metrics of describing a galaxy can be plotted against each other in an attempt to find correlations between them, and determine how the average galaxy behaves. In turn, outliers from these correlations allow for populations of unusual galaxies to be isolated. In studying these unusual galaxies, a greater understanding of the remainder can be attained.

1.1.1 Morphological classifications

One of the earliest methods of classifying galaxies was presented in Hubble (1926). Hubble (1926) devised a classification scheme for what were then known as 'extra-galactic nebulae' by visually separating out the thin, spiral–armed galaxies from the roughly spherical, elliptical galaxies. A further division was imposed on the spiral galaxies based on whether or not they had a strong visible bar feature in their centres. This kind of classification became known as the 'Hubble Sequence', or the 'tuning fork' diagram, due to its unique shape. One such diagram, showing both

a schematic view of the galaxies of each classification and a sample image of a real galaxy, is presented in Figure 1.1. Elliptical galaxies were further divided according to their elongation, whereas both kinds of spiral galaxies were further classified according to the tightness of the winding of their spiral arms. Hubble (1926) states that the tuning fork diagram describes an evolutionary process, whereby these 'extragalactic nebulae' acquire more complex structure as they evolve. Therefore, ellipticals, which lack the strong structure of the spiral galaxies, were dubbed 'early types', and spirals as 'late types'.

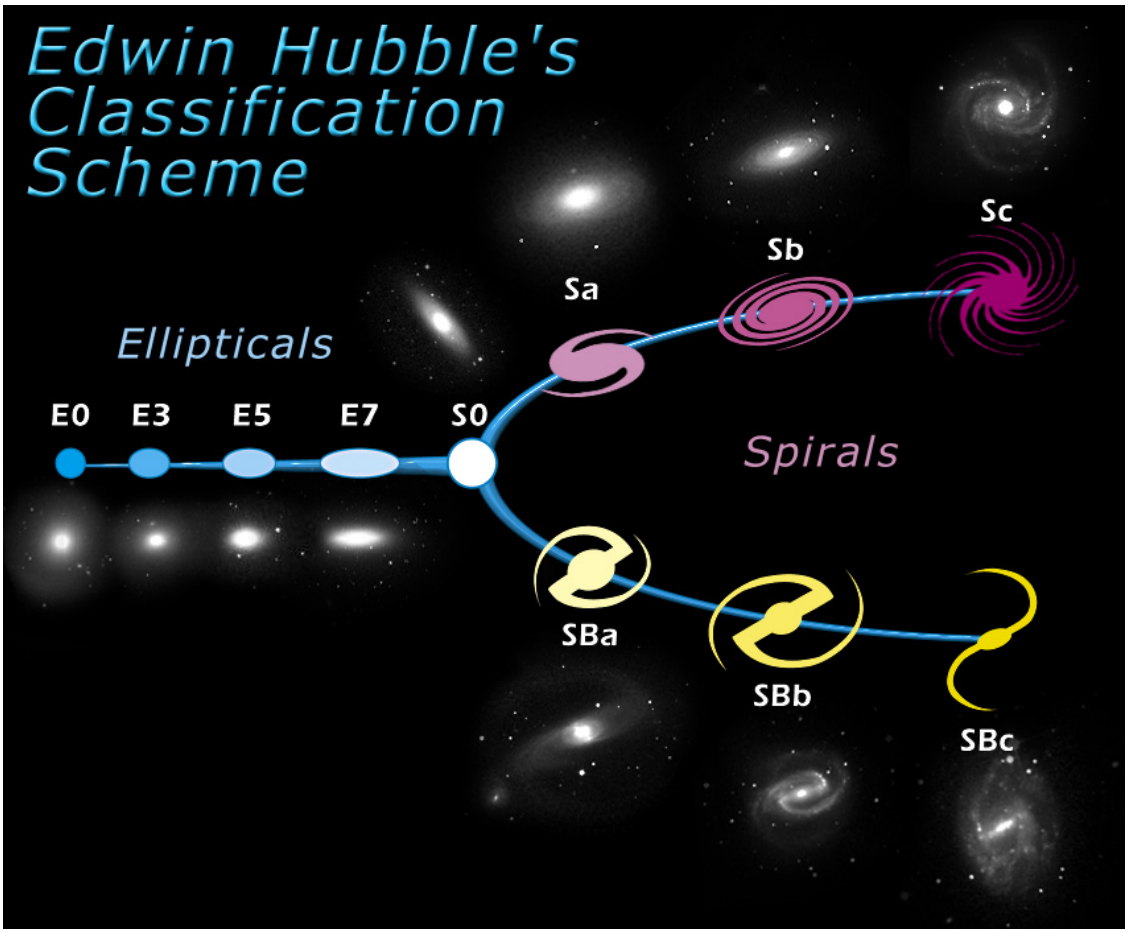


Figure 1.1: The Hubble tuning fork diagram, as classified in Hubble (1926). Ellipticals form the handle of the tuning fork, and are further classified by elongation. Spirals are split into barred and unbarred, forming the two prongs, and are further classified by the tightness of the spiral arms. Image credit: NASA

The evolution presented in Hubble (1926) is now broadly believed to be backwards, and that spiral galaxies are the more fundamental building block used to construct ellipticals. This revision notwithstanding, the nomenclature proposed in Hubble (1926)

of 'early' and 'late' type galaxies referring to ellipticals and spirals respectively is still in use.

1.1.2 Island Universes

Galaxies are seldom truly the 'island universes' they were originally thought to be. However, a small fraction of galaxies can truly be considered part of an isolated 'field' population, free from external influences. Changes to these isolated galaxies are dominated by internal, 'secular' processes (e.g., Kormendy & Kennicutt, 2004, and references therein). With no outside forces impacting them, these galaxies will follow a number of scaling relations based upon their stellar mass.

Galaxy colours are a useful metric to probe a galaxy's current star forming status, as stars of different ages contribute different colours to the observed light of a galaxy. Blue light comes predominately from hot, young, high mass stars with short lifetimes. Red light originates from cooler, lower mass stars with much longer lifetimes. The redder the colour of the galaxy, the older its population of stars, as an increasing fraction of the short-lived blue stars would have disappeared. According to the galactic Colour-Magnitude Diagram (CMD), a galaxy has an increasing chance of being a red, passive, elliptical galaxy in lieu of a star forming spiral as its stellar mass increases. A sample CMD is shown in Figure 1.2, which shows galaxies from the Sloan Digital Sky Survey as described in Patton et al. (2011). This figure is reproduced from Figure 6 of Patton et al. (2011). Points coloured in red describe the red sequence; the density of points in this region is higher than those points coloured in blue, which form the 'blue cloud', lower on the diagram. The blue cloud is comprised of a diffuse population of blue, typically slightly lower-mass galaxies, when compared to the red sequence. The vast majority of galaxies are found populating either the blue cloud or the red sequence. The red sequence contains primarily passive, red galaxies. The high magnitude end of this diagram, which corresponds to the high mass end, is almost entirely contained in the red sequence (Baldry et al., 2004).

The CMD can also be thought of as an evolutionary diagram. The high mass end is almost entirely made of the reddest and least star-forming galaxies; in order to create those high mass galaxies, their mass must be built up over time by other galaxies in the diagram. By definition, these building block galaxies must be less massive. If blue cloud galaxies exhaust their ability to continue forming stars (for instance, by consuming their gas after merging with another galaxy), the resulting drop in the

amount of blue light they produce will push the galaxy vertically towards the red sequence. Further support of the construction of red galaxies over time is found in the evolution of the stellar mass of the galaxies found in the red sequence; the mass in the red sequence has increased by a factor of two since redshift ~ 1 (Bell et al., 2004). Whatever process is driving galaxies from the blue cloud to the red sequence must operate on relatively fast timescales, otherwise the region between blue cloud and red sequence (often called the 'green valley') would be better populated with galaxies in the process of shutting down their star formation (Baldry et al., 2004).

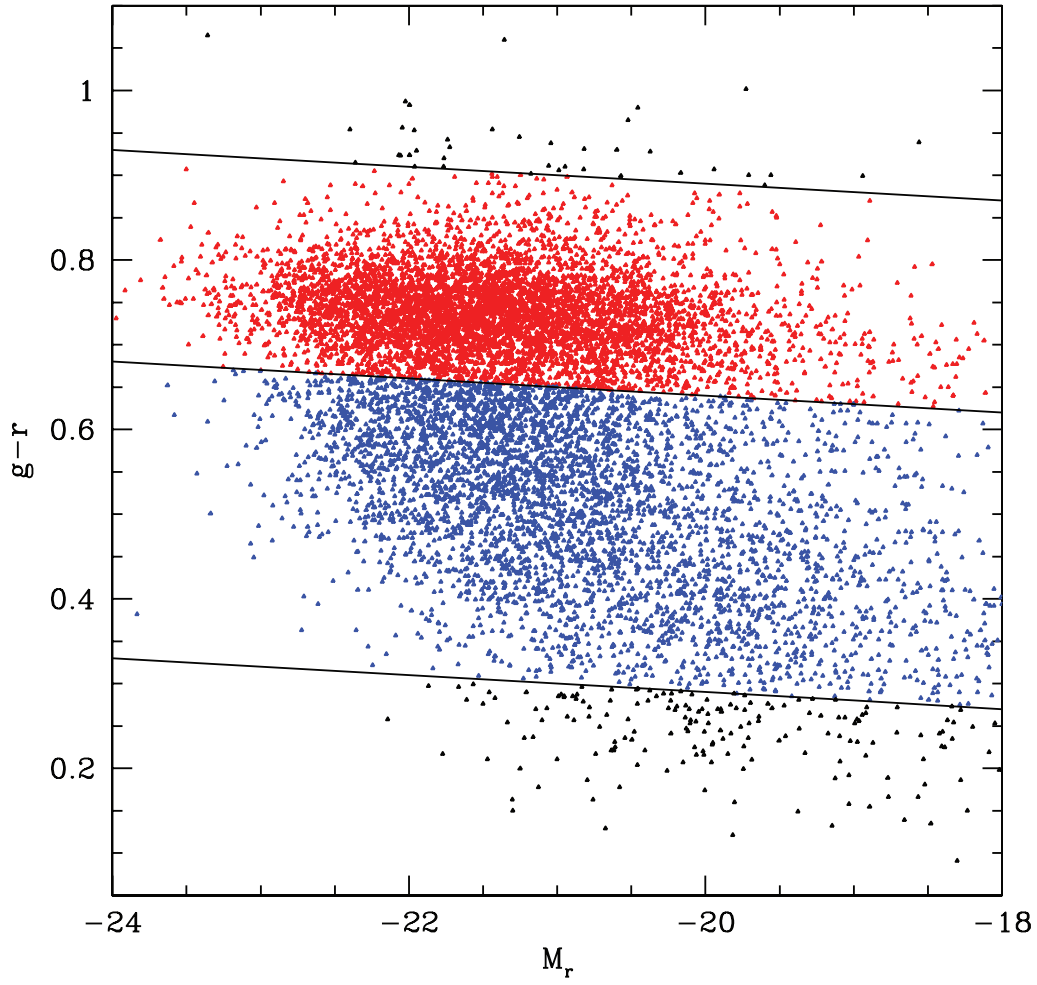


Figure 1.2: Figure 6 of Patton et al. (2011). The Colour Magnitude Diagram for a sample of galaxies taken from the Sloan Digital Sky Survey. The division between red sequence and blue cloud galaxies is given by the middle black line; all red sequence galaxies are plotted in red, and blue cloud galaxies in blue.

If a galaxy is a star-forming, its star formation rate will follow the star forma-

tion rate mass relation, otherwise known as the star formation 'main sequence' (e.g., Noeske et al., 2007b; Wuyts et al., 2011). This relation spans many orders of magnitude in both star formation rate and stellar mass. The relation between stellar mass and star formation rate is thought to arise from an increased total volume of gas within high mass galaxies, relative to low mass galaxies. This higher volume allows a higher star formation rate in the high mass galaxies (Boselli et al., 2001; Brinchmann et al., 2004). Figure 1.3 shows one depiction of the star forming main sequence, adapted from Figure 1 of Wuyts et al. (2011). The primary star forming main sequence is described by the white line. The coloured background points indicate the Sérsic index of the galaxy, which measures the shape of the intensity of light coming from a galaxy as a function of radius. A Sérsic index of 1 indicates a galaxy that is an exponential disk, with a negligible bulge component. A Sérsic index of 4 indicates a galaxy which follows a de Vaucouleurs profile, which describes an elliptical galaxy (de Vaucouleurs, 1948). The galaxies with high Sérsic index, indicative of being primarily bulge galaxies, fall below the star forming main sequence. These are also the galaxies that would fall along the 'red sequence' in the CMD. These galaxies are forming stars at a much lower rate than they would if their mass was in the form of a star-forming disk. These two properties, taken together, form the reason elliptical galaxies are often dubbed 'red and dead': their red colour, along with a star forming 'death'.

In light of the strong dependence of a galaxy on its stellar mass, some studies prefer to use specific star formation rates (SSFRs), which are defined as the star formation rate per unit stellar mass ($SFR/M_{\odot}$). If specific star formation rates are plotted against stellar mass, the star formation mass relation inverts its correlation; high mass galaxies show lower star formation rates per unit mass than low mass galaxies (Zheng et al., 2007). Specific star formation can be used as a slightly more fair way of comparing the star formation rates of two galaxies of different masses, since it is a comparison of the rate at which the two galaxies are forming stars per unit mass, instead of a comparison between the raw star formation rates.

An isolated galaxy's metal content can be described by the mass-metallicity relation (e.g., Lequeux et al., 1979; Tremonti et al., 2004; Moustakas et al., 2011), which, like the star forming 'main sequence' spans many orders of magnitude in stellar mass and slightly over an order of magnitude in metallicity. The mass-metallicity relation (or MZR) shown in Figure 1.4 is adapted from Figure 6 of Tremonti et al. (2004), and shows a sample of galaxies from the Sloan Digital Sky Survey Data Release 2.

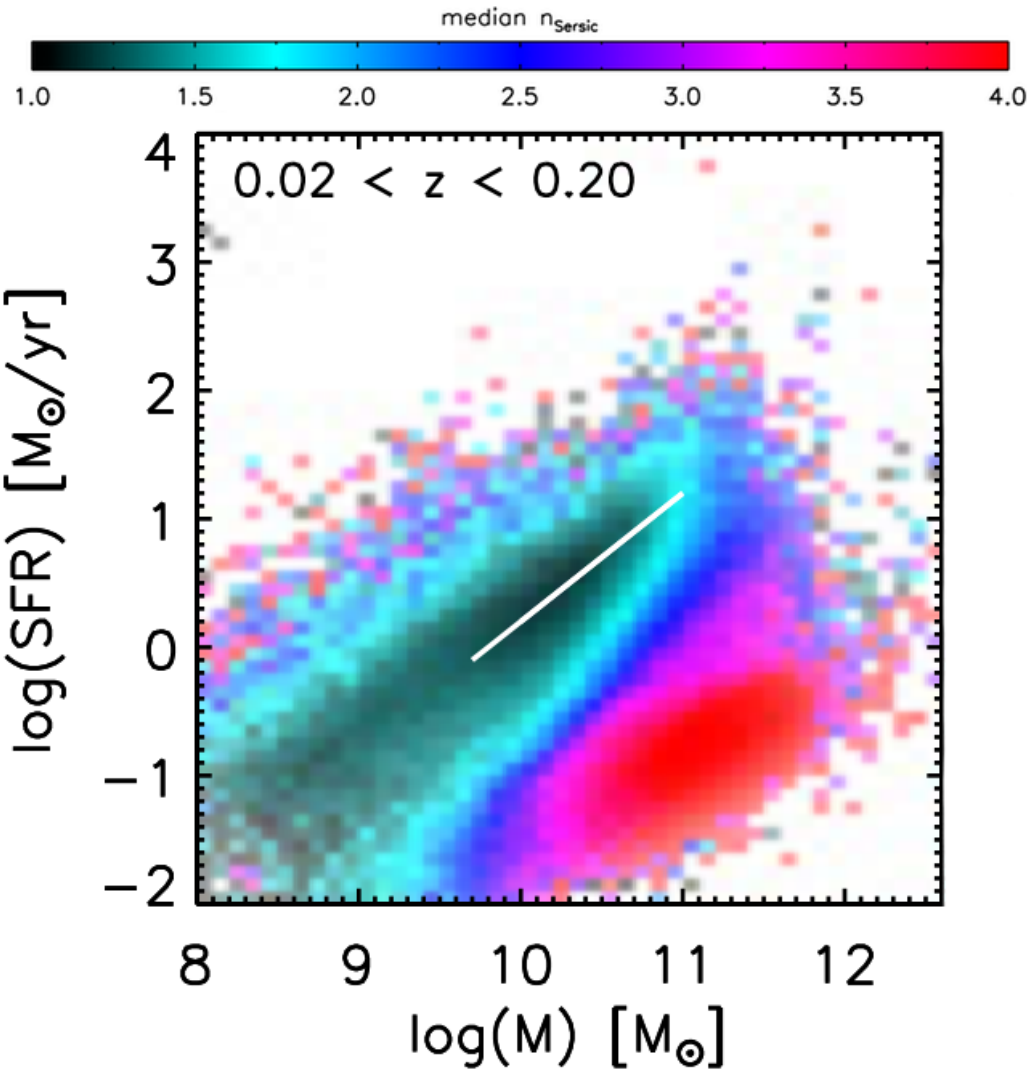


Figure 1.3: Adapted from Figure 1 of Wuyts et al. (2011). The star formation rate-mass relation, also known as the star forming 'main sequence', for galaxies in the SDSS. The colour bar indicates Sérsic index, where 1 is an exponential disk galaxy, and 4 is an elliptical. For those galaxies which are disk-dominated, the relation between star formation and mass is relatively tight, as indicated by the white line. Elliptical galaxies tend to sit below this sequence.

The metal[1] content of a galaxy is a byproduct of stellar evolution, as the metals are produced within the cores of stars, released back into the gaseous medium of the galaxy, either through a stellar wind or in a supernova. A more complete description of metallicities and how they are calculated is presented in Chapter 2.

The MZR indicates that galaxies of low stellar masses are preferentially more metal-poor than their high mass counterparts. While the existence of the mass metallicity relation is well established, the physical processes that create it are still under debate. One explanation for the mechanisms underlying the MZR invokes the varying impact of galaxy scale winds in different stellar mass galaxies as the main driver. The low metal content in low mass galaxies has been suggested to be the result of the shallower gravitational potential in low mass galaxies; a shallower potential would make it easier for gas to be ejected from a galaxy through galactic scale outflows (e.g., Larson, 1974; Tremonti et al., 2004). Galaxies with higher stellar masses will live in correspondingly deeper potential wells, which will require more energy to escape. As a result, the high mass galaxies are efficient at holding on to their metals by trapping the galactic-scale winds that would otherwise transport metals out of a galaxy. Low mass galaxies, by contrast, do not have the same depth to their gravitational potentials needed to trap the metals produced by their stars, and the metals produced are lost from the galaxy.

An alternate explanation for the mass metallicity relation is a difference in the evolutionary stage of galaxies at different masses. High mass galaxies may have begun forming their stars at an earlier time than lower mass galaxies (e.g., Noeske et al., 2007a); this delay in the start of star formation will have kept the gas content of the lower mass galaxies pristine for a longer period of time. Since the low mass galaxies have had a smaller amount of time to enrich their gas content, they would appear to have fewer metals relative to higher mass galaxies. Working in combination with this effect is the prediction from some simulations that lower mass galaxies should have lower star formation rate efficiencies; in other words, that low mass galaxies are less efficient at turning their gas reservoir into stars. This lower level of star formation efficiency results in a lower level of metal production in the lower mass galaxies. (Brooks et al., 2007; Mouhcine et al., 2008). Brooks et al. (2007) finds that in low mass galaxies, the energy of supernovae is sufficient to lower the efficiency with which

[1]In an astronomical context, a 'metal' is any element heavier than helium. Oxygen, as the most abundant 'metal', is frequently used to trace the overall abundance of metals, relative to the abundance of hydrogen.

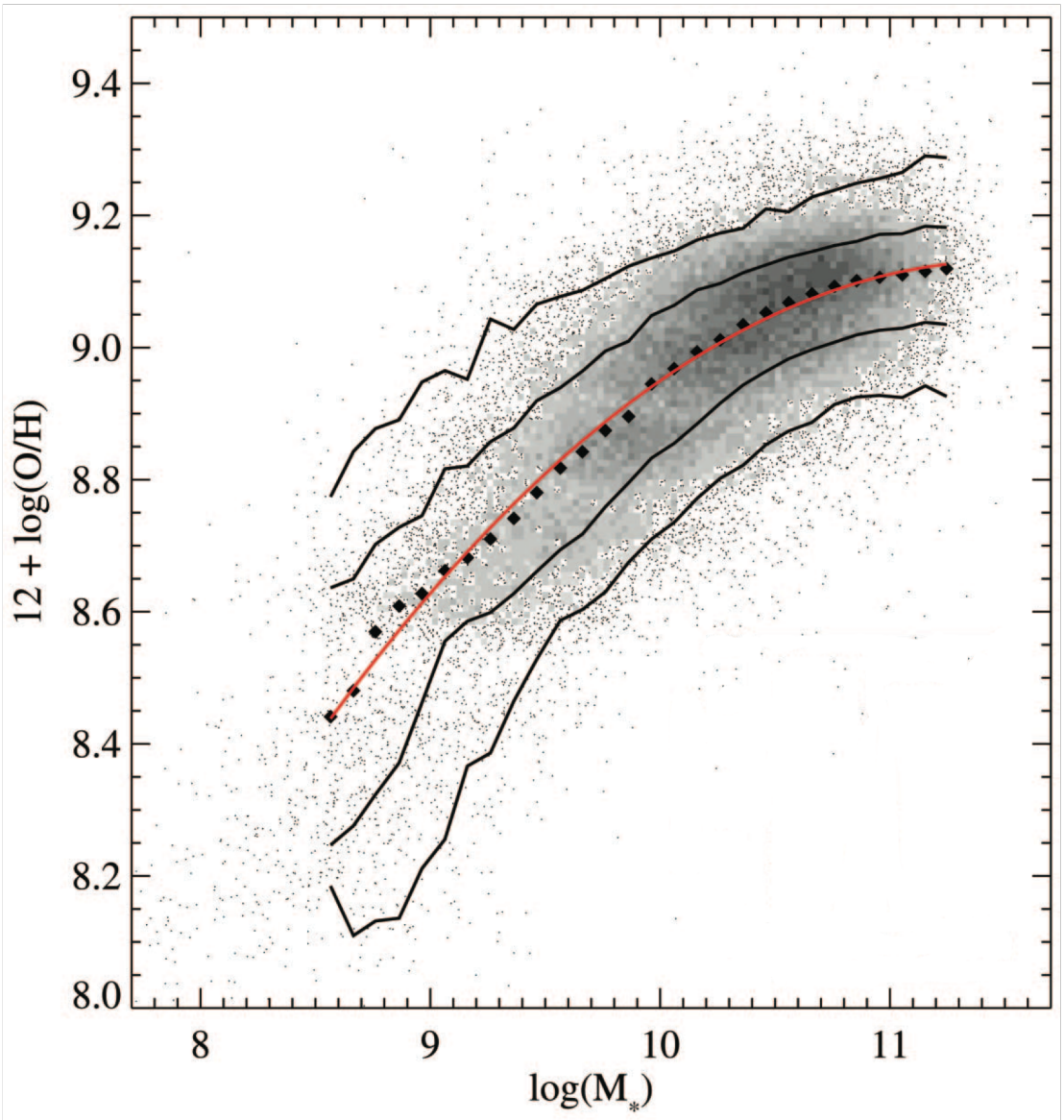


Figure 1.4: The mass-metallicity relation for galaxies in the SDSS DR2, from Figure 6 of Tremonti et al. (2004). Black diamonds show median metallicities as calculated by Tremonti et al. (2004) for that mass bin; the red line shows a polynomial fit to the data. The black solid lines delineate 68% and 95% of the sample.

galaxies form stars. Since the low mass galaxies are forming stars less efficiently than the high mass galaxies, they are producing fewer metals, resulting in the trend of gas phase metallicity with stellar mass.

As a galaxy's mass increases, the chances of it hosting an actively accreting black hole at its nucleus (also known as an Active Galactic Nucleus, or AGN) increases as well (Kauffmann et al., 2003). In Figure 1.5, which is adapted from Figure 5 of Kauffmann et al. (2003), the fraction of galaxies which show significant line emission (dotted histogram) as a function of stellar mass is presented, along with the fraction of galaxies classified as containing an AGN as a function of stellar mass (dashed histogram) and the fraction of the overall population (i.e., the dotted histogram) which appears as an AGN (solid histogram) as a function of stellar mass is shown. Galaxies with stellar masses lower than 10^{10} $M_{\odot}$ only make up a few percent of the total AGN population. By contrast, emission line galaxies with stellar masses of 10^{11} $M_{\odot}$ are 80% AGN. A more detailed description of AGN classifications will be undertaken in Chapter 2.

1.2 The impact of environment

As previously mentioned, isolated galaxies are the exception to the general galaxy population. Most galaxies find themselves either in small groups or clusters, where they cannot evolve in a purely secular manner, avoiding other galaxies entirely. External environmental effects usually play a significant role in the evolution of a galaxy.

1.2.1 Local vs. Global environment

Environment is often roughly divided into two camps: 'global' and 'local'. While the exact definitions of what local and global mean usually differ from study to study (an issue which will be tackled in greater detail in Chapter 3), from a conceptual standpoint, the *intent* of these metrics is usually the same between studies.

Global environment is usually intended to indicate the effects of the environment on scales of $\sim$ 1 Mpc, regardless of the clustering of galaxies on smaller scales. For instance, the impact a galaxy feels simply by being part of a large cluster would be a global environment effect. The local environment, by contrast, is usually a metric of how closely packed the nearest 5 neighbouring galaxies are, and, in principle, measures galaxy clustering on scales significantly smaller than the 1 Mpc commonly used to

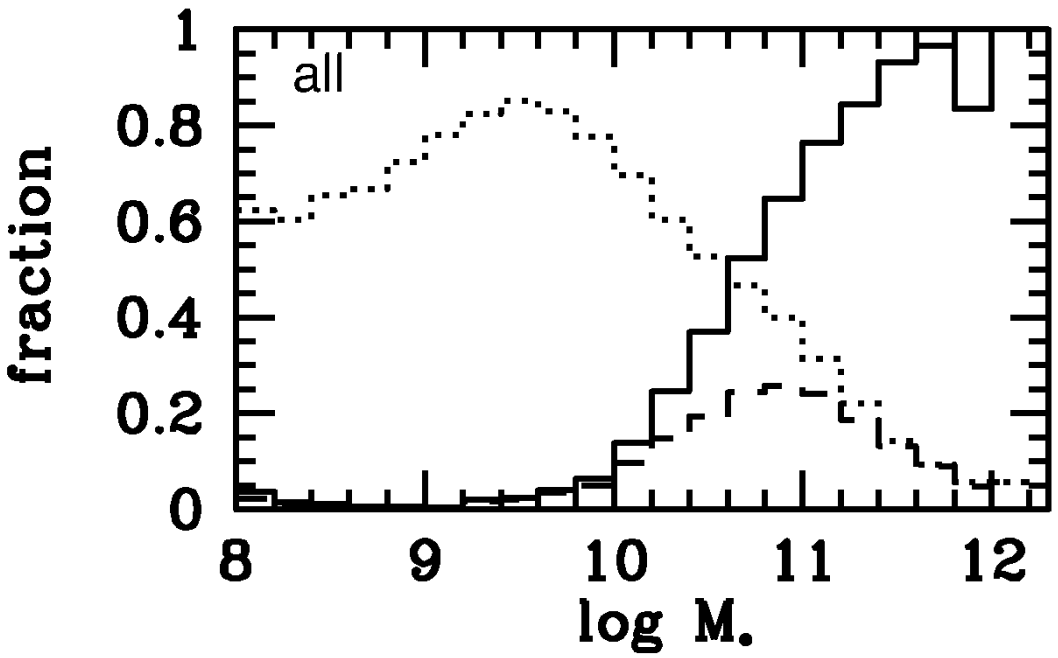


Figure 1.5: This figure is adapted from Figure 5 of Kauffmann et al. (2003). The dotted histogram shows the fraction of all galaxies which have S/N > 3 in the required emission lines for the BPT classification in each bin of stellar mass. The dashed histogram shows the subset of the dotted histogram (galaxies classifiable on the BPT diagram) which are AGN, and the solid histogram shows the fraction of the classifiable emission line galaxies (i.e., the dotted histogram) which are classified as AGN. Above a stellar mass of log $M_* = 10$, the AGN fraction of the galaxy population rises significantly.

describe global environment. Fundamentally, local environment is a probe of how likely a galaxy is to have undergone a recent direct interaction with another galaxy.

1.2.2 The cluster environment

As the galaxy population living in extremely isolated environments is compared to those living in very dense environments such as a cluster, a number of well known relations come into play. One such relation is the morphology–density relation (Dressler, 1980; Postman & Geller, 1984), which states that as the galaxy population moves to higher density environments, the overall fraction of spiral galaxies within that population drops, to be replaced with a much higher fraction of elliptical galaxies. Along a similar line is the colour–density relation, which states that the fraction of blue galaxies declines as higher density environments are probed. The colour–density relation (Baldry et al., 2006; Tanaka et al., 2004; Skibba et al., 2009), like the morphology–density relation, can be explained through an increasingly high fraction of elliptical galaxies as higher density environments are investigated, as the ellipticals which replace the spirals generally have significantly redder colours than the spiral galaxies (Balogh et al., 2004). Figure 1.6, which reproduces Figure 2 of Tanaka et al. (2004), shows one version of the colour–density relation.

A galaxy cluster is usually thought to bring three additional physical effects into play that isolated galaxies do not encounter: ram-pressure stripping, harassment, and strangulation. These additional physical processes are the result of the combination of a high density of galaxies with high relative velocities and the interactions between those galaxies and the cluster medium through which they are travelling.

Ram-pressure stripping

A sufficiently massive galaxy cluster usually contains a hot Intra-Cluster Medium (or ICM), which can act as a physical barrier to the galaxy. The stellar component of a galaxy is usually unaffected by the ICM, as the stars within a galaxy are generally collisionless. However, the gaseous component of a galaxy reacts strongly to the presence of a dense ICM. The pressure of the ICM acts like a very strong wind when combined with the velocity with which the galaxies themselves are moving through the cluster. This wind can cause dramatic changes to a galaxy by forcibly removing the cold gas component of a galaxy. This process of cold gas removal has been dubbed ram pressure stripping (e.g., Gunn & Gott, 1972; Kapferer et al., 2008; Bekki, 2009;

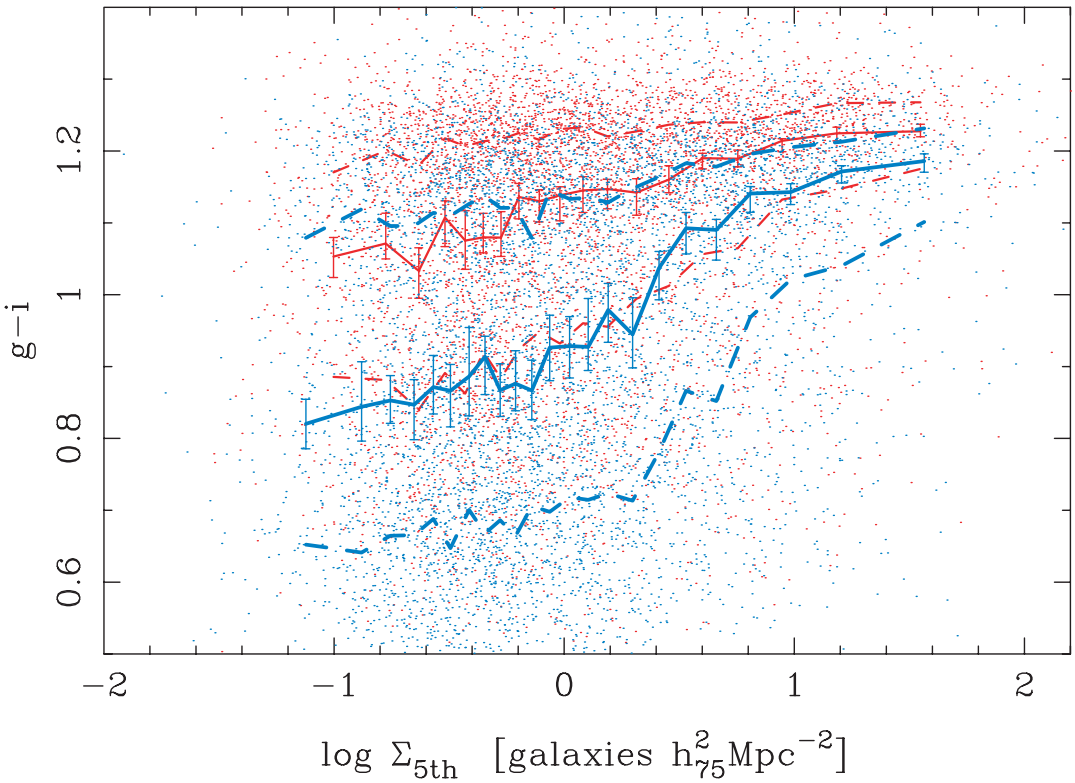


Figure 1.6: This figure is reproduced from Figure 2 of Tanaka et al. (2004). $g - i$ colour is plotted versus density, as measured by the distance to the fifth nearest neighbour. The red and blue lines represent a bright and faint galaxy sample respectively (determined by absolute magnitude measurements). As density increases, the colours of galaxies of all magnitudes tend to become more red.

Tonnesen & Bryan, 2012). Importantly, ram pressure stripping is only effective within a rich cluster that can host a dense ICM. Looser clusters or poor groups that do not have the dense ICM required to effectively remove gas from a galaxy will be unable to induce changes in a galaxy population through ram pressure stripping.

Tonnesen & Bryan (2012) provide a theoretical look into the morphological changes that can be driven by ram pressure stripping; Figure 1.7 reproduces their Figure 11. While this theoretical result is unconstrained by the limitations of observational studies, such as surface brightness and resolution limits, signatures of ram pressure stripping in the nearby dense Virgo cluster have nonetheless been identified (e.g., Chung et al., 2009; Abramson et al., 2011). Disk galaxies in clusters tend to have smaller or displaced HI reservoirs, compared to their optical disk or with galaxies further from the centre of the cluster potential (Chung et al., 2009).

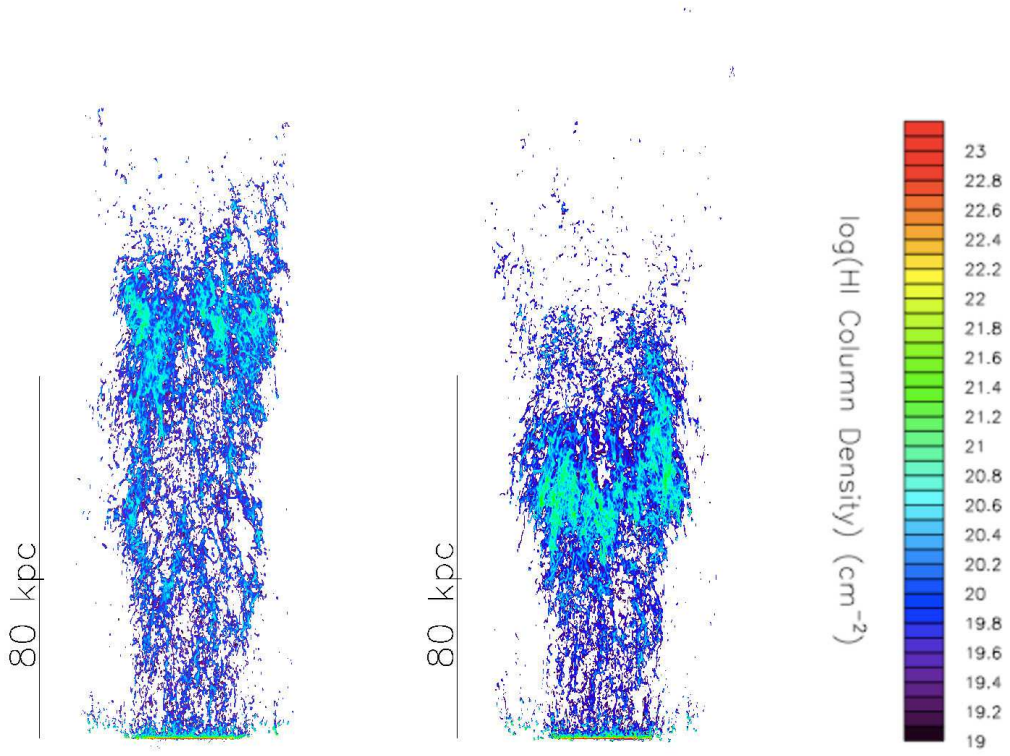


Figure 1.7: Figure 11 of Tonnesen & Bryan (2012). Result of ram pressure stripping for two different galaxy models. Colour corresponds to the HI column density of the galaxy. The disk of the galaxy is found at the bottom of the image.

Harassment

Any cluster with a high velocity dispersion can introduce another force onto the galaxies within it. The high galaxy density present in clusters, in combination with the high velocities with which the galaxies are moving, lends itself to a series of high velocity encounters. A high velocity encounter is unlikely to result in the two galaxies combining to form a single object, but the abrupt tidal interaction between the two can destabilize the internal kinematics of the galaxy, along with a potential shearing off of the external regions of the galaxy. The high velocity encounters a galaxy will undergo as part of cluster membership were dubbed galaxy 'harassment' (Moore et al., 1996).

With a sufficient number of high velocity encounters, the gas content of the galaxy would entirely fall towards the centre of the galaxy, after transferring its angular momentum to the stellar component of the galaxy. Such an energy transfer would efficiently scramble the orbits of the stars in the galaxy, transforming a galaxy that was once a disk-like spiral into something resembling a compact elliptical (Moore et al., 1996, 1998; Fujita, 1998). If the cluster potential is particularly efficient at transforming previously star forming spiral galaxies into gas-exhausted ellipticals, then galaxy harassment may be one of the physical mechanisms driving the morphology–density relation.

Strangulation

The third commonly invoked mechanism by which galaxies in clusters are altered by their dense environment is a process known as 'strangulation' (Larson et al., 1980; Balogh et al., 2000). Strangulation operates under the assumption that most star forming galaxies, if left to produce stars at a consistent rate with no additional source of gas available to them, will run out of gas on timescales of a few Gyr. Therefore, most blue spiral galaxies have not been deprived of additional sources of gas, whether the gas has arrived through minor mergers or ambient filaments of gas, these galaxies have been receiving additional fuel from a diffuse, gas-rich envelope.

Strangulation consists of the removal of the gas-rich envelope of a galaxy (Balogh et al., 2000; Bekki et al., 2002; Kawata & Mulchaey, 2008). The removal of this envelope of gas will slowly shut down star formation in a galaxy by removing the reservoir of gas from which additional gas can be accreted into a galaxy. In this way, the star formation will gradually exhaust the existing reservoir of cold gas in the disk,

and be unable to accrete any further material. While the galaxy's envelope would be removed as the galaxy entered the cluster, its star formation would be able to continue for another ∼few Gyr before quietly shutting down (Larson et al., 1980; Bekki et al., 2002). Since strangulation does not affect the luminous portion of the galaxy, it should not alter the morphology of a galaxy, unlike ram pressure stripping, which leads to an asymmetric distribution of the cold gas within the galaxy, or harassment, which leads to asymmetry in the stellar component of a galaxy. Instead, strangulation should result in a slow shutdown of the star formation within a galaxy, transitioning it from a blue spiral to a red one (Bekki et al., 2002; Kawata & Mulchaey, 2008).

1.2.3 Changes to the galaxy population

The end result of these effects is that, if the entire galaxy population is considered (regardless of the strength of its star formation), in high density environments galaxies tend to have lower average star formation rates than the galaxy population in low density environments. This SFR–density dependence has been measured in a number of studies (Hashimoto et al., 1998; Poggianti et al., 2008; Kauffmann et al., 2004; Cooper et al., 2008) but is considered to primarily be a reflection of the changing galaxy population in different environments.

Quantifying changes exclusively within the star forming population are harder measurements. Some studies find evidence that star forming galaxies form fewer stars if found in dense clusters (Balogh et al., 1998; Pimbblet et al., 2002; Gómez et al., 2003; Welikala et al., 2008), while a similar number find that the star forming population has no dependence on its environmental density (Couch et al., 2001; Balogh et al., 2004; Tanaka et al., 2004; Weinmann et al., 2006; Patiri et al., 2006; Park et al., 2007; Peng et al., 2010; McGee et al., 2011; Ideue et al., 2012); in other words, that the trend of decreasing star formation rates as density increases is entirely due to the changing fraction of star forming galaxies.

The two environmental parameters (global and local) are certainly linked. Dense global environments are often associated with dense local environments as well. A galaxy living inside a cluster is both dealing with the extra physical processes associated with the cluster potential (described above), and with a large number of galaxies in its immediate area. The inverse is not always true; global environments that are extremely isolated do not mandate that the galaxy has few nearby companions.

The exact interplay between the local and global environment remains unclear.

Existing studies have come to discrepant conclusions about the relative importance of the global environment relative to the local environment (e.g., Hashimoto et al., 1998; Carter et al., 2001; Lewis et al., 2002; Goto et al., 2003; Kauffmann et al., 2004; Blanton et al., 2006; Blanton & Berlind, 2007; Park & Hwang, 2009), but for the moment it appears that the local and global environmental effects are too strongly intertwined to be able to cleanly separate them.

1.3 The Impact of Interactions

While the impact of a galaxy cluster's potential on a galaxy can be large, the influence of the nearest neighbouring galaxy as a force of galaxy evolution should not be underestimated. The current model of galaxy evolution, that of hierarchical structure formation (White & Rees, 1978), relies on the buildup and transformation of galaxies through a series of galaxy–galaxy mergers; the final result of a sufficiently low-energy encounter between two galaxies. It is useful to define a few terms which, while they may seem interchangeable, imply a different set of physical conditions within the timeline of a galaxy–galaxy interaction.

A galaxy which is undergoing a gravitational perturbation due to the presence of another nearby galaxy is said to be *interacting*. The second galaxy in the pair is generally called the *companion* to the galaxy in question, regardless of its mass. A galaxy *merger* can be used to refer to a set of two galaxies which will ultimately merge into a single object. This definition is most useful for theoretical studies, as the simulations can be set up such that the galaxies will merge, a luxury unavailable to the observers. Therefore, observationally, a merger is usually limited to galaxies which are currently in the final stages of combining into a single object. This final stage of merging galaxies is also referred to as *coalescence*. Other commonly used terms for particular stages of an interaction are pericentre (the closest separation that a set of interacting galaxies has reached), and apocentre (the furthest separation the two galaxies will reach after a close pass).

Finding galaxy *pairs* is the best observational method to select a sample of interacting galaxies. These are galaxies which appear close to each other on the sky, and may be at any stage of an interaction. They may be physically bound systems, but it is impossible to determine if the galaxy is before a first tidal encounter, after a first encounter, or approaching coalescence. There is also a significant fraction of projected pairs, which are galaxies which are not physically associated with each

other, but merely appear close together on the sky (e.g., Patton et al., 2000; Patton & Atfield, 2008). The fraction of these interloping projected pairs can be limited by requiring that the relative velocities of the two galaxies be relatively small, but even for galaxies at small separations and low velocities (i.e., $r_p < 20$ kpc h^{-1} and $\Delta v < 500$ km s^{-1}), the projected pair fraction can be 50% (Patton & Atfield, 2008).

The first collection of interacting galaxies was compiled in Arp (1966), into what is known as Arp's Catalogue of Peculiar Galaxies. Arp (1966) contained a sample of particularly dramatic, nearby, perturbed galaxies. These galaxies showed strong morphological changes when compared to the standard structure of a galaxy. Wide tidal arms, shells, and other morphological distortions were clearly visible in the selected 'peculiar galaxies'. Explaining these images became the basis of the theoretical work of Toomre (1977). Toomre (1977) selected 11 of the galaxies presented in Arp's Catalogue to represent a timeline of how the morphology of a galaxy changes over the course of an interaction and subsequent merger with another galaxy (Figure 1.8[2]). The Toomre Sequence, as this timeline became known, was the first attempt to understand how galaxies may evolve through the course of an interaction.

Not only is the galaxy's morphology dramatically altered in the course of the interaction, but a number of other observational properties are altered in galaxy pairs by their companion. Star forming galaxies in pairs universally show both bluer than average in their optical colours (Larson & Tinsley, 1978; Patton et al., 2011) and higher than average star formation rates, both effects which are primarily centrally concentrated (e.g., Donzelli & Pastoriza, 1997; Barton et al., 2000; Lambas et al., 2003; Alonso et al., 2004, 2006; Woods & Geller, 2007; Ellison et al., 2008b, 2010; Darg et al., 2010; Xu et al., 2010).

A galaxy in a pair is also generally found to have a central gas-phase metallicity which is lower than expected for its mass and redshift (Kewley et al., 2006; Ellison et al., 2008b; Michel-Dansac et al., 2008). This is explained through an overall flattening of the normally steep metallicity gradient in spiral galaxies. As gas flows to the centre from the outer regions, the lower metallicity gas from larger radii dilutes the metallicity of the gas at the centre of the galaxy. This flow of gas means that the entire metallicity gradient of the galaxy flattens as the interaction progresses (Rupke et al., 2010b; Kewley et al., 2010; Perez et al., 2011).

[2]Taken from: `http://burro.cwru.edu/Talks/GLC03/toomre.html`

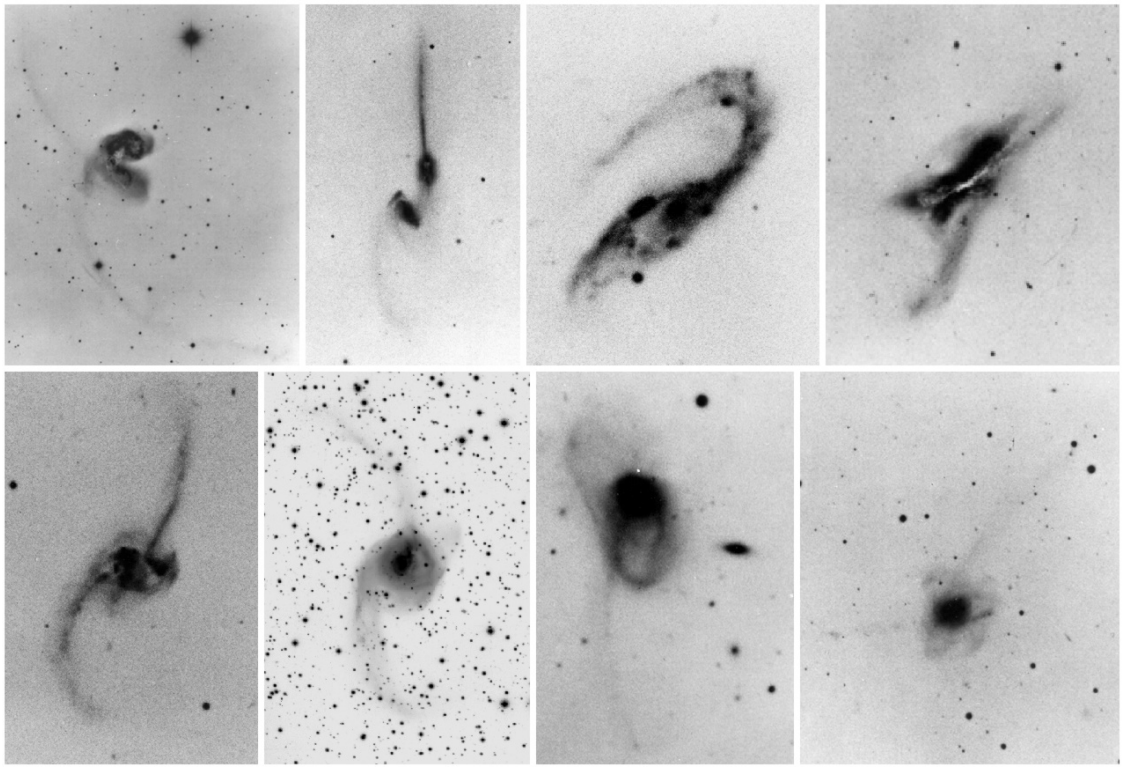


Figure 1.8: The Toomre Sequence, illustrated through the images of Arp (1966). Interacting galaxies progress from left to right across the top row of images, and then along the bottom row as their nuclei merge.

1.3.1 A theoretical understanding

Theoretical simulations have improved dramatically since the first attempts to simulate a galaxy interaction, and the current understanding of the internal machinations of perturbed galaxies is largely derived from a careful comparison between the simulations and observational results.

Toomre & Toomre (1972) showed, with simple gravitational arguments, that the extended features remarked upon by Arp (1966) can be explained as the result of tidal gravitational forces exerted during galaxy interactions. Many further simulations have reinforced the idea that the stellar arms visible in perturbed galaxies are the results of strong gravitational tidal forces, lending the structures the name of 'tidal arms' (e.g., Barnes, 1988; Barnes & Hernquist, 1996; Mihos et al., 1998; Di Matteo et al., 2007; D'Onghia et al., 2010). With the advent of more detailed simulations, a more thorough theoretical understanding of the internal kinematics of a galaxy as it is perturbed is now available.

Gas Flows

The majority of the observational perturbations to a galaxy as the result of an interaction are currently understood to be the result of a large scale tidal perturbation of the galaxy, which results in bar instabilities developing in the nucleus of the galaxy. (Barnes & Hernquist, 1996; Mihos & Hernquist, 1996; Cox et al., 2006; Di Matteo et al., 2007; Montuori et al., 2010; Rupke et al., 2010a; Torrey et al., 2012) These bar instabilities are the response of the galaxy to a non-axisymmetric gravitational potential; the nearby gravitational potential of the companion galaxy means that the galaxy is no longer sitting in a smooth potential well. Both the gas and the stars within a galaxy create bar instabilities in response to the perturbation. However, the gaseous bar and stellar bar are usually misaligned by a few degrees, with the stellar bar lagging slightly behind the gaseous bar (Barnes & Hernquist, 1996; Hopkins et al., 2009). The stellar bar is effectively non-collisional; even in the centre of the galaxy, the density is not high enough for stars to collide. Unlike the stellar component, the gaseous component of a galaxy is highly collisional, and due to the misalignment between the gaseous and stellar bars, a torque is exerted on the gas by the stars. The gaseous component transfers energy to the stellar bar, which decreases the angular momentum of the gas. This efficient loss of angular momentum due to the torque exerted by the stellar bar misalignment means that the gas can no longer support its

previous orbit. Lacking the angular momentum to stay at its current radius, the gas flows towards the centre of the galaxy (Barnes & Hernquist, 1991, 1996).

This gas flow to the centre results in a large volume of gas collecting near the nucleus of the galaxy. As expected for a large volume of dense gas, this gas then begins to turn into stars. A starburst as a result of gas inflow is a consistent result of the simulations (Barnes & Hernquist, 1996; Mihos & Hernquist, 1996; Cox et al., 2006; Di Matteo et al., 2007; Montuori et al., 2010; Rupke et al., 2010a; Torrey et al., 2012) and adroitly provides a physical mechanism to explain the observational results of nuclear starbursts, and corresponding shift towards bluer colours (e.g., Patton et al., 2011), in galaxies. This same physical mechanism also explains the observation of lower than expected nuclear metallicities (e.g., Kewley et al., 2006; Ellison et al., 2008b; Michel-Dansac et al., 2008). Given that the average spiral galaxy shows a negative gas-phase metallicity gradient with increasing metallicity (i.e., gas phase metallicity decreases as the distance from the centre increases; Rupke et al., 2010b; Kewley et al., 2010; Perez et al., 2011), any gas flowing to the centre from a larger radii will be a flow of low-metallicity gas, relative to the nuclear metallicity. This is observed in a number of simulations (e.g., Torrey et al., 2012); the evolution of the metallicity gradient is modelled explicitly by Rupke et al. (2010a).

An example of interaction-induced changes within a galaxy over the full course of a merger is shown in Figure 1.9, where Figure 8 of Torrey et al. (2012) is reproduced. This figure shows the change in nuclear (defined as a 1 kpc h^{-1} sphere) metallicity, the mass flux over the same region, and the nuclear star formation rate, along with the separation between galaxies, as a function of time. The morphology of the galaxy's gas density is shown for five different points in time along the top of the figure. The mass within the nucleus begins to increase quickly after the first passage, with a corresponding drop in metallicity, while the star formation rate peaks at slightly later times. As the galaxies approach coalescence, the nuclear mass flux, star formation rates, and metallicities are all more strongly affected than after a close passage.

Modulating effects

The effectiveness of the tidal perturbation depends on the orbit of the two galaxies. The closer together these galaxies are at their closest passage, the stronger the tidal impulse. Furthermore, the relative rotations of the galaxies can play a strong role. If the two galaxies are both rotating in the same direction relative to each other, the

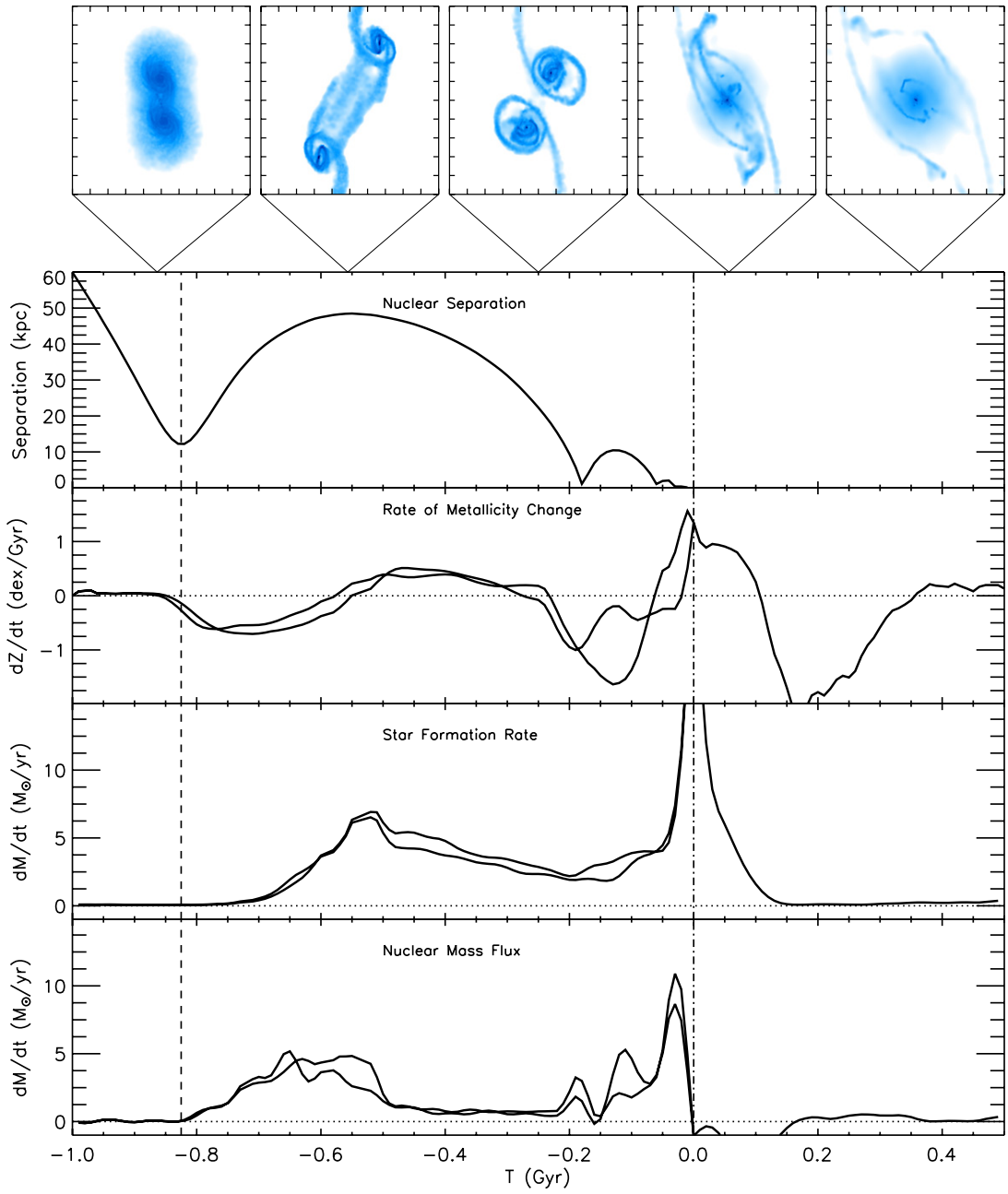


Figure 1.9: This figure reproduces Figure 8 of Torrey et al. (2012). The top five images indicate the morphology of the gas density at different time steps. The four panels below the images are the galaxy–galaxy separation, change in gas-phase metallicity, star formation rate, and nuclear mass flux as a function of time, determined over a sphere of 1 kpc h^{-1} in all panels.

effectiveness of the merger's ability to trigger star formation can be a factor of two lower than if the galaxies were rotating in opposite directions (Di Matteo et al., 2007; D'Onghia et al., 2010). If the galaxies pass very near to each other at the first close pass, the tidal forces can be so extreme that gas is preferentially scattered outwards into tidal arms, reducing the amount of gas in the main part of the galaxy. This can result in a significantly weaker starburst being triggered at coalescence, as less gas is available to be turned into stars (Di Matteo et al., 2007).

The strength of the tidal perturbation is also partially driven by the relative masses of the two galaxies in the interaction. Both in theoretical and in observational works, equal mass interactions drive the strongest gas inflows (e.g., Cox et al., 2006; Di Matteo et al., 2007; Woods & Geller, 2007; Ellison et al., 2008b). However, the expected level of star formation in galaxies in very unequal mass interactions is still uncertain. Very few simulations have been run to determine the expected starburst strength, and those that have, vary the gas fraction and the mass ratio simultaneously, making the effect of the uneven mass ratio difficult to disentangle (Cox et al., 2006).

1.3.2 The physics of coalescence

Three physical mechanisms drive the rapidity with which the galaxies merge together: dynamical friction, violent relaxation, and tidal stripping. In the next few sections, a brief overview of each of these physical processes is presented.

Dynamical Friction

Dynamical friction was first proposed by Chandrasekhar (1943) as a mechanism of slowing the motion of a star as it moves through a stellar cluster densely populated with other stars. Chandrasekhar (1943) proposed that dynamical friction should be considered the result of a series of small, slowing, gravitational tugs exerted upon the star of interest by its neighbouring stars. Chandrasekhar (1943) noted that only stars travelling slower than the target star will contribute to this slowing effect, as they will be marginally accelerated by the gravitational pull of the faster target star. Under the assumption that there is no particular distribution to the neighbouring stars, the direction of the target star's travel should not be affected by this force. Over time, each of these small gravitational tugs backwards on the target star will gradually cause the target star to lose energy, donating that energy to the acceleration of the slower stars with which it is interacting.

The same principle can be applied to entire galaxies, instead of simply to individual stars. The physics involved is identical; the change is simply that instead of one star of interest moving through a field of stars, all of the stars in the galaxy feel the gravitational pull from the field of stars in the disk of the other galaxy. Since the galaxies have relative motions of hundreds of km s^{-1}, a large fraction of the stars in the companion will appear to be moving at a slower rate than the stars in the incoming disk. Each star in a galaxy that acts as a braking force on the stars in the other galaxy will gain a small amount of energy as a result of the energy transfer, but contribute to an overall loss of energy from the other system.

The dynamical friction force can be approximated with the following form:

$$f_{dyn} \approx C\frac{G^2M^2\rho}{v^2} \tag{1.1}$$

C is a constant, G is the gravitational constant, M is the mass of the target particle, v is the velocity of the target particle, when the cloud through which it is moving is considered to be at rest, and ρ is the density of the field through which the target particle is moving.

The efficiency with which dynamical friction operates is directly dependent on the masses of the galaxies involved in the interaction. The more equal mass the galaxies are, the faster the two galaxies will merge into a single galaxy, as the efficiency of dynamical friction increases. A small galaxy will take much longer to lose energy through dynamical friction; an unequal mass interaction will take a significantly longer time to coalesce than an equal mass interaction. Dynamical friction also depends on the density of stars through which the perturbing galaxy must pass; very dense host galaxies increase the efficiency with which energy can be transferred, as more stars will exert a force on the other galaxy with each close passage. Further, as dynamical friction is fundamentally a gravitational effect, the slowing of the galaxies' orbits around each other will be more effective the closer the galaxies come to each other in each of their close passages.

Violent Relaxation

The principle of violent relaxation was introduced some time later, by Lynden-Bell (1967). Most isolated galaxies live in gravitational potential wells that are both relatively stable as a function of time, and symmetric in every direction. However, when a galaxy encounters another galaxy, this description of the potential well is no

longer valid. The gravitational potential of the companion drives a strong asymmetry in the overall gravitational potential through which both galaxies are travelling. Since both galaxies are moving relative to each other, this potential is also strongly evolving as a function of time.

As stars orbit in a changing gravitational potential, they will not be able to maintain the same ordered rotation that they could in the stable, spherical potential of an isolated galaxy. As the stars adapt to the new shape of the gravitational potential, their orbits become increasingly randomized. This scrambling of stellar orbits can change the stars in a galaxy from proceeding according to the ordered rotation of a spiral galaxy to the random orbits of the stars in an elliptical galaxy. The scrambling of the orbits of stars within a galaxy due to a changing gravitational potential is the process of violent relaxation. It is through the mechanism of violent relaxation that elliptical galaxies were thought to be built from spiral galaxies. Indeed, early simulations supported this idea; Hernquist (1992) found that it was possible to produce nearly pure ellipticals as a result of a merger between two disk galaxies, since the merger event scrambled the previous orbits of the stars into something much less ordered.

Tidal Stripping

The third factor that comes into play with the close passage of two diffuse objects is tidal stripping. Tidal stripping consists of the removal of material from the outermost radii of a satellite galaxy as it passes by another, larger host galaxy. When the tidal force exerted by the host can overcome the gravitational binding energy holding that material to the satellite galaxy, that material is lost from the satellite to the host.

Since the tidal effect is driven by a gravitational differential in forces, the tidal forces exerted depend primarily on the masses of the two galaxies, and the distance between them. The closer the galaxies come together, the stronger the tidal effect felt by the galaxies. These tidal forces cause the growth of distortions in the galaxies' morphologies.

The tidal force felt by the satellite as it orbits the host can be expressed as the difference in gravitational pull between the centre of the satellite and the edge of the satellite, where a test particle u is located. The tidal force can be expressed as:

$$F_T = \frac{2GM_h ur}{R^3} \tag{1.2}$$

G is the gravitational constant, M_h is the mass of the host itself, r is the radius of the satellite, and R is the distance between satellite and host. If M_s is the mass of the satellite galaxy, the gravitational force binding the test particle u to the satellite can be written as:

$$F = \frac{GM_s u}{r^2} \tag{1.3}$$

If the gravitational force F binding the test particle u is less than the tidal force F_T exerted by the companion, u will become unbound. To determine the radius within which a satellite will not lose material due to tidal stripping, F and F_T are set equal.

$$\begin{aligned}
\frac{GM_s u}{r^2} &= \frac{2GM_h ur}{R^3} \\
\frac{M_s}{r^3} &= \frac{2M_h}{R^3} \\
r^3 &= R^3 \frac{M_s}{2M_h} \\
r &= R\left(\frac{M_s}{2M_h}\right)^{1/3}
\end{aligned}$$

Any material at a distance from the centre of the satellite less than radius r is stable, and will not be stripped from the galaxy while the galaxies are at distance R from each other. Any material orbiting at radii greater than r from the satellite will be lost. Tidal stripping can be responsible for causing bridges of material to develop between two interacting galaxies, but will be particularly effective at consuming galaxies which are either very diffuse or very close to the host galaxy.

1.3.3 Merger remnants

The galaxy which remains as the end result of a merger has long been held to be an elliptical galaxy (e.g., Hernquist, 1992; Rossa et al., 2007), which points to galaxy-galaxy interactions as a major driver of galaxy evolution from blue, star forming galaxies to red, dead elliptical galaxies. In support of the theoretical models suggesting that ellipticals are built up from bluer spirals, many elliptical galaxies are found to host faint shells of material surrounding them when imaged with sufficiently deep observations (e.g., Malin & Carter, 1983; Atkinson et al., 2013). These shells of

material are thought to be the remnants of previous mergers that have not quite dissipated. Galaxy–galaxy mergers have also been used to help explain the rather sharp dichotomy between the red sequence and the blue cloud, as defined by the galactic CMD. Very few galaxies populate the green valley (as the region of overlap between the two has been dubbed), indicating that galaxies must move from the blue cloud to the red sequence quickly (Baldry et al., 2004).

The Toomre sequence shown in Figure 1.8 was originally meant to give a visual example of how a spiral galaxy could be naturally transformed into an elliptical using observed galaxies, but new simulations now suggest that not all galaxy mergers result in an elliptical remnant. It has been suggested that with a sufficiently high gas fraction, galaxies could remain as star-forming disk galaxies even after a merger, either by maintaining their disk throughout the interaction, or by rebuilding a disk after the interaction is complete (e.g., Hopkins et al., 2009).

Regardless of the whether or not a galaxy is able to re-build a disk component to the galaxy or the disk is destroyed, any merger should strengthen the bulge component of a galaxy. The processes of violent relaxation and dynamical friction will be effective no matter what the gas content of a galaxy is, and will serve to randomize the orbits of some fraction of the stars within the galaxy. These stars with randomized orbits will construct or build up the pressure-supported bulge component of the galaxy. If the galaxy is gas rich, and manages to reconstruct a disk component, the end result of the merger will not be completely transformed morphologically from spiral to elliptical, but will instead increase the fraction of mass in the bulge component (Kannappan & Fabricant, 2001; Kannappan et al., 2004).

1.3.4 Duration of interaction-triggered changes

While simulated studies of galaxy interactions generally use time as the fundamental parameter by which to track changes within the galaxy, observational studies of galaxy interactions with respect to time are impossible. As each galaxy interaction is, to an observer, frozen in time, the only way to compile a timeline of an interaction is statistically, through other metrics.

The observational tracer of choice is projected separation (r_p). Since time is unavailable for an observational study, the next best thing is to observe distance; in principle, the further the distance travelled, the more time the galaxy must have spent travelling to get there. Observational studies are unable to measure the proper

distance between two objects, but again are limited to what is directly observable: the separation between two galaxies as projected upon the plane of the sky. This projection effect means that galaxy pairs are generally observed to be closer in projection than they would be if an objective measurement could be made.

Even given the limitations of projected separations, on average it is still the case that if a galaxy has had a previous close encounter, larger separation galaxies have likely travelled longer to reach those separations than galaxies which appear close together on the sky. (Projected pairs and galaxies which are moving towards a first encounter will not follow this logic.) As a result, observational studies of the colours or star formation rates of galaxy pairs which intend to determine the length of time over which a galaxy is perturbed by an encounter with a companion have universally used projected separation as their metric of determining timescales.

Under the assumption that a galaxy pair is not moving along the plane of the sky, the relative velocities of the galaxies can be used to estimate a timescale for distance. The conversion between km s^{-1} and distance per million years works out to be approximately 1 kpc h^{-1} = 1 pc / 1 Myr. This means that a galaxy moving at 300 km s^{-1} should move 300 pc/Myr, or 30 kpc h^{-1} per 100 Myr. A galaxy moving at 100 km s^{-1} will move 10 kpc h^{-1} in 100 million years.

Most existing studies are in relative agreement that galaxies show significant enhancements of their star formation rates at small separations. Usually these enhancements are present at separations of $r_p <$ 30 kpc h^{-1}, which smoothly decline to a value in agreement with the average galaxy population at separations beyond 30 kpc h^{-1} (Lambas et al., 2003; Alonso et al., 2006; Nikolic et al., 2004; Ellison et al., 2008b; Li et al., 2008). This trend with projected separation has traditionally been interpreted as a sharp spike in star formation immediately after a close passage, with the strength of this starburst dwindling as the galaxies become more widely separated (e.g., Barton et al., 2000; Barton Gillespie et al., 2003; Nikolic et al., 2004; Woods et al., 2010). This interpretation further indicates that interacting galaxies return to their fiducial values very quickly; the interaction is therefore not driving long-term changes within the galaxy.

There are a few existing studies which suggest that the effect of a galaxy interaction ought to continue to slightly wider separations, although this is extended only slightly further, to 50 kpc h^{-1} (e.g., Barton et al., 2000; Lin et al., 2007; Robaina et al., 2009; Wong et al., 2011). These studies found a similar trend of gradual decline to an average value, but progressing slightly slower down to the average value. Patton

et al. (2011), however, in investigating the colours of galaxies in pairs, found that galaxies found in the blue cloud had significantly bluer than average nuclear colours to separations of 80 kpc h^{-1} (the widest separations in the sample of galaxy pairs), which would represent a significant lengthening of the duration of interaction-induced changes to a galaxy's internal processes.

1.4 Thesis Goals

In this thesis, new research covering both the impact of large scale environment and the impact of local interactions on samples of star forming galaxies is undertaken. This work is furthermore extended into radio wavelengths to investigate the impact of the gas content, as measured through HI masses, on the star formation rates of a sample of galaxies in close pairs. The basic structure of this thesis is outlined below.

Although the perturbations of a galaxy can be measured through many observational parameters, in this thesis, the main metrics will be star formation rates and metallicities. As both Chapter 3 and Chapter 4 rely upon the star formation rates and metallicities of the galaxies in their respective samples, in **Chapter 2**, the data from the Sloan Digital Sky Survey Data Release 7 used in both chapters will be described, along with the quality control requirements imposed upon the data.

In **Chapter 3**, an attempt to determine whether or not the presence of a large scale cluster potential has any impact upon a galaxy's star formation rates and metallicities is undertaken. By selecting a very particular local environment, the interdependence of local and large scale density should be disentangled, allowing for a unique determination of the impact of a nearby cluster on a star forming galaxy.

In **Chapter 4**, the duration of the alterations in star formation and metallicity induced by an interaction is investigated, using projected separation as a proxy for timescale. The extent of changes induced in the gas phase metallicities as a function of r_p is plotted for the first time, and a novel interpretation of trends with r_p is suggested after a comparison with new simulations.

Chapter 5 presents new data from the Jansky Very Large Array radio interferometer. Based on the results of Chapter 4, HI gas fractions are obtained for a subsample of galaxy pairs to determine whether the gas fraction of a galaxy governs the level of star formation enhancement. The data reduction steps, along with the subsequent analysis, are described.

Finally, in **Chapter 6**, the main results of this thesis are summarized, along with a brief discussion of possible expansions for future work.

Chapter 2

Sloan Digital Sky Survey: Data Analysis

2.1 Roadmap

In Chapters 3 and 4 that follow, both star formation rates and metallicities will be used as important tracers of gas motions within the galaxy samples. Robust star formation rate values are already publicly available in existing catalogues, so this thesis will not re-calculate these values. However, the metallicity values from the preferred calibration are not pre-existing in a public catalogue. In order to obtain a reliable catalogue of metallicities for as many galaxies as possible, the first step of this thesis work is to create such a catalogue from a large galaxy sample. The spectroscopic catalogue from the Sloan Digital Sky Survey Data Release 7 (henceforth SDSS DR7, and described in more detail in Section 2.2) is therefore used as a starting point. Using emission line fluxes available in a publicly available catalogue, these metallicities may be calculated after a series of quality control steps. In bullet point form, these steps are:

- Eliminate unreliable flux measurements
- Apply dust corrections
- Eliminate Active Galactic Nuclei contamination
- Eliminate duplicate objects from the sample
- Calculate metallicity values

Each of these steps will be described in detail later in this chapter before undertaking the metallicity calculation. As these metallicities will be used for multiple samples of galaxies, metallicities are calculated for all possible galaxies. The sample of galaxies with calculated metallicities will then be able to be used as a master sample from which subsamples of interest can be drawn.

2.2 Survey Basics

The Sloan Digital Sky Survey (York et al., 2000) consists of a photometric and spectroscopic survey of over 8000 square degrees of the northern sky. All data in the SDSS was taken by the Sloan 2.5 metre optical telescope at Apache Point Observatory in New Mexico. The Sloan telescope surveyed the sky between 2000 and 2008, with regular public updates released by the survey team (Stoughton et al., 2002; Abazajian et al., 2003, 2004, 2005; Adelman-McCarthy et al., 2006, 2007, 2008). The final data release, Data Release 7 was released in 2009 (Abazajian et al., 2009). The SDSS DR7 contains imaging of over 230 million objects and spectra for 1,374,000 objects, of which 928,567 are classified as galaxies. The average redshift of the main galaxy sample is $z \approx 0.1$, with a redshift accuracy of ~ 30 km s^{-1}.

Spectra for the SDSS are acquired by a pair of fibre-fed multi-object spectrographs. The two spectrographs allow for the wide wavelength coverage in the SDSS data, which spans from 3800Å to 9200Å, with a resolution[1] of ~ 2000. The bluer spectrograph covers the wavelengths between 3900-6100Å, while the redder spectrograph covers 5900-9100Å. Light from the telescope travels down the fibre towards the spectrograph. Sets of 20 fibres are bundled together so that the fibres are slightly more stable than they would be if they needed to be individually fed to the spectrograph. These bundles are attached to a 'slithead', which feeds the light into the spectrograph system. As shown in Figure 2.1, once light enters the spectrograph system, it is passed to a collimating mirror, which reshapes the beam of light. The light is then passed through a beam splitter, which reflects light of a frequency higher than 6000Å, and transmits light of a frequency longer than 6000Å. Both beams of the divided light are then passed through a grism, which splits the light according to frequency, focusing each wavelength at a different spatial location, which is then captured by one of the two cameras[2].

[1] Resolution is calculated as R=$\lambda/\Delta\lambda$, or R=$c/\Delta v$

[2] Further information about the spectrograph is available in the Program Booklet, hosted at

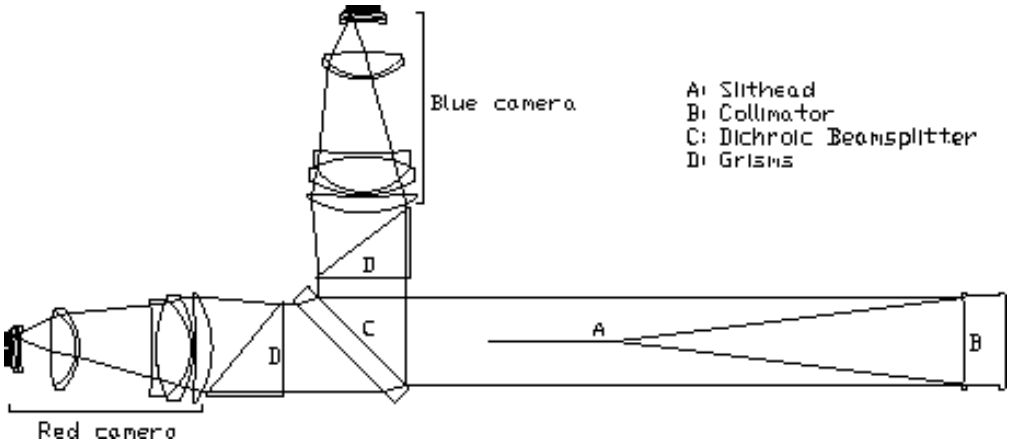


Figure 2.1: SDSS light path from fibre to camera. Light enters from the fibre via the 'slithead', labeled A in this image. It then passes to a collimating mirror (B), and through a beam splitter (C) which reflects light bluewards of 6000Å, and transmits light redwards of 6000Å. Each of the divided light paths is passed through a grism (D) in order to separate the light by frequency, and the photons are captured by either the red or blue camera, depending on its wavelength. Image credit: SDSS Program Booklet. URL: `http://www.astro.princeton.edu/PBOOK/spectro/spectro.htm`

The spectroscopic targets were observed 640 objects at a time, with all 640 fibres plugged into a custom-made aluminum plate, which was attached to the telescope for each observation (Blanton et al., 2003). Each spectroscopic exposure is 15 minutes long, and repeated until a signal to noise of at least 4.0 is acquired; typically this requires 45 minutes of exposure time (Strauss et al., 2002). While a minimum S/N of 4.0 is obtained, the average S/N is closer to 10.0 in the SDSS DR7 (Abazajian et al., 2009).

Each individual spectroscopic fibre has a diameter of 3 arcseconds, with a minimum sky separation between fibres of 55 arcseconds; placing fibres closer together than this limit was impossible due to the protective casing surrounding the fibres themselves. Objects which are closer together on the sky than 55″ are therefore preferentially missed by the spectroscopic survey. If two galaxies are located closer than this separation, the galaxy selected as a spectroscopic target is randomly drawn. However, approximately 30% of the survey area is observed more than once due to overlap in spectroscopic observation scans (Strauss et al., 2002), which means that there is information for objects at separations below the collision limit, but in fewer numbers than are present beyond the collision limit. Strauss et al. (2002) comment that only 6% of the total galaxy population is missed due to fibre collisions. The main galaxy catalogue is intended to be a magnitude limited, complete survey of the galaxy population above a limiting magnitude $m_r \leq 17.77$. However, at magnitudes brighter than $m_r > 14.5$, the galaxies are often confusing to the automated deblender, which can begin to assign more than one spectroscopic target to a single, large object. (Strauss et al., 2002)

For samples of galaxy pairs (as will be investigated in Chapter 4), the incompleteness due to fibre collisions is significantly more important. Patton & Atfield (2008) find that the ratio of photometric to spectroscopic pairs drops from 80% to 26% at the 55 arcsecond collision limit. Ellison et al. (2008b) find that this incompleteness can be accounted for by randomly selecting and excluding 67.5% of all galaxy pairs which are separated on the sky by more than 55 arcseconds, removing the bias introduced by the fibre collisions.

Individual sources are labeled by a unique combination of plate, fibre, and MJD identifiers. The plate value is the serial number of the plate upon which the fibres were placed. The fibre value is a unique identifier for each object based upon the fibre used to observe that target. MJD stands for Modified Julian Date, and marks

---

`http://www.astro.princeton.edu/PBOOK/spectro/spectro.htm`

the time of observation in units of days. The combination of these three identifiers provides a unique way of labelling a particular observation within the spectroscopic catalogue.

The photometry compiled as part of the SDSS is obtained via drift-scanning. Drift scanning is a method of observing the sky by which the telescope does not slew to track a particular object, but allows the rotation of the earth to shift the direction the telescope is pointing. Images are taken by reading information off of the CCD at the same rate as the apparent motion of the sky. The Sloan imaging is taken in five bands: *ugriz* (Fukugita et al., 1996; Smith et al., 2002). Due to the nature of drift-scanning observations, the photometric survey is relatively shallow, as each pixel was on sky in a given filter for ~ 54 seconds.

2.2.1 MPA-JHU Catalogue

In the work that follows, the catalogue of extragalactic spectroscopic sources from the SDSS compiled by the Max Planck Institute for Astrophysics (MPA) in collaboration with Johns Hopkins University (JHU) is used. The MPA-JHU catalogue is a publicly available catalogue with a number of additional properties derived from the SDSS Data Release 7, and is available at the following url: `http://www.mpa-garching.mpg.de/SDSS/DR7/`. This catalogue is split into both raw and derived quantities. Among the raw data are the emission line fluxes (along with their error scalings), identifying information about the galaxy such as its unique plate, fibre, and MJD values, along with continuum values and continuum errors for each galaxy in their catalogue. The derived quantities are the results of further analysis of the raw data, and include stellar masses, metallicities, and star formation rates for all galaxies in the SDSS spectroscopic sample.

Pre-existing metallicity calibrations

The metallicities presented in the MPA-JHU catalogue use the Tremonti et al. (2004) metallicity calibration. However, Kewley & Ellison (2008) found that the Tremonti et al. (2004) produces systematically higher metallicity values than other methods, and has significant residuals when attempting to convert into any other commonly used metallicity calibration. The Tremonti et al. (2004) relation also requires a detection of [SII] in the spectrum, which is often a relatively weak line. Requiring the [SII] line therefore limits the sample of possible metallicity calibrations, as other cal-

ibrations do not require the sulfur line. To avoid the complications of the Tremonti et al. (2004) method, a different metallicity calibration is desired. The 'recommended method' of Kewley & Ellison (2008) will be used instead, which is a modification of Kewley & Dopita (2002). This metric requires fewer emission lines than the Tremonti et al. (2004) calibration, and also has the smallest residuals when converting into other metrics. A full description of the Kewley & Ellison (2008) method is presented in §2.5.

Star Formation Rates

The star formation rate values used for the remainder of this thesis are fully described in Brinchmann et al. (2004). Brinchmann et al. (2004) uses an emission line modelling strategy developed by Charlot & Longhetti (2001), which uses the population synthesis models of Bruzual & Charlot (2003) to subtract off any absorption in the spectrum, and then model the emission lines by using Bruzual A. & Charlot (1993) models in conjunction with emission line strength predictions from the spectral synthesis code CLOUDY. This model uses a Kroupa (2001) initial mass function (IMF) to describe the distribution of stellar masses formed within a galaxy. Brinchmann et al. (2004) state that a conversion to a Salpeter (1955) IMF would increase the SFR estimates by a factor of 1.5. The models used to output a SFR optimize on the following parameters: metallicity, ionization parameter (the number of ionized to neutral atoms), dust attenuation, and the dust to metals ratio in the ionized gas. This grid results in a total suite of $\sim 2 \times 10^5$ models. This modelling approach can only be used for galaxies with high signal to noise in the required emission lines; for this work, the Hα, Hβ, [OIII]λ5007 and [NII]λ6584 lines are required to have S/N > 3.0 in order to qualify as 'high S/N'. The median uncertainty on an individual SFR value in this high S/N category is approximately 0.9 dex.

Star formation rates are calculated differently depending on the signal to noise of the emission lines of that galaxy. If the Hα, Hβ, [OIII]λ5007 and [NII]λ6584 emission lines have S/N < 3, but the galaxy still has S/N > 2 in the Hα line, they are classified as 'Low S/N Star forming'. Galaxies in this category are unable to be fed into the full modelling suite, but as Brinchmann et al. (2004) find that calibrations using only the Hα line strength match the results of the models very well, the low S/N galaxies are assigned a SFR value by using the 'average likelihood distribution' SFR for all galaxies in the high S/N sample selected to have stellar masses within a factor of 3 of

the mass of the low S/N galaxy. In the work that follows, the imposed quality control will require S/N values in these lines of > 3, meaning that the SFR values used will be exclusively selected from the high S/N star-forming sample, where emission line modelling is used to calculate SFRs.

In both cases, the emission line fluxes can only be modelled for the central region of the galaxy encompassed by the SDSS fibre. In order to obtain information about the SFR outside the fibre region, a further aperture correction is required. The MPA-JHU group uses a method that fits models to the photometry outside the fibre region in order to make the correction from the fibre SFR to the total SFR, according to the suggestion provided by Salim et al. (2007). Both the fibre and total SFR values are presented in the MPA-JHU catalogue, and both are used in this thesis, although the fibre values are more heavily used.

Unlike the SFR values, for the purposes of metallicity calibration, there is no correction to the spectra which can account for the flux lost outside the fibre region. Both the emission line fluxes and the subsequently calculated metallicities that rely upon them are limited to the central region of the galaxy only.

Stellar Masses

The stellar masses presented by the MPA-JHU group are used only in Chapter 3. In Chapters 4 and 5, different mass calibrations are used. The stellar mass values presented in the MPA-JHU catalogue are created by fitting a large suite of Bruzual & Charlot (2003) models to the total magnitudes of the 5-band DR7 photometry. The best-fit model is determined through a χ^2 minimization procedure. The MPA-JHU group prefer the photometrically determined masses as they are less dependent on the region of the galaxy covered by the fibre, where spectra can be obtained. The photometric masses are also less reliant on an accurate elimination of emission line features within the spectrum. Furthermore, no correction is needed to go from a fibre mass to a total mass, since the total mass can be directly calculated from the total photometry. The spectroscopically determined masses, by contrast, require aperture corrections which may introduce additional assumptions. However, both the photometrically and spectroscopically derived mass values are generally in good agreement with each other. An average uncertainty of approximately 0.1 dex is presented for these mass measurements. A more complete comparison between the MPA-JHU models and those calculated spectroscopically by Kauffmann et al. (2003) is available

here: http://www.mpa-garching.mpg.de/SDSS/DR7/mass_comp.html.

2.3 Line Flux Quality Control

The ultimate goal of the quality control measures presented here is to construct a large sample of robust metallicity values from the MPA-JHU's catalogue of SDSS galaxies. Significant quality control restrictions need to be imposed before the metallicities can be reliably calculated. Metallicity calibrations rely upon a set of emission line fluxes; if these flux values are unreliable or missing, the metallicity values will also be unreliable or impossible to calculate. With this in mind, the first set of quality control parameters develop; only galaxies within a certain range of redshifts will be useful, as galaxies that are too close or too distant will shift some of the necessary emission lines outside of the wavelength range of the SDSS spectra. Not all galaxies will have reliable detections in the needed emission lines within the permitted redshift range, as some emission lines may be weak or undetected. Ensuring that the line fluxes are reliable requires the development of a second set of quality control criteria, which use the limitations of the data to determine the restrictions which should be put into place. The end result of the imposed criteria should be a set of galaxies with robust emission line fluxes.

2.3.1 Negatives & zeros

The most obviously unreliable fluxes are those which were recorded as either a negative value or as exactly zero. Flux values of exactly zero were often the signature of problems with the spectra: e.g., if a flaw in the spectrum had clipped out all or part of an emission line. Some of the lines with zero flux recorded were for galaxies where the line had been redshifted out of the spectral range of the instrument, so a flux measurement could not be calculated. Negative emission line flux values are unphysical, and result from incomplete removal of the Balmer absorption line. The flux measurement of a poorly subtracted residual spectrum would result in a negative value. As neither zero nor negative values are physically meaningful, galaxies with these flux measurements were excluded from the sample.

2.3.2 Flux extremes

In order to ensure that the remaining flux values (i.e., with fluxes > 0) are reliable, the distributions of both flux and flux error were plotted for the Hα & Hβ lines in Figure 2.2. The raw flux errors presented by the MPA-JHU catalogue have been found to slightly underestimate the true uncertainty on the fluxes. The MPA-JHU group therefore recommends a scaling factor to be applied to each of the emission lines to more accurately describe the emission line quality. These scaling factors have been determined by investigating duplicate observations to determine the empirical scatter in the observations, relative to the formal errors quoted in the catalogue itself. The per-emission line flux error scalings are available at the following URL: `http://www.mpa-garching.mpg.de/SDSS/DR7/raw_data.html`. These scalings are adopted for all presented flux errors.

Both the Hα and Hβ lines have a prominent secondary peak in the distribution of flux errors at extremely high values (flux error $> 10^{-11}$ erg/s/cm^2, or $> 10^6$ in Figure 2.2, where fluxes are measured in units of 10^{-17} erg/s/cm^2), and a similar secondary distribution peak at high flux values (flux $> 10^{-12}$ erg/s/cm^2; $> 10^5$ in Figure 2.2). A check of the original spectra indicated that these high flux values were indeed spuriously high, as the raw spectra did not show peak values in the Hα and Hβ lines that aligned with the recorded value. Rather than making an *a priori* cut in the recorded flux of these lines to eliminate the anomalously high values, more fundamental cuts were sought to naturally remove them.

Redshift cuts

Sky lines are a complex series of emission lines produced by our atmosphere. Poor sky line subtraction from the observed spectrum could lead to a miscalculation of the flux of an extragalactic emission line. Due to their complex nature, sky lines are difficult to remove without leaving residual flux in the spectra. Imposing a redshift cut to exclude galaxies which have emission lines falling in a region of the spectrum with strong sky line residuals should remove some of the high flux galaxies. To check if this is likely to to be driving some fraction of the high flux peak ($> 10^{-12}$ erg/s/cm^2) in Figure 2.2, in Figure 2.3, the redshift distribution of those galaxies found within the high flux peak is plotted as a red dashed line. For comparison, the redshift distribution of the total sample is also plotted (black solid line), where both samples are normalized to the number of galaxies in each sample. This plot demonstrates

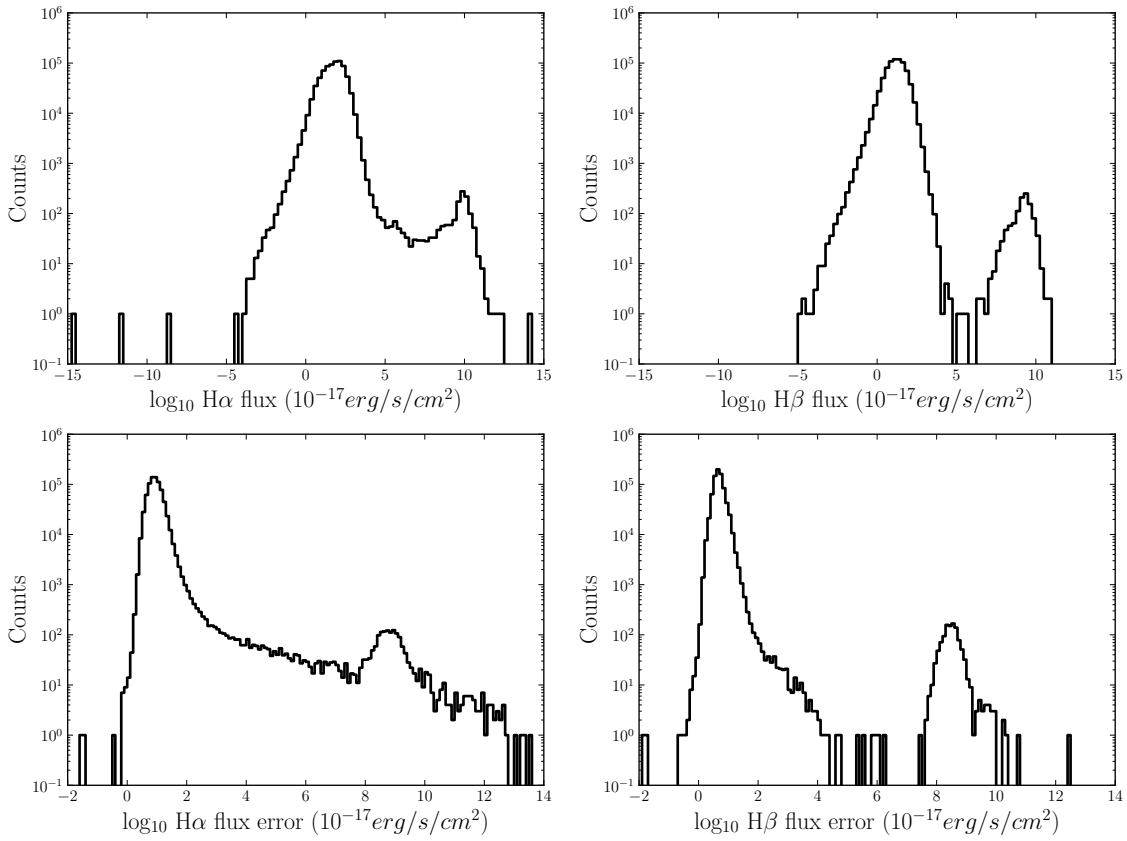


Figure 2.2: Distribution of the flux values in the MPA-JHU catalogue, in log-log space. Left panels are the log of Hα flux (top) and error (bottom), and right panels are Hβ flux (top) and error (bottom). Secondary peaks at flux values $> 10^{-12}$ erg/s/cm^2 ($> 10^5$ in the plotted units) are visible in both Hα and Hβ, but more prominently in Hβ.

that there is a redshift beyond which the high flux peak is a dominant contributor to the distribution; this transition occurs at a redshift of $z \gtrsim 0.2$. However, a more precise cut can be effected by using the wavelength of the sky lines as a constraint on viable wavelengths λ. The spectra show considerable sky line residuals beginning at 8200Å and continuing to higher wavelengths. If the redshift range is constrained such that the Hα line (centred at 6563Å) cannot be redshifted into these sky lines, the maximum reliable redshift is calculated to be 0.25. This is slightly more generous than is indicated by the histogram, but as the distribution of the total sample in Figure 2.3 indicates that there should be a gradual tail of galaxies with redshifts up to $z \sim 0.25$, in order to avoid excluding more galaxies than necessary, the less conservative cut is adopted.

A lower limit on redshift is put into place where the [OII]λ3727 line would be redshifted out of the lower end of the spectral range, which would occur at z= 0.02. The final redshift cut put into place is therefore $0.02 < z < 0.25$, with the additional criterion that the MPA catalogue parameter `z_warning` must be 0; a non-zero value for this parameter indicates that the redshift solution is unreliable.

Continuum errors

Another potential cause of error in the flux measurements is the continuum fitting that takes place before the flux measurements are possible. The distributions of continuum measurements and errors on those measurements are presented in Figure 2.4. A clear bimodality in both errors and total continuum values is visible. A population of galaxies with extremely high recorded continuum fluxes (at least 2 orders of magnitude above the main distribution) is correlated with the high flux error (F_{err}) peak in Figure 2.2. Eliminating the high flux error peak from the sample also eliminates the high continuum flux peak. Based upon this distribution, a natural cut in continuum error (C_{err}) is imposed, such that only the lower continuum error distribution remains. As such, the cut of 10^{-2} erg/s/cm^2< C_{err} < 10^2 erg/s/cm^2 is imposed. This particular cut has a dramatic effect on the overall flux distribution shown in Figure 2.2, and cuts out the vast majority of the high flux values. Some high flux values lingered, but as the remaining high fluxes were still associated with high flux error, these lines are likely to be unreliable. To remove the remaining high fluxes, a third cut was imposed on the flux error: 0.1 erg/s/cm^2< F_{err} < 10^4 erg/s/cm^2.

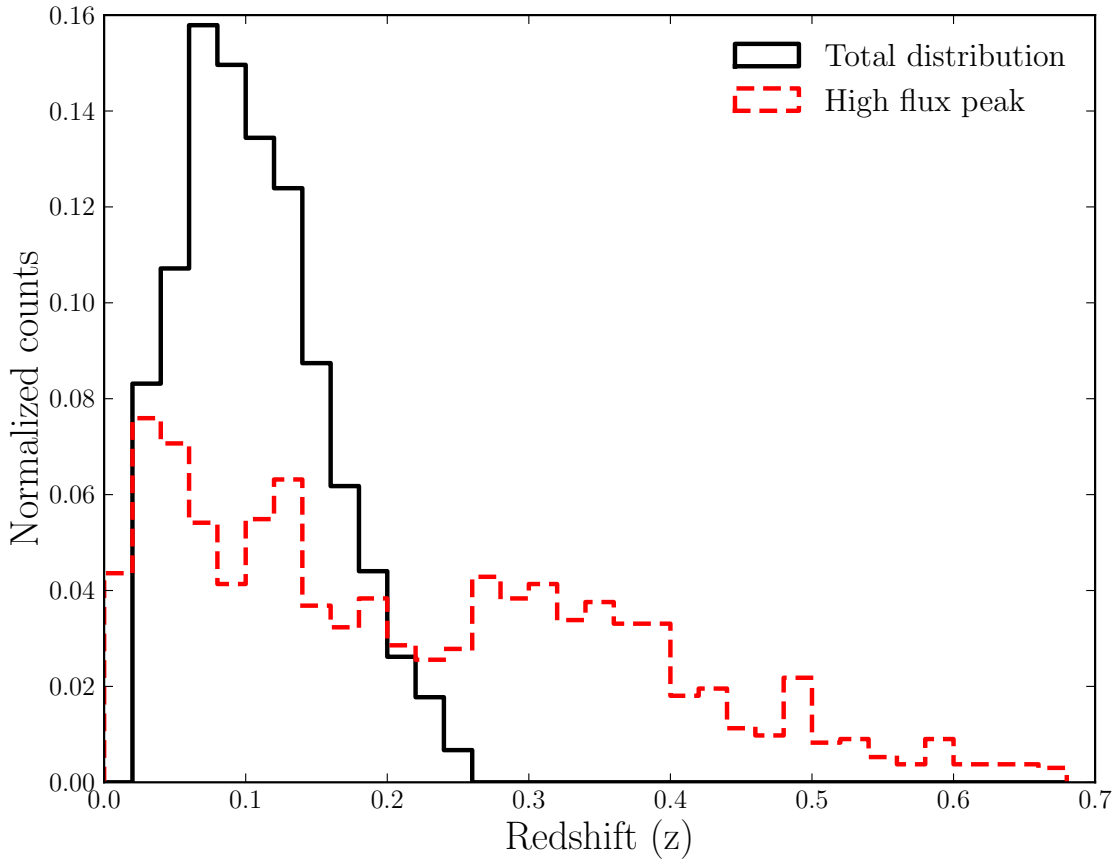


Figure 2.3: The normalized redshift distribution over the full sample (black solid line) as compared to the distribution in the high flux peak (red dashed). The high flux peak begins to dominate over the total distribution at $z \gtrsim 0.2$, indicating that the high flux values preferentially correlate with higher redshift objects.

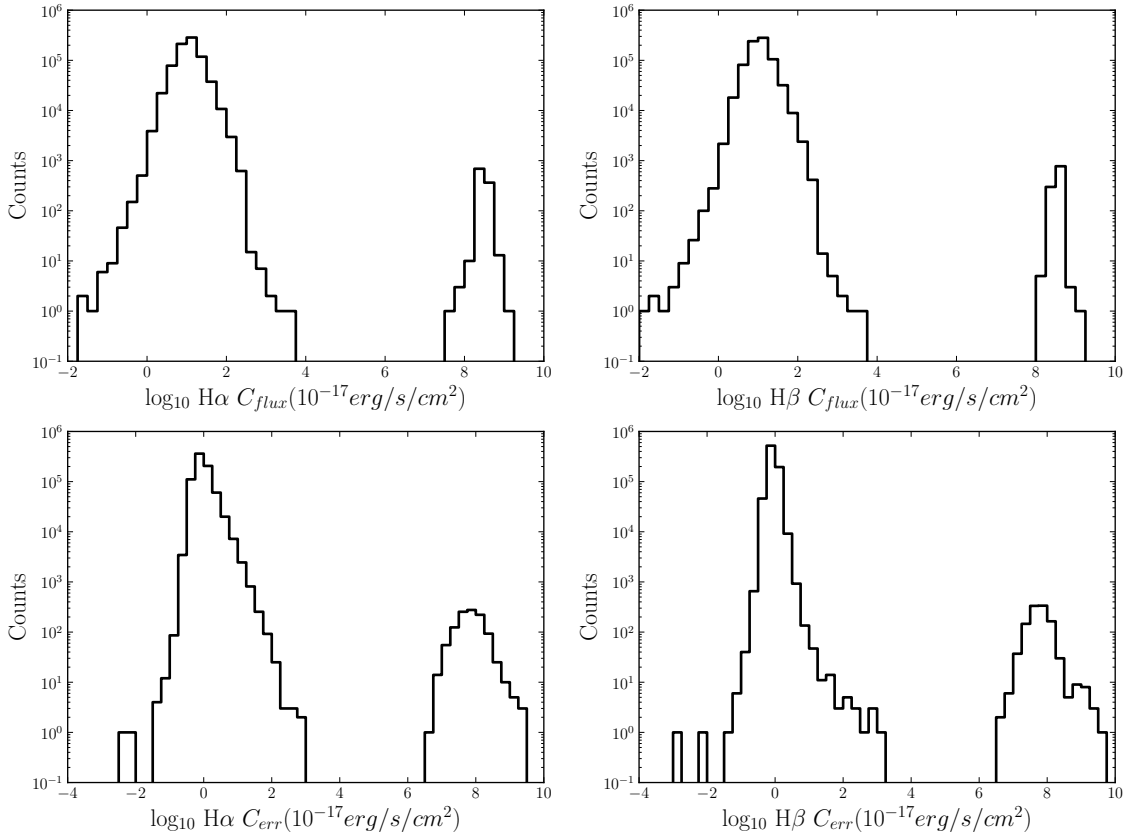


Figure 2.4: Total continuum error & continuum flux distribution. Panels are organized in the same way as Figure 2.2. The continuum error has a strong bimodality, with a secondary peak at errors $> 10^{-11}$ erg/s/cm^2; the high continuum flux error galaxies directly correspond to galaxies with continuum fluxes of $> 10^{-9}$ erg/s/cm^2.

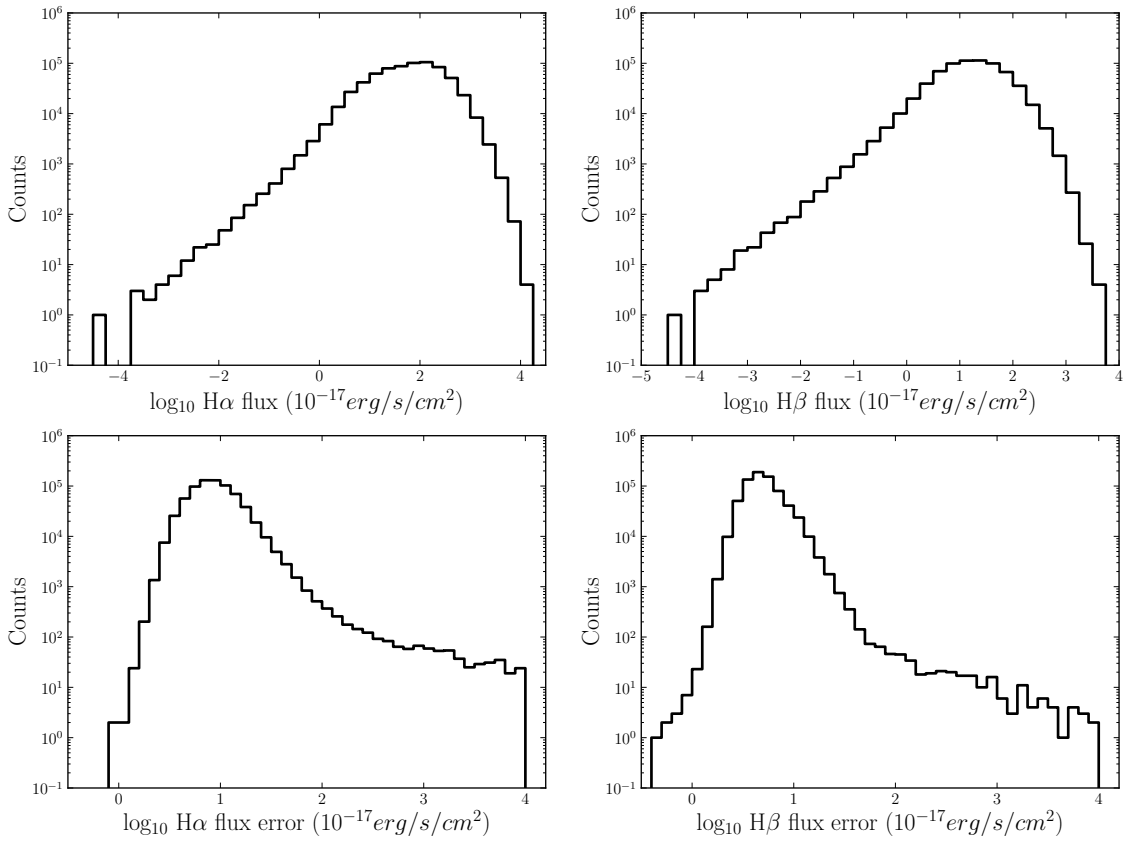


Figure 2.5: Final flux distribution of the quality controlled sample, with cuts as in Table 2.1. The panels are the same as Figure 2.2.

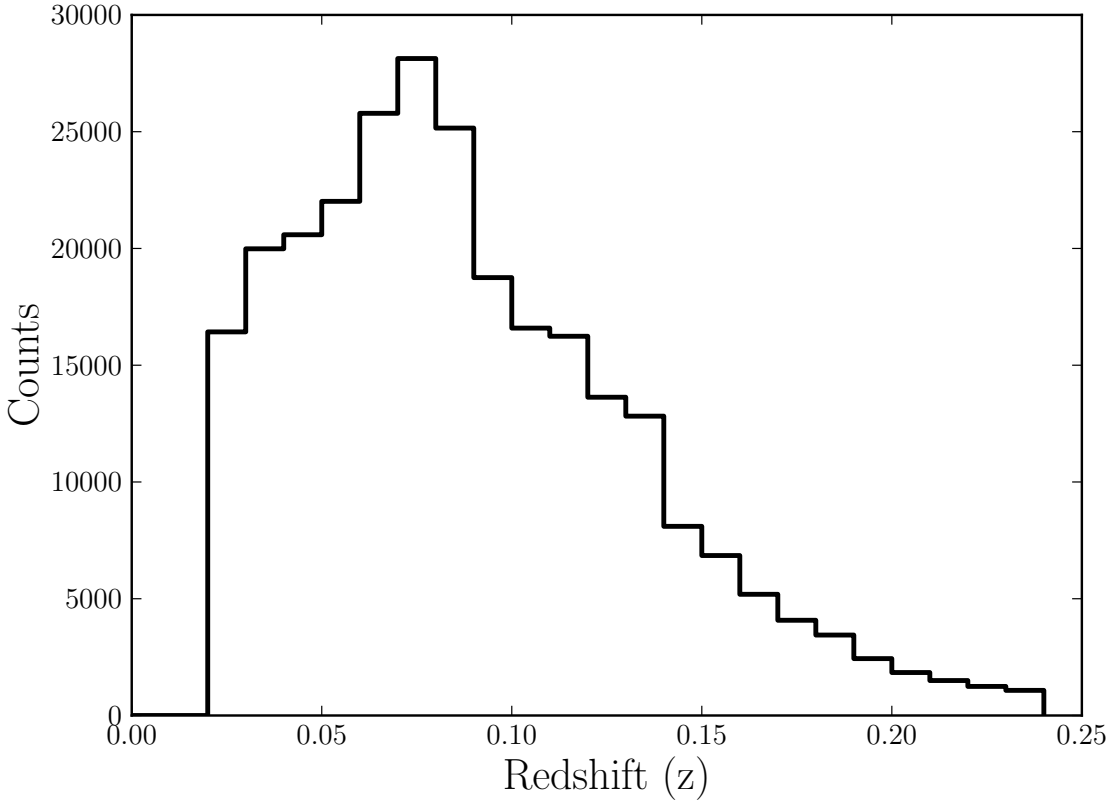


Figure 2.6: The final redshift distribution of the quality controlled sample.

Signal to Noise

With these cuts on continuum error, redshift, and flux error in place, the final galaxy flux distributions for Hα and Hβ are shown in Figure 2.5; the flux values that remain are within reasonable bounds. The final redshift distribution is shown in Figure 2.6.

However, the reliability of the detection itself is also of concern for the steps which follow; the assessment of a Signal-to-Noise (or S/N) ratio cut on the emission lines is therefore undertaken. The data is used as a guide to dictate the strictness of the S/N cut required. If the data is well behaved at low S/N, a more lax cut may be put into place without introducing unphysical values into the sample. If, however, the data does not become well behaved until a much stricter S/N cut is imposed, then a harsher S/N criterion would be required to ensure that the sample does not contain a significant fraction of unreliable values.

Reliable Balmer ratios are required in order to proceed with a correction for the dust intrinsic to the galaxies in the sample. Determining that the Balmer ratio itself has strong signal to noise is therefore critical in order to continue. Under the assumption that galaxies are radiating as optically thick clouds (Case B recombination; Baker & Menzel, 1938), the Balmer ratio has a minimum value of 2.85. It might be expected that as the signal to noise of the Balmer ratio decreases, the adherence to this lower limit may become weaker. To make sure that the data behaves as expected, the S/N of the Hα and Hβ line ratio is calculated as follows:

$$\mathrm{S/N}_{H\alpha/H\beta} = \frac{1}{\sqrt{(1/(S/N_{H\alpha}))^2 + (1/(S/N_{H\beta}))^2}} \tag{2.1}$$

Figure 2.7 shows the S/N of the Balmer ratio as a function of the Balmer ratio itself. It is clear that at low S/N of the Hα/Hβ ratio, a very wide range of Balmer ratios can be determined. Even at a S/N of 1, some galaxies show Balmer ratios that are a factor of 100 smaller than predicted by the theoretical lower limit. The peculiar shape of this figure in log-log space merited further investigation. It was determined that the galaxies populating the diagonal to the lower left all had poor Hα S/N and the diagonal towards the lower right was populated with galaxies with poor Hβ S/N. The top peak, with a Balmer ratio centred around ~ 4, is the only location where good signal to noise for the Hα and for Hβ lines overlapped.

Fortunately, a stricter cut in S/N eliminates the majority of the unphysical scatter below the theoretical limit due to poor S/N. Such a S/N cut will also eliminate much of the scatter towards extremely large Balmer ratios. A S/N cut of 5 (shown in

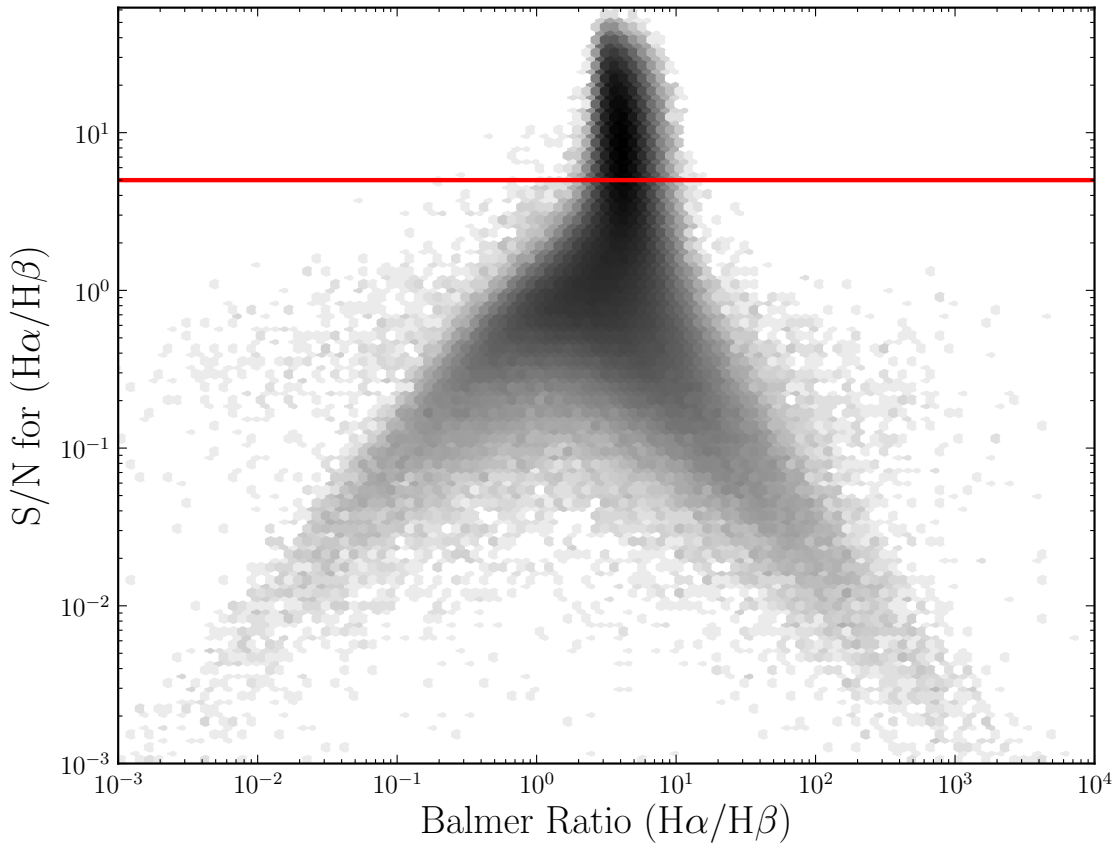


Figure 2.7: Log-log plot of Balmer ratio (Hα/Hβ) versus the signal to noise of the Balmer ratio. The red line marks a S/N in the Balmer ratio of 5.

Figure 2.7 as a horizontal red line) serves as a good compromise between maintaining the sample size and excluding unphysical Balmer ratios. Increasing this S/N cut may improve the sample quality but would primarily serve to limit the sample size. Allowing the S/N cutoff to drop below 5 swiftly increases the fraction of galaxies with unphysically low Balmer ratios into the sample.

With a Balmer ratio S/N > 5 cut in place, a large portion of the sample was excluded, leaving a sample of galaxies tightly clustered around the theoretical lower limit for Case B (optically thick clouds) of 2.85, shown in Figure 2.8, with the theoretical lower limit of 2.85 plotted as the vertical red line. Some galaxies are expected to scatter to the left of this lower limit, but this is expected to be a random effect, and can be limited with a further significance cut. A 3σ cut for galaxies falling to the left of 2.85 is imposed (plotted as the blue dashed line in Figure 2.8), which eliminates ~ 50 galaxies from the sample. The final set of imposed criteria results in a much tighter distribution of Balmer ratios, with a range of only 12.95.

Final Quality Control	
Redshift	$0.02 < z < 0.25$
Continuum error	10^{-2} erg/s/cm^2$< C_{err} < 10^2$ erg/s/cm^2
Flux error	.1 erg/s/cm^2$< F_{err} < 10^4$ erg/s/cm^2
Hα/Hβ Signal/Noise	S/N > 5
Scatter below Hα/Hβ =2.85	$< 3\sigma$

Table 2.1: Final quality control parameters.

2.3.3 Extinction corrections

Once these criteria are in place, the sample is reduced to 270,553 galaxies with robust Hα and Hβ. However, the measured flux values are still reddened from the dust present in each galaxy; metallicity calculations require intrinsic (de-reddened) fluxes. No Galactic extinction corrections are required, as the fluxes provided by the MPA-JHU catalogue have already been corrected for Galactic extinction. The intrinsic reddening of a galaxy ($E(B-V)$) can be calculated for all galaxies with reliable Balmer ratios, as determined in the previous section. With a calculated reddening factor in hand, the fluxes for each galaxy's emission lines may be corrected to an intrinsic (de-reddened) value. To calculate the $E(B-V)$ values, a reddening curve must be chosen.

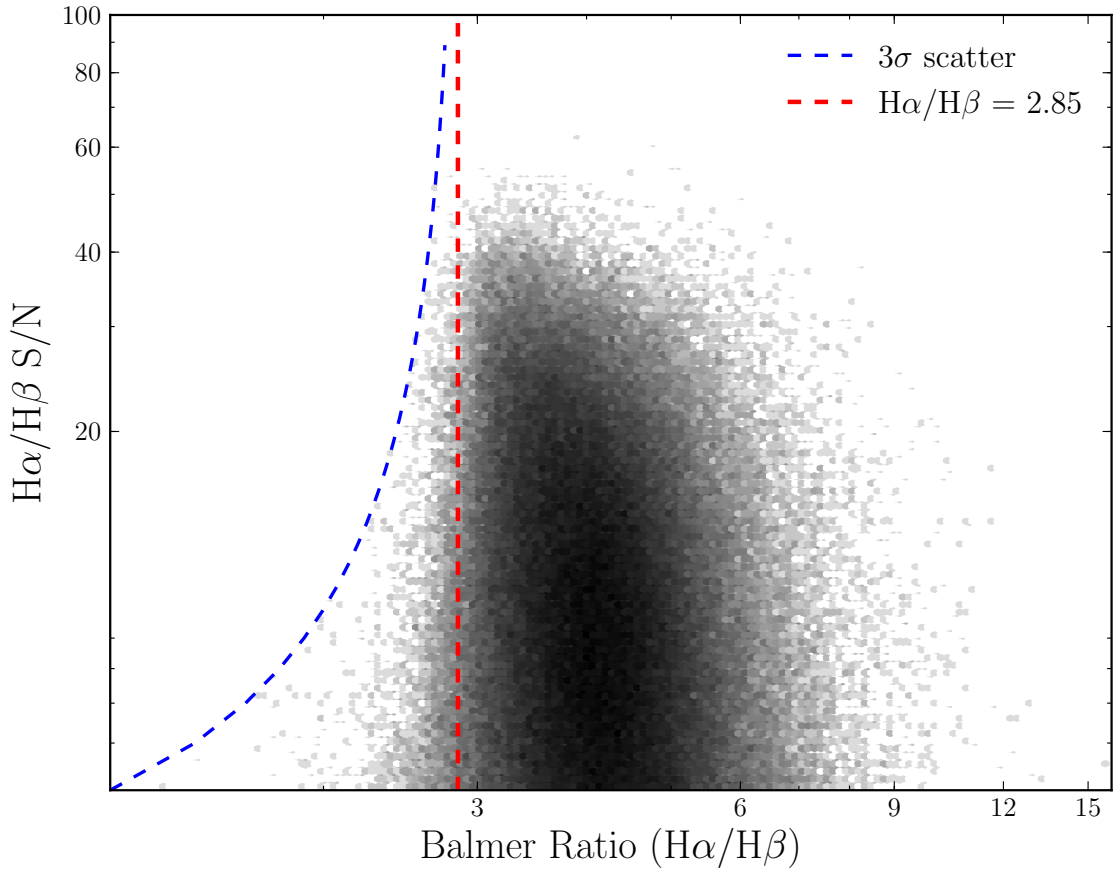


Figure 2.8: Balmer ratios vs. Hα/Hβ S/N for the quality controlled sample. The theoretical lower limit is plotted as a red vertical line, with the 3σ lower limit for scatter below this value plotted as a blue dashed line.

There are three primary reddening curves; that of the Milky Way (MW), the Large Magellanic Cloud (LMC), and the Small Magellanic Cloud (SMC). Each curve varies in form; the steeper the reddening curve is, the more dramatic the wavelength dependence of the correction to an emission line will be. The Milky Way reddening curve is commonly used; however the SMC extinction curve lacks the significant features present in the Milky Way curve, potentially making it a more generic curve. $E(B-V)$ values are calculated for both the Milky Way extinction curve and the SMC extinction curve. In Chapters 3 and 4 that follow, $E(B-V)$ values from the SMC curve are adopted.

Within the wavelength range of the emission lines in the sample, the three extinction curves are nearly identical. Figure 2.9 shows a comparison between the $E(B-V)$ values calculated via the Milky Way and SMC reddening curves. The blue solid line shows the results of the Milky Way curve, while the green dashed line shows the SMC curve; the two distributions are nearly identical. The distribution match between the DR7 SMC and MW $E(B-V)$ curves in Figure 2.9 (green dashed and blue solid, respectively) is an *a posteriori* confirmation that the differences between the SMC and MW extinction curves do not have a significant effect on the data.

The $E(B-V)$ of a galaxy takes the following form:

$$E(B-V) = \frac{\log(\frac{2.85}{(H\alpha/H\beta)})}{0.4[k(\lambda_{H\alpha}) - k(\lambda_{H\beta})]} \tag{2.2}$$

2.85 is the expected Balmer ratio for Case B optically thick clouds[3] and $k(\lambda)$ is:

$$k(\lambda) = \xi(\lambda) R_v, \tag{2.3}$$

R_v is a constant and $\xi(\lambda)$ is the reddening curve.

Using the reddening curves presented in Pei (1992), $\xi(\lambda)$ takes the following form:

$$\xi(\lambda) = \sum_{i=1}^{6} \frac{a_i}{(\lambda/\lambda_i)^{n_i} + (\lambda_i/\lambda)^{n_i} + b_i}. \tag{2.4}$$

The $k(\lambda)$ and $\xi(\lambda)$ values only need to be calculated once per emission line, as they are only a function of λ, which is fixed for each emission line. By contrast, the $E(B-V)$ must be separately calculated for every galaxy in the sample, as each

[3]For the small number of galaxies which have Balmer ratios less than 2.85, the $E(B-V)$ values calculated wind up being very close to zero (e.g., $\sim 10^{-5}$). There are no negative values of $E(B-V)$ in the sample.

Parameter i	a	λ_i	b_i	n_i
Milky Way galaxy				
1	165.	0.047	90.	2.
2	14.	0.08	4.	6.5
3	0.045	0.22	-1.95	2.
4	0.002	9.7	-1.95	2.
5	0.002	18.	-1.80	2.
6	0.012	25.	0.00	2.
Small Magellanic Cloud				
1	185.	0.042	90.	2.
2	27.	0.08	5.5	4.
3	0.005	0.22	-1.95	2.
4	0.010	9.7	-1.95	2.
5	0.012	18.	-1.80	2.
6	0.030	25.	0.00	2.

Table 2.2: Table of coefficients for the reddening curve $\xi(\lambda)$ from Pei (1992) for both SMC and MW curves. SMC coefficients are used in Chapters 3 and 4.

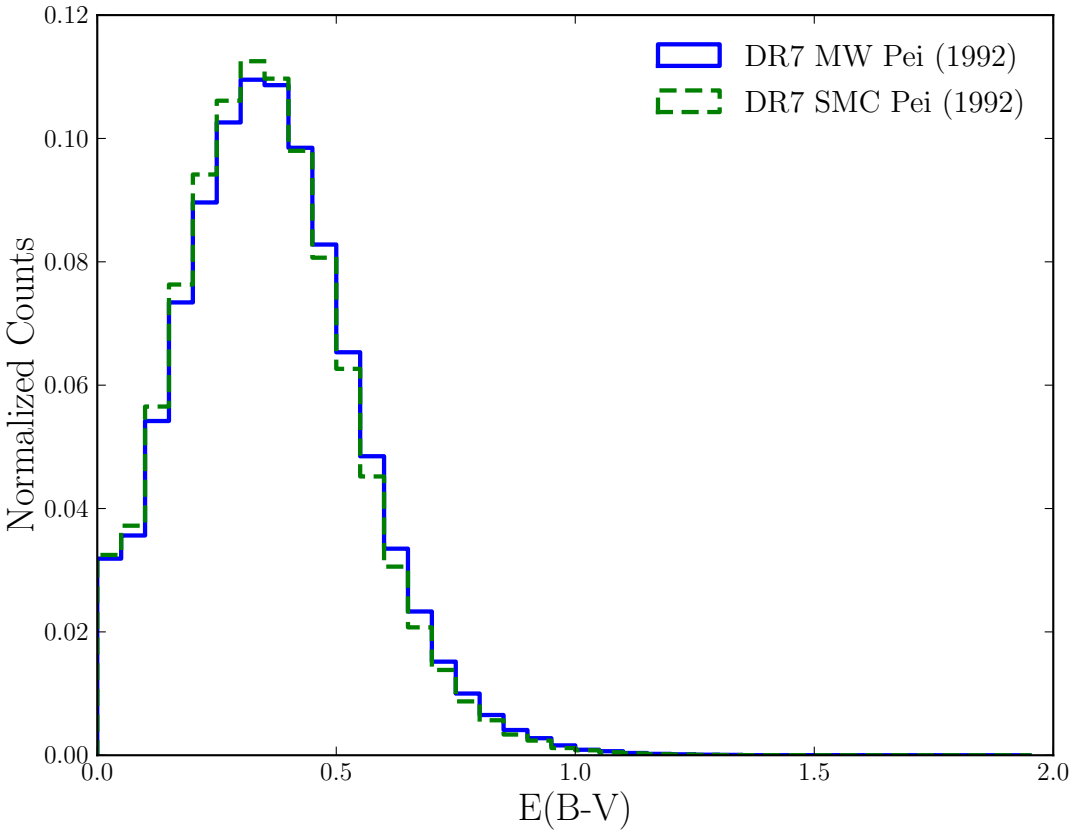


Figure 2.9: $E(B-V)$ distributions using both the Pei (1992) SMC & Milky Way curves for the SDSS DR7 sample. The $E(B-V)$ values for the Milky Way curve are solid blue, and results of the SMC curve are plotted in a green dashed line.

galaxy will have a different Hα/Hβ ratio. As the $E(B-V)$ corrections are expanded to correct the reddening present in the other emission lines needed for metallicity calibrations, each galaxy will require all $E(B-V)$ calculations to be re-run with the new $k(\lambda)$ and $\xi(\lambda)$ values for the wavelengths of the additional lines.

Each emission line must also pass a signal to noise cut to ensure that $E(B-V)$ calculations are only made for reliable line detections. However, at this stage, the lack of a given emission line (beyond the Hα and Hβ lines, which are required) does not eject the galaxy from the sample, but simply marks the line as a non-detection. $E(B-V)$ values are not calculated for emission lines marked as non-detections. Two separate S/N cuts are imposed on the emission lines: S/N > 1 and S/N > 5. The results presented in the rest of this thesis rely exclusively on the S/N > 5 sample, but for completeness, the sample sizes for S/N > 1 are also reported here, as the S/N > 1 sample has been used in other published works (e.g., Patton et al., 2013). The S/N > 1 sample cannot be used for metallicity values (hence its disuse for the rest of this thesis); however, if more basic quantities are required, such as a line flux or AGN classification (discussed in Section 2.4), the larger number statistics in the lower S/N sample can be useful.

Each galaxy's emission lines are extinction-corrected using the following equation:

$$F_i = F_o \times 10^{0.4k(\lambda)E(B-V)}, \tag{2.5}$$

where F_i is the intrinsic (extinction-corrected) flux, F_o is the observed flux, $k(\lambda)$ is as defined in Equation 2.3, and the $E(B-V)$ is as calculated for the individual galaxy's emission lines (Equation 2.2). This correction leaves a sample of galaxies with robust, extinction corrected, fluxes. The number of galaxies with extinction corrections for each emission line are presented in Table 2.3 for the S/N > 1 sample, and in Table 2.4 for the S/N > 5 sample.

Line	Number of galaxies
Hα	516,659
Hβ	516,659
[OII]λ3726	400,696
[OII]λ3729	420,932
[OIII]λ4959	329,395
[OIII]λ5007	481,381
[NII]λ6584	508,533
[SII]λ6717	480,605
[SII]λ6731	453,096
[OI]λ6300	378,131

Table 2.3: Final number of galaxies left in the sample after quality control and extinction corrections. This sample requires S/N > 1 in the emission lines to remain in the sample, and is not used for the remainder of this thesis.

Line	Number of galaxies
Hα	272,788
Hβ	272,788
[OII]λ3726	101,741
[OII]λ3729	118,124
[OIII]λ4959	73,033
[OIII]λ5007	162,938
[NII]λ6584	264,705
[SII]λ6717	245,352
[SII]λ6731	211,096
[OI]λ6300	63,387

Table 2.4: Final number of galaxies left with robust emission lines (S/N > 5) for given lines, after quality control and extinction corrections. This sample is used throughout the rest of this thesis.

2.4 Active Galactic Nuclei classification

With the corrected line fluxes for the emission lines needed for the metallicity calibrations, and a clean quality controlled sample, the next step is to identify and exclude any galaxies harbouring an Active Galactic Nucleus (AGN). Galaxies hosting an AGN are thought to contain a central black hole that is actively accreting material. AGN are particularly important to identify in studies requiring emission line fluxes, as emission line fluxes excited by an AGN are produced under fundamentally different physical conditions than the emission line fluxes produced by a star forming region. These different physical conditions mean that many of the underlying assumptions fed into emission line flux analysis will no longer be valid in the AGN case.

Unlike the thermally produced ionizing radiation of bright stars, an AGN will emit ionizing radiation according to a non-thermal power-law continuum; this difference changes the ratios of the line strengths emitted by the gas surrounding it. Photoionisation from the UV radiation produced by hot O & B type stars obeys the physics of a Strömgren sphere. The UV photons from (typically) O or B type stars are energetic enough to ionize hydrogen, and they do so in a smoothly expanding sphere, which is almost entirely a region of fully ionized hydrogen, with a minimally thin partially ionized region (Strömgren, 1939). By contrast, an AGN produces many more high energy photons, which do not gently ionize a bubble, but penetrate much further into the surrounding gas than the photoionization of the O & B type stars will. This results in a much thicker region of partial ionization surrounding the fully ionized region.

This physical distinction between the radiation produced by hot stars and that of an AGN allows for diagnostics to be created which reflect the changing strength of the overall ionization state of the gas and the dominance of the partially ionized region. This is most easily accomplished by using forbidden emission line fluxes, relative to the strength of a hydrogen Balmer line. A forbidden line is the result of an ion spontaneously de-exciting; it is termed forbidden due to the extreme unlikeliness of the line occurring. They only occur in extremely low-density environments, as the ion must be excited to reach a high energy state, but subsequently spend a sufficiently long period of time undisturbed so that the ion can spontaneously de-excite and radiate a photon, instead of becoming collisionally de-excited. On earth, the high energy state would be collisionally de-excited long before the ion has a chance to radiate, so these lines are almost exclusively observed in an astrophysical context.

For both oxygen and nitrogen, the forbidden lines are collisionally excited as well as collisionally de-excited. In order to observe these lines, the time between collisions must be sufficiently long that the atoms have time to radiate between collisions, and the energy of impact must be of sufficient energy that it can excite an electron. Selecting a ratio of forbidden line to Balmer line avoids an additional complication of abundance dependence, which would be a concern if two forbidden lines of different elements were chosen to calculate a ratio. The choice of emission lines to use as a diagnostic should be carefully considered. Ideally, the selected ratio should use lines which are close in wavelength, to avoid a dependence on the reddening of the spectrum (Baldwin et al., 1981).

A single line ratio using [OIII] is sufficient to be able to determine whether there are high or low ionization levels within a galaxy. However, since the [OIII] lines can be excited to equally high flux levels in strong AGN and in regions around sufficiently energetic O & B type stars, an additional metric is required to separate star forming galaxies from those with AGN dominated fluxes.

To perform this separation, it is useful to consider an emission line that primarily arises in the partially ionized region that exists around AGN, but not around thermally ionized regions. The [NII] line fits this requirement; it is primarily produced as a result of recombination in the partially ionized regions of gas found preferentially around AGN. Star-forming galaxies almost universally have low [NII] fluxes, whereas the harder ionization fields around an AGN will generally push the [NII] emission line to high values (Baldwin et al., 1981; Veilleux & Osterbrock, 1987). The [NII] line is very close in wavelength to the Hα line, and [NII] is usually a relatively strong line in a spectrum, making the [NII]/Hα ratio a practical one.

To more completely characterize the AGN, the information available through the [OIII] line is not discarded. In order to distinguish between Seyfert and LINER (Low Ionization Nuclear Emission Region) galaxies, the second collisional line is needed to discriminate between their ionization strengths. Seyfert galaxies are characterized by high ionization fields, whereas LINER galaxies, as the name implies, have lower ionization fields. These differences will be reflected in the flux values for the [OIII] lines. The resulting diagnostic, plotting [OIII]/Hβ against [NII]/Hα, is referred to as the BPT diagram, after the authors of the Baldwin et al. (1981) work in which it was first proposed.

When enough statistics are obtained, galaxies plotted on this diagram result in what has been dubbed the 'seagull plot', due to the two distinct wings that form in

the data. The left hand wing is comprised of star forming galaxies, and the right hand wing of galaxies with AGN. Where the two lines meet is a jumble of 'composite' galaxies, where the star forming and AGN wings scatter into each other. The decision of where to draw a line to distinguish between the AGN and star forming galaxies then becomes highly dependent on whether a clean AGN or clean star-forming sample is desired. Some diagnostic lines are designed to exclude all possible star forming galaxies, leaving a clean AGN sample by excluding the majority of composite galaxies. A line designed for such a clean AGN sample would cut through the right hand wing, and only galaxies to the right of it would be kept. This would lead to a star-forming sample that could be heavily contaminated by AGN sources. Similarly, the opposite can be done; by drawing a line which very tightly follows the star forming wing, a clean sample of star forming galaxies would be selected, but the AGN population may have contamination from star forming galaxies. Three variations on diagnostic lines are plotted in Figure 2.10: Kewley et al. (2001, K01), Kauffmann et al. (2003, K03), and Stasińska et al. (2006, S06).

The diagnostic line proposed by K01 is drawn along a theoretical maximum starburst contour. K01 determined that any galaxy undergoing even an extreme starburst event would not fall to the right of this line. Plotted in blue in Figure 2.10, it cuts deeply into the right hand wing as an upper limit to the possible star forming galaxies. K03, by contrast, is a semi-empirical line, drawn to more evenly divide the two wings of the BPT diagram. It is clearly more conservative in what it labels a star forming galaxy than the line drawn by K01 (see the green line in Figure 2.10). Stasińska et al. (2006) uses upper limits on new photoionisation models to determine an upper bound for purely star-forming galaxies. The calibration of Stasińska et al. (2006) is thus distinctly more stringent, and cuts out approximately 20% of the star forming sample determined by Kauffmann et al. (2003). These classifications therefore increase in the permissiveness of the star forming sample from S06 to K03 to K01.

For this sample, the Kewley et al. (2001) line is used as the master sample of star forming galaxies, with subsamples of the galaxies passing Kauffmann et al. (2003) and Stasińska et al. (2006) individually flagged such that switching between samples can be easily done. In order to be flagged as star forming, galaxies must be able to be plotted on the BPT diagram (i.e., all four emission lines must pass the previous quality control), shrinking the sample further to 156,599 galaxies, with the Kewley et al. (2001), Kauffmann et al. (2003) and Stasińska et al. (2006) subsamples numbering 146,753, 127,679, and 106,545 respectively.

By requiring a less conservative limit of S/N of > 1 in the lines required for the AGN diagram, this sample increases to 476,138 classifiable galaxies, with 401,126 passing the SF cut of Kewley et al. (2001), 320,680 passing for Kauffmann et al. (2003), and 219,309 for Stasińska et al. (2006).

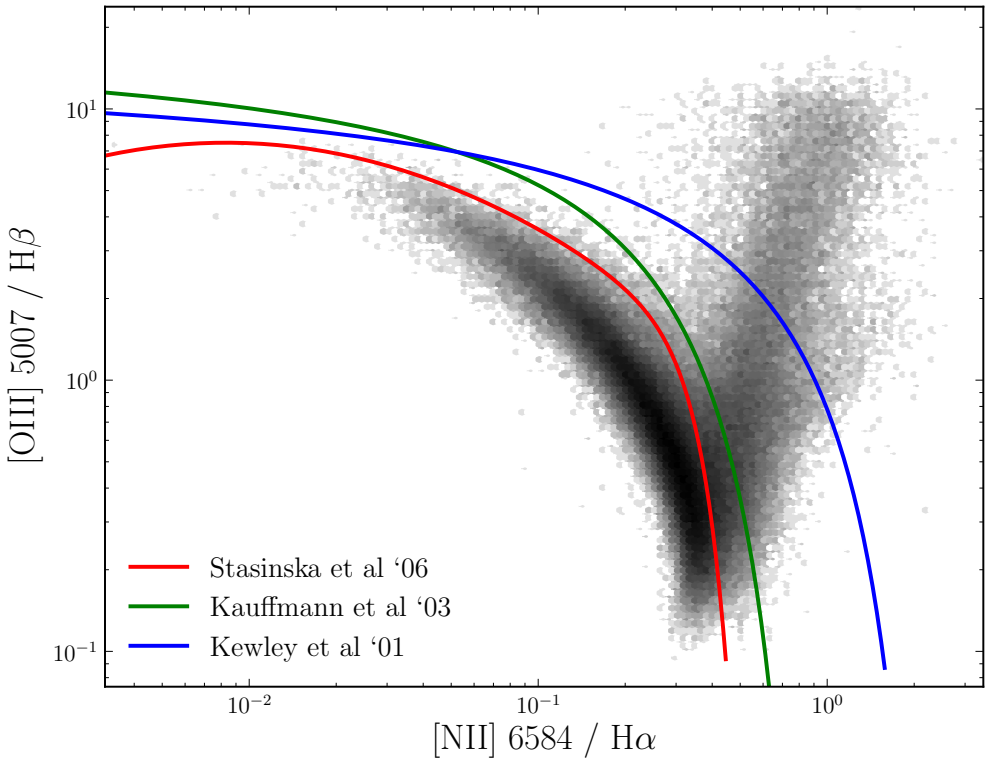


Figure 2.10: AGN BPT Diagnostic diagram. The blue line is Kewley et al. (2001), the green line is Kauffmann et al. (2003) and the red line is the Stasińska et al. (2006) curve. A total of 156,599 galaxies (shown in the grayscale 2-D histogram) have significant detections for the emission lines needed to be plotted on the diagram.

2.5 Metallicity Calibration

As a sample of star forming galaxies with reliable line fluxes has now been obtained, it is possible to proceed to the metallicity calibration itself.

Gas-phase metallicity is usually defined as the ratio of the number of metal atoms to the number of hydrogen atoms in any given HII region. Oxygen is often used as the tracer element when investigating the metallicity of interstellar gas, due to its abundance in the universe. Metallicity (Z) using oxygen as a tracer is expressed as $Z = \log(\mathrm{O/H}) + 12$. Metallicity calibrations are still poorly constrained in terms of producing an absolute measure of the metal content of the gas. However, as a way of calculating relative metal-richness or metal-poverty for consistently performed calculations within a specific calibration, most metallicity calibrations are reasonably reliable (Kewley & Ellison, 2008).

In order to obtain a metallicity directly, the gas temperature is required; the simplest way of acquiring a gas temperature is to use the ratio of the strengths of two forbidden lines produced by the same ion species. Using the same species ensures that the calculation is not dependent on both ions being present in the gas cloud at exactly the same abundance (generally an unlikely prospect). The flux of the higher energy line, relative to the lower energy line can provide a handle on the temperature of the gas; the higher the temperature, the more dominant the high energy flux will be. For oxygen, the commonly used forbidden lines are [OIII]λ4363 and [OIII]λ5007. By taking the ratio of the [OIII]λ5007 to [OIII]λ4363 lines, it should be possible to directly work out the temperature. In practice, however, the [OIII]λ4363 line is difficult to detect. Oxygen is a good radiative coolant, so the oxygen lines are easiest to detect at low metallicities, where the temperatures of the gas are still high enough to excite the collisional forbidden lines. As the metallicity increases, the gas will have more coolant atoms, and will no longer have the temperature required to continue to excite the high energy states. Even in low metallicity regions, however, the [OIII]λ4363 line is hard to detect, so methods using either different oxygen lines or different elements altogether are usually required.

When the direct method is unavailable, there are two alternatives. The first is an empirical calibration wherein the direct method described above is calibrated to the emission line ratios of other metals (such as nitrogen) to hydrogen, where both calibrations are possible. The second is to use the results of photoionization models. The predicted strengths of the line ratios from the photoionization models

depend both on the metallicity and on the ionization parameter (the fraction of ionizing photons relative to the number of hydrogen atoms). One of the most common calibrations is R_{23} (Pagel et al., 1979), which uses the following line ratios:

$$R_{23} = \frac{[\mathrm{OII}]\lambda 3727 + [\mathrm{OIII}]\lambda\lambda 4959 + 5007}{\mathrm{H}\beta} \tag{2.6}$$

This metallicity calibration can be used either with the empirical calibration, or with photoionisation models. The models produce an expected curve for how R_{23} varies with metallicity. Unfortunately, the R_{23} method is degenerate with very high and very low metallicities, as shown in Figure 2.11. An additional set of line ratios is needed to determine whether the gas sits on the upper or lower branch (which, in this figure, are on the right and left sides of the diagram respectively). Nitrogen, like oxygen, has a collisionally excited forbidden line ([NII]λ6583), and the ratio between the [NII] and [OII] lines is often used to break this degeneracy in the R_{23} method, in part because the [NII]/[OII] ratio is not as sensitive to the ionization parameter as [NII]/Hα, but is highly sensitive to metallicity, as shown in Figure 2.12.

For the calculation that follows, a metallicity calibration using the results of the photoionization models will be applied to the data. This will require the emission line fluxes in Table 2.5; note that not all of the galaxies have every emission line. However, since the first step of the metallicity calibration is to break the degeneracy of the R_{23} calibration, depending on whether the galaxy falls on the 'upper branch' and 'lower branch' of the metallicity calibration, a different set of emission lines will be required. Only the emission lines required for the branch in which the galaxy falls are required for the metallicity calculation, so the metallicity calibration will not necessarily be limited by (for instance) the number of galaxies with a reliable [OIII]λ4959 flux, which is the emission line with the lowest number statistics.

2.5.1 KD02-KE08 method

The metallicity calibration that was chosen is the adaptation of the Kewley & Dopita (2002, henceforth KD02) recommended method presented in Kewley & Ellison (2008, KE08). This calibration has the advantage of having low intrinsic scatter, and is easily converted into a number of other metallicity calibrations without significant residuals (KE08). It also does not require the presence of an [SII] line, which will help keep the final sample large. This adaptation of the KD02 method will be referred to as the KD02-KE08 calibration. The metallicity calibration for the upper branch

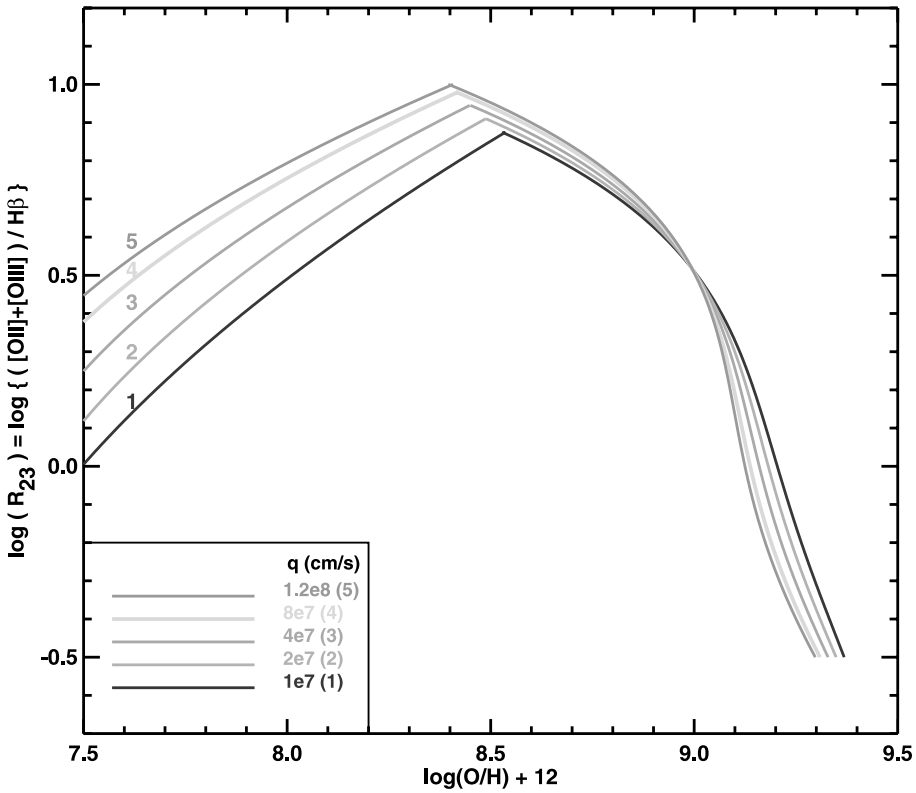


Figure 2.11: Figure 7 of Kobulnicky & Kewley (2004). R_{23} values are plotted for changing ionization parameter, as a function of resultant metallicity. As R_{23} can produce both a high and low metallicity value, an additional line ratio is needed to determine the side of the R_{23} curve upon which a galaxy falls.

Line	K01
Hα	146,753
Hβ	146,753
[OIII]λ5007	146,753
[NII]λ6584	146,753
[OII]λ3726	92,435
[OII]λ3729	104,073
[OIII]λ4959	61,405

Table 2.5: The number of star forming galaxies with significant emission line detections, after the Kewley et al. (2001) lines diagnostic line is imposed. The first four emission lines were required to create the diagnostic diagram, so these define the sample. The remaining 3 lines are the last of the emission lines needed for metallicity calibrations.

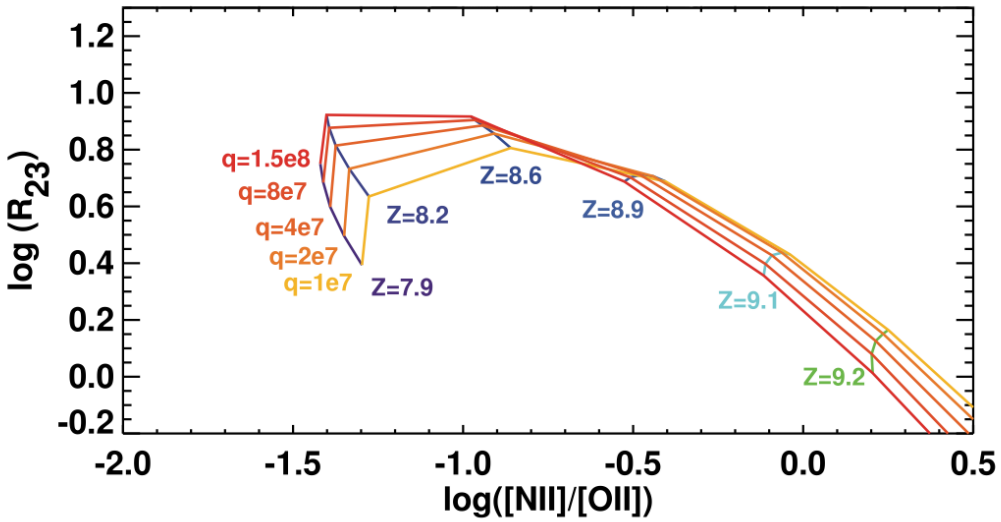


Figure 2.12: Figure 8 of Kewley & Ellison (2008), which shows the results of the photoionization models described in Kewley & Dopita (2002). R_{23} values are plotted as a function of the [NII]/[OII] ratio. The red to yellow lines describe different ionization parameters, while the purple to blue lines describe different metallicities.

remains the same as the KD02 method, but the lower branch has changed. Instead of the average of the Zaritsky et al. (1994, Z94), which is an R_{23} calibration (see Equation 2.9), and McGaugh (1991, M91) methods, KE08 use the average of the Kobulnicky & Kewley (2004, KK04) and M91 methods. This calibration requires line fluxes for Hα, Hβ, [OII]λ3727, [OIII]λ4959, [OIII]λ5007, and [NII]λ6584.

The upper branch from the KD02 recommended method, defined for those galaxies which have $\log([\text{NII}]/[\text{OII}]) > -1.2$ is described by:

$$\begin{aligned}\log([\text{NII}]/[\text{OII}]) = 1106.8660 - 532.15451Z + 96.373260Z^2 - \\ -7.8106123Z^3 + 0.23928247Z^4\end{aligned} \tag{2.7}$$

where $Z = \log(\text{O/H}) + 12$, which is the metallicity value.

For the lower branch, for galaxies with $\log([\text{NII}]/[\text{OII}]) < -1.2$, the averages of the values given by Equations 2.12 & 2.8 are taken.

The M91 lower branch is described by:

$$\begin{aligned}12 + \log([\text{NII}]/[\text{OII}]) = 12 - 4.944 + 0.767x + 0.602x^2 - \\ -y(0.29 + 0.332x - 0.331x^2)\end{aligned} \tag{2.8}$$

where x = $\log(R_{23})$, y = $\log(O_{32})$, and:

$$R_{23} = \frac{([\text{OII}]\lambda3727 + [\text{OIII}]\lambda4959 + [\text{OIII}]\lambda5007)}{\text{H}\beta} \tag{2.9}$$

$$O_{32} = \frac{([\text{OIII}]\lambda4959 + [\text{OIII}]\lambda5007)}{[\text{OII}]\lambda3727}. \tag{2.10}$$

The KK04 lower branch is described by:

$$\begin{aligned}\log(\text{q}) = 32.81 - 1.153y^2 + [Z](-3.396 - 0.025y + 0.1444y^2)\times \\ \times\{4.603 - 0.3119y - 0.163y^2 + [Z](-0.48 + 0.0271y + 0.02037y^2)\}^{-1}\end{aligned} \tag{2.11}$$

$$Z = 9.40 + 4.65x - 3.17x^2 - \log(\text{q})(0.272 + 0.547x - 0.513x^2) \tag{2.12}$$

Where x and y are, as above, x = $\log(R_{23})$, y = $\log(O_{32})$, and q is the ionization parameter. q is defined as the ratio of the number of ionizing photons over the number of hydrogen atoms. Equations 2.11 & 2.12 require an iterative approach;

3 iterations of the two equations are generally required before q and 12+log(O/H) converge (Kewley & Ellison, 2008).

The master star forming sample of galaxies (i.e., star forming according to the Kewley et al., 2001 curve), is run through this algorithm. Only galaxies with significant detections for the lines required for their calculations ([NII]λ6584, [OII]λ3727, & [OII]λ3729 for all galaxies, plus [OIII]λ4959, [OIII]λ5007, and Hβ for lower branch galaxies) had metallicities calibrated for them. [OII]λ3727 was calculated by summing the [OII]λ3726 and [OII]λ3729 line fluxes.

To ensure that these metallicities had been properly calibrated, the galaxies in this sample were cross-matched to the galaxies present in Ellison et al. (2008b). The metallicities for Ellison et al. (2008b) use a previous data release of the SDSS (DR4), and were plotted against the DR7 values calculated here in Figure 2.13. The gap at log(O/H) + 12 $\sim$ 8.4 is a function of the calibrations switching equations from upper to lower branch. It is a feature intrinsic to the calibration, as evidenced by the gap's presence in both horizontal and vertical directions (and thus both the DR4 and DR7 samples.) The calibrations are approximately 1-1; the higher log(O/H) + 12 values for DR7 are largely due to higher (or newly significant) [OII]λ3727 fluxes in the DR7.

After all metallicity calibrations are complete for those galaxies for which the values could be computed, a sample of 109,647 galaxies remains in the Kewley et al. (2001) sample.

2.6 Duplicates

Due to plate overlap in the SDSS, some galaxies have been observed more than once, and assigned a second, unique plate-fibre-MJD number. The MPA-Garching team provides a list of duplicates[4] so that the duplicated galaxies can be identified. To identify which of the two or more galaxies was kept as the primary, the sample of galaxies with metallicities were matched, using the plate-fibre-MJD identifier to a unique SDSS object id. Once the duplicates were removed from the sample, the final metallicity samples are indicated in Table 2.6.

The same process can be undertaken to identify the unique galaxies with AGN classifications. Once the duplicates were removed from the AGN classified sample, the final sample of classifiable galaxies is reduced to 124,042 galaxies with S/N $>$ 5, and

[4]Available at this url: `http://www.mpa-garching.mpg.de/SDSS/DR7/Data/all_matches_dr7.dat`

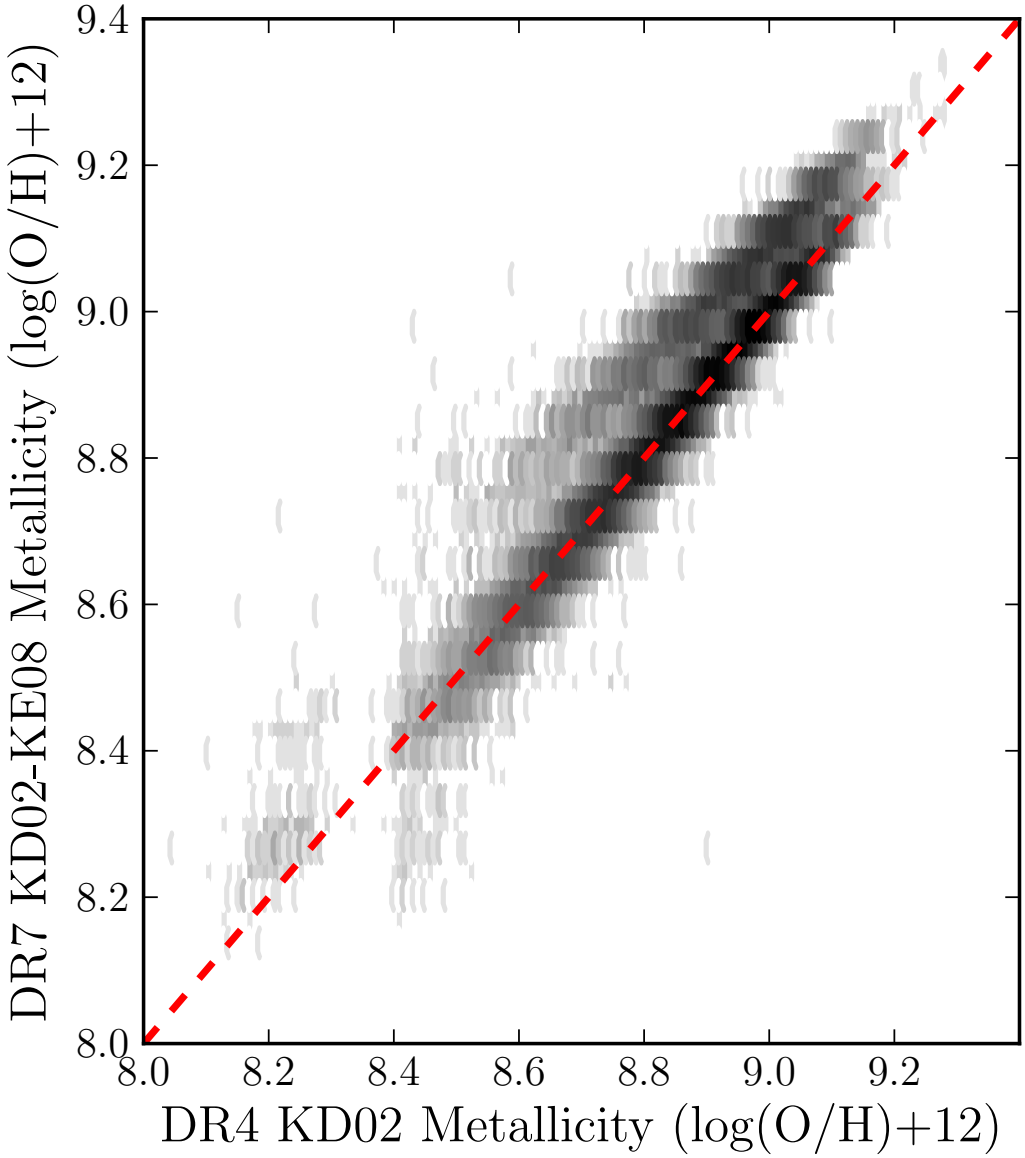


Figure 2.13: Metallicity calibrations in the DR7 (following the Stasińska et al., 2006 curve) vs. the sample of galaxies from Ellison et al. (2008b) that exist in both samples. From left to right, the KD02-KE08 calibration is plotted against the DR4 KD02, DR4 KK04, and DR4 T04 samples. The red line is 1-1.

Stasińska et al. (2006)	Kauffmann et al. (2003)	Kewley et al. (2001)
65,360	75,859	83,543

Table 2.6: Unique metallicity sample.

400,283 galaxies in the S/N > 1 sample. The final star forming sample is therefore shown in Table 2.7 for the three diagnostic lines and the two S/N samples, independent of whether or not a metallicity is calculable.

S/N	Stasińska et al. (2006)	Kauffmann et al. (2003)	Kewley et al. (2001)
> 5	81,259	97,478	115,361
> 1	173,054	234,471	332,797

Table 2.7: Number of galaxies classified as star forming, for the three different classifications.

Throughout this thesis, $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$ and $H_0 = 70$ km s^{-1} Mpc^{-1} are assumed.

Chapter 3

The dependence of galaxy group star formation rates and metallicities on large scale environment

In the previous chapter, the compilation of physical properties from the MPA-JHU catalogues, as well as the detailed process behind the AGN classifications and calculation of metallicity values was described for a sample of 115,361 star-forming galaxies. This large data set now allows for a determination of how the properties of these galaxies are affected by their environment. In this chapter, the role of large scale versus small scale structure is investigated. Since local and large scale effects can be difficult to disentangle and compare cleanly, a sample of galaxies found within a consistently extreme, and relatively isolated, local environment was used. By comparing the broader environment in which these compact groups of galaxies were found, and the star forming and metallicity properties of the galaxies within them, it was possible to determine that the star forming galaxies do, in fact, feel the influence of the large scale environment in which they are found. The results presented in this chapter are published in Scudder et al. (2012a)[1].

[1]Scudder, J. M., Ellison, S. L., & Mendel, J. T. 2012a, MNRAS, 423, 2690

3.1 Introduction

The evolution of a galaxy is fundamentally shaped by its environment. Qualitative evidence of the influence of environment has been known since the 1930s (e.g., Hubble & Humason, 1931); the large surveys now available have added their statistical weight to its influence, allowing a quantification of environmental changes. Nearly every observable quantity of a galaxy has been shown to vary with environment, including star formation rates (SFRs), morphologies, colours, active galactic nucleus (AGN) fractions, and mean stellar mass (e.g., Dressler, 1980; Balogh et al., 2004; Kauffmann et al., 2004; Croton et al., 2005; Baldry et al., 2006; Park et al., 2007). The primary mechanisms that can operate in denser environments are strangulation, harassment, and ram-pressure stripping (e.g., Gunn & Gott, 1972; Larson et al., 1980; Moore et al., 1996). However, disentangling which of these might be a fundamental variable and which are corollary effects is a non-trivial task (e.g., Whitmore & Gilmore, 1991; Blanton et al., 2005; Blanton & Berlind, 2007; Skibba et al., 2009). Moreover, relatively simple processes such as individual galaxy–galaxy interactions become complicated by the influence of a local, high density environment. For example, at low densities, close galaxy pairs show enhanced SFRs (Lambas et al., 2003; Alonso et al., 2004; Ellison et al., 2010). However, once the galaxy pair is found within a high density environment, the strength of the SFR enhancement diminishes or becomes undetectable (Alonso et al., 2004, 2006; Baldry et al., 2006; Ellison et al., 2010). Clearly, galaxies are sensitive to their environment beyond the scale of their nearest neighbour. The question remains, how far beyond the nearest neighbour does environment matter? Is it merely a local density effect, or does cluster membership, with scales far beyond the local, also impact a galaxy's evolution? This work will attempt to address this question.

Several density trends are well documented in the literature for low redshift galaxy samples. The morphology-density and colour-density relations describe the declining fractions of blue, late-type galaxies with increasing density (Dressler, 1980; Postman & Geller, 1984; Balogh et al., 1997, 1998; Martínez et al., 2002; Tanaka et al., 2004; Baldry et al., 2006). The overall distribution of star formation rates also correlates with density, with the average galactic SFR declining as density increases (Hashimoto et al., 1998; Poggianti et al., 2008; Kauffmann et al., 2004; Cooper et al., 2008). Interpretations of the SFR-density relation are complicated by the increasing fraction of low- (or non-) star forming elliptical galaxies as density increases. The contribution

of early type galaxies will bias the average SFR to lower values at higher densities (e.g., Balogh et al., 2004).

The dependence of SFR on density becomes much less clear for only the star forming subsample of galaxies, with many conflicting interpretations. Some studies find that among the star-forming fraction, there is no residual dependence on density (Couch et al., 2001; Balogh et al., 2004; Tanaka et al., 2004; Weinmann et al., 2006; Patiri et al., 2006; Park et al., 2007; Peng et al., 2010; McGee et al., 2011; Ideue et al., 2012). According to these works, the observed decrease in SFR with increasing density is entirely due to the changing fraction of star-forming galaxies with density, and once this dependence is eliminated, the distribution of SFRs is independent of density. Other studies counter these results, finding that the population change is not enough to account for the entire SFR-density relation, and that a density dependence remains (Balogh et al., 1998; Pimbblet et al., 2002; Gómez et al., 2003; Welikala et al., 2008). Moreover, the scale over which the SFR is influenced by its surroundings is widely debated. Many authors find that it is primarily on the local scale (generally considered to be of order 1 Mpc or less) that the residual SFR-density dependence is found (Hashimoto et al., 1998; Carter et al., 2001; Lewis et al., 2002; Kauffmann et al., 2004; Blanton et al., 2006; Blanton & Berlind, 2007). By contrast, some studies indicate that the environment on scales of order several Mpc is of greater importance than that of sub-Mpc scales (Goto et al., 2003; Park & Hwang, 2009).

A further probe into the scales involved in environmental effects comes from the gas phase metallicity, which, like the SFR, can be modulated by the gas supply (Ellison et al., 2008a; Mannucci et al., 2010; Lara-López et al., 2010; Yates et al., 2012). In samples of interacting pairs, changes in the gas phase metallicity and SFR are linked, and have been shown to respond to some of the same physical processes (Scudder et al., 2012b). Theory offers the picture of an interaction triggering large scale gas flow from the outer regions of a galaxy to the centre, and the enhanced densities of gas sparking star formation (Mihos & Hernquist, 1996; Barnes & Hernquist, 1996; Barnes, 2004; Rupke et al., 2010a; Montuori et al., 2010; Torrey et al., 2012). In agreement with this theoretical picture, close galaxy pairs show enhanced SFRs (Larson & Tinsley, 1978; Donzelli & Pastoriza, 1997; Barton et al., 2000; Lambas et al., 2003; Alonso et al., 2004, 2006; Woods & Geller, 2007; Ellison et al., 2008b, 2010) and diluted metallicities (Kewley et al., 2006; Ellison et al., 2008b; Michel-Dansac et al., 2008). This indicates that gas phase metallicities ought to change, as the SFRs do, with density. In confirmation of this idea, studies of galaxies in clusters and other

dense environments have shown metal enhancement relative to the field (Mouhcine et al., 2007; Cooper et al., 2008; Ellison et al., 2009).

In this chapter, a new tactic to investigate the effects of large scale environmental dependences is employed. Two samples of galaxies are selected with similarly high small scale densities, but with different large scale environments. To this end, the catalogue of 3491 compact group (CG) galaxies in 828 CGs presented in Mendel et al. (2011) is used as a refinement on the sample of McConnachie et al. (2009, henceforth M09). Although the original CG selection criteria were designed to identify groups that are not part of a larger overdensity (Hickson 1982; M09), Mendel et al. (2011) find that the SDSS CG sample can be divided into two distinct populations. One population appears to be truly isolated, whereas the other population appears embedded within a large scale structure, but isolated from other galaxies within the group. Mendel et al. (2011) show that despite the strong differences in their large scale environments, the galaxies in these 'embedded' and 'isolated' groups have similar morphological properties. In this chapter, the spectroscopic properties of the embedded and isolated CG galaxies are investigated, focussing on their star formation rates and metallicities. The photometric catalogue of CG galaxies is therefore cross-matched with a spectroscopic sample of emission line galaxies from the SDSS DR7 (Abazajian et al., 2009). In Section 3.2, the sample selection is described, including metallicity and AGN calibrations. In Section 3.3, the methods for quantifying differences between the isolated and embedded CG galaxies and a sample of control galaxies is described. In Section 3.4, the implications of these results are discussed, along with a comparison to previous works, and present the final conclusions in Section 3.5.

3.2 Sample Selection

As described in Chapter 2, the spectroscopic sample is taken from the publicly available MPA-JHU SDSS DR7 catalogue [2] of 927,552 galaxies. This catalogue provides measurements of up to 12 emission lines per galaxy, corrected for stellar absorption lines and Galactic reddening. Of these lines, robust flux measurements are required to be present in Hα, Hβ, [OII]λ3727, [OIII]$\lambda\lambda$4959, 5007, and [NII]λ6584 for metallicity

[2] Available at: `http://www.mpa-garching.mpg.de/SDSS/DR7/raw_data.html`

calibrations[3]. As some of these galaxies are not unique objects within the catalogue, the MPA-JHU's duplicate catalogue [4] is used to remove duplicate galaxies from the sample. The quality control measures described in Chapter 2 are applied to the full MPA-JHU catalogue in order to select a reliable sample of star-forming galaxies with measurable metallicities. A summary of the final quality control requirements are listed in Table 2.1, and a detailed description of the metallicity calibrations are provided in Section 2.5.

Final Metallicity sample

The final metallicity sample is 75,863 galaxies and represents both the sample with which the CG sample is cross-matched and the basis of the pool from which a control sample is constructed. Stellar mass[5] values are as calculated by the MPA/JHU group from fitting models to the 5-band SDSS DR7 photometry [6]. These masses are, in general, in good agreement with spectroscopically derived values calculated by Kauffmann et al. (2003). Star formation rates within the SDSS fibre (3″) are taken from Brinchmann et al. (2004), who use a set of 6 emission line fits to determine the SFR. Aperture corrected SFRs are also present in the catalogue. These SFRs are corrected from fibre to total values by calculating the galaxy light not contained within the fibre; by fitting models to the photometry of the galaxy outside the fibre, the aperture correction can be effectively made (See Brinchmann et al. 2004 for a full discussion of their aperture correction methodology). Salim et al. (2007) compare the results of this methodology with the SFRs obtained from UV flux, and find that for star forming galaxies the two methods agree very well, with no bias introduced by the aperture corrections. Aperture-corrected SFR values are used for the rest of this chapter.

[3]Recently, Groves et al. (2012) have found that the Hβ equivalent widths from this catalogue are systematically underestimated by 0.35Å. This should not affect the results of this chapter, as all comparisons made here are relative, and the underestimation found is constant across the entire sample.

[4]Available at: `http://www.mpa-garching.mpg.de/SDSS/DR7/Data/all_matches_dr7.dat`

[5]Unless otherwise stated, all mass values in this chapter are stellar masses.

[6]A discussion of their method is available at: `http://www.mpa-garching.mpg.de/SDSS/DR7/mass_comp.html`

3.2.1 Compact Group Sample

The parent CG sample consists of 3491 CG galaxies in 828 CGs (Mendel et al., 2011). This sample is a refinement of the Catalogue A of M09. Typical intergalactic distances within the groups are a few tens of kpc, ranging up to 140 kpc. The M09 CG sample applies a modification of the original criteria laid out by Hickson (1982) to the SDSS DR6, with isolation, minimum number of galaxies, and density requirements:

1. $N(\Delta m = 3) \geq 4$
2. $\theta_N \geq 3\theta_G$
3. $\mu_e \leq 26.0$ mag arcsec^{-2}

$N(\Delta m = 3)$ is the number of galaxies within 3 magnitudes of the brightest galaxy within the group, μ_e is the surface brightness of the group, θ_G is the angular size of the group, and θ_N is the size of the circle beyond which there are no galaxies within the 3 magnitudes required for group membership. These criteria guarantee that there must be at least 4 galaxies in the group and the group itself must be separated by at least 3 times its own radius from any other equally bright galaxies. The surface brightness criterion guarantees the compactness of the group. Applying these criteria, M09 find 2297 CGs, containing 9713 galaxies, down to a limiting magnitude of $r = 18$. However, there is no redshift criterion in the identification process for these galaxies; galaxies identified as a CG are done so based only on their photometry, regardless of whether or not they have concordant redshifts. Approximately 50% of the original 9713 galaxies have reliable spectroscopic redshifts in the DR7 (Mendel et al., 2011).

In order to address the issue of false (due to projections) CGs in the M09 sample, Mendel et al. (2011) apply a statistical likelihood restriction to the master CG sample. Briefly, by looking at the probability density functions for the combinations of spectroscopic and photometric redshifts available for CG galaxies found within a given group, a large fraction of interloping galaxies can be removed. If this process of interloper rejection results in the group failing the richness criterion used to define the sample, the entire group is rejected. Mendel et al. (2011) estimate that there could be 20%-30% contamination remaining in the cleaned sample. Mendel et al. (2011) identified a distinction in the distribution of distances between the CG and the next nearest cluster or rich group. Approximately 50% of the cleaned CG sample can be classified as within ≤ 1 Mpc h^{-1} of a rich group, using the Tago et al. (2010)

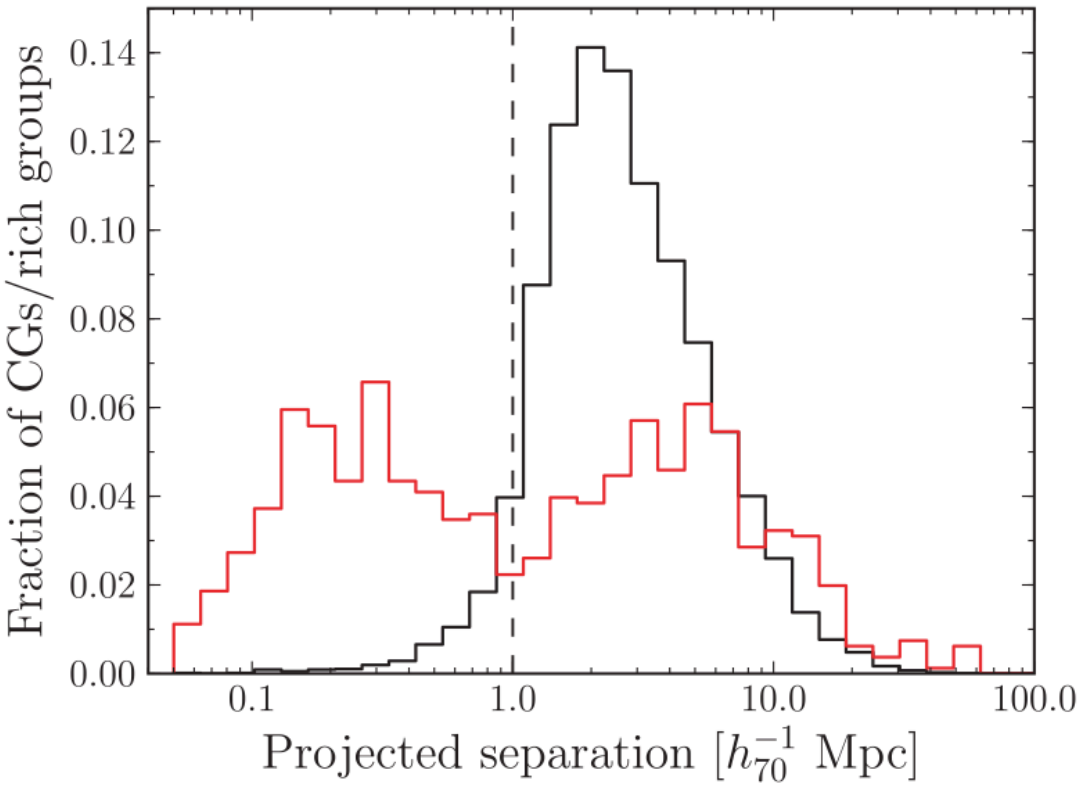


Figure 3.1: Figure 5 from Mendel et al. (2011). "Distribution of rich group–CG (red line) and rich group–rich group (black line) projected separations. Projected separations are given to the nearest rich group within $\Delta z = 0.02$. The distribution of rich group–CG separations is found to be bimodal, suggesting that CGs can be separated into those residing within the halo of a rich group (left peak) and those more likely associated with structures of comparable richness to the CGs themselves (right peak); the vertical dashed line shows the adopted division between these two CG populations."

catalogue compiled from the DR7 (see Figure 5 in Mendel et al., 2011). The other half of the sample is distributed at distances > 1 Mpc away from group structures. Placing a dividing line in the CGs at the minimum between the two distributions, Mendel et al. (2011) describe two populations: 'embedded' and 'isolated' CGs (Figure 3.1). This terminology is adopted for the galaxies residing within those groups. Mendel et al. (2011) shows that the galaxies residing within the two density bins have systematically different photometric properties (i.e., an increase in the fraction of blue galaxies within isolated CGs).

The galaxies belonging to either embedded or isolated CGs are cross-matched with the final metallicity and SFR sample to select only the galaxies residing in CGs

that also have robust metallicities and SFRs. This cross-matching results in our final CG galaxy sample of 112 galaxies. Splitting the sample into embedded and isolated CGs results in 62 galaxies in isolated CGs, and 50 galaxies in CGs embedded within a large scale structure. For a list of the embedded and isolated CG galaxies and their properties, see Tables A.1 & A.2 respectively. SDSS thumbnails of 4 random embedded CGs and 4 isolated CGs are presented in Figures 3.2 & 3.3 respectively.

3.2.2 Matching to Controls

In order to make an effective comparison between the CG sample of galaxies and a control sample of galaxies, any major sources of bias must be eliminated. The three major sources of bias are stellar mass, redshift, and environment (Perez et al., 2009). As this work is looking for an effect that varies with environment, it is impossible to match in that property, but a control sample will be matched to the CG sample in stellar mass and redshift, such that the distribution of stellar mass and redshift in the control match the distributions in the CG sample. The control pool is defined as any galaxy in the final metallicity sample that is not flagged as belonging to a CG.

Galaxies in the CG sample are matched to galaxies in the control pool simultaneously in total stellar mass and redshift, in a similar way to Ellison et al. (2008b). Briefly, this algorithm finds the best simultaneous match in mass and redshift to each galaxy in the CG sample. Once every galaxy has been matched to a control, a Kolmogorov-Smirnov (KS) test is used on the total distributions of the CG sample and control. If the KS test finds that the distributions are consistent with being drawn from the same parent distribution at $> 30\%$, then the matching continues without replacement until a maximum of 50 control galaxies have been matched to each CG galaxy in the sample, or the KS-test results in a probability $< 30\%$. Since the pool of possible control galaxies is large compared to the CG galaxy sample, the matching procedure reaches the full capacity of 50 matches per CG galaxy. The final CG samples of 62 isolated CGs and 50 embedded CGs thus have control samples of 3100 and 2500 galaxies respectively. The masses and redshifts of the control galaxies are typically matched to within 0.05 dex and 0.003 respectively of the CG galaxy value. The E(B−V) values of the galaxies are found to be consistent between all four samples when a KS-test is applied.

The resulting normalized distributions of mass and redshift are shown in Figure 3.4. The control samples and the CG samples can be seen to match very well in both

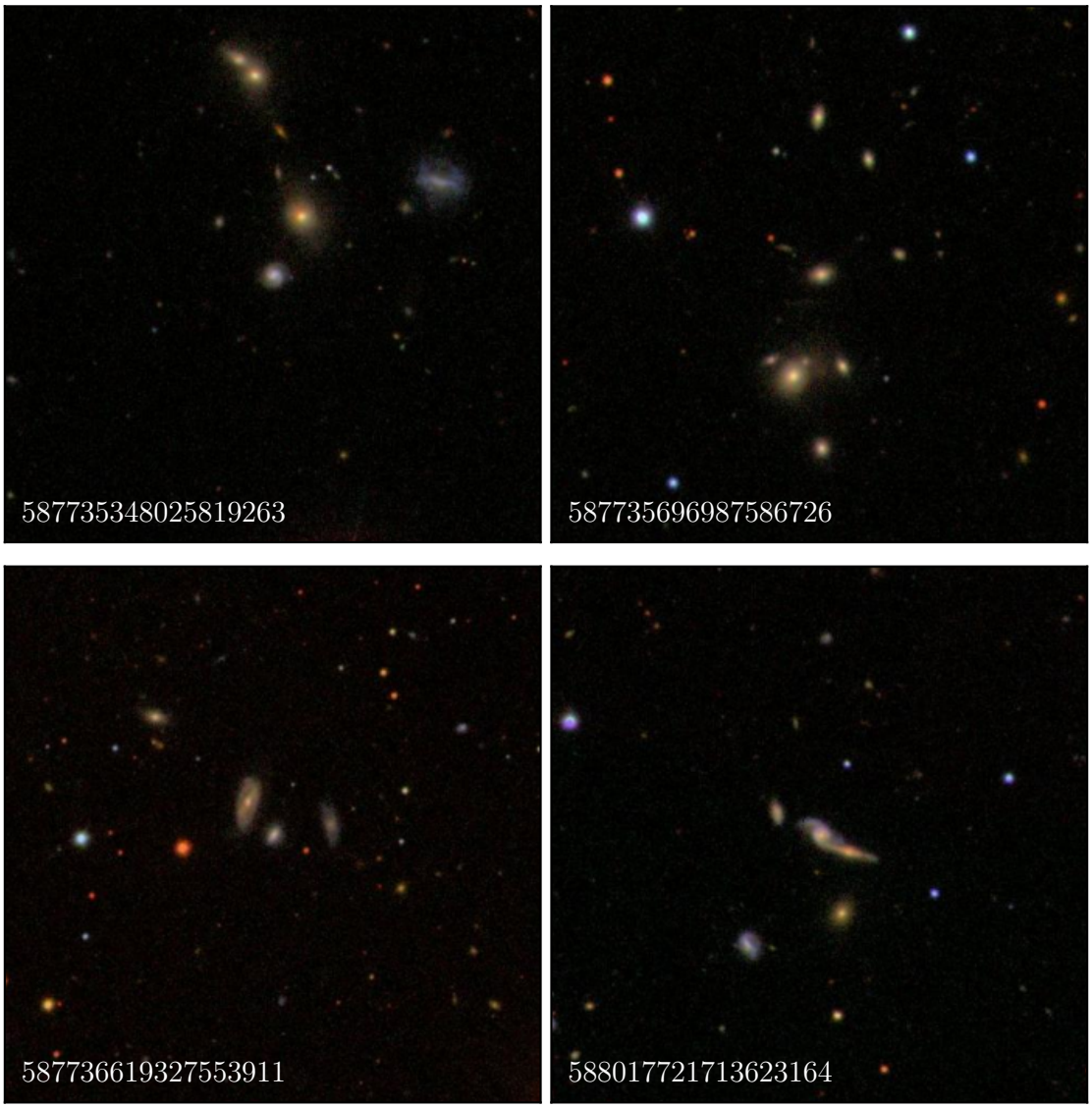


Figure 3.2: 4 embedded CG galaxies, selected at random. The images are centred on the galaxy in the spectroscopic sample (identified by SDSS objID in the lower left corner), and are approximately 200 arcseconds to a side.

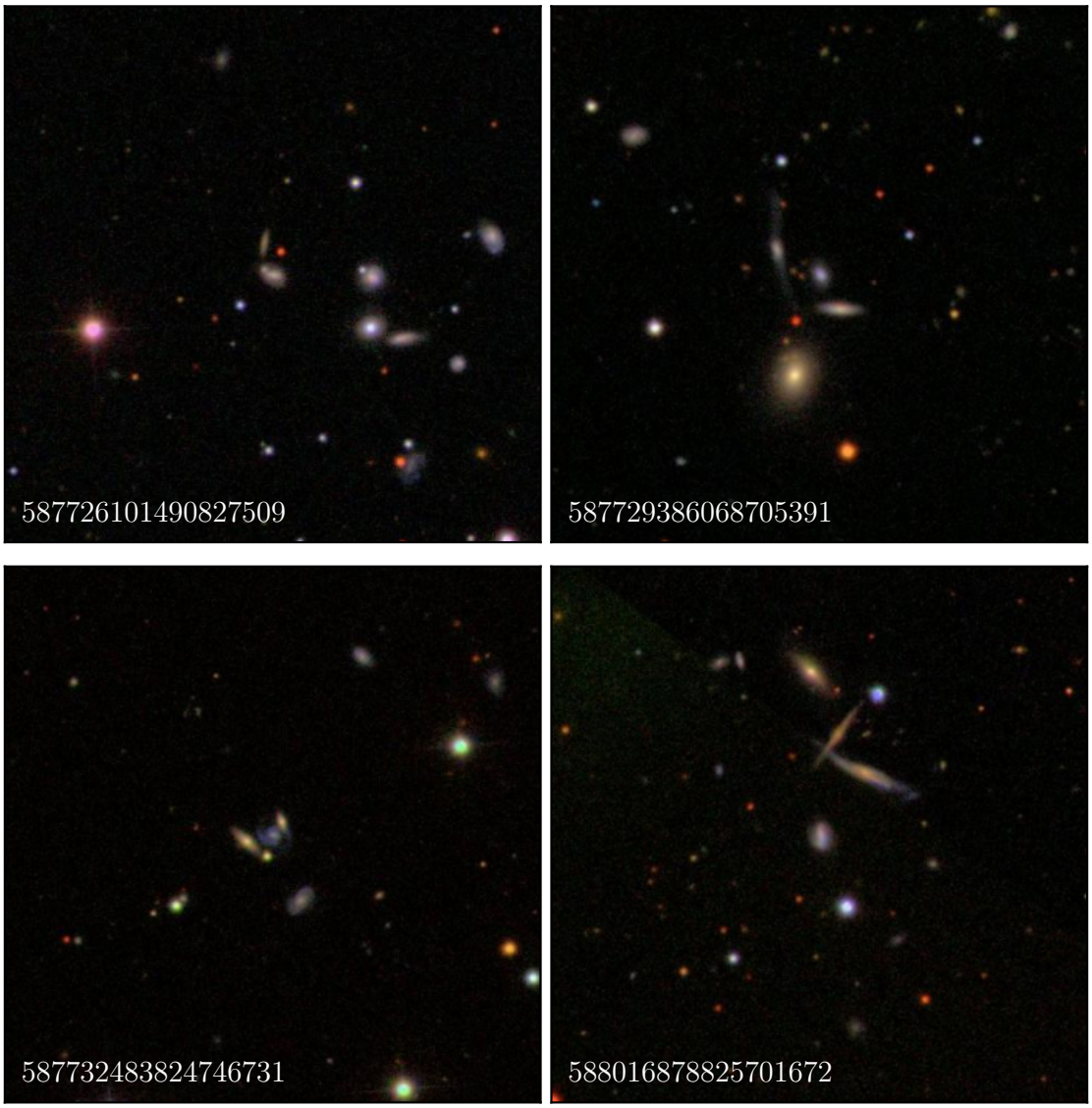


Figure 3.3: 4 isolated CG galaxies, selected at random. The images are centred on the galaxy contained in the spectroscopic sample (identified by SDSS objID in the lower left corner), and are approximately 200 arcseconds to a side.

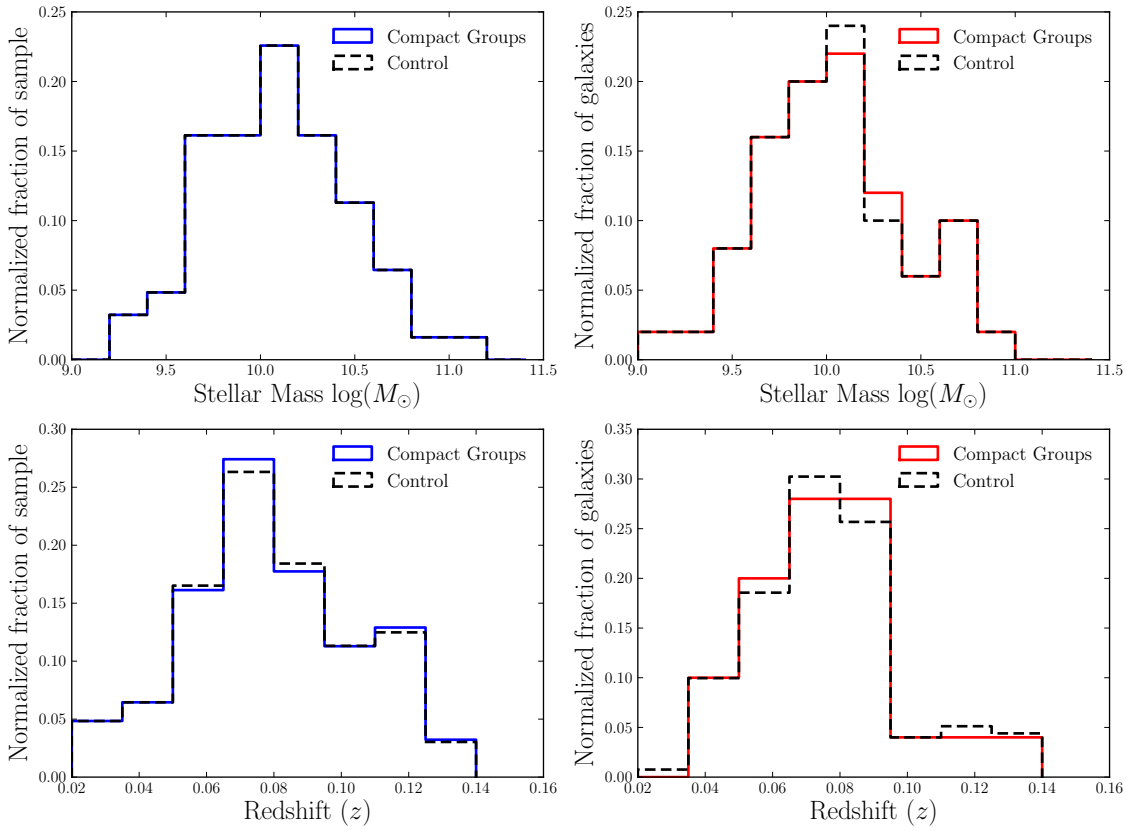


Figure 3.4: Normalized distributions of isolated (left panels) and embedded (right panels) CG galaxies and their controls for total stellar mass (upper panels) and redshift (lower panels). Galaxies are matched to controls in total stellar mass and redshift simultaneously until 50 control galaxies are matched to each group galaxy. KS-test probabilities that the samples are drawn from the same population are $> 99\%$ for the control and CG galaxy distributions in all panels.

mass and redshift (the KS-test probabilities are $> 99\%$ between CG galaxies and their controls for both mass and redshift distributions). The embedded and isolated galaxies are also statistically similar to each other in both stellar mass and redshift, with KS-test values of 32.57% and 49.16% respectively.

3.3 Offsets in the SFR and Metallicity of Compact Group Galaxies

In this section, the methodology for finding and quantifying differences between the SFRs and metallicities of compact group galaxies and their controls is presented. Also

presented are a set of bootstrap simulations to quantify the significance of a given measured offset.

3.3.1 Offset methodology

To quantify shifts away from the expected metallicities and SFRs at a given mass, a modification of the offset method from Patton et al. (2011) is used. For each galaxy in the CG sample, the median (metallicity or SFR) value of all 50 control galaxies to which it was matched is taken, and that value is subtracted from the observed value of the CG galaxy. An offset is therefore defined as:

$$\Delta\log(\mathrm{O/H}) = (\log(\mathrm{O/H}) + 12)_{observed} - \mu_{1/2}\,(\log(\mathrm{O/H}) + 12)_{controls}\,, \tag{3.1}$$

$$\Delta\log(\mathrm{SFR}) = \log(\mathrm{SFR})_{observed} - \mu_{1/2}\log(\mathrm{SFR})_{controls}, \tag{3.2}$$

where $\mu_{1/2}$ signifies the median. The calculated offsets are presented in Figures 3.5 and 3.6. The median metallicity offset for the isolated CG galaxies (solid vertical line in Figure 3.5) is -0.02 ± 0.04 dex (metal poor), whereas the embedded population has a median metallicity offset of 0.00 ± 0.02 dex (no offset; dashed vertical line).

The uncertainty on the median is calculated using a jackknife technique. For each sample (embedded and isolated) the median offset is re-calculated by systematically removing one galaxy. For a sample of N galaxies, the jackknife error on the median is given by

$$\sigma = \sqrt{\frac{N-1}{N} \times \sum^{N} \left(\mu_{1/2}(N) - \mu_{1/2}(N-1)\right)^2} \tag{3.3}$$

where $\mu_{1/2}(N)$ is the median value for the full sample of N galaxies, $\mu_{1/2}(N-1)$ is the median value for $N-1$ galaxies, and each of N galaxies is removed from the sample once.

The SFR offset distribution is shown in Figure 3.6; the embedded CG galaxies have SFRs that are lower than the control, with a median offset of -0.03 ± 0.05 dex. Isolated CG galaxies, by contrast, have a median offset of $+0.07\pm0.03$. Again, uncertainties are determined from the jackknife statistic.

Since the jackknife statistic can be inconsistent for a distribution of medians, the

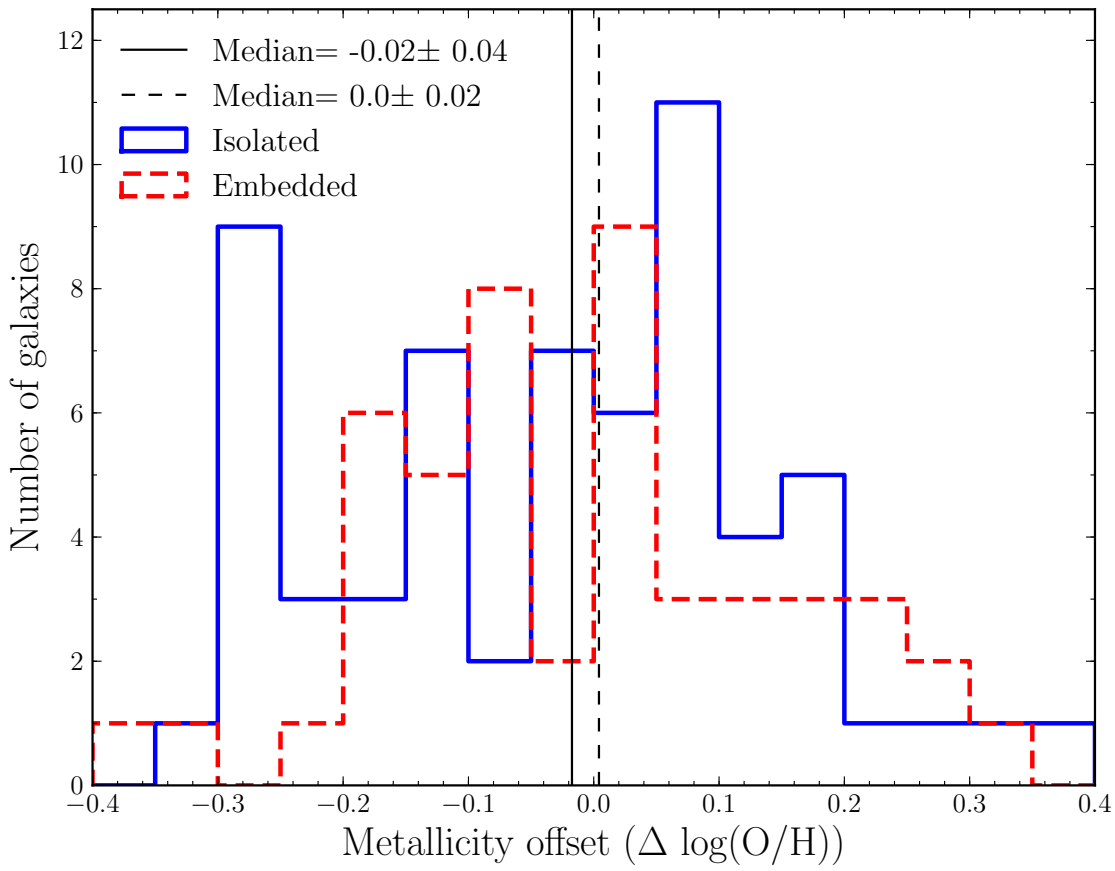


Figure 3.5: Offsets from the control MZR as defined by the median value per set of control galaxies for the embedded (red dashed line) and isolated (blue solid) compact group sample (50 and 62 galaxies, respectively). Black solid and dashed lines show the median offsets for the isolated and embedded compact group galaxies. Embedded galaxies show no median metallicity enrichment, contrasted with the median metallicity offset of isolated groups, which show a dilution of -0.02 ± 0.04 dex compared to the overall relation.

errors on these median values are recalculated using a standard error on the median methodology. The significance (or lack thereof) of the results presented above is unchanged. The isolated median metallicity quoted as -0.02 ± 0.04 using the jackknife method changes to -0.02 ± 0.02, whereas the embedded median metallicity changes from 0.00 ± 0.02 to 0.00 ± 0.01 when the standard error method is used. Neither of the median metallicity measurements are significantly affected by the change in statistic.

Similarly for the SFRs, the embedded CG galaxies show an offset of -0.03 ± 0.05 dex when measured with the jackknife statistic. When this is changed to the standard error on the median, the error is only slightly affected, increasing by 0.01 dex to -0.03 ± 0.06. The isolated CG galaxies go from $+0.07\pm0.03$ to $+0.07\pm0.04$. While still significantly enhanced over the controls, this standard error on the median indicates a slightly lower level of significance than the jackknife error. As the jackknife statistic does not significantly bias the error estimates provided in Figures 3.5 & 3.6, or change whether or not the samples are significantly offset from their controls, for consistency with the published work, the error on the medians quoted for the remainder of this chapter will be that of the jackknife.

In order to determine whether the measured median offsets are strongly influenced by the errors on the SFRs, a Monte-Carlo resampling simulation is used to test the sensitivity of these results. The distribution of possible SFR values is defined as a gaussian with mean of the original SFR, and with width defined as the 1σ errors on the SFRs, taken from Brinchmann et al. (2004). Median errors on any individual galaxy are ~ 0.1dex. New SFR values are drawn for both the CG and the control samples, and the offsets are recalculated for the CG samples. The median offsets are not significantly changed. After 10,000 iterations, the median Δlog(SFR) for isolated CG galaxies is 0.07 with a 1σ scatter in the distribution of 0.05. Embedded CG galaxies show a median Δlog(SFR) of -0.06 ± 0.06 (see Figure 3.7). These simulated median offsets and their uncertainties are consistent with the medians in Figure 3.6; the errors on the SFR measurements are therefore not artificially enhancing the separation between the two samples.

The uncertainties on the median offsets indicate that the metallicities of the galaxies within both embedded and isolated CGs are consistent with a mass and redshift-matched control sample of non-CG galaxies. While the median SFR of the embedded CG galaxies is consistent with the control, the median SFR of the isolated CG galaxies is significantly offset from its control at $\sim 2\sigma$. Further, the isolated and embedded medians are significantly different from each other at the 1σ level.

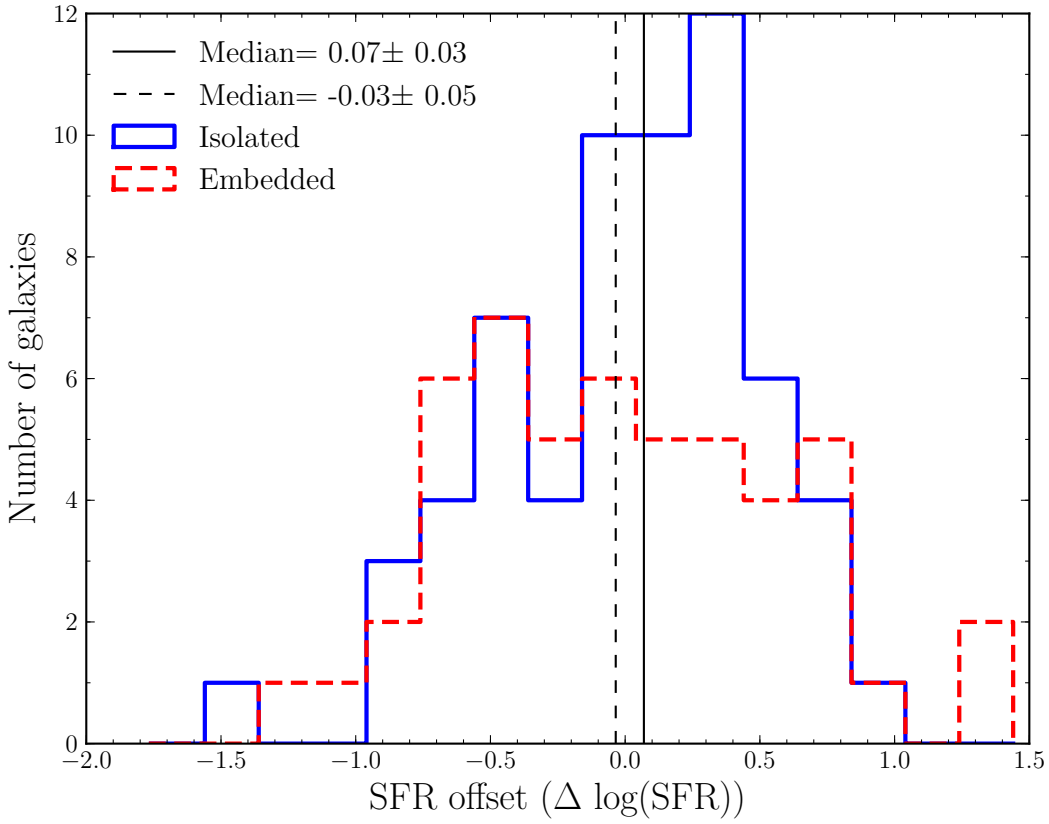


Figure 3.6: Same as Figure 3.5, but now showing the offset of CG galaxies from the SFR − mass relation. Isolated CGs show a median SFR enhancement of $+0.07\pm0.03$ dex, while embedded CG galaxies have SFRs depressed by -0.03 ± 0.05 dex below the control sample.

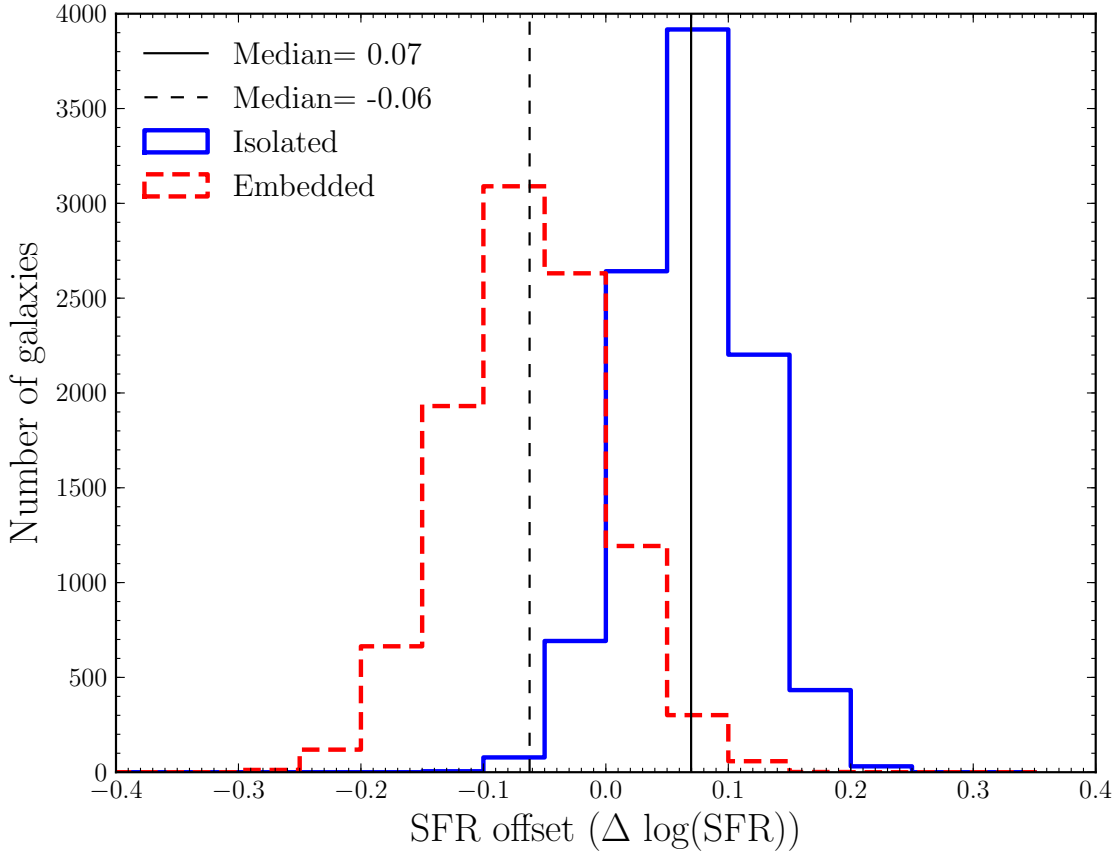


Figure 3.7: Results of a Monte-Carlo simulation on the SFR values of CG galaxies. The distribution of median offsets for the isolated (blue solid) and embedded (red dashed) CG galaxies is shown after 10,000 resamplings of the error distribution. Both control and CG galaxies were re-drawn, and their offsets recalculated. The median of the distribution of median offsets for isolated CGs is $+0.07\pm0.03$ dex, where the uncertainty is a 1σ width of the distribution. Embedded CG galaxies show median SFRs of -0.06 ± 0.06 dex. These distributions indicate that the effect of errors on the SFRs are not likely to change the results of the analysis. KS tests indicate a 0% chance that the two distributions were drawn from the same parent sample.

3.3.2 Significance simulations for the SFRs

The jackknife technique estimates the statistical error on the samples of ~ 60 galaxies. However, a quantification of how often a given median offset will occur by chance (due to the intrinsic scatter in the SFR-mass relation) is also desired. A random sample of 60 galaxies is therefore bootstrapped from the control sample 10,000 times. Recall that each control is one of 50 that has been matched to a given CG galaxy. When one of them is selected at random, the 49 others in the set which are not selected as the 'sample' are left as the 'control'. Offsets are then calculated between the randomly selected galaxy and each of the remaining controls. The median offset for the sample of 60 galaxies is determined, along with the jackknife error on that median. As the galaxies selected are part of the control pool, and no offset is artificially introduced, the majority of these offsets should be centred around zero. For each median, the significance σ is calculated by dividing the measured median offset by the jackknife error on the median. The fraction of the total number of runs with significant offsets (defined as $\sigma > 1, 2, 3$, or 4) can then be calculated.

For any given offset, it is possible to determine the fraction of the total number of bootstraps which returned significant medians at or beyond the offset being investigated. The results are shown in the top panel of Figure 3.8. The different lines indicate different values of σ. To compare to the results presented here, the SFR enhancement seen in isolated CGs would be along the 2σ line. This simulation indicates that there is a 0.19% chance of detecting a SFR enhancement of $+0.07$ ± 0.03 dex at random. Therefore, the SFR enhancement detected, at its given observed significance, has a very low probability of occurring by chance due to scatter.

Whilst the above procedure assesses the probability that a given offset is found between the CG and control samples, it is also possible to quantify the significance of the difference *between* the embedded and isolated samples. This is achieved by repeating the bootstrapping procedure described above, but this time selecting 2 separate samples of 60, and the absolute value of the difference in their median offsets is found. The percentage of iterations where the simulated differences in median is greater than a given offset is then found. The results of the two sample bootstraps are shown in the lower panel of Figure 3.8. The probability of a range of 0.10 dex between two samples drawn at random from the control SFR-mass relation is 0.67%, which corresponds to $> 2\sigma$ significance.

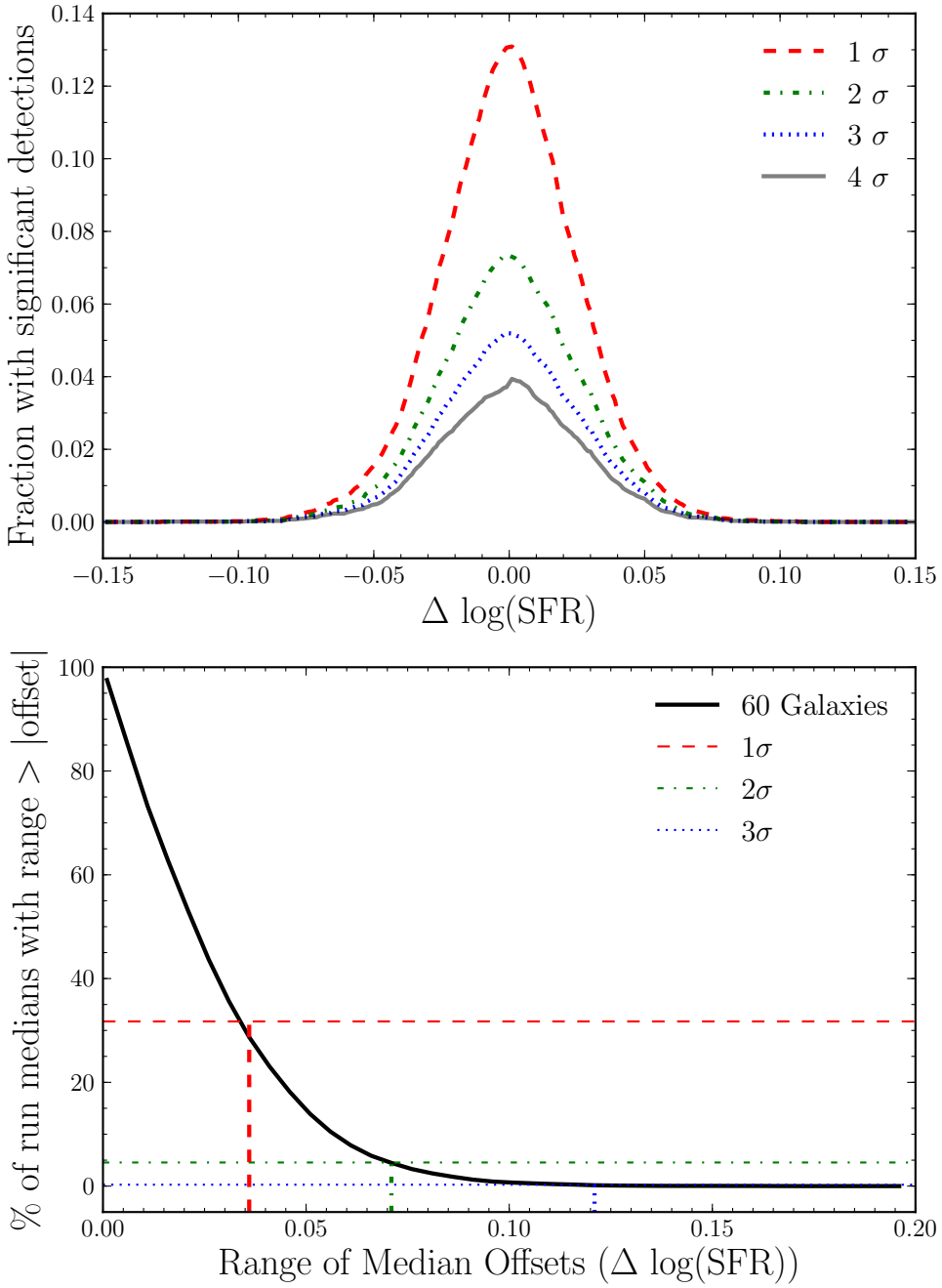


Figure 3.8: Results of the bootstrapping simulation for SFRs. A randomly selected sample of 60 galaxies is pulled from the control pool, and their offsets from their matched controls recorded, repeated 10,000 times. The top panel is the fraction of the total number of runs where a significant offset (i.e., $\sigma > 1, 2, 3$, or 4) greater than a given offset i is found. Offsets of -0.04 and $+0.08$ dex at 2 and 4σ significance have 2.14% and 0.02 % likelihood of arising by chance, respectively. The bottom panel tests the percentage of times that a range in medians between samplings occurs; a difference in median of 0.10 dex could occur randomly $\sim 0.67\%$ of the time ($> 2\sigma$).

3.3.3 Upper limit simulations for metallicities

Although a statistically significant difference between the SFRs of embedded and isolated CGs is found, the metallicity offsets are inconclusive (-0.02 ± 0.04 and 0.00 ± 0.02 for isolated and embedded CGs respectively). Metallicity offsets in both embedded and isolated structures are statistically consistent with their control samples. To test the sensitivity of the null result given the sample size, a set of simulations similar to the bootstrap used in §3.3.2 is run, but now introducing an artificial offset. In this simulation, 60 galaxies are randomly selected from the control sample and an artificial, single value offset (i) is added to all the metallicities of the selected galaxies. The resultant metallicities are then run through the offset calculation algorithm. Offsets are calculated between each of the 60 randomly selected galaxies and the remaining 49 in its matched set. The median measured offset (m) of the 60 galaxies is recorded. This step allows for a quantification of the difference between the input offset i and the recovered offset m via the algorithm. $i - m = 0$ is expected if the sample can accurately recover the input metallicity offset. This bootstrap process is repeated 5,000 times. The jackknife error on each of the median values returned by the bootstrap is calculated. For each realization, we calculate a significance σ by dividing the measured offset m by the jackknife error. The fraction of the total number of bootstrapped samples which return significant offsets for a given i can then be found. This fraction is then recorded, and plotted in the top panel of Figure 3.9. Red, green, and blue points reflect the fractions for $\sigma > 1, 2$, and 3, respectively. Each inserted offset requires a separate, independent run of 5,000 iterations of the bootstrap, and is plotted as a separate point in the figure.

In the bottom panel of Figure 3.9, the median difference between the input and measured offsets ($i - m$) is shown for those runs which returned significant medians at the 3σ level, as a function of the inserted offset i. This panel indicates that it is possible to accurately recover an offset, if it is systematically different from the control. For values of i smaller than ±0.05, ($i - m$) shows an asymmetric shift away from zero. At very small i, it becomes increasingly unlikely for an offset m to be measured as significant. Therefore, those simulations which are measured as significant at small i are likely to be those which have scattered to median values further from zero. Negative values of i therefore preferentially return more negative values of m, and positive values of i return more positive values of m.

With these simulations as a base, the fractions of significant measured offsets re-

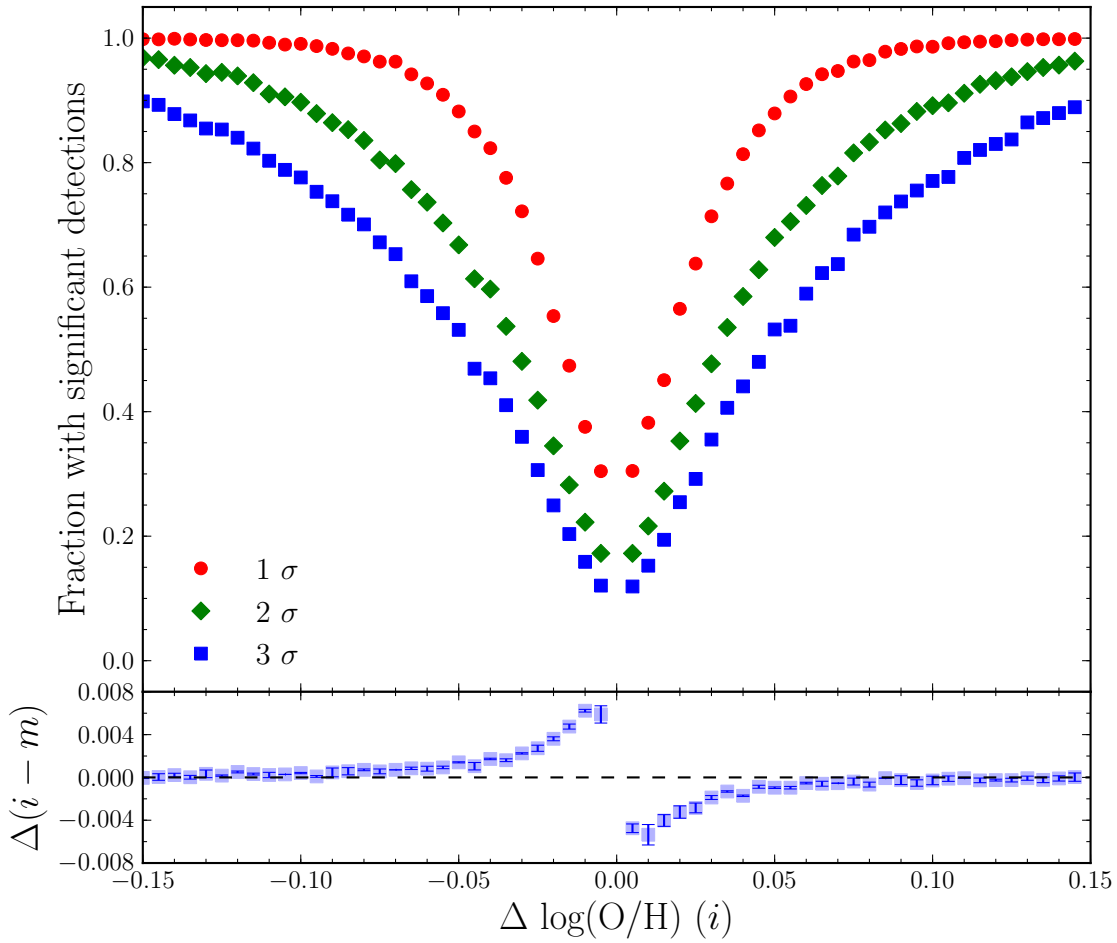


Figure 3.9: Results of the sensitivity simulation for the MZR offsets. Each point is the result of an independent simulation, bootstrapped 5,000 times. Each simulation selects a random sample of 60 galaxies from the control sample, and inserts an artificial offset i into the metallicities. The offsets are then calculated as normal, resulting in a measured offset m. The fraction of those runs which result in significant detections at the 1,2, and 3 σ level are plotted in black, red, and blue, respectively. The median difference between the measured offset m and the input offset i for the fraction of significant runs (at 3σ) is displayed as the blue squares in the lower panel as a function of i. Jackknife errors on the medians are smaller than the points.

turned by any given iteration of the bootstrap may be compared with the magnitude of offset that might be expected. The findings of previous studies of metallicity changes in high density environments may be used as a guideline for reasonable magnitudes of offsets. The metallicity suppressions found in the pairs sample of Ellison et al. (2008b) and of Michel-Dansac et al. (2008) are of order 0.05 dex, and the enhancements in overdensities in Cooper et al. (2008) are ∼0.03 dex. Similar magnitudes are found by Ellison et al. (2009), who find offsets of 0.04 dex in galaxies in high density regions, relative to their control. Mouhcine et al. (2007) find slightly higher differences in metallicity of 0.06–0.08 dex between high and low density environments. For offsets in the range 0.03 – 0.05 dex, there is only a 34.8–51.1% chance of detecting a 3σ result. 99% confidence levels of detecting a 1σ offset with the given sample size are not reached until offsets of 0.13 dex are tested, which is a much larger effect than anything previously discussed in the literature. The lack of significant metallicity offset in the CG sample may therefore be a limitation of the sample size. The fact that the metallicity results are both consistent with the control, combined with the results of this simulation, indicate that the sample is unlikely to result in a significant offset of the magnitude suggested by the literature.

In summary, the offsets in SFR in the CG samples appear to be statistically distinct from each other, and galaxies in isolated CGs have significantly enhanced SFRs, whereas embedded CGs show no enhancement. However, the measured metallicity offsets are not sufficiently sensitive to probe metallicity differences between embedded and isolated compact group galaxies and their controls.

3.4 Discussion

The varying large scale environments of a consistently selected sample of compact, high density groups has been used to probe the influence of the presence of a rich group or cluster. The compact group selection criteria allows the selection of a consistently dense local environment on scales of $\sim$ 100 kpc (see Figure 6 of Mendel et al., 2011) which, ignoring all external influences, ought to be subject to the same internal evolutionary processes. Galaxies in CG systems typically have small projected separations from each other; for the samples presented here, the median projected separation to the nearest neighbour is 46 kpc for both isolated and embedded samples. In galaxy pairs separations $\lesssim$ 30 kpc result in significantly enhanced SFRs (e.g., Barton et al., 2000; Lambas et al., 2003; Alonso et al., 2004, 2006; Ellison et al., 2008b,

2010; Scudder et al., 2012b). This sample has a significant population of galaxies with separations of this order, making galaxy–galaxy interactions an appealing explanation for the SFR enhancement seen in the isolated CG sample. The level of offset seen in the sample of isolated CG galaxies (+0.07 dex) may appear modest, but is comparable to studies of merger-induced star formation, which generally find median SFR enhancements of around +0.1 dex (e.g., Ellison et al., 2008b; Scudder et al., 2012b). If the isolation criterion of the CG selection is effective at limiting outside influences, and interactions are the only dominant mechanism for transformations, both samples of CG galaxies might be expected to show consistent SFR enhancements over the control.

Galaxies in CGs in close proximity to a cluster or rich group (<1 Mpc) do not show the same preferential enhancement as their counterparts in isolated CGs. The embedded CG galaxies instead show consistency with the average galaxy in the control. Galaxies in isolated CGs (>1 Mpc from large scale structures), by contrast, show SFR enhancements over the control, consistent with the general trend presented by the SFR-density relation. This SFR disparity indicates that the local overdensities identified by the CG selection criteria cannot be considered in isolation. The presence of a nearby rich group exerts *some* influence on the star forming galaxies found within CGs, either by introducing an additional quenching mechanism, or because the galaxies within rich groups are physically distinct from their isolated counterparts. The SFRs and metallicities of the galaxies within those groups therefore serve as diagnostics of the impact of the larger scale environment. *However, given the small magnitude of the difference between the two CG samples, the large scale environment does not appear to drive major changes in a galaxy's evolution at fixed local density.*

3.4.1 Density scale dependences

Having determined the sensitivity and significance of the metallicity and SFR offsets, a comparison of these results to previous work is desired. However, previous metrics of environment usually fall into one of two categories; either an n^{th} nearest neighbour approach, or calculating galaxy density within annuli of increasing radius around a given point. In this work, a new method of determining the large scale environment of a galaxy has been used. The division of galaxies into embedded and isolated systems is a binary assignment, based upon the presence (or lack) of rich group structure within a cylinder in redshift and projected separation space, centred upon the CG. This is

an entirely different method of characterizing the nature of a galaxy's environment than has been typically used in the past.

Comparisons to the results found using existing measures of environment will require attention to the methodology of those metrics to be sure that they can probe the same kinds of overdensities. Since the n^{th} neighbour and annulus methods are calculated using entirely different algorithms, the meaning of the densities returned (as it applies to the CG sample) will vary, and comparisons to the results presented here must bear this in mind. In addition, most previous work has focused upon the *relative* importance of local and large scale dependences. This relative comparison is one that cannot be addressed in this work, as the classifications of embedded and isolated are not relative selections, and the local density - that of the CG itself - is relatively constant across the CG sample. Therefore, in comparing to previous work, it is only possible to determine whether or not the previously obtained results are consistent with some level of SFR dependence on large scale environment as measured by this new method, to the extent that the density metrics may be compared.

The n^{th} nearest neighbour approach identifies high density regions of space by returning a shorter distance to the n^{th} neighbour in dense regions, and a longer distance in lower density regions (e.g., Hashimoto et al., 1998; Carter et al., 2001). Varying the value of n effectively changes the distance scales over which one wishes to probe the densities. With small values of n, the densities returned will be more reflective of local scale densities. Larger values of n will necessarily wander further afield to reach more distant neighbours. These distant neighbours will return a much larger volume of space over which to average the galaxy density, relative to their locally dense counterparts. The larger distances for isolated systems means that the densities derived from the identified n^{th} neighbour will be more reflective of a large scale density than those in locally dense environments (which return smaller distances), even if those local overdensities are relatively isolated from other structures (Weinmann et al., 2006). Since the distance to n^{th} nearest neighbour can vary smoothly from system to system, the distance scales being probed will also vary strongly as a function of local galaxy density. Furthermore, since the distance to n^{th} nearest neighbour is converted into a density, information about the distribution of those galaxies on scales less than that distance is lost. For the purposes of comparison to the metric used here, the n^{th} nearest neighbour algorithm simply needs to be able to distinguish between isolated and embedded CG systems. Both isolated and embedded CGs have similar densities in the cores, but exhibit different galaxy surface densities on scales of 0.1

– 1 Mpc (Mendel et al., 2011). As long as the distances returned by the n^{th} nearest neighbour probe these scales, they should be able to distinguish between the isolated and embedded systems.

This relatively straightforward criterion for comparison with the n^{th} nearest neighbour provides a sharp contrast with the other main metric of galaxy density. The annulus approach is more difficult to compare to the method used here. In this approach, the galaxy density within annuli of increasing radius is measured (e.g., Lewis et al., 2002; Kauffmann et al., 2004; Blanton et al., 2006). Many works place their dividing line between 'local' and 'large' scale at 1 Mpc. The densities within the annulus of radius 1–3 or 1–6 Mpc is therefore used as their measure of large scale environment. 1 Mpc is also a natural dividing line in the CG sample, but only in the sense that this is the distance to the nearest group that distinguishes isolated and embedded CGs. The annulus approach is sensitive to the presence of broad overdensities beyond the distance of the defined 'large scale', generally 1–6 Mpc away from the selected galaxy. The method used here provides the ability to define a system as embedded even when the overdensity is entirely on scales < 1 Mpc. In this case, the CG would still distinctly be embedded in a more massive halo than its own, but would not be distinguished as 'large scale' structure by an annulus of radius >1 Mpc. If some fraction of the rich group extends beyond 1 Mpc distant from the CG galaxy, then it would be detected by the >1 Mpc annulus. However, Mendel et al. (2011) shows that rich groups and CGs in any environment converge to similar galaxy densities at separations ~ 1 Mpc. With the median distance from CG centre to a rich group for embedded groups being of order ~ 0.3 Mpc, and the majority of rich groups's strong galaxy surface density visible on scales < 1 Mpc, it might be expected that only a small fraction of the galaxies within a rich group to extend beyond the 1 Mpc threshold. Furthermore, the annulus method loses all information on scales smaller than the area of the annulus. If an annulus contains an overdensity, but also contains a very low density region, these two extremes will average to a value that is not representative of either region.

With these differences in mind, it is possible to attempt to compare the results presented here of a SFR dependence on large scale environment with previous studies. It is important to note that this investigation focuses only on the the dependence on those galaxies which are star forming relative to a star forming control, not the overall trend of star formation rate with density. Those works which conclude that the large scale environment affects the evolution of star forming galaxies by way of

the cluster-centric radius (Park & Hwang, 2009) or dark matter halo mass (Goto et al., 2003; Weinmann et al., 2006) are compatible with these results. However, a large body of work in the literature (using a mixture of methodologies) has concluded that the local scale effects dominate galaxy evolution (e.g., Hashimoto et al., 1998; Carter et al., 2001; Lewis et al., 2002; Kauffmann et al., 2004; Blanton et al., 2006; Blanton & Berlind, 2007). These analyses consistently find a strong SFR dependence upon local galaxy density, with either very weak or absent dependence upon galaxy densities beyond the 1 Mpc scale. For example, Blanton et al. (2006) constrains the effects of large scale (1-6 Mpc) effects to be small, but does not exclude them as a minor contributor. The results found here are not in disagreement with those studies which found that the local scales dominate galaxy evolution, consigning the large scale effects to a minor role. *The small difference in $\Delta log(SFR)$ between star forming galaxies in isolated and embedded CGs surely confirms the minor nature of the large scale environment. The current work indicates simply that there is a non-zero effect upon the galaxies within a larger structure beyond local density effects; it does not require that the effect must be the dominant one.* Other works in the literature are less conservative, and interpret the lack of a strong dependence upon the galaxy densities at > 1 Mpc radii as evidence that the large scale environment does not have an effect upon a galaxy. Blanton & Berlind (2007) find that for the star forming fraction, large scale structure (defined as structure in an annulus > 1 Mpc) has no discernible effect upon the galaxy, with a bias in their measurement of no more than 5%, and density error of order 15%. Kauffmann et al. (2004) found that local densities dominate the SFR evolution; at low masses, the SFR of a galaxy was up to 10 times lower in high local densities than it was in low local densities. When testing the spatial scales between $1 - 3$ Mpc, they found a very noisy relation in the 4000 Å break strength, their metric for SFR, with density: consistent with no dependence. In contrast, this work shows that SFRs do depend, albeit at a low level, on their association with a large scale structure.

Reconciling the current results with those in the literature likely comes down to the nature of the 'environment' identified by the different density metrics that have been previously used. The division in environment used in this chapter identifies galaxies which are within 1 Mpc of a larger group than the CG itself. By the time scales of > 1 Mpc from the centre of a group are reached, the galaxy densities for isolated, embedded, and rich groups have all converged (Mendel et al., 2011). This indicates that most of the galaxy overdensity due to the rich group has also dwindled

by 1 Mpc from the centre. If the groups are then sufficiently far away from the CG, they might cross over into the > 1 Mpc annulus, and be detected as an overdensity by that metric. However, the peak of the distribution of the distance to the nearest rich group is $\sim$0.3 Mpc for embedded structures (Mendel et al., 2011). At 0.3 Mpc, the majority of the rich group will be found within 1 Mpc of the CG, and would not significantly increase the measured density for annuli >1 Mpc. Therefore, even though the embedded CGs are situated in a significantly different environment than the isolated CGs, the overall galaxy density no longer reflects this difference at scales > 1 Mpc. Dividing environmental factors into large and local scale effects simply by using a cut in radius may blend extremely local drivers of evolution and some larger scale environmental factors. This metric therefore cleanly separates those groups embedded in a more massive halo from those which are independent halos, no matter what the projected scale of the more massive group may be.

3.4.2 Physical Drivers

The interpretation of the difference between embedded and isolated galaxy SFRs depends on whether the galaxies are intrinsically and physically different when they are found within an overdensity, relative to in isolation, or whether the overdensity is imposing an additional processing mechanism on an initially similar set of galaxies. If the former, then galaxies found within large scale structure are reacting to the same set of evolutionary processes as galaxies in isolation, but their intrinsic properties make SFR enhancement (for example) more difficult to achieve.

One mechanism by which a physical difference in the galaxy populations between embedded and isolated CGs might arise is assembly bias (e.g., Croton et al., 2007; Cooper et al., 2010). Assembly bias states that galaxies in high mass potentials tend to be older, as they collapsed and began to cluster at earlier times. In this scenario, one might expect clusters to contain a higher fraction of old, so-called 'red and dead' galaxies, galaxies which have undergone many mergers, or galaxies which have been affected by the cluster potential through quenching mechanisms for a longer period of time. Galaxies in rich groups tend to be preferentially gas poor (e.g., Chung et al., 2009) as a result. This framework offers one explanation for the SFR suppression in overdense regions. If galaxies within rich groups are simply older than their counterparts in the isolated groups, then they would be further along in their evolution, and may be further along the process of gas exhaustion and consumption. If the CG

galaxies, due to their age, in addition to any environmental effects, are preferentially gas poor, then it will be significantly more difficult to trigger a SFR enhancement after an interaction (e.g., Alonso et al., 2006). Their isolated counterparts, having collapsed at later times, are younger, more gas-rich, and at a much earlier stage in their evolution; these galaxies will show SFR enhancements much more easily, as their gas reservoir is less depleted. This explanation does not require a difference in the mechanisms that modulate star formation rates in the two environments; the difference in intrinsic galaxy properties is enough to alter the galaxy's response to the same large scale environmental mechanisms. However, all the galaxies in the CG sample are emission line galaxies. While their star formation rates may be lower than the average galaxy, they still have significant amounts of gas. They cannot have been so affected by quenching and stripping mechanisms as to remove all their gas, if they can maintain strong emission lines. A large portion of the galaxies in the CG sample, regardless of their classification as embedded or isolated, retain relatively late-type morphologies (see Figures 3.2 & 3.3). The maintenance of the emission lines and their late-type morphologies limits the magnitude of the environmental and age-related evolution that the galaxies could have undergone.

Turning to consider processes that could alter the SFRs of a given galaxy in a different environment, quenching mechanisms that function on very small scales such as those associated with interactions can be eliminated; these should be consistently affecting both isolated and embedded systems. Whatever mechanism is preventing SFR enhancement in embedded CG galaxies must be associated with the larger scale cluster potential, and not simply with the high densities of the CG system. One explanation could be the difference in evolutionary processes in central and satellite galaxies. Galaxy populations divided in this way show that galaxies which are not the dominant galaxy in the dark matter halo undergo stronger environmental processes than the central galaxy does (Haines et al., 2006; Kimm et al., 2009; Peng et al., 2012; Tinker et al., 2011). Embedded CGs would, by definition, be a satellite halo within the larger cluster halo. The enhanced processing seen in most satellites would therefore be systematically applied to all galaxies in the embedded systems. Blanton & Berlind (2007) find that high density substructures (on scales of < 300 kpc) within a cluster are more correlated than expected with hosting red galaxies, i.e., these galaxies are further along in the quenching process. By contrast, the probability of the dominant galaxy in the potential being in the CG sample is significantly higher for isolated systems, where the CG environment is more likely to be an isolated halo.

If the difference between galaxies in isolated and embedded CG systems is not due to a systematic physical difference in galaxy population, then the galaxies in both systems ought to have begun their evolution in the CG as a roughly uniform population of galaxies. In order to explain the absence of the SFR enhancement seen in the isolated CG galaxies, an additional evolutionary force must be at work in embedded CG galaxies due to the cluster potential that is not in effect in isolated halos. This brings into suspicion the main drivers of galaxy evolution in clusters: ram-pressure stripping, galaxy-galaxy harassment, and strangulation. The isolation criterion for CG selection means that external galaxy harassment is unlikely to be a significant influence. Galaxy harassment within the CG would contribute similar levels of evolutionary processing to both galaxy populations of CGs, and is therefore not a useful explanation for the difference in SFR response. Further, CGs have relatively low velocity differences, so harassment within the group itself is unlikely to dominate over lower velocity tidal interactions. This leaves ram-pressure stripping (removal of cold gas from the disk) and strangulation (removal of the external gas supply). Weinmann et al. (2006) argue that ram-pressure stripping should affect low mass galaxies more quickly, as they have a smaller potential and are more easily stripped of their gas, within a halo of given mass. This ought to result in a higher fraction of low mass early type galaxies relative to the fraction of high mass early types. Their sample, taken from the SDSS DR2, contradicts this prediction, finding no evidence for preferential changes to the lower mass galaxies. Wetzel et al. (2012) find that galaxies introduced into a larger halo can retain their SFR at pre-accretion levels for 1-2 Gyr before rapid quenching takes place; if ram-pressure stripping were the dominant mechanism, the galaxies would have their cold gas removed on much quicker timescales. Ram-pressure stripping is also excluded by a number of other authors (e.g., Tanaka et al., 2004; Blanton & Berlind, 2007) due to the large radial extent of SFR suppressions, and SFR suppressions in low mass halos. Neither of these environments are expected to have the dense intra-cluster medium needed to support effective ram-pressure stripping (e.g., Rasmussen et al., 2008). This leaves strangulation as the prime suspect for SFR suppression in cluster environments, as it could still remove the hot halo of gas around a galaxy in lower density or lower mass environments (Balogh et al., 2000). Furthermore, it does not affect the morphologies or cold gas reservoirs of the galaxies involved (Weinmann et al., 2006). Mendel et al. (2011) find that the morphologies within embedded and isolated CGs are very similar, which also supports the idea of an environmental process being at work, rather than

different evolutions due to age or gas exhaustion. Strangulation therefore appears to be an appealing mechanism for long-term suppression of SFR in emission line galaxies.

3.5 Conclusions

A sample of robustly calculated gas-phase metallicities and star formation rates for a sample of 75,863 star-forming galaxies in the SDSS DR7 is presented. These metallicities use emission line fluxes which passed rigorous quality control, and are cleaned of any AGN contribution. From this sample, a compact group sample of 112 galaxies uniformly selected on criteria for local richness and cleaned of likely interlopers is defined. The compact group sample is further split by large scale environment into isolated and embedded subsamples, with 62 and 50 galaxies in each sample respectively. These two samples of star forming CG galaxies therefore have the same (very high) local density, but different large scale environments. The samples of compact group galaxies are simultaneously matched in mass and redshift with 50 non-compact group galaxies each in order to form robust control samples. By comparing the control samples with the compact group galaxy properties, the difference in SFR and metallicity at fixed mass can be measured. The main conclusions of this work are as follows:

1. Isolated and embedded compact group galaxies are offset from each other, and isolated CG galaxies are significantly enhanced relative to the field SFR-mass relation. Embedded galaxies do not show a similar enhancement. The SFRs of the galaxies in isolated CG systems show a median enhancement of $+0.07\pm0.03$ dex, whereas the galaxies in embedded systems show a median offset of -0.03 ± 0.05 dex. A series of bootstrap and Monte Carlo simulations indicate that this offset is robust. These results indicate that, even though previous works have found a primary dependence on local density, the SFRs of star forming galaxies in locally overdense regions are mildly sensitive to large scale environment at fixed local density.

2. No evidence for a significant metallicity offset is found in either the isolated or embedded samples relative to their control samples. However, this non-detection is likely to be a result of small number statistics, as the CG sample is not likely to detect offsets of order 0.03–0.05 dex, which is the typical metallicity offset expected due to interactions or density driven effects. The sample is therefore

inconclusive with regards to the metallicity offsets. The simulations indicate that at 99% confidence offset of 0.13 dex would have been detected.

Criteria classically used to select CGs define a consistent set of high density groups of galaxies. However, when these systems are divided according to the presence of a nearby cluster or rich group, the SFRs of the star forming galaxies within these systems become divided. The environmental split in median SFR demonstrates that large scale structure can exert an effect upon substructure found within it, even if at a low level, and that substructure is relatively isolated from other galaxies. The two possible explanations are: galaxies found within overdense regions are intrinsically different from galaxies in low density regions, which alters their response to the same local overdensities, or that the rich group structure imposes an additional evolutionary mechanism onto the galaxies in embedded CGs. This may give further support to the differences in satellite and central galaxy evolution, with strangulation, found only in cluster environments, as an effective process in slowly shutting down star formation.

Chapter 4

Tracing changes in star formation rate and metallicity out to separations of 80 kpc

In the previous chapter, the focus of the investigation was on the importance of large and small scale structure on the star formation rates and metallicities of galaxy groups. It is now possible to turn to the most dramatic environments; that of galaxy interactions. As in the previous chapter, star formation rates and metallicities are used as tracers of gas motions. These now-familiar tracers can be used to determine the level of perturbation and gas flow within interacting galaxies, allowing a study of the effects of a close tidal encounter on a galaxy. A sample of 1899 galaxies with a close companion selected to be interacting is used to determine the range of projected separations over which a galaxy is affected by a close encounter with another galaxy. The results in this chapter are published in Scudder et al. (2012b)[1].

4.1 Introduction

Galaxies that experience a close encounter with a companion are expected to undergo significant changes. Both observations and simulations have been used to probe the internal properties of galaxies as they progress through a merger, with simulations providing a framework through which the observational results may be interpreted.

[1]Scudder, J. M., Ellison, S. L., Torrey, P., Patton, D. R., & Mendel, J. T. 2012b, MNRAS, 426, 549

Currently, theoretical models present a consistent general picture of the evolution of a galaxy in a merger. Strong tidal interactions may trigger bar instabilities in the central regions of the galaxy in both the stellar and gaseous components. These bars are misaligned, and torques exerted on the gas by the stars result in the loss of angular momentum in the gas at larger radii. This gas then falls towards the nucleus of the galaxy, efficiently funnelled by the bar instabilities (Barnes & Hernquist, 1996; Mihos & Hernquist, 1996; Cox et al., 2006; Di Matteo et al., 2007; Montuori et al., 2010; Rupke et al., 2010a; Torrey et al., 2012). Indirect signatures of this picture are visible in the gas-phase metallicities and star formation rates of the interacting pairs. Gas inflowing from the outer regions is generally of a lower metallicity than the nuclear regions, so inflow from larger radii results both in a diluted nuclear metallicity (Kewley et al., 2006; Ellison et al., 2008b; Michel-Dansac et al., 2008) and in a flattening of the standard metallicity gradient present in spiral galaxies (Rupke et al., 2010b; Kewley et al., 2010; Perez et al., 2011). This new central concentration of gas provides the ideal catalyst for a significant starburst. Higher than average central star formation rates occur in nearly all simulations of galaxy mergers (e.g., Mihos & Hernquist, 1994, 1996; Di Matteo et al., 2007; Montuori et al., 2010), and are also a ubiquitous feature in observational studies (e.g., Larson & Tinsley, 1978; Donzelli & Pastoriza, 1997; Barton et al., 2000; Lambas et al., 2003; Alonso et al., 2004, 2006; Woods & Geller, 2007; Ellison et al., 2008b, 2010; Darg et al., 2010; Xu et al., 2010).

The strength of the induced star formation can vary dramatically from merger to merger. A significant contributing factor to this variation is predicted to be the mass ratio of the two galaxies (Cox et al., 2006). Galaxies in major mergers, i.e., with masses within a factor of 3, have previously been reported to show the strongest star formation rate (SFR) enhancements, on average, with both galaxies showing comparable enhancements (Woods & Geller, 2007; Ellison et al., 2008b; Xu et al., 2010; Lambas et al., 2012). Additionally, simulations suggest that a number of parameters within the merger exert a significant influence upon the extent to which gas is funnelled to the central regions of a galaxy. Most notably, orbital parameters can control the strength of the SFR response (Di Matteo et al., 2007; D'Onghia et al., 2010). The difference between a prograde–prograde and prograde–retrograde merger can alter the triggered star formation by a factor of 2, and the separations of the nuclei at first passage can inhibit SFR at coalescence if the tidal forces are so strong that gas is preferentially ejected into tidal features, rather than collecting in the central regions. Gas fractions also seem to play a role, with lower gas fraction galaxies

showing weaker SFR triggering (Di Matteo et al., 2007).

Observationally, studies of the SFR in galaxy pairs with projected separation (r_p) are largely in agreement with each other. Galaxies in pairs show enhanced SFRs out to separations of $\sim$ 30 kpc h^{-1}, with the strongest enhancement at the smallest separations, and a smooth decline to a fiducial value (Lambas et al., 2003; Alonso et al., 2006; Nikolic et al., 2004; Ellison et al., 2008b; Li et al., 2008). Some studies have found evidence for SFR enhancements at wider separations (e.g., Barton et al., 2000; Lin et al., 2007; Robaina et al., 2009; Wong et al., 2011), although generally this only weakly extends the trend to $\sim$50 kpc h^{-1}. Galaxies are generally expected to be most strongly enhanced just after first passage, and again at coalescence, but the timescales for catching a galaxy during a significant starburst are also tied to the strength of the burst (Torrey et al., 2012). The highest SFR enhancements are generally of the shortest duration, whereas lower-level enhancements can persist for longer periods of time (Di Matteo et al., 2007; Montuori et al., 2010). The trend of SFR with projected separation is therefore likely to be a combination of merger-induced changes and the timescale on which the observational snapshot is taken.

In practice, the ability to determine the merger phase of a galaxy pair is severely limited, as the only observable indicators are that of the projected separations (r_p) of the two galaxies, or the morphological disturbance of the galaxy. However, even this latter metric is fraught with problems. Depending on the orbit of the encounter, galaxies do not always show strong tidal features after an interaction (e.g., Toomre & Toomre, 1972; D'Onghia et al., 2010), and the ability to classify galaxies, either visually or using an automated method, is strongly dependent on the surface brightness of the tidal features relative to the limitations of the data. With large data sets, such as those obtained from large surveys, the imaging is often not very deep, and the samples are large enough that visual classifications quickly become impractical (although see Darg et al. 2010). As a result, many studies rely entirely upon the projected separations as a metric for determining the merger phase of a galaxy pair, by correlating large r_p with a large time elapsed since pericentric passage, and small separations with galaxies actively undergoing a close passage. The SFR enhancement trends with r_p are thus usually interpreted as the signature of a sharp increase in SFR at pericentric passage, with the strength of the enhancement dwindling as the separations between the galaxies increases (e.g., Barton et al., 2000; Barton Gillespie et al., 2003; Nikolic et al., 2004; Woods et al., 2010). This physical model would indicate that galaxies are able to rapidly funnel gas to the central regions of a galaxy after

an interaction, and that gaseous concentration is rapidly converted to stars, with the SFR declining as the galaxies separate and the burst ceases.

To date, there has been no simultaneous statistical study of both the gas phase metallicity and the star formation rates in a sample of galaxy pairs. Both the SFR and the gas phase metallicity are expected to change significantly throughout the merger as gas flows are induced through the galaxy; studying both of these quantities simultaneously allows a gain in insight into when changes are induced. In this chapter, this issue will be tackled by using a sample of pair galaxies from the Sloan Digital Sky Survey Data Release 7 (SDSS DR7; Abazajian et al., 2009). The sample presented here will be used to precisely quantify differences between the pair and the control samples over a wide range of projected separations and mass ratios. A suite of simulations based on the work in Torrey et al. (2012) is developed to allow an analysis of the trends with r_p for characteristic markers of the merger stages.

To this end, the sample of galaxy pairs in the SDSS DR7 compiled by Patton et al. (2011) and the metallicities calculated in Chapter 2 is used to construct a clean pairs sample with stringent quality control measures. In Section 4.2 the sample selection is reviewed and the control sample is defined, along with a description of the metallicity and SFR values. In Section 4.3, the methodology for quantifying the changes in the SFR and the metallicity relative to the control sample is described and the changes for a subsample of morphologically disturbed galaxies is further quantified. In Section 4.4 the diagnostic statistics of the sample are described in more detail, and in Section 4.5, the results are compared to simulations in order to develop an interpretative framework for trends with projected separation. Comparisons to previous work and an analysis of the physical picture the results suggest are presented in Section 4.6 by way of a conclusion.

4.2 Sample Selection

In this section, a description of the criteria applied to the spectroscopic pool of galaxies in the SDSS DR7 from which the final pairs and control samples can be compiled is presented. The goal of this selection is to identify a tightly controlled set of pair galaxies with a high probability of being a clean selection of physically associated and interacting systems. Consistently calculated quantities across the pairs and the controls are therefore required in order to be able to accurately compare between the samples.

In order to qualify for the pairs sample or the control sample, a galaxy must have a stellar mass[2] and sufficiently strong emission lines so that a gas-phase metallicity and star formation rate can be calculated. In order to prevent the issues with Sloan photometry detailed in Simard et al. (2011) from affecting the mass estimates, the masses used here have been recalculated from the MPA catalogue[3], using their inferred mass to light ratios (M/L). The relationship between the MPA-derived M/L and $(g-r)$ colour is determined for the total MPA catalogue. This relationship is then used to derive masses from the $g-r$ colours derived from the updated photometry presented in Simard et al. (2011). With a new M/L value for each galaxy, the luminosities from Simard et al. (2011) are used to calculate a mass without needing to rely on the Sloan magnitudes. The SFR values presented in the MPA catalogue are used, calculated according to the Brinchmann et al. (2004) work. For galaxies with sufficiently high emission line signal to noise (S/N), Brinchmann et al. (2004) uses a set of 6 emission line fits to the SDSS spectra to calibrate the SFR; lower S/N galaxies are calibrated with the strength of the Hα line. Given the high S/N required for emission lines in this sample, all SFR values will be from the multi-emission line fits. The median 1σ error on any individual SFR value in the sample is ~ 0.09 dex. The SFR values taken from Sloan's 3 arcsecond fibre can be corrected to total values using models fit to the photometry of the galaxy outside the fibre, taking into account the fraction of light not contained within the fibre. A full description of the aperture correction methodology and bias testing is present in the Brinchmann et al. (2004) work. However, for this chapter, only the fibre values are used. Each galaxy is further required to have a consistently calculated metallicity, following the criteria laid out in Chapter 2.

The galaxy pairs sample used here is based upon the spectroscopic catalogue of Patton et al. (2011). For a full description of the sample selection algorithm, the reader is referred to that work. Briefly, Patton et al. (2011) select galaxies with a spectroscopic close companion within $r_p < 80$ kpc h^{-1}, with a velocity difference $\Delta v < 10{,}000$ km s^{-1}, and within a mass ratio of 10:1. This results in a pairs sample of 23,397 galaxies in pairs or higher order multiples[4]. The master pairs catalogue is the result of a re-running of the algorithm of Patton et al. (2011) with a minor alteration to the culling process meant to account for fibre collisions in Sloan. Due to the fixed fibre collision constraint of $55''$ (Strauss et al., 2002), galaxies with a projected

[2]Unless otherwise stated, all mass values in this chapter are stellar masses.

[3]Available here: http://www.mpa-garching.mpg.de/SDSS/DR7/

[4]$\sim$8% of these galaxies have more than one close companion; 1.6% have more than 2 companions. Excluding these from the sample does not significantly impact the results.

companion nearer than 55″ will be preferentially missed in the spectroscopic sample. Some galaxies with separations less than the collision limit do have spectroscopic companions, primarily as a result of overlap between plates. In order to compensate for this incompleteness effect, 67.5% of galaxies in pairs with separations $> 55''$ are randomly culled from the sample, according to the incompleteness calculations of Patton & Atfield (2008). In Patton et al. (2011), this cull was done without taking into account the culled galaxy's companion; i.e., a galaxy could remain in the sample while its companion was culled. For the sample used in this chapter, the culling has been repeated, but once one galaxy in a pair has been excluded from the sample, its companion is also removed.

Galaxies with Δv values at the high end of the distribution are unlikely to be physically associated. To select a sample of systems which are more likely to be physically associated, a $\Delta v < 300$ km s^{-1} limit is imposed to minimize the influence of projected pairs (Patton et al., 2000). This cut is in line with Δv cuts in other works, which usually range between 250 – 500 km s^{-1} (e.g., Lambas et al., 2003; Ellison et al., 2008b; Patton et al., 2011). 1,899 galaxies in the master pairs catalogue pass all criteria, and comprise the final galaxy pairs sample[5].

4.2.1 Matching to Controls

In order to determine whether an interaction leads to changes in galactic properties, a control sample must also be defined. To this end, a sample of non-pair galaxies is defined as any galaxy which meets the mass, SFR and metallicity requirements of the pairs sample, but is not part of the full pairs catalogue (the 'control pool').

The largest potential sources of bias between a sample of galaxies and its control are differences in the stellar masses, redshifts, and environments of the galaxies in the two samples (Perez et al., 2009). In order to eliminate these sources of bias from the sample, the control sample should be matched in all three parameters, such that the normalized distribution of the final control sample matches that of the pair galaxies. Simply matching in stellar mass and redshift could bias the results, as it has been shown that pairs tend to exist in higher density environments than non-pairs (e.g., Barton et al., 2007; Patton et al., 2011), and there is a well known trend of SFR with density (e.g., Kauffmann et al., 2004; Poggianti et al., 2008; Cooper et al., 2008). The density metric used is calculated as an overdensity by taking the 5th nearest

[5]The selection criteria does not require that both a galaxy and its companion pass all criteria.

neighbour density and normalizing by the median density in a redshift slice of ± 0.01 around the galaxy.

Galaxies in the pairs sample are therefore matched to galaxies in the control pool simultaneously in mass, redshift, and local density, in a similar way to Ellison et al. (2010). Briefly, this algorithm finds the galaxy from the control pool which is the simultaneous best match in all three parameters for each pair galaxy. After every pair galaxy has been matched to a control, a Kolmogorov-Smirnov (KS) test is used on the total distributions of pairs and control. If the KS test finds that the distributions are consistent with being drawn from the same parent distribution at $> 30\%$, then the tentative controls become part of the control sample. Matching continues without replacement until 10 control galaxies have been matched to each pair galaxy in the sample, or the KS test returns a probability $< 30\%$. For the sample of galaxy pairs, the algorithm ran without failing the KS criterion; the final control sample therefore contains 18,990 galaxies.

The normalized distributions of stellar mass, redshift, and density for the pairs and control samples are displayed in Figure 4.1. KS tests on the final distributions result in probabilities of 99.77% for the stellar masses of the two samples, 83.84% for the redshifts, and 99.85% for the density distribution, indicating that the galaxy pair and control samples are very unlikely to have been drawn from different parent populations. Pairs and their controls are typically matched to within 0.001 dex in total stellar mass, 6×10^{-6} in redshift, and 0.0008 dex in density, with a median range within the set of control galaxies of 0.07 dex in mass, 0.005 in redshift, and 0.09 dex in density.

4.3 SFR and metallicity offsets

With a well-matched control sample of non-pair galaxies, it is now possible to proceed to compare the SFRs and metallicities of the pairs and controls. As the fibre values for the SFRs are being used, both the SFRs and metallicities are values for the central $3''$. The adopted methodology follows the methodology presented in Chapter 3, which itself is a modification of the method presented in Patton et al. (2011). Briefly, each galaxy in a pair is matched to 10 controls, which all have metallicity and SFR values. The median SFR or metallicity of the 10 controls is taken, and subtracted from the value for the matched pair galaxy, as expressed in Equations 4.1 & 4.2. The resultant

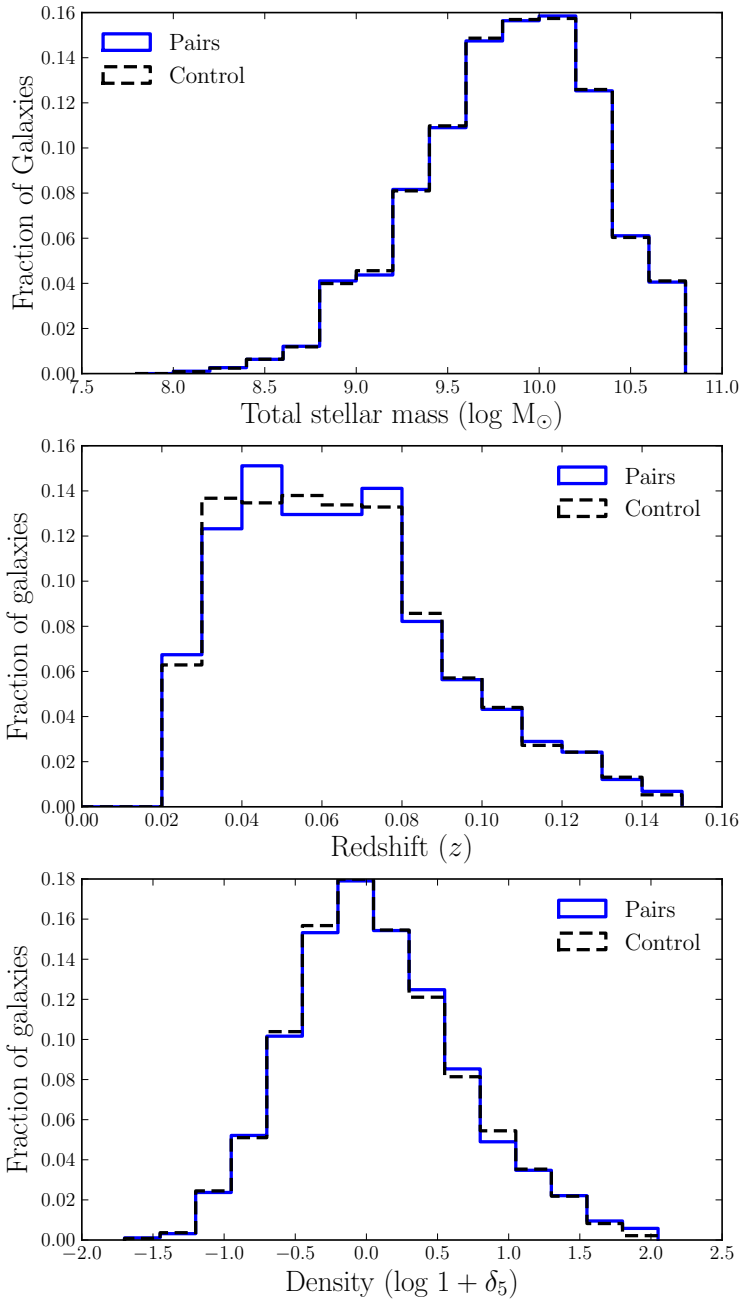


Figure 4.1: Normalized distribution of the pair galaxies (solid blue) and the control sample (dashed black), matched in mass (top), redshift (middle), and density (bottom). Density is measured via the distance to the 5th nearest neighbour. Each pair has 10 controls. KS tests result in a probability of $\sim$ 99.8% that both the stellar mass and density distributions were drawn from the same parent distribution, and a probability of $\sim$ 83.8% that the redshift distributions were drawn from the same parent population.

value, $\Delta\log(\mathrm{SFR})$ or $\Delta\log(\mathrm{O/H})$, is defined to be the 'offset' value from the control.

$$\Delta\log(\mathrm{O/H}) = (\log(\mathrm{O/H}) + 12)_{pair} - \mu_{1/2}\,(\log(\mathrm{O/H}) + 12)_{controls}\,, \tag{4.1}$$

$$\Delta\log(\mathrm{SFR}) = \log(\mathrm{SFR})_{pair} - \mu_{1/2}\log(\mathrm{SFR})_{controls}, \tag{4.2}$$

where $\mu_{1/2}$ signifies the median.

Positive offset values indicate enhancements over the control sample, i.e., metallicity enrichment or SFR enhancement, whereas negative offset values indicate suppression relative to the control sample, i.e., metallicity dilution or suppressed SFR. This method has the advantage of allowing the quantification of the changes in SFR and metallicity on a galaxy by galaxy basis, relative to the 10 control galaxies to which it is matched. Both SFR and metallicity are known to have strong mass dependences (Tremonti et al., 2004; Ellison et al., 2008a; Noeske et al., 2007b; Elbaz et al., 2011), but as pair galaxies and their controls are tightly matched in mass, a comparison to the matched control galaxies is effectively a comparison at fixed mass.

The calculated SFR and metallicity offsets are plotted against projected separation (r_p) in Figure 4.2. The distribution of $\Delta\log(\mathrm{SFR})$ is already visibly offset from zero. This trend is made clearer in Figure 4.3, which shows the same sample of galaxies, but binned in r_p. SFRs are significantly offset from the control sample out to 80 kpc h^{-1}, the widest separation probed in this sample, with an increase in the offset magnitude at the smallest separations ($r_p \leq$ 15 kpc h^{-1}) to $\sim$ 0.25 dex (about a factor of two). At separations $\gtrsim$ 30 kpc h^{-1}, the offsets maintain a roughly constant magnitude of $\sim$0.11 dex (a 30% enhancement).

$\Delta\log(\mathrm{O/H})$ displays a much weaker trend with r_p than the SFR, but similarly shows a sharp increase in offset magnitude at the smallest separations. Metallicity values are significantly suppressed by -0.02 dex out to $\sim$ 60 kpc h^{-1}. The plateau seen in the SFR offsets is not apparent in the metallicity offsets for the total sample of pairs. The relative weakness of this trend is not surprising, as metallicity shifts due to interactions are usually found to be a small magnitude effect, generally of order -0.03 to -0.05 dex (e.g., Michel-Dansac et al., 2008; Cooper et al., 2008; Ellison et al., 2008b, 2009).

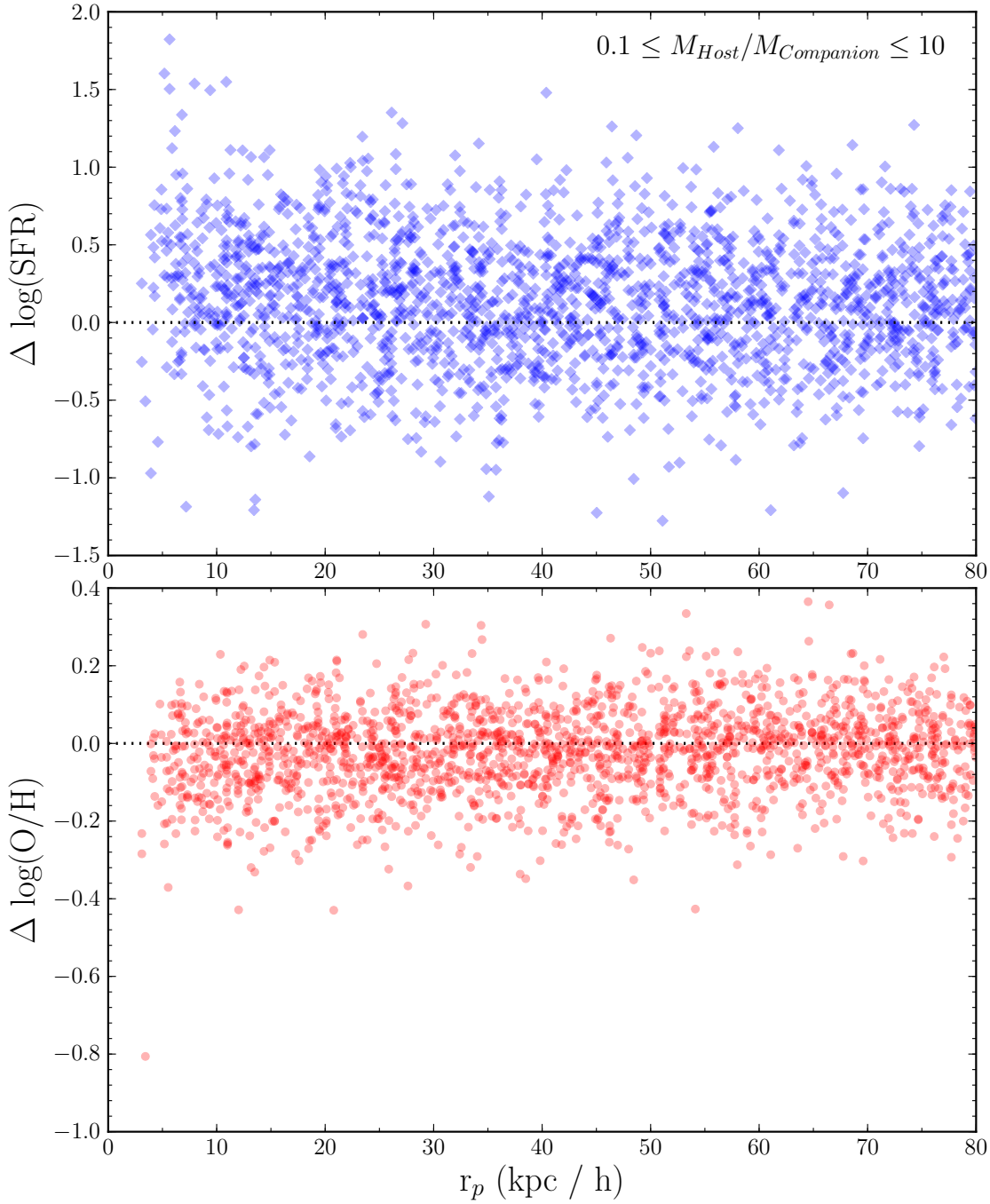


Figure 4.2: Offset values for all 1899 galaxies in pairs sample as a function of projected separation. The top panel shows the SFR offsets, and the bottom panel shows metallicity offsets. The horizontal black dotted lines indicate the zero line.

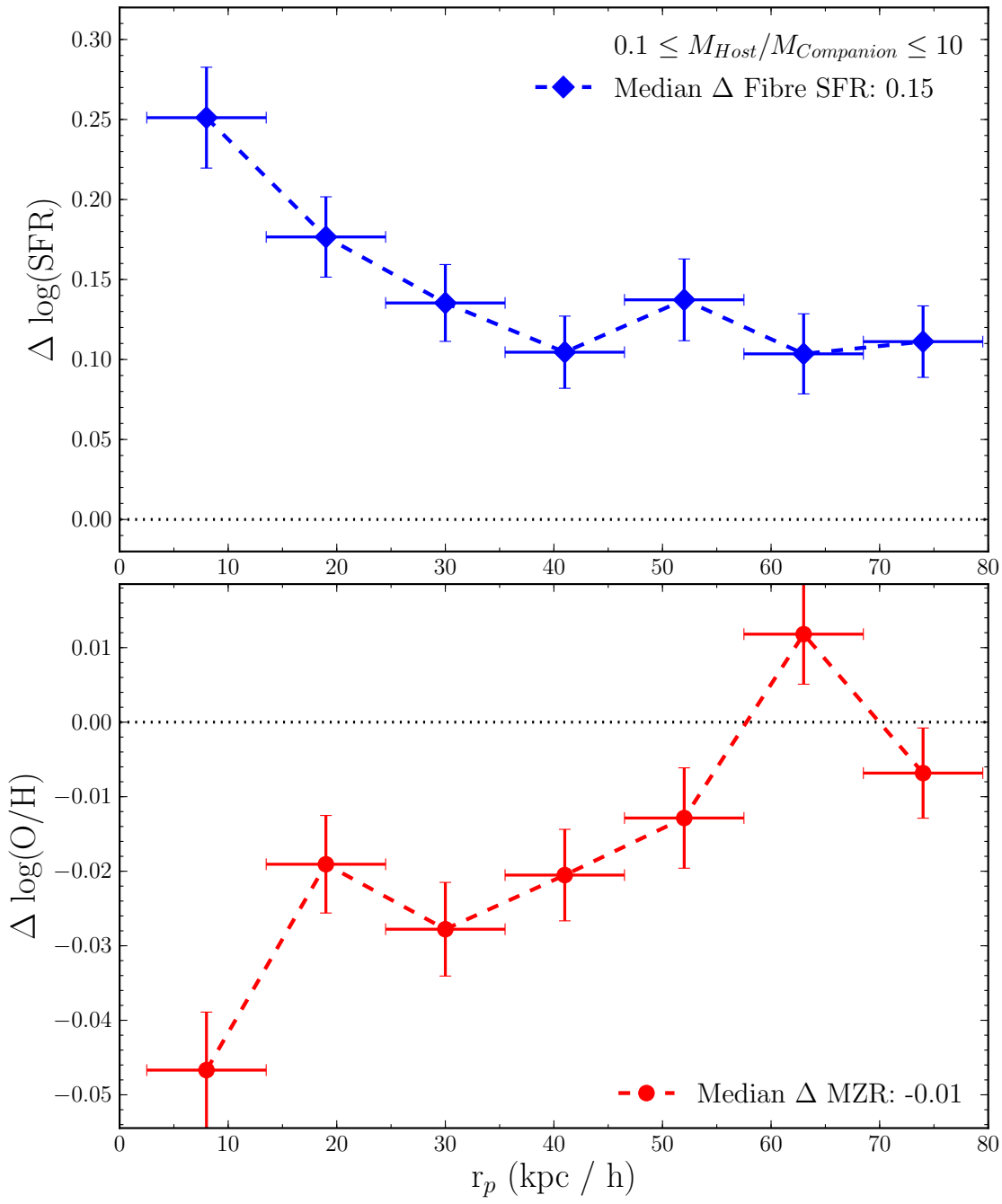


Figure 4.3: All 1899 galaxies in pairs sample. The top panel shows the SFR offsets, bottom panel shows metallicity offsets. Points are median values for the bin, and error bars are standard error on the median. Horizontal black dotted lines indicate the zero line.

4.3.1 Major & Minor Mergers

It is interesting to consider the difference in SFR trends for major versus minor mergers. The sample is large enough that the sample can be divided according to the mass ratios of the interacting galaxies. Note that the galaxies in this sample are divided by stellar mass instead of by magnitude (cf. Lambas et al., 2003; Woods & Geller, 2007; Lambas et al., 2012); see Ellison et al. (2008b) for a more extensive discussion of the use of luminosity, rather than mass, ratios. Galaxies within a mass ratio $0.33 \leq M_{host}/M_{companion} \leq 3.0$ are considered major mergers, $M_{host}/M_{companion} > 3.0$ are more massive companions in a minor merger, and $M_{host}/M_{companion} < 0.33$ are less massive companions in a minor merger. The pairs sample contains 1116 galaxies in major mergers, and 783 in a minor merger, of which 184 are the more massive companion, and 599 are the less massive companions. The smaller number of more massive companions in minor mergers is likely due to the increased probability of high mass galaxies hosting an AGN (e.g., Kauffmann et al., 2003) and thereby being removed from the star-forming sample.

The trends in Δlog(SFR) in major, more massive companion, and less massive companion subsamples with r_p are plotted in the top, middle, and bottom panels of Figure 4.4 respectively. Grey background points are the SFR offsets from Figure 4.3 in all panels. Notably, galaxy pairs of all mass ratios are enhanced relative to the control at all separations. The shape of the offset vs. r_p relation seen in Figure 4.3 is apparently driven by the major mergers. This is unsurprising, as major mergers are generally observed to show the strongest effects (e.g., Woods et al., 2006; Woods & Geller, 2007; Ellison et al., 2008b), and they make up the majority (59%) of the pairs sample.

With the exception of the more massive companion bin at 50 kpc h^{-1}, both the more massive and less massive companions in a minor merger show relatively flat enhancements at the same magnitude ($\sim$ 25% enhancement) at wider separations. However, the innermost bin shows a different response between more massive and less massive companions in a minor merger. The less massive companion shows no strong enhancement at small separations, distinguishing it from both the more massive companions and the major mergers, both of which show an increase in the magnitude of the SFR enhancement. More massive companions, on the other hand, show similar enhancement as the major mergers at small separations, at 2 times stronger than the control; major pairs are enhanced by a factor of 1.9. However, as

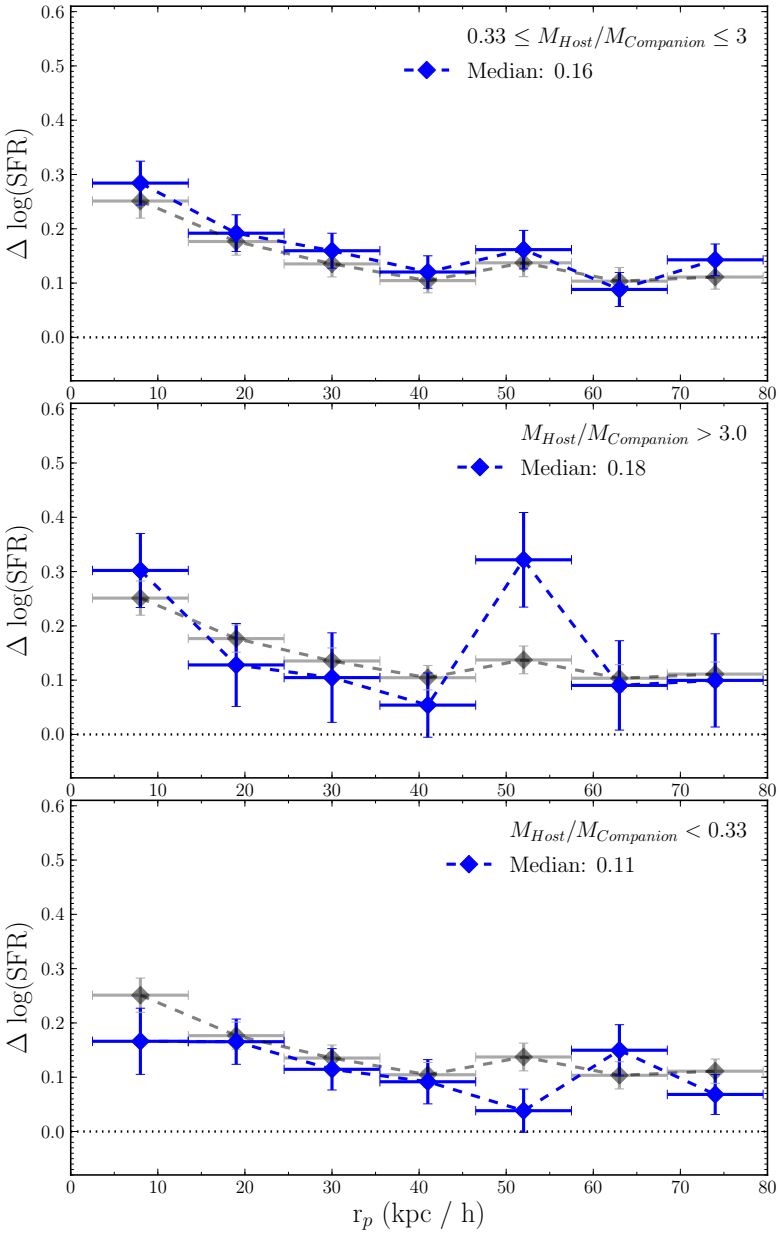


Figure 4.4: SFR offsets, separated by mass ratio. Grey background points in all panels show the trends from the total sample (Figure 4.3). The top panel shows the trend for all galaxies in major mergers (1116 galaxies). Middle panel shows the more massive companion in a minor merger (184 galaxies) and the bottom panel shows the trend for the less massive companion in a minor merger (599 galaxies). The more massive galaxies in a minor merger show an extremely strong enhancement at small separations, whereas the less massive companions show a much more consistent effect without any small r_p increase in offset. Both the more massive and less massive companions show similar levels of offset at wide separations (~ 0.1 dex).

the more massive companions have the smallest sample, each binned point in Figure 4.4 only has 20-30 galaxies. While the use of a median means that the sample is not biased by one or two outlying points, the poor number statistics may result in an anomalously high point due to poor sampling of the total distribution. A more detailed discussion of the offset distributions is presented in §4.4.1.

As the significance of the metallicity offsets is much weaker than that of the SFR offsets, splitting the metallicity offsets into mass ratio bins does not provide any additional information.

4.3.2 Visual Classifications

If the SFR enhancements and metallicity dilutions are truly being driven by the tidal interactions of galaxies in pairs, then selecting a subsample which shows morphological evidence of a recent tidal interaction ought to amplify the effects seen in Figure 4.3 (e.g., Michel-Dansac et al., 2008; Lambas et al., 2012, which found stronger effects in a morphologically disturbed subsample). To this end, only those galaxies which show strong tidal arms or other asymmetries induced by an interaction are selected. Although the pairs sample is designed to minimize the inclusion of projected pairs, physically bound galaxy pairs which have not undergone their first pass should have SFRs and metallicities close to the control, and will weaken the interaction-triggered signal. Any remaining projected pairs with low Δv values would not be excluded from the sample, and, as physically dissociated systems, would also weaken the signal from the interacting systems.

In order to eliminate the weakening effect due to either pre-interaction or projected pairs, all 1899 galaxies in the sample were visually classified for signs of morphological disturbances, as neither sample would be expected to show morphological signs of recent interactions. Galaxies were flagged as either 'visibly disturbed' or 'not visibly disturbed'. Galaxies which fall into the category of not being visibly disturbed will include pre-interaction galaxies and interacting galaxies whose tidal features are below the surface brightness limit of the SDSS imaging, along with galaxies which are not truly interacting, so this category is not a useful diagnostic on its own. However, the set of galaxies which falls into the disturbed galaxy classification should be a clean sample of galaxies which have already had a close encounter with a companion. 1105 galaxies are classified as disturbed, and the remaining 794 as 'not visibly disturbed' (See Figure 4.5 for 5 randomly selected examples of each classification).

Figure 4.5: SDSS thumbnails of 10 randomly selected galaxies flagged as 'visibly disturbed' (top row) or 'not visibly disturbed' (bottom row). All galaxies are labeled with their SDSS objids.

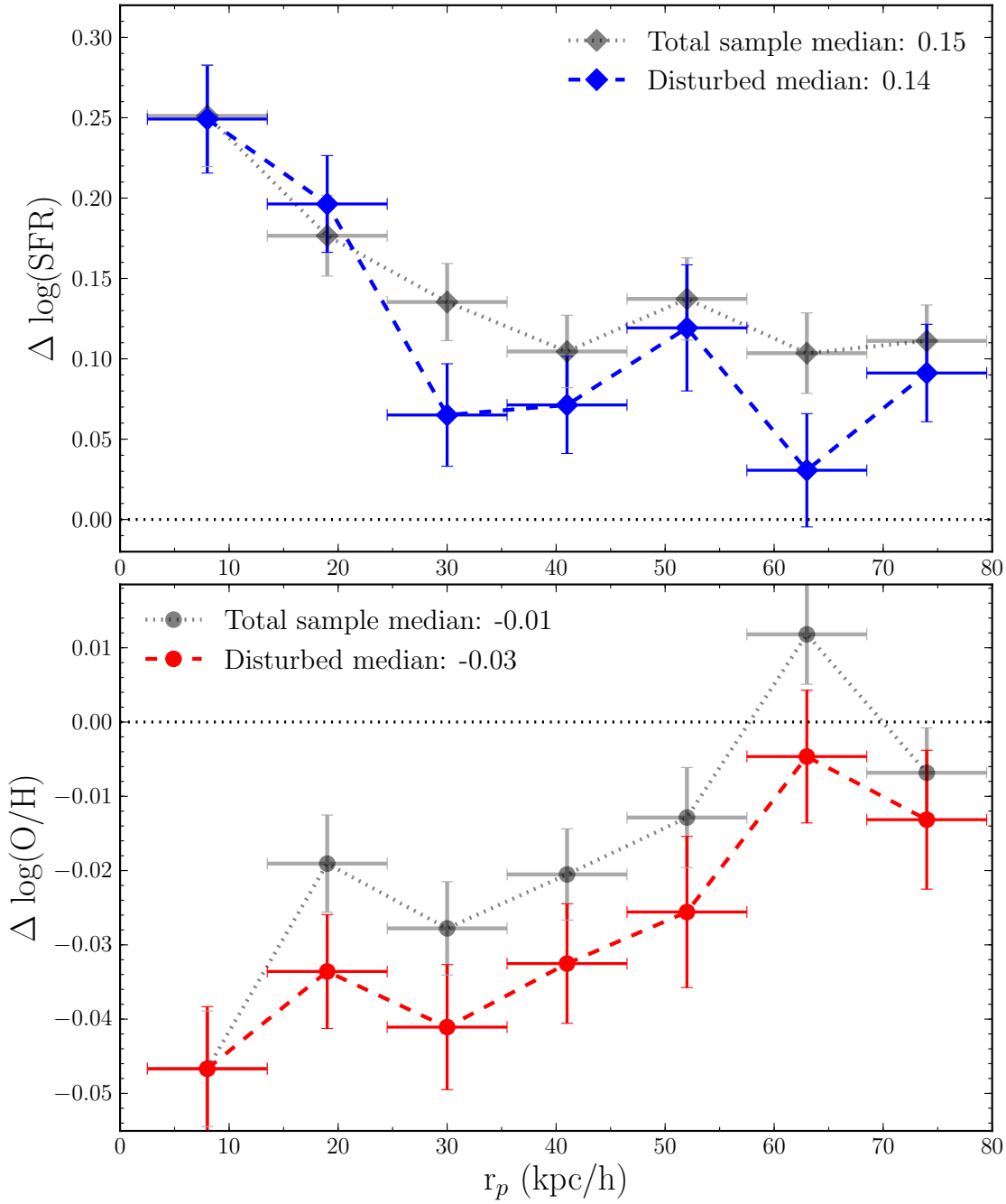


Figure 4.6: SFR and metallicity offsets for 1105 pair galaxies flagged as disturbed. Grey background points are the trends from the total sample (i.e., Figure 4.3). SFR offsets are of approximately equal magnitude in the disturbed and total samples, but the metallicity offsets are significantly larger in magnitude for the disturbed sample than for the total sample. The median value of $\Delta\log(\mathrm{O/H})$ triples in magnitude when the disturbed sample only is taken.

The SFR and metallicity trends with r_p for the subsample of visibly disturbed galaxies are shown in Figure 4.6. Grey points are the total sample points from Figure 4.3; the disturbed subsample is overplotted in coloured points. The SFRs do not show a systematic boost to their offsets in the morphologically disturbed sample; most points in the disturbed subsample are consistent with, or slightly below, the total SFR offsets. A KS test reveals an 89.78% chance that the disturbed and total samples were drawn from the same parent population. However, the median metallicity has shifted to more metal-poor values in all r_p bins, in some cases by nearly -0.02 dex. The median $\Delta\log(\mathrm{O/H})$ has tripled in magnitude to -0.03 dex for the disturbed sample. The disturbed and total metallicity samples have a KS probability of being drawn from the same distribution of only 1.29%. Within this subsample, galaxies are significantly metal poor out to 80 kpc h^{-1}, the same range over which the SFRs show enhancements.

The fact that the SFRs show no significant enhancement in the disturbed subsample is a puzzling contrast to the significantly lower metallicities. To ensure that the shifts between the total sample and the disturbed subsample are not simply due to the smaller size of the disturbed subsample, a random sample of 1105 galaxies is selected from the full pairs sample, and find the median value of the random subsample in both SFR and metallicity. This bootstrap resampling was then repeated 20,000 times. The results of this test indicate that it is very easy to obtain a SFR offset of +0.14 dex through random sampling of the $\Delta\log(\mathrm{SFR})$ distribution, but extremely difficult to reach median $\Delta\log(\mathrm{O/H})$ values of –0.03 dex at random. The bootstrap test indicates that the lower metallicity in the disturbed subsample is unlikely to be a statistical fluke due to a smaller sample size. However, the median SFR offset in the disturbed subsample is consistent with the total sample, and no significant difference in the SFR offsets of the disturbed sample is seen.

One potential interpretation is that the timescales over which galaxies are visibly disturbed are more strongly correlated with the timescales in which metallicity dilution is the strongest, and only weakly correlated with strong star formation enhancements. There is some suggestion from simulations that tidal features are visible for both a shorter period of time than the SFR burst, and at earlier times than the SFR enhancement (Lotz et al., 2008). If this is the case, then a sample of galaxies selected to be morphologically disturbed would not necessarily be expected to identify the galaxies with the largest $\Delta\log(\mathrm{SFR})$.

4.4 Distributions of SFR and metallicity offsets

While the median offset values as a function of r_p prove useful as a metric of the typical merger, it is interesting to explore the range of metallicity and SFR offsets. The distributions will provide, for example, insight into how often the most extreme offsets occur, as well as how frequently no statistical change is seen. Furthermore, since simulations suggest that the two galaxies involved in a major merger ought to show similarly enhanced SFRs (Montuori et al., 2010; Torrey et al., 2012), it is possible to search for observational evidence of simultaneous triggering on a merger-by-merger basis.

4.4.1 Offset distributions: Pairs vs. Controls

The pairs sample is first split into two bins of r_p. All galaxies with $r_p < 30$ kpc h^{-1} are defined as 'close pairs', and any galaxy pairs with $r_p > 30$ kpc h^{-1} as 'wide pairs'. The distributions of Δlog(SFR) and Δlog(O/H) are shown in Figures 4.7 & 4.8 respectively, with blue solid lines for the close pairs sample and red dashed lines for the wide pairs. The control sample is also divided by taking all control galaxies matched to close pairs as the $r_p < 30$ kpc h^{-1} control sample, and all galaxies matched to wide pairs as the $r_p > 30$ kpc h^{-1} control sample. Control offsets were calculated in a similar way as the pair offsets. Each galaxy in the control sample was compared to the set of other galaxies matched to the same pair galaxy. As control galaxies were matched in sets of ten, every control galaxy has 9 galaxies of similar mass, redshift, and local density, to which it can be compared. Therefore, the median of the 9 other control galaxies is taken, and the difference between that median value and the selected control galaxy is found. This difference (analogous to the calculations in Equation 4.1) produces a control offset for every control galaxy in both SFR and metallicity. As the control galaxies are tightly matched, the control offsets ought to be centred around zero offset in both SFR and metallicity, with no significant shift away from zero or between the close and wide control samples. Figures 4.7 & 4.8 show the control samples as the solid grey (close controls) and dashed black (wide controls) lines. The two control samples trace each other extremely well, and are centred around zero.

The top and middle panels of Figures 4.7 & 4.8 offer two ways of viewing the distribution of offsets for SFR and metallicity. The top panel shows the offsets in a discrete way, with the overall form of the offset distributions for the two subsamples

of both pairs and controls, normalized to the sample size. The middle panel shows the cumulative distribution of the same samples, which illustrates the differences between the distributions more clearly. Figures 4.7 & 4.8 show the total pairs sample rather than the disturbed subsample in order to maintain the number statistics of the pairs sample. While the full sample will contain a small fraction of galaxies which are not physically associated, and a larger fraction of galaxies which have not yet interacted, this should only increase the number of galaxies with offsets near zero, and will not affect the distribution of offsets further from zero.

In the top panel of Figure 4.7, it is already clear that both the close and wide pair samples are shifted systematically to higher offsets than their respective control samples, a trend which is even more dramatic in the middle panel. The cumulative distribution is calculated as the fraction of the total sample which has an offset greater than a given threshold. As the threshold values increase, the fraction of the control samples which have strong offsets decreases; the pairs decline more slowly. KS tests confirm this visual offset; both close and wide pairs are inconsistent with being drawn from the same parent population as their controls at $> 6\sigma$. To quantify this in a different way, the median Δlog(SFR) of the close pairs sample is +0.21 dex (or a 60% enhancement over the control; see also Figure 4.3), whereas the wide pairs sample has a median Δlog(SFR) of +0.11 dex (a 30% enhancement).

The bottom panels of Figure 4.7 shows the ratio of the pair fraction greater than a given offset (seen in the centre panel of Figure 4.7) relative to its respective control fraction, as a function of offset value. The black dotted line indicates a ratio of 1:1, i.e., that the pair and control samples have exactly the same fraction of the total sample at that offset or greater. This panel illustrates the relative frequency with which a given offset will appear in the pairs sample instead of the control. Values higher than one indicate that the offset is preferentially found in the pairs sample. The shaded regions indicate the range of 1σ errors, calculated from $\sqrt{N}$ statistics. This panel indicates that as Δlog(SFR) becomes more extreme, it is increasingly likely to observe these offsets in the close pairs sample, rather than the controls or the wide pairs sample. The wide pairs sample also shows an excess of positive Δlog(SFR) relative to the control sample, although only up to offsets less than +0.85 dex (7 times the control value) Therefore, while the median SFR offset is only an increase of 40%, there is an excess of close pairs with SFR enhancements up to a factor of 10 stronger than the control. Although only $\sim 3\%$ of pairs have excesses at this level, this is 3 times more than exist in the control. Wide pairs, by contrast, only show excesses up to a factor

of 7 (0.85 dex).

Figure 4.8 is set up in the same way as Figure 4.7. The systematic shift of the metallicities in the pairs sample is less pronounced than that of the SFR offsets, but is still visible. KS tests also verify this visual trend, with the metallicities for both close and wide pairs being inconsistent with being drawn from their parent distribution at $> 4\sigma$. Here the cumulative distribution is calculated as the fraction of galaxies with offsets *less* than the tested offset. The close pairs show a distinct shift to lower values relative to the control, whereas the wide pairs do not show a particularly strong shift. The difference in median is also much weaker; close pairs have a median offset of -0.034 dex (8%), and the median for the wide pairs is -0.01 dex (2%).

This distinction between close and wide pairs in the metallicities is confirmed in the bottom panel, which shows the relative frequency of a given offset between the pair and control samples. Here it is visible that the the wide pairs do not show a significant excess of metallicity offsets relative to the control at any magnitude (see also Figure 4.3). However, the close pairs show a significant excess of metal-poor galaxies at all negative offsets up to -0.25 dex (up to 1.78 times lower than the control). Offsets more extreme than -0.25 dex are so rare that the significance is overwhelmed by poor number statistics. The excess of very low metallicity galaxies is interesting, considering that the median offset for the pairs is ~ -0.02 dex, or 5% lower than the control. The discrepancy between the magnitudes of the excess offsets in the pairs sample and the overall median seems to indicate that the median offset is at least partially driven by a small number of galaxies which are very strongly offset from the controls.

If, instead of splitting the total pairs sample into close pairs and wide pairs, the sample is divided into bins of mass ratio, it is possible to investigate how the offset distributions depend on the mass ratio of the merger. The pairs sample is divided into major mergers ($0.33 \leq \mathrm{M}_{Host}/\mathrm{M}_{Companion} \leq 3$), less massive galaxies in a minor merger ($\mathrm{M}_{Host}/\mathrm{M}_{Companion} < 0.33$), and the more massive galaxies in a minor merger ($\mathrm{M}_{Host}/\mathrm{M}_{Companion} > 3$). The control samples are split so that the control galaxies matched to pair galaxies in each bin in mass ratio are assigned to their respective bins. Figure 4.9 is constructed identically to Figures 4.7 & 4.8. The top panel of Figure 4.9 shows that all three mass ratio bins (in colour) are systematically shifted to higher SFR values than their control samples (in grey). The control samples overlay each other reasonably well and are centred around zero, as expected. Given the similarity of the median values in Figure 4.4, it is unsurprising to see that the 3 distributions are

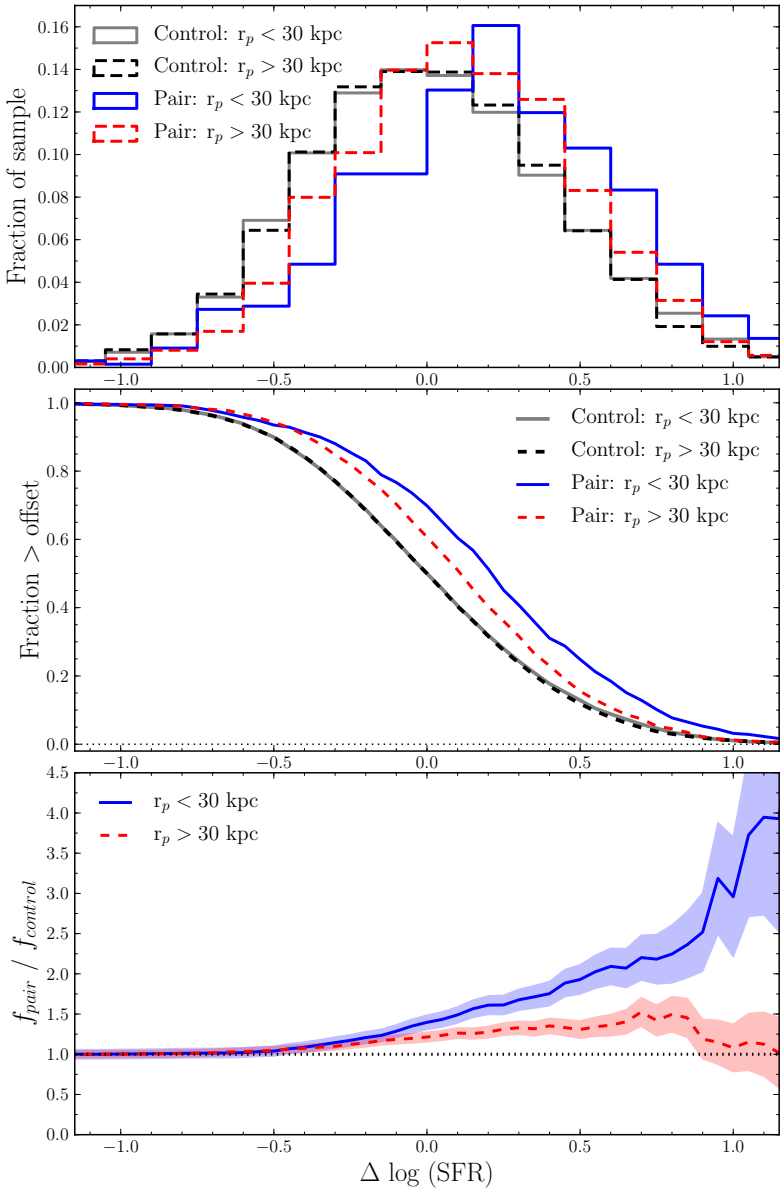


Figure 4.7: Top panel: histogram of SFR offsets for close pairs (blue solid), wide pairs (red dashed), and their controls. Control offsets are calculated between all controls matched to the same galaxy. Both close and wide pair galaxies are visibly shifted to higher offset values than the controls. Middle panel: cumulative distribution of SFR offsets, indicating the fraction of the galaxies in the sample with offsets greater than a given value. Galaxies are divided in the same way as the top panel. At all offset values, the close pairs sample shows a higher fraction of galaxies with strong SFR enhancement. The bottom panel shows the ratio of the fraction of pair galaxies above the threshold offset to the fraction of the control above that offset, with the shaded region indicating $\sqrt{N}$ errors. Black dotted line shows a ratio of 1.

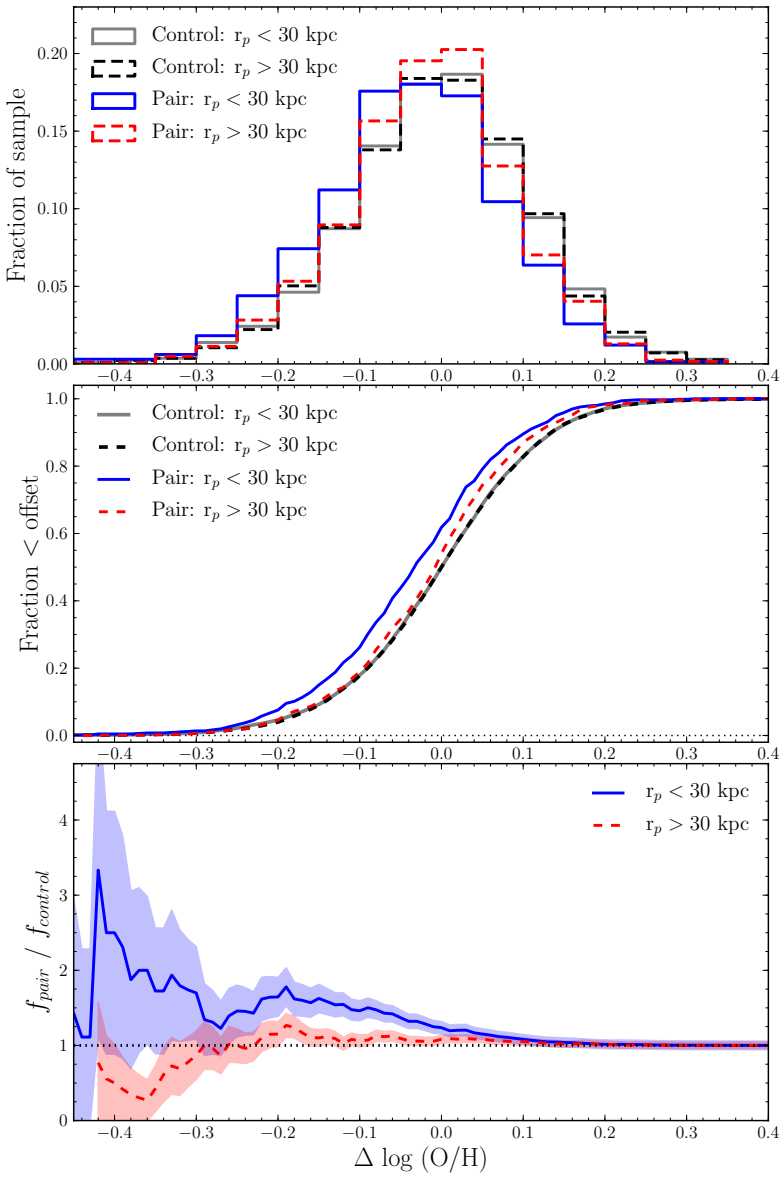


Figure 4.8: Same as Figure 4.7, but for metallicities. The top panel shows the distribution of the pairs and control samples, split into $r_p < 30$ kpc and $r_p > 30$ kpc bins, coloured same as Figure 4.7. Middle panel shows the cumulative distribution of metallicity offsets, and bottom panel shows the ratio of the pair and control fractions for > 30 and < 30 kpc bins. The shaded region indicates $\sqrt{N}$ errors.

shifted by approximately the same amount relative to the control samples. KS tests give probabilities indicating that none of the galaxy pair subsamples are likely to be drawn from the same parent distribution as their respective control samples at $> 5\sigma$ in all cases. The middle panel of Figure 4.9 shows the fraction of each mass ratio subsample which has an offset greater than a given threshold, as a function of that threshold. Here again, the three distributions are roughly consistent in their shift to higher Δlog(SFR), relative to the control sample. KS-tests on the distributions of pair offsets in the 3 mass ratio subsets indicate that the 3 distributions are consistent with being drawn from the same parent population; none of them are inconsistent with the null hypothesis at $> 3\sigma$ confidence.

The similarity between the offset distributions between mass ratio bins is reinforced in the bottom panel of Figure 4.9. All three samples show an excess of positive values of Δlog(SFR) relative to the control. Both major and minor mergers seem equally effective at inducing offsets up to 0.45 dex above the control value, or an enhancement of a factor of ~ 3. These intermediate offsets are roughly 1.3 times as likely to occur in the pairs sample as the control in all samples, with no significant distinction between the more massive and less massive companions in the minor mergers, or between the minor mergers and the major mergers. SFR enhancements of a factor of 2 (+0.3 dex) occur in 30-35% of the pairs sample, and are 1.35–1.65 times more likely to occur in the pairs than in the control. However, offsets more extreme than 0.8 dex are reached almost exclusively through major mergers. These offsets are just as rare as before, occurring in roughly 5% of galaxies, but are more than twice as likely to occur in major pairs as in the control.

In summary, extreme offsets are rare in the pairs sample, but are most likely to be found in the close pairs with approximately equal masses. SFR enhancements of +1.0 dex (a factor of 10 higher than the control) are 3 times as likely to occur in the pairs as in the control, but occur in only 3% of the total pairs sample. Similarly, metallicity offsets of –0.25 dex (a factor of 1.78 lower than the control) are 1.5 times as likely to occur in the pairs as in the control, but also are only found in 3% of the pairs sample. Conversely, modest SFR enhancements, up to a factor of ~ 3 beyond the control, can be found as readily in the minor mergers as the equal mass pairings.

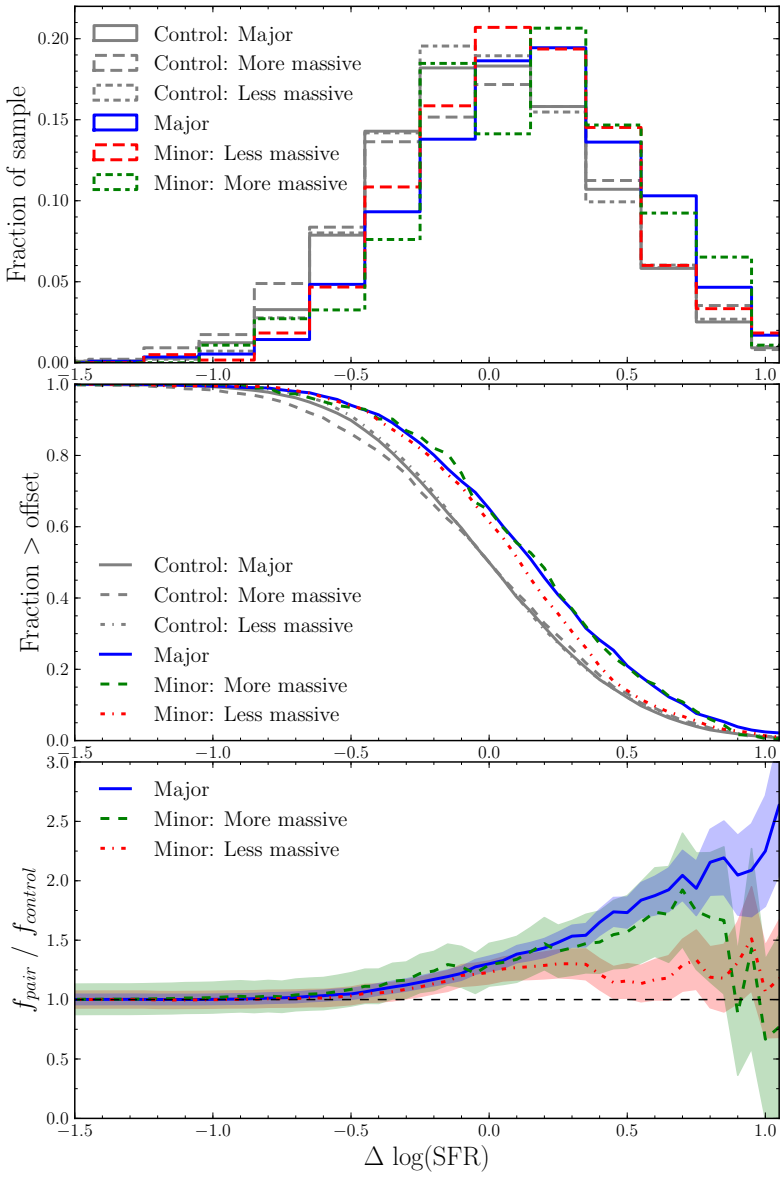


Figure 4.9: Same as Figure 4.7, but split by mass ratio instead of r_p. The top panel shows the distribution of offsets for the three mass ratio bins: major mergers ($0.33 \leq \mathrm{M}_{Host}/\mathrm{M}_{Companion} \leq 3$), the less massive companion in a minor merger ($\mathrm{M}_{Host}/\mathrm{M}_{Companion} < 0.33$), and the more massive galaxy in a minor merger ($\mathrm{M}_{Host}/\mathrm{M}_{Companion} > 3$). The control galaxies for the galaxies in each mass ratio bin are plotted in grey. The centre panel shows the fraction of each of the three samples (and respective controls in grey) which has an offset greater than a given value. The bottom panel shows the ratio fraction of the pairs galaxies above the threshold offset to the fraction of the control above that offset. Shaded regions indicate $\sqrt{N}$ errors.

4.4.2 Evidence for synchronized SFR triggering

Previous work has found that galaxies in pairs often have both galaxies show enhanced star formation more frequently than would be expected at random (Xu et al., 2010). This is interpreted as the signature of the two galaxies in an interaction undergoing synchronized SFR enhancement, as both galaxies undergo similar tidal torques due to their companion. This physical picture is supported by simulations of equal mass mergers, where both galaxies tend to show similar responses to the interaction (e.g., Torrey et al., 2012). Some observational evidence for merger-driven synchronicity between galaxies has already been found in galaxy pairs with AGN. Ellison et al. (2011) found that in galaxy pairs at small separations, a galaxy hosting an AGN was twice as likely to have a companion also hosting an AGN than would be expected at random. Therefore it is pertinent to investigate whether the Δlog(SFR) values also show evidence of synchronized enhancements.

First, only those galaxies from the total pairs sample where both the pair and the companion exist in the sample are selected. This reduces the sample to 45% of the original sample; 425 galaxy pairs remain in the sample (850 galaxies). In Figure 4.10, the host Δlog(SFR) versus its companion's Δlog(SFR) is plotted, where the points are colour-coded according to projected separation. Each galaxy pair is plotted only once on this diagram. It is clear that there is an overabundance of points in the double-enhancement (top right) quadrant of the figure; this quadrant contains 48.0% of the total pairs sample. The double-deficit quadrant contains only 14.59% of the sample. To gain a sense of what would be expected at random, random control galaxies are paired with each other, and find the distribution of points for the control offsets calculated in Section 4.4. As expected, each quadrant contains 25% of the control pair distribution. This indicates that the pairs are 1.9 times as likely to fall in the double-enhanced quadrant than would be expected from the control. Splitting this sample into galaxies in major mergers and those in minor mergers results in an almost identical fraction of galaxies with doubly enhanced SFRs (47.42% in major mergers vs 49.25% in minor mergers). The minor mergers show a slightly lower fraction of galaxies with double suppressions (2%) relative to the major mergers.

However, since the Δlog(SFR) values are higher in the pairs sample than they are in the control sample, simply having an excess of galaxies with doubly enhanced SFRs does not necessarily indicate that there is synchronous triggering. This could instead be the result of comparing between two samples which are not drawn from the

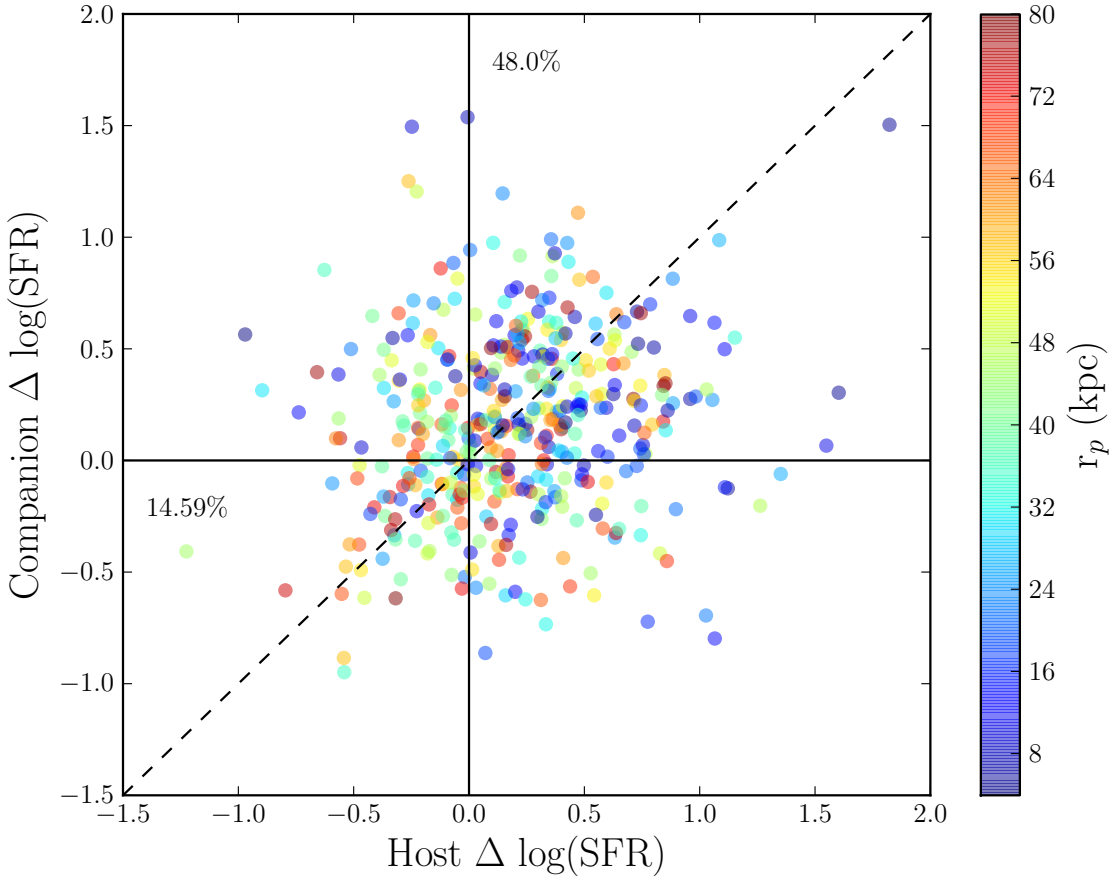


Figure 4.10: For the subset of galaxy pairs where both the host and the companion are in the sample, (∼45% of the sample; 425 galaxy pairs), the distribution of the galaxy pair Δlog(SFR) vs. its companion's Δlog(SFR). The colour bar indicates the projected separations of the pairs. 48% of the sample is found where both the pair and its companion have Δlog(SFR) > 0 (top right quadrant). Using a random galaxy pair match of control sample galaxies, each quadrant is equally populated. Galaxies in pairs are 1.9 times as likely to show double SFR enhancements than a random control sample, and 0.6 times as likely to show double suppressions (bottom left quadrant). Relative to a false pairs sample, the galaxies in true pairs are >17% more likely to show double enhancements.

same parent distribution, rather than indicating that pairs of galaxies are showing correlated SFR enhancement. In order to eliminate the issue of the differing parent distributions, instead of comparing the population of the pairs in this diagram to the control sample, the pairs sample must be compared to itself. Each galaxy in the sample is reassigned to a random companion, thereby scrambling the pairs sample. This results in a set of uncorrelated pair Δlog(SFR)s which can be plotted on the same diagram, but which will have the same distribution of offsets as the true pairs sample. The scrambled pairs populate the double enhancement quadrant with 40.99% of the total sample. 13.28% of the scrambled pairs sample falls in the double-suppression quadrant, identical to the true pairs sample. However, the double enhancement quadrant has 17% more galaxies in the true galaxy pair sample than in the scrambled pairs.

The SFR offsets have some intrinsic scatter, and galaxies very close to the zero line may simply be scattered to one side of zero or another. To minimize the effect of this scatter, a buffer around the zero value can be included, such that galaxies within a certain range around the 0 offset value are not counted in this fractional counting. If the buffer is used, the strength of the double SFR enhancement increases. If galaxies which have offsets beyond $\pm$0.13 dex (the 25th percentile for the control distribution) are considered in the fractional calculation, then the true pairs are 34% more likely to show double enhancements.

4.5 Comparison with Theoretical Models

The SFRs and metallicities of the pairs are significantly offset from the control over significantly larger distance scales than has been previously seen (e.g., Lambas et al., 2003; Alonso et al., 2006; Woods & Geller, 2007; Ellison et al., 2008b). The interpretation of the smoothly declining trend with increasing r_p has traditionally been that galaxies promptly undergo a burst of star formation as the galaxies reach pericentre. This burst would then decline to a fiducial value as the galaxies separate. However, Figure 4.6 indicates that galaxies are still offset from the control values at $\sim$80 kpc h^{-1} separations, and the wider separations show a plateau instead of a smooth decline to control values. The existing interpretation is insufficient to account for the wide separation plateau, as even a long lasting starburst should still show a smooth decline to the control values. A turn to theoretical models to aid in the interpretation of these trends is therefore required.

The simulations of major galaxy mergers developed in Torrey et al. (2012) are used to help interpret these results.[6] A detailed analysis of varying mass ratios and orbital parameters has previously been explored elsewhere (e.g., Di Matteo et al., 2007; Cox et al., 2008). Here, the goal is to take a general look at the SFR and metallicity changes in a merger as a function of r_p. The reader is referred to Torrey et al. (2012) for a detailed description of the simulation setup and parameters. Briefly, Torrey et al. (2012) presents a set of N-body/Smooth Particle Hydrodynamics simulations using GADGET-2 (Springel, 2005). These models include cooling, star formation, feedback, and chemical enrichment. Within these simulations, the metallicities are defined as the mass-weighted average of the metallicities of all gas particles within a sphere of 1 kpc around the centre of the galaxy. Increasing the radius of this sphere only mildly alters the results of the simulations, and does not affect the conclusions drawn. (As a comparison, the physical diameter of the SDSS fibre is generally of order of a few kpc h^{-1}.) Galaxies are shown to be stable (i.e., do not develop a bar) for at least 2 Gyr when modelled in isolation. For the purposes of constructing an interpretive framework for the SDSS data presented here, a suite of 16 galaxy mergers is used, varying only the galaxy orientations between mergers. As a result, their orbits are kept constant through the suite. The model galaxies are merged with identical copies of themselves in all 16 simulations[7].

An example of the evolution of one of the 16 mergers is shown in Figure 4.11. This figure shows the time evolution of the separation of the two nuclei in the top panel, the change in metallicity in the centre panel, and the change in SFR for both galaxies in the bottom panel, where $t = 0$ is scaled to coalescence. Snapshots of the galaxies' gas densities are shown as insets at the top of the figure. Comparable values to the observational $\Delta\log(O/H)$ and $\Delta\log(SFR)$ values are extracted from the simulations. $\Delta\log(O/H)$ and $\Delta\log(SFR)$ are calculated as the difference between the merging galaxy's SFR or metallicity at any given time and the SFR or metallicity of a model quiescent disk at that same time. The two lines in the bottom two panels show the responses of the two galaxies in the merger. For major mergers, the close tracking of the two lines is a general feature of the merger simulations. The magnitudes of the SFR enhancement and metallicity dilution may shift slightly from

[6]The simulations described here were set up and run by Paul Torrey, although the specific suite of simulations described here was developed collaboratively between Paul Torrey and JMS specifically for its inclusion in this work as a point of comparison.

[7]The model galaxy is Disk B in Table 1 of Torrey et al. (2012), but with a 25% gas fraction, and merged on all 16 orientations in Table 2 of Torrey et al. (2012).

merger to merger, but the overall shape of these tracks with time is consistent across the suite of simulations. Therefore, in some runs, the metallicities may re-enrich past the initial value, but this re-enrichment is almost always bracketed by periods of metallicity dilution. Similarly, the strength of the initial SFR burst may vary slightly from merger to merger, but always diminishes again, and is dwarfed in magnitude by the peak in SFR as the galaxies reach the end of their merger.

Of particular note in Figure 4.11 is the time delay between first passage and the strongest metallicity dilutions and highest SFR enhancements after pericentric passage. Metallicity dilutions do not reach their strongest values until $\sim$ 200 Myr after first passage; the peak in SFR enhancement is slightly longer, at $\sim$ 275 Myr[8]. This delay means that the galaxies have had time to progress out to wider separations by the time the triggered SFR enhancement or metallicity dilution will be strongest. There also is a period of metallicity re-enrichment after the first starburst, prior to coalescence. Recently formed stars will return enriched gas to the interstellar medium as they reach the end of their lifespans, which drives the galaxies to nearly return to their initial metallicity. The flow of metal-poor gas to the central regions of the galaxies continues throughout the interaction, re-diluting the central metallicities. As galaxies progress into final coalescence, metallicity offsets drop to their lowest values, and Δlog(SFR) reaches the strongest enhancement of the merger sequence, due to the large torques exerted as the nuclei coalesce.

In order to more directly compare the results of the simulation to the data, the Δlog(O/H) and Δlog(SFR) tracks must be folded into a plot as a function of separation, rather than as a function of time elapsed. Figures 4.12a and 4.13a show metallicity and SFR tracks respectively for one galaxy's response to a merger, as a function of true physical separation. The metallicity and SFR are measured at fixed time steps (10 Myrs) throughout the merger, and points are colour-coded as a function of time (from dark blue at early times to yellow at late times). Figure 4.12a shows that the maximal dilution in Δlog(O/H) after the first close passage seen in Figure 4.11 occurs at a separation of $\sim$ 60 kpc (in this particular merger), with the final metallicity drop due to coalescence seen at real separations of < 15 kpc, as the galaxies coalesce. Similarly, the strongest SFR enhancement after first passage occurs in a relatively narrow peak between $45-65$ kpc, with the coalescent peak appearing at separations < 15 kpc. Interestingly, between the two peaks, the SFR appears to drop

[8]These timescales are primarily set by the free-fall timescales of the systems, given a rotational velocity of the simulated galaxy of 130 km/s.

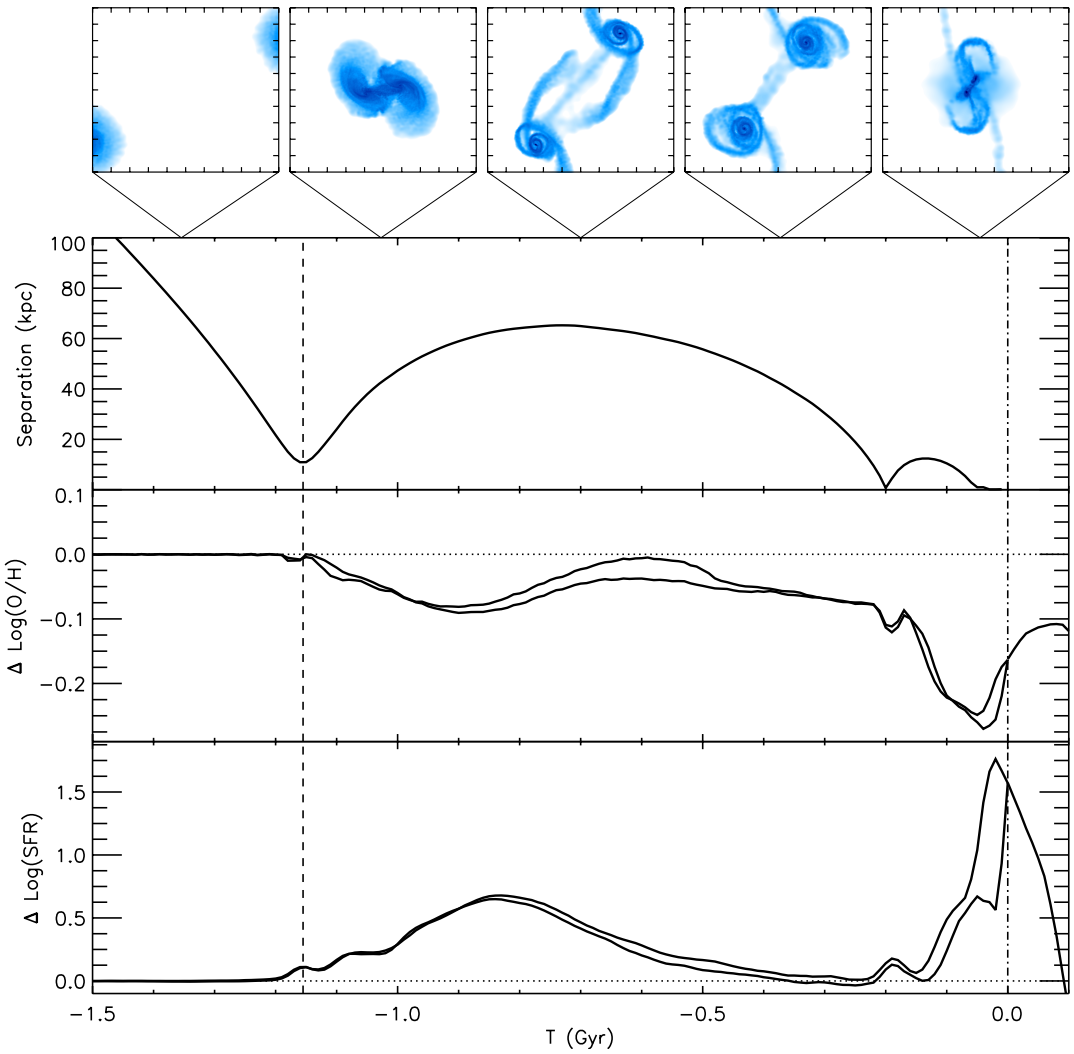


Figure 4.11: Simulation snapshots, Δlog(SFR) (within the central 1 kpc), & Δlog(O/H) as a function of time. The dashed line marks first pericentre, and the dot-dashed line indicates coalescence. The top series of panels shows the surface density of the gas within the galaxy at different points in the merger. Δlog(O/H) is the difference between the metallicity of a quiescent disk and the metallicity of the merging galaxy at any given time step, and Δlog(SFR) is the difference between the current SFR at a given time and the SFR of a galaxy modelled in isolation. The two solid lines in the metallicity and SFR panels indicate the responses of the two galaxies in the simulation. The separation panel shows the intergalactic distance as a function of time, which remains fixed through the full suite of simulations; Δlog(O/H) andΔlog(SFR) change between simulations as the galaxy orientations change.

to nearly its original value; high SFR offsets are generally present at wide separations after a close passage, or at very small separations due to coalescence. However, the narrowness of the peak in SFR at 60 kpc h^{-1} in separation space is a result of the coincidence of the galaxies being at apocentre during the period of time when the SFR enhancement strengthens. There is no reason to expect the apocentre and the SFR burst to coincide; this is a coincidence of the merger orbital parameters. If the galaxy takes longer to reach apocentre, the SFR enhancement would begin prior to apocentre, and this SFR peak would have a broader distribution in separation space.

Physical separations are not an observable quantity, so in order to place the results of the simulations on more comparable footing with the data, the full suite of 16 mergers is introduced, and the physical separation tracks are converted into projected separation space. The conversion takes each point from the simulations and multiplies by a random viewing angle in 3-dimensional space, defined as $|\cos(\phi)|$, where $\phi = sin^{-1}(R)$ and R is a different random value between 0 and 1 for each point in the simulations. The cosine of a random angle accounts for the 2 dimensional spin of the viewing angle; to account for the distribution of angles over the surface of a sphere, the arcsine of a random number between 0 and 1 is used, which should properly account for the 3 dimensional distribution of possible viewing angles. The simulated Δlog(O/H) and Δlog(SFR) can then be plotted as a function of projected separation. The contours of the scattered points are shown in Figures 4.12b & 4.13b respectively. The projected separation contours are time-weighted, as each point in the real separation simulations is measured at fixed time intervals. In regions of real space where the galaxies spend most of their time, there is a corresponding increase in the number of points at that distance.

Strikingly, these contours outline the same general form as is seen in the SDSS sample of pair galaxies. The metallicity contours outline a low level plateau of metallicity dilution out to 70 kpc h^{-1}, with a sharper drop at very small separations. The contours do not show a marked change from the individual tracks, other than an expected blurring of minor features due to projection effects and the scatter introduced by including several merger orientations. The SFR enhancements undergo a more dramatic transformation between real and projected space. The reasonably tight peak in SFR at wide separations has scattered to smaller separations, resulting in a distinct plateau of enhanced SFR out to 70 kpc h^{-1}. Converting to projected separations will scatter galaxies to smaller separations, so the wide separation peak has filled in the previous gap between the high r_p post-pericentre peak and the peak

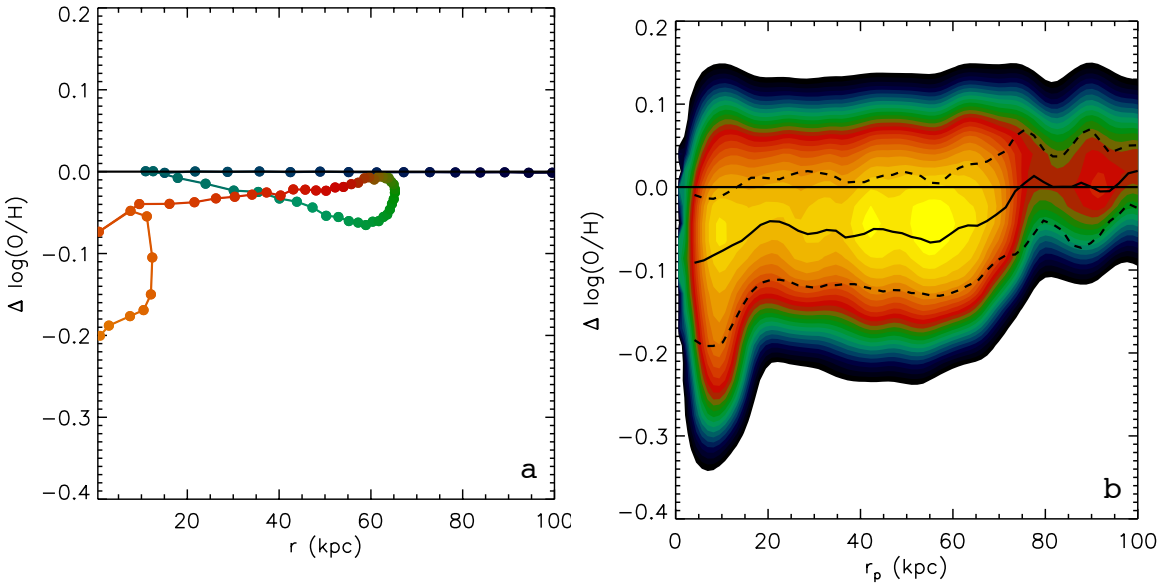


Figure 4.12: Panel a: 1 sample track of the metallicity offset as a function of true separation. Colour indicates the progression of time through the merger, with dark blue at early times to yellow at late times. Each dot indicates a fixed time step of 10 Myrs as the simulation progresses. Panel b: 16 tracks, converted into projected separations. Colours indicate the density within the contours, with yellow indicating the highest density of points from the simulation and blue indicating the lowest density. The solid black line overlaid on the contours indicates the median value, and the dashed black lines indicate the 25^{th} and 75^{th} percentile range.

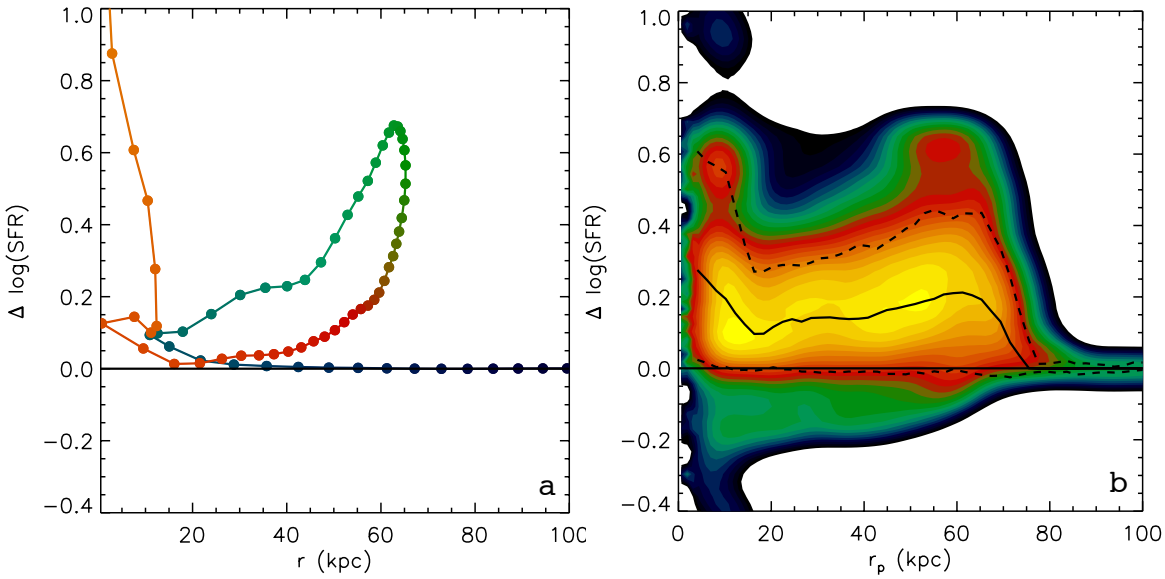


Figure 4.13: Same as Figure 4.12, but for SFR. The contours in Panel b outline the same general form as is seen observationally, with a wide separation plateau, and higher SFR enhancements at small separations.

due to coalescence.

The cutoff of the offset plateau at $\sim$ 70 kpc h^{-1} is a result of the simplified merger suite. As the simulation suite does not explore a large range of potential orbits, varying only the galaxy inclination, none of the mergers reach separations beyond $\sim$70 kpc h^{-1} after their first pass. With a wider suite of simulations, this low level plateau would be expected to naturally extend as galaxies reach larger distances after their first pass, as seen in Patton et al. (2013). Observationally, this would correspond to galaxies in either more weakly bound interactions than those simulated here, or galaxies with higher initial angular momenta. The residual peak at 60 kpc h^{-1} is also a function of the fixed initial orbital energy and angular momentum in all 16 of the simulated mergers; keeping these two parameters the same results in fixed orbits. With a wider range of orbits, this feature should weaken into a flatter plateau, as galaxies with smaller and wider apocentres blend together. Since the strength of the tidal interaction is expected to directly impact the strength of the induced SFR (e.g., Di Matteo et al., 2007; Cox et al., 2008), it would also be expected that with a broader range of tidal interactions, a corresponding increase in the range of SFR enhancements induced in the simulations should be visible.

The qualitative similarity between the contour plots resulting from this simple suite of 16 major mergers and the observational results makes the framework of these simulations an appealing one for the interpretation of the observed trends. In this scenario, the innermost peak in offset values should be due almost entirely to galaxies approaching coalescence, while the wide separation plateau is due to galaxies which have gone through a close passage, and are only showing the SFR and metallicity response at wider separations. These galaxies could ultimately merge, or be part of a population of fly-by encounters, which are also expected to show SFR enhancements (e.g., Di Matteo et al., 2007; Montuori et al., 2010).

It is necessary to caution that these simulations are not intended to function as a direct quantitative comparison to the data, but simply as a theoretical framework to help interpret the form of the signal observed in the SDSS data. With a basic set of simulations, the suite has not reproduced (nor indeed, attempted to generate) a representative sample of the range of mergers that exist within the SDSS sample. The galaxies in the simulations are major mergers only, with a single progenitor galaxy, on a constant merging orbit. The analysis can state only that the simulations do not lie in a region of parameter space devoid of points in the SDSS data. With these concerns in mind, the SDSS data are not overplotted on the simulation contours, nor

do is a direct comparison of the magnitude of these effects encouraged. There are a large number of parameters which could change the magnitude of the signal seen in this suite of simulations (e.g., a varying fraction of interloping or pre-pericentre galaxies, the influence of minor mergers, gas fractions, or initial orbits), many of which have been the subject of other simulation studies (e.g., Di Matteo et al., 2007; Cox et al., 2008). However, the comparison of the general form of the trends with r_p seen in the Δlog(SFR) and Δlog(O/H) figures is robust. The fact that the trends seen in the SDSS sample is reproducible using a straightforward set of simulations is encouraging. This match in form indicates that the simulation does not need to be tuned to the particulars of the pairs sample to observe the same general trends.

4.6 Discussion

In a strictly selected sample of galaxy pairs from the SDSS DR7, the star formation rates are typically enhanced by at least 30% out to separations of 80 kpc h^{-1}. The metallicities are suppressed by –0.02 dex ($\sim$ 5%) within 60 kpc h^{-1}. All galaxies in the sample are visually classified for signs of morphological disturbance to further clean the sample of galaxies which have not yet undergone an interaction or of the remaining fraction of interloping galaxies. When only the disturbed subsample is taken, the metallicity trend increases in significance over the entire range of r_p and is offset from the control values out to at least 80 kpc h^{-1} by -0.03 dex. Within the inner 30 kpc h^{-1}, the disturbed sample is offset by –0.04 dex (9%) in metallicity, and enhanced by 65% in SFR, relative to the control sample. Although previous studies have found enhanced SFRs and diluted metallicities in samples of close pairs (e.g., Ellison et al., 2008b; Kewley et al., 2006), this is the first time that the changes in SFR and metallicity have been studied simultaneously as a function of projected separation.

4.6.1 Star formation rate enhancements out to 80 kpc.

Most previous work using large sample statistics has found that the enhancements seen in SFR are restricted to within 30–40 kpc h^{-1} separations (e.g., Xu & Sulentic, 1991; Lambas et al., 2003; Alonso et al., 2006; Ellison et al., 2008b; Robaina et al., 2009). Indeed, in the pairs sample, the strongest signals are present at the smallest separations. However, statistically significant offsets in metallicity and SFR are

present out to the maximum separations of the pairs sample (80 kpc h^{-1}).

Many previous works have used a mean control value as the point of comparison for their SFR enhancements (e.g., Lambas et al., 2003; Alonso et al., 2006; Ellison et al., 2008b; Lambas et al., 2012), whereas this project has used 'offsets' for individually matched galaxies. As a test, the analysis is repeated using the mean of the control sample as the baseline for the SFR enhancements. The results are broadly unchanged; significant SFR enhancement at 80 kpc h^{-1} is still observed. Evidently, the change in methodology is not entirely responsible for the increase in sensitivity to changes in SFR at wide separations. If the methodology is not the dominant improvement, then the other major set of differences between this work and previous works is in the definition of a control sample. While most of the previous works match their pair and control samples in redshift, few of them are matched in stellar mass. As the SFR of a galaxy is strongly dependent on mass, if the distributions of mass between the pairs and the control galaxies are not well matched, this could well influence the strength of the overall SFR enhancement measured. For instance, Lambas et al. (2003) and Darg et al. (2010) match to a control sample only in redshift. Other works have used magnitudes as a proxy for mass (Woods & Geller, 2007; Wong et al., 2011; Alonso et al., 2006). However, Ellison et al. (2008b) showed that this method can dilute trends with galaxy properties with a mass dependence, such as SFR. The conversion from magnitude ratios to mass ratios is not 1:1 (see their Figure 3), so taking magnitude ratios as a direct proxy for mass will introduce a significant fraction of galaxies of different masses than intended. Perez et al. (2009) compares different methods of selecting a control sample, and finds that matching only on luminosity and redshift introduces significant biases into the result, resulting in artificially redder, more massive pairs, which live in higher density regions of space, relative to their controls. The biases introduced by matching in luminosity and redshift can be reduced by 70% by instead matching a control sample in mass, redshift, and local density.

The pairs sample is matched to a control in total stellar mass, redshift, and in local density. Patton et al. (2011) and Ellison et al. (2008b) also match in mass and redshift. The extent of the SFR offset trend is remarkably similar to the results found in Patton et al. (2011). The Patton et al. (2011) study used a superset of the sample presented here, as it did not require the gas phase metallicity calculations, and uses a comparable offset methodology to ours; direct comparisons between the results are robust. Patton et al. (2011) find that the central $(g-r)$ colours of blue cloud galaxies are consistently bluer than the control values out to 80 kpc h^{-1} (see

their Figure 15). Ellison et al. (2008b) sees a hint of this trend in major mergers, but the errors are large enough that the trend is not systematically significant to the widest separations, and the full sample only shows significant enhancement at $r_p <$ 30 kpc h^{-1}. The pairs sample used here is moderately larger than that of the Ellison et al. (2008b) work, at 1899 galaxies versus their 1719, and this work has a slightly more stringent Δv requirement of 300 km/s instead of 500 km/s, both of which would help in reducing error bars. Furthermore, Ellison et al. (2008b) use total star formation rates, rather than the fibre values used in this chapter. Patton et al. (2011) determined that the strong offsets in colour are primarily a nuclear effect, as the total $(g - r)$ colours of the galaxy showed smaller offsets across all r_p by a factor of ~ 4, and lacked the small separation peak to bluer colours. The aperture corrected SFRs provided in Brinchmann et al. (2004)[9] are used to test whether an analogous weakening of the magnitude of the SFR offsets is seen. Consistent with the Patton et al. (2011) work, the galaxy's total SFR shows a weaker trend than the fibre values, and lacks the central spike to higher SFR. However, SFR offsets for the total values are still visible at 80 kpc h^{-1}, so while using the fibre values increases the magnitude of the SFR enhancements, the use of the fibre values alone is not enough to account for the new sensitivity. A similar comparison is impossible for the metallicities, as there is no correction from fibre metallicities to total metallicities. The nuclear concentration of the SFR enhancement indicates that it is much more likely to detect strong SFR offsets when looking at the fibre values rather than the total values, as in Ellison et al. (2008b). Therefore, it is proposed that one of the main factors in increasing sensitivity to small effects in SFR and metallicity is due to the tightly mass-matched control sample. Further aiding the sensitivity of this work is the use of fibre SFRs rather than the aperture corrected values.

4.6.2 Mass Ratios

As suggested by many previous studies, the form of the trend with projected separation is primarily driven by the contribution of major mergers (e.g., Woods et al., 2006; Woods & Geller, 2007; Ellison et al., 2008b). Investigating the more massive and less massive companions in minor mergers (mass ratios more extreme than the 3:1 major merger criterion), both the less massive and more massive companions show SFR

[9]Brinchmann et al. (2004) use model fits to the photometry outside the fibre to make the correction from fibre to total values. A complete discussion of potential biases is presented in that work.

enhancement at all r_p. The more massive galaxies distinguish themselves from the less massive galaxies only in the smallest separations (see Figure 4.4). Major mergers have slightly higher median offsets, but the median value is very comparable to that of the minor merger enhancements. Lambas et al. (2012) find a similar effect; minor mergers show enhanced SFRs at all masses, relative to the control. Their sample of major mergers is found to have SFRs enhanced at a slightly stronger level than the minor mergers, but in large part the offsets of the two samples are consistent within error bars. Minor mergers are just as effective as major mergers at inducing SFR enhancements of less than a factor of 3 over the control. However, the rarest, most extreme starbursts occur almost entirely within major mergers. Since these extreme starbursts are so rare, the overall offset distributions are consistent between major mergers and minor mergers.

While previous studies of SFR enhancement as a function of mass ratio have drawn conflicting interpretations from their data (e.g., Lambas et al. 2003 determined that the more massive galaxy in a minor merger is more strongly affected, while Woods & Geller 2007 found the inverse), a closer inspection of the data indicates that these two previous observational results and that of the current work are not inconsistent. Woods & Geller (2007) found no evidence for a correlation between small r_p and high SFR for the higher mass companion in a minor merger, whereas there was some evidence of correlation for the less massive companion, and conclude that the less massive companion is more strongly affected by a merger than the high mass companion. However, since the SFRs of the galaxies in pairs are not directly compared to a control sample, it is impossible to judge whether or not these galaxies are also systematically enhanced over the control value. For example, the current sample of less massive galaxies in a minor merger shows systematic SFR enhancement, but no correlation with the r_p of the galaxy pair, so on this point the data presented here do not necessarily conflict.

Woods & Geller (2007) also found that when comparing the distributions of specific star formation rates (SSFRs), the distributions of both the more massive and less massive galaxies in minor pairs were statistically consistent with the field sample. Lambas et al. (2003) also find that galaxies in a minor merger show no significant SFR enhancement at any r_p, compared with the average control SFR value. In contrast to this previous result, the distribution of star formation rates at a given mass found here are statistically different for galaxies in minor pairs and the control sample; KS tests indicate that the less massive galaxies are unlikely to be drawn from the same

parent distribution as the control at $\sim 6\sigma$, and at $> 5\sigma$ for the more massive galaxies in a minor merger. Both Lambas et al. (2003) and Woods & Geller (2007) are subject to the same issues with matching a control sample in magnitudes instead of using stellar masses described in Ellison et al. (2008b), which will weaken the sensitivity of their measurements.

While these results are not in conflict with the observational results, they do not appear to align with the expectations from simulations. There are relatively few simulations which have investigated the SFR enhancements of galaxies in unequal mass mergers. Cox et al. (2008) find that high mass galaxies in unequal mass mergers are less likely to be tidally perturbed and drive strong SFR enhancements when interacting with a low mass companion, whereas the low mass companion will be strongly perturbed by a massive companion. Both Cox et al. (2008) and Mihos & Hernquist (1994) find that minor mergers can drive some gas inflows in the massive companion, but the inflowing gas is not necessarily converted into stars. Mihos & Hernquist (1994) find that the massive companions only show enhanced SFR at coalescence, which does not help explain the trends observed at wide separations. Cox et al. (2008) is meant as an improvement upon the Mihos & Hernquist (1994) work, and finds that the massive galaxy is unlikely to undergo a starburst except in very specific cases. If this theoretical model were borne out, it might then be expected to see significantly stronger $\Delta\log(\mathrm{SFR})$ in the less massive companions, particularly at small separations, where the galaxies are likely to be in the final stages of a merger. In contrast, the results found here indicate that the more and less massive galaxies show similar levels of SFR enhancement over most of the range in r_p probed by the sample, with the *more* massive companions displaying higher SFR enhancement at the smallest separations.

4.6.3 Magnitude of the SFR enhancement

With these results in hand, a comparison of the magnitude of the SFR enhancements found here to those found in previous works can be made. A series of statistical tests beyond the median values is used to determine which magnitude of offsets are preferentially found in the pairs sample instead of the control. Close pairs ($r_p< 30$ kpc h^{-1}) preferentially contain the strongest SFR enhancements (up to a factor of 10 over the control), and galaxies at wider separations show a statistical excess of more moderate enhancements (up to a factor of 7 over the control). These tests

indicate that while the median values are kept at a much lower Δlog(SFR) due to the relative scarcity of extreme offsets, galaxies in pairs preferentially produce strong SFR enhancements. However, since this is a test unique to this work, the most straightforward comparison to previous studies is a comparison of median values.

The median SFR enhancement in these results is 40% over the control; this roughly corresponds to the value of the wide separation plateau. The enhancements at smaller r_p are between 60% and 80% enhancement over the control values. The magnitude of the SFR enhancement found here is broadly consistent with that found in previous studies. Robaina et al. (2009) find enhancements of order 80% over the control values for galaxies with $r_p <$ 40 kpc h^{-1}, within mass ratios of 1:4, and between redshifts of $0.4 \leq z \leq 0.8$, consistent with the enhancements seen in the inner peak. Lin et al. (2007) find that galaxies up to $z \sim 1$ and $r_p <$ 50 kpc h^{-1} are enhanced by a factor of 1.9 ± 0.4, relative to a control sample. This level of offset is also consistent with the SFR enhancements in this sample. Wong et al. (2011) find that the average enhancement is generally 15-20% for galaxies within 50 kpc h^{-1}, and 25–30% for galaxies within 30 kpc h^{-1} for a sample of galaxies $0.25 \leq z \leq 0.75$. These values are slightly lower than even the wide separation plateau enhancements found here. However, Wong et al. (2011) use a very generous Δv cut of $\Delta v \leq$ 3000 km s^{-1}, which will introduce a non-negligible fraction of interloping galaxies into their sample, their results are a lower limit. Some differences in the observed SFR enhancement between works may also be introduced based on differences in the method used to obtain the SFR values, as different methods are well known to give different results (e.g., Kennicutt & Evans, 2012, and references therein). For instance, Wong et al. (2011) use dust-corrected ultraviolet colours as a tracer of SFR, while Lin et al. (2007) use infrared luminosities to calculate the SFRs. Both the effectiveness of the dust corrections and the robustness of the control sample will influence how robust the enhancements are to comparisons between studies. With these caveats in mind, the relative strength of the results found here is consistent with those of Wong et al. (2011). It is intriguing that the average SFR enhancement in galaxy pairs is consistent with these results even at redshifts up to $z \sim 1$.

4.6.4 Metallicity offsets

There has been very little work on the dependence of metallicity on projected separation, and certainly none with the statistics presented here. Kewley et al. (2006)

have a sample of 86 galaxies in pairs with $r_p < 50$ kpc h^{-1} and found that metallicities were only significantly diluted at separations closer than 20 kpc h^{-1}. Galaxy pairs with $r_p < 20$ kpc h^{-1} have an average suppression of -0.2 dex relative to non-pairs. However, Kewley et al. (2006) uses the luminosity-metallicity relation in lieu of a mass-metallicity relation to calculate their changes in metallicity, and Ellison et al. (2008b) showed that $\sim 50\%$ of this offset was likely due to an increase in luminosity due to enhanced SFRs. While the median values found here are not nearly as low as those found in Kewley et al. (2006), (between -0.03 and -0.04 dex within 20 kpc h^{-1} for the total and disturbed subsamples respectively), the unbinned points in Figure 4.2 show a significant population of galaxies at offsets of -0.1 to -0.2 dex, with extreme outliers down to -0.7 dex. Changes in metallicity may also be sensitive to internal galaxy parameters such as the gas fraction, similar to the dependence for SFR suggested by Di Matteo et al. (2007). Metallicity dilutions have been shown to be strongest in low density environments, with metallicity enhancements present in cluster environments, where the gas fraction is likely to be low (Ellison et al., 2009). In support of this idea, Skillman et al. (1996) find that the most gas-deficient galaxies in the Virgo cluster were also the most metal-rich. Including low gas fraction galaxies in the sample may therefore shift the average $\Delta\log(\mathrm{O/H})$ to values closer to the control sample.

Rather than probing metallicity as a function of separation, most prior work has has focused on the mass–metallicity relation in a sample of closer pairs, selecting only galaxies within 25–30 kpc h^{-1} (Ellison et al., 2008b; Michel-Dansac et al., 2008) where the effects of the merger are expected to be the most visible. In this context, it has been demonstrated that galaxies in pairs are systematically lower in metallicity at fixed mass, relative to a non-pair control, and that the magnitude of this effect is approximately -0.03–0.05 dex (Ellison et al., 2008b; Michel-Dansac et al., 2008). If the total sample is limited to the galaxies with $r_p < 30$ kpc h^{-1}, a median $\Delta\log(\mathrm{O/H})$ of -0.03 dex is obtained; for the disturbed subsample, the median offset increases to -0.04 dex. Michel-Dansac et al. (2008) calculates metallicity offsets in a similar way to this work, using the mean metallicity of a set of control galaxies of similar mass as the zero point. However, their metallicity offsets are only visible in the strongly merging sample, and not the tidally disturbed sample. The division by morphological disturbances presented here does not distinguish between these two types, so the higher magnitude metallicity offset seen in Michel-Dansac et al. (2008) (-0.05 dex vs. -0.04 dex) is perhaps unsurprising. Therefore, while the form of the trend with

r_p is novel, both the sign and magnitude of the offsets are consistent with previous studies of metallicities in galaxy pairs.

4.6.5 Simulations

That the trends observed in the SDSS data are broadly reproduced by the simple suite of simulations presented here is evidence that the merger tracks offered by the models may be an accurate framework within which to interpret the results. No effort has been made to tune the simulations to the data set, using the same galaxy model for all interactions. As mentioned in Section 4.5, the suite of simulations is not reproducing a representative sample of mergers, or fully exploring the parameter space of the simulations. For instance, the location of the large separation peak seen in Figure 4.13a is partially a function of the initial angular momentum given to the galaxy pairs. Further, the simulations in this simple suite do not include any galaxies which progress further than $\sim$65 kpc h^{-1} away from their companion after the first passage; as the plateau continues until at least 80 kpc h^{-1} in the data, some fraction of galaxies in the pairs sample would be expected to be involved in more weakly bound (or unbound) interactions, which can carry the galaxies to much wider separations (Patton et al., 2013). With these caveats in mind, the fact that a similar trend to the data is reproduced, but using a minimum of additional assumptions, is very reassuring. This indicates that the models do not need to precisely reproduce the exact mergers in the data to provide a theoretical model of what the galaxies, on a statistical level, are doing as they progress through a merger. Using these models as an interpretive tool, the increase in the magnitude of offsets in the smallest r_p bins is due to the sharp spike seen due to galaxies near the end of their merger at those separations, when large amounts of metal poor gas are dumped into the nuclear region of a galaxy, and a large SFR burst is triggered. This dramatic peak due to coalescence at small separations has been seen before in both theoretical (e.g., Di Matteo et al., 2007; Montuori et al., 2010; Torrey et al., 2012) and observational works (e.g., Larson & Tinsley, 1978; Donzelli & Pastoriza, 1997; Barton et al., 2000; Lambas et al., 2003; Alonso et al., 2004, 2006; Woods & Geller, 2007; Ellison et al., 2008b, 2010; Darg et al., 2010; Xu et al., 2010). By contrast, the wide separation offsets are due to a combination of galaxies making their way out after first passage, or possibly back in for a second or final passage, depending on the orbits of the interacting galaxies. The influence of projected pairs and galaxies that have not yet interacted would result in

an overall lowering of the magnitude of the offsets at all r_p.

It is important to note that the absolute values of offsets depend on the definition of the control pool. In this case, the control pool galaxies have been restricted to have S/N> 5 in the emission lines required to accurately determine metallicities. However, this has the effect of selecting control galaxies that are slightly biased towards high SFR. In turn, this means that the calculated delta SFRs are likely to be slightly under-estimated. Indeed, by calculating SFR offsets from the full star-forming main sequence, as described in Ellison et al. (2013a), the SFR offsets are typically 0.15 dex higher.

4.7 Conclusions

A sample of 1899 close pairs from the Sloan Digital Sky Survey's Data Release 7 was used to study SFR and metallicity offsets as a function of projected separation. A simple suite of simulations is used to interpret these results. The main conclusions of this work are as follows:

1. Galaxies undergoing an interaction show significant metallicity depressions and SFR enhancements, relative to a control sample that is tightly matched in stellar mass, redshift, and local density. Galaxy pairs show enhanced SFRs out to projected separations of 80 kpc h^{-1}, the widest separations in this sample, while significant metallicity dilution is observed out to ∼60 kpc h^{-1}. Within 30 kpc h^{-1}, SFRs are enhanced, on average, by ∼0.21 dex, or by 60%, over the control. At separations wider than 30 kpc h^{-1}, the SFRs are enhanced by ∼0.1 dex (25% above the control). Metallicities are found to be diluted by –0.02 dex (∼5% lower than the control) within 60 kpc h^{-1}.

2. When only the morphologically disturbed subsample of galaxies is taken, the metallicity offsets are significantly offset from the control out to 80 kpc h^{-1}, and the median offset within 30 kpc h^{-1} drops to –0.04 dex (∼9% lower than the control). Galaxies are visually classified as either visibly disturbed or not visibly disturbed to attempt to reduce the fraction of galaxies which have not yet interacted.

3. The form of the $\Delta\log(\mathrm{SFR})$ and $\Delta\log(\mathrm{O/H})$ offsets as a function of r_p are primarily driven by galaxies in major mergers. Both the more massive and less

massive companions in a minor merger show similar SFR enhancements to each other and to the major mergers. The less massive companions show a minor enhancement at all r_p, but do not show the same increased SFR enhancement at small r_p as is visible in the major mergers or more massive companions. Furthermore, the most extreme offsets are preferentially found in major mergers, but minor mergers are equally effective at inducing moderate starbursts in both the less and more massive companions.

4. Galaxies with strong starbursts, e.g., 10 times as strong as the control, are relatively rare, occurring in only 3% of galaxy pairs, but are 3 times as likely to be found in the close pairs sample (separations < 30 kpc h^{-1}) than in the wide pairs or control sample. Extremely diluted metallicities (e.g., 1.8 times as metal poor as the control) are rare (3%), but are twice as likely to be found in the close pairs sample than in the control. There is evidence for synchronous starbursts as a result of a galaxy pair's interaction. Galaxies in pairs are at least 17% more likely to show significantly enhanced SFR (relative to a scrambled pairs sample) if their companion galaxy is also enhanced.

5. A simple suite of major merger simulations (Torrey et al., 2012) is used to construct a new interpretation of how interactions affect the SFRs and metallicities of a galaxy, as a function of projected separation. The Δlog(SFR) and Δlog(O/H) offsets from the simulations are calculated for a qualitative comparison with the observational trends. The models are able to reproduce the small separation peak in offset value and wider separation offset plateau seen in the SDSS data. The simulations offer the interpretation that the wide separation plateau is caused by projection effects blurring post-pericentre starbursts at wide separations to smaller separations. The large offset peak at small separations, by contrast, is primarily due to the extreme starburst induced at the end of a merger, near coalescence, and is less subject to projection effects.

Galaxy mergers clearly induce large scale gas inflows in galaxies. However, the observations presented here, combined with the theoretical framework of hydrodynamical simulations, indicates that these inflows occur on longer timescales than has been previously assumed. With metal poor gas reaching the central regions of a galaxy at a slower rate, the nuclear gas-phase metallicities will take longer to dilute significantly, and the gas reservoir necessary to fuel a significant starburst will take longer to accumulate. By the time these signatures of gas flow in a galaxy are

measurable after the first encounter, the two galaxies have progressed out to wide separations. As the galaxies reach the end of their merger, the tidal forces inducing these gas inflows become significantly stronger, and the gas phase metallicities and star formation rates change both more rapidly and more dramatically. Future work will quantify the extent of separations over which it is possible to see the effects of an interaction, and attempt to determine what parameters govern the highest SFR galaxies.

Chapter 5

HI gas fractions in close galaxy pairs

5.1 Introduction

In the previous chapter, it was determined that galaxies are strongly affected by interactions out to separations of at least 80 kpc h^{-1}, and that the driving forces behind the strength of the triggered starburst appear broadly coherent. From Figures 4.7 and 4.8, it is apparent that galaxies in equal mass (major) mergers at the smallest separations (< 30 kpc h^{-1}) are the most likely to be the most strongly starbursting. Only galaxies at small separations in equal mass mergers are able to produce SFR enhancements of a factor of ten above the control.

However, selecting galaxies from the pairs sample of Chapter 4 which meet the criteria of being in major mergers at small separations does not identify a clean sample of strongly starbursting systems. Instead, a distribution of SFR enhancements that spans more than two orders of magnitude remains, ranging from a factor of ten below an undisturbed control to a factor of ten above it (see Figure 5.1). On average, the equal mass, small separation galaxies in pairs are enhanced over their controls by a factor of 1.6 (or 0.2 dex), slightly higher than the average for the full sample presented in Chapter 4. The fact that the range of the distribution remains wide indicates that the mass ratio of the pair and the proximity of the galaxies in projected separation are not the only driving forces behind the strength of the starburst. In this chapter, another potential driver of starburst strength will be investigated as a possible cause of the scatter in the distribution: gas fraction.

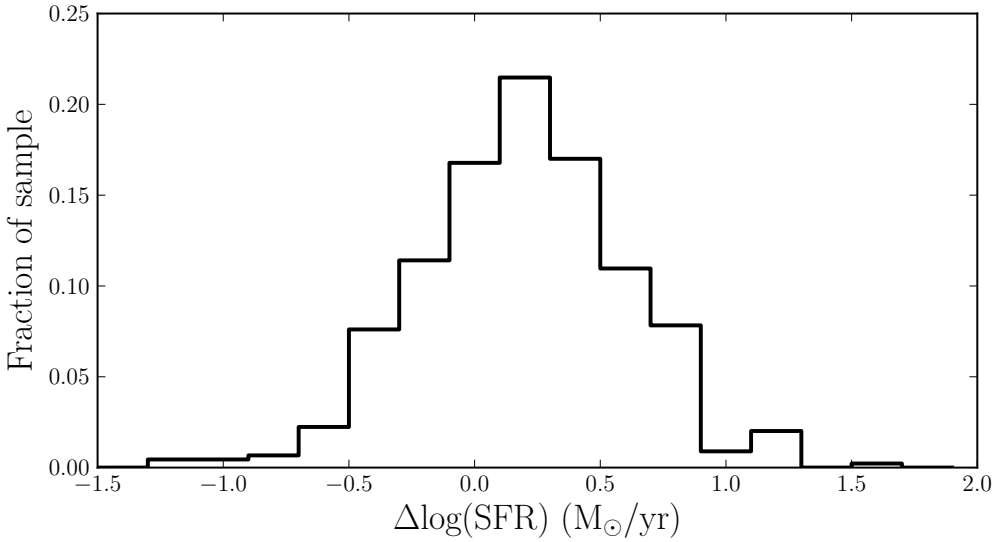


Figure 5.1: Distribution of Δlog(SFR) for the subsample of galaxies from Chapter 4 found in interactions with mass ratios less extreme than 4:1, and within 30 kpc h^{-1}. Δlog(SFR) values remain in the sample which are a factor of ten above or below the control.

Gas Fractions

For this work, the gas fraction is defined as the HI gas mass divided by the stellar mass (M_*)[1]. By this metric, a gas fraction of 1 indicates that the galaxy has an equal mass in HI gas as it does in stars. A gas fraction between 0 and 1 indicates that the galaxy contains more mass in stars than in gas, and any gas fraction above 1 is a galaxy which contains more mass in gas than in stars.

An important distinction between observational and theoretical studies is the typical definition of gas fraction. While the observational definition of gas fraction is usually as adopted above, the simulations tend to adopt an equation of gas fraction similar to:

$$f_{gas} = \frac{M_{gas}}{(M_{gas} + M_*)} \tag{5.1}$$

This equation will output a significantly different value than the previously adopted definition. Under this equation, the maximum possible gas fraction is 1.0, for 100% gas and no stars. Because of the difference in definition, in order to compare observational gas fractions to the results of simulations, a conversion is required. A 50% gas fraction as defined by Equation 5.1 is equivalent to $M_{HI}/M_* = 1.0$.

In the Chapter that follows, theoretical definitions of gas fraction will be represented by f_{gas}, while the observational definition will be annotated as M_{HI}/M_*.

Star formation rate offsets

The values of Δlog(SFR) used in this chapter are distinct from those used in Chapter 4. In Chapter 4, only those galaxies which passed the relatively strict quality control outlined in Chapter 2 were used as control values. This requirement means that, as referenced in Section 4.6, the offset values may be biased towards lower values. If the control values are selected from the upper edge of the star forming main sequence (as these criteria select strong emission line galaxies), then offsets from the true 'average' population would be underestimated. To mitigate this effect, in Ellison et al. (2013a), the Δlog(SFR) values were recalculated using a less stringent S/N cut on the the emission lines. This allows a more representative sample of galaxies through the quality control steps to populate the main sequence. Recalculating the Δlog(SFR) values for all galaxies allows for a more accurate determination of the star formation rate offset from the average population. As in Ellison et al. (2013a), the control sample

[1]Note that this definition of gas fraction does not account for the molecular gas content of the galaxy.

of non-pair galaxies is matched in mass, redshift, and local density, with a typical control sample of > 50 galaxies per pair galaxy. The updated Δlog(SFR) values for both fibre and total SFR quantities are presented in Table 5.4.

5.1.1 Theoretical predictions

The results from theoretical simulations of galaxies do not yet appear to have converged on the importance of gas fraction in a galaxy. Positive, negative, and no correlation options have all been suggested in the literature, each with a distinct and plausible physical explanation.

The physical picture behind a positive correlation is relatively straightforward. Due to the increased gas reservoir in the galaxy, the perturbation provided by the merger has a much larger supply of material available to be funnelled towards the centre of the galaxy, and from which stars can then be formed. Relative to galaxies with less of a gaseous component, these gas rich galaxies can form a considerably larger starburst simply based on the larger amount of raw materials available. Bournaud et al. (2007), Bournaud et al. (2008), Bournaud et al. (2011) and Perez & Sanchez-Blazquez (2011) all suggest that the fact that the perturbation of the merger should be able to drive a greater amount of gas towards the centre with a greater volume available should result in strong positive correlations between the gas fraction of a galaxy and its star formation rate enhancement. Perez & Sanchez-Blazquez (2011) find that galaxies with $f_{gas} = 0.5$ show significantly stronger gas flows to the central region of a galaxy (measured across a 4 kpc h^{-1} sphere) during an interaction than the simulated galaxies with f_{gas} of only 0.2. Bournaud et al. (2011) find that the high gas fractions in mergers at higher redshifts (simulated as $f_{gas} = 0.7$), where the gas fractions of galaxies are, on average, much higher than in local galaxies, can lead to average SFR enhancements of a factor of 10 during the merger, relative to the initial value. Relative to the average enhancement factor of ~ 2 in local mergers, the high redshift enhancement in SFRs is much higher than what is found in local galaxies.

By contrast, Di Matteo et al. (2007) finds that the gas fraction is unimportant to a galaxy's peak star formation rate over the entire merger, as shown in Figure 5.2. For the purposes of comparison with the data presented in this thesis, only the green and orange circles (which describe interactions where both galaxies are late-types) in Figure 5.2 are comparable. The y axis shows the peak SFR found across the span of the entire interaction, and the gas mass plotted along the x axis is normalized to the

amount of gas present in the system before the interaction which triggered that peak value began. For those galaxies which merge, the strongest burst of star formation occurs at coalescence; the pre-interaction gas fraction is lower because the galaxies have already consumed a significant fraction of their gas in a previous close pass and starburst. Those galaxies in flybys have not consumed a significant amount of their gas, as only a small amount of time has passed between the initial state of the galaxy and first pericentre.

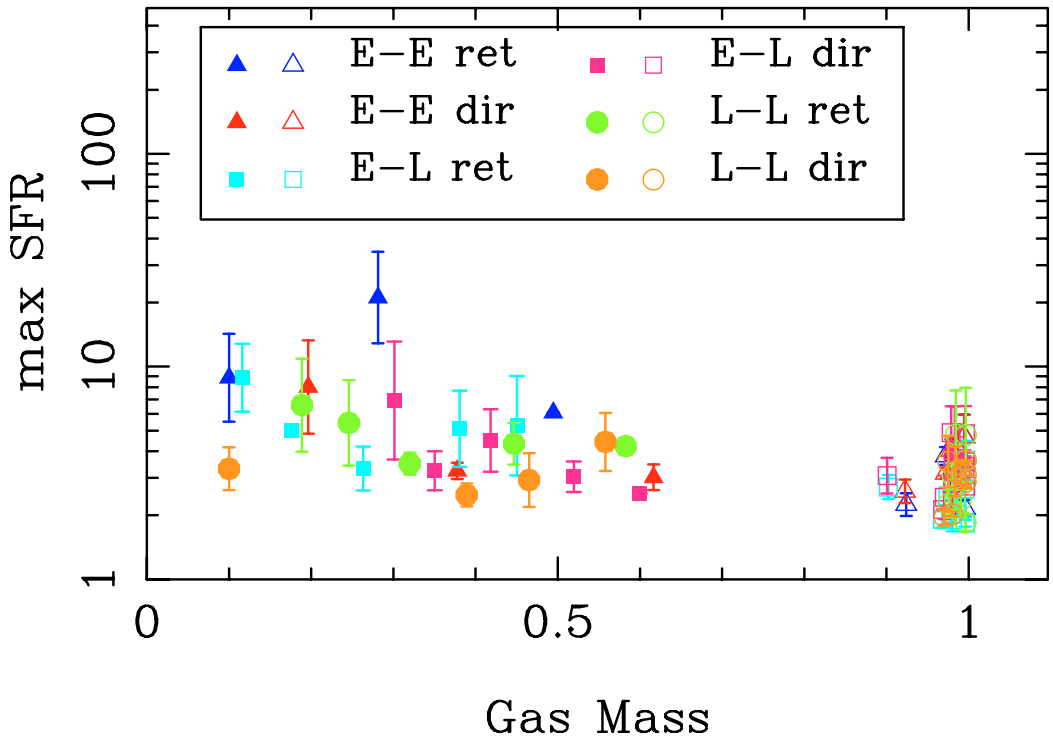


Figure 5.2: This figure reproduces Figure 20 of Di Matteo et al. (2007), which shows the maximum SFR (normalized to a non-interacting galaxy) of the interaction or merger as a function of the gas mass of that galaxy 50 Myr before that burst. Open circles indicate flyby encounters, whereas solid points indicate galaxies which merge. E in the legend indicates an elliptical galaxy, while L indicates a spiral. 'ret' indicates a retrograde encounter (galaxies rotating in opposite directions), while 'dir' indicates a direct encounter, where both galaxies are rotating in the same direction.

Di Matteo et al. (2007) finds that other parameters of the interactions, such as the relative orientation of the two galaxies and the geometry of their orbits, far outweighed the impact of the gas fraction. The strength of the SFR burst at coalescence correlated with the distance of the first pericentric passage, the amplitude of the

tidal forces, and with the orbital orientation, but not with the total gas fraction in the galaxy. An encounter in which the galaxies are rotating in opposite directions (retrograde) often results in a SFR enhancement that is a factor of two higher than direct encounters, where both galaxies are rotating in the same direction. If this dependence on orbital parameters is the driving force behind the SFR enhancement in a galaxy, one would not expect to find any signature of a correlation between the gas fraction and Δlog(SFR).

Hopkins et al. (2009) offers a different perspective. Instead of considering the larger volume of gas as a larger reservoir from which to form stars as Perez & Sanchez-Blazquez (2011) and Bournaud et al. (2011) do, Hopkins et al. (2009) considers the physics of the angular momentum loss of the gas itself. Since gas loses angular momentum via the torque of a gaseous bar against a stellar bar, if the stellar bar is massive relative to the mass of the gas bar, this angular momentum loss will be efficient. This efficient loss of angular momentum permits a significant flow of gas towards the centre of the galaxy. However, if the gas bar is much more massive than the stellar bar, the torque will be much less efficient. This inefficiency translates into a smaller fraction of the galaxy's gas flowing to the centre, which, in turn, means that there is less gas accumulating in the centre of the galaxy to fuel a strong centrally concentrated starburst. In the extreme where the galaxy is 100% gas and 0% stars, there is no stellar bar at all to induce the necessary torque for angular momentum loss within the gas component, and this mechanism will not facilitate gas flow to the central regions. If this physical picture dominates, then a negative correlation between gas fraction and Δlog(SFR) should be present in observations; as gas fraction increases, the efficiency of gas funnelling should decrease, thereby limiting the starburst.

In their Figure 13, reproduced here in Figure 5.3, Hopkins et al. (2009) addresses the impact of fly-by or post-pericentre encounters, which is most relevant for a sample of well-separated galaxy pairs. Figure 5.3 shows the dependence between gas fraction and burst fraction over the complete range of gas fractions, from a galaxy with $f_{gas} = 0$ (0% of the galaxy's mass in gas) to an $f_{gas} = 1.0$ (100% gas). The x axis displays the gas fraction prior to a first passage with the companion galaxy, while the y axis shows the ratio of the fraction of gas consumed in the burst relative to the fraction of gas prior to the burst. Each of the four panels shows a different orbit. The lines plotted in each panel show the analytic prediction based upon the model proposed within Hopkins et al. (2009); this scaling of burst fraction with gas fraction goes as

$1 - f_{gas}$. Between gas fractions of 0.1 and 0.5 ($0.11 < \mathrm{M}_{HI}/\mathrm{M}_{*} < 1$), the burst fraction declines by $\sim$0.1, after which point the burst fraction drops precipetously. The coloured points show the data from hydrodynamical simulation runs after first pericentre passage. Hopkins et al. (2009) notes that there is significant scatter from the simulation points, which is a factor of a few in magnitude, and significantly larger than the scatter found for a complete merger.

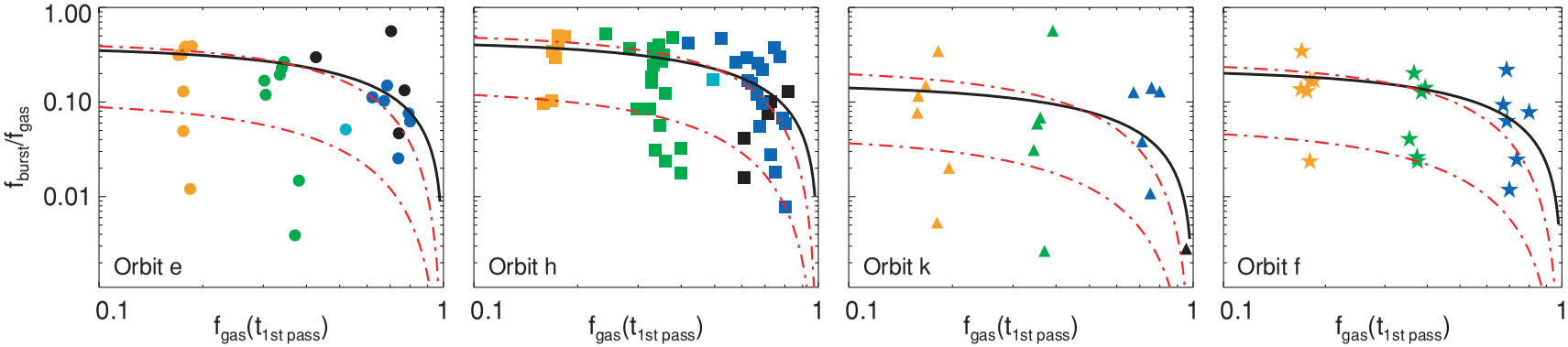


Figure 5.3: This figure reproduces part of Figure 13 of Hopkins et al. (2009), which shows the burst fraction relative to gas fraction as a function of pre-interaction gas fraction for a series of galaxies after first passage. Four different galaxy orbits are shown. The solid and dashed lines describes the models presented in Hopkins et al. (2009) with the dots as the results of the hydrodynamical simulations after first passage.

With the three conflicting physical models proposed by varying theoretical works indicating that the Δlog(SFR) should be either strongly dependent on gas fraction in either positive or negative directions, or not dependent at all on the gas fraction, it is up to the observers to more directly test whether a strong correlation exists in the data. Thus far, very few observational studies have investigated the role of gas fraction on the enhancement of an interaction-triggered burst. However, one exception is Xu et al. 2012, which compares triggered star formation rates in high redshift galaxy mergers to their low redshift counterparts, finding that the higher redshifts show lower levels of enhancement; in Xu et al. (2012) this is tied to the changing gas fraction within those galaxies. The aim of the work presented in this chapter is to begin to tackle this question of the SFR enhancement's dependence on gas fraction directly, using new observations of the gas fractions of a sample of local, interacting galaxy pairs with a wide variety of Δlog(SFR) values.

To determine the gas fraction of a sample of galaxies, the 21cm emission line that traces neutral hydrogen will be observed. Most radio telescopes are sensitive to this line; however, in order to correlate Δlog(SFR) with the gas fractions of each galaxy individually, high spatial resolution (i.e., significantly smaller than 30 arcseconds) will be required in order to distinguish emission from each galaxy in the pair individually. Single-dish telescopes such as the Arecibo Observatory (among others) do not have this level of spatial resolution; only an interferometric telescope will be able to obtain the necessary spatial resolution. The Jansky Very Large Array (VLA), based in New Mexico, emerges as the optimal observatory for these observations.

5.2 VLA telescope functionality

The VLA is a radio frequency interferometer, situated 50 miles outside of Socorro, New Mexico. It was constructed in 1973, with first light in 1976[2]. The telescope consists of 27 individual antennas in a Y configuration, each of which is 25 metres in diameter. These antennas are situated on mobile platforms, which allows the telescope to expand and contract over the course of the year. In 2002, the VLA began work on a large upgrade to the technology that had been in place for almost thirty years. The original analog cabling was replaced with fibre-optic cables, along with more sensitive receivers, and a new supercomputer correlator was installed. The last of the upgrades

[2] `http://www.vla.nrao.edu/genpub/overview/`

were completed in mid-2012. With the end of the upgrade process, the telescope was officially renamed the Karl Jansky Very Large Array (VLA) from its temporary name of the 'Expanded Very Large Array' (EVLA)[3].

5.2.1 Basics of Interferometry

Resolution at long wavelengths

The angular resolution of a telescope θ is defined both by the wavelength of light being observed, and the size of the aperture available to collect photons. This relation can be written as $\theta \propto \lambda/D$, where D is the diameter of the aperture. In order to maintain the same angular resolution, the size of the telescope must increase linearly with the wavelength of the light. In principle, continuing to increase the aperture size will continue to shrink the angular resolution. In practice, it often becomes increasingly (or prohibitively) difficult to build sufficiently precise mirrors at large sizes, as the weight of the mirror itself can cause deformations in the shape of the mirror.

For telescopes of all wavelengths, the surface accuracy of the mirror is determined by the wavelength of the light it intends to observe. The Hubble Space Telescope was built to have a surface accuracy of 1/70 λ, which puts the permitted roughness of the mirror in the nanometre range. A mirror of equal quality in the radio wavelength would have roughness on the scale of millimetres, but building a telescope of sufficient size to equal the angular resolution of Hubble in the radio regime would be impossible[4]. Generally, single dish telescopes at longer wavelengths are of the order of tens of metres across, and as a result have relatively poor angular resolution (of order several arcminutes).

The solution to the issue of poor angular resolution is to build a telescope that links a number of widely spaced antennas together, synthesizing a telescope of much larger aperture. This 'aperture synthesis' approach allows much better resolution to be reached, as the results of the synthesis constructs a telescope with a primary aperture several kilometres in size, an object which would be impractical to build to sufficient precision.

[3] `http://www.nrao.edu/pr/2012/jansky/`

[4] In order to keep the angular resolution of Hubble, a 2.4 metre telescope functioning at wavelengths of $\sim$ 500 nanometers, a radio telescope operating at 21 cm wavelengths (a factor of five orders of magnitude longer) would have to have an aperture of 1008 km.

Baselines

A 'baseline' in an interferometer is the term describing the separation between any two antennas in the array. This distance determines the sensitivity of the telescope to power in the sky at a certain scale. The shorter the baseline, the more sensitive the telescope is to large scale structure, and the less sensitive the telescope is to small scale structure. The critical parameter here is the 'fringe spacing', which determines the scale of power observable by the interferometer. If there is power on scales smaller than the fringe spacing, this power is effectively invisible to the telescope. Fringe spacings are inversely proportional to the distance between antennas. If the baseline distance doubles, the fringe spacing shrinks by half, which increases the telescope's sensitivity to smaller scale structure. The tradeoff is that with larger baselines, the telescope's sensitivity to large scale power decreases. The VLA has 27 antennas with 351 unique baselines, which range from very short spacings to widely spaced, allowing the telescope to be sensitive to power on a relatively wide range of scales.

Configurations

By enabling the telescope's 27 antennas to be moved, the telescope can morph into a series of different configurations. Each configuration is pertinent for different types of science, so the adaptability this introduces to the telescope makes it a much more broadly useful tool. The configurations of the VLA are labeled from A to D, with a few 'hybrid' arrays. Hybrid arrays have one arm of the Y in a different extension from the remaining 2 arms. A configuration is the most extended, with a maximum baseline of 36.4 km; this extension gives it the best small scale sensitivity, but the poorest sensitivity to larger structures. The configurations become gradually more compact until D configuration, which has a maximum separation of 1.03 km. D configuration therefore has the weakest sensitivity to small scale structure, but the best sensitivity for diffuse, large scale power.

The data presented in this chapter were taken in C configuration. The minimum separation between antennas in C configuration is 0.035 km and the maximum separation is 3.4 km. This results in a synthesized beam of 14 arcseconds, and a maximum large scale sensitivity of 970 arcseconds at the frequencies observed ($\sim$1.5 GHz). At the typical redshift of the sample ($z = 0.02$), the angular scale is 0.4 kpc/arcsecond; the beam thus has a physical diameter of 5.6 kpc at $z = 0.02$.

Designing an interferometer

Since interferometers technically take data in the Fourier domain, the goal of the interferometer is to provide as uniform and broad a coverage of the Fourier UV plane. To address the issues of broad, uniform coverage, there are two main considerations when designing the layout of an interferometric array. One is the physical orientation of the telescope with respect to the rotation of the earth, and the second the variation between baseline lengths.

To provide broad coverage, the rotation of the Earth during the observation can be used to provide curved tracks through UV space. If the observer were at the location of the target source, it would be possible to watch the Earth rotate, and the individual tracks of the baselines will appear to trace out a curve as the planet rotates and the apparent baseline length between antennas changes. The lower a source is in the sky, the more foreshortened the array of detectors will appear to the source. To avoid the worst of the foreshortening effect, most interferometric arrays are not built so that all the antennas are aligned east-west, but have baselines extending to the north and south, which will never line up and foreshorten severely as the source is rising or setting.

To provide uniform coverage, the aim is to avoid duplicating baselines. Since the length of the baseline determines a unique frequency of power that is observed at that baseline, it is advantageous to have as many baselines of differing length as possible, to avoid redundancy in the Fourier UV plane. This includes the distances between an antenna and its immediate neighbour, so many interferometric arrays are designed so that no two baselines will be the same length. For arrays with east-west alignments, it is still possible to avoid some redundancy by spacing the telescopes irregularly. This will not avoid the foreshortening problem, but it will help populate the UV plane more quickly.

The VLA has a Y shaped configuration, which is slightly rotated (by 5°) from being aligned due north, to avoid any perfectly East-West baselines from occurring between the arms. This allows the telescope to observe objects low on the horizon without the foreshortening problems of many east-west arrays. Each array element n (progressing from the centre) has increased its spacing by $n^{1.716}$ relative to the previous one, which means that the baseline separations between antennas are unique (Thompson, 1999).

Correlating the signal

To go from an array of individual antennas to one synthesized large telescope, the data from each individual antenna must be correlated with every other antenna in the array. This complex function is undertaken by the supercomputing correlator, and must take into account a number of considerations. Unlike a single dish telescope, where the wavefront of light hits the telescope mostly at a single point in time, the manipulation of the wavefront for an interferometer is much more complex.

As an electromagnetic wave hits the telescope's dish, it is reflected from the primary mirror into the secondary, and back down through a hole in the centre of the dish and into the appropriate feed horn for the frequency range of the project. The feed horn then feeds the astronomical signal into an amplifier, which amplifies the signal and then splits the data into sufficient portions to be correlated with every other antenna. This information is then fed along fibre-optic cables to the correlator.

However, due to the physical separation of the antennas, the wave front will arrive at different antennas at different times. The time delay between the moment the wave front from the source reaches one antenna and the moment it reaches the second antenna is called the geometric delay. This can be easily calculated using the physical separation of the antennas and the position of the source in the sky. This time delay must be removed from the data before correlation.

The correlator must also account for phase delay. Similar to the geometric delay, if the source is not immediately at the centre of the antenna's pointing, there will be an additional phase delay due to the additional time it will take the wave front to reach one side of the dish of the antenna relative to the other side of the dish. This is usually calculated based upon calibration of a source of known phase.

In addition to accounting for the above delays, the correlator also averages the correlated signal over a 1-5 second interval, depending on the project. It is the correlated, averaged, signal that is retrieved by the scientist, ready for calibration and imaging.

5.3 Sample Selection

A sample of galaxies which allows an observation of as broad a range of $\Delta\log(\mathrm{SFR})$ as possible in a set of close galaxy pairs should be selected. As detailed in Section 5.1, galaxies in pairs which are separated by less than 30 kpc h^{-1} are most likely to show

extremely high star formation rate enhancements, and galaxies with wider separations are much less likely to show strong SFR enhancements. In order to select a sample of galaxies which has the highest possible chance to be undergoing a strong starburst, galaxies in this small physical separation regime ($r_p <$ 30 kpc h^{-1}) are selected. As roughly equal mass interactions also tend to drive the strongest starbursts, the mass ratio of the interaction is limited to those less extreme than 4:1. These galaxies are required to have calculated stellar masses and star formation rates, though the galaxies are not required to have been part of the sample discussed in Chapter 4. In order to ensure that the star formation rates are reliable, at least one of the galaxies in each pair is required to be star forming as classified by Kauffmann et al. (2003). The sample presented in Chapter 4 is smaller than necessary for the project presented here, as the gas-phase metallicities required are not required for this work, and will only serve to unnecessarily limit the sample size.

In addition to the above requirements, an angular separation requirement is also imposed on the galaxy pairs sample. The sample should be separated widely enough on the sky that the two galaxies can be individually spatially resolved by C configuration. C configuration was preferable to B configuration, as the exposure times for individual galaxies jump from $\sim$ 30 minutes in C configuration to tens of hours for each pair if B configuration were to be used. Furthermore, the finer spatial resolution of the B configuration is not required for a simple estimation of a total gas mass. If the science goal had been to map the gas within the galaxy rather than simply get an estimate of the total volume of gas present within the galaxy, B configuration would have been more appropriate.

The typical angular separations of the total galaxy pairs sample with projected separations less than 30 kpc h^{-1} are between $10 - 20$ arcseconds. These separations are too close to be spatially resolved by the VLA's C configuration, so an additional constraint on the sky separation of these galaxies will be required. The C configuration of the EVLA has a beam size of 14 arcseconds. In order to be sure that the two galaxies in a given pair would be cleanly spatially resolved, the criterion that the galaxies must be separated by more than 35 arcseconds on the sky is imposed, so that the galaxies are separated by at least 2.5 times the size of the beam (see Figure 5.4). In addition to the spatial resolution, the galaxies will also be resolved in velocity space, as the majority of them are separated by several tens of kilometres per second, which is significantly larger than the velocity resolution of the VLA of $\sim$3.5 km s^{-1}.

These criteria result in a sample of 17 galaxy pairs. The combination of the

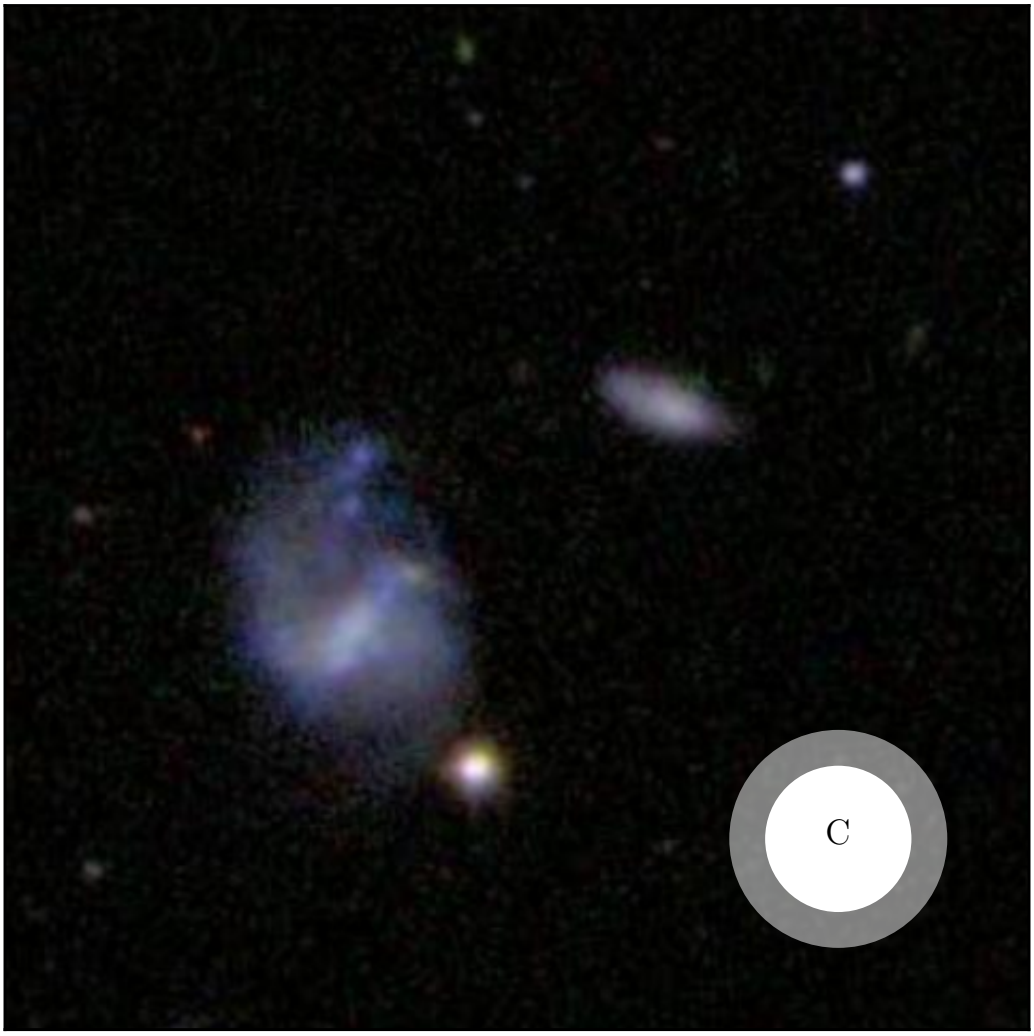


Figure 5.4: Size of the VLA beam for C configuration compared with a typical galaxy pair in the observed sample (588017702411763744 & 588017702411763872). The solid inner circle indicates the raw beam size; the transparent outer circle indicates the size of the beam after natural weighting, which increases the size of the beam by a factor of 1.5. Image is originally 400×400 px at a scale of 0.25 arcseconds/pixel, or 100 arcseconds to a side.

35″ limit and the 30 kpc h^{-1} projected separation limit result in a relatively low redshift sample ($z < 0.04$) with separations of 15 - 30 kpc h^{-1} and Δv between 15 and ∼300 km s^{-1}. Four of the galaxy pairs are shown in Figure 5.6. All images, both in Figure 5.6 and in the Appendix, are kept at the same angular resolution for consistency. With a sample of 17 galaxy pairs, gas fractions will be determinable for 34 individual galaxies in the sample, while maximizing the likelihood for strong Δlog(SFR) enhancements in these galaxy pairs. This sample of 34 galaxies is a representative sampling of the overall distribution of Δlog(SFR) shown in Figure 5.1; Figure 5.5 shows the distribution of Δlog(SFR) for the VLA sample. Approximately half of the sample show SFRs that are enhanced by at least a factor of two, with the highest enhancement at a factor of 20 over the control, and the lowest a factor of four lower than the control.

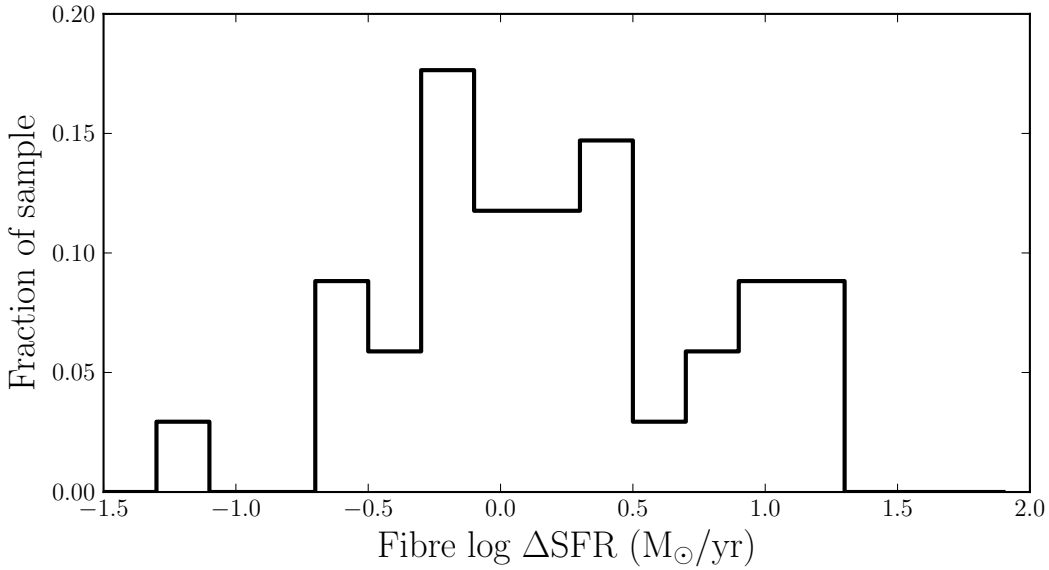


Figure 5.5: Distribution of the fibre Δlog(SFR) values for the sample of galaxies selected for VLA observations.

This sample of galaxy pairs was checked against the archive of publicly available VLA data, and there was no overlap; all data obtained for this project were new to the science archive.

Exposure times

Exposure times were calculated based upon a gas mass sensitivity limit. Using a limiting gas mass of 10% of the higher mass galaxy, the appropriate exposure time

for each pair was determined. The 10% limit was chosen as galaxies below this limit are very gas poor and generally fall along the red sequence, indicating that they are not likely to be strongly star forming (Kannappan, 2004). The time estimates (described below) assume that the entire HI gas content of the galaxy will be present on angular scales smaller than a single beam. This is a reasonable assumption, as all of the galaxies in the sample have optical half-light radii that are smaller than the beam for C configuration. The half-light radii range from 10% to 93% of the beam. All but two galaxies are less than 55% of the beam size, with a median value of 35%. Unless the HI disk of the galaxy is dramatically larger than the optical half-light size of the galaxy, the majority (if not all) of the HI mass should be contained within a single beam.

In some previous studies of the HI content of individual interacting galaxies, it has been found that tidal tails in the HI can cover a very wide region of space (e.g., Smith, 1991; Hibbard & Yun, 1999). Generally, these HI tails are spread over both a spatial region and velocity range. If the HI tidal features exist primarily in the velocity regime, and do not span a wide range of spatial coordinates, this flux ought to be recoverable by increasing the size of the velocity bins. If HI tidal features are limited in velocity and diffusely scattered in their spatial component, it is possible that some fraction of the total HI of the system may be lost to the noise of the data cube. However, if a significant fraction of the gas is removed in tidal features, this may still be detected, but spatially offset from the optical counterpart.

From the base calculation of the observing time required to be sensitive to a 10% mass limit at 3σ, each time estimate is rounded up to a minimum of 30 minutes exposure, as 30 minutes is the minimum scheduling block allotted by the VLA. Required overheads are approximately 50% of a 30 minute scheduling block, but the fraction of time required by overheads drops as the length of observation increases. For exposure times that cannot be fit into a single 30 minute scheduling block, a 25% overhead is assumed.

Mathematically, the sensitivity limit can be worked out using the following steps. Given the HI mass to which sensitivity is desired, the target noise level in the data can be determined by working backwards from the following equation relating total flux (in Jy/km s^{-1}) to the HI mass:

$$M_{HI} = 2.36 \times 10^5 D^2 \int S(v) dv \tag{5.2}$$

M_{HI} is the HI mass in units of solar masses ($M_\odot$), D is the distance to the galaxy in Mpc, and the integral of $\int S(v)dv$ is the flux in the emission line in units of Jy km s^{-1}. For galaxies at low redshifts, D may be approximated as $D = cz/H_0$, where c is the speed of light in km s^{-1}, z is the redshift of the galaxy, and H_0 is the Hubble constant (assumed throughout this thesis to be 70 km s^{-1} Mpc^{-1}). This approximation is only valid for local galaxies where the recession velocity v is well matched by cz.

Once rearranged, this equation allows a determination of the desired flux limit, $S(v)dv$, for each individual galaxy, given its redshift. As the signal is spread over a $14'' \times 14''$ beam, so to convert $S(v)dv$ into Jy km s^{-1} beam^{-1} from Jy km s^{-1}, the flux is divided by the area of the beam, which can be assumed to be circular.

$$\frac{M_{HI}}{2.36 \times 10^5 D^2 \times 14'' \times 14''} = \int S(v)dv \tag{5.3}$$

Since 3 sigma detections are the goal, this integrated flux must be divided by three to arrive at the target RMS noise level for the data set. In addition, this flux should be spread out over 3 channels, so the resultant RMS value must be divided by an additional three to arrive at the final target RMS. With the addition of the velocity width of the channels and the sky frequency of the galaxy, this data can be input into the exposure time calculator available on the NRAO website to determine the time required to reach this 3σ limit.

As recommended on the NRAO observation preparation website, accounting for the additional time required by overhead observations is a 25% increase in the observing time output by the exposure time calculator. The final time request is then this overhead-included time estimate, rounded up to a half hour interval, the smallest time allotment interval granted by the VLA.

5.4 Data Acquisition

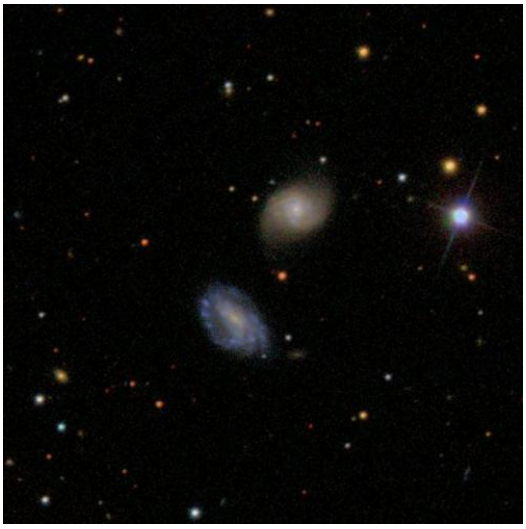
For the 17 galaxy pairs (34 galaxies) for which C array observations with the VLA were possible, the data was acquired in two separate proposals. The two successful proposal IDs are 12A-061, and a Director's Discretionary Time proposal, 13A-537. For all galaxies in the sample, the corresponding proposal ID is listed in Table 5.1. All observations were undertaken in L–band; this band's frequency range (1–2 GHz) brackets the frequency of the HI 21 cm spectral line, which has a rest frequency of 1.420 GHz.

Table 5.1: SDSS objids, redshifts, RA & Declination, and proposal ID for all galaxy pairs in the sample: 34 galaxies in 17 pairs.

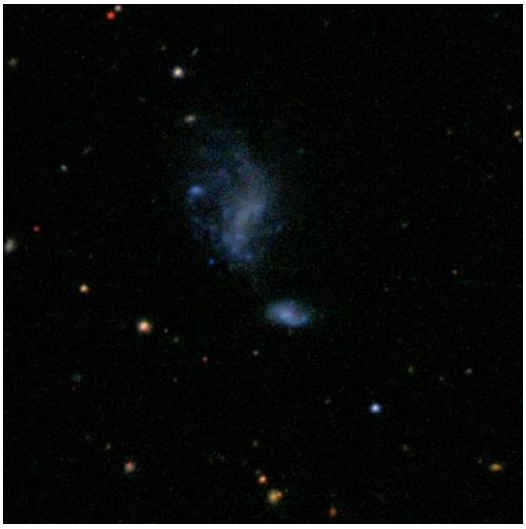
SDSS ObjID	Redshift (z)	RA	Dec	Proposal ID
587726033341776175	0.027775	14:58:21.64	+02:32:48.94	12A-061
587726033341776191	0.027899	14:58:23.24	+02:32:18.82	
587739609695453284	0.027482	11:18:11.21	+31:59:16.32	12A-061
587739609695453281	0.02726	11:18:10.97	+31:58:35.93	
587744873717563559	0.019347	09:01:30.48	+12:39:43.63	12A-061
587744873717563471	0.019761	09:01:33.86	+12:39:31.67	
587727179536859247	0.018063	02:06:21.33	−08:52:19.26	12A-061
587727179536859227	0.018439	02:06:20.21	−08:52:57.50	
587729160043757697	0.023771	13:32:02.01	+05:32:12.73	12A-061
587729160043757707	0.023608	13:32:04.62	+05:32:37.14	
587741489815027774	0.03761	08:05:53.02	+14:00:06.34	12A-061
587741489815028146	0.037924	08:05:55.42	+13:59:59.04	
587739303684866183	0.017681	13:12:07.51	+34:16:13.78	12A-061
587739303684866173	0.016686	13:12:05.26	+34:15:23.16	
587726033308680234	0.031671	09:55:22.79	+02:34:14.25	12A-061
587726033308680320	0.031426	09:55:23.34	+02:34:58.81	
588018056204780081	0.03083	17:05:01.61	+23:09:26.43	12A-061
588018056204780049	0.030114	17:04:59.9	+23:10:08.57	
587729158970867777	0.023125	13:39:27.31	+04:30:38.03	12A-061
587729158970867792	0.022964	13:39:29.71	+04:30:40.95	
587733605328093368	0.026354	13:59:27.75	+58:19:33.72	12A-061
587733605328093256	0.025908	13:59:31.93	+58:18:52.55	
588848899908370674	0.025371	11:20:57.36	−00:05:49.48	12A-061
588848899908370505	0.025317	11:20:57.06	−00:05:03.73	
588017605758025795	0.019123	12:18:38.72	+45:46:28.13	13A-537
588017605758025732	0.019291	12:18:39.14	+45:47:06.90	
588017702411763744	0.029248	14:48:19.69	+09:07:02.14	13A-537
588017702411763872	0.029015	14:48:17.59	+09:07:23.48	
587727178473930875	0.033197	03:45:43.95	−06:54:46.02	13A-537
587727178473930886	0.032945	03:45:46.7	−06:54:56.24	
587742901789589569	0.025918	12:39:35.85	+16:35:16.14	13A-537
587742901789589575	0.026322	12:39:38.35	+16:35:06.34	
588023670245949622	0.021097	12:20:37.91	+20:40:21.19	13A-537
588023670245949625	0.021278	12:20:39.01	+20:39:38.69	

Table 5.2: SDSS objids, projected separations, velocity differences, mass ratios, and stellar masses for all 17 galaxy pairs in the final sample.

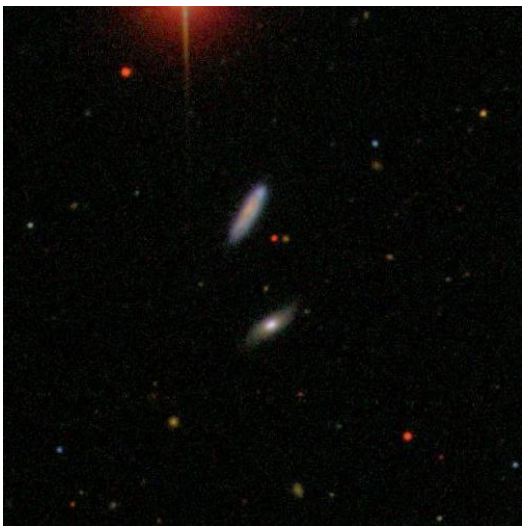
SDSS ObjID	r_p (kpc h^{-1})	Δv (km s^{-1})	Mass ratio	log M_* ($M_\odot$)
587726033341776175	21.51	36	3.6550	9.6463
587726033341776191				10.2092
587739609695453284	22.27	65	2.5073	9.7674
587739609695453281				10.1666
587744873717563559	20.18	122	4.2567	9.1412
587744873717563471				8.5121
587727179536859247	15.45	111	3.6954	8.8972
587727179536859227				8.32951
587729160043757697	22.03	48	2.8969	8.6136
587729160043757707				9.0755
587741489815027774	26.78	91	1.1217	9.8316
587741489815028146				9.7817
587739303684866183	20.2	293	3.0329	8.6097
587739303684866173				9.0916
587726033308680234	28.57	71	1.2667	9.6466
587726033308680320				9.7492
588018056204780081	29.46	208	4.6377	9.8134
588018056204780049				10.4797
587729158970867777	16.77	47	1.3052	9.8747
587729158970867792				9.7590
587733605328093368	27.7	130	1.1507	9.1776
587733605328093256				9.1166
588848899908370674	23.46	16	2.2012	9.1912
588848899908370505				9.5338
588017605758025795	15.2	49	1.0854	8.4618
588017605758025732				8.4974
588017702411763744	22.04	68	3.6570	9.63962
588017702411763872				9.0765
587727178473930875	27.87	73	1.2448	8.9420
587727178473930886				8.8469
587742901789589569	19.55	118	3.5917	9.7253
587742901789589575				9.1700
588023670245949622	19.39	53	1.6263	8.3517
588023670245949625				8.5629



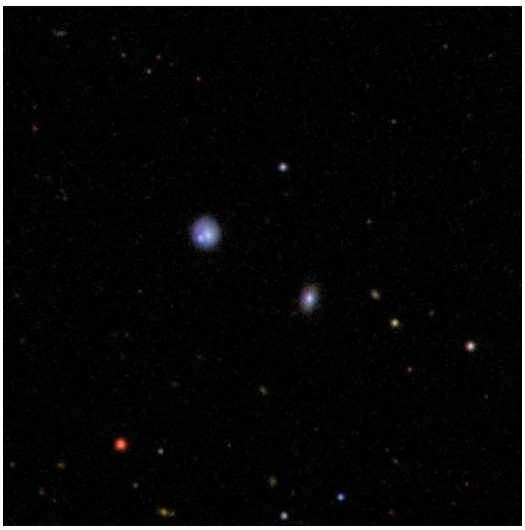
(a) 588018056204780081
& 588018056204780049
(bottom left and top right, respectively).



(b) 587727179536859247
& 587727179536859227
(top left and bottom right, respectively)



(c) 587726033308680234
& 587726033308680320
(bottom and top, respectively)



(d) 587729160043757697
& 587729160043757707
(lower right and upper left, respectively)

Figure 5.6: Four of the 17 galaxy pairs in the sample. Images of the full sample are presented in the Appendix. All images were originally 500×500 pixels across, imaged at $0.5''$/pixel (i.e., 250 arcseconds across). For consistency, all images are at the same angular scale.

5.4.1 Proposal 12A-061

The first proposal was submitted during the normal cycle on August 1st, 2011, and awarded a combination of Priority A, Priority B, and Priority C time. Priority A time is almost guaranteed to be observed, barring any unforeseen circumstances. Priority B time is observed on a 'best-effort' basis, and Priority C time is scheduled as filler between larger projects, and not guaranteed to be observed. Time was awarded for 12 out of the 17 pairs, with the remaining five listed as 'not scheduleable' due to scheduling pressure at those RAs. Part of this pressure was due to continued commissioning of new instrumentation on the VLA; the telescope was still operating in Open Shared Risk Observing (ORSO) mode at this time. Of the 12 galaxies which were awarded time at any priority, all were observed, for a total observing time of 7.5 hours.

Observational setup: Instrument configuration

L-band observations were obtained of these 12 galaxy pairs, using dual polarization data in the Open Shared Risk Observing (OSRO) configuration. This configuration was later retired after Open Shared Risk operations completed.

Dual polarization is recommended by the NRAO[5] as the optimal configuration for science targets which are not expected to be strongly polarized. The wavefront of light detected by the antenna is passed through a circular polarizer; depending on the number of polarization products requested, a varying number of the polarization products are kept. The two polarization products are labeled as R and L. In single polarization mode, either the correlation between R products (RR) or L products (LL) is kept. Dual polarization keeps both RR and LL correlations between polarization products, and will help to decrease the noise level of the resultant cube, as the two correlations function as relatively independent data products in the case where the emission is unpolarized. For all of the observations presented here, data is taken in dual polarization mode.

In each observing session, the correlator integration time was five seconds long. For each galaxy, a frequency span of 16 MHz was needed in order to be sure that both galaxies in the pair, which could be separated by up to 300 km s^{-1}, were both fully captured in velocity space. With a typical rotational velocity of $\sim$ 200 km s^{-1} for

[5]See here: https://science.nrao.edu/facilities/vla/docs/propfaq/polarization-products-and-noise

both galaxies, and the potential for additional tidal features, the velocity range needed simply to capture emission from the pair is well over 1000 km s^{-1}. In order to be sure that there was ample space on the high and low velocity end of the observation to have a stable emission free region, this velocity range was expanded. Given that 1 MHz at 21cm[6] is approximately 210 km s^{-1}, 16 MHz corresponds to $\sim$ 3360 km s^{-1}. A frequency resolution of 15.6 kHz (or $\sim$ 3.5 km s^{-1} at $z \sim 0.02$) was used to be sure to have the velocity resolution to resolve the emission from the galaxy pairs, where the lowest velocity difference was $\sim$ 15 km s^{-1}. Under OSRO observing modes, such a frequency span with the desired frequency resolution was only possible by using 8 adjacent 2 MHz wide subbands (or spectral windows), each with 128 spectral channels and dual polarization. The centre frequency of this set of 8 subbands was determined by inputting the sky frequency of the HI line for each galaxy, given the redshift of that galaxy.

Each 2 MHz subband, however, has its own characteristic sensitivity pattern that results in noisier channels at their edges. To counter this, another set of eight subbands, 2 MHz each, was invoked through a second data path in the VLA system. This extra set of subbands were places 1 MHz off in frequency (i.e., half a subband width) with respect to the first set of subbands. With the staggering of the two sets of subbands, any frequency observed in a low signal to noise section of one set ought to be observed at high signal to noise in the other. The resulting total frequency coverage from this setup was 17 MHz.

5.4.2 Director's Discretionary Time: 13A-537

To complete the sample, a Director's Discretionary Time Exploratory Proposal was submitted on June 28th, 2013 for additional time with the then-current C configuration VLA. Priority C time was awarded for all five of the targets, summing to a total of 6.5 hours of L–band observations. Nearly all of the requested scheduling blocks were observed; one galaxy pair was only observed for three out of five requested half hour scheduling blocks, resulting in shallower than anticipated data on that galaxy pair.

[6]As estimated here: `https://science.nrao.edu/facilities/vla/docs/manuals/obsguide/modes/line/velocity`

Observational setup: Instrument configuration

In order to simplify the data reduction process, the telescope setup was changed relative to the 12A-061 observations. Where the previous observations had used two sets of eight subbands offset with respect to each other by half a subband width in frequency space, in observations for 13A-537 this complication was removed. Instead, a single subband or spectral window was used, placing the frequency of the galaxy near the centre of the spectral window.

This meant that the second set of observations was half as data intensive. Only one subband was used, 16 MHz wide with 2048 channels to deliver a spectral resolution identical to that in the 12A-061 observations (i.e., 15.6 kHz). The centring of the subband was determined automatically by the online observation preparation tool, based upon the HI line's rest frequency and the redshift of the source. The correlator integration times were kept consistent between the two sets of observations, with five second integration times for both.

Since low priority time was being sought, in order to maximize the chances that the data would be taken, the observations were split into 30 minute scheduling blocks so that they could more easily be scheduled between higher priority projects. This increased the overhead associated with each science target as a whole, since each scheduling block had to observe all the required calibration sources independently, decreasing the amount of time able to be spent on the science target. This meant that the integrated time spent on the science target is less for the set of three 30 minute blocks for pair 588017605758025795 & 588017605758025732 (as detailed in Table 5.3) than was spent on the science target of the pair 587727179536859247 & 587727179536859227, which had a contiguous 1h30m scheduling block.

5.5 Observational structure: Scheduling Blocks

The structure of each scheduling block is consistent across all observations, as low-frequency observations all require the same set of calibrators to be observed in the same order. The basic observing strategy is summarized as follows:

- Setup scan
- Flux density calibrator/Bandpass calibrator
- Complex gain calibrator

Table 5.3: Record of data taken for both 12A-061 and 13A-537 data.

SDSS ObjID	.ms name	Date Observed (UTC)	Duration
587726033341776175	402775.ms	Feb 06 2012 14:35:39	30m
587726033341776191			
587739609695453284	37037.ms	Feb 22 2012 11:03:28	30m
587739609695453281			
587744873717563559	84028.ms	Feb 20 2012 02:12:44	1h
587744873717563471			
587727179536859247	25232.ms	Mar 17 2012 23:57:42	1h30m
587727179536859227			
587729160043757697	88885.ms	Feb 07 2012 13:02:15	30m
587729160043757707			
587741489815027774	78236.ms	Feb 06 2012 03:37:43	30m
587741489815028146			
587739303684866183	04398.ms	Feb 07 2012 12:32:23	30m
587739303684866173			
587726033308680234	62963.ms	Feb 09 2012 10:54:42	30m
587726033308680320			
588018056204780081	14352.ms	Feb 05 2012 17:09:28	30m
588018056204780049			
587729158970867777	31481.ms	Feb 07 2012 13:32:15	30m
587729158970867792			
587733605328093368	21528.ms	Feb 06 2012 14:06:02	30m
587733605328093256			
588848899908370674	62037.ms	Feb 10 2012 10:50:47	30m
588848899908370505			
588017605758025795	04167.ms	Aug 6 2013 00:11:35	30m
588017605758025732	71528.ms	Aug 09 2013 23:55:51	30m
	91435.ms	Aug 10 2013 01:25:38	30m
588017702411763744	38889.ms	Jul 26 2013 06:38:53	30m
588017702411763872			
587727178473930875	55787.ms	Aug 18 2013 12:37:22	30m
587727178473930886	84375.ms	Aug 19 2013 13:48:09	30m
	89354.ms	Aug 27 2013 08:17:19	30m
587742901789589569	80555.ms	Aug 13 2013 23:40:12	30m
587742901789589575	89352.ms		
588023670245949622	79398.ms	Aug 10 2013 03:11:20	30m
588023670245949625	19676.ms	Aug 13 2013 23:44:03	30m
	19907.ms	Sep 02 2013 00:43:11	30m

- Science target
- Complex gain calibrator

The setup scan is required in order to set the attenuator levels[7] for each antenna, given the specific frequency of the observations to follow. It is usually only 1 minute in duration, and should be flagged out of the data set automatically by the pipeline.

The absolute flux density calibrator, which also functions as a bandpass calibrator, is one of five precisely known calibration sources on the sky. As there are only five of them in the entire sky, only some fraction of them will be available for observation. The nearest flux density calibrator to the science target is usually chosen to minimize telescope slewing time. These five sources are relatively bright targets at the frequencies of the current observations, and allow for the determination of the spectral response in each subband. The shape of the instrumental response over a range in frequency is known as the 'bandpass'. Since the behaviour of the flux density of these targets is very well modelled across a wide range of frequencies, the flux density calibrator sources can not only be used to set the absolute flux density scales, but also to calibrate the bandpass shape.

If a particular antenna's bandpass and flux density levels cannot be calibrated through the recorded bandpass calibration (e.g., if the antenna was offline for some reason during observation of the bandpass calibrator), the data from that antenna cannot be used for the remainder of the observation block. The combination of the length of time on the setup scan and the flux density calibrator/bandpass calibration scan, should equal about ten minutes to allow enough time for the initial slew of the antennas (for the current observations, this calibrator is always observed at the start of the observing session) and on-source time on the calibrator.

For L-band observations in C configuration of the VLA, the complex gain calibrator must be observed both before and after the science target, and at 30-40 minute intervals during science target observation, if more than 30 minutes is spent on the science target. Since most of the data in this Chapter was taken using 30 minute scheduling blocks, in most cases, the full observation plan resembled the list above. For longer observations, a third complex gain calibration step was inserted between science target scans. Generally, integration times for the complex gain calibrator are of order 2 minutes. The complex gain calibrator's purpose is to account both

[7]The attenuation levels set help to modulate the amplification of the signal received by the antenna, and are partially determined by the gain of the antenna at a given frequency.

for the effect of the atmosphere on the phase of the incoming wavefront. Since the flux/bandpass calibrator is usually quite distant from the location of the target source, the phase of the wavefront coming through the atmosphere for the science target is usually somewhat different between the science target location and the bandpass calibrator location. The complex gain calibrator is therefore chosen to be the closest, reasonably bright calibration source to the target source. The proximity of the complex gain calibrator to the science pointing means that both the calibrator and the science target will be observed through similar atmospheric conditions. Generally, for low frequency observations, the atmosphere does not undergo dramatic changes over the course of the observations; bracketing the observations allows for an interpolation between the two calibration solutions. In principle, these solutions should be very closely matched. At high frequencies, the phases are much more variable, as the atmosphere is increasingly turbulent at higher frequencies, and more frequent observations of the complex gain calibrator are required.

The remaining time in the scheduling block can then be dedicated to the science target of interest. For a 30 minute scheduling block, the remaining time is usually $\sim$ 15 minutes, meaning 50% of all observing time is lost to overheads. It can often be wise to divide up longer periods of time spent on a single target into shorter, $\sim$ 5 minute scans. The shorter scans serve as a safeguard against catastrophic data loss. If a particular scan is not delivered properly by the correlator, that scan will have to be discarded; if the data that is lost is a critical calibrator source, the entire data set may not be able to be properly calibrated. Any scans longer than 10 minutes in length are discouraged in order to maintain reasonable file sizes for each scan.

5.6 Data reduction steps

Once the scheduling block has been planned, submitted, and observed, the resulting data can be obtained from the NRAO Science Archive, where it is stored as proprietary data for a period of one year. Once the data is downloaded, data reduction can begin. Unless otherwise noted, all data reduction was performed using version 4.1.0 of the Common Astronomy Software Applications (CASA) software, provided by NRAO.

Before beginning any calibration steps, several key reference files must be either obtained or produced. The first is the log file produced by the telescope operator, which states which antennas were in operation, or if anything failed during obser-

vation. Listed in the log file are any antennas which are missing receivers, had abnormally low amplitudes, or data corruption problems. The file produced by the task LISTOBS is also required, which is a summary of the observation. It lists which scans of the telescope correspond to which source, which antennas took part in the observations and their location in the array, and the RA and DEC of the pointings.

5.6.1 Flagging

As a first pass, the data should be visualized as amplitude vs frequency for a wide range of frequencies. For 12A-061, where 16 spectral windows were contained in the data, the data were visualized in two sets: 0–7, followed by 8–15. For 13A-537, only one spectral window existed in the data, so this extra iteration was not required. Iterating through each antenna's set of baselines to all other antennas allows for a determination of whether any antennas are malfunctioning. For certain antennas in the array, common problems included low to no amplitude in one correlation (RR or LL), or a missing receiver. These malfunctioning antennas are not contributing valid data to the observation, and should be flagged as bad data, thereby removing them from the data set. Due to the way that CASA handles the data, in order to create an image from any given frequency and time visibility, both RR and LL correlations are needed. If an antenna has poor behaviour in one correlation, the entire antenna is therefore flagged.

At this stage it is also possible to find strong interference signals, if it exists, in any of the scans. Interference is discussed in further detail in §5.6.2. If the interference is sufficiently strong, it will alter the scale of the data plotting subroutine to such an extent that the unaffected data cannot be properly viewed. The channels or time ranges of this interference should be flagged before proceeding on with other calibrations. More subtle data issues are often discovered later at the time of applying the bandpass calibration to the data, and the troublesome sections of the data will be identified and flagged through the use of the FLAGDATA function.

5.6.2 Data problems

Radio Frequency Interference

The most common issue with low frequency radio data is interference from other (non-astronomical) sources of radio waves, known as Radio Frequency Interference (RFI).

Many of these extra sources of RFI are communications satellites in orbit around the earth; another common source is radar interference from the ground. If one of these satellites becomes visible to the telescope, and is emitting an interfering signal in the same frequency range as the telescope is observing, a very bright, noisy signal can appear in the data.

RFI can manifest itself in a number of different ways. RFI can either be limited to a small number of channels (in some cases a single channel only may be affected), or be present across a wide swath of frequencies. It is also possible for the RFI source to be observed as 'continuous' interference, which is present across the entire length of time of an observation, or 'intermittent'. Intermittent RFI may only be present in the data across a single discrete period of time, or it may be present at multiple time steps across the data.

The NRAO provides a list of known RFI sources in the L–band sky at the following URL: `https://science.nrao.edu/facilities/vla/observing/RFI/L-Band`. For instance, between frequencies of 1376-1386 MHz, there is intermittent interference from an object listed as 'GPS L3 NUDET'. This signal results from communication between nuclear detonation detection software on board a GPS satellite and the ground. For observations between this range of frequencies, there is the possibility that this communication will interfere with and overwhelm the desired astronomical source data. Where RFI is present in the data, all affected times and frequencies must be flagged from the data and removed.

Mystery interference

For two of the data sets in 12A-061, and all of the data sets in 13A-537, a characteristic issue with the raw data was present that was never understood.

For specific baselines, the raw data showed regularly spaced diagonal stripes in amplitude when the data cube was plotted in frequency vs time projection, iterating through baselines. Both the width of the diagonal stripe and the angle of the stripes relative to the axes varied from baseline to baseline. This effect was strongest at the smallest separations; all baselines connecting the most densely packed antennas were usually strongly affected. This pattern was present in both correlations of the recorded data, and was often most apparent in the flux density calibrator scans, although it was generally present across scans of the complex gain calibrator and the science target as well.

Figure 5.7 illustrates the nature of the problematic data, and shows one of the two data sets from 12A-061 affected by the patterning of the data. This data is shown as visualized by the Astronomical Image Processing System (AIPS) software, which can display all 16 spectral windows for a given baseline (in one correlation) in one viewer window. Figure 5.8 shows a more characteristic example of the data failure. This figure shows the normal bandpass sensitivity pattern of weaker sensitivity at early and late channels (the edges of the spectral window in frequency space), but superimposed on top are a series of regular stripes. Figure 5.9 shows an even noisier symptom; on top of the stripes, it appears that this data set was affected by severe, chaotic RFI. In all cases, this effect was limited to individual baselines that were preferentially concentrated at the smallest separations. Longer baselines from the same antennas appeared unaffected by this issue.

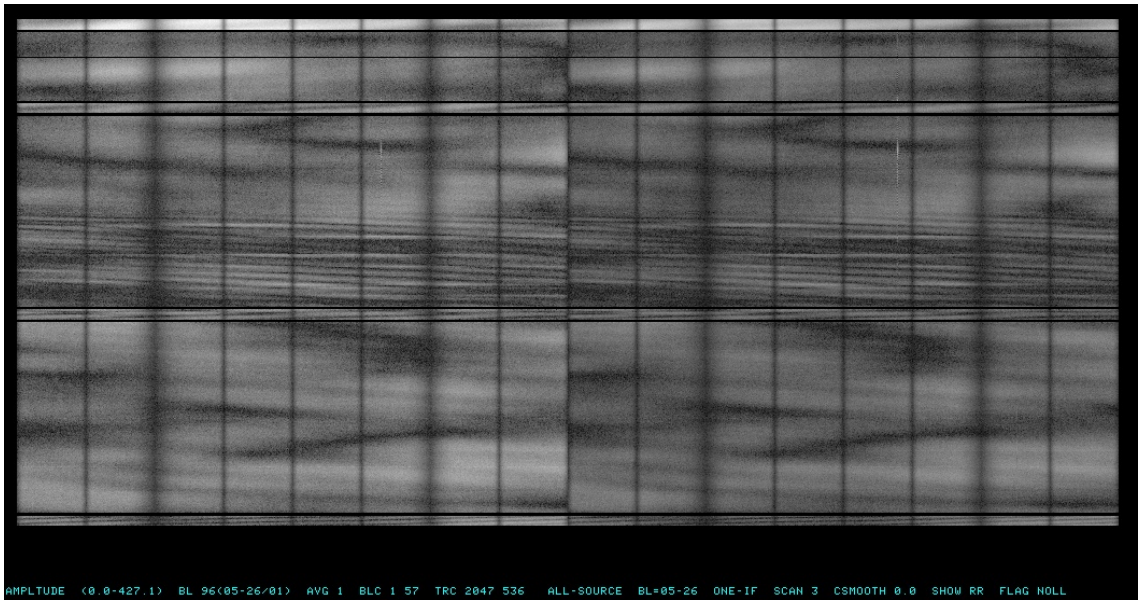


Figure 5.7: Channel vs. time for the ea05&ea26 baseline on one polarization within the data set for 587727179536859247 & 587727179536859227 (25232.ms). In addition to the large scale structure present in the data, RFI is present as a vertical line in spectral windows 5 and 12, which correspond to the same frequency. This data was imaged in AIPS, and shows the complex gain calibrator and science target scans, omitting the flux density calibrator due to its higher amplitudes.

Initial troubleshooting indicated that this pattern in the data could not be due to interference from the sun, as the sun was always at least 5 or 6 hours away from the flux density calibrator on the sky. The data themselves were taken at different

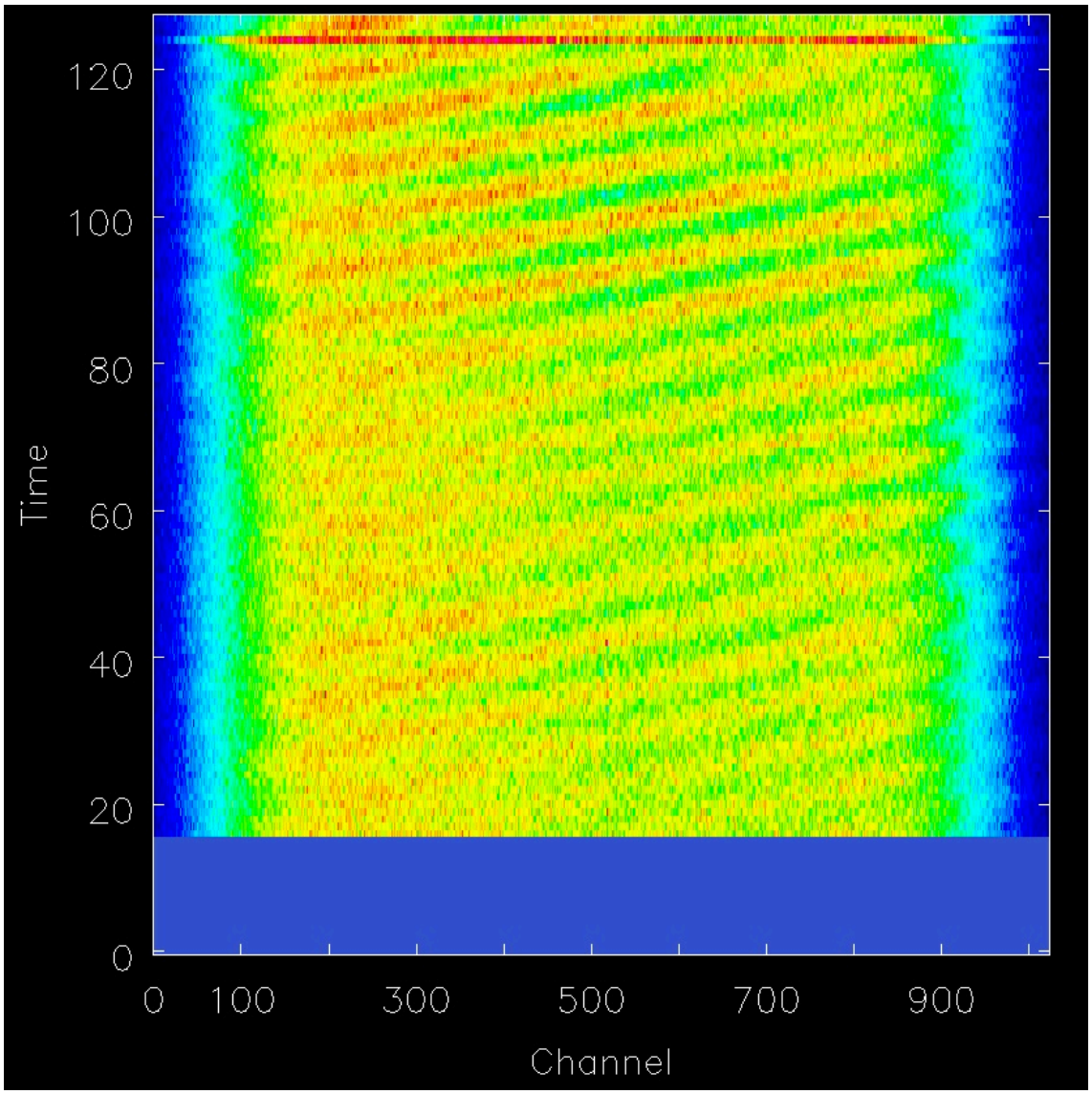


Figure 5.8: A representative baseline displaying the issue with the data. Regular stripes appeared in the data in time vs. frequency space. This data is from 19676.ms, for the pair 588023670245949622 & 588023670245949625, antenna baseline 12-14. Visible at the top is a 5 second integration interval affected by RFI across all frequencies in the data cube. The solid blue section at the bottom of the figure represents data that has already been flagged.

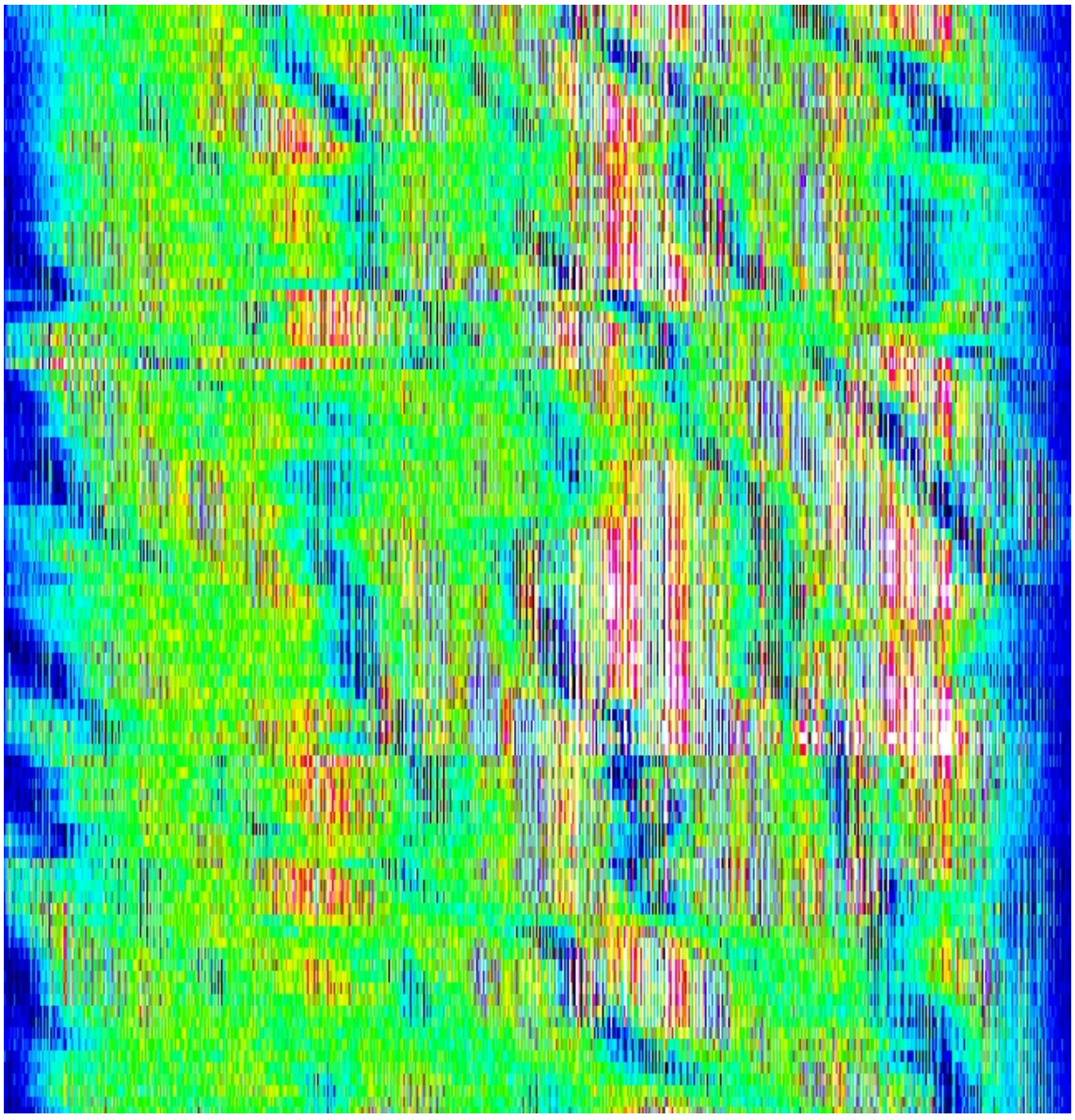


Figure 5.9: This particular data set (71528.ms) from the DDT appeared to have a more chaotic RFI signal imposed on top of the base issue of diagonal stripes. 91435.ms showed the same behaviour. Both of these data sets were targeting 588017605758025795 & 588017605758025732.

times of day and night, and no correlation appeared between the hour of observations and the strengths of the pattern. The potential influence of solar flares was also investigated, but the data did not appear to be taken during the period of time after a strong solar flare when the atmosphere would be expected to be affected by a burst of charged particles.

With sufficiently careful flagging, the data taken in 13A-537 was successfully calibrated. One observation was sufficiently strongly affected by this issue that re-observation was necessary, as calibration failed to complete. Although the re-observation was still affected by the patterned data, calibration was successful after careful flagging of the data. For the affected data in 12A-061, calibration using the CASA data reduction software failed to produce phase calibration solutions even after careful flagging, and it was necessary to calibrate the data using AIPS instead of CASA. The calibration steps for these two data sets (25232.ms and 78236.ms) were completed by Emmanuel Momjian. Once calibration was completed, the data was converted back into a CASA readable format, and imaging was completed using the CASA suite.

5.6.3 Antenna position corrections

Since the exact location of the antennas in the array is critical to correctly accounting for the geometric delay of the arriving wave front (as described in §5.2.1), the first step of calibration is to check if the positions of the antennas have been updated to a more accurate value since the data were taken. To test this, the task GENCAL is run in 'antpos' mode. If the program quits with a severe error, no updates to the recorded position of the antennas have been made. If, however, the antenna positions were updated, then GENCAL will output a calibration file containing the adjustments to the antenna positions. This output calibration file (if it exists) should be used with in all following calibration steps.

5.6.4 SetJy

The task SETJY determines the absolute flux density of the flux density calibrator used in the observations. There are only five flux density calibrators that can be used, and their absolute magnitudes and slopes as a function of frequency are known precisely. SETJY takes the model data for the calibrator used in the observations and appends the flux densities to the spectral window (or windows). The model values can

be visualized through PLOTMS by plotting the model data column against frequency. For the flux density calibrators used in these observations, the results of the task should be a straight line smoothly transitioning across all frequencies.

5.6.5 Delay calibration

Since the wave front will reach different antennas at different times, due to the physical separation between antennas, there is an instrumental delay built into the telescope to account for the difference in arrival times. However, this instrumental delay is not always perfectly accurate, so a delay calibration is necessary to account for the differences in arrival times.

This calibration uses the GAINCAL task. If any antenna positions have been recorded, the position alterations are also input into the delay calibration. The results of the delay calibration are stored in a file, and used for all further calibrations, to account for the differences in delay time. Typically, delay calibration corrections are on the order of nanoseconds.

5.6.6 Bandpass calibration

Bandpass calibration simultaneously solves for the amplitude and the phase using the bandpass calibrator, and accounts for the shape of the instrumental response as a function of frequency. Since the data is divided up by the correlator in several different ways, some channels may have a steeper drop in sensitivity than others, depending on how the data were sliced. Bandpass calibration accounts for this shape and removes it, and produces a table which is applied to every other source within a given observation. This table is effectively a clean description of the signal response as a function of frequency, which should be constant as a function of time, and therefore a valid description of the signal response for the remaining time of the observation.

Applying the calibration

After a calibration solution has been found for the data, the calibration solutions themselves are first checked to ensure that there are no large-scale errors in the solution. In some cases, the shape of the bandpass can be altered by extremely strong RFI that must be flagged before the calibration can successfully complete. If something abnormal is noticed in the calibration solution, it is usually most straightforward to

visualize the data either in PLOTMS or with the VIEWER tasks. Very often, lower-level RFI may be present in the data that was not picked up on at the beginning of data reduction. If further flagging is required, the calibration must be re-run and re-checked to make sure that the major issues have been eliminated.

Once the calibration solutions have been verified, the calibration can be applied to the data directly, which should result in fully calibrated bandpass data. This data should be concentrated around the true flux density of the calibrator, with no spikes, or unusual behaviour in certain baselines. In most cases, additional, subtle data errors are revealed. Low-level RFI, or specific channels that behave in unexpected ways should be flagged out of the data and removed. Any additional flagging indicates that the calibration should be re-run. These steps are repeated until the calibrated bandpass data no longer appears to be poorly behaved. Once the bandpass calibration is complete, a similar process will be implemented for the next calibration source, the complex gain calibrator.

5.6.7 Complex Gain calibration

In order to calibrate the phases and the amplitudes of the visibilities, a temporary 'true' flux density for the complex gain calibrator of 1Jy is assumed. The behaviour of the bandpass is already accounted for from the bandpass calibration steps; the bandpass calibration solutions, along with the delay calibration solutions, are applied to the data as the solutions for the complex gain calibration are derived. As mentioned previously, the phase solutions of the complex gain calibration should be very closely matched. A large mismatch in phase solution is often a sign of failed calibration due to a problem in the data, such as RFI.

Similar to the bandpass calibration, if any RFI or poorly behaved baselines are noticed either in the calibration solutions or in the calibrated data, the poorly behaved data is flagged, and calibration, checking, and calibration application is iterated until only well-behaved data remains.

5.6.8 Flux scaling

Flux scaling allows the complex gain calibrator flux density to be adjusted from its temporary value of 1 Jy to its actual physical amplitude. As a result of the completed bandpass calibration and the complex gain calibration, the ratio of 'true flux' to 'observed flux' was measured for the observation. This ratio is termed the 'antenna

gain', and changes as a function of elevation due to distortions of gravity, along with other quantities that change the gain of the antenna. This efficiency factor should not change significantly over the course of the observation, this ratio between 'true' and 'observed' fluxes can be applied to the all the data observed. This amplitude correction is then applied through the task FLUXSCALE, and the flux density of the complex gain calibrator is scaled to its absolute value.

The value of the complex gain calibrator can then be compared to the flux density stated in the VLA Calibrator Catalogue. Discrepancies between the catalogue values and the values output by the flux density scaling routine are not of great concern, as many catalogue sources are somewhat variable.

Once the final flux density scaling has been completed, one final calibration is run to apply any the final set of amplitude adjustments to all targets in the observation. This should result in a fully calibrated bandpass calibrator, complex gain calibrator, and properly flux density scaled science target data.

5.6.9 Splitting out the target source

Once all calibration is complete, the target field should be stripped out of the full data set. This has the advantage of lowering the size of the file with which subsequent reduction steps must deal. A task called SPLIT is designed specifically for this task, and will select out the regions of the data cube identified.

The science data must also be inspected for signs of RFI, poorly behaved channels, or misbehaving baselines. However, in contrast to the two previous calibrator targets, nothing needs to be re-run after the flagging of any poorly behaved data, as all scalings are determined by the previously edited and calibrated bandpass and complex gain sources.

5.6.10 Preparing for Imaging

Regridding

CVEL is a task that allows the data to be transformed from the native frame of reference, called TOPO (topocentric, or telescope-centred), to a reference frame of choice. The optimal choice is BARY, or barycentric, which is in reference to the centre of mass between the Earth and the Sun. CVEL also requires a velocity mode if the user wishes to convert the data cube from the frequency space in which it was

observed to a velocity gridding. Since radio astronomy and optical astronomy use different definitions for velocity, the distinction between the two is required if CVEL is operating in velocity mode. Optical velocities are calculated as:

$$v_{optical} = cz = c\left(\frac{\lambda_{obs} - \lambda_{rest}}{\lambda_{rest}}\right) \tag{5.4}$$

In the radio convention, velocity is determined in frequency space as the following:

$$\begin{aligned} v_{radio} &= c\left(\frac{\nu_{rest} - \nu_{obs}}{\nu_{rest}}\right) \\ &= c\left(\frac{\lambda_{obs} - \lambda_{rest}}{\lambda_{obs}}\right) \\ v_{radio} &\neq v_{optical} \end{aligned}$$

For very low redshift objects (e.g., within the Galaxy), radio and optical velocities are approximately equal. However, for the purposes of extragalactic astronomy, radio and optical definitions become increasingly discrepant. Since the optical definition is more intuitive for extragalactic observations, the optical definition of velocity is chosen. CVEL also combines multiple spectral windows (if they exist in the data) and combines them together into one wider spectral window.

For this work the data cube was regridded from frequencies observed in TOPO frame to optical velocities in BARY frame. Channels were simultaneously averaged together during this process. The velocity resolution in the raw data was 3.6 km s^{-1}; this was averaged to a set of 10 km s^{-1} slices. For all observations in 12A-061, the data were also combined into a single spectral window. For 13A-537, this additional data manipulation was not required, as the data were taken with only a single spectral window.

Continuum subtraction

Continuum subtraction is necessary for any spectral line observations. Since the continuum around an extragalactic source can be filled with rather strong continuum sources, in order to observe the relatively weaker extragalactic source, these continuum sources must first be removed from the data. UVCONTSUB takes a data set which consists of one spectral window, and with the input of sections of the data which are free of emission line flux density (generally near the extremes of the velocity width in the data), fits a model to the continuum and extrapolates across the region where

the HI emission is expected to be found.

Testing the continuum subtraction

In order to verify that the continuum subtraction has worked properly, several additional steps can be implemented. First, the original file produced by CVEL can be imaged using CLEAN. CLEAN performs a Fourier transform on the data to convert from the UV plane to the image plane.

This produces a continuum image of the target field. The NRAO VLA Sky Survey (Condon et al., 1998, hereon NVSS) is a low frequency continuum survey of the Northern sky, conducted with the VLA in D configuration, and can be used as a point of reference to ensure that the data have been properly calibrated. With the help of an NVSS thumbnail, the continuum sources in the field of view of the telescope can be compared to the continuum image. If there are large artifacts or other mismatches between the NVSS field and the field within the EVLA data, additional calibration of the EVLA data may be required.

The cleaned image should show peaks in all the locations of continuum sources detected in the NVSS field. As the NVSS was observed with D configuration, the data obtained in the current project is slightly more resolved than the NVSS field; certain continuum sources are therefore weaker and can also be resolved into multiple components because the C-configuration observations have higher angular resolution (with a beam size of $14''$) than the D-configuration observations (with a beam size of $46''$) of the NVSS. The noise level of the cleaned cube can also be compared to the predicted noise level of the data from the Observing Time Calculator as an estimate of how much the flagged data has affected the sensitivity of the data. For some data sets, a significant fraction of the data had to be flagged, so this comparison was not always trivial.

To test that the continuum subtraction has worked as expected, the target field can be imaged after subtraction. By definition, this field should contain no major sources. If the image of the continuum-subtracted cube does have major sources of emission, the UVCONTSUB program has not functioned as planned.

5.6.11 'Imaging' the science target

At this point, the CLEAN task is required once more to create a spectral line cube of the science target. Setting the cleaning mode to 'velocity' (with a width in velocity either

of the same size as produced by CVEL, or a multiplicative factor of that number, ensures that the data cube that is produced can be iterated over velocity instead of frequency. CLEAN will also take into account a weighting scheme for the beam. For this work, the 'natural' weighting scheme is used, along with a cell size of 4.0 arcseconds. The cell size is determined by the size of the synthesized beam (which for C configuration is ~ 15 arcseconds) and the Nyquist theorem, which states that one should sample the beam by at least $2.4\times$ the cell size. With a grid of nine pixels to the beam size, this results in a beam size of approximately 4 arcseconds. Since the galaxies comprise a relatively small portion of the sky, a relatively small image pixel size can be used, and 128×128 pixels is used for the image size. The output of CLEAN is an image of the targeted sky across a wide range of velocities, where the velocities are binned into 'channels' according to the input of CLEAN.

Using the RA and DEC of the galaxies taken from Sloan, it is then possible to search for a spectral line feature within the data. If no spectral line is immediately apparent, due to low signal to noise, the data may be further smoothed within CLEAN by specifying a broader velocity width until either a line is apparent or it is clear that no line is present.

In general, the detections in the data presented here are not strong enough to require any further processing, and the spatial component of the emission does not add to the information about the galaxy. In this case, the spectrum at the location and velocity of the galaxy may be used as a tentative detection on the mass of the HI associated with the galaxy. For the remainder of this thesis, the spectra are considered the final data products of the observations obtained.

For those galaxies with strong emission line fluxes, further cleaning is required to deconvolve the source from the point spread function. In this case, and in order to determine the spatial extent of the line emission, a moment map can be used to determine the spatial extent of the emission line. By selecting the velocity range over which the emission line is present, and excluding all pixels from the moment map which are less significant than 2 times the noise level of the cube, the flux density remaining should primarily be associated with emission from the galaxy [8].

[8]In some cases, CASA is unable to produce a moment map due to the way that it defines the beam size. It selects the size of the beam from the first channel of the image cube and attempts to smooth all other channels to this size. The beam size is recorded as a function of frequency, and is stored in each velocity channel. If the beam becomes larger later on in the cube due to a significant amount of flagged data, the moment map creation will fail. If this occurs, the solution is to determine the maximum values of the beam's major and minor axis, along with position angle, and manually smooth the data to that value using IMSMOOTH. This will produce a new image cube which contains

The root mean squared (RMS) noise level is determined most accurately by using a number of channels on either side of the emission and the task IMSTAT. IMSTAT will output the RMS values of the selected channels. The RMS of the channels on either side of the emission region are usually of similar magnitude and a reasonably accurate estimate of the RMS of the cube as a whole. The spatial extent of the HI emission remaining in the moment map can be used to design a clean mask for further cleaning of the image cube. This final cleaning step ought to remove the residual ringing effects from the image cube, lowering the RMS of the cube as a whole.

5.7 Science products

5.7.1 Extracting a Spectrum

In order to obtain the spectrum of the galaxy, the final data cube as produced by CLEAN is loaded into VIEWER. Using the coordinates of the galaxy as a first pass, a point can be placed at the exact RA & Dec of the two galaxies in the cube. The point uses only the fluxes contained within the beam centred upon the pixel on which it is placed.

Once all units are properly aligned, to locate the optimal spatial region to include in the spectrum, some experimentation is required. Generally, scrolling through the image cube in velocity space can help to determine how broadly distributed the emission from the galaxy appears. In some cases, the centre of the HI emission is not exactly coincident with the RA & Dec centre as determined by the SDSS. The optimal selected region may be irregularly shaped, rectangular, oval, or a simple point, whichever appears to minimize noise in the non-emission portion of the data cube and capturing the most significant emission signal. This is done independently for both galaxies in the data cube.

When the spatial region is selected, the spectrum output in CASA can be saved to a machine-readable .txt file which contains the amplitudes as a function of velocity. It is at this stage that CASA is no longer used, and data analysis switches back to self-coded python scripts. Python scripts are used to produce plots of the spectrum itself, along with estimations of the signal to noise level of the emission recorded, and the HI gas mass estimates based upon the integrated flux.

---

identical beam sizes at every channel, and the moment map creation can run smoothly.

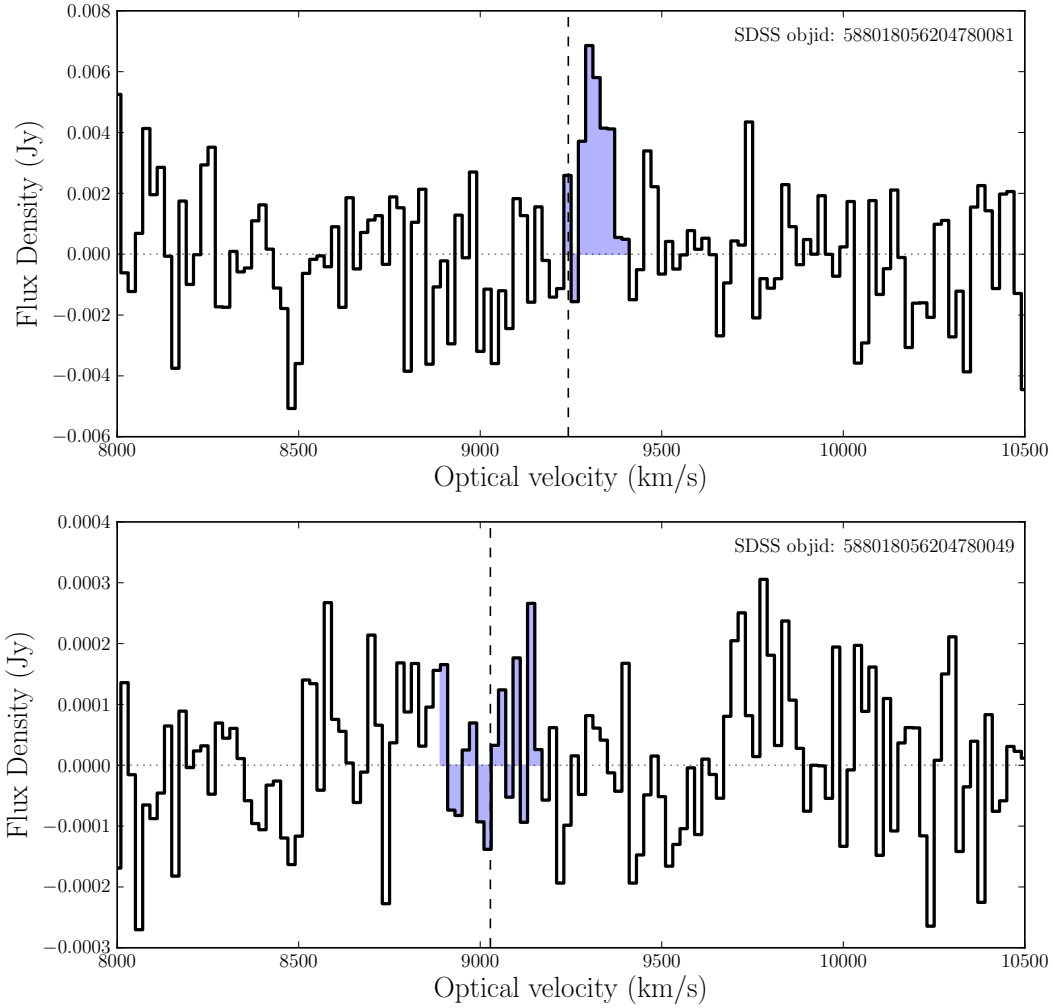


Figure 5.10: Flux density vs. optical velocity for galaxy pair 588018056204780081 & 588018056204780049. The vertical dashed line indicates the velocity of the galaxy, horizontal dotted line indicates 0 flux. The shaded region indicates the region of the spectra used for signal to noise and gas mass calculations. The SDSS image for this pair is presented in Figure 5.6a. The upper spectrum has a S/N of 3.19, while the lower has a S/N of 1.07. For a complete illustration of the spectra of the full sample, see the Appendix.

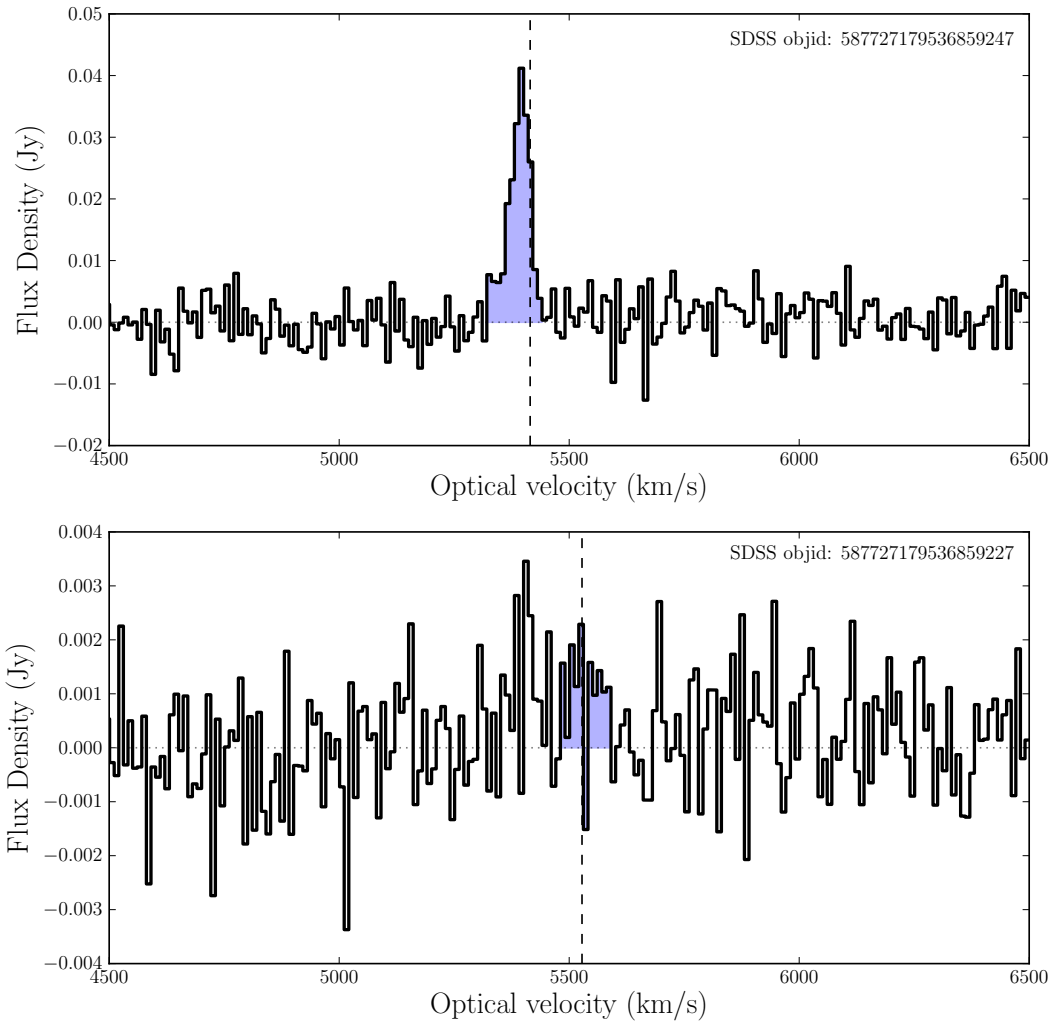


Figure 5.11: Same as in Figure 5.10; Flux density vs. optical velocity for galaxy pair 587727179536859247 & 587727179536859227. The SDSS image for this pair is presented in Figure 5.6b. The upper spectrum has a S/N of 11.40, while the lower spectrum has a S/N of 2.02.

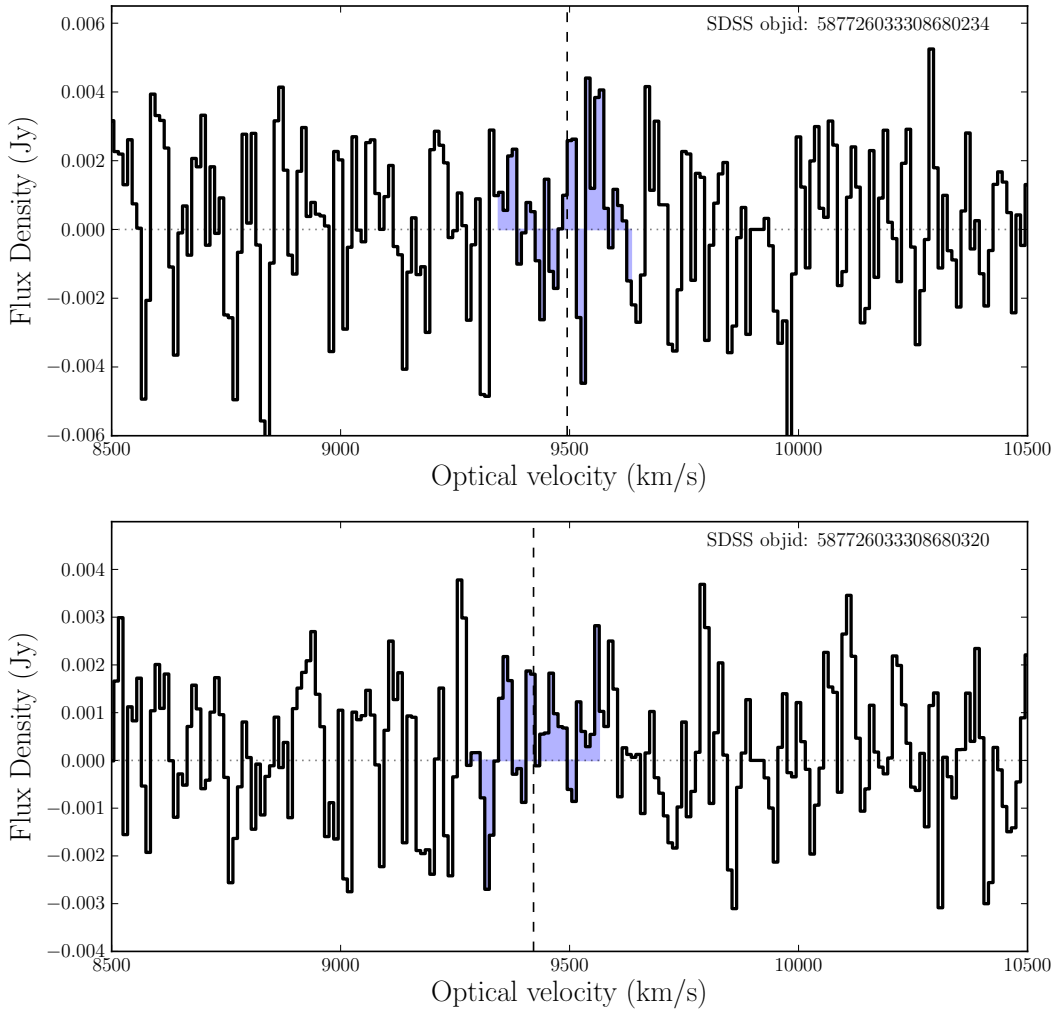


Figure 5.12: Same as in Figure 5.10; Flux density vs. optical velocity for galaxy pair 587726033308680234 & 587726033308680320. The SDSS image for this pair is presented in Figure 5.6c. The upper spectrum has a S/N of 1.97, while the lower spectrum has a S/N of 2.00.

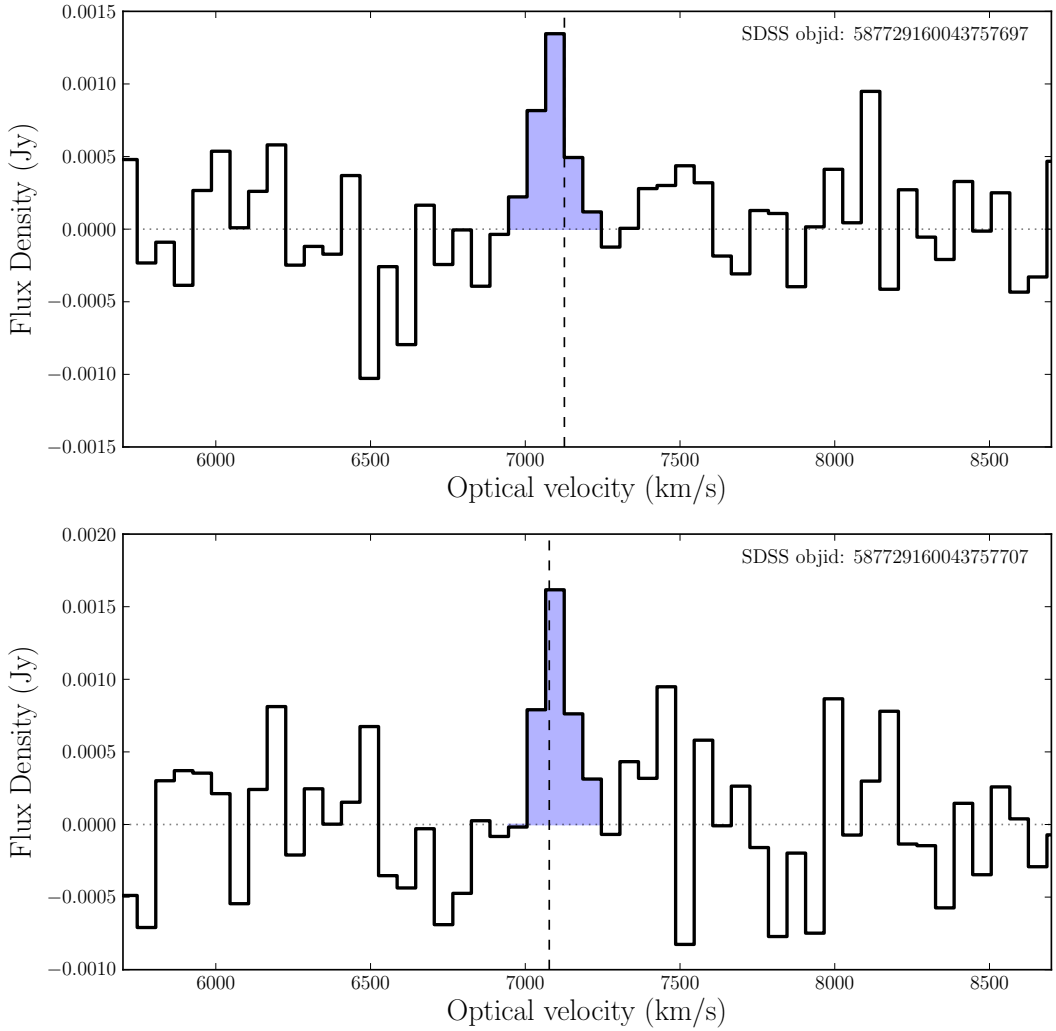


Figure 5.13: Same as in Figure 5.10; Flux density vs. optical velocity for galaxy pair 587729160043757697 & 587729160043757707. SDSS image for this pair is presented in Figure 5.6d. The upper spectrum has a S/N of 3.76, while the lower spectrum has a S/N of 3.16.

5.7.2 Signal to noise calculations

In order to determine the significance of the emission lines recorded in the data cube, the signal to noise value for each spectrum must be calculated. There is no clear 'best method' for determining the significance of line emission in radio data, but two methods are primarily used.

The first is a peak over RMS calculation. This is the more straightforward of the calculations; the peak flux value is divided by the RMS of the spectrum outside of the emission line region. The emission line region used is illustrated in Figures 5.10 – 5.13 as the shaded blue region.

The second method comes from the ALFALFA (Arecibo Legacy Fast ALFA, where ALFA is 'Arecibo L-band Feed Array') survey papers (e.g., Saintonge, 2007; Haynes et al., 2011).

$$S/N = \left(\frac{1000 S_{21}}{W_{50}}\right) \frac{w_{smo}^{1/2}}{\sigma_{RMS}} \tag{5.5}$$

Here, S_{21} is the integrated flux under the line, in Jy km s^{-1}. The factor of 1000 is to convert the flux value into mJy. W_{50} is the width of the spectral feature in km s^{-1}. Collectively, the portion in brackets is flux density across the whole spectral feature, in units of mJy. w_{smo} is the smoothing width, and is the number of bins that make up half the width of W_{50}. For the ALFALFA papers, a bin is 10 km s^{-1} in width, so w_{smo} is also written as $W_{50}/(2 \times 10)$. As long as W_{50} is less than 400 km s^{-1}, this approximation is used; if W_{50} is $>$ 400 km s^{-1}, W_{50} is replaced in the w_{smo} expression by 400 km s^{-1}. w_{smo} is divided by two in order to account for the additional Hanning smoothing done at the very beginning of data reduction, which improves the quality of the spectra by a factor of two. This ALFALFA expression was developed in part based on the ALFALFA team's automated emission line finder algorithm. This algorithm attempts to fit a model emission line template to the form of the emission line, rather than looking at the peak flux of the emission.

Both the results of a peak flux over RMS S/N calculation, and the results of the S/N calculation according to the ALFALFA expression are presented in Table 5.4. For the data analysis that follows, the peak/RMS S/N values are used.

Non-detections

Any galaxy whose emission line has a signal to noise of less than 3.0 (as calculated by peak/RMS) should be considered to be a non-detection. Therefore, galaxies whose spectra have a signal to noise less than 3.0 have an upper limit calculation made based upon the RMS of the data cube outside the line region. An estimate of the maximum amount of flux that could be in the data over the same emission line region at 3σ can be calculated. The RMS of the region of the spectrum outside of the emission region is calculated. The amount of flux that could be present over the same velocity range as the emission feature, is then calculated by multiplying the RMS value by a factor of 3, and integrating over the velocity width of the failed detection. This assumes a step function emission feature, which is not very realistic, but it should give a conservative upper limit to the HI mass present in the system. With the flux limits, an upper limit gas mass may be calculated using the standard formula, thereby providing a gas fraction limit in addition to the flux and HI mass limits. These limits are presented in Table 5.4; all upper limits are indicated by the $<$ in front of the estimate.

Gas masses

Gas masses are determined from the spectrum by scaling the integrated total flux in the emission line across the velocity width over which it is present, using Equation 5.2. The results of this calculation are presented in Table 5.4.

Stellar Masses

The photometry of the SDSS has been recalculated by Simard et al. (2011) to account for crowding issues in the previous SDSS photometry that were particularly poor for galaxies at small separations. Simard et al. (2011) improved upon the existing sky subtraction and deblending procedures, which particularly affected objects in close apparent association to each other, such as galaxy pairs. To be consistent, the stellar mass calculations, which are based upon the photometry, are repeated. The masses from Mendel et al. (2014) are therefore used. This calibration uses both the single Sérsic fits and the bulge/disk decompositions of Simard et al. (2011) to calculate masses for the galaxies in the photometric sample. For the work that follows, the results of the single Sérsic fits are used.

The fitting was done using a large grid of synthetic SEDs based on the Chabrier (2003) models, which were simultaneously fitted to the 5 band photometry. This fitting routine outputs the stellar masses by using the best fit SED model that matches the photometry. The parameters that go into the SED fitting are the $E(B-V)$ values, the stellar metallicity, the rate at which star formation declines, and the population age. By finding the best-fit model over all these parameters, the stellar mass can be estimated. The models predict certain fluxes per unit mass, so with a scaling to match the amplitude of the fluxes that are observed, the stellar mass of the galaxy can be estimated by scaling the normalized stellar mass of the best fit SED by the same factor.

Table 5.4: SDSS ObjID, Δlog(SFR) values for both fibre and total, HI gas mass, gas fraction, and S/N for both peak/RMS and ALFALFA calculations.

SDSS ObjID	$\Delta\log(\mathrm{SFR})_{tot}$	$\Delta\log(\mathrm{SFR})_{fib}$	M_{HI} ($\mathrm{M}_{\odot}$)	($\mathrm{M}_{HI}/\mathrm{M}_{*}$)	$\mathrm{S/N}_{(pk/RMS)}$	$\mathrm{S/N}_{ALFALFA}$
587726033341776175	0.2501465	−0.154705	< 9.31700	<0.46848	2.85421	1.70510
587726033341776191	−0.108226	0.107319	8.68459	0.02988	3.42757	2.28776
587739609695453284	0.236737	0.087317	9.18843	0.26365	4.07996	3.81064
587739609695453281	−0.415163	−0.631508	9.10137	0.08605	3.68863	2.14184
587744873717563559	0.000605	0.430205	8.48027	0.21831	5.35473	2.97228
587744873717563471	0.158192	0.48887	8.32039	0.64309	3.50965	2.07175
587727179536859247	−0.0478615	−1.105805	9.47965	3.82368	11.39616	11.59608
587727179536859227	0.0169275	−0.570885	< 8.77782	< 2.80742	2.02263	1.96433
587729160043757697	0.599746	1.11029	8.64328	1.07084	3.76537	2.65002
587729160043757707	0.510587	0.996827	8.70024	0.42146	3.16271	1.92981
587741489815027774	0.268483	0.799247	< 9.92147	< 1.22989	2.69840	1.79509
587741489815028146	0.541953	1.221671	< 9.84139	< 1.14727	2.22529	0.95586
587739303684866183	-0.1814955	0.068245	< 9.18656	< 3.77447	2.35447	1.66279
587739303684866173	−0.123769	−0.277825	8.72869	0.43364	3.56959	1.99486
587726033308680234	0.316141	0.996821	<9.94104	<1.97003	1.97173	0.79083
587726033308680320	0.406928	0.281004	<9.75795	<1.02026	2.00384	1.40004
588018056204780081	0.177184	−0.62394	9.31691	0.31879	3.18756	2.75781
588018056204780049	0.33618	0.14028	< 8.63737	< 0.01438	2.16678	0.75422
587729158970867777	0.024606	−0.021075	9.71241	0.68819	4.09109	5.76301
587729158970867792	−1.101018	−0.41663	9.11680	0.22793	3.70391	3.31781

Table 5.4 – continued from previous page

SDSS ObjID	$\Delta\log(\mathrm{SFR})_{tot}$	$\Delta\log(\mathrm{SFR})_{fib}$	M_{HI} ($\mathrm{M}_{\odot}$)	($\mathrm{M}_{HI}/\mathrm{M}_{*}$)	$\mathrm{S/N}_{(pk/RMS)}$	$\mathrm{S/N}_{ALFALFA}$
587733605328093368	−0.158113	−0.16979	< 8.83705	< 0.45650	2.25933	1.15455
587733605328093256	0.281566	−0.27453	9.66408	3.52714	4.03562	3.17385
588848899908370674	0.218891	−0.27249	9.60111	2.57018	4.541817	3.43467
588848899908370505	0.149803	0.27798	8.57115	0.10898	3.25502	1.43386
588017605758025795	−0.256845	−0.492335	< 9.17308	< 5.63927	1.93141	0.94320
588017605758025732	0.700759	1.007085	9.21998	5.27950	3.62006	3.52394
588017702411763744	0.682228	0.43367	< 10.24940	< 4.07176	2.67099	2.89599
588017702411763872	0.082843	0.4314	< 9.96207	< 7.68384	2.78635	1.52566
587727178473930875	0.66972	0.805306	< 10.01761	< 11.90185	2.62493	3.52301
587727178473930886	0.135028	0.499731	< 9.49656	< 4.46324	2.24308	2.00069
587742901789589569	0.542769	1.105814	9.54155	0.65500	3.20105	2.54339
587742901789589575	0.1562905	0.67437	< 9.98504	< 6.53173	2.70261	2.84665
588023670245949622	0.1646025	0.07104	< 9.00592	< 4.51044	1.61425	1.18894
588023670245949625	−0.3055225	−0.282695	< 9.38744	< 6.67621	2.73038	1.67577

5.8 Data analysis

The primary objective of determining whether or not strong correlations exist between the level of triggered star formation and the gas fraction of the galaxy can now be undertaken, as all necessary data are in hand. In Figure 5.14, the gas fraction of each galaxy is plotted against the total Δlog(SFR) of this galaxy sample. All points are colour-coordinated to match between galaxies of the same pair. Large circles indicate that the detection is above the peak/RMS S/N > 3.0 threshold, while smaller triangles indicate upper limits as calculated in §5.7.2. Error bars on the gas fraction are calculated in a similar way to the upper limits for the non-detections; the RMS value of the data cube was taken as a 1σ level of flux, and converted into a gas mass. This gas mass for the 1σ level was then divided by the stellar mass of the galaxy to obtain the 1σ error on each of the detections. 1σ error bars on the Δlog(SFR) values were assumed to correspond to the uncertainty on an individual SFR value, which is approximately 0.09 dex (see also Chapter 4). When stellar mass is plotted, the uncertainty on the stellar mass is taken to be 0.15 dex (Mendel et al., 2014). 17 galaxies (exactly half of the sample) have detections above S/N= 3.0. Relatively few galaxies populate the region corresponding to gas fractions > 0.5 and star formation rates below 0.0.

For clarity, in Figure 5.15, the galaxies that do not meet the S/N > 3.0 criterion are excluded from the sample. Without the non-detections present, there continues to exist a suggestive lack of points populating the region of the diagram where galaxies with gas fractions > 0.5 and star formation rate offsets of < 0.0. This gas fraction limit corresponds to a theoretical f_{gas} of 0.33, well within the range probed by the simulations.

To determine whether or not this lack of low Δlog(SFR), high gas fraction data points is statistically significant given the small sample size, a Spearman rank correlation test is applied to the S/N> 3.0 data (i.e., the upper limits shown in Figure 5.14 are not included in the correlation testing). This statistical test outputs both the rank correlation coefficient, which indicates the degree to which the two parameters tend to increase or decrease with each other, and a probability that the given distribution could arise from an uncorrelated distribution, which tests the null hypothesis.

For the two parameters displayed in Figure 5.15, gas fraction and total Δlog(SFR), the Spearman rank correlation coefficient is 0.5637 for the detections at S/N > 3.0. A coefficient of 0 indicates no correlation; the two parameters in that case do not

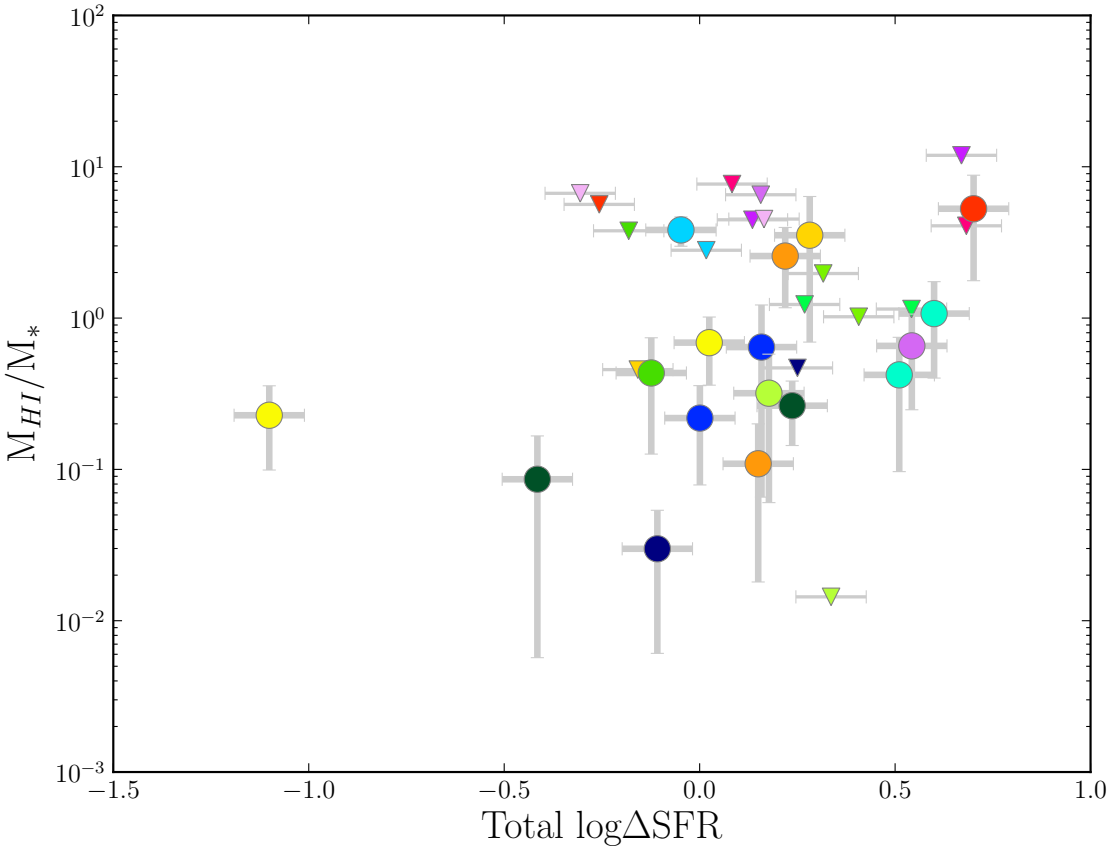


Figure 5.14: The log of the gas fraction vs. total Δlog(SFR). Points are colour-coded according to pair. Large circles indicate S/N $>$ 3, with smaller triangles indicating upper limits for the maximum amount of flux that could exist in the spectrum at 3σ.

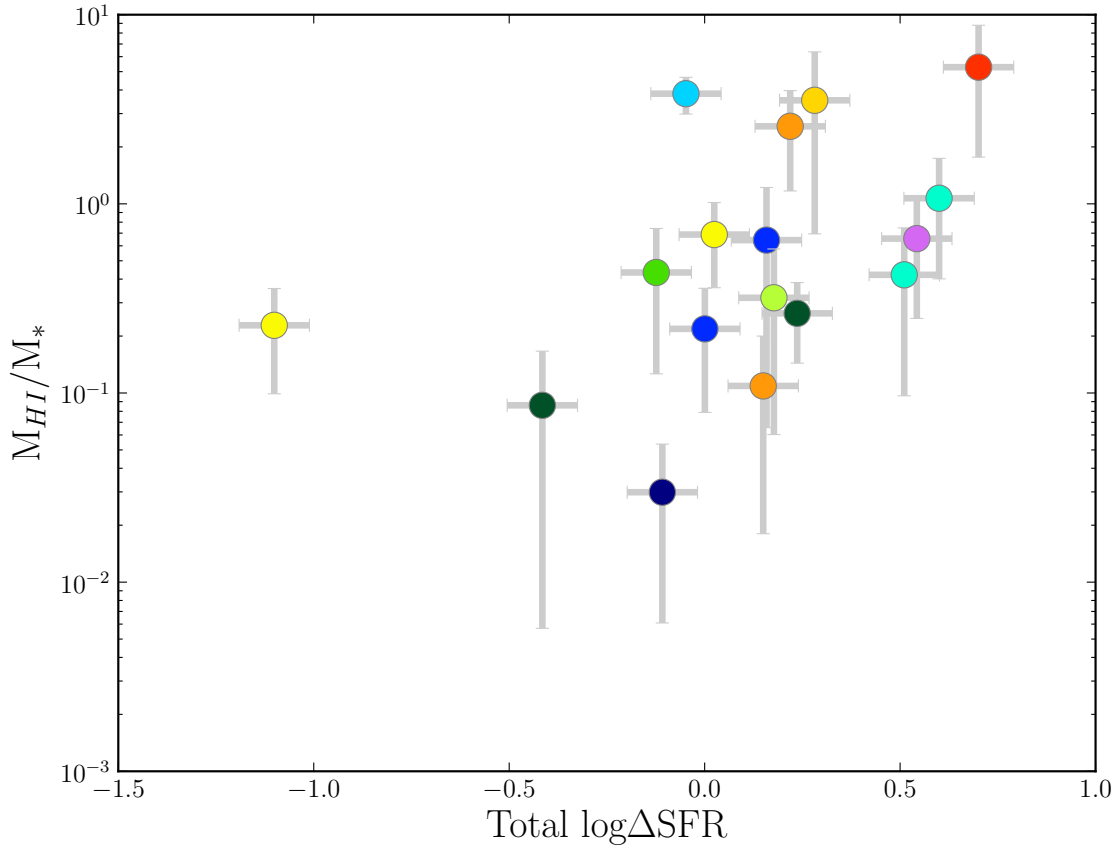


Figure 5.15: The log of the gas fraction vs. total Δlog(SFR). Points are colour-coded according to pair. Only galaxies with HI detections with S/N $>$ 3 are shown.

preferentially increase or decrease in tandem with each other. A coefficient of 1 or -1 indicates perfect correlation, although this correlation does not need to be linear. A coefficient of ± 1 would be obtained under any conditions where the secondary parameter *always* increased or decreased as the first increased. A rank coefficient of 0.56 is indicative that the two parameters generally increase with each other, although there is significant scatter in the relation.

Potentially more compelling is the null hypothesis test. The p-value output by the Spearman rank test indicates the likelihood with which a random sample of uncorrelated data would appear to have the same distribution, given the sample size. For this sample (of S/N $>$ 3.0 galaxies only), the p-value output is 0.0184. This is approximately 2.5σ (which would have been a p-value of 0.0124). This test indicates that the null hypothesis of no correlation between the total Δlog(SFR) and the HI gas fraction is ruled out at greater than 2σ for those galaxies which have detections at S/N $>$ 3.0.

A linear regression to the points in Figure 5.15 results in a slope of 0.9, with a y-intercept of -0.15 for a line of functional form $y = 10^{Ax+b}$. The fitted line is therefore $M_{HI}/M_* = 10^{0.9\Delta SFR-0.15}$, and is shown in Figure 5.16. The covariance matrix output by the fitting algorithm (in this case, Python's CURVE_FIT), indicates that the variance on the slope is 0.24 and the variance on the y-intercept is 0.069. This fitting algorithm uses a least-squares fitting method to minimize the distance between the points and the line.

To test whether this correlation is primarily driven by star formation in either the fibre region of the galaxy (i.e., centrally triggered star formation), or by star formation outside the fibre region, a similar plot can be made using fibre Δlog(SFR) or a newly calculated outer Δlog(SFR). The outer Δlog(SFR) is calculated as in Ellison et al. (2013a) as the total value minus the fibre value. This should leave only the enhancement that is due to the region of the galaxy outside the central 3 arcsecond fibre.

Figure 5.17a shows the log of the gas fraction versus fibre Δlog(SFR) for the full sample, including all upper limit non-detections, as in Figure 5.14. Figure 5.18a shows only those galaxies which have a peak/RMS S/N of greater than 3.0, as in Figure 5.15. The same Spearman rank correlation test that was applied to the total Δlog(SFR) correlation is applied to the fibre Δlog(SFR) correlation. The Spearman rank correlation coefficient is 0.1275, with a p-value for the null hypothesis of 0.6259. This p-value is completely consistent with the distribution being drawn from an un-

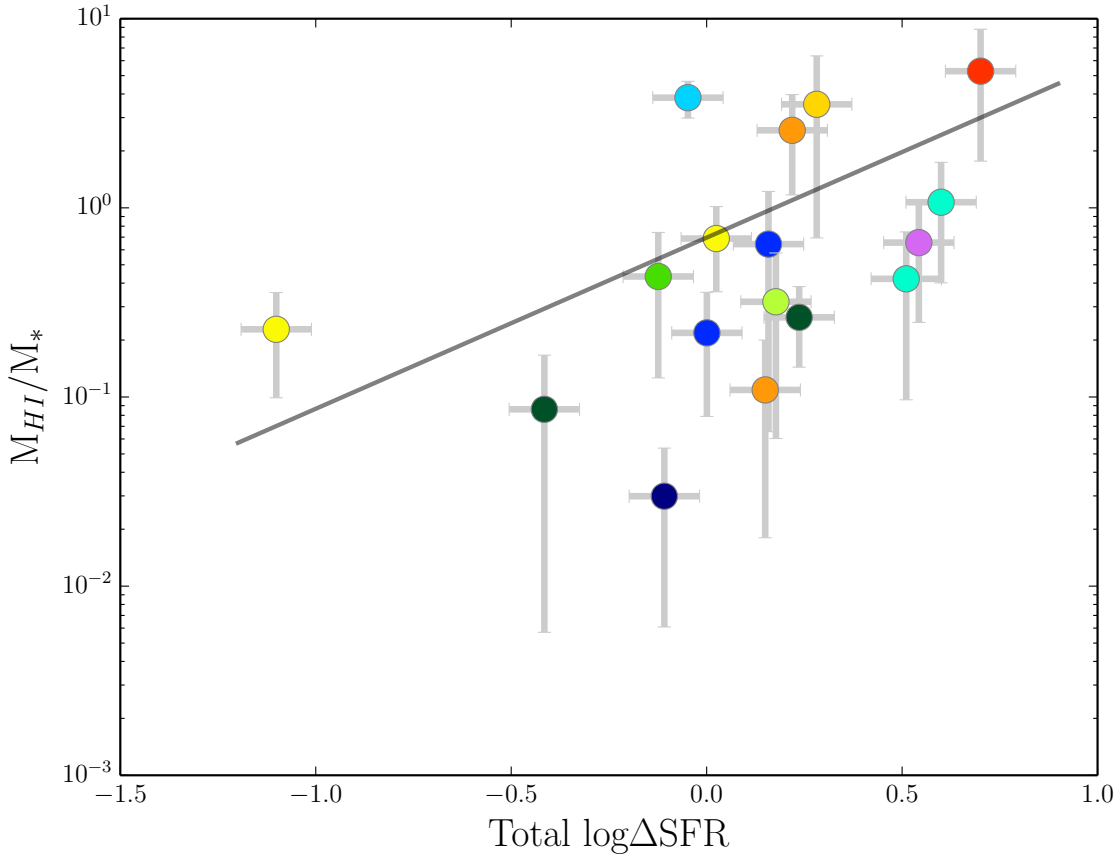
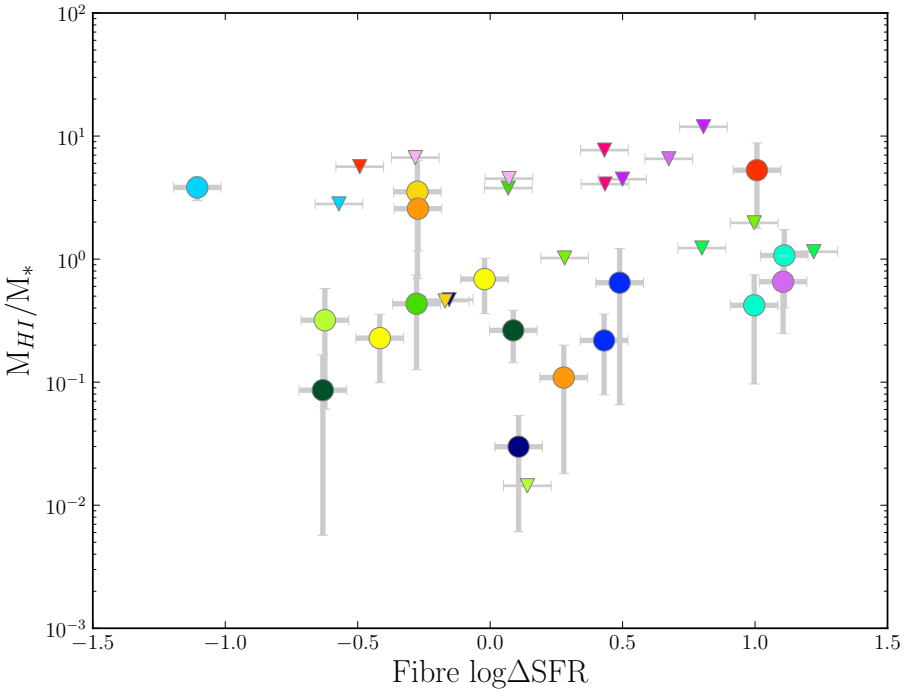
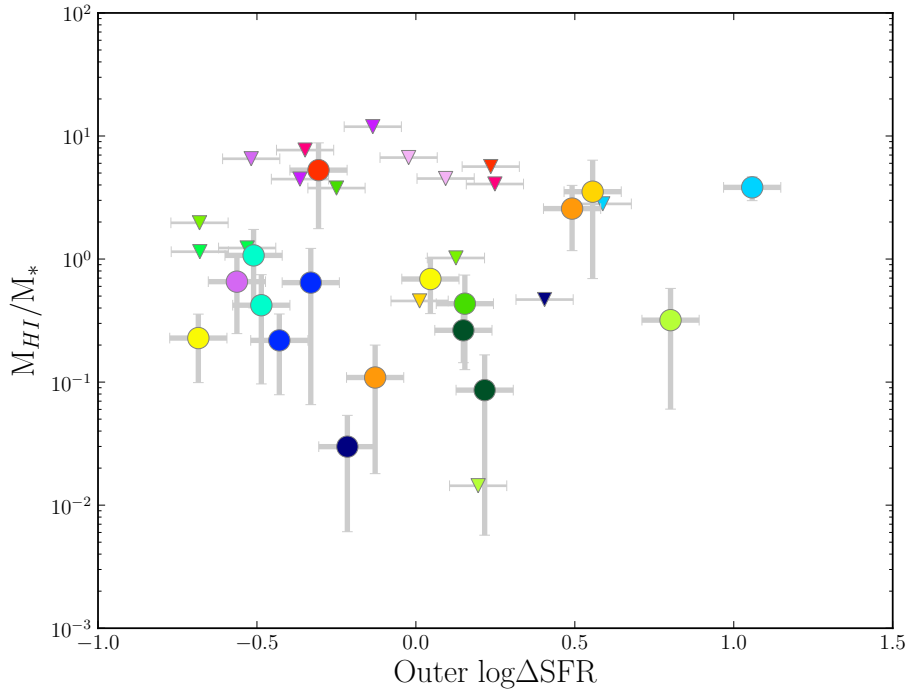


Figure 5.16: Same as Figure 5.15; The log of the gas fraction vs. total Δlog(SFR), with points colour-coded according to pair, for galaxies with HI detections with S/N > 3. The gray line overlaid on the points shows the results of the linear regression, using a least-squares method.



(a) Gas fraction versus SFR enhancement in the fibre.



(b) Gas fraction versus SFR enhancement outside the fibre.

Figure 5.17: Panel a: Gas fraction versus fibre Δlog(SFR). Panel b: Gas fraction versus outer Δlog(SFR). Both upper limits (smaller triangles) and S/N > 3.0 detections (larger circles) are plotted. Points are colour-coded according to pair.

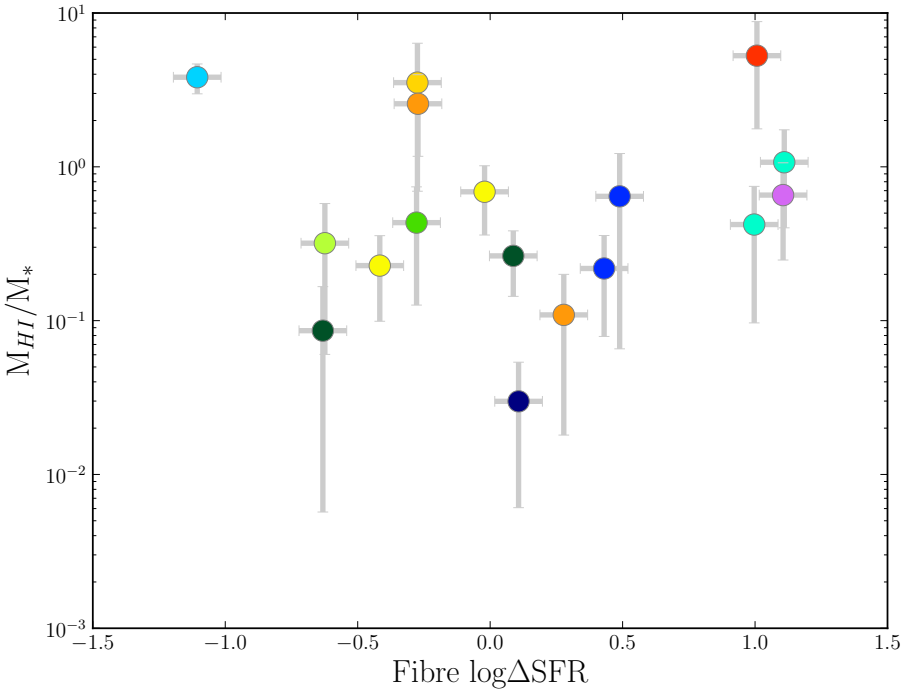
correlated distribution. The correlation coefficient is also much closer to 0 than was found for the total Δlog(SFR) values; the fibre values are therefore not significantly correlated with the gas fraction of the galaxy.

Similar results are obtained for the outer Δlog(SFR). As with the fibre values, the log of the gas fraction of the galaxy is plotted against the outer Δlog(SFR) in Figure 5.17b, which shows the entire sample, including the upper limits for the non-detections, and Figure 5.18b shows only the galaxies with S/N$>$ 3.0. When the Spearman rank correlation is applied to the outer Δlog(SFR) values, a rank correlation coefficient of 0.1961 is determined, along with a p-value of 0.4507, which is not a significant correlation.

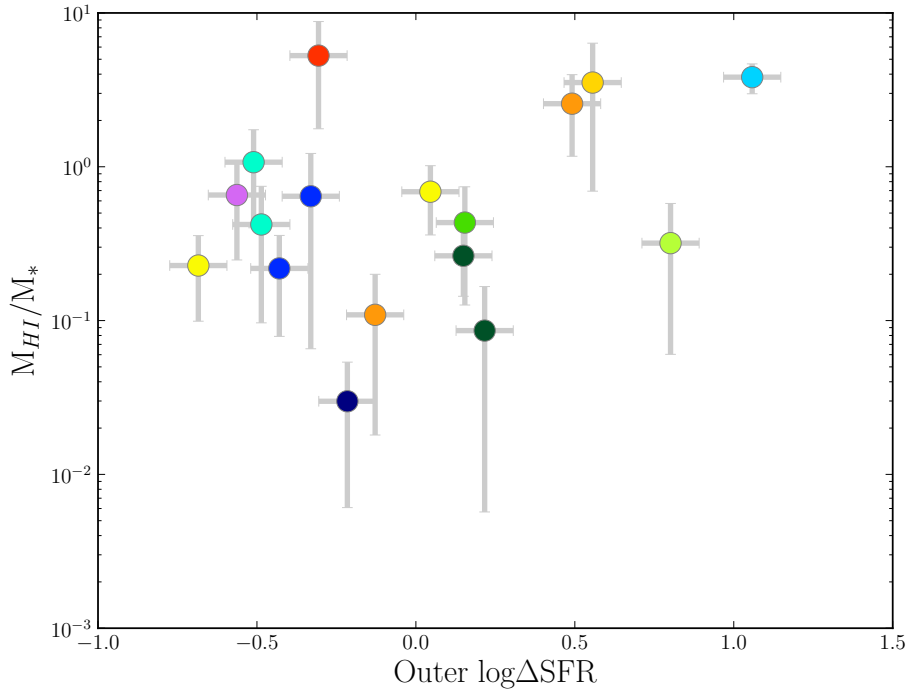
It appears that only when the SFR enhancement of the galaxy as a whole is considered, does the Δlog(SFR) show any hint of significant correlations between the SFR enhancement of the galaxy and its gas fraction. In Figure 5.19, it can be seen that galaxies with high total Δlog(SFR) values correlate with galaxies with high fibre Δlog(SFR)s. This result is not highly surprising, as it has been previously shown that galaxies undergoing an interaction tend to show the majority of their triggered star formation within the central region (e.g., Patton et al., 2011). However, at fixed total Δlog(SFR), there is a wider range of fibre Δlog(SFR) values than is true in the converse case. Some galaxies appear to have triggered a larger amount of star formation in the outskirts of the galaxy, which accounts for those galaxies with total enhancements of $\approx$ 0.0, but fibre enhancements well below 0. The fibre Δlog(SFR) and outer Δlog(SFR) are strongly anticorrelated, but since the outer Δlog(SFR) values are calculated from subtracting the fibre value from the total value, this anticorrelation is simply a feature of the correlation between fibre and total Δlog(SFR).

5.8.1 Monte Carlo tests of the Spearman rank metric

In order to test the robustness of the Spearman rank correlation statistic, a Monte Carlo resampling of the data points with S/N $>$ 3.0 is undertaken. Each of the data points is resampled within the error bars, assuming gaussian uncertainties with width equal to the size of the error bar. Resampling is done for both the Δlog(SFR) and the H I gas fraction. Each set of 17 re-drawn galaxies is fed to the Spearman rank correlation statistic, and the p-value indicating the probability of no correlation is recorded. This process is repeated 10,000 times, and the histogram of the resulting



(a) Gas fraction versus SFR enhancement in the fibre.



(b) Gas fraction versus SFR enhancement outside the fibre.

Figure 5.18: Panel a: Gas fraction versus fibre Δlog(SFR). Panel b: Gas fraction versus outer Δlog(SFR). Only galaxies with HI detections with S/N > 3 are shown. Points are colour-coded according to pair.

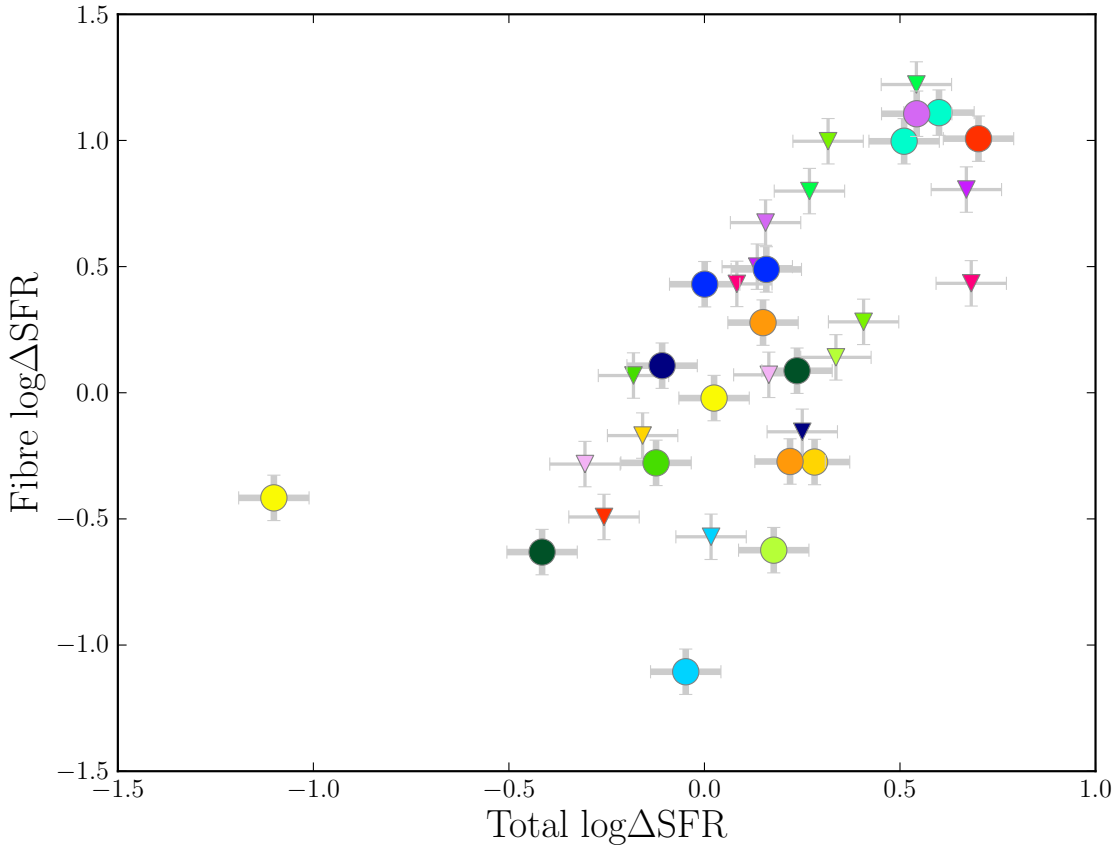


Figure 5.19: Fibre Δlog(SFR) vs. total Δlog(SFR) for all galaxies in the sample. High total Δlog(SFR) galaxies tend to be associated with the highest fibre Δlog(SFR) galaxies.

distribution is plotted in Figure 5.20, normalized to the number of runs. Figure 5.20 is truncated at a p-value of 0.5 for visibility purposes; the tail beyond 0.5 continues out to 1.0, but the fraction of runs at high p-values is low. 76.6% of all runs are found at $> 1\sigma$. 50% of all runs have p-values less than 0.124, which is approximately 1.5σ. 25.4% of all runs are found at $> 2\sigma$, and only 0.8% of all runs were found to be $> 3\sigma$. At the p-value presented in the previous section, 11.0% of all runs were found to exclude the null hypothesis at higher significance. The mode of this distribution is found at the original, measured p-value.

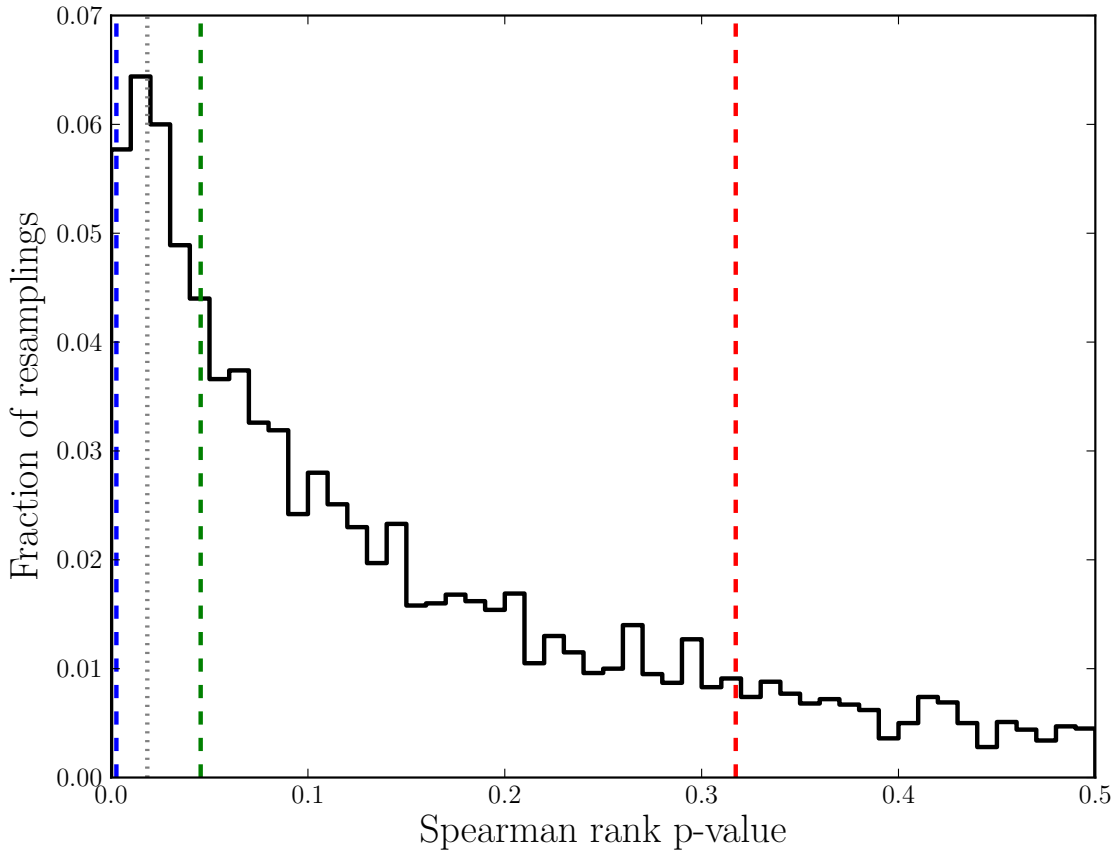


Figure 5.20: Normalized histogram of the p-values of the Spearman Rank correlation test after each of 10,000 resamplings of the error bars in Figure 5.15. For visibility, this figure is truncated at a p-value of 0.5, but the tail continues out to a p-value of 1.0. The red, green, and blue vertical dashed lines indicate p-values of 1σ, 2σ, & 3σ respectively. The gray dotted line indicates the measured p-value from the data. The mode of the distribution is centered around the measured value.

Taken as a whole, high gas fraction galaxies appear to be preferentially forming stars at a higher rate than expected for galaxies of their mass, redshift, and local environment. However, this trend of higher gas content in galaxies which show stronger overall triggered star formation does not appear to be more tightly correlated with SFR enhancements in the central region of the galaxy.

5.9 Comparison with ALFALFA

To determine whether the galaxies observed here are unusual with respect to the average galaxy population, a rough comparison with the galaxy population detected with by the ALFALFA survey is undertaken. The ALFALFA survey is a drift-scanning survey undertaken at the single dish Arecibo observatory in Puerto Rico. ALFALFA is a relatively shallow survey, so while the VLA data is expected to detect galaxies at much lower HI masses than the ALFALFA survey, a rough comparison to galaxies which are not likely to be in galaxy pairs may serve as a useful comparison sample, particularly at the high HI mass end. The current publicly available data release contains 40% of the total sample (ALFALFA.40), and contains over 15,000 extragalactic HI sources, of which > 8000 can be reliably matched to an optical counterpart from SDSS. Derek Fertig has cross-matched the sources within the ALFALFA.40 catalogue to the SDSS by looking for optical counterparts within the ALFALFA beam for significant HI detections within the ALFALFA survey.

To construct a comparison sample, galaxies which are not in pairs and have no nearby companion within the size of the ALFALFA beam ($\sim$ 3.5 arcminutes) are selected. Using the terminology of the ALFALFA.40 data release, this means galaxies classified as code 1 (a detection with S/N > 6.5, as calculated through Equation 5.5), ocode = I ('optical code', indicating a matched SDSS source), and with N_gals = 1, which restricts the matching to galaxies which are the only optical counterpart within 1.5 arcminutes and 500 km s^{-1}. The first two criteria restrict the sample to robust HI detections of SDSS galaxies, while the latter criterion serves as an isolation requirement. Galaxies must also have calculated stellar masses and star formation rates; Δlog(SFR) values are calculated in an identical manner as for the VLA sample. This results in a comparison sample of 4595 SDSS galaxies with no close companion and robust HI masses.

When considering the gas fraction detected as a function of stellar mass, a broad trend of higher stellar masses being correlated with lower gas fractions is evident,

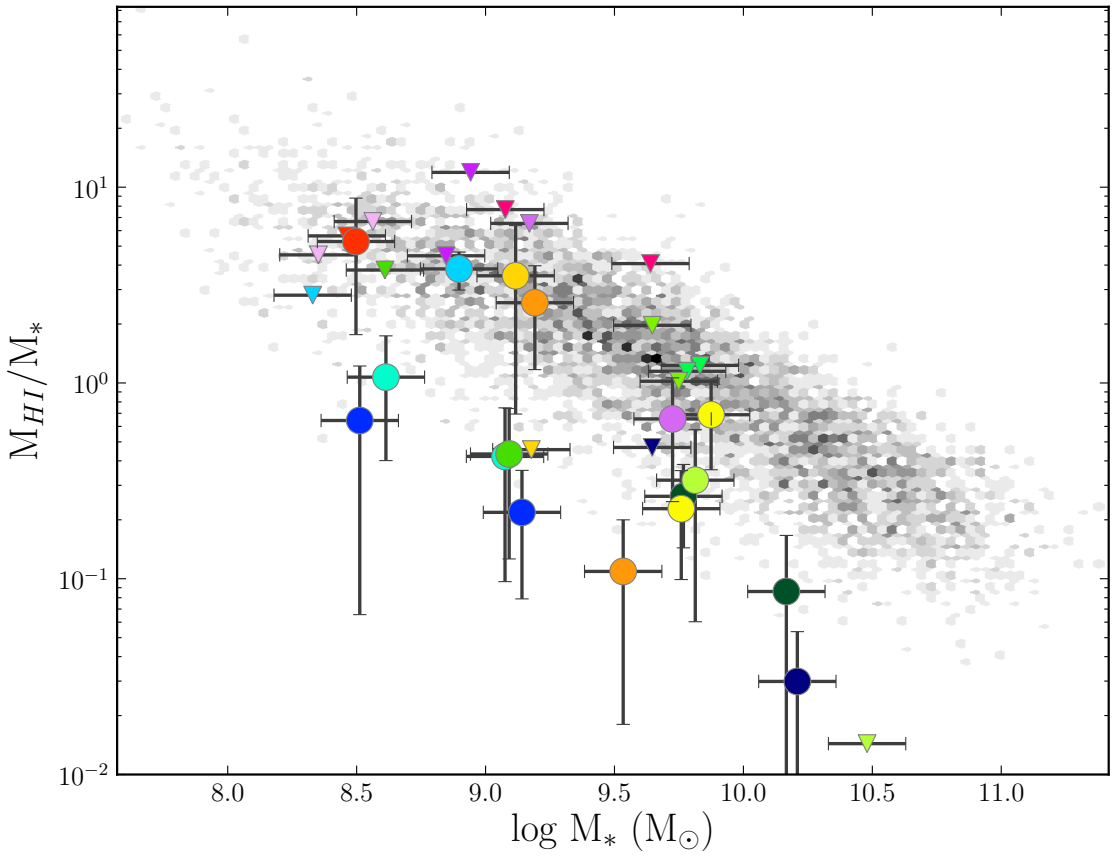


Figure 5.21: The log of the gas fraction versus stellar mass for both the ALFALFA & the VLA samples. The ALFALFA sample is shown as a 2-D grayscale histogram, where darker points indicate bins with a larger number of galaxies. Colour coding and symbols of the coloured points are as in Figure 5.14.

as expected from previous studies (e.g., Kannappan, 2004). All gas fractions within the VLA sample detected at S/N $>$ 3.0 either fall within the main envelope of the ALFALFA detections or to lower gas fractions. Lower gas fractions at fixed stellar mass (which in this figure, indicates smaller HI gas masses for a given stellar mass) than those detected by ALFALFA are to be expected, as the VLA galaxies have been observed for a longer period of time than the ALFALFA detections, and are therefore deeper observations. However, Figure 5.21 indicates that the high gas fraction end of the VLA sample of galaxies presented in Figure 5.15 is not an abnormal population, given the stellar masses of those galaxies.

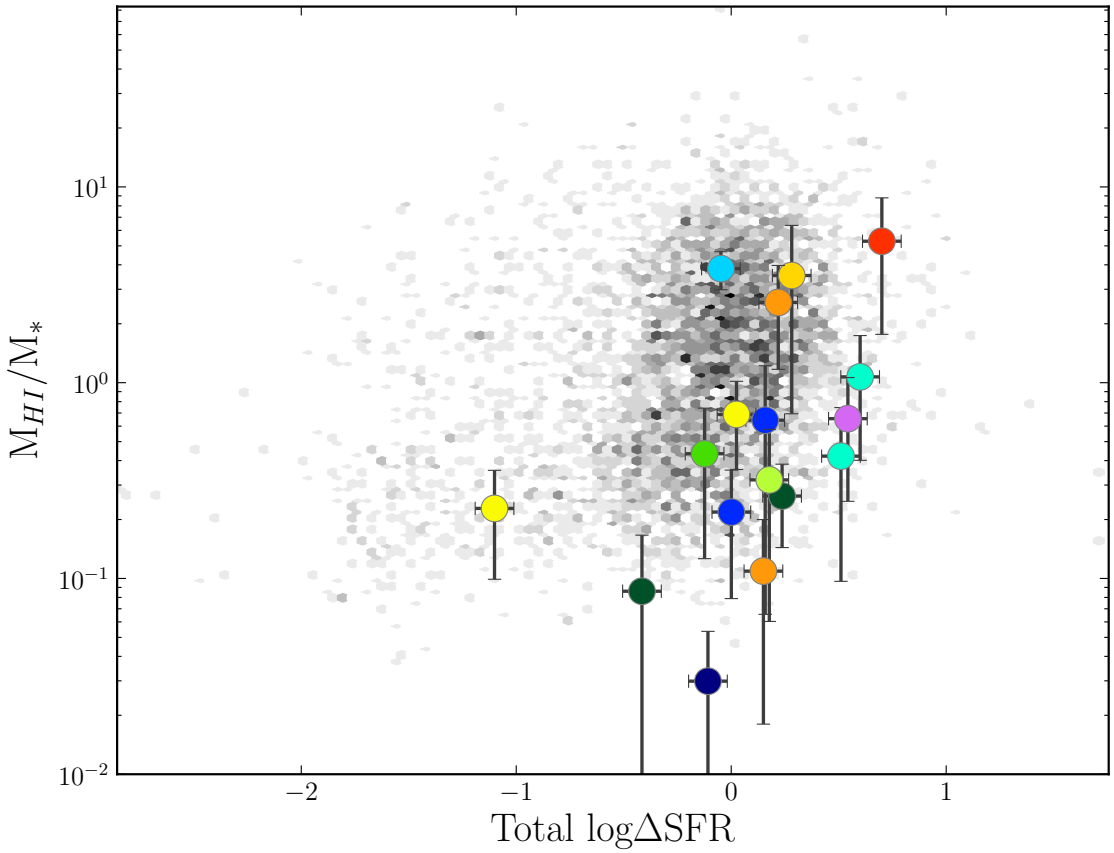


Figure 5.22: The log of the gas fraction versus stellar mass for both the ALFALFA & the VLA samples. The ALFALFA sample is shown as a 2-D grayscale histogram, where darker points indicate bins with a larger number of galaxies. Colour coding as in Figure 5.15

In general, it appears that the galaxies in the VLA sample sit to slightly higher total Δlog(SFR) values than the peak of the ALFALFA distribution. Given that these galaxy pairs are being compared to a sample of non-pairs, which have statistically

lower Δlog(SFR) values (as shown in Chapter 4), these higher Δlog(SFR) values are not surprising. Outliers to the lower edge of $\log(M_{HI}/M_*)$ are also expected given that the VLA sample is a much deeper investigation of gas masses.

5.10 Discussion & further work

The current results are suggestive of a positive correlation between the HI content of a galaxy and the level of triggered star formation within that galaxy. In §5.1, three conflicting physical models outlined by the theoretical studies of galaxy interactions were presented. Hopkins et al. (2009) suggested that the increasing gas fraction of a galaxy ought to lower the effectiveness with which torques between gaseous and stellar bars could funnel gas towards the centre of the galaxy during an interaction. Di Matteo et al. (2007) found that the gas fraction did not appear to be correlated with increasing starburst strength at all; the parameters of the orbit itself were more dominant in driving strong or weak enhancements in SFR, while in Perez & Sanchez-Blazquez (2011) and Bournaud et al. (2007), high gas fractions were found to be correlated with extremely strong starbursts when compared to lower gas fraction galaxies.

5.10.1 Comparison with & between simulations

The results of the work presented here show a 2σ positive correlation between increasing gas fraction and increasing starburst strength, which aligns well with the physical model presented by Perez & Sanchez-Blazquez (2011) and Bournaud et al. (2007). However, it is not immediately apparent why the physical models presented by the theoretical works of Hopkins et al. (2009), Di Matteo et al. (2007), or Perez & Sanchez-Blazquez (2011) come to such disparate conclusions about the importance of gas fraction. Direct comparisons with the existing simulations are difficult to put into practice, as these simulations were not designed with applications to the observations in mind; a brief discussion of the difficulties is summarized here.

The Hopkins et al. (2009) models

The Hopkins et al. (2009) work is primarily concerned with the gas fractions of galaxies as they proceed through coalescence. As the majority of the observed VLA galaxies are well separated, a comparison to the merger or post-merger stage is less applicable.

Figure 5.3 is, however, relevant, as this figure is for post-pericentre or flyby encounters only. However, the panels of Figure 5.3 seem to indicate that until very high gas fractions ($\sim$ 0.7) are reached, the dependence of the burst fraction on gas fraction is relatively weak. The models for full hydrodynamic simulations are constructed to have gas fractions among the following: f_{gas} = 0.05, 0.1, 0.2, 0.4, 0.6, 0.8, and 1.0, where gas fractions are defined here as in Equation 5.1. The dynamic range of the gas fractions presented in Figure 5.3 is quite large; an $f_{gas} = 0.1$ corresponds to $M_{HI}/M_* = 0.11$, and an $f_{gas} = 0.95$ corresponds to $M_{HI}/M_* = 19.0$.

The most extreme galaxy in the VLA sample has $f_{gas} = 0.83$ ($M_{HI}/M_* = 5.2$), but the majority of the VLA sample has $M_{HI}/M_* < 1.0$ (an f_{gas} of < 0.5), where the models predict a relatively weak dependence on gas fraction. More problematic is that the f_{gas} in Hopkins et al. (2009) refers to the gas fraction at the moment of the interaction, relative to the fraction of stars formed in a burst after the interaction. This use of multiple time steps in a single variable is impossible to compare with observations, which are only able to constrain the current star formation rate and current gas fraction.

The results of the hydrodynamic simulations of galaxies after first passage (for galaxies which will eventually merge) are plotted on these axes and show significant scatter of order $\sim$a few. It seems that the scatter at this stage of an interaction is canceled out by other factors later on in the merger, such that the scatter along this relation at coalescence is greatly reduced. This reduction in scatter is explained by stating that galaxies '...retaining more gas on first passage will yield a larger supply for the second burst...' (Hopkins et al., 2009). This comment is not elaborated upon, but seems to suggest a competing positive correlation within the timeframe of the merger itself.

The Di Matteo et al. (2007) models

There are four separate galaxy models used in the set of simulations presented in Di Matteo et al. (2007), and includes both elliptical and spiral galaxy models. All four models have a combined mass of star and dark matter particles totalling 2.3×10^{11} $M_\odot$. The elliptical model contains a gas fraction of 0.0, no stars in a disk, and splits its mass between a stellar bulge (70%) and dark matter halo (30%). The three disk galaxy models contain gas fractions of 0.1, 0.2, and 0.3 to model an Sa, Sb, and Sd type spiral galaxy respectively. However, in addition to varying the gas content of

these galaxies, the bulge to disk ratio is also changed between models. The Sa galaxy model is split between bulge, dark matter halo, and stellar disk components with 10%, 50%, and 40% respectively, with a gas mass of 9.2×10^9 $M_\odot$ in addition to the fixed stellar mass component of 2.3×10^{11} . In comparison, the Sd galaxy divides its stellar & dark matter mass into 0% bulge, 75% dark matter halo, and 25% stellar disk, while containing an additional 1.7×10^{10} $M_\odot$ in gas.

This morphological shift between galaxy models means that the stellar mass of the galaxy, the dominance of the bulge component, and the gas fraction are all changing between simulations. Since no distinction between late-type models is made in Figure 5.2, it is difficult to disentangle the starting gas fraction from the starting bulge fraction or stellar mass. Furthermore, while the gas content of the galaxy does not appear to correlate with the depletion of the gas since the start of the simulation in Figure 5.2, the dynamic range of gas fractions is still quite limited, as the starting gas fractions range from 0.1 to 0.3, a much narrower range than found either in the observational VLA sample of galaxy pairs or the ALFALFA sample of non-pairs, and also significantly narrower than the range of gas fractions presented in Hopkins et al. (2009).

The Perez & Sanchez-Blazquez (2011) & Bournaud et al. (2011) models

The work presented in Perez & Sanchez-Blazquez (2011) is primarily focused on the chemical evolution of interacting galaxies instead of on the star formation rates of those galaxies. As a result, only the raw SFR values are plotted as a function of the merger time, instead of making a comparison to the SFRs of an isolated galaxy. While it does appear that the magnitude of the burst is stronger from the baseline level of star formation in the high gas fraction case, this is difficult to quantify, and Perez & Sanchez-Blazquez (2011) comment only that that the bursts peak at higher values in the high gas fraction simulation. Perez & Sanchez-Blazquez (2011) only has two gas fractions in the simulations: the 'low' gas fraction of 0.2, and the 'high' gas fraction at 0.5. These gas fractions are calculated as the fraction of the total mass of the disk (gas + stars) is present as gas. Both high and low gas fraction galaxies were kept at a fixed baryonic mass ($\sim 5 \times 10^{10}$ $M_\odot$), which means that the high gas fraction galaxy had a lower stellar mass at $M_* \approx 3.2 \times 10^{10}$ $M_\odot$ than the low gas fraction galaxy, at $M_* \approx 4.4 \times 10^{10}$ $M_\odot$.

While this change to include a lower gas fraction in a high stellar mass galaxy

may be somewhat physically motivated, as at low redshift, high stellar mass galaxies tend to be detected at lower gas fractions (e.g., Figure 5.21), it also complicates the comparison between the high and low gas fractions difficult, as high mass galaxies are expected to show higher SFRs than their low mass counterparts, without the additional complication of the interaction. The normalization of the SFR in the merger to an undisturbed galaxy's star formation rate is especially critical for high gas fraction galaxies, where the gas disks can be unstable and begin to collapse even before the interaction takes place. Indeed, Perez & Sanchez-Blazquez (2011) notes that their gas-rich disk shows significant fragmentation of the gas disk well before the first encounter, and would occur even without the interaction.

The Bournaud et al. (2011) work cited as a reference for an explanation of the higher peak values of SFR in Perez & Sanchez-Blazquez (2011) is designed to model high redshift galaxies ($z \sim 2$). These galaxies have a stellar mass of 4×10^{10} $M_{\odot}$ and a fixed gas fraction of 70% (defined by Equation 5.1), which corresponds to $M_{HI}/M_{*} = 2.34$. This again is a higher gas fraction than the majority of the galaxies in the VLA sample. This work has a similar problem with regards to unstable disks as Perez & Sanchez-Blazquez (2011); a significant amount of star formation is triggered by disk instabilities unrelated to the interaction between the two galaxies. In spite of the instability of the high redshift disk prior to any interaction, Bournaud et al. (2011) find that the merger event triggers more massive and denser clumps, which contributes to stronger star formation in these galaxies. The enhancements are measured relative to an 'initial value', and are found to show enhancements of order a factor of ten higher. However, this initial value is not likely to be a representative measurement of the time evolution of the star formation of a high gas fraction galaxy, and as there is no direct comparison to a lower gas fraction galaxy, the conclusions drawn are qualitative in nature.

Model differences

Given that the existing theoretical models all investigate galaxy models of different stellar masses, and investigate gas fractions which often do not overlap with each other, and that the parameters used to quantify the impact of gas fraction are both very different between studies and often unobservable, it is extremely difficult to make a fair comparison between the simulation or between any given simulation and the observations. With the current status of the existing simulations, it is still possible to

hypothesize a scenario in which the results of the simulations are not in direct conflict with each other, but a comparison to the data will have to wait for an appropriately designed simulation.

For instance, it is possible that all three suggested physical models are in play at varying strengths during a merger. The burst efficiency drop presented in Hopkins et al. (2009) is most severe for galaxies which have extremely high gas fractions (approaching 100% gas). Even within the models of Hopkins et al. (2009), certain orbits appear to be more prone to maintaining high starburst fractions (e.g., $\sim$40% of the total amount of gas in the galaxy) across a wide range of gas fractions, whereas others have lower burst fractions ($\sim$20%). This scatter induced by varying the orbits of the merging galaxies corresponds well with the importance of orbital parameters proposed by Di Matteo et al. (2007). Given the statement in Hopkins et al. (2009) that a smaller burst at first passage will likely be made up for by a larger one at coalescence given the remaining gas available, clearly the gas content of the galaxy remaining matters, in accordance with the suggestions of Perez & Sanchez-Blazquez (2011) and the Bournaud et al. (2011) works. However, since the Perez & Sanchez-Blazquez (2011) and Bournaud et al. (2011) simulations are replicating extremely high redshift galaxies with physical parameters unlikely to exist in lower-redshift galaxies, the results of these simulations may not be directly applicable to the results of the VLA sample.

A speculation could be made that the example of high gas content lending itself well to higher star formation rates of Perez & Sanchez-Blazquez (2011) and Bournaud et al. (2011) could also dominate for galaxies which are not as turbulent as those simulated for the high redshift regime. For local galaxies such as those found in the current sample, the majority of them are found with gas masses less than their stellar masses. Within this range of gas fractions, the burst fraction efficiently will not yet have suffered significant degradation in the models of Hopkins et al. (2009), so sufficiently strong torques ought to be present to drive gas to the central regions of the galaxy. The orbital parameters of the merger can still introduce scatter along this broad correlation, as proposed in Di Matteo et al. (2007). Under these circumstances, the galaxies with higher gas fractions will indeed find themselves with a large, dense gas reservoir from which to form stars. Even if the efficiency with which the gas is driven to the central regions of the galaxy is slightly lower, if the increase in gas fraction can compensate for a small drop in efficiency, a larger number of stars can still be formed.

5.10.2 The Impact of Molecular Gas

As mentioned in Section 5.1, the adopted definition of gas fraction in this work does not include any contribution from the molecular gas component of a galaxy. However, the impact of the exclusion of molecular gas is unclear. In non-interacting galaxies, the relationship between the molecular hydrogen and neutral hydrogen content of a galaxy appears to be weak, if present. Saintonge et al. (2011) finds that, while on average the molecular gas mass of a galaxy is about 30% that of the neutral gas mass, there is a great deal of scatter around this average value, and the two quantities are, at best, only weakly correlated.

However, the behaviour of the molecular gas in interacting galaxies is still very unclear. It has been suggested that in galaxy mergers, the relationship between SFR and the molecular gas content is systematically changed such that interacting galaxies have significantly faster depletion times; i.e., they turn more gas into stars per unit time, when compared to their isolated counterparts (Daddi et al., 2010; Genzel et al., 2010). Data from the IRAM 30m telescope has been obtained[9] in order to test the dependence of SFR on the H_2 content of a sample of galaxy pairs, but until this data has been fully analyzed, it remains unclear how the molecular gas will correlate with the SFR enhancement of a galaxy. If the molecular gas content is anti-correlated with increasing $\Delta\log(\mathrm{SFR})$, or if the H_2 simply adds scatter to the total gas mass of a galaxy (i.e., $(M_{HI} + M_{H_2})/M_*$), then it is quite possible that the trend presented here will be appreciably reduced in significance.

5.10.3 Ideal simulation suite

The above physical picture will remain a speculation until a more consistent set of simulations can be developed for the purposes of comparison with the observations.

A suite of simulations which uses similar initial stellar masses and gas fractions to observed local galaxies would need to be used in order to be able to compare more directly with the observations. To isolate the influence of gas fraction, a set of isolated galaxy models of fixed stellar mass but with varying gas fraction could be developed such that in isolation the galaxies are stable. The star formation rates of these non-interacting galaxies would then be used as a benchmark for the interacting suite. The approach of Patton et al. (2013) was to use two galaxy models of different stellar mass such that their mass ratios were 2.5:1, which matches a typical inter-

[9]Proposal 216-11, PI: Kristin Coppin, Sara Ellison, Jillian Scudder

action in the SDSS. While the range of orbits probed by the simulations presented in Patton et al. (2013) are a good exploration of eccentricity, closest approach, and total energy of the encounter, for the purposes of examining gas fraction, a smaller suite of orbital encounters would be sufficient. Varying the impact parameter would result in varying strengths of a tidal interaction, which would help to disentangle the relative dominance of gas fraction versus the tidal interaction strength suggested in Di Matteo et al. (2007). Similarly, the relative galaxy orientations have been shown to drive significant changes in the star formation response of a galaxy.

Therefore, a suite of simulations which, at fixed mass, determines the star formation rates throughout an encounter for varying gas fractions, closest encounters, and relative galaxy orientations would be an ideal suite through which to compare the results of the simulated galaxy pairs with the results of observations. If 5 gas fractions, 3 closest approach distances, and 3 relative disk orientations are included, this results in a simulation suite of 45 interaction tracks.

Until a more consistent theoretical model is developed, either through a similar suite to the one suggested here, or through another approach, the interpretation of the physical processes underlying the results in this Chapter will remain hazy.

Chapter 6

Discussion & Summary

This thesis presents new results which interpret observational tracers of gas motions within interacting galaxies. A brief overview of the main results presented in this thesis, along with a short discussion of the implications of those results is undertaken in this chapter.

Galaxy pairs

The results presented in Chapter 4 on the changes induced in the star formation rates and metallicities of a set of galaxy pairs indicate that the previous physical picture of how galaxies react to local-scale perturbations was incomplete. Previous studies (e.g., Lambas et al., 2003; Alonso et al., 2006; Nikolic et al., 2004; Ellison et al., 2008b; Li et al., 2008) indicated that star formation rates returned back to normal after a relatively short period of time, as changes from the average galaxy were only detected over projected separations < 30 kpc h^{-1}. The physical picture derived from these results indicated that galaxies underwent a strong burst of star formation shortly after a close passage, which declined back to pre-interaction levels as the galaxies separated.

However, the results found here indicate that galaxies are influenced by an interaction over much larger timescales than was previously indicated. At the limits of the sample presented in Chapter 4 (80 kpc h^{-1}), galaxies still show significant enhancement over a mass, redshift, and environment-matched control sample of non-pair galaxies. The continuation of this enhancement beyond the limits of the current work prompted the study of Patton et al. (2013), in which a new pair selection algorithm was developed which could extend the sample out to arbitrarily large r_p.

In Patton et al. (2013), the influence of the pairwise interaction is followed out to separations of 2 Mpc, with a careful matching to the control sample on mass, redshift, distance to second nearest neighbour, and the number of neighbours within 2 Mpc. This matching criteria ensures that galaxies in pairs are matched to a control galaxy whose environment matches the paired galaxy as closely as possible, with the only difference being the lack of the companion galaxy in the pair. In spite of the new control matching methodology, the results presented in Patton et al. (2013) are consistent with those of Scudder et al. (2012b) in the range of projected separations where the samples overlap ($r_p < 80$ kpc h^{-1}). Patton et al. (2013) indicate that galaxies are affected by their interactions out to separations of 150 kpc h^{-1}. This result is a significant spatial extension of the work presented in Chapter 4. Both the results of Patton et al. (2013) and those of Chapter 4 indicate that the impact of a merger is that of a dramatic and enduring effect upon a galaxy.

Further studies have reconfirmed the extent of a galaxy's disturbance after an interaction; in re-inspecting the impact of a merger on the AGN fraction of the galaxy population, the range of separations over which a higher than expected AGN fraction was increased from 40 kpc h^{-1} (Ellison et al., 2011) to beyond 80 kpc h^{-1} (Ellison et al., 2013a). A study of the fraction of galaxies classified as Luminous Infra-Red Galaxies (LIRGs), which are a class of galaxy with infra-red luminosities exceeding 10^{11} $L_{\odot}$, Ellison et al. (2013b) find that the fraction of galaxies in interactions which are classified as LIRGs is higher than a sample of non-pairs over all separations < 80 kpc h^{-1}, although this fractional excess is highest at the smallest separations.

Since the study presented in Chapter 4 and in Patton et al. (2013) are limited to well-resolved pairs ($r_p \gtrsim 4$ kpc h^{-1}), the final stage of a merger is omitted from these studies. At the final stages, when the galaxies have combined to become a single object, theoretical models predict the most extreme tidal forces on the galaxy, resulting in extremely high star formation rates and very low nuclear metallicities (Torrey et al., 2012). To verify the theoretical predictions, Ellison et al. (2013a) extend the study of galaxy pairs inwards to galaxies which have recently coalesced by using a sample of galaxies selected as post-mergers in GalaxyZoo (Darg et al., 2010). These galaxies are selected to be highly morphologically disturbed but lacking a nearby companion. Ellison et al. (2013a) finds that the theoretical predictions are in fact borne out in the observations. Post-merger galaxies have strong star formation rate enhancements (a factor of 3.5 over the control galaxies) and prove to have very dilute gas-phase metallicities at -0.09 dex below the control. These galaxies are also

3.75 times more likely to contain an AGN than the control (non-pair) sample; the wider pairs sample peaks at an increase in the AGN fraction of 2.5.

The consequence of this recent body of work indicates that galaxy–galaxy interactions are very important events in a galaxy's lifetime, altering a wide range of observational properties. The galaxies that undergo a close tidal interaction show enhanced star formation rates (Scudder et al., 2012b; Patton et al., 2013; Ellison et al., 2013a), diluted metallicities (Scudder et al., 2012b), and trigger accretion onto a central supermassive black hole (Ellison et al., 2011, 2013a). Galaxies at separations of up to 150 kpc h^{-1} may only just be recovering from a previous close passage, instead of being well beyond the influence of the other galaxy, as would have previously been thought.

The newly determined extent of interaction-induced alterations in a galaxy also required a new physical model for the phase of the merger being observed at each separation. The physical model suggested in Scudder et al. (2012b) and further developed in Patton et al. (2013) indicates that these interactions are much more complex than had previously been assumed. While in previous studies it has been assumed that galaxies with SFR enhancements measured as a function of projected separation were measuring galaxies immediately prior to or after a close passage, the results of the more recent studies indicate that galaxies are likely to cross any given projected separation multiple times during an interaction, contributing different levels of star formation enhancement at each stage of an interaction. Galaxies showing strong star formation rate enhancements at very small separations are likely to be approaching final coalescence, whereas galaxies which are widely separated are likely to be at an earlier stage; either entering into the interaction for the first time, separating after a close pass, or returning for a second passage.

HI content of galaxy pairs

While the strength of an interacting galaxy's starburst is well known to correlate with the mass ratio of the two galaxies involved (Woods & Geller, 2007; Ellison et al., 2008b; Xu et al., 2010; Lambas et al., 2012), and the strongest starbursts are more likely to occur in galaxies found at small separations (Lambas et al., 2003; Alonso et al., 2006; Nikolic et al., 2004; Ellison et al., 2008b; Li et al., 2008), it has also been suggested that the gas fraction of the galaxies involved in the interaction should play a role in determining the strength of starburst that a galaxy is able to drive

after a perturbation (e.g., Cox et al., 2006). Previous studies have found that high redshift galaxies tend to show higher average star formation rates than local galaxies, which is generally ascribed to higher gas fractions, the additional *triggering* in star formation due to interactions has received relatively little study. The exception is Xu et al. (2012), which finds that galaxy pairs at higher redshifts do not show strong triggering in their SFRs, a phenomenon ascribed to the higher gas fractions at high redshift. However, as the results presented here show a correlation between high gas fraction and higher triggered star formation, the lack of triggered star formation in their high redshift sample must be due to some other physical process.

Historically, a study of the gas content of galaxies has been limited to either relatively shallow surveys such as the HI Parkes All-Sky Survey (e.g., Barnes et al., 2001), or in-depth mappings of very local galaxies (e.g., Peterson & Shostak, 1974; Smith, 1991, 1994; Hibbard & Yun, 1999; Hibbard et al., 2001; Torres-Flores et al., 2012). Relatively deep mapping of the HI content of a large number of galaxies has been difficult due to the time constraints. For example, in order to map a single galaxy's HI content, Smith (1991) required more than 8 hours of VLA time.

Several surveys are beginning to tackle this issue, such as THINGS (The HI Nearby Galaxy Survey, Walter et al., 2008). THINGS is a detailed HI survey of 34 local galaxies which was undertaken by the Very Large Array in multiple configurations in order to obtain high resolution images without losing sensitivity to flux on large scales. While THINGS does contain a single interacting galaxy, the majority of the sample is comprised of undisturbed, face-on spiral galaxies. A study of that single interacting galaxy undertaken in Holwerda et al. (2011) indicated that the morphology of the HI gas is a good indicator of disturbance from a previous interaction. The statistical limitations of the VLA data prohibit large-scale statistical comparisons to the SDSS.

Studies of the HI content of galaxies in mergers on a case-by-case basis have proven the existence of extremely long HI tidal tails in some interacting systems. These HI tails are not always accompanied by a strong optical counterpart, and can extend for many tens of kpc h^{-1}. Smith (1991) discovered an HI tail which contained 40% of the mass of the HI gas in the system which extended across $\sim$ 54 kpc h^{-1}. In extreme cases, the HI content of the galaxy can be spread over even wider regions of space; Hibbard & Yun (1999) reports the discovery of an HI tidal tail which spans 180 kpc h^{-1}. Another study found that the HI content of interacting systems could be entirely displaced from the optical counterpart. Sengupta et al. (2013) found the bulk of the HI content in the interaction located 70 kpc h^{-1} away from the galaxies .

The ALFALFA survey is also beginning to compile a deeper sample of galaxies than has been previously available. Since the beam size of the Arecibo telescope is 3.5 arcminutes, studies of galaxy pairs in close proximity to each other are limited to pairwise properties, where both galaxies are considered as a unit. Indeed, such work can be undertaken (e.g., Fertig et al., in preparation), but for a study of individual galaxy properties, only the widest pairs could be disentangled. As the wide pairs are not generally hosts to extreme starbursts (see Chapter 4), this does not help to understand what drives the starburst in cases where strong starbursts are expected. Furthermore, the ALFALFA survey is relatively shallow. By redshift $z \sim 0.06$, it is only sensitive to the highest gas fraction galaxies, which is significantly lower than the average redshift of the galaxies in Chapter 4.

The work in Chapter 5 is a first attempt to observationally quantify whether HI gas correlates with the starburst strength in a sample of locally selected galaxy pairs. A selection of the galaxies which are most likely to have strong enhancements in their star formation rates - with projected separations less than 30 kpc h^{-1}, and with mass ratios close to unity - does not cleanly select galaxies containing strong starbursts. The wide range of star formation rate enhancements remaining in the sample after selecting close galaxy pairs can be used to help probe the impact of gas fraction upon a galaxy's observed triggered SFR. The work presented in Chapter 5 finds that galaxies with higher gas fractions also tend to have higher star formation rate enhancements with 2σ confidence.

A full understanding of the physical picture behind these results will likely require the development of new simulations that more consistently test the effect of gas fraction upon an interacting galaxy's star formation rate. As a first pass, it seems that increasing the gas fraction within a local galaxy interaction can lead to higher levels of triggered star formation. It is unclear how extensions to the high redshift regime will complement this work.

In Chapter 5, a set of simulations which would be able to more adeptly disentangle the effects of gas fraction from other properties of the galaxy is outlined. A similar suite will hopefully be undertaken in the near future by theoretical colleagues. However, in the interim, Moreno et al. (in prep) are developing a set of galaxy interactions simulations that output spatial information about the galaxy via artificial Integral Field Unit (IFU) observations of the galaxy, projected on the sky. With the spatial information, it may be possible to determine if certain galaxies are particularly prone to star formation outside the central region, for instance. Some of the galaxies in the

JVLA sample were found to show very high levels of star formation outside the central region; if a certain property of the galaxy is more effective at driving large-scale star formation across the disk, a comparison with these simulations will be able to help isolate these factors.

The effect of environment

In spite of the dramatic effects of a local scale interaction on a galaxy, and the modulating effects of the gas fraction of a galaxy, the large scale environment in which a galaxy finds itself is also a force of change to a galaxy. A galaxy pair found within a large cluster potential will not show the same enhancements in star formation rate as a similar pair outside the cluster potential (Alonso et al., 2004, 2006; Baldry et al., 2006; Ellison et al., 2010).

The relative importance of the local and large scale environments in which a galaxy is found has been debated between works for some time (Hashimoto et al., 1998; Carter et al., 2001; Lewis et al., 2002; Goto et al., 2003; Kauffmann et al., 2004; Blanton et al., 2006; Blanton & Berlind, 2007; Park & Hwang, 2009). Both the local and the global environment have been found to be either dominant or entirely responsible for changes in the galaxy population between low galaxy density environments and high galactic density environments such as a cluster environment.

In the work presented in Chapter 3, a determination of the *relative* importance of local versus large scale environment is impossible to undertake. However, it is possible for a determination of whether or not the presence of a large scale environment drives any significant changes in a population of star forming galaxies in a consistent local environment. Since the local environment is held fixed in this work, comparisons with the impact of local scale effects are not feasible. The work in Chapter 3 indicates that the presence of a nearby, large scale overdensity does have a statistically significant impact upon a galaxy's star formation rate.

This indication of the importance of large scale structure indicates that galaxies, even if their local environments are consistent and relatively isolated from other galaxies, are, if found within a large-scale overdensity, likely to show lower star formation rates than galaxies in the same local environment but without the influence of large scale structure.

6.0.4 Summary of findings

This thesis has determined that galaxy–galaxy interactions are far more effective at driving long term changes in the internal gas motions of a galaxy than has previously been established. These gas motions result in higher than average star formation rates and lower than average metallicities, among other properties independently confirmed by other works.

Within the context of galaxy interactions, galaxies with the strongest star formation enhancements are also the galaxies with the highest gas fractions. Existing simulations indicate that this positive correlation has a straightforward physical explanation; higher gas fractions mean a higher volume of gas from which to form stars. However, the body of theoretical work thus far is extremely heterogeneous, and a more robustly constructed set of simulations will be required in order to more fully develop a physical understanding of this result.

This thesis also determined that the large scale environment in which a galaxy finds itself drives changes in the star formation rate of a galaxy. High density environments host star forming galaxies with statistically lower star formation rates than galaxies living in low-density environments.

6.1 Future work

The work presented in this thesis would be well suited to extensions in a number of directions. As stated in Chapter 5, a more robust comparison with theoretical works to explore the physical conditions in an interaction as a function of gas fraction is a natural expansion of the work presented here. An expansion of the work in both Chapter 4 and Chapter 5 would be to probe to higher redshifts, where gas fractions are expected to be higher, and star formation rates are higher, but the role that interactions play in triggering star formation is less clear.

Continuing the multi-wavelength element of this thesis, further studies using radio wavelength observations are also possible. For instance, as AGN are frequently seen to be radio loud, it may be possible to determine if radio-loud AGN appear preferentially at a certain stage (or r_p) of an interaction, or if they can be triggered equally efficiently at all stage of an interaction.

Integrated Field Spectroscopy (IFS) observations are becoming less time-intensive and more easily integrated into large scale surveys, an endeavour being undertaken by

both the Mapping Nearby Galaxies at Apache Point (MaNGA) and Sydney Australian Astronomical Observatory Multi-object Integral Field Spectrograph (SAMI) surveys. With the advent of IFS surveys, large volumes of data will be available to study the internal kinematics in much greater detail than has previously been possible. With such data, it may be possible to begin to disentangle the influence of AGN and star formation, by inspecting a spectrum from each spatial resolution element separately; alternately, gas motions within galaxy interactions could be traced directly through the kinematic data.

With increasing collaboration between theoretical and observational colleagues, the increasing importance of multi-wavelength studies, and the availability of new survey technologies, opportunities to address questions previously impossible are now feasible projects to undertake.

Bibliography

Abazajian, K., et al. 2003, AJ, 126, 2081

—. 2004, AJ, 128, 502

—. 2005, AJ, 129, 1755

Abazajian, K. N., et al. 2009, ApJS, 182, 543

Abramson, A., Kenney, J. D. P., Crowl, H. H., Chung, A., van Gorkom, J. H., Vollmer, B., & Schiminovich, D. 2011, 22

Adelman-McCarthy, J. K., et al. 2006, ApJS, 162, 38

—. 2007, ApJS, 172, 634

—. 2008, ApJS, 175, 297

Alonso, M. S., Lambas, D. G., Tissera, P., & Coldwell, G. 2006, MNRAS, 367, 1029

Alonso, M. S., Tissera, P. B., Coldwell, G., & Lambas, D. G. 2004, MNRAS, 352, 1081

Arp, H. 1966, Atlas of peculiar galaxies, ed. Arp, H.

Atkinson, A. M., Abraham, R. G., & Ferguson, A. M. N. 2013, ApJ, 765, 28

Baker, J. G., & Menzel, D. H. 1938, ApJ, 88, 52

Baldry, I. K., Balogh, M. L., Bower, R. G., Glazebrook, K., Nichol, R. C., Bamford, S. P., & Budavari, T. 2006, MNRAS, 373, 469

Baldry, I. K., Glazebrook, K., Brinkmann, J., Ivezić, Ž., Lupton, R. H., Nichol, R. C., & Szalay, A. S. 2004, ApJ, 600, 681

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

Balogh, M., et al. 2004, MNRAS, 348, 1355

Balogh, M. L., Morris, S. L., Yee, H. K. C., Carlberg, R. G., & Ellingson, E. 1997, ApJL, 488, L75

Balogh, M. L., Navarro, J. F., & Morris, S. L. 2000, ApJ, 540, 113

Balogh, M. L., Schade, D., Morris, S. L., Yee, H. K. C., Carlberg, R. G., & Ellingson, E. 1998, ApJL, 504, L75

Barnes, D. G., et al. 2001, MNRAS, 322, 486

Barnes, J. E. 1988, ApJ, 331, 699

—. 2004, MNRAS, 350, 798

Barnes, J. E., & Hernquist, L. 1996, ApJ, 471, 115

Barnes, J. E., & Hernquist, L. E. 1991, ApJL, 370, L65

Barton, E. J., Arnold, J. A., Zentner, A. R., Bullock, J. S., & Wechsler, R. H. 2007, ApJ, 671, 1538

Barton, E. J., Geller, M. J., & Kenyon, S. J. 2000, ApJ, 530, 660

Barton Gillespie, E., Geller, M. J., & Kenyon, S. J. 2003, ApJ, 582, 668

Bekki, K. 2009, MNRAS, 399, 2221

Bekki, K., Couch, W. J., & Shioya, Y. 2002, ApJ, 577, 651

Bell, E. F., et al. 2004, ApJ, 608, 752

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

Blanton, M. R., Eisenstein, D., Hogg, D. W., Schlegel, D. J., & Brinkmann, J. 2005, ApJ, 629, 143

Blanton, M. R., Eisenstein, D., Hogg, D. W., & Zehavi, I. 2006, ApJ, 645, 977

Blanton, M. R., Lin, H., Lupton, R. H., Maley, F. M., Young, N., Zehavi, I., & Loveday, J. 2003, AJ, 125, 2276

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

Bournaud, F., Elmegreen, B. G., & Elmegreen, D. M. 2007, ApJ, 670, 237

Bournaud, F., et al. 2008, A&A, 486, 741

—. 2011, ApJ, 730, 4

Brinchmann, J., Charlot, S., White, S. D. M., Tremonti, C., Kauffmann, G., Heckman, T., & Brinkmann, J. 2004, MNRAS, 351, 1151

Brooks, A. M., Governato, F., Booth, C. M., Willman, B., Gardner, J. P., Wadsley, J., Stinson, G., & Quinn, T. 2007, ApJL, 655, L17

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

Bruzual A., G., & Charlot, S. 1993, ApJ, 405, 538

Carter, B. J., Fabricant, D. G., Geller, M. J., Kurtz, M. J., & McLean, B. 2001, ApJ, 559, 606

Chabrier, G. 2003, PASP, 115, 763

Chandrasekhar, S. 1943, ApJ, 97, 255

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

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

Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q. F., Perley, R. A., Taylor, G. B., & Broderick, J. J. 1998, AJ, 115, 1693

Cooper, M. C., Gallazzi, A., Newman, J. A., & Yan, R. 2010, MNRAS, 402, 1942

Cooper, M. C., Tremonti, C. A., Newman, J. A., & Zabludoff, A. I. 2008, MNRAS, 390, 245

Couch, W. J., Balogh, M. L., Bower, R. G., Smail, I., Glazebrook, K., & Taylor, M. 2001, ApJ, 549, 820

Cox, T. J., Jonsson, P., Primack, J. R., & Somerville, R. S. 2006, MNRAS, 373, 1013

Cox, T. J., Jonsson, P., Somerville, R. S., Primack, J. R., & Dekel, A. 2008, MNRAS, 384, 386

Croton, D. J., Gao, L., & White, S. D. M. 2007, MNRAS, 374, 1303

Croton, D. J., et al. 2005, MNRAS, 356, 1155

Daddi, E., et al. 2010, ApJL, 714, L118

Darg, D. W., et al. 2010, MNRAS, 401, 1552

de Vaucouleurs, G. 1948, Annales d'Astrophysique, 11, 247

Di Matteo, P., Combes, F., Melchior, A.-L., & Semelin, B. 2007, A&A, 468, 61

D'Onghia, E., Vogelsberger, M., Faucher-Giguere, C.-A., & Hernquist, L. 2010, ApJ, 725, 353

Donzelli, C. J., & Pastoriza, M. G. 1997, ApJS, 111, 181

Dressler, A. 1980, ApJ, 236, 351

Elbaz, D., et al. 2011, A&A, 533, A119

Ellison, S. L., Mendel, J. T., Patton, D. R., & Scudder, J. M. 2013a, MNRAS, 435, 3627

Ellison, S. L., Mendel, J. T., Scudder, J. M., Patton, D. R., & Palmer, M. J. D. 2013b, MNRAS

Ellison, S. L., Patton, D. R., Mendel, J. T., & Scudder, J. M. 2011, MNRAS, 418, 2043

Ellison, S. L., Patton, D. R., Simard, L., & McConnachie, A. W. 2008a, ApJL, 672, L107

—. 2008b, AJ, 135, 1877

Ellison, S. L., Patton, D. R., Simard, L., McConnachie, A. W., Baldry, I. K., & Mendel, J. T. 2010, MNRAS, 407, 1514

Ellison, S. L., Simard, L., Cowan, N. B., Baldry, I. K., Patton, D. R., & McConnachie, A. W. 2009, MNRAS, 396, 1257

Fujita, Y. 1998, ApJ, 509, 587

Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider, D. P. 1996, AJ, 111, 1748

Genzel, R., et al. 2010, MNRAS, 407, 2091

Gómez, P. L., et al. 2003, ApJ, 584, 210

Goto, T., Yamauchi, C., Fujita, Y., Okamura, S., Sekiguchi, M., Smail, I., Bernardi, M., & Gomez, P. L. 2003, MNRAS, 346, 601

Groves, B., Brinchmann, J., & Walcher, C. J. 2012, MNRAS, 419, 1402

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

Haines, C. P., La Barbera, F., Mercurio, A., Merluzzi, P., & Busarello, G. 2006, ApJL, 647, L21

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

Haynes, M. P., et al. 2011, AJ, 142, 170

Hernquist, L. 1992, ApJ, 400, 460

Hibbard, J. E., van der Hulst, J. M., Barnes, J. E., & Rich, R. M. 2001, AJ, 122, 2969

Hibbard, J. E., & Yun, M. S. 1999, AJ, 118, 162

Hickson, P. 1982, ApJ, 255, 382

Holwerda, B. W., Pirzkal, N., de Blok, W. J. G., Bouchard, A., Blyth, S.-L., van der Heyden, K. J., & Elson, E. C. 2011, MNRAS, 416, 2401

Hopkins, P. F., et al. 2009, MNRAS, 397, 802

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

Hubble, E. P. 1926, ApJ, 64, 321

Ideue, Y., et al. 2012, ApJ, 747, 42

Kannappan, S. J. 2004, ApJL, 611, L89

Kannappan, S. J., & Fabricant, D. G. 2001, AJ, 121, 140

Kannappan, S. J., Jansen, R. A., & Barton, E. J. 2004, AJ, 127, 1371

Kapferer, W., Kronberger, T., Ferrari, C., Riser, T., & Schindler, S. 2008, MNRAS, 389, 1405

Kauffmann, G., White, S. D. M., Heckman, T. M., Ménard, B., Brinchmann, J., Charlot, S., Tremonti, C., & Brinkmann, J. 2004, MNRAS, 353, 713

Kauffmann, G., et al. 2003, MNRAS, 341, 33

Kauffmann, G., et al. 2003, MNRAS, 346, 1055

Kawata, D., & Mulchaey, J. S. 2008, ApJL, 672, L103

Kennicutt, R. C., & Evans, N. J. 2012, ARA&A, 50, 531

Kewley, L. J., & Dopita, M. A. 2002, ApJS, 142, 35

Kewley, L. J., & Ellison, S. L. 2008, ApJ, 681, 1183

Kewley, L. J., Geller, M. J., & Barton, E. J. 2006, AJ, 131, 2004

Kewley, L. J., Heisler, C. A., Dopita, M. A., & Lumsden, S. 2001, ApJS, 132, 37

Kewley, L. J., Rupke, D., Zahid, H. J., Geller, M. J., & Barton, E. J. 2010, ApJL, 721, L48

Kimm, T., et al. 2009, MNRAS, 394, 1131

Kobulnicky, H. A., & Kewley, L. J. 2004, ApJ, 617, 240

Kobulnicky, H. A., & Kewley, L. J. 2004, ApJ, 617, 240

Kormendy, J., & Kennicutt, Jr., R. C. 2004, ARA&A, 42, 603

Kroupa, P. 2001, MNRAS, 322, 231

Lambas, D. G., Alonso, S., Mesa, V., & O'Mill, A. L. 2012, A&A, 539, A45

Lambas, D. G., Tissera, P. B., Alonso, M. S., & Coldwell, G. 2003, MNRAS, 346, 1189

Lara-López, M. A., et al. 2010, AAP, 521, L53+

Larson, R. B. 1974, MNRAS, 169, 229

Larson, R. B., & Tinsley, B. M. 1978, ApJ, 219, 46

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

Lequeux, J., Peimbert, M., Rayo, J. F., Serrano, A., & Torres-Peimbert, S. 1979, A&A, 80, 155

Lewis, I., et al. 2002, MNRAS, 334, 673

Li, C., Kauffmann, G., Heckman, T. M., Jing, Y. P., & White, S. D. M. 2008, MNRAS, 385, 1903

Lin, L., et al. 2007, ApJL, 660, L51

Lotz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2008, MNRAS, 391, 1137

Lynden-Bell, D. 1967, MNRAS, 136, 101

Malin, D. F., & Carter, D. 1983, ApJ, 274, 534

Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, MNRAS, 408, 2115

Martínez, H. J., Zandivarez, A., Domínguez, M., Merchán, M. E., & Lambas, D. G. 2002, MNRAS, 333, L31

McConnachie, A. W., Patton, D. R., Ellison, S. L., & Simard, L. 2009, MNRAS, 395, 255

McGaugh, S. S. 1991, ApJ, 380, 140

McGee, S. L., Balogh, M. L., Wilman, D. J., Bower, R. G., Mulchaey, J. S., Parker, L. C., & Oemler, A. 2011, MNRAS, 413, 996

Mendel, J. T., Ellison, S. L., Simard, L., Patton, D. R., & McConnachie, A. W. 2011, MNRAS, 418, 1409

Mendel, J. T., Simard, L., Palmer, M., Ellison, S. L., & Patton, D. R. 2014, ApJS, 210, 3

Michel-Dansac, L., Lambas, D. G., Alonso, M. S., & Tissera, P. 2008, MNRAS, 386, L82

Mihos, J. C., Dubinski, J., & Hernquist, L. 1998, ApJ, 494, 183

Mihos, J. C., & Hernquist, L. 1994, ApJL, 425, L13

—. 1996, ApJ, 464, 641

Montuori, M., Di Matteo, P., Lehnert, M. D., Combes, F., & Semelin, B. 2010, A&A, 518, A56

Moore, B., Katz, N., Lake, G., Dressler, A., & Oemler, A. 1996, Nature, 379, 613

Moore, B., Lake, G., & Katz, N. 1998, ApJ, 495, 139

Mouhcine, M., Baldry, I. K., & Bamford, S. P. 2007, MNRAS, 382, 801

Mouhcine, M., Gibson, B. K., Renda, A., & Kawata, D. 2008, A&A, 486, 711

Moustakas, J., et al. 2011, ArXiv e-prints

Nikolic, B., Cullen, H., & Alexander, P. 2004, MNRAS, 355, 874

Noeske, K. G., et al. 2007a, ApJL, 660, L47

—. 2007b, ApJL, 660, L43

Pagel, B. E. J., Edmunds, M. G., Blackwell, D. E., Chun, M. S., & Smith, G. 1979, MNRAS, 95

Park, C., Choi, Y.-Y., Vogeley, M. S., Gott, III, J. R., & Blanton, M. R. 2007, ApJ, 658, 898

Park, C., & Hwang, H. S. 2009, ApJ, 699, 1595

Patiri, S. G., Prada, F., Holtzman, J., Klypin, A., & Betancort-Rijo, J. 2006, MNRAS, 372, 1710

Patton, D. R., & Atfield, J. E. 2008, ApJ, 685, 235

Patton, D. R., Carlberg, R. G., Marzke, R. O., Pritchet, C. J., da Costa, L. N., & Pellegrini, P. S. 2000, ApJ, 536, 153

Patton, D. R., Ellison, S. L., Simard, L., McConnachie, A. W., & Mendel, J. T. 2011, MNRAS, 412, 591

Patton, D. R., Torrey, P., Ellison, S. L., Mendel, J. T., & Scudder, J. M. 2013, MNRAS, 433, L59

Pei, Y. C. 1992, ApJ, 395, 130

Peng, Y., et al. 2010, ApJ, 721, 193

Peng, Y.-j., Lilly, S. J., Renzini, A., & Carollo, M. 2012, ApJ, 757, 4

Perez, I., & Sanchez-Blazquez, P. 2011

Perez, J., Michel-Dansac, L., & Tissera, P. B. 2011, MNRAS, 417, 580

Perez, J., Tissera, P., & Blaizot, J. 2009, MNRAS, 397, 748

Peterson, S. D., & Shostak, G. S. 1974, AJ, 79, 767

Pimbblet, K. A., Smail, I., Kodama, T., Couch, W. J., Edge, A. C., Zabludoff, A. I., & O'Hely, E. 2002, MNRAS, 331, 333

Poggianti, B. M., et al. 2008, ApJ, 684, 888

Postman, M., & Geller, M. J. 1984, ApJ, 281, 95

Rasmussen, J., Ponman, T. J., Verdes-Montenegro, L., Yun, M. S., & Borthakur, S. 2008, MNRAS, 388, 1245

Robaina, A. R., et al. 2009, ApJ, 704, 324

Rossa, J., Laine, S., van der Marel, R. P., Mihos, J. C., Hibbard, J. E., Böker, T., & Zabludoff, A. I. 2007, AJ, 134, 2124

Rupke, D. S. N., Kewley, L. J., & Barnes, J. E. 2010a, ApJL, 710, L156

Rupke, D. S. N., Kewley, L. J., & Chien, L.-H. 2010b, ApJ, 723, 1255

Saintonge, A. 2007, AJ, 133, 2087

Saintonge, A., et al. 2011, MNRAS, 415, 32

Salim, S., et al. 2007, ApJS, 173, 267

Scudder, J. M., Ellison, S. L., & Mendel, J. T. 2012a, MNRAS, 423, 2690

Scudder, J. M., Ellison, S. L., Torrey, P., Patton, D. R., & Mendel, J. T. 2012b, MNRAS, 426, 549

Sengupta, C., Dwarakanath, K. S., Saikia, D. J., & Scott, T. C. 2013, MNRAS, 431, L1

Simard, L., Mendel, J. T., Patton, D. R., Ellison, S. L., & McConnachie, A. W. 2011, ApJS, 196, 11

Skibba, R. A., et al. 2009, MNRAS, 399, 966

Skillman, E. D., Kennicutt, Jr., R. C., Shields, G. A., & Zaritsky, D. 1996, ApJ, 462, 147

Smith, B. J. 1991, ApJ, 378, 39

—. 1994, AJ, 107, 1695

Smith, J. A., et al. 2002, AJ, 123, 2121

Springel, V. 2005, MNRAS, 364, 1105

Stasińska, G., Cid Fernandes, R., Mateus, A., Sodré, L., & Asari, N. V. 2006, MNRAS, 371, 972

Stoughton, C., et al. 2002, AJ, 123, 485

Strauss, M. A., et al. 2002, AJ, 124, 1810

Strömgren, B. 1939, ApJ, 89, 526

Tago, E., Saar, E., Tempel, E., Einasto, J., Einasto, M., Nurmi, P., & Heinämäki, P. 2010, A&A, 514, A102+

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

Thompson, A. R. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 180, Synthesis Imaging in Radio Astronomy II, ed. G. B. Taylor, C. L. Carilli, & R. A. Perley, 11

Tinker, J., Wetzel, A., & Conroy, C. 2011, ArXiv e-prints

Tonnesen, S., & Bryan, G. L. 2012, MNRAS, 422, 1609

Toomre, A. 1977, in Evolution of Galaxies and Stellar Populations, ed. B. M. Tinsley & R. B. G. Larson D. Campbell, 401

Toomre, A., & Toomre, J. 1972, ApJ, 178, 623

Torres-Flores, S., de Oliveira, C. M., de Mello, D. F., Scarano, S., & Urrutia-Viscarra, F. 2012, MNRAS, 421, 3612

Torrey, P., Cox, T. J., Kewley, L., & Hernquist, L. 2012, ApJ, 746, 108

Tremonti, C. A., et al. 2004, ApJ, 613, 898

Veilleux, S., & Osterbrock, D. E. 1987, ApJS, 63, 295

Walter, F., Brinks, E., de Blok, W. J. G., Bigiel, F., Kennicutt, Jr., R. C., Thornley, M. D., & Leroy, A. 2008, AJ, 136, 2563

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

Welikala, N., Connolly, A. J., Hopkins, A. M., Scranton, R., & Conti, A. 2008, ApJ, 677, 970

Wetzel, A. R., Tinker, J. L., & Conroy, C. 2012, MNRAS, 424, 232

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

Whitmore, B. C., & Gilmore, D. M. 1991, ApJ, 367, 64

Wong, K. C., et al. 2011, ApJ, 728, 119

Woods, D. F., & Geller, M. J. 2007, AJ, 134, 527

Woods, D. F., Geller, M. J., & Barton, E. J. 2006, AJ, 132, 197

Woods, D. F., et al. 2010, AJ, 139, 1857

Wuyts, S., et al. 2011, ApJ, 742, 96

Xu, C. K., & Sulentic, J. W. 1991, ApJ, 374, 407

Xu, C. K., et al. 2010, ApJ, 713, 330

—. 2012, ApJ, 760, 72

Yates, R. M., Kauffmann, G., & Guo, Q. 2012, MNRAS, 422, 215

York, D. G., et al. 2000, AJ, 120, 1579

Zaritsky, D., Kennicutt, Robert C., J., & Huchra, J. P. 1994, ApJ, 420, 87

Zheng, X. Z., Bell, E. F., Papovich, C., Wolf, C., Meisenheimer, K., Rix, H.-W., Rieke, G. H., & Somerville, R. 2007, ApJL, 661, L41

Appendix A

Additional Information

In the Appendix, additional tables and figures are included for completeness purposes. Tables A.1 and A.2 contain information on the embedded and isolated CG galaxy samples (respectively) presented in Chapter 3.

Figures A.1 through A.34 show the full set of SDSS thumbnails and spectra for the sample presented in Chapter 5.

SDSS ObjID	CG ObjID	Redshift	Stellar Mass	Metallicity	SFR
		z	$\log(M_*)$ ($M_\odot$)	12+ log(O/H)	log(SFR) ($M_\odot$ yr^{-1})
587735348567736612	SDSSCGA00374.4	0.066	9.70	8.97	−0.23
587732470923591912	SDSSCGA02222.4	0.089	9.91	8.96	0.22
587742551749427305	SDSSCGA01310.2	0.072	10.09	9.09	0.58
587739629557907538	SDSSCGA00209.3	0.063	9.78	9.00	0.51
587736525909393590	SDSSCGA00436.4	0.109	10.79	9.29	1.59
587742614558212294	SDSSCGA00213.1	0.127	10.85	9.23	1.29
587731499184423092	SDSSCGA02083.4	0.050	9.52	8.96	−0.43
588017702397476999	SDSSCGA01110.2	0.067	10.02	9.04	0.34
587741600963297361	SDSSCGA00933.2	0.059	10.46	8.99	0.29
587729157905514683	SDSSCGA01012.4	0.069	9.64	9.01	−0.05
587736619327553911	SDSSCGA01186.2	0.089	10.13	9.03	0.61
587744727687954659	SDSSCGA01185.4	0.092	10.07	9.20	0.31
587729160048148639	SDSSCGA01232.2	0.091	9.88	8.94	0.38
587742551749427435	SDSSCGA01310.3	0.071	9.51	8.91	0.03
587742062159593500	SDSSCGA00253.2	0.077	10.15	9.17	0.72
587739648886505610	SDSSCGA00838.4	0.133	10.26	9.16	0.60
587741421632356738	SDSSCGA00979.4	0.051	9.59	8.99	−0.42
588017116128870540	SDSSCGA01825.1	0.071	10.04	8.96	0.62
588848901523439828	SDSSCGA01598.4	0.082	10.44	9.17	0.33
587729388223201436	SDSSCGA01434.3	0.035	9.80	9.11	−0.34
588017116128870551	SDSSCGA01825.2	0.071	10.22	9.24	0.42
587731891650494559	SDSSCGA02027.5	0.065	9.73	9.04	−0.06

Table A.1 – continued from previous page

SDSS ObjID	CG ObjID	Redshift	Stellar Mass	Metallicity	SFR
587732578296529206	SDSSCGA00435.4	0.048	9.61	9.05	−0.46
587734894367539290	SDSSCGA01858.4	0.095	10.00	9.05	0.45
587741489294475468	SDSSCGA01272.3	0.042	9.17	8.71	−0.83
587738372745855292	SDSSCGA01487.3	0.096	9.98	9.00	0.47
587734894367539289	SDSSCGA01858.3	0.095	10.00	9.03	0.52
588017721713623164	SDSSCGA00494.1	0.082	10.69	9.31	0.96
587726033854398529	SDSSCGA01063.3	0.078	10.48	9.23	1.03
587745403074248994	SDSSCGA02266.3	0.116	10.32	9.05	0.48
588017979961901100	SDSSCGA00904.4	0.067	9.99	9.10	0.08
587739609162317939	SDSSCGA02209.3	0.078	9.72	8.97	0.17
587742060554879300	SDSSCGA01274.3	0.085	10.08	9.01	0.26
587736542020042979	SDSSCGA01833.2	0.059	10.10	9.12	0.02
587735696987586726	SDSSCGA01070.2	0.083	10.65	8.86	−0.44
587736584977711632	SDSSCGA02108.1	0.063	9.87	9.13	0.13
588017721713688616	SDSSCGA00494.3	0.082	9.93	9.01	0.23
587742060554944652	SDSSCGA01274.4	0.082	10.66	9.07	0.74
587735696987586734	SDSSCGA01070.4	0.082	10.38	9.31	0.58
587741600963297386	SDSSCGA00933.4	0.059	9.97	9.11	0.10
587731869627252855	SDSSCGA02011.4	0.140	10.31	9.15	0.85
587733429770846395	SDSSCGA01195.2	0.073	10.70	9.29	0.87
587738065653334220	SDSSCGA01301.2	0.077	10.20	9.14	0.65
587741490354389310	SDSSCGA01109.4	0.091	10.08	9.13	0.43

Table A.1 – continued from previous page

SDSS ObjID	CG ObjID	Redshift	Stellar Mass	Metallicity	SFR
588023722317840542	SDSSCGA00728.4	0.044	9.30	8.91	−0.41
588023722317840548	SDSSCGA00728.3	0.044	9.90	9.07	−0.29
587745403074314378	SDSSCGA02266.4	0.116	10.04	9.10	0.22
587733429770911869	SDSSCGA01195.4	0.074	9.56	8.91	0.05
588016891713880190	SDSSCGA01137.3	0.056	9.64	9.11	−0.22
587735348025819263	SDSSCGA01108.4	0.054	9.81	9.15	0.06

Table A.1: Embedded Compact Group galaxy properties. SDSS Objid is the unique identifier for that galaxy within Sloan, CG Objid is the Galaxy ID from the M09 catalogue where numbers before the decimal indicate the group ID, and the number after the decimal varies between galaxies within one group. Metallicities calculations are described in §2.5. SFR values come from the calculations of Brinchmann et al. (2004).

SDSS ObjID	CG ObjID	Redshift	Stellar Mass	Metallicity	SFR
		z	$\log(M_*)$ ($M_\odot$)	12+ log(O/H)	log(SFR) ($M_\odot$ yr^{-1})
587739844854939883	SDSSCGA01177.2	0.114	10.61	9.10	0.82
588023720707424428	SDSSCGA01088.4	0.061	10.11	8.98	−0.51
587736975813836873	SDSSCGA00878.2	0.030	9.97	9.22	0.72
587739405707116711	SDSSCGA01128.1	0.054	10.48	9.04	0.41
587732471457186108	SDSSCGA00146.1	0.041	10.24	9.21	0.13
587725552264479032	SDSSCGA01533.2	0.066	10.16	9.18	0.35
587726101490827509	SDSSCGA00370.4	0.085	10.25	9.23	0.57
587725980687925438	SDSSCGA02231.4	0.097	10.27	9.18	0.50
588017724946972875	SDSSCGA00942.4	0.083	9.79	8.87	−0.07
587735344792928602	SDSSCGA00422.5	0.046	9.23	8.92	−0.48
587732702864080970	SDSSCGA02097.1	0.052	9.67	8.87	0.39
587736543623446530	SDSSCGA00970.4	0.087	9.89	8.97	0.57
587725041698799708	SDSSCGA01076.4	0.063	9.61	8.87	0.12
587725041698799706	SDSSCGA01076.3	0.064	9.62	8.87	0.18
587729386610557106	SDSSCGA01836.3	0.072	10.08	9.09	−0.41
587739377773576383	SDSSCGA00513.3	0.079	10.13	9.19	0.49
587728930803744828	SDSSCGA00362.1	0.112	10.05	9.06	0.74
587745244704473323	SDSSCGA01087.3	0.045	9.84	9.03	−0.19
587742062678900857	SDSSCGA01309.3	0.068	9.69	8.80	0.07
587745245240295524	SDSSCGA00601.3	0.113	10.36	9.20	0.48
587739827674022085	SDSSCGA00041.2	0.072	10.39	9.23	0.82
587729386068705391	SDSSCGA00078.3	0.056	9.37	8.84	0.01

Table A.2 – continued from previous page

SDSS ObjID	CG ObjID	Redshift	Stellar Mass	Metallicity	SFR
587739844310007987	SDSSCGA01088.3	0.061	9.93	9.02	0.06
587731886268088671	SDSSCGA01783.2	0.078	10.75	9.25	1.16
587745539432776118	SDSSCGA02092.1	0.052	9.48	8.80	−0.11
587742774022045799	SDSSCGA01953.4	0.104	9.88	8.99	0.34
587742550686368002	SDSSCGA00687.2	0.087	10.18	9.19	0.34
588017627221721269	SDSSCGA01937.2	0.074	10.25	9.07	0.64
588848899373072598	SDSSCGA00713.3	0.118	10.50	9.01	0.50
588017990162710837	SDSSCGA00718.2	0.139	10.84	9.32	1.11
587732483824746731	SDSSCGA00305.2	0.084	9.76	8.88	0.25
587742013820305436	SDSSCGA01570.2	0.106	9.95	8.90	0.84
587742577521000590	SDSSCGA01809.3	0.093	10.33	9.19	0.48
588017627221721251	SDSSCGA01937.3	0.074	10.09	9.11	0.53
588007006048485395	SDSSCGA00936.1	0.074	10.19	9.07	0.80
587736975813836874	SDSSCGA00878.3	0.031	9.71	9.16	−0.50
587738947198910752	SDSSCGA01170.2	0.043	9.71	8.94	−0.19
587734893827129437	SDSSCGA00945.1	0.060	10.56	9.16	0.79
587741720678826052	SDSSCGA00562.2	0.067	9.50	8.90	0.39
587726101490827456	SDSSCGA00370.3	0.086	10.20	9.00	0.56
587725469057351815	SDSSCGA01097.1	0.132	11.09	8.89	0.87
587736782540964307	SDSSCGA01542.3	0.124	10.69	9.27	0.76
587739115237015957	SDSSCGA01970.3	0.099	9.74	8.88	0.21
587739610230554751	SDSSCGA00820.2	0.073	10.20	9.09	0.37

Table A.2 – continued from previous page

SDSS ObjID	CG ObjID	Redshift	Stellar Mass	Metallicity	SFR
588017627221721162	SDSSCGA01937.4	0.074	10.40	9.03	1.03
588848900472111426	SDSSCGA01473.2	0.076	10.11	9.18	0.48
587739405707116699	SDSSCGA01128.2	0.054	10.46	9.08	0.67
588848899392340185	SDSSCGA00476.3	0.093	10.44	9.21	0.62
587729407546228763	SDSSCGA02056.3	0.029	9.82	9.02	−0.46
587739811028992231	SDSSCGA01347.2	0.068	9.57	8.93	0.23
587722984441119019	SDSSCGA01703.2	0.120	10.36	9.13	0.74
587722984441119010	SDSSCGA01703.3	0.121	10.68	9.27	0.72
587736546852864070	SDSSCGA00337.4	0.069	9.79	9.01	0.49
587737810644697402	SDSSCGA01583.3	0.103	9.81	8.95	0.46
587742578054201471	SDSSCGA01218.3	0.099	10.53	9.20	0.52
588016878825701672	SDSSCGA00227.3	0.082	10.12	9.05	0.71
587739610230554752	SDSSCGA00820.3	0.072	9.91	8.92	0.35
588009368008261825	SDSSCGA02225.5	0.092	10.13	9.04	0.44
587738411408294053	SDSSCGA00614.3	0.079	10.00	8.93	0.33
587742013297131696	SDSSCGA01348.3	0.102	9.96	8.96	0.54
587741603095707807	SDSSCGA00610.3	0.081	10.08	8.99	0.19
587742774555508821	SDSSCGA01132.3	0.123	10.22	9.14	0.59

Table A.2: Same as Table A.1, but for the isolated CG galaxies

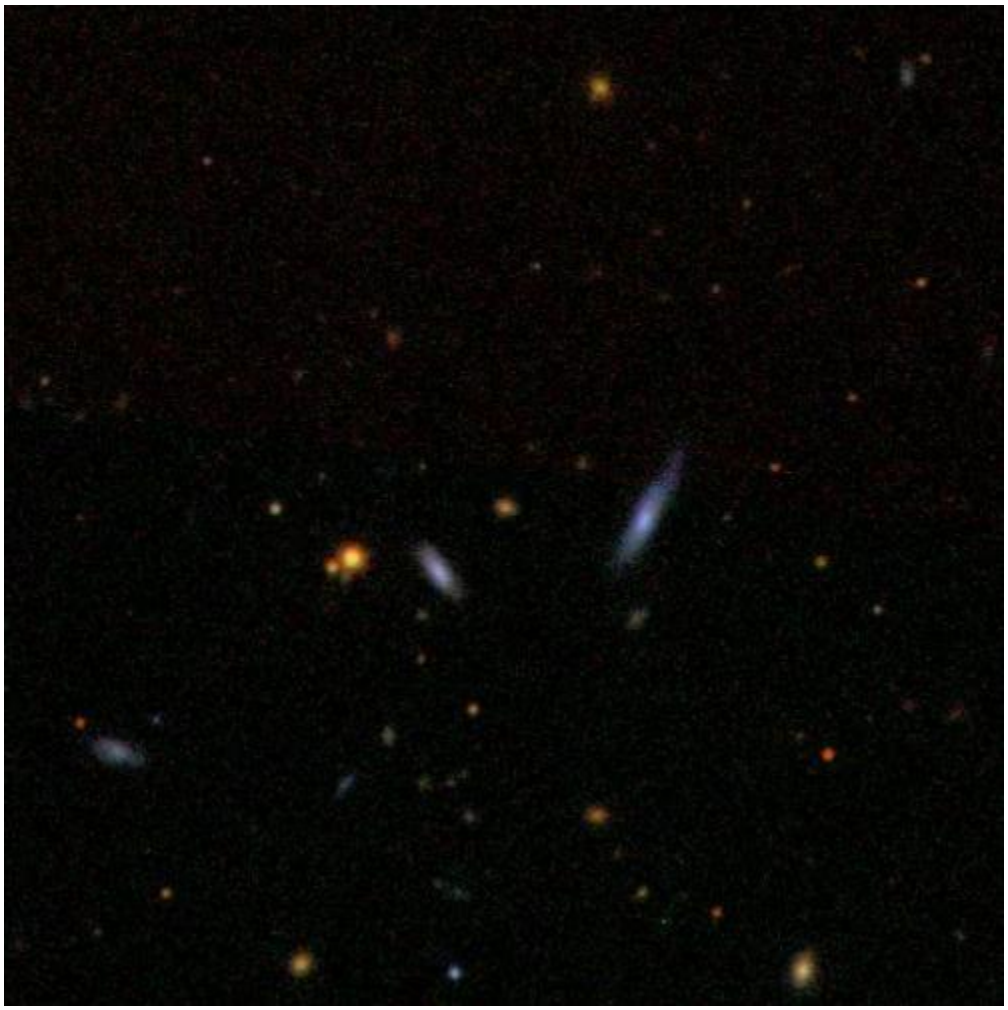


Figure A.1: SDSS Thumbnail of galaxy pair 587727178473930875 (right) & 587727178473930886 (left).

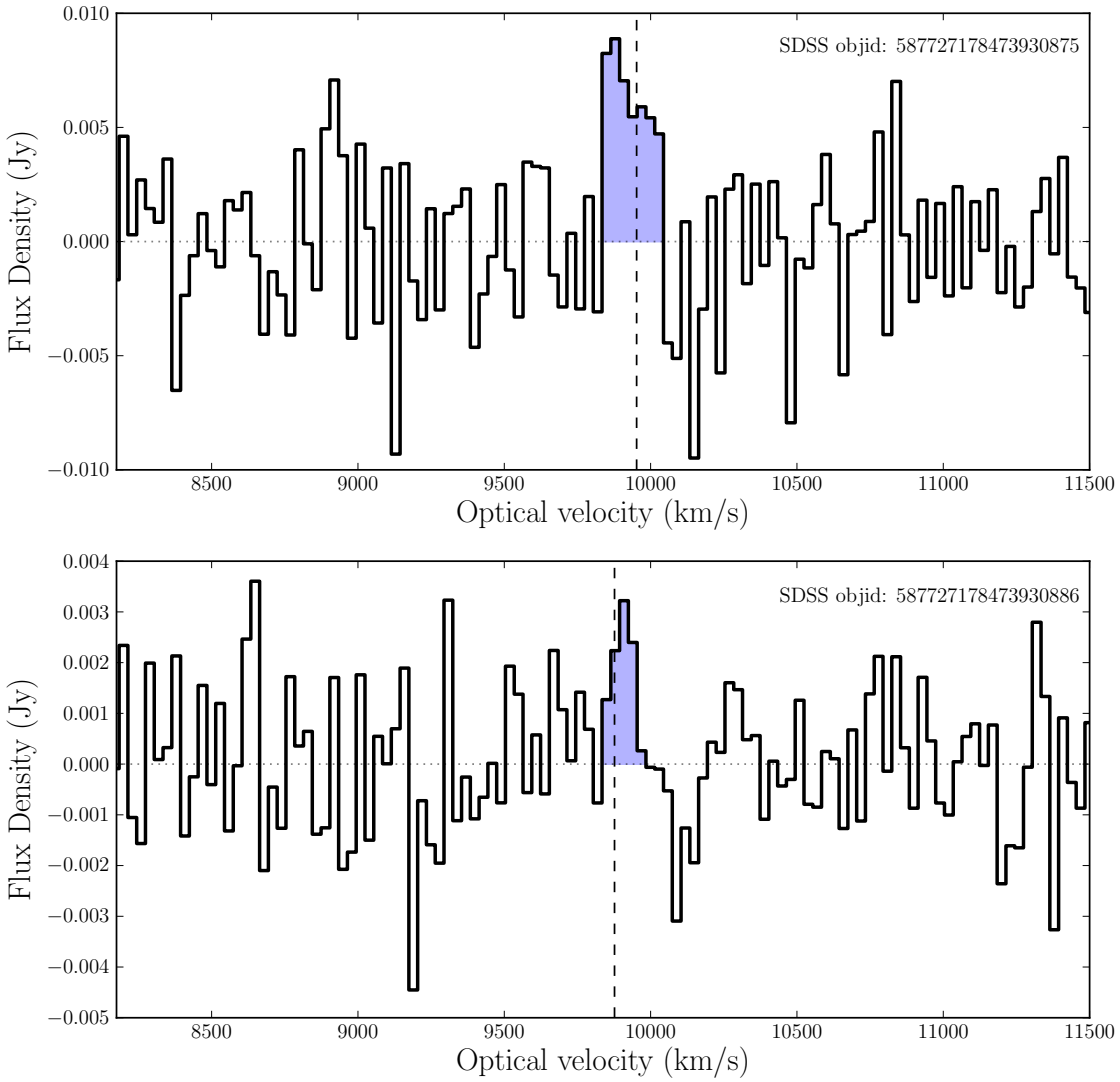


Figure A.2: Flux density vs. optical velocity for galaxy pair 587727178473930875 & 587727178473930886. The vertical dashed line indicates the velocity of the galaxy, horizontal dotted line indicates 0 flux. The shaded region indicates the region of the spectra used for signal to noise and gas mass calculations. The SDSS image for this pair is presented in Figure A.1. The upper spectrum has a peak/RMS S/N of 2.62, while the lower spectrum has a peak/RMS S/N of 2.24. SDSS image presented in Figure A.1.

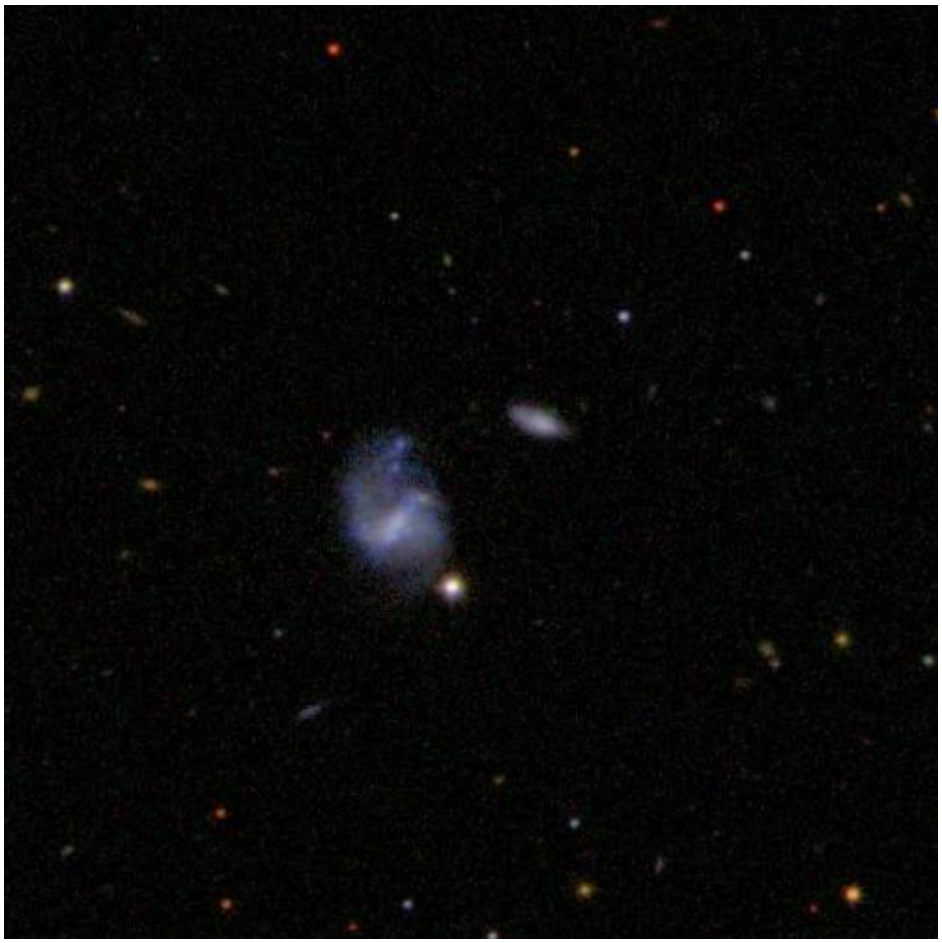


Figure A.3: SDSS Thumbnail of galaxy pair 588017702411763744 (bottom left) & 588017702411763872 (top right).

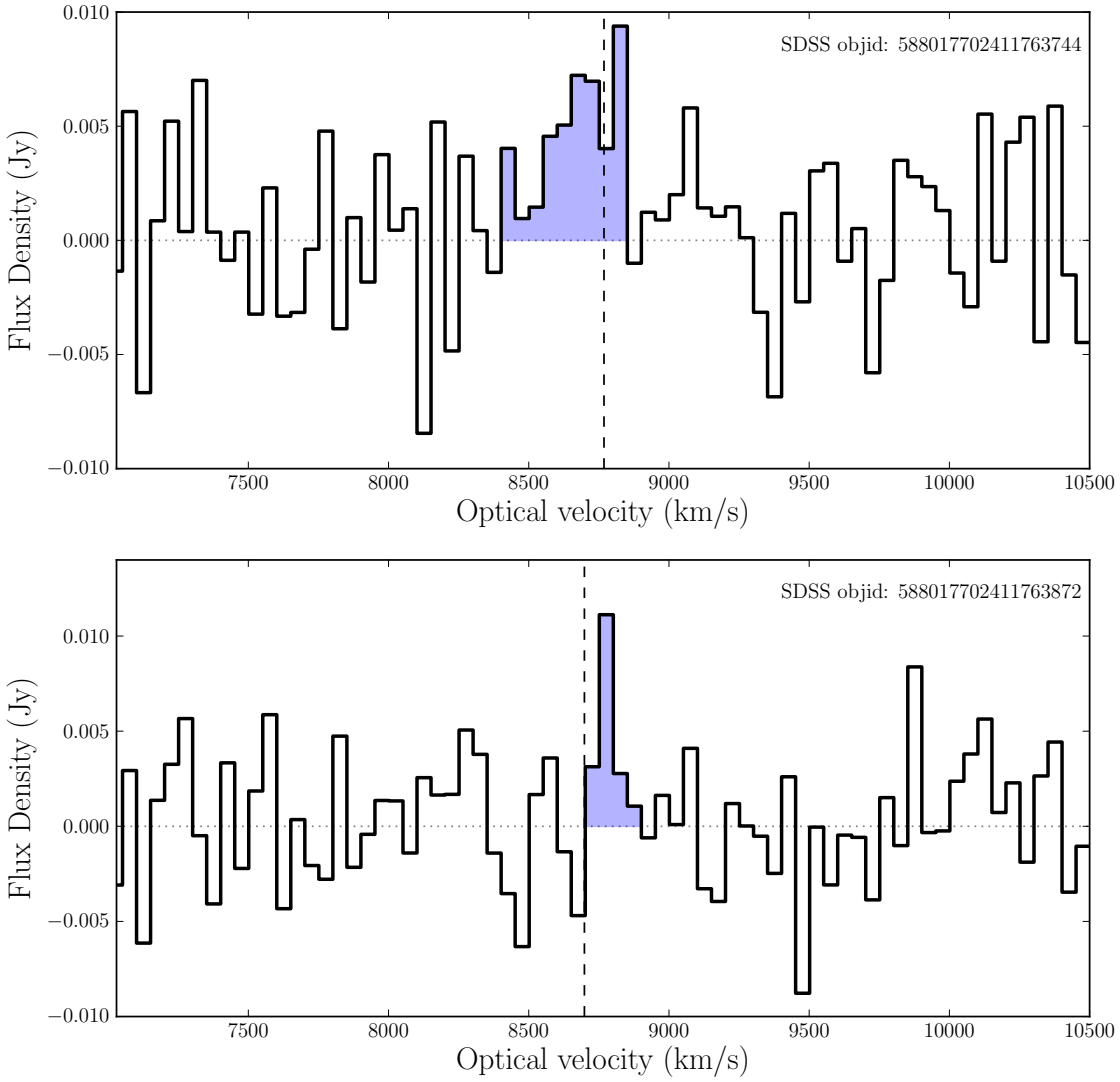


Figure A.4: Same as in Figure A.2; Flux density vs. optical velocity for galaxy pair 588017702411763744 & 588017702411763872. The upper spectrum has a peak/RMS S/N of 2.67, while the lower spectrum has a peak/RMS S/N of 2.79. The SDSS image for this pair is presented in Figure A.3.

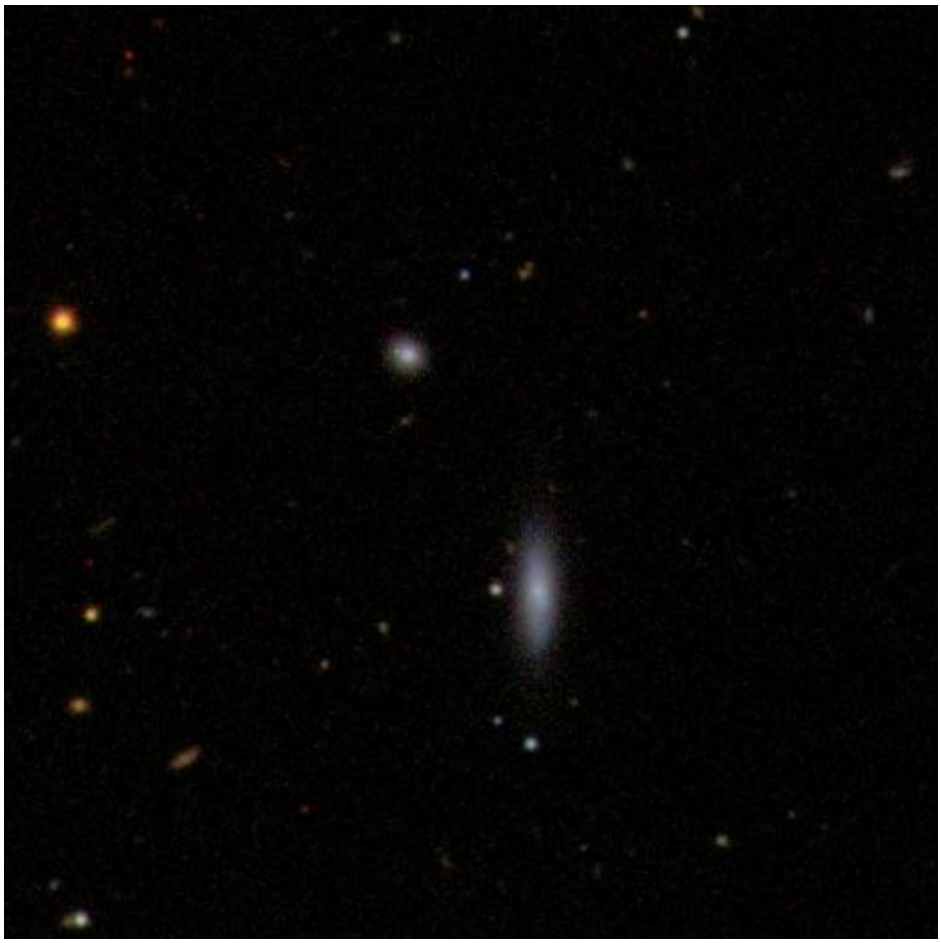


Figure A.5: SDSS Thumbnail of galaxy pair 587739303684866183 (top left) & 587739303684866173 (bottom right).

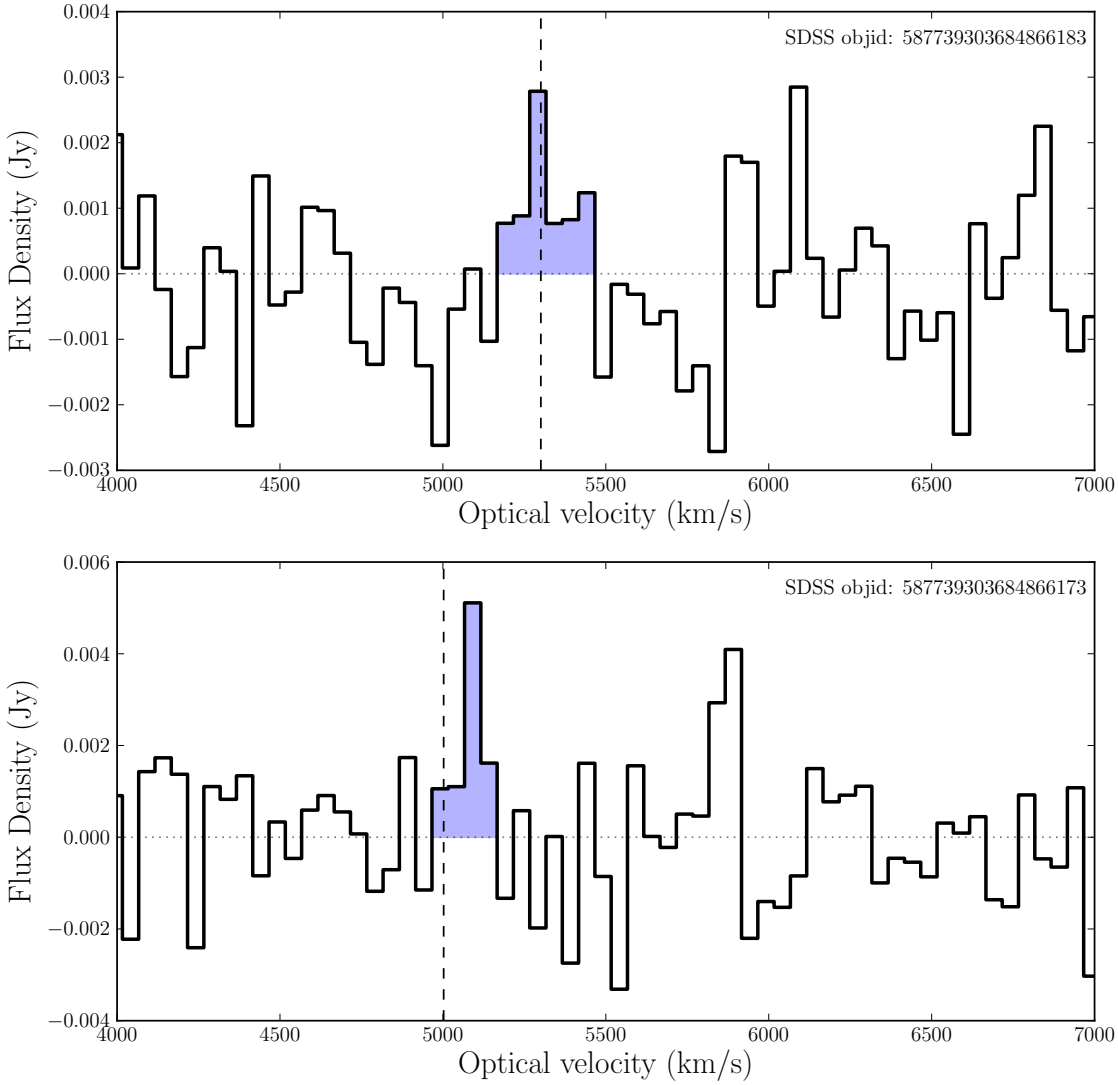


Figure A.6: Same as in Figure A.2; Flux density vs. optical velocity for galaxy pair 587739303684866183 & 587739303684866173. The upper spectrum has a peak/RMS S/N of 2.35, while the lower spectrum has a peak/RMS S/N of 3.57. The SDSS image for this pair is presented in Figure A.5.

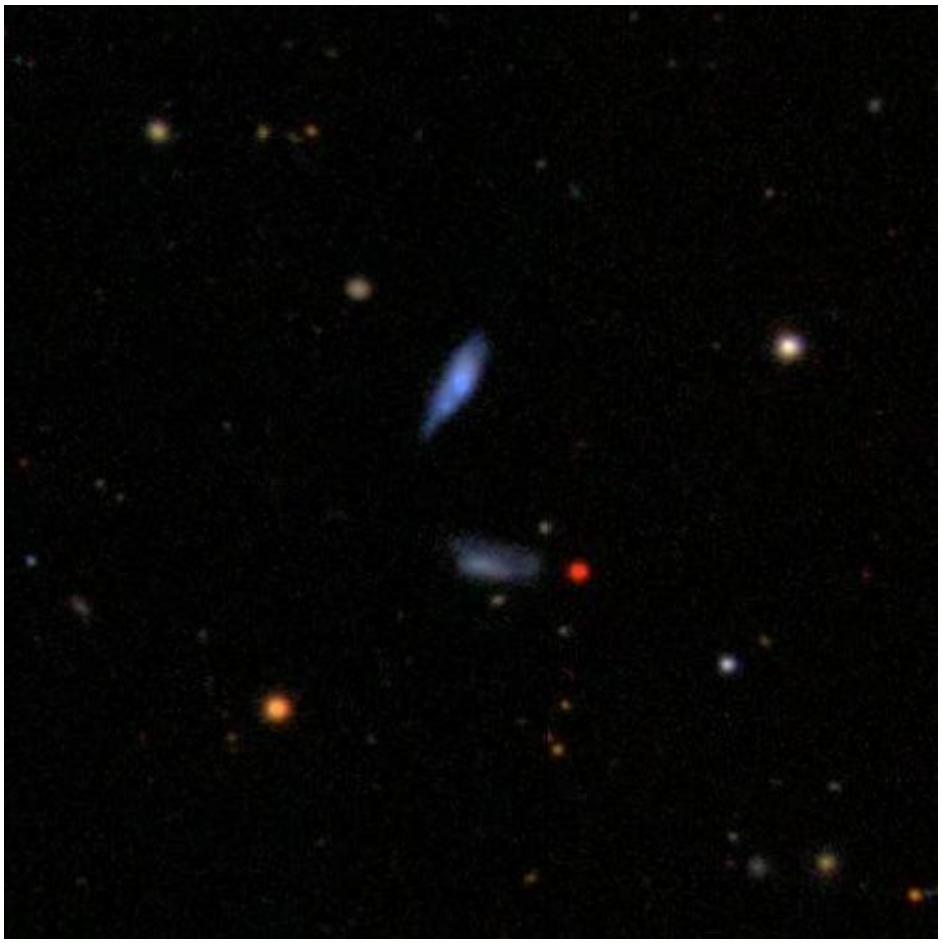


Figure A.7: SDSS Thumbnail of galaxy pair 588017605758025795 (bottom) & 588017605758025732 (top).

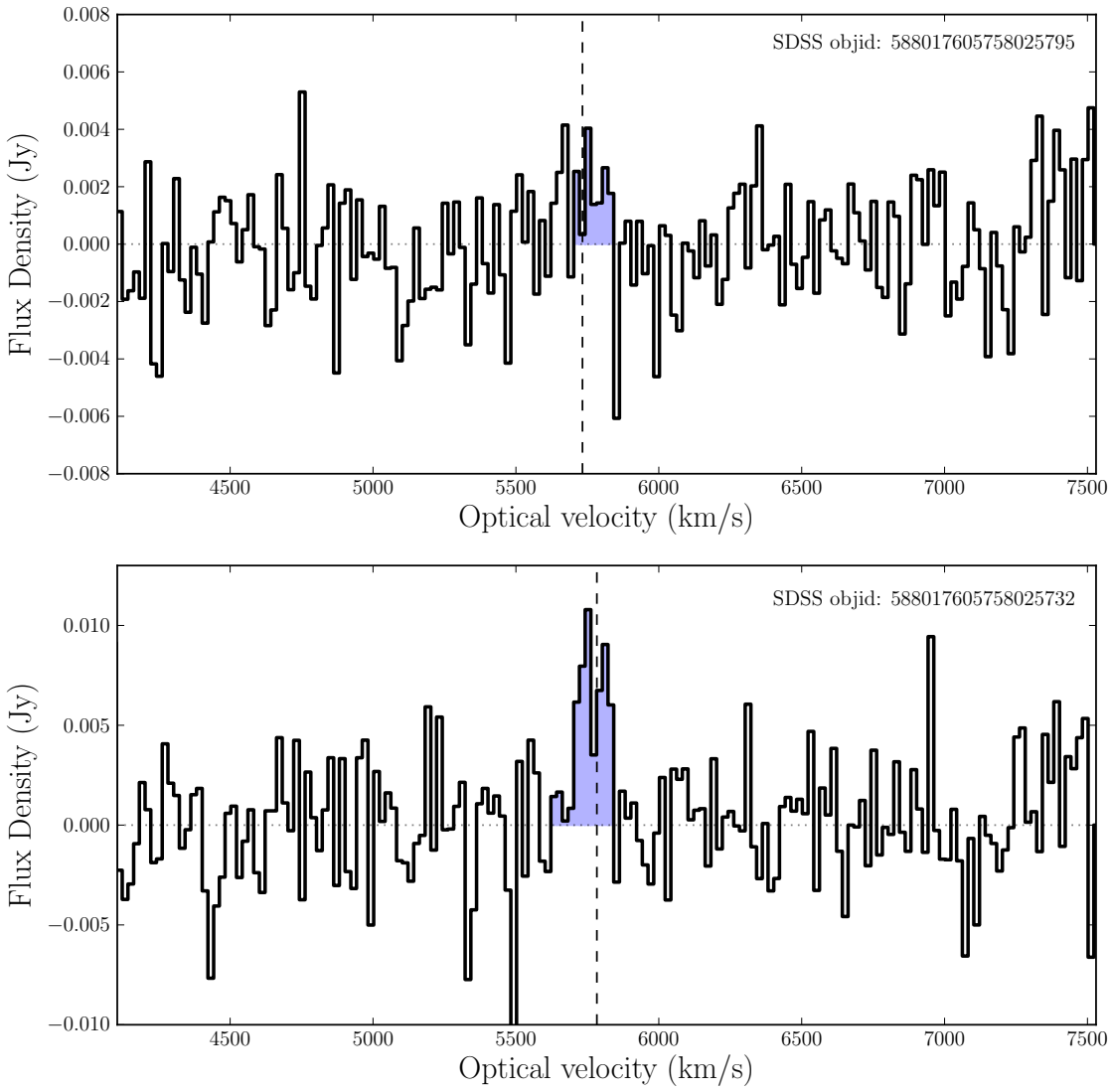


Figure A.8: Same as in Figure A.2; Flux density vs. optical velocity for galaxy pair 588017605758025795 & 588017605758025732. The upper spectrum has a peak/RMS S/N of 1.93, while the lower spectrum has a peak/RMS S/N of 3.62. The SDSS image for this pair is presented in Figure A.7.

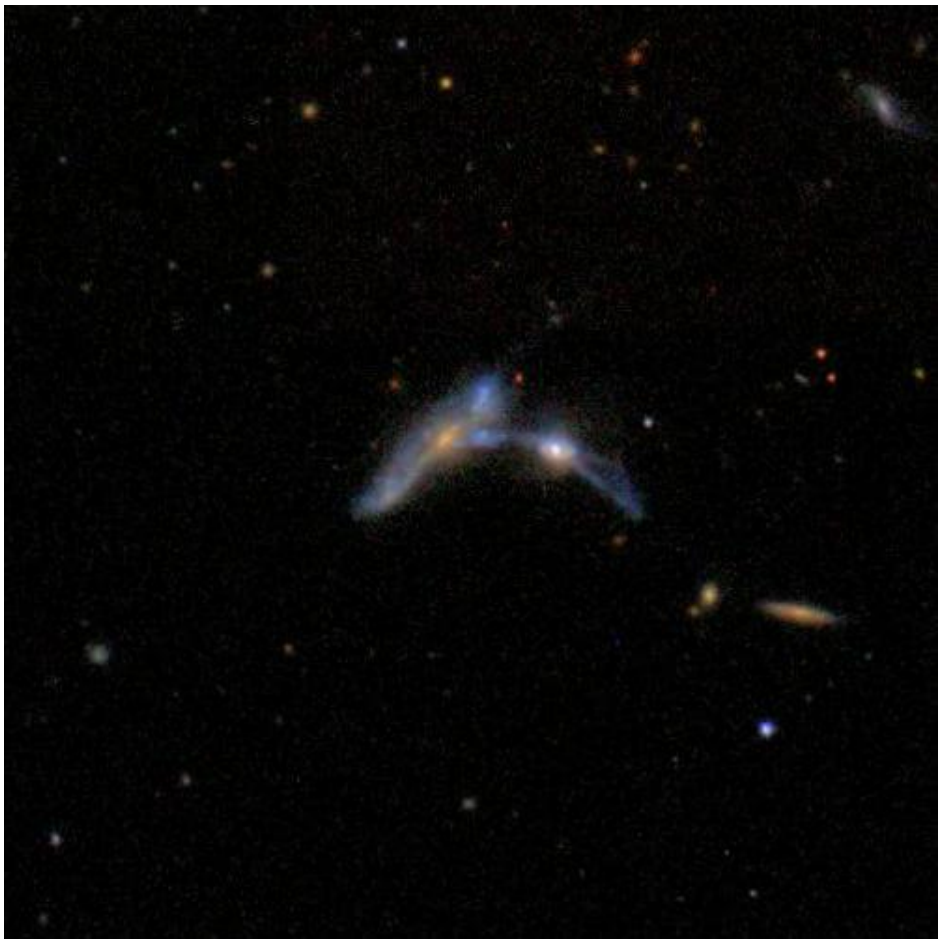


Figure A.9: SDSS Thumbnail of galaxy pair 587742901789589569 (right) & 587742901789589575 (left).

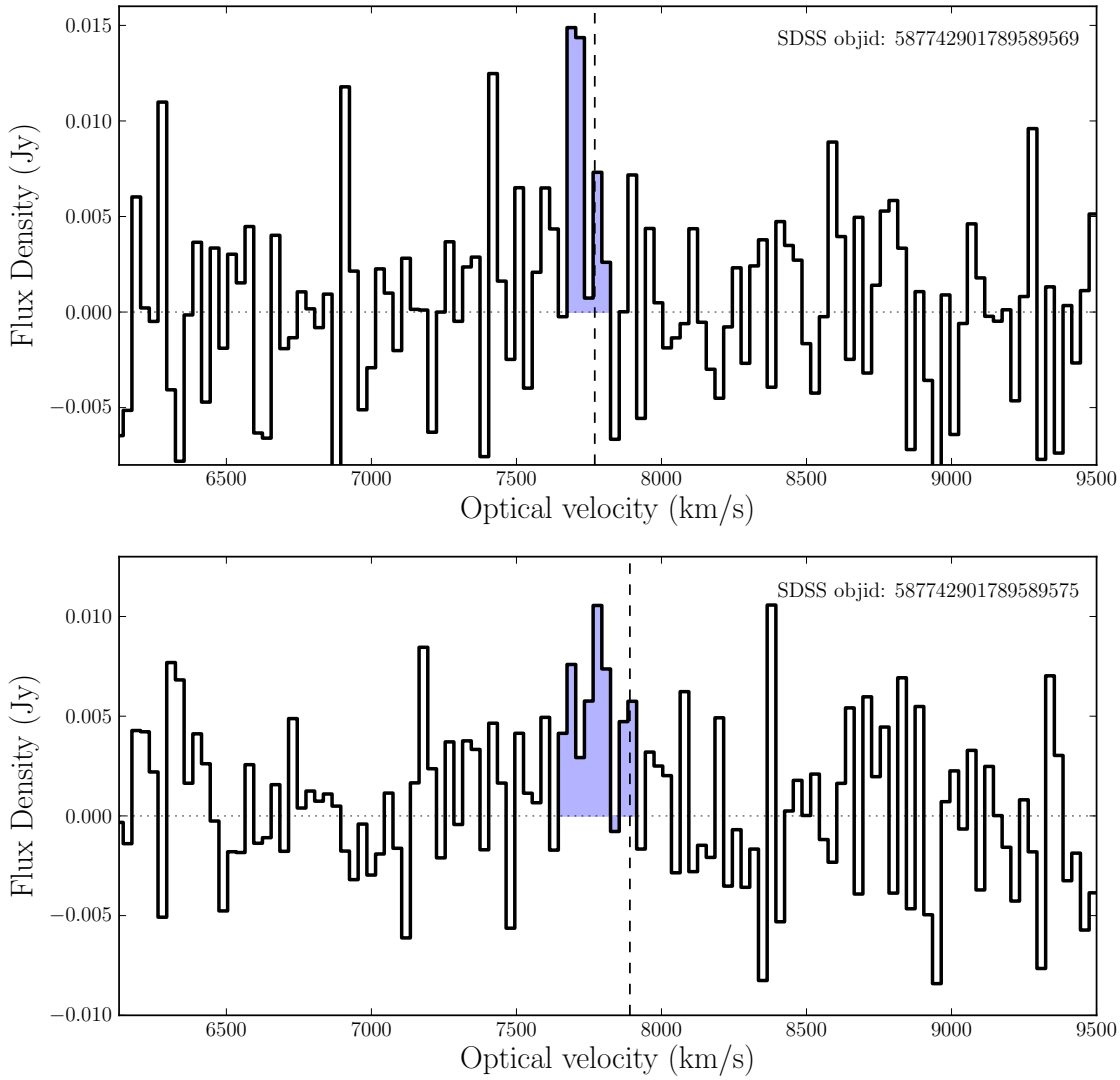


Figure A.10: Same as in Figure A.2; Flux density vs. optical velocity for galaxy pair 587742901789589569 & 587742901789589575. The upper spectrum has a peak/RMS S/N of 3.20, while the lower spectrum has a peak/RMS S/N of 2.70. The SDSS image for this pair is presented in Figure A.9.

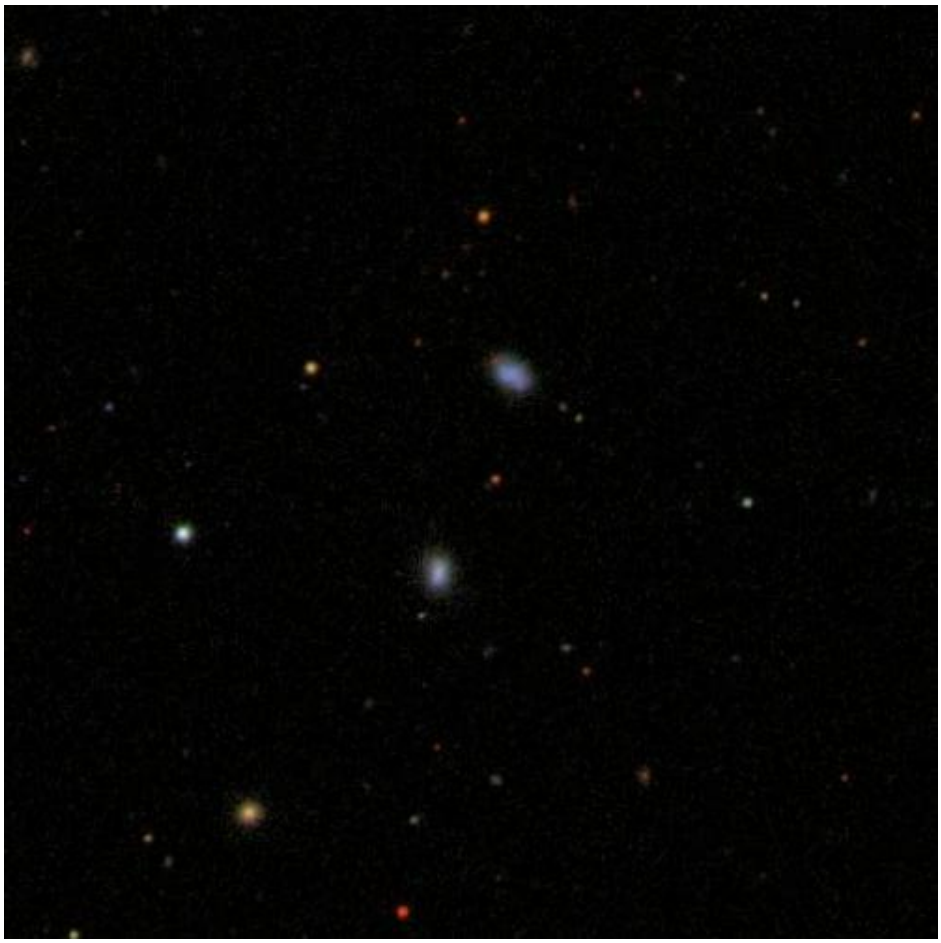


Figure A.11: SDSS Thumbnail of galaxy pair 588023670245949622 (top) & 588023670245949625 (bottom).

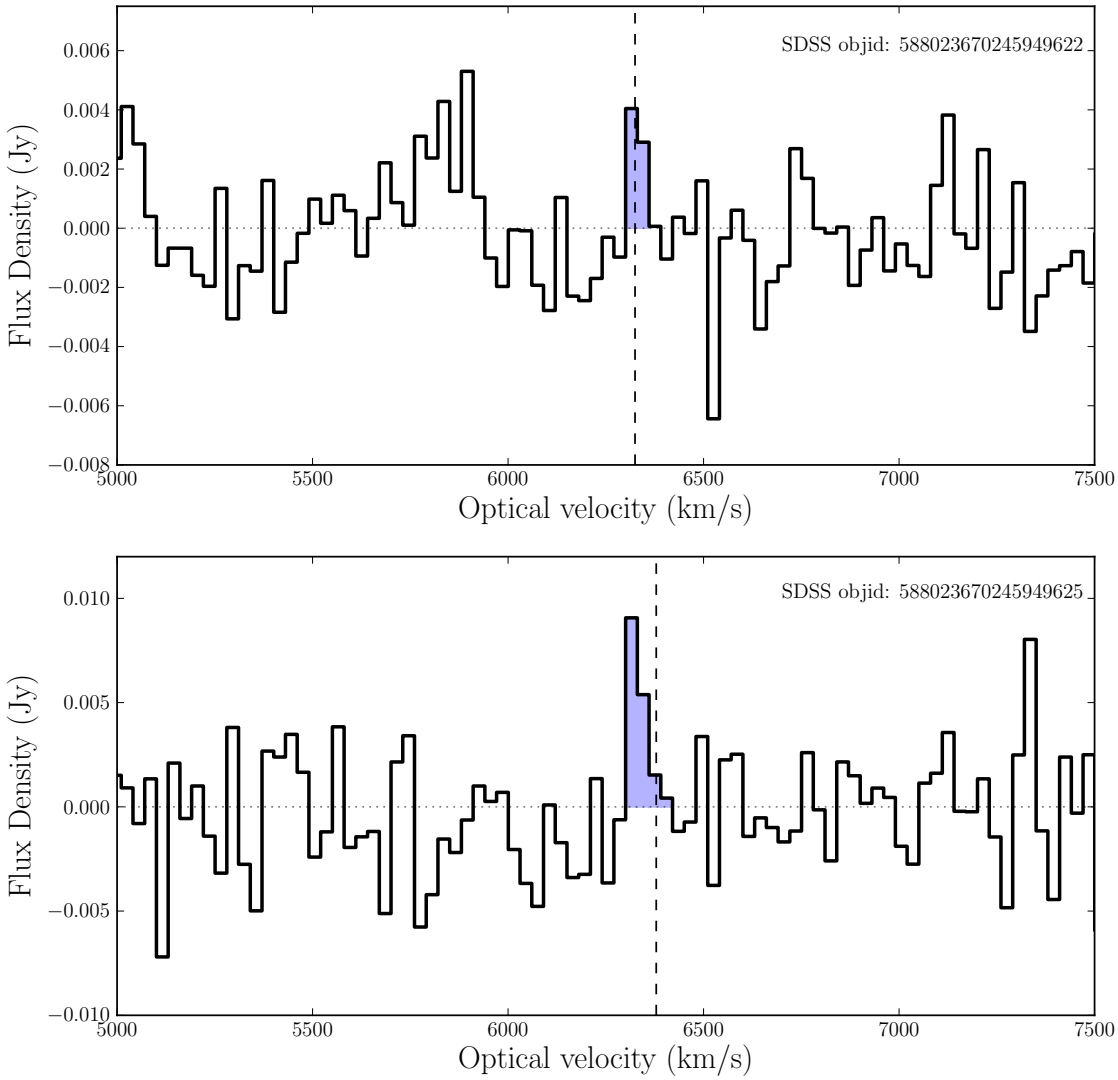


Figure A.12: Same as in Figure A.2; Flux density vs. optical velocity for galaxy pair 588023670245949622 & 588023670245949625. The upper spectrum has a peak/RMS S/N of 1.61, while the lower spectrum has a peak/RMS S/N of 2.73. The SDSS image for this pair is presented in Figure A.11.

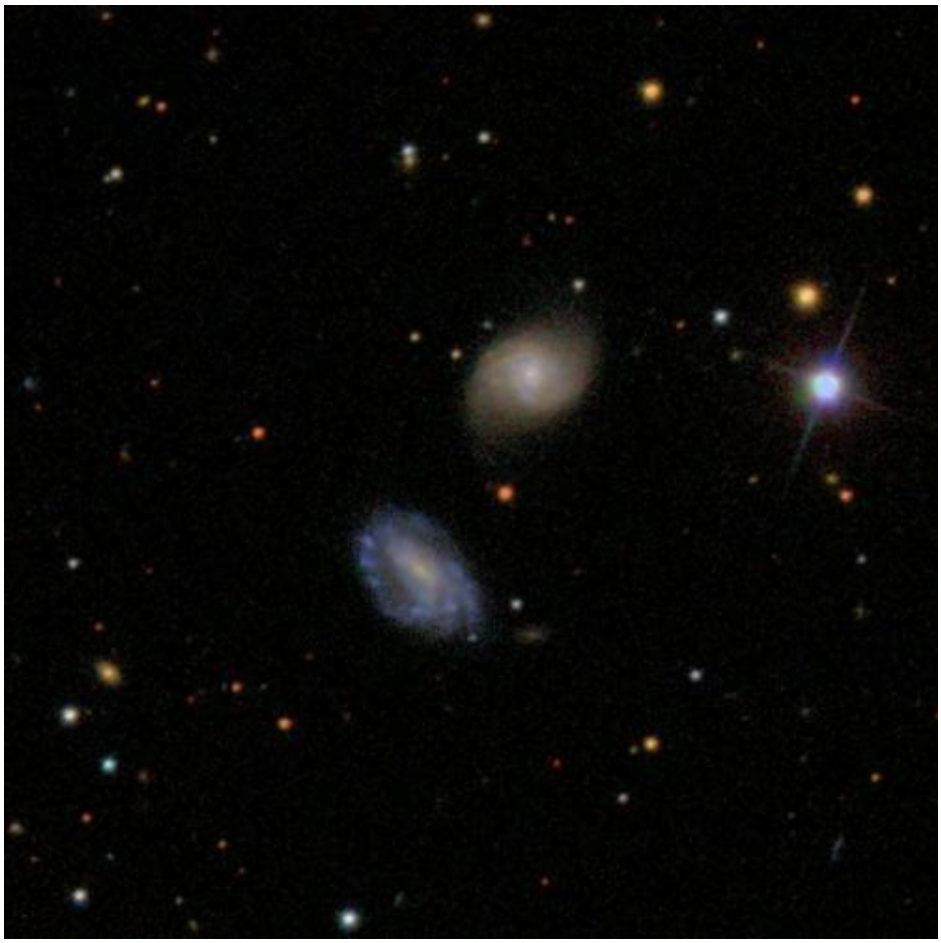


Figure A.13: SDSS Thumbnail of galaxy pair 588018056204780081(bottom left) & 588018056204780049 (top right).

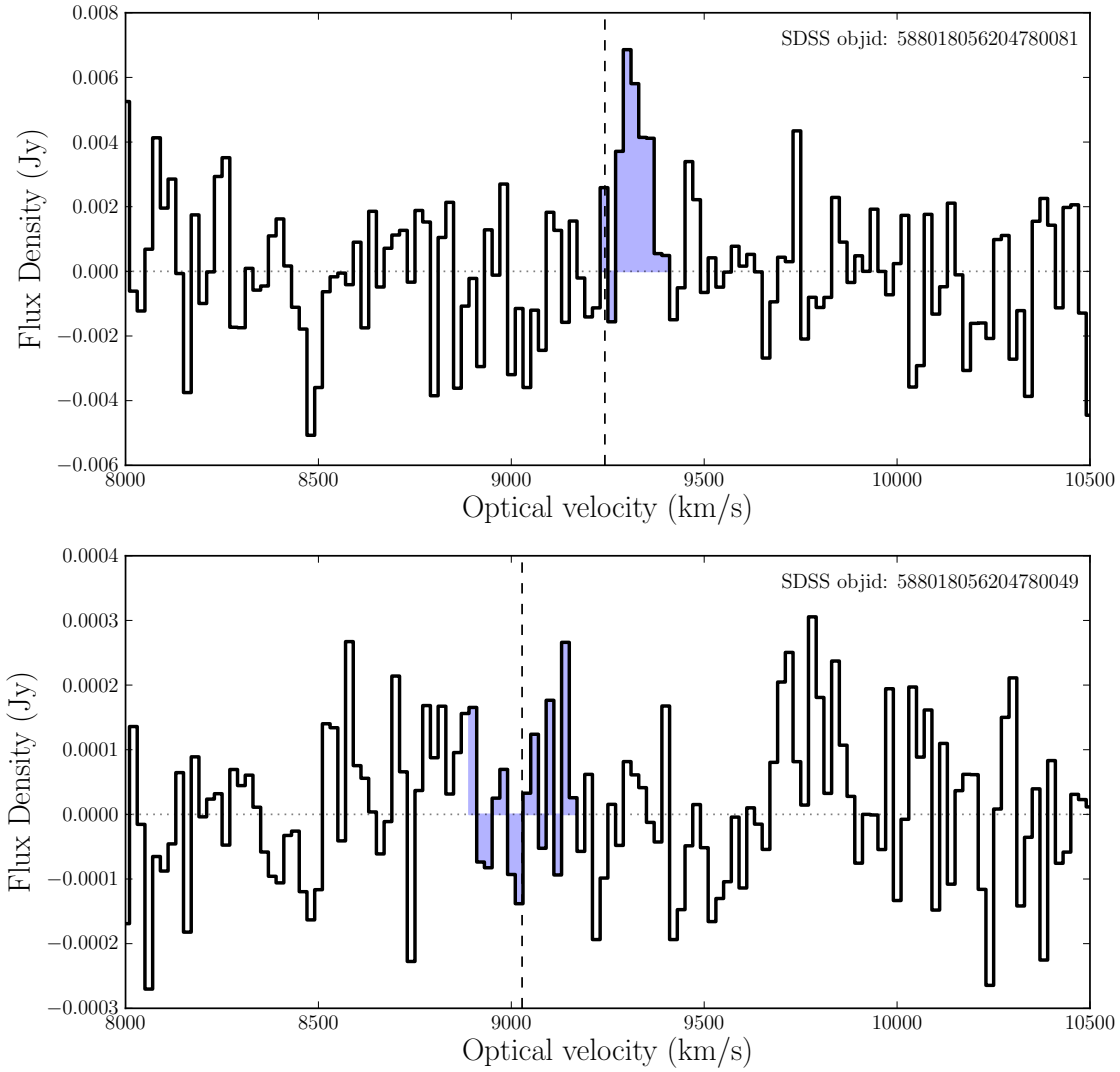


Figure A.14: Same as in Figure A.2; Flux density vs. optical velocity for galaxy pair 588018056204780081 & 588018056204780049. The upper spectrum has a peak/RMS S/N of 3.19, while the lower has a S/N of 1.07. The SDSS image for this pair is presented in Figure A.13.

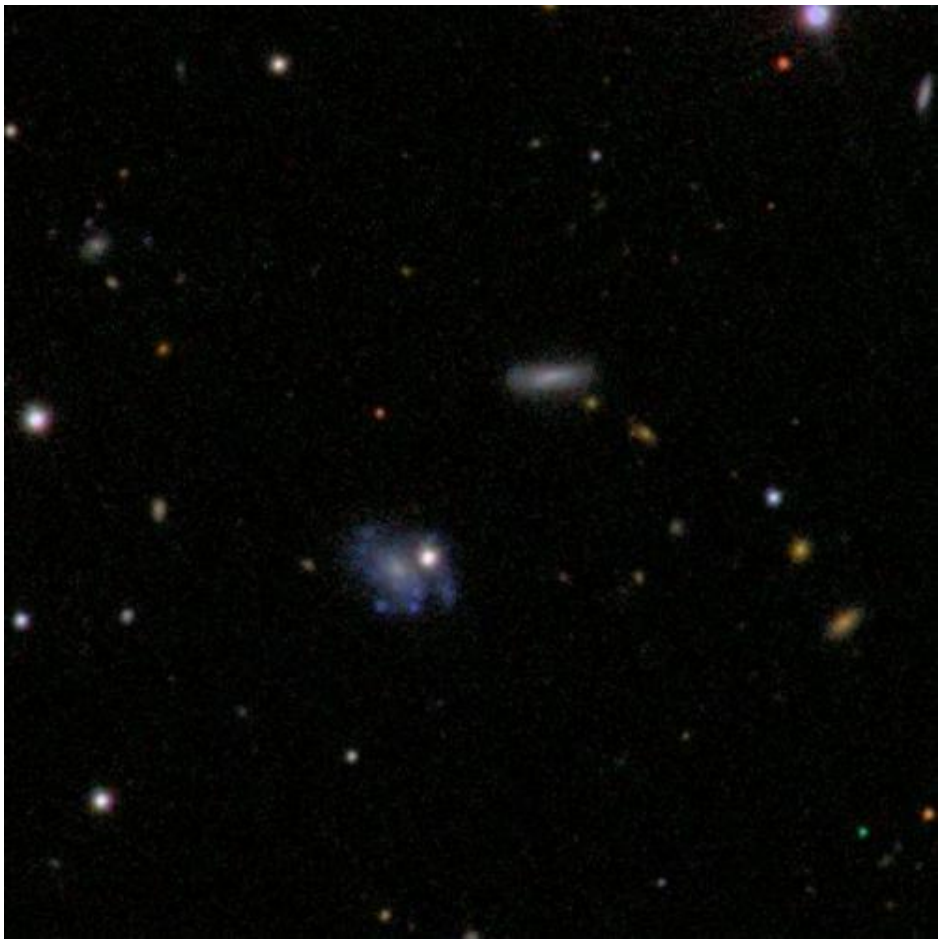


Figure A.15: SDSS Thumbnail of galaxy pair 587733605328093368 (top right) & 587733605328093256 (bottom left).

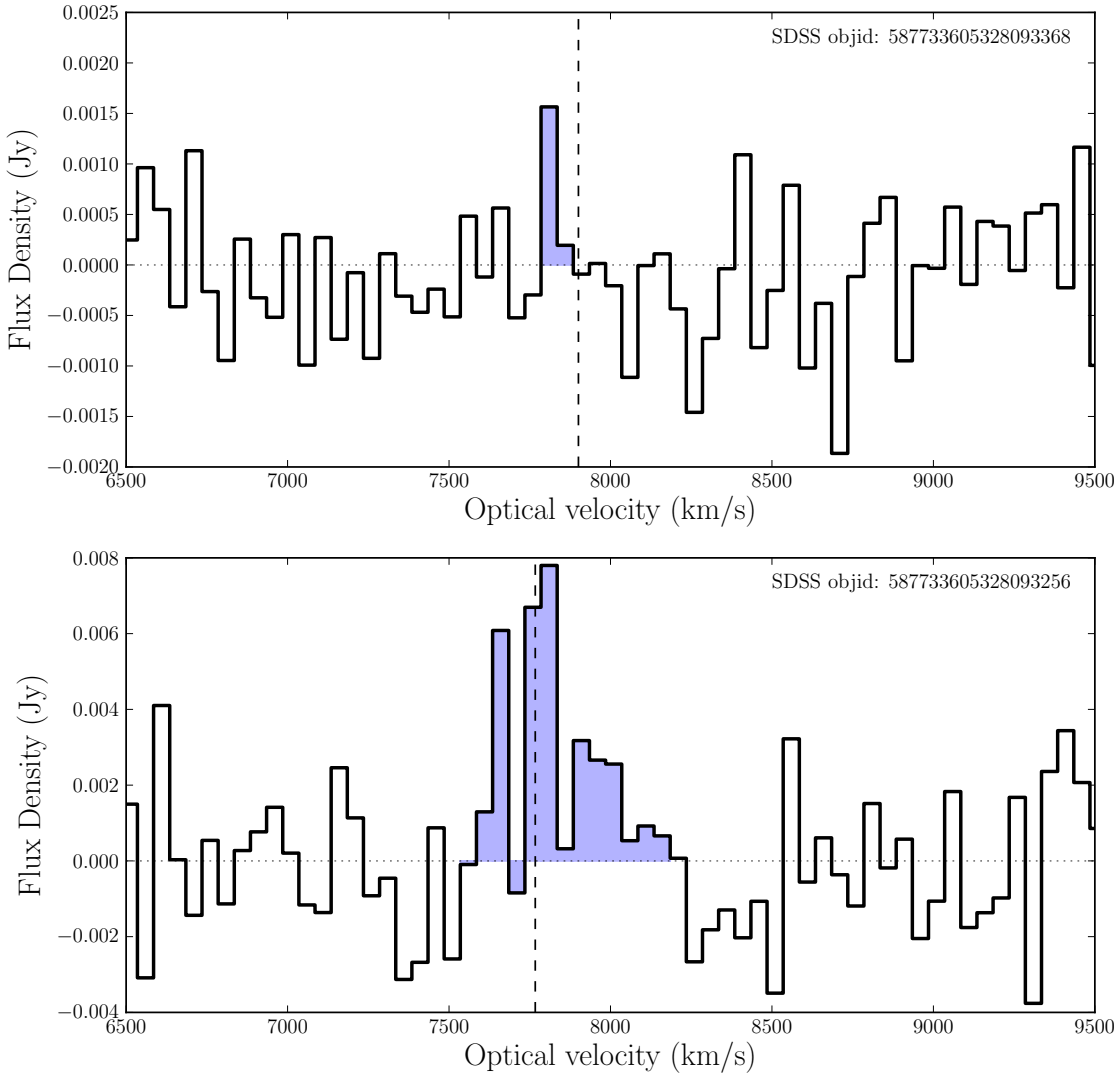


Figure A.16: Same as in Figure A.2; Flux density vs. optical velocity for galaxy pair 587733605328093368 & 587733605328093256. The upper spectrum has a peak/RMS S/N of 2.26, while the lower spectrum has a peak/RMS S/N of 4.04. The SDSS image for this pair is presented in Figure A.15.

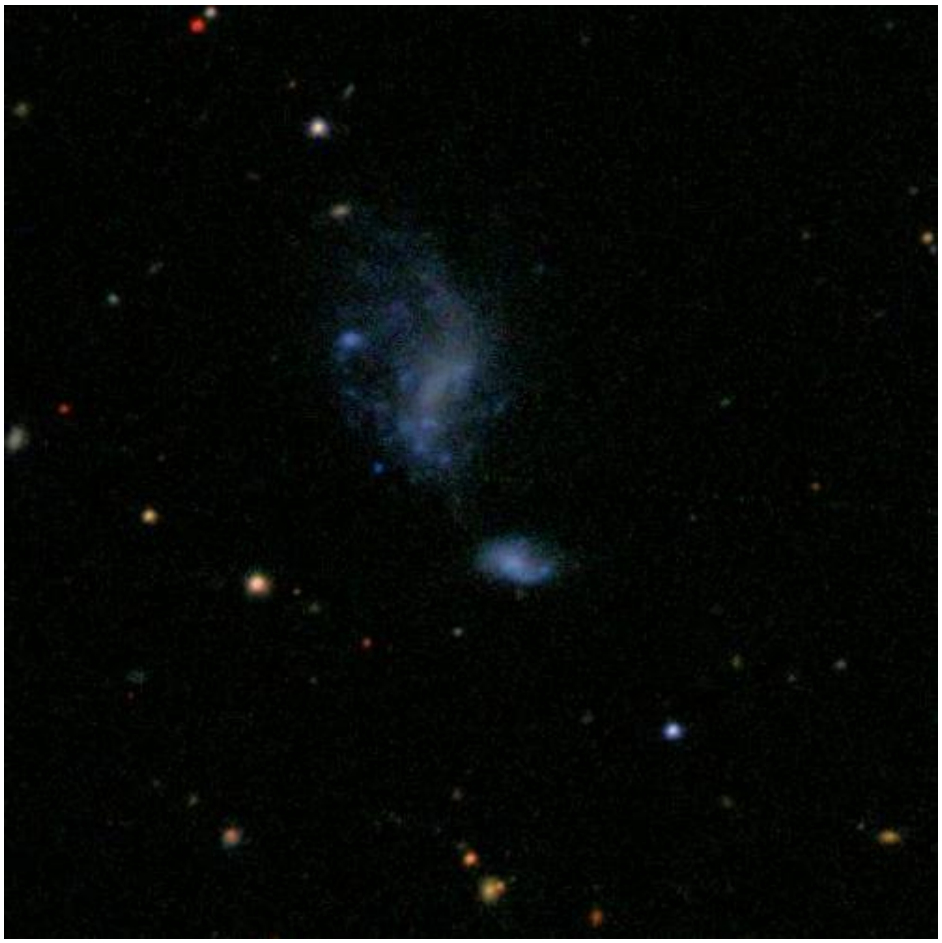


Figure A.17: SDSS Thumbnail of galaxy pair 587727179536859247 (top left) & 587727179536859227 (bottom right).

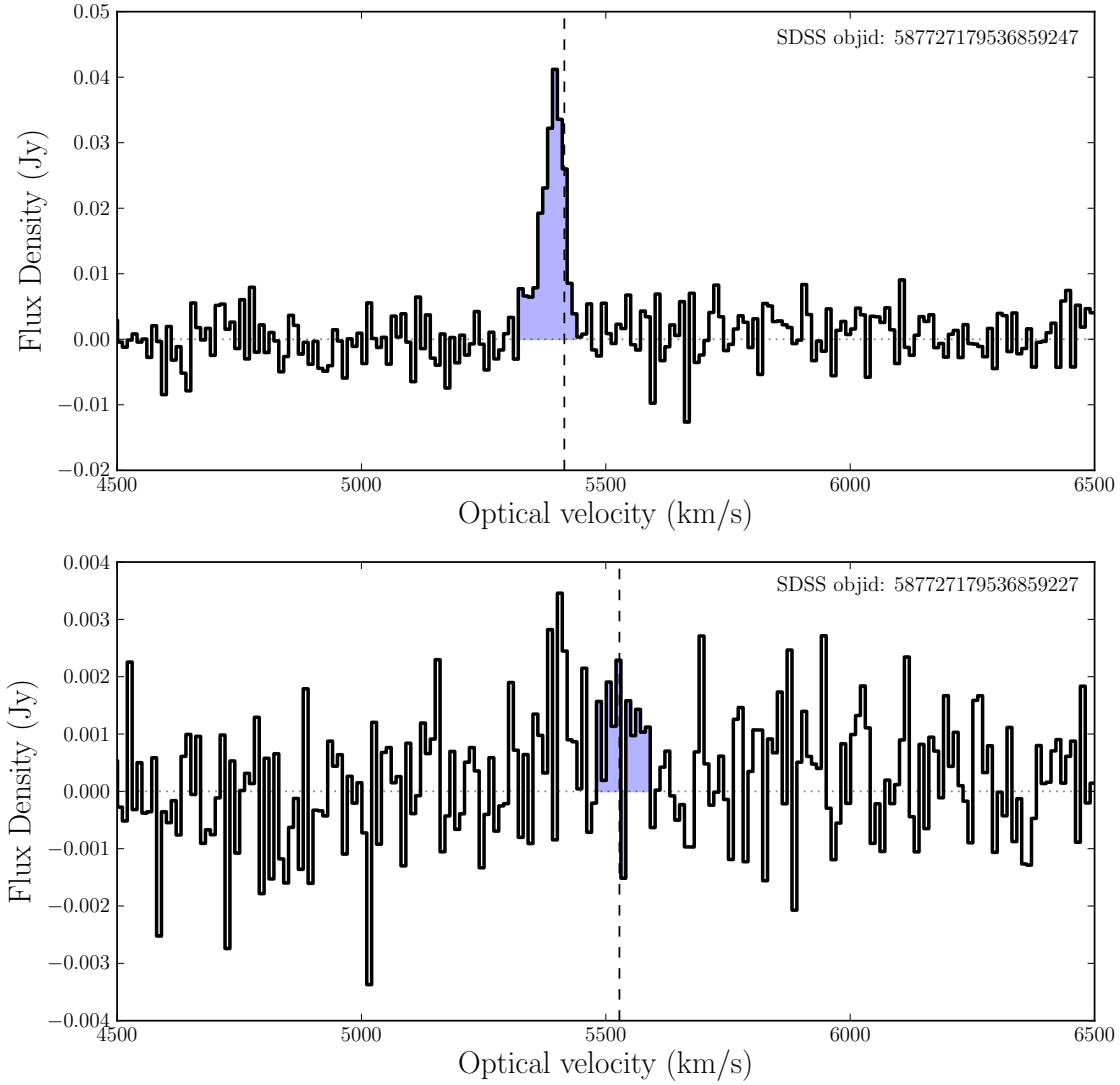


Figure A.18: Same as in Figure A.2; Flux density vs. optical velocity for galaxy pair 587727179536859247 & 587727179536859227. The upper spectrum has a peak/RMS S/N of 11.40, while the lower spectrum has a peak/RMS S/N of 2.02. The SDSS image for this pair is presented in Figure A.17.

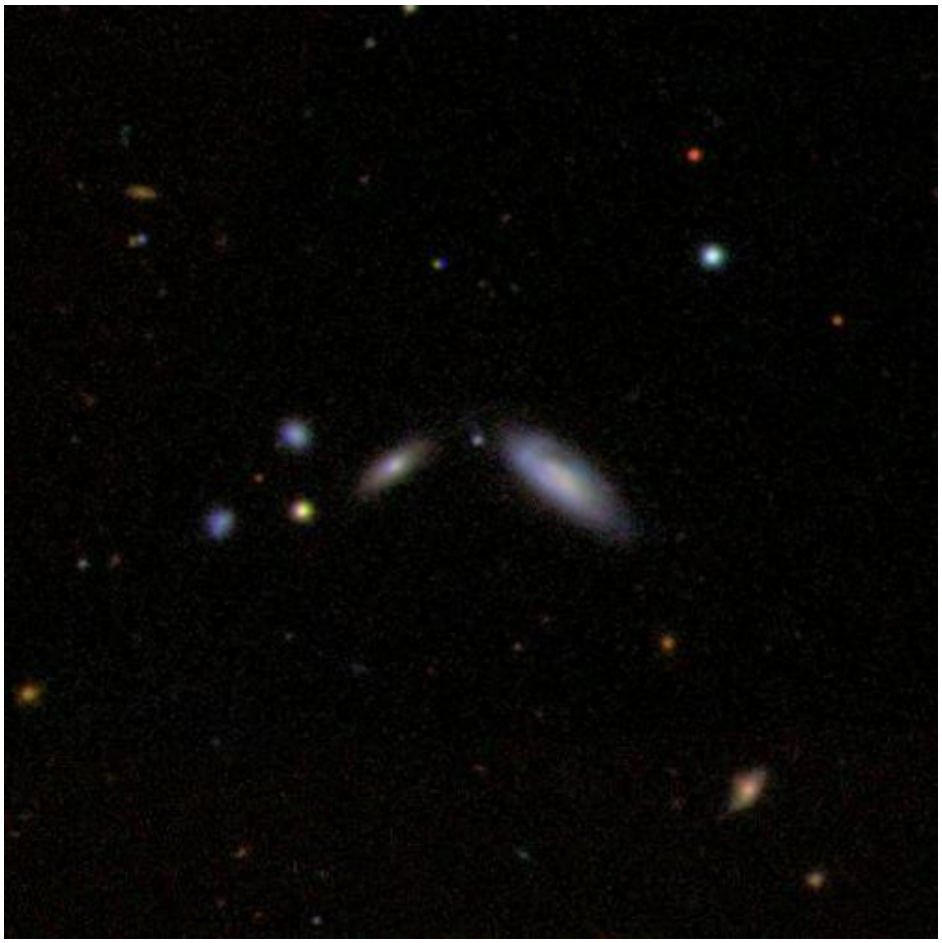


Figure A.19: SDSS Thumbnail of galaxy pair 587729158970867777 (right) & 587729158970867792 (left).

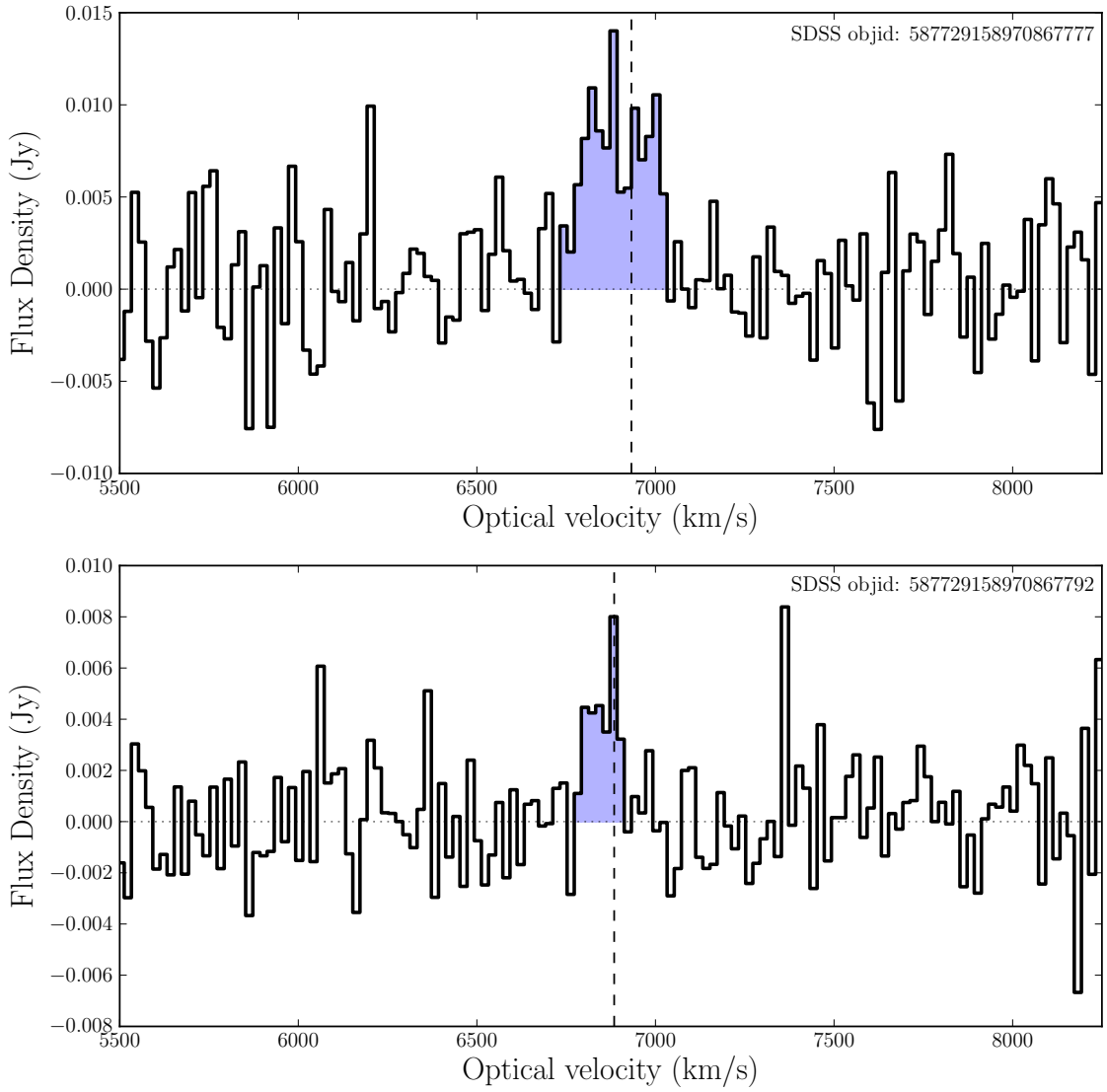


Figure A.20: Same as in Figure A.2; Flux density vs. optical velocity for galaxy pair 587729158970867777 & 587729158970867792. The upper spectrum has a peak/RMS S/N of 4.09, while the lower spectrum has a peak/RMS S/N of 3.71. The SDSS image for this pair is presented in Figure A.19.

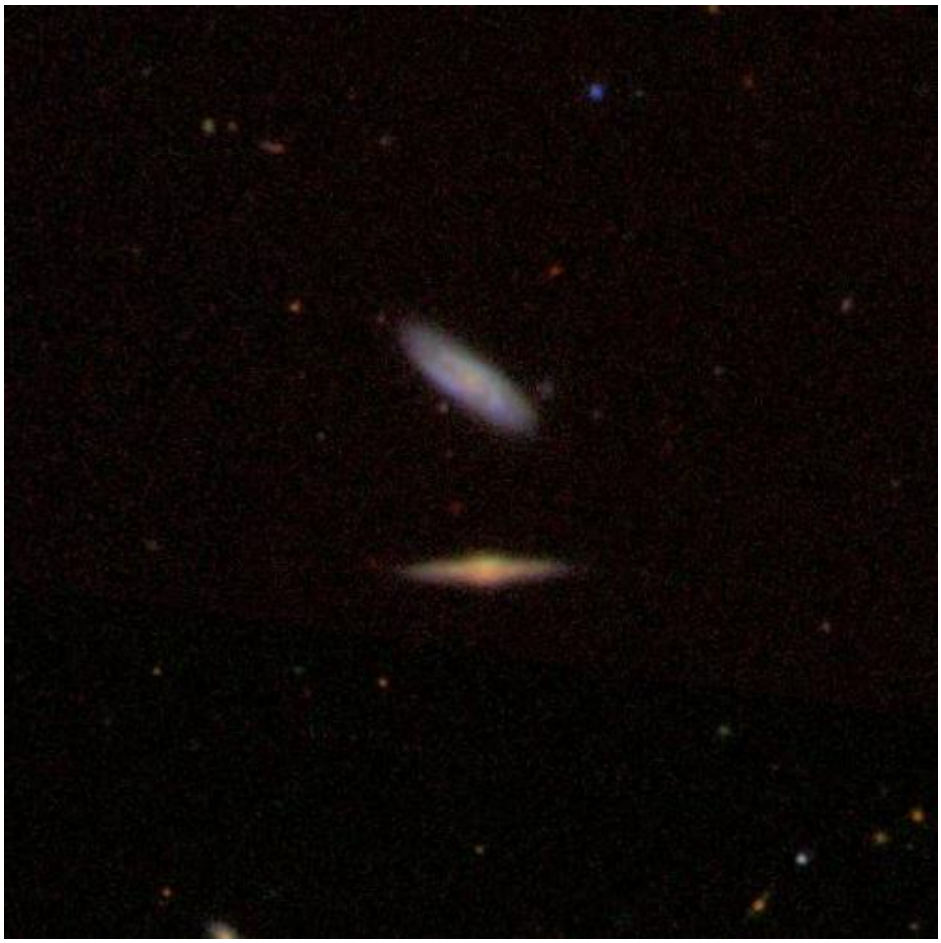


Figure A.21: SDSS Thumbnail of galaxy pair 587739609695453284 (top) & 587739609695453281 (bottom).

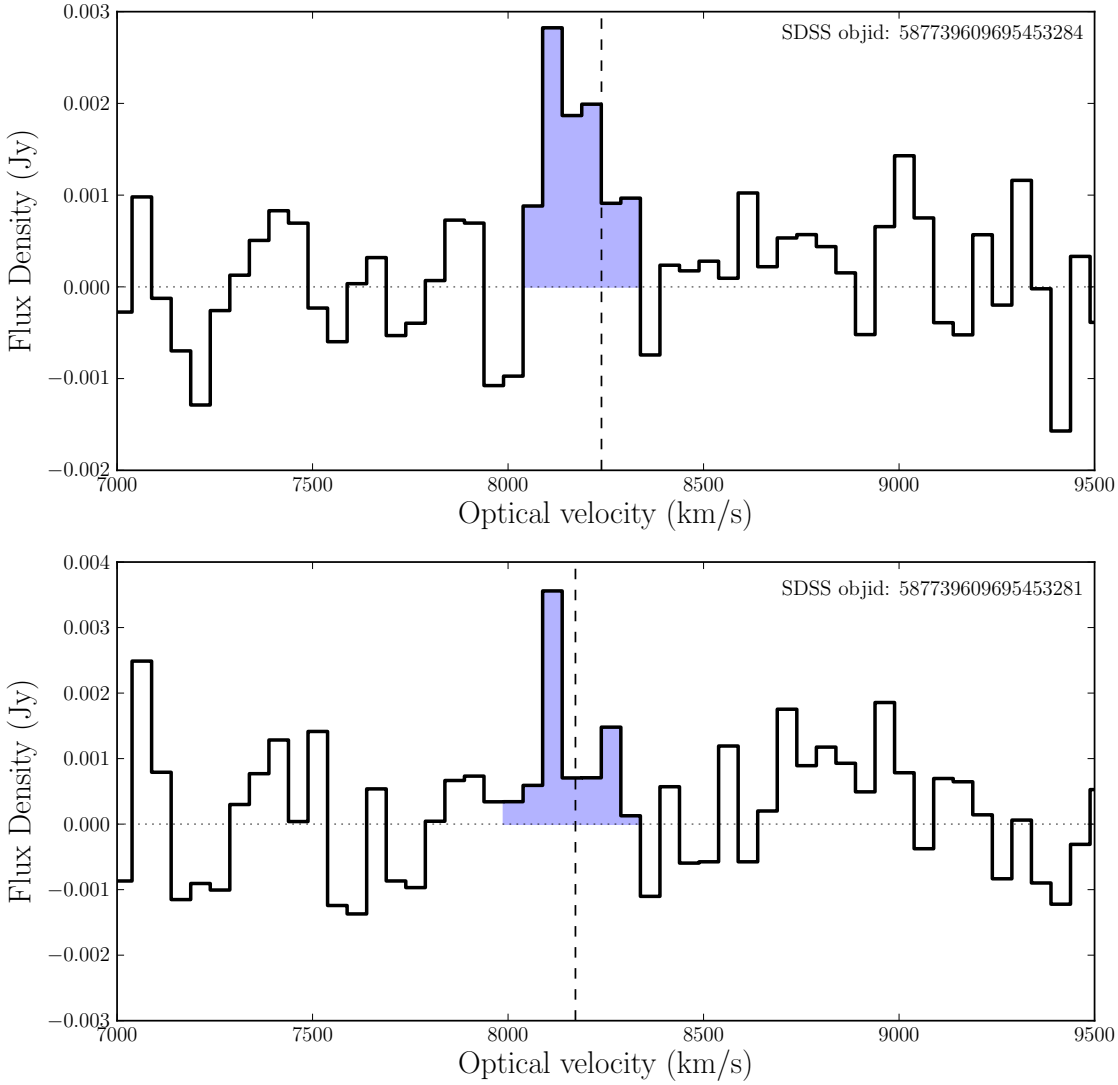


Figure A.22: Same as in Figure A.2; Flux density vs. optical velocity for galaxy pair 587739609695453284 & 587739609695453281. The upper spectrum has a peak/RMS S/N of 4.08, while the lower spectrum has a peak/RMS S/N of 3.69. The SDSS image for this pair is presented in Figure A.21.

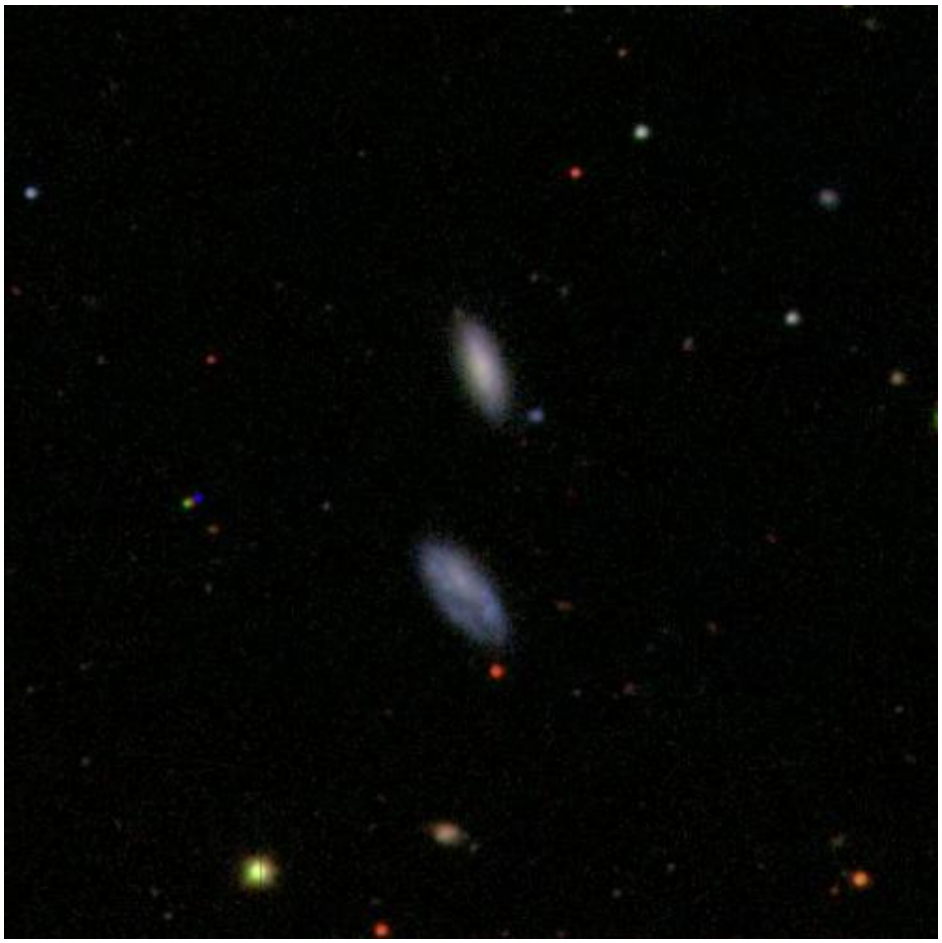


Figure A.23: SDSS Thumbnail of galaxy pair 588848899908370674 (bottom) & 588848899908370505 (top).

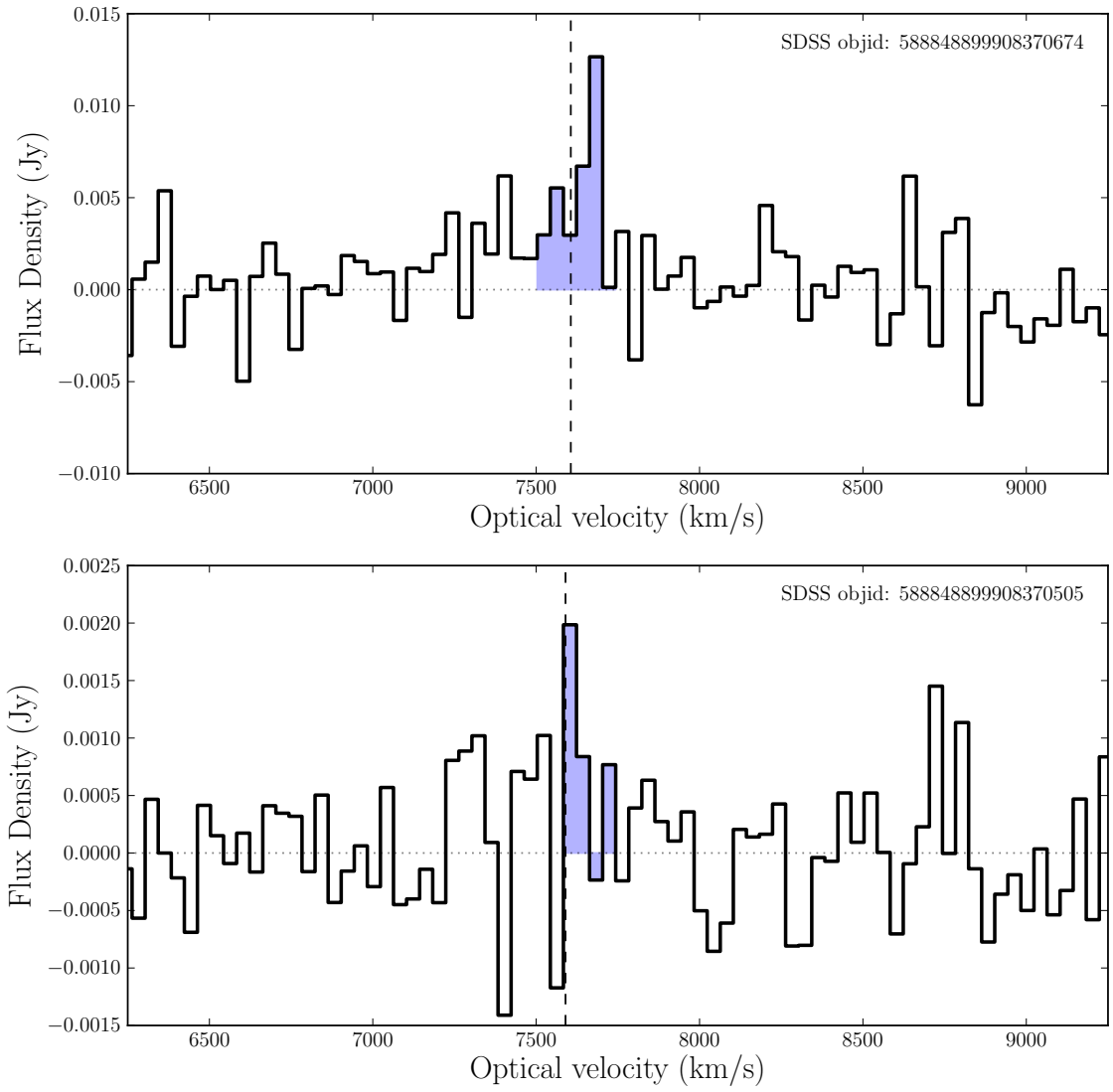


Figure A.24: Same as in Figure A.2; Flux density vs. optical velocity for galaxy pair 588848899908370674 & 588848899908370505. The upper spectrum has a peak/RMS S/N of 4.54, while the lower spectrum has a peak/RMS S/N of 3.26. The SDSS image for this pair is presented in Figure A.23.

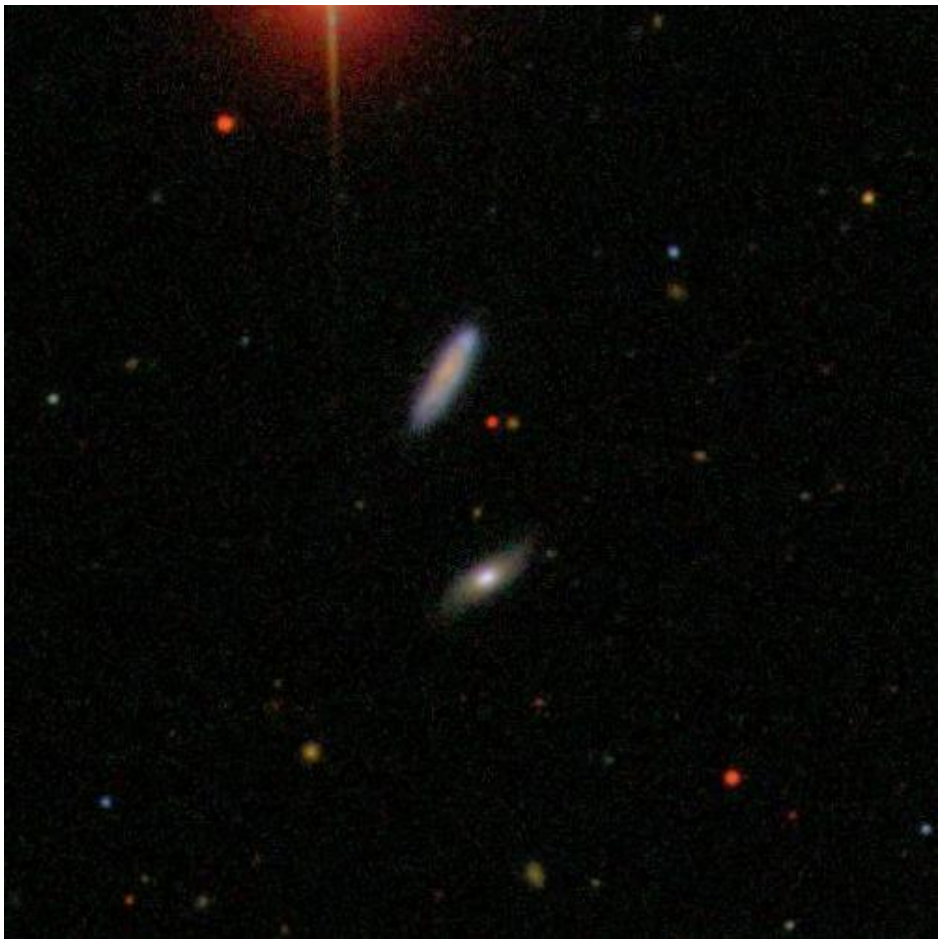


Figure A.25: SDSS Thumbnail of galaxy pair 587726033308680234 (bottom) & 587726033308680320 (top).

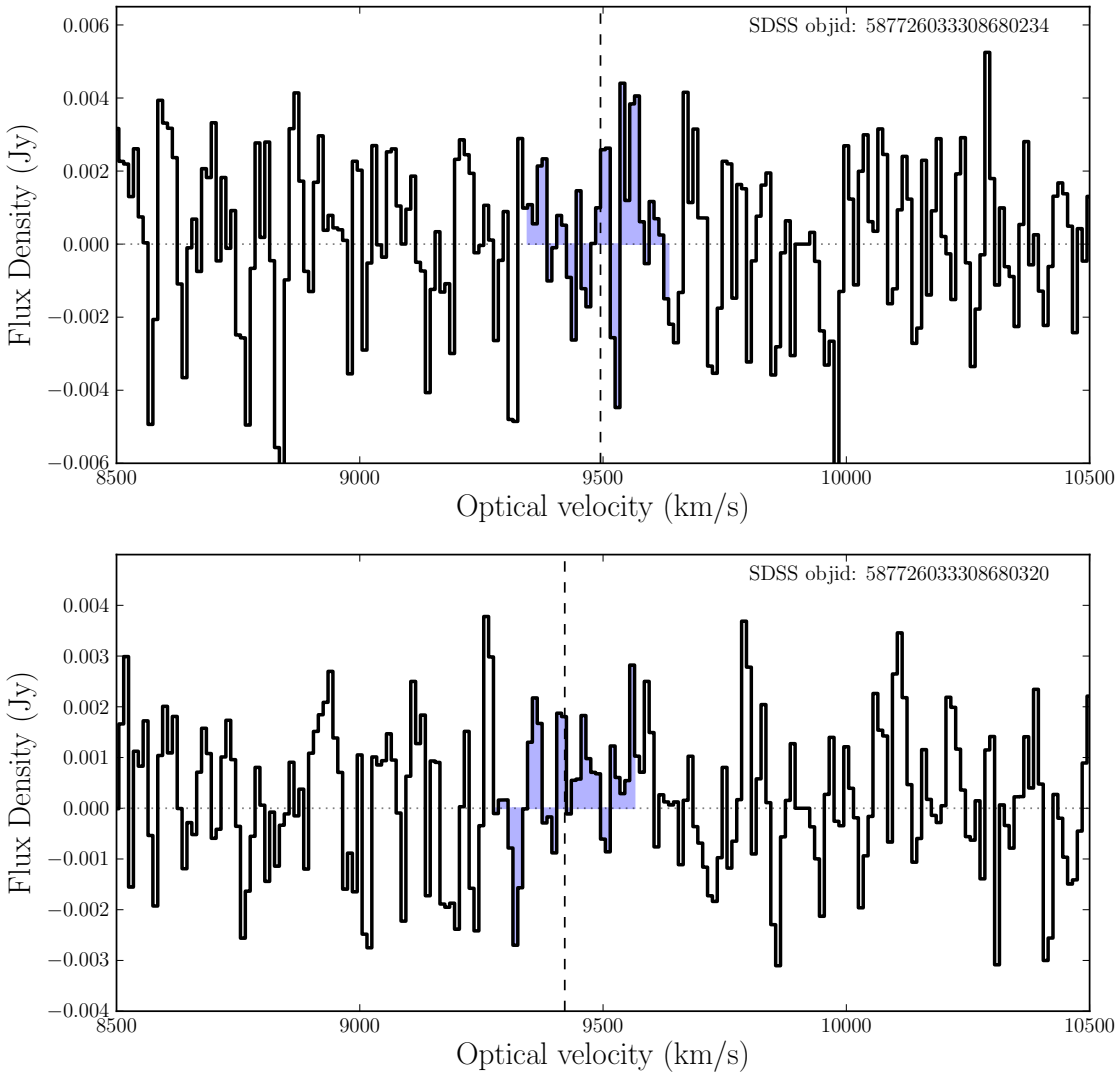


Figure A.26: Same as in Figure A.2; Flux density vs. optical velocity for galaxy pair 587726033308680234 & 587726033308680320. The upper spectrum has a peak/RMS S/N of 1.97, while the lower spectrum has a peak/RMS S/N of 2.00. The SDSS image for this pair is presented in Figure A.25.

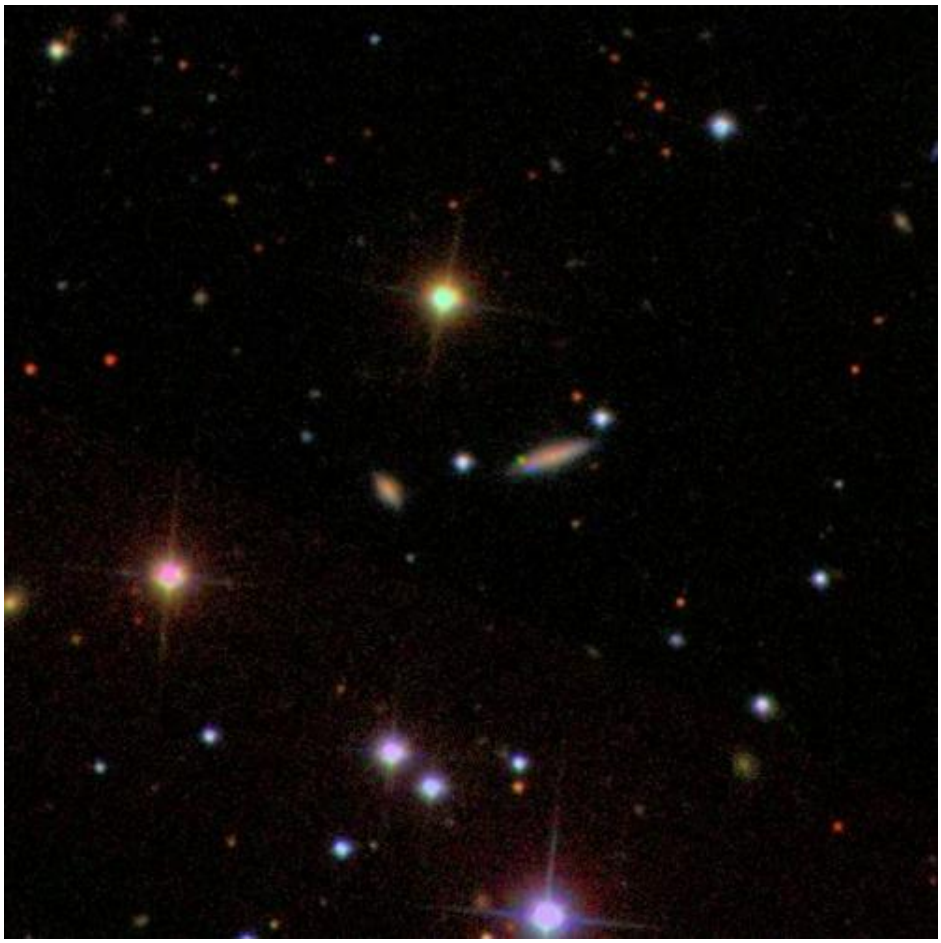


Figure A.27: SDSS Thumbnail of galaxy pair 587741489815027774 (right) & 587741489815028146 (left).

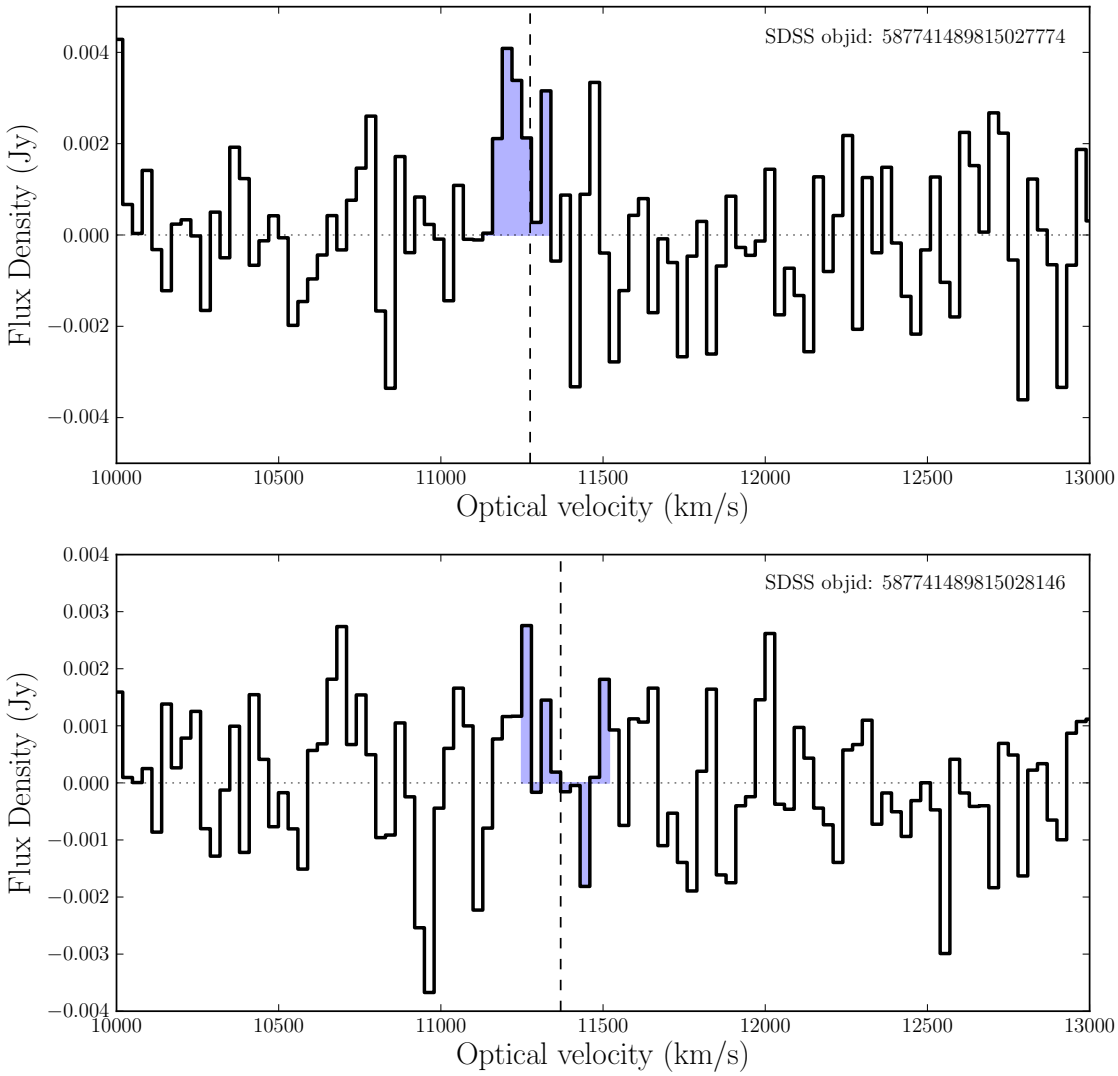


Figure A.28: Same as in Figure A.2; Flux density vs. optical velocity for galaxy pair 587741489815027774 & 587741489815028146. The upper spectrum has a peak/RMS S/N of 2.70, while the lower spectrum has a peak/RMS S/N of 2.23. The SDSS image for this pair is presented in Figure A.27.

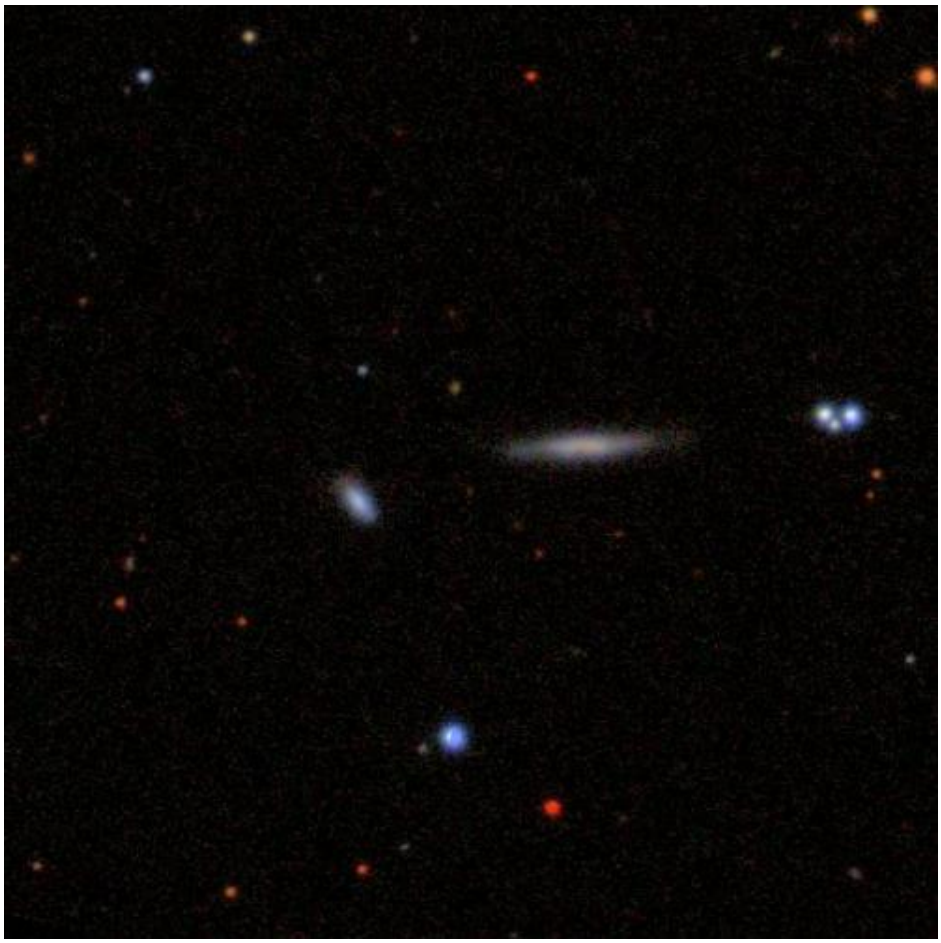


Figure A.29: SDSS Thumbnail of galaxy pair 587744873717563559 (right) & 587744873717563471 (left).

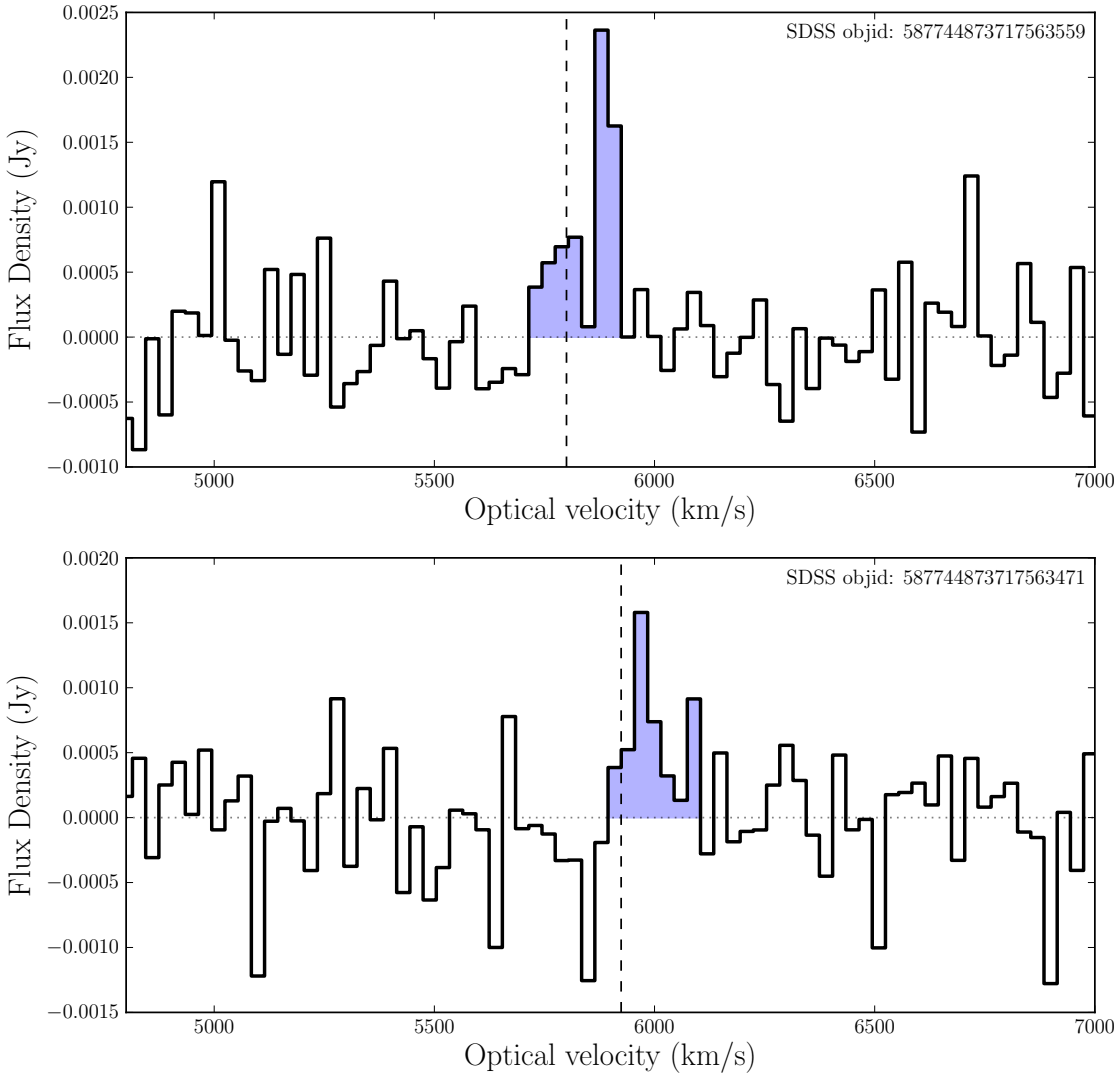


Figure A.30: Same as in Figure A.2; Flux density vs. optical velocity for galaxy pair 587744873717563559 & 587744873717563471. The upper spectrum has a peak/RMS S/N of 5.35, while the lower spectrum has a peak/RMS S/N of 3.51. The SDSS image for this pair is presented in Figure A.29.

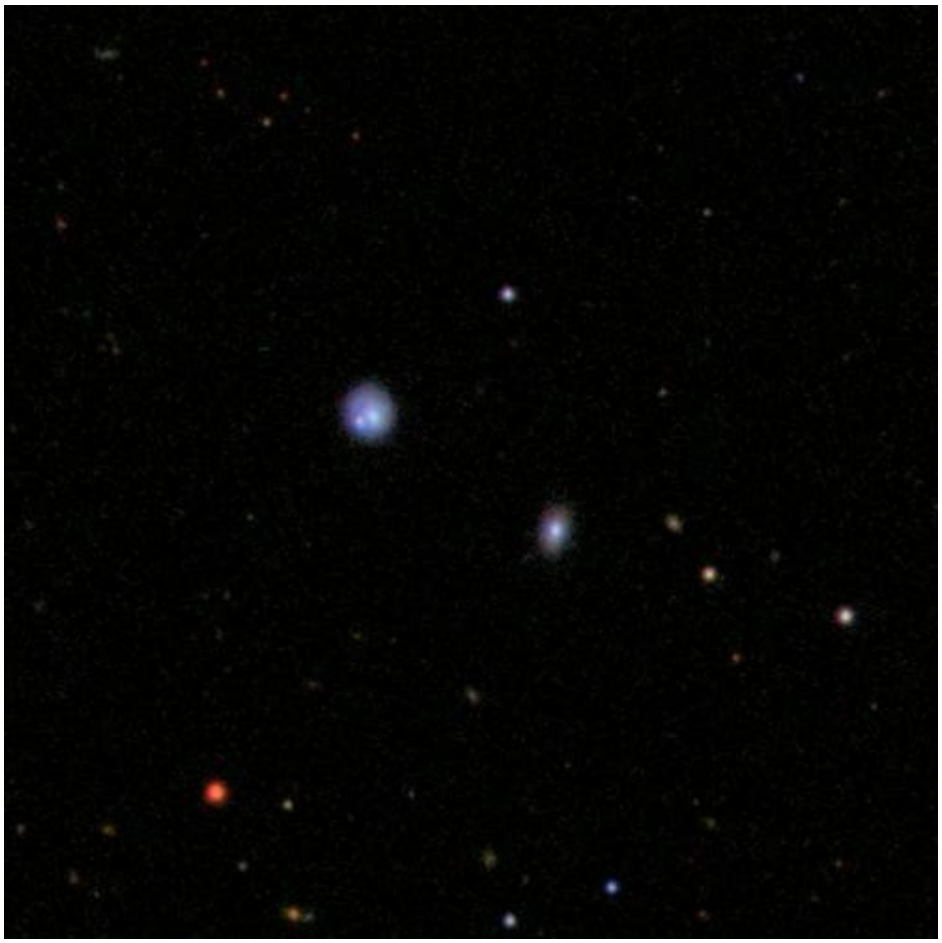


Figure A.31: SDSS Thumbnail of galaxy pair 587729160043757697 (lower right) & 587729160043757707 (upper left)

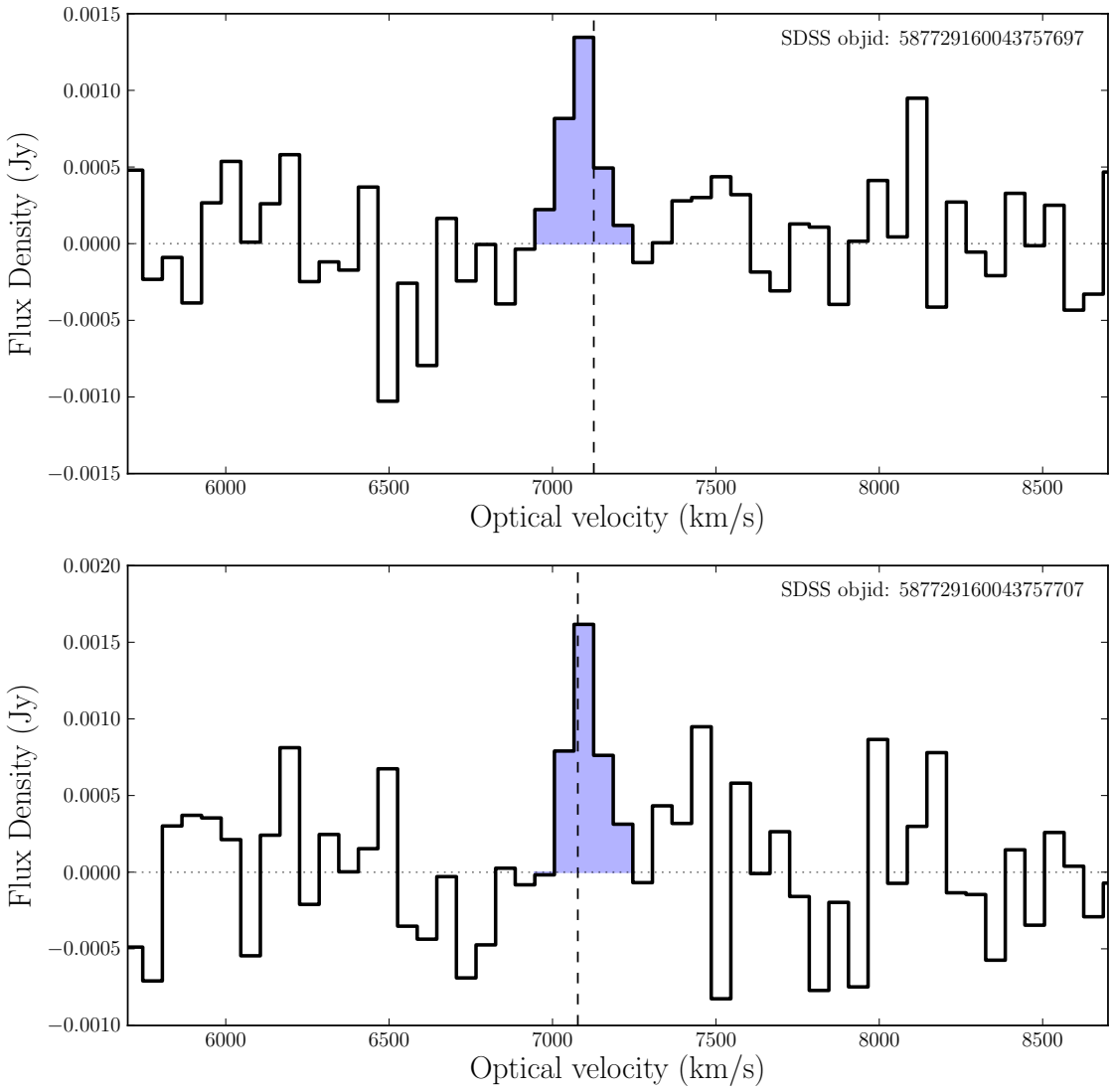


Figure A.32: Same as in Figure A.2; Flux density vs. optical velocity for galaxy pair 587729160043757697 & 587729160043757707. The upper spectrum has a peak/RMS S/N of 3.76, while the lower spectrum has a peak/RMS S/N of 3.16. The SDSS image for this pair is presented in Figure A.31.

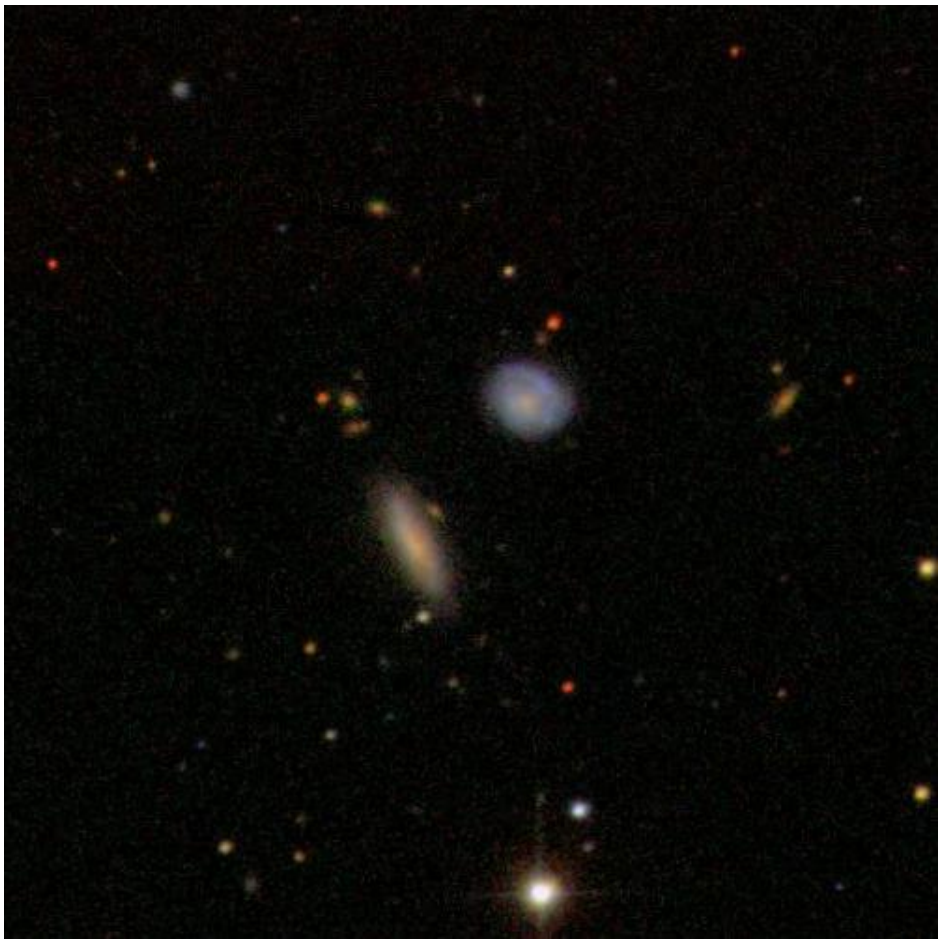


Figure A.33: SDSS Thumbnail of galaxy pair 587726033341776175 (upper right) & 587726033341776191 (lower left)

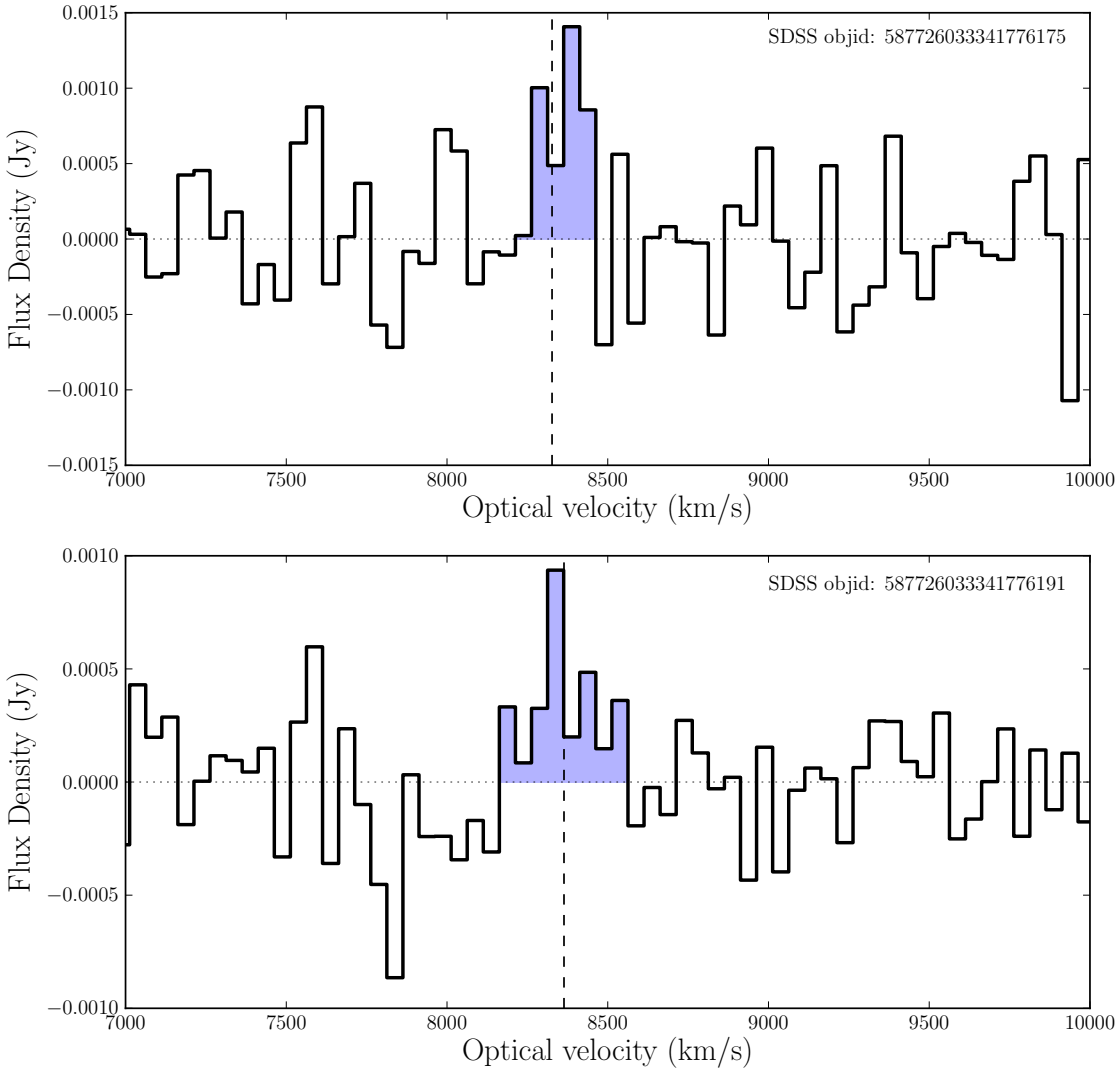


Figure A.34: Same as in Figure A.2; Flux density vs. optical velocity for galaxy pair 587726033341776175 & 587726033341776191. The upper spectrum has a peak/RMS S/N of 2.85, while the lower spectrum has a peak/RMS S/N of 3.42. The SDSS image for this pair is presented in Figure A.33.

Appendix B

Glossary of Terms

In this Glossary, terms and abbreviations are listed along with a brief definition.

AGN: Active Galactic Nucleus; an actively accreting black hole at the centre of a galaxy.

Continuum: The spectral continuum is the flux produced by a galaxy's stellar component. Any emission lines from the gaseous component of a galaxy are in addition to this stellar component. The continuum must be well fit by a model in order to extract emission line fluxes. If the level of continuum flux is substantially over or under estimated, the emission line flux will be under or over estimated, respectively.

dex: 'decimal exponent'. The difference between two log quantities. 10^1 and 10^2 are separated by 1 dex.

HII: A region of ionized hydrogen.

$M_\odot$: Mass, in units of the mass of the Sun (astronomical abbreviation: $\odot$), or 'solar masses'.

M_*: Stellar mass: the mass of a galaxy which is contained in stars. Usually in units of solar masses ($M_\odot$).

O & B type stars The first two of the stellar classification scheme OBAFGKM. These are the hottest, bluest, and shortest lived of the stars. Our sun is a G-type star, and K & M type stars are the coolest and longest lived of all the stellar types.

Redshift (z): The apparent shift in the wavelength of a spectral feature (such as an emission line), due to the recession of the galaxy with respect to us. This is determined through the equation $z = \frac{\lambda_{obs} - \lambda_{rest}}{\lambda_{rest}}$, where λ_{obs} is the observed wavelength, and λ_{rest} is the rest wavelength.

SFR: Star Formation Rate, usually in units of solar masses per year ($M_\odot$/yr).

B.0.1 Emission line abbreviations

All emission lines written with square brackets (e.g., [NII]) indicate that the emission is a forbidden line, which is to say that it is only produced in extremely low density environments.

Hα: Hydrogen alpha, the first emission line of the Balmer series. Emits at a wavelength of 6563Å.

Hβ: Hydrogen beta, the second emission line of the Balmer series. Emits at a wavelength of Å.

H I: Neutral Hydrogen. Emits at a wavelength of 21cm, or a frequency of 1.420 GHz.

[NII]λ6584: Singly ionized nitrogen. Specifically refers to the emission line at a wavelength of 6584 Å.

[OI]λ6300: Neutral oxygen. Specifically refers to the emission line at a wavelength of 3726 Å.

[OII]λ3726: Singly ionized oxygen. This emission at 3726 Å is part of a doublet with [OII]λ3729.

[OII]λ3729: Singly ionized oxygen. This emission at 3729 Å is part of a doublet with [OII]λ3726.

[OIII]λ4959: Doubly ionized oxygen. Specifically refers to the emission line at a wavelength of 4959 Å.

[OIII]λ5007: Doubly ionized oxygen. Specifically refers to the emission line at a wavelength of 5007 Å.

[SII]λ6717: Singly ionised sulfur. Specifically refers to the emission line at a wavelength of 6717 Å.

[SII]λ6731: Singly ionised sulfur. Specifically refers to the emission line at a wavelength of 6717 Å.