MAPPING THE LINKS BETWEEN STARS, GAS, AND DARK MATTER WITH LARGE SURVEYS

by

Ting-Wen Lan

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

March, 2016

Abstract

The evolution of galaxies is driven by complex interplay between multiple components of the Universe, including stars, gas, and dark matter. To advance our understanding of galaxy evolution, it is essential to characterize the interactions between these components. This dissertation explicates the applications of statistical techniques for extracting new information from large survey data in order to investigate the relationships between stars, gas and dark matter.

By employing the cross-correlation technique which is developed to extract relevant signals in spectroscopic and photometric datasets, I study the connections between galaxies and dark matter halos. I characterize the galaxy luminosity functions in halos with different masses. This new observation uncovers an ubiquitous upturn at the faint-end of the luminosity functions, suggesting that it is formation that regulate the stellar mass of satellite galaxies. With a simple model, I show that the shape of the luminosity functions can be used to infer the formation history of galaxies.

Applying the same technique, I study the relationships between galaxies and gas around them, the circumgalactic medium. I robustly characterize the spatial distri-

bution of gas and reveal that the cool gas around star-forming galaxies and those around passive galaxies exhibit different behaviors. This result demonstrates that the dichotomy of galaxy types is reflected in the circumgalactic medium.

Finally, I explore the diffuse interstellar bands, a set of mysterious absorption lines ubiquitously seen in the interstellar medium of our Milky Way. I map the distribution of 20 different DIBs across the sky by using about half a million of SDSS spectra of quasars, galaxies, and stars. Using the map, I show that the DIB carriers have different sensitivity to the physical conditions in the ISM.

The results of the dissertation demonstrate that combining large survey data and statistical techniques is a powerful and promising approach to study galaxy formation and evolution. Pursuing this direction, one can uncover new astronomical phenomena and make great progress in our understanding of galaxy formation and evolution.

for my family and Yao-Ying

Acknowledgments

Acknowledgement 1

This dissertation would not have been possible without helps from many people. I would like to first thank my advisor Brice Ménard for his constantly support in my pursuit of academic career. I would like to thank the group members, Guangtun Zhu, Gail Zasowski, Mubdi Rahman, Alex Mendez, and Murdock Hart for their kindness and intellectual wisdom. It has been a great fortune for me to work with such an amazing group.

It is an indispensable life experience for being a member of JHU physics and astronomy department. I would like to thank Tim Heckman, Marc Kamionkowski, Nadia Zakamska, and Joe Silk for their encouragement and support. I am also grateful to the interaction with Katy, Liang, Rachael, Raymond, Kirill, Mike, Ali, Zhilei, Mohammad, Ting, Edwin, and Zack.

I also want to thank my collaborators, Houjun Mo, Nao Suzuki, and Masha Murga. It has been a great pleasure to work with them. I have learned so much from them.

ACKNOWLEDGMENTS

The life in the USA would be much difficult without friends. I would like to thank Chih-Han, Po-Yao, Reiko, Fengwei, Para, Andrew, Yi-An, Jeff, Li-Fang, John, Tzuyu and many other Taiwanese friends.

Finally, I would like to thank my family and Yao-Ying. Their constant love is my guide. They give me courage and faith to overcome every challenge in my life.

Acknowledgement 2

In the past five years, I was supported by Johns Hopkins University through the William Gardner fellowship, and by NSF and NASA through grants awarded to Professor Brice Ménard: NSF grant AST-1109665 and NASA grant 12-ADAP12-0270.

This thesis makes use of data from the Sloan Digital Sky Survey. Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is `www.sdss.org`.

SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German

ACKNOWLEDGMENTS

Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

This thesis also makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration.

I also use data from Planck. Based on observations obtained with Planck (`http://www.esa.int/Planck`), an ESA science mission with instruments and contributions directly funded by ESA Member States, NASA, and Canada.

I also want to thank John Moustakas and PRIMUS team for providing their galaxy catalog.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The formation and evolution of galaxies are driven by complex interplay between multiple components of the Universe, including stars, gas, and dark matter. For example, dark matter provides the gravitational potential where galaxies form and evolve through regulating gas via various physical processes. To understand how galaxies form and evolve, it is essential to disentangle the interplay between these components by investigating their connections. To this end, the goal of this dissertation is to exploit novel statistical techniques which allow us to extract relevant signals from large surveys and to reveal new physical processes that drive the evolution of galaxies.

1.1 A brief overview of galaxy formation and evolution

Galaxies are the building blocks of the Universe. How they form and evolve is one of the most important questions in modern astronomy. It was first raised by Edwin Hubble in 1930s (Hubble, 1936) when he found that galaxies have diverse morphologies; some have smooth elliptic light profiles while some have spiral arms and bulges in the center of galaxies. He classified those galaxies according to their shapes, and placed those morphological types into a sequence—the Hubble sequence. This sequence establishes the first classification of galaxy types and suggests that each type of galaxies may represent an unique stage of the evolution of galaxies. What is the origin of these morphological features? Why do galaxies have such diverse properties? How can we make sense of the Hubble sequence? Astrophysicists have been trying to address these questions from both observational and theoretical aspects since then.

With decades of observational efforts, we have made great progress in characterizing the properties of galaxies. In addition to their morphology, galaxies have a wide range of properties, including luminosity, color, star formation rate, stellar mass, and size, etc. They also live in diverse density environments in the Universe. Using the data from the Sloan Digital Sky Survey (York et al., 2000), Blanton & Moustakas (2009) summarized the properties of galaxies at the local Universe and their depen-

dence on the density environments. The development of large optical telescopes and the Hubble space telescope in 90s also allows us to detect light from galaxies at the early Universe and to map the evolution of galaxy properties with cosmic time. For example, the recent CANDLES survey (Koekemoer et al., 2011) with Hubble space telescope provides the detail properties of galaxies at redshift 2. Besides optical astronomy, multi-wavelength observations open other avenues toward revealing different components consisting of galaxies in addition to stars, such as atomic and molecular gas and dust.

In order to interpret the observed properties of galaxies, galaxy formation theory is established within the framework of the cosmological model of the Universe. In the current paradigm of structure formation (see Mo, van den Bosch & White, 2010, for an overview), galaxies are final products of the initial density fluctuations of the Universe. High density regions at the early Universe grow through gravitational attraction of matters. Those regions with high enough matter densities will eventually collapse into gravitational bounded systems, the so-called dark matter halos. These halos, which are stable against the expansion of the Universe, provide deep gravitational potential wells where various baryon processes take place, including gas flows in the halos. Numerical simulations have been developed to model these processes and to predict the properties of galaxies with this paradigm. However, even the state-of-art simulations, such as EAGLE (Schaye et al., 2015) and Illustris simulations (Vogelsberger et al., 2014), can only reproduce a part of the observed galaxy properties

with certain assumptions on their models. Until today, we still do not know how and why galaxies have such diverse properties as Hubble presented in 1930s. How galaxies evolve is still poorly known.

1.2 Mapping the links between stars, gas, and dark matter

To provide new insights on galaxy evolution, one promising, but challenging, avenue is to study the relationships between each component of galaxies, including stars, gas, and dark matter. As mentioned above, the evolution of galaxies is strongly connected to dark matter and gas halos. If we can probe the connections and characterize them from observations, we will be able to reveal and identify key physical processes that are currently missed in our galaxy formation theory. However, such observations are challenging and therefore limited progress has been made.

1.2.1 The observational challenges

There are at least two challenges in studying the links between stars, gas, and dark matter. First of all, the properties of gas and dark matter halos can only be obtained by specific methods. Because gas in halos emits essentially no light, one needs to get the spectrum of a background object in order to extract absorption-line

features, the signature of gas. Similarly, the properties of dark matter halos can only be probed via the observable features of baryons that trace gravity, e.g., the velocity dispersion of galaxy groups and clusters. To robustly characterize the properties of gas and dark matter halos, we have to obtain a significant number of spectra—a main constraint of previous studies.

The second challenge is that one needs to detect not only the signatures of halos but also their associated galaxies in order to study the correlations between the two. Circumventing this challenge requires yet a much larger set of data to be mined—a great challenge for all researchers. These two obstacles call for new approaches to study the links between stars, gas, and dark matter in a more systematic and detailed way.

1.2.2 Combining large sky surveys and cross-correlation techniques

A new way to conduct astronomical research has emerged in the past decades. Instead of observing certain area of sky with specific targets, telescopes have been dedicatedly built to scan across entire available sky. These sky surveys detect all objects within the depth of the observations and measure their properties. For example, the Sloan Digital Sky Survey (SDSS; York et al., 2000) observes about 10,000 square degree of the sky and detects photometrically hundred millions of objects, including

stars, galaxies, and quasars in optical wavelengths. Moreover, it obtains millions of spectra of objects,which allow us to detail their properties.

Using millions of spectra, we can extract the properties of gas and dark matter halos. The mass of dark matter halos can be estimated based on the coupling of velocity dispersion of galaxy groups/clusters and total stellar mass of the systems (e.g. Yang et al., 2007). Also, the properties of gas can be learned from absorption signatures of gas imprinted in hundred thousands of bright quasar spectra in the SDSS (e.g. Zhu & Ménard, 2013). In addition, the photometric datasets inform the properties of galaxies. These large datasets provide a chance for us to overcome the challenges of studying the links between multiple components of the Universe.

Cross-correlation as an exploratory tool

To fully make use of the datasets and extract relevant signals, I apply cross-correlation techniques. By cross-correlating an observable whose properties are not well-known with many other well-known variables, we can gain insights on its nature. In general, cross-correlation is a measure that indicates the degree of how two variables vary together. It can be estimated by the covariance of two variables ,

$$\sigma_{XY} = \langle \Delta X \cdot \Delta Y \rangle = \int \int (X - \mu_X)(Y - \mu_Y) f(X, Y) dX dY, \tag{1.1}$$

where μ_X and μ_Y are the mean values of X and Y, and $f(X, Y)$ is their joint probability distribution. The brackets $\langle\rangle$ signal the expectation value of the enclosed quantity. Note that one can estimate covariance with only one variable centering with its mean

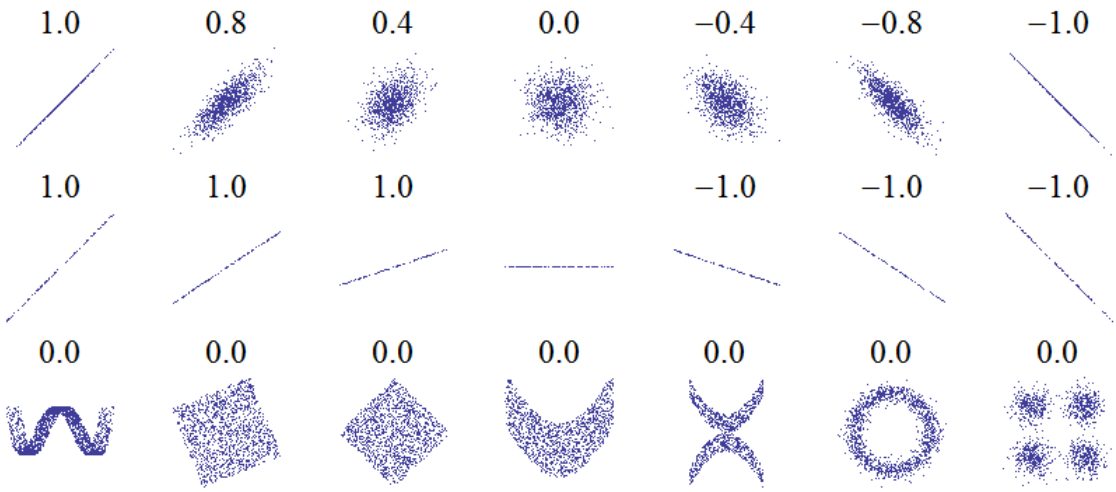

Figure 1.1: Pearson correlation coefficients for different distributions. Credit: `https://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient`

value:

$$\langle \Delta X \cdot Y \rangle = \langle (X - \mu_X) \cdot Y \rangle = \langle X \cdot Y \rangle - \mu_X \langle Y \rangle = \langle X \cdot Y \rangle - \mu_X \mu_Y = \langle \Delta X \cdot \Delta Y \rangle. \quad (1.2)$$

Non-zero covariance indicates that the X and Y variables depend on each other. Additionally, the sign of covariance reflects the direction of the correlation. Cross-correlation based on the covariance contains the units of the two variables, while some estimators of cross-correlation involve a normalization to remove the units of the variables and their individual variations (see below). However, in some cases, one might prefer to use covariance to extract quantities with physical meaning. In what follows, I will first discuss the main properties of cross-correlation by using the Pearson correlation coefficient as an example and then illustrate how we can use the concept of cross-correlation in astronomical data analysis.

The Pearson correlation coefficient is one frequently-used estimator of cross-correlation

with the units of the variables and their individual variations normalized. The correlation coefficient is defined as

$$\rho_{X,Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}, \tag{1.3}$$

where σ_{XY} is the covariance between the X and Y variables as shown in Eq. 1.1, σ_X and σ_Y are the standard deviations of X and Y. The variance—the square of the standard deviation—can be expressed as

$$\sigma_X^2 = \int (X - \mu_X)^2 p(X) dX, \tag{1.4}$$

where $p(X)$ is the probability distribution of X. The same formula applies to σ_Y^2.

The main properties of the Pearson correlation coefficient are illustrated in Figure 1.1. The top row demonstrates that if the distribution of X and Y perfectly follows a linear function, one will get the Pearson correlation coefficient of value 1 for a positive correlation and value -1 for a negative one. When the distribution of X and Y deviates from a straight line, the absolute value of the Pearson correlation coefficient decreases. The coefficient value will approach to 0 when there is no correlation between X and Y, as shown in the middle plot of the top row. In other words, the correlation coefficient describes the ability that one can infer the values of one variable from the values of the other one through a linear function. However, it is important to note that the Pearson correlation coefficient does not contain information about the underlying linear function of two variables, such as the slope of the function. This is illustrated in the middle row of Figure 1.1. As long as the X and Y variables follow an exact linear relationship, the corresponding correlation coefficients

are equal to 1 or -1. Finally, the main limitation of the Pearson correlation coefficient is that it is not optimal to probe non-linear distributions as shown in the bottom row of Figure 1.1. When the underlying relationship between two variables is nonlinear, it is possible that the correlation coefficient will predict no correlation between them. To put it differently, zero Pearson correlation coefficient value between two variables does not necessarily entail that there is no dependence between the two.

Another common estimator of cross-correlation is the Spearman's correlation coefficient. The Spearman's correlation coefficient uses the same formula of the Pearson correlation coefficient as shown in Eq. 1.3, but it replaces the values of the two variables with their ranking values. In terms of probing the correlation between two variables from a monotonic non-linear function, this difference makes the Spearman's correlation coefficient more optimal than the Pearson correlation coefficient.

In astronomical data analysis, sometimes we want to address questions about whether or not the presence of a type of objects will affect another observable. In this case, one variable in Equation 1.1 can be considered as delta functions at the locations of those objects, X. The covariance becomes the expectation value of the fluctuation of Y with respect to its mean at the locations of the objects, $\langle \Delta Y \rangle_X$:

$$\langle \delta_X \cdot \Delta Y \rangle = \int \int \delta_X (Y - \mu_Y) f(X, Y) dX dY = \langle \Delta Y \rangle_X. \tag{1.5}$$

This demonstrates that we can measure the mean change of an observable due to the presence of a type of objects through cross-correlation analysis.

The main advantage of this method is that the signal to noise ratio of the cross-

correlation signal, $\langle \Delta Y \rangle_X$, scales with the square root of the number of measurements, $\sqrt{N}$, because of the central limit theorem. This improvement holds as long as the measurements are independent. This suggests that if one can obtain 100 times more measurements than previous experiments, the ability to probe correlation signals will increase by a factor of 10. This property makes the method a powerful tool to extract weak signals from a dataset even when (1) the background signals dominate the measurements ($\mu_Y \gg \Delta Y$) and (2) the desired signal is below the noise level of individual measurements ($\sigma_Y \gg \Delta Y$). In other words, by applying cross-correlation techniques, we can substantially reduce the noise level and reveal underlying correlations between two variables from noise-dominated measurements.

Table 1.1: Examples of previous applications of the cross-correlation techniques

$\langle$Galaxy $\cdot$ Galaxy$\rangle$	Galaxy clustering
$\langle$Galaxy $\cdot$ Galaxy shape$\rangle$	Weak lensing (shear)
$\langle$Galaxy $\cdot$ Background brightness$\rangle$	Weak lensing (magnification)
$\langle$Galaxy $\cdot$ Background color$\rangle$	Dust reddening
$\langle$Galaxy $\cdot$ CMB fluctuation$\rangle$	ISW and Sunyaev-Zel'dovich effects

Thanks to large datasets provided by sky surveys in the past few decades, many discoveries and observational achievements have been made by applying the cross-correlation method as shown in Equation 1.5. For example, Fischer et al. (2000) use the SDSS data and correlate the presence of galaxies with the shapes of background

galaxies. They are able to measure the mean distortion of galaxy shapes due to weak lensing effects and infer the properties of dark matter halos. Applying the same concept, Scranton et al. (2003) cross-correlate the presence of luminous red galaxies with the flux fluctuation of the cosmic microwave background (CMB) taken by the Wilkinson Microwave Anisotropy Probe (WMAP; Bennett et al., 2003). This allows them to detect the signature of the integrated Sachs-Wolfe (ISW) effect imprinted in the CMB. Table 1.1 lists some examples of applications of the cross-correlation techniques. The left column indicates the observables and the right column shows the physical information that can be extracted from the correlation signals. All the applications make use the fact that cross-correlation allows us to enhance the signal-to-noise ratio of the desired signals, scaling with $\sqrt{N}$. These demonstrate that cross-correlation techniques are powerful tools to extract weak signals in different datasets, revealing new astronomical phenomena.

The goal of this dissertation is to utilize the cross-correlation method and maximize its applications in studying galaxy formation. A brief summary of the applications of the cross-correlation techniques in this dissertation is presented in Table 1.2 and discussed in the followings.

In Chapter 2, I cross-correlate the presence of dark matter halos traced by galaxy groups and clusters with the over density of galaxy counts above the mean of the Universe. This allows me to investigate the properties of the galaxy population residing in dark matter halos. I obtain the luminosity functions of galaxies as a function

Table 1.2: Summary of the applications of the cross-correlation techniques in this dissertation

Chapter 2	$\langle \text{Dark matter} \cdot \text{Galaxy} \rangle$	Galaxy and dark matter halo connections
Chapter 3	$\langle \text{Gas} \cdot \text{Galaxy} \rangle$	Galaxy and gas halo connections
Chapter 4	$\langle \text{Dust} \cdot \Delta\text{flux} \rangle$	ISM weak mysterious absorption lines

of dark matter halo mass with the largest dynamical range ever measured. I demonstrate that there is an ubiquitous upturn of the faint end of the luminosity functions, suggesting that it is formation, rather than halo-specific environmental effect, that plays a major role in regulating the stellar masses of faint satellites. With a simple model, I show that the shape of the luminosity functions can be used to infer the formation history of galaxies.

In Chapter 3, I study the links between the properties of galaxies and gas around them by again cross-correlating the presence of gas halos traced by their absorption signatures detected in the background quasar spectra with the over-density of galaxy counts. Using the technique and large datasets, I create an effective sample which is 10 times larger than any previous study, and characterize the distribution of the cool gas and its dependencies on galaxy properties. As a result, the cool gas of star-forming galaxies and that of passive galaxies exhibit different behaviors. This shows that the dichotomy of galaxy types is reflected in the distribution of halo gas.

In Chapter 4, I focus on detailing baryon components of galaxies. I treat our Galaxy as an unique laboratory and investigate a set of weak absorption lines, called

diffuse interstellar bands (DIBs). I introduce a new data-driven method based on cross-correlating the amount of dust, $E_{\mathrm{B-V}}$, along a line of sight with the flux decrement, Δflux, in the background spectrum. I apply this method to approximately 500,000 SDSS spectra of all types of objects and robustly detect 20 different DIBs, absorbing only a fraction 0.001 of the source flux. I produce the largest map of DIB absorption in our Milky Way with this approach. Using this map, I study the relationships between DIBs and interstellar medium (ISM) tracers, such as atomic hydrogen, molecular hydrogen, and dust. Results show that the carriers of DIBs prefer to exist in different ISM environments.

Finally, in Chapter 5, I summarize the key results of these studies and discuss the future applications of the developed methods.

The results presented in Chapter 2, Chapter 3, and Chapter 4 are based on published or submitted studies.

- Chapter 2: submitted to the Monthly Notices of the Royal Astronomical Society.
- Chapter 3: Astrophysical Journal, Volume 795, Issue 1, article id. 31, 13 pp. (2014)
- Chapter 4: Monthly Notices of the Royal Astronomical Society, Volume 452, Issue 4, p.3629-3649

Chapter 2

Probing galaxy and dark matter halo connections

2.1 Abstract

We characterize the luminosity functions of galaxies residing in $z \sim 0$ groups and clusters over the broadest ranges of luminosity and mass reachable by the Sloan Digital Sky Survey. Our measurements cover four orders of magnitude in luminosity, down to about $M_r = -12$ mag or $L = 10^7\,L_\odot$, and three orders of magnitude in halo mass, from 10^{12} to $10^{15}\,\mathrm{M}_\odot$. We find a characteristic scale, $M_r \sim -18$ mag or $L \sim 10^9\,L_\odot$, below which the slope of the luminosity function becomes systematically

This work is submitted to the Monthly Notices of the Royal Astronomical Society.

steeper. This trend is present for all halo masses and originates mostly from red satellite galaxies. The ubiquitous presence of this faint-end upturn suggests that it is formation, rather than halo-specific environmental effect, that plays a major role in regulating the stellar masses of faint satellites. We show that the observed luminosity functions of satellite galaxies can be described in a simple manner by a double Schechter function with amplitudes scaling with halo mass over the entire range of observables. Combining these conditional luminosity functions with the dark matter halo mass function, we can accurately recover the entire field luminosity function measured over 10 visual magnitudes. This decomposition reveals that the field luminosity function is dominated by satellite galaxies at magnitudes fainter than -18 mag or $L < 10^9\,L_\odot$ and central galaxies above. We find that the luminosity functions of blue and red satellite galaxies show distinct shapes and we present estimates of the stellar mass fraction as a function of halo mass and galaxy type. Finally, using a simple model, we show that the average number and the faint-end slopes of blue and red satellite galaxies can be interpreted in terms of their formation history, with two distinct modes separated by some characteristic time.

2.2 Introduction

The galaxy luminosity function is one of the most fundamental quantities describing the observable Universe. Its study was initiated by Hubble (1936) and has contin-

ued to the present day (e.g. Baldry et al., 2012; Blanton et al., 2003, 2005; Efstathiou et al., 1988; Kochanek et al., 2001; Loveday et al., 1992, 2012, 2015; McNaught-Roberts et al., 2014; Moorman et al., 2015; Norberg et al., 2002). Characterizing the luminosity function, and especially its lower-order moments, allows us to estimate important quantities describing the Universe we live in: galaxy number counts which can be related to the cosmic mean mass density and the luminosity density which can be related to the overall production of the heavy elements and the surface brightness of the night sky (e.g. Fukugita & Peebles, 2004). In addition, the luminosity function provides us with insight into the physics of galaxy formation and with constraints on the corresponding theoretical models.

It was realized long ago that the shape of the luminosity function depends on galaxy type and environment (Abell, 1962; Holmberg, 1950). Since the introduction of the Schechter function (Schechter, 1976), observed luminosity functions have traditionally been described by an amplitude, a characteristic luminosity and a faint end slope. These three parameters are believed to carry important information about the physical processes relevant to galaxy formation and evolution (e.g. Benson et al., 2003; Cooray & Milosavljević, 2005; Croton et al., 2006; Trayford et al., 2015). However, establishing such a connection can be done meaningfully only when the galaxies contained in a luminosity function all form in a similar fashion. It is therefore important to first identify the different building blocks giving rise to the overall galaxy population and then measure their respective luminosity functions separately. An im-

portant step in this direction is to study the conditional luminosity functions (CLFs) of galaxies (e.g. Cooray, 2006; Hansen et al., 2009; van den Bosch et al., 2003; Yang et al., 2003), i.e. the luminosity distributions of galaxies in systems representing the building blocks within which galaxies form and evolve.

In the current paradigm of structure formation (see Mo, van den Bosch & White, 2010, for an overview), galaxies are assumed to form in dark matter halos (e.g. Cole & Lacey, 1996; Kauffmann et al., 1993; Navarro et al., 1995; Somerville & Primack, 1999; White & Frenk, 1991), the building blocks of the cosmic web, whose mass function $\mathrm{n}(M_h)$ is thought to be known with high accuracy (Press & Schechter, 1974, and later extensions). It is then natural to introduce a mapping between the mass function and the luminosity function through

$$\Phi(L) = \int \mathrm{d}M_h \; \mathrm{n}(M_h) \; \Phi(L|M_h) \; , \tag{2.1}$$

where the conditional luminosity function $\Phi(L|M_h)$ describes the luminosity distribution of galaxies in halos of a given mass (van den Bosch et al., 2003; Yang et al., 2003). So defined, the conditional luminosity function takes us one step closer towards the understanding of galaxy formation and evolution in dark matter halos. For example, it describes the overall efficiency of star formation as a function of halo mass and halo formation histories (e.g. Yang et al., 2012). Another important dichotomy required to describe galaxy formation is the separation of centrals and satellite galaxies sharing a common dark matter halo:

$$\Phi(L|M_h) = \Phi_{\rm cen}(L|M_h) + \Phi_{\rm sat}(L|M_h) \; , \tag{2.2}$$

as it is known that their formation processes differ (e.g. Peng et al., 2010; Tal et al., 2014; Weinmann et al., 2006). Finally, considering separately passive and star forming galaxies (i.e. red/blue) is another required step, as it may provide important information about how star formation proceeds in halos of different masses at different epochs. Once each component giving rise to the overall ensemble of galaxies is characterized, detailed inferences about galaxy formation processes can be made from the observed luminosity functions. In addition, Eq. 2.1 provides us with an integral constraint or consistency check on the relationships between $\Phi(L)$, $\Phi_{\rm cen}(L|M_h)$ and $\Phi_{\rm sat}(L|M_h)$ within the current paradigm of structure formation. We will investigate this property in the present study.

The existence and possible origin of a faint end upturn in the luminosity function has been a matter of debate (e.g. Blanton et al., 2005; Loveday et al., 2012). Accurate characterization requires large, complete samples of galaxies with reliable photometry and redshift determinations. In order to bypass the need for redshift determination, most of the observational work regarding low-luminosity galaxies has concentrated on photometric galaxies in rich clusters for which contamination by interlopers is thought to be small and may be characterized. Investigations carried out so far have focused on luminosity functions in clusters of galaxies down to $M_r \sim -14$ mag. Some authors claimed the detection of a slope changing at luminosities fainter than $M_r \sim -18$ mag (e.g. Agulli et al., 2014; Banados et al., 2010; Barkhouse et al., 2007; de Propris et al., 1995; Driver et al., 1994; Jenkins et al., 2007; Milne et al., 2007; Moretti et al.,

2015; Popesso et al., 2006; Wegner, 2011). However this result has been debated (e.g. Boue et al., 2008; Harsono & De Propris, 2009; Rines & Geller, 2008).

In this chapter we attempt to settle the debate on the faint end upturn by measuring the conditional luminosity function over a wide range of halo masses and galaxy luminosities using the statistical power provided by the Sloan Digital Sky Survey (SDSS; York et al., 2000). We measure and characterize $\Phi(L|M_h)$ using galaxies selected in groups and clusters spanning three orders of magnitude in mass ($10^{12} - 10^{15}\mathrm{M}_{\odot}$) at low redshift, $z < 0.05$ or within a distance of about 200 Mpc. Using photometrically selected galaxies down to $r = 21$ we are able to probe a range of luminosities spanning over four orders of magnitude, reaching an absolute magnitude of about $M_r = -12$ mag or a luminosity of $10^7\,L_{\odot}$. Our analysis capitalizes on the method developed in Lan et al. (2014). This method can handle background subtraction accurately and has been applied successfully.

After describing the datasets in Section 2.3 and analysis method in Section 2.4, we present the measurements of conditional luminosity functions in Section 2.5 and discuss their physical interpretation in Section 2.6. Our main finding are summarized in Section 2.7. Throughout the chapter, all physical quantities are obtained by using a cosmological model with $\Omega_{\mathrm{m},0} = 0.275$, $\Omega_{\Lambda,0} = 0.725$, $h = 0.702$ (WMAP7; Komatsu et al., 2011). As a convention, halos are defined by an average mass density which is 200 times the mean density of the Universe. We note that we use M_h and M_{200}

During the completion of this chapter we became aware of a recent work done by Rodriguez et al. (2015) who present similar measurements but focus their analysis on comparisons with simulations and an interpretation in the context of halo occupation distributions.

interchangeably. Magnitudes are in AB magnitude system.

2.3 The data

2.3.1 The group catalog

To select halos as a function of mass we make use of the group catalog constructed by Yang et al. (2007) from the SDSS spectroscopic data release 7 (DR7, Abazajian et al., 2009). Galaxy groups are identified with the halo-based group finder developed by Yang et al. (2005) which assigns galaxies into groups on the basis of the size and velocity dispersion of the host dark halo represented by the current member galaxies of a group, and an iteration is used until the identification of member galaxies and the estimation of halo mass converge. Three catalogs are constructed based on three samples of galaxies: (I) SDSS galaxies with spectroscopic redshifts (spec-z) from SDSS only, (II) SDSS galaxies with SDSS spec-z plus about 7000 galaxy redshifts from other surveys, and (III) SDSS galaxies with spec-z plus galaxies which do not have redshifts due to fiber collisions but have assigned redshifts according to the redshifts of their nearest neighbors. These three samples provide nearly identical catalogs in terms of the group properties used here, namely the location, the central galaxy, and the estimated halo mass. Throughout this work, we use the catalog constructed from

We correct the offset of SDSS u-band magnitude to AB magnitude with $u_{\rm AB} = u_{\rm SDSS} - 0.04$.
`http://gax.shao.ac.cn/data/Group.html`

Sample II. We have also tested other samples and found consistent results.

The halo masses in the catalog are based on two measurements: the total luminosity or total stellar mass of all group members brighter than $M_r < -19.5$. Yang et al. (2007) showed that the two estimators provide consistent halo mass estimates. For our analysis, we adopt the halo masses, M_{200}, based on the total stellar mass and the corresponding radius, r_{200}. Following these authors, we identify the central galaxy to be the most massive member. At a given redshift, we only use groups with halo masses higher than the completeness limit presented in Eq. 9 of Yang et al. (2007). We focus on the redshift range $0.01 < z < 0.05$ so that the sample is complete for all groups with $M_{200} \geq 10^{12}\,\mathrm{M}_\odot$. The lower redshift limit is chosen to reduce the effect of distance uncertainties due to peculiar velocities. The upper limit is set by the lowest luminosities we wish to probe in this study (see below). The mass function corresponding to these groups is shown in the inset of Figure 2.1. Note that systems with $M_{200} \sim 10^{15}\,\mathrm{M}_\odot$ are one thousand times rarer than those with $M_{200} \sim 10^{12}\,\mathrm{M}_\odot$.

2.3.2 The SDSS photometric galaxies

We measure conditional luminosity functions by counting galaxies from the SDSS DR7 photometric catalog (Abazajian et al., 2009). We select galaxies with r-band model-magnitude (corrected for Galactic extinction) brighter than 21 to ensure completeness. The selection yields about 46 million galaxies within a sky coverage of about 8500 deg^2.

We estimate the absolute magnitude M_r of a galaxy around a group located at redshift z as

$$M_r = r - DM(z) - K(z), \tag{2.3}$$

where r is the reddening-corrected r-band magnitude, $DM(z)$ is the distance modulus at the redshift of the group, and $K(z)$ is the K-correction of the galaxy. We use the K-correction estimate provided by Blanton & Roweis (2007). To reduce the computing time, we use the SDSS main galaxy sample from the NYU value-added galaxy catalog (Blanton et al., 2005) with redshift from 0.01 to 0.05 and create a grid with bin size about 0.3 mag in the observed $(u - r)$ and $(g - i)$ color-color space to obtain the median K-correction of each band for each color-color bin. The apparent magnitudes of the photometric galaxies are then corrected based on the K-correction values at the nearest $(u - r)$ and $(g - i)$ bins on the grid. Because our sample has a narrow redshift range, we do not apply correction for redshift evolution. At $z = 0.01$ an apparent magnitude of $r = 21$ corresponds to an absolute magnitude of $M_r \simeq -12$ mag. As the high-mass systems selected in the group catalog are much less numerous, they tend to be found at the high end of the redshift interval which probes a larger volume. At $z = 0.05$, this reduces our ability to detect faint galaxies and allows us to reach only an absolute magnitude of $M_r \simeq -14$ mag.

`http://sdss.physics.nyu.edu/vagc/`

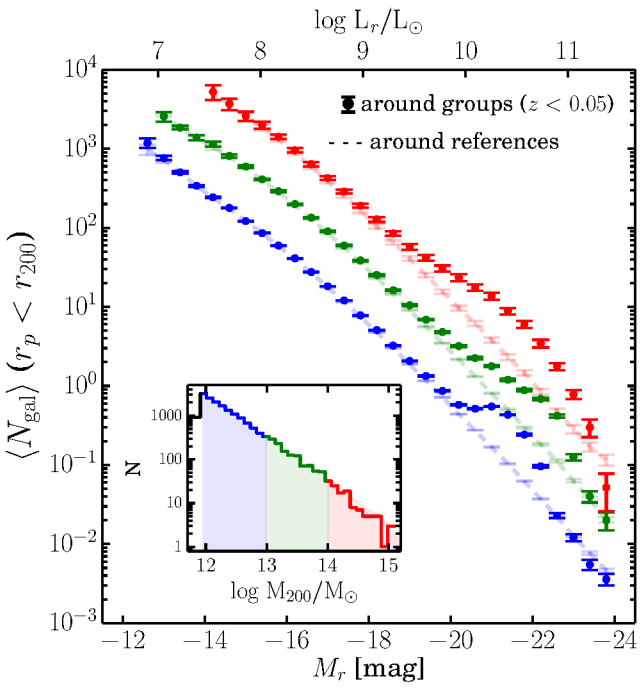

Figure 2.1: Average numbers of galaxies measured in projection around groups and random positions as a function of magnitude, for halos in three mass bins from 10^{12} to $10^{15}\,M_\odot$. The excess above random counts seen at the bright end corresponds to galaxies physically associated with the groups. The inset shows the number of groups as a function of halo mass with colors indicating the three halo mass bins.

2.4 Analysis

To infer conditional luminosity functions we cross-correlate systems selected from the group catalog (for which we have spectroscopic redshifts) with galaxy counts from the SDSS DR7 photometric dataset. For each selected galaxy group with redshift z_i, we search all photometric galaxies with projected distances within r_{200} of the halo. We convert their apparent magnitudes into absolute magnitudes with distance modules and K-corrections at z_i according to Eq. 2.3. We then estimate and subtract the contribution of uncorrelated interlopers. To do so, for a selected set of halos within a given mass bin, we first estimate the mean number of galaxies per unit magnitude in excess with respect to the background:

$$\frac{dN}{dM}(M_r) = \frac{1}{dM}\left[\langle N_{\rm gal}^{\rm grp}(M_r)\rangle - \langle N_{\rm gal}^{\rm ref}(M_r)\rangle\right], \tag{2.4}$$

where $\langle N_{\rm gal}^{\rm grp}(M_r)\rangle$ is the average number of galaxies with absolute magnitude $M_r \pm dM/2$ detected around groups in a given halo mass bin and $\langle N_{\rm gal}^{\rm ref}(M_r)\rangle$ is the average number of galaxies with the same inferred absolute magnitudes but around reference points. Subtracting the interloper contribution needs to be done carefully so as to take care of possible systematic effects due to the inhomogeneities of the photometric data produced by photometric calibration errors and by uncertainties in Galactic dust extinction correction. To test the validity of our analysis, we use two approaches:

- a global estimator: for each group we assign the redshift and the halo mass to eight random points in the SDSS footprint and use the same aperture size to

estimate the background contribution.

- a local estimator: we estimate the background contribution by counting the number of galaxies around groups from $2.5\,r_{200}$ to $3.0\,r_{200}$. This allows us to capture possible large-scale fluctuations of the zero point of the photometry.

We find that these two approaches generally yield consistent results. In Appendix A.2.1, we compare the luminosity functions derived from the two estimators as well as from the spectroscopic galaxy sample. For small halos ($M_{200} < 10^{13} M_{\odot}$), the global estimator tends to slightly underestimate the background in comparison to the local estimator. This is due to the fact that the global estimate can not account for the contribution of galaxies from nearby large scale structure of a halo even though we have attempted to exclude known large groups and clusters around small halos (see below). This effect is found to become more important for smaller halos. In what follows, results for halos with $M_{200} < 10^{13} M_{\odot}$ are obtained from the local estimator, while those for more massive halos are from the global estimator. We note that the conclusions of our analysis are unchanged with the use of either background estimator.

Figure 2.1 shows an example of the number counts within the halo radius r_{200} for halos selected in 3 bins of mass. The data points show average counts of photometric galaxies around halos and the dashed lines show the counts around reference positions. The excess seen around halos corresponds to galaxies belonging to these groups. This gives us the ability to probe the luminosity functions for galaxies spanning a range of 10 magnitudes without the need for individual redshifts.

When measuring the luminosity functions, we only use groups that are not located around the vicinity of imaging artifacts, bright stars and the edge of the survey footprint. For each group, we calculate the fraction of unmasked area within r_{200} using the STOMP library and we only use groups for which this fraction is above 95%. In addition, in order to reduce the contamination from nearby massive groups, we also exclude groups with $M_{200} < 10^{13} M_{\odot}$ that are located within r_{200} of a more massive group with $M_{200} > 10^{13} M_{\odot}$.

To estimate errors on galaxy number counts, we bootstrap the group catalog 200 times. The bootstrapping errors include both Poisson noise and cosmic variance. We also estimate the Poisson errors of the number counts and confirm that the bootstrapping errors dominates the error budget. In the analysis, we adopt the bootstrapping errors as the errors of our measurements. In Appendix A.2.2, we show tests using group samples in different redshift ranges to explore the effects of projection and sample variance. We find that different samples give similar results in the luminosity ranges they can probe, demonstrating the reliability of our results.

2.5 Measured luminosity functions

In this section we present our measurements of conditional luminosity functions. After presenting the overall behaviors, we examine in detail the behaviors at both the bright and faint ends. We then show how these conditional luminosity functions can

developed by Ryan Scranton and available at: `https://code.google.com/p/astro-stomp/`

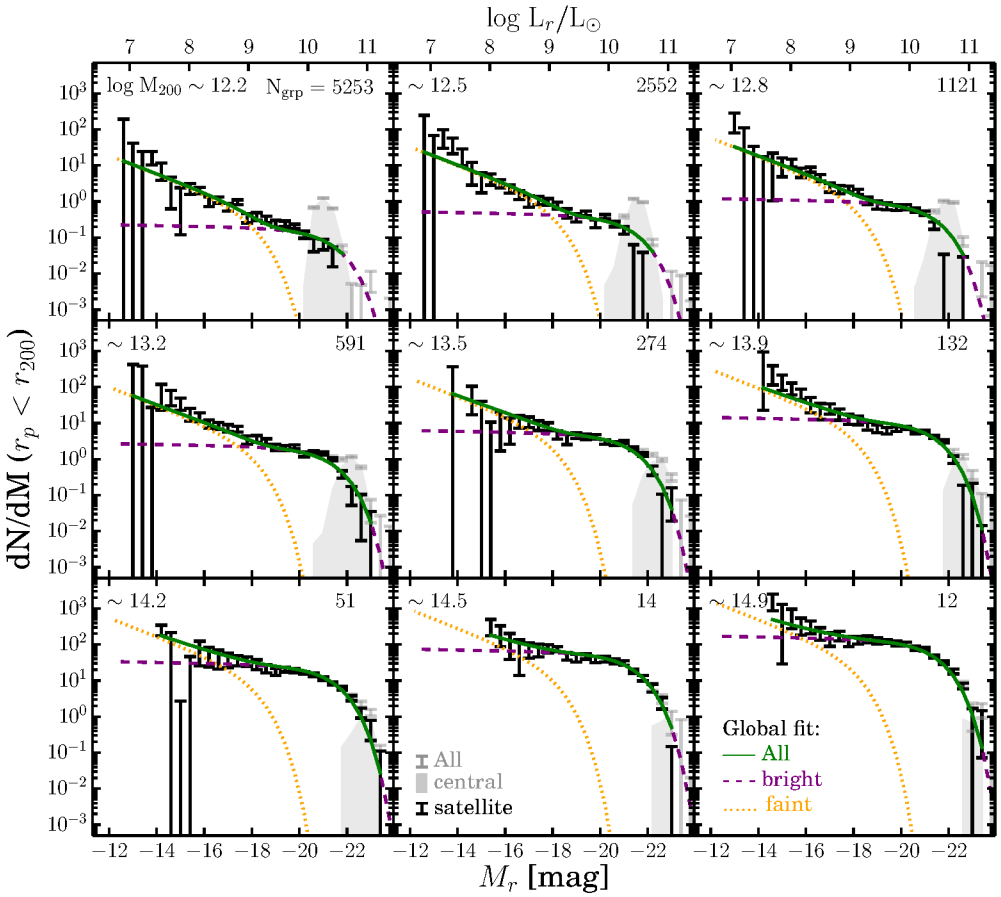

Figure 2.2: Conditional luminosity functions of galaxies in halos of mass M_{200}, shown with and without central galaxies (grey and black data points). The shaded regions highlight the contributions from central galaxies only. The global best-fits for satellite luminosity functions are shown with green solid lines. The purple and orange lines show the individual components of the double Schechter fit. Best-fit parameters are presented in Table A.2. Counts from galaxies brighter than the central galaxies are due to Poisson fluctuations introduced by the background subtraction method and are not included in the fitting procedure. The errors are estimated by bootstrapping the group sample.

be combined with the halo mass function to recover the field luminosity function of galaxies. Finally we present results separately for red and blue galaxies.

2.5.1 Overall behavior

In Figure 2.2 we present our measurements of the conditional luminosity functions in different halo mass bins. As one can see, our results cover about 10 magnitudes or 4 orders of magnitudes in luminosity, and about 3 orders of magnitude in halo mass. In each panel, the number shown at the top left indicates the mean halo mass, while the number of halos used in the corresponding mass bin is indicated at the top right. The grey data points show the luminosity functions including both central and satellite galaxies, with the grey shaded regions showing the contribution of central galaxies as obtained directly from the group catalog. The black data points show the satellite luminosity functions obtained by subtracting the contribution of central galaxies from the total luminosity functions. The color lines are the results of a global, double Schechter function fit to the conditional luminosity functions of satellite galaxies, as to be detailed in Section 2.5.3. Note that the signal to noise ratio of the luminosity functions decreases towards the faint end as a smaller fraction of groups (at the lowest redshifts) contributes to the measurements.

Inspecting these distributions, we notice the following properties:

- There appears to be a characteristic magnitude, $M_r \sim -18$ mag or $L \sim 10^9\,L_\odot$, at which the slope of the luminosity function becomes steeper toward the fainter

end. The behavior is consistent with luminosity functions of massive clusters ($M_{200} \geq 10^{14}\,M_{\odot}$) measured in previous studies (e.g. Agulli et al., 2014; Barkhouse et al., 2007; Moretti et al., 2015; Popesso et al., 2006). Here, our analysis extends these measurements to much lower halo masses, with $M_{200} \sim 10^{12}\,M_{\odot}$.

- Above this scale, the satellite luminosity functions remain flat over a few magnitudes and then decline exponentially at the bright ends $M_r < -21$ mag, as usually observed.

- There is a continuous change in the overall shape of the luminosity function with halo mass. Among all satellites, the fraction of the 'dwarf' population (e.g. $M_r > -20$) decreases with increasing halo mass. However, the trend reverses when centrals are included, reflecting that centrals are the dominant component in lower mass halos (see the shaded regions).

2.5.2 The bright end

Let us first focus on galaxies with $M_r < -18$ mag (or $L > 10^9\,L_{\odot}$). The corresponding parts of the luminosity functions are shown in Figure 2.3. Following the conventions introduced in Fig. 2.2, the black data points show the measured values for the satellite galaxies, the grey points include the contribution from centrals, and the shaded regions indicate the contribution of the central galaxies identified directly

from the group catalog.

To describe the behavior of this collection of luminosity functions of satellite galaxies, we use a Schechter function to fit the data:

$$\Phi(M_r) = N_{\rm b}\, \mathcal{F}\left(M_r; \alpha_{\rm b}, M_{\rm b}^*\right), \tag{2.5}$$

with $N_{\rm b}$ being the overall amplitude. $\mathcal{F}$ is the functional form of the Schechter function given in terms of absolute magnitude:

$$\mathcal{F}(M_r; \alpha, M^*) \equiv 10^{0.4(M^*-M_r)(\alpha+1)} \exp\left[-10^{0.4(M^*-M_r)}\right], \tag{2.6}$$

where M^* is the characteristic absolute magnitude and α is the faint-end slope. For each halo mass bin, we fit the measured satellite luminosity function over the range $M_r < -18$ mag. We exclude data points brighter than central galaxies, as they are expected to originate from Poisson errors introduced by the background subtraction. The best fit Schechter function for each halo mass is shown with the solid black line, with the best fit values presented in Table A.1 and displayed in Figure 2.5 with black data points. In the left panel of Figure 2.5, we find that $M_{\rm b}^* \sim -21.3$ mag over a large halo mass range at $M > 10^{13}\, M_\odot$, with a tendency toward fainter magnitudes for halos with lower masses. In the middle panel of Figure 2.5, the black data points show that the slope $\alpha_{\rm b}$ is roughly constant over the entire range of halo masses, with a value consistent with -1. This is in line with the observation that the satellite conditional luminosity functions appear flat over the magnitude range $-18 > M_r > -21$ mag. The right panel of Figure 2.5 shows that $N_{\rm b}$ as a function of M_{200} is well constrained,

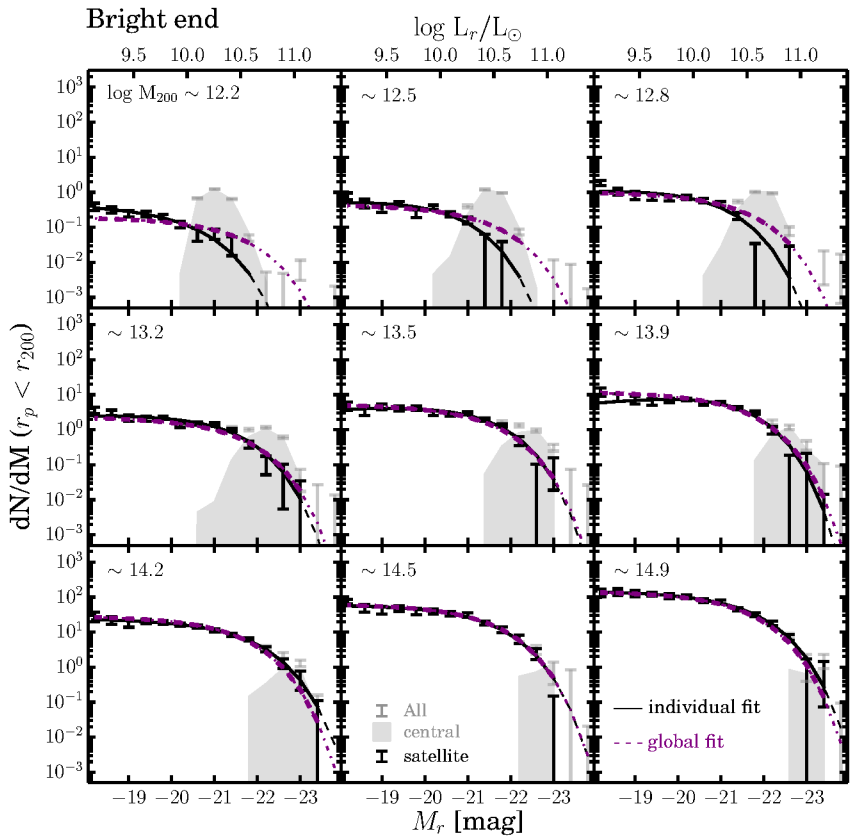

Figure 2.3: The bright end ($M_r < -18$ mag) of the conditional luminosity functions for galaxies as a function of halo mass using the same conventions as in Fig. 2.2.

and the relation can be described by

$$N_{\rm b} = A_{\rm b} \times \left(\frac{M_{200}}{10^{12} M_\odot}\right)^{\gamma_{\rm b}}, \tag{2.7}$$

where $A_{\rm b}$ is the overall normalization and $\gamma_{\rm b}$ the power index. This trend is consistent with the work of Yang et al. (2009).

These results indicate that the bright-end of the satellite luminosity functions can be characterized by four parameters: $(\alpha_{\rm b}, M^*_{\rm b})$ determining the shape of the luminosity function and $(A_{\rm b}, \gamma_{\rm b})$ governing the overall amplitude as a function of halo

mass. This simple behavior motivates us to describe the *global* behavior of the bright parts of the nine conditional luminosity functions using a single functional form with these 4 parameters. We perform such a global fit and show the best-fit luminosity functions in Figure 2.3 with the purple dashed lines. Overall this 4-parameter model provides a reasonable description of the data. The best-fit parameters are listed in Table A.2 and presented visually in Figure 2.5 as the purple dashed lines with shaded regions indicating the corresponding errors.

We note that the global fit tends to slightly overestimate the bright ends of luminosity functions for halos with $M_{200} < 10^{13}\,\mathrm{M}_{\odot}$, clearly owing to the use of a single M_{b}^{*} for all halo masses. The bright ends can be better modelled by introducing extra parameters. However, given the large error bars at the very bright ends, a Schechter function with a single M_{b}^{*} is still consistent with the data. Since this study focuses on the behaviors of the faint ends of the conditional luminosity functions, in the following we will use this simple formalism but refrain from making any strong statements about the behaviors of the satellite conditional luminosity functions at the very bright end.

2.5.3 The faint end

Next we examine the satellite luminosity functions at the faint end, i.e. with $-18 < M_r < -12$ mag (approximately $10^7 < L < 10^9\,M_{\odot}$). The existence of a steepening of the luminosity function toward the faint end can already be seen in Figure

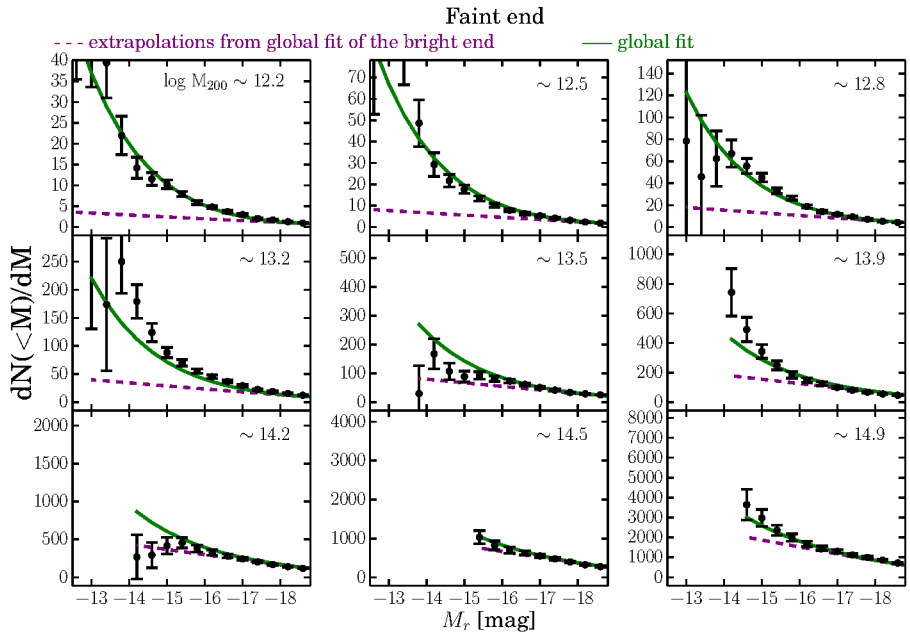

Figure 2.4: The faint end ($-12 > M_r > -18$ mag) of the conditional luminosity functions for galaxies as a function of halo mass shown in cumulative manner. The purple dashed lines show extrapolations of the best fit bright-end Schechter functions. A faint-end upturn is apparent across a wide range of halo mass.

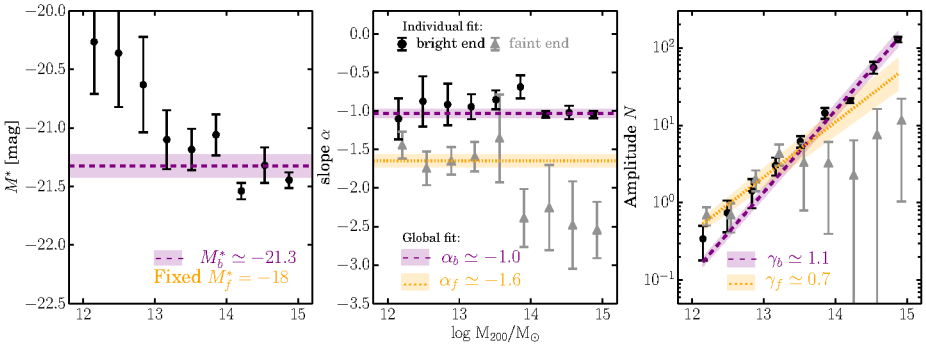

Figure 2.5: Best-fit parameters for the conditional luminosity functions as functions of halo mass described with a double Schechter function. The panels show the characteristic magnitude $M^\star$, the slopes α_b and α_f for the two luminosity ranges and the normalizations N. In each panel, best fit parameters for individual mass bins are shown with data points and the global best-fit values (over the entire halo mass range) are shown with lines and one sigma contours.

2.2. In order to demonstrate this more clearly, we show the cumulative luminosity functions for groups with different halo masses in Figure 2.4. The purple dashed lines show the values expected if one extrapolates the global best fit functions derived from the bright end with $\alpha_b = -1$. As can be seen, the measurements clearly depart from these trends, indicating a change of slope at the faint end.

In order to characterize the luminosity functions including the faint components, we choose to use the sum of two Schechter functions. Thus, for a given halo mass M_{200}, the conditional luminosity function is written as

$$\Phi(M_r) = N_{\rm b}\ \mathcal{F}(M_r; \alpha_{\rm b}, M^*_{\rm b}) + N_{\rm f}\ \mathcal{F}(M_r; \alpha_{\rm f}, M^*_{\rm f})\,, \tag{2.8}$$

where the subscripts 'f' and 'b' indicate the faint and bright components, respectively. The combination of two Schechter functions leads to a high degree of degeneracies between model parameters. To simplify the problem and limit potential degeneracies, we use two simplified assumptions motivated by the data: (i) For the bright component, we use the global bright-end best fit parameters obtained in the previous section, namely we take $M^*_{\rm b} = -21.27$, $\alpha_{\rm b} = -1.01$, $A_{\rm b} = 0.12$ and $\gamma_{\rm b} = 1.06$; (ii) We fix the characteristic magnitude to be $M^*_{\rm f} = -18$ mag where the slope appears to change. This leaves us with two free parameters $(N_{\rm f}, \alpha_{\rm f})$ to describe the faint end behavior for a given halo mass. The best fit parameters are shown as the grey points in Figure 2.5 and their values are listed in Table A.1. As shown in the middle panel of Figure 2.5, the slopes of the faint ends $\alpha_f \sim -1.6$ are steeper than the slopes of the bright components $\alpha_b \sim -1$ for all halo masses, demonstrating the ubiquitous

upturn of the conditional luminosity functions shown in Figure 2.4. The right panel of Figure 2.5 shows $N_{\rm f}$ as a function of halo mass. This relation is consistent with a power law. As for the bright parts, we also perform a global fitting to the faint ends of the conditional luminosity functions with three free parameters, ($\alpha_{\rm f}$, $A_{\rm f}$, and $\gamma_{\rm f}$). The parameters obtained from the fit are shown as the orange dashed lines, with the shaded regions indicating the errors. The values are listed in Table A.2.

Together with the global best-fit parameters for the bright ends, we have a double Schechter function (Eq. 2.8) which is specified by eight parameters $(M^*_{\rm b}, \alpha_{\rm b}, A_{\rm b}, \gamma_{\rm b})$ and $(M^*_{\rm f}, \alpha_{\rm f}, A_{\rm f}, \gamma_{\rm f})$. The global best fit functions are shown as the solid green lines in Figure 2.2 and 2.4. As can be seen, this functional form provides a reasonable description of the data over the entire range of halo masses The reduced χ^2 of the fit is 1.72 for a total of more than 200 data points. For reference, the bright and faint components are plotted separately as the purple dashed and orange dotted lines in Figure 2.2. The results indicate that the simple functional form and the parameters obtained are adequate to describe the luminosity functions of the satellite galaxies in the luminosity range $-12 > M_r > -23$ mag in halos with masses spanning 3 orders of magnitude. This suggests that the satellite population has a simple relation to the host dark matter halos, as to be discussed in Section 2.6.3.

2.5.4 Decomposition of the field luminosity function

The general (field) galaxy luminosity function has been measured by numerous authors (see Johnston, 2011, for a review). With current large surveys, the field luminosity function can now be measured down to ~ -12 mag (e.g. Blanton et al., 2005; Loveday et al., 2015). As indicated in Introduction, the conditional luminosity functions are related to the field luminosity function according to Eq. 2.1. One can therefore use this relation to (i) test the validity of the paradigm of galaxy formation within dark matter halos and (ii) explore the halo mass range that effectively contributes to the observed luminosity function. Separating galaxies into centrals and satellites we can write

$$\Phi(L) = \int_{M_{min}}^{\infty} dM_{200}\, \mathrm{n}(M_{200}) \times \left[\Phi_{\mathrm{cen}}(L|M_{200}) + \Phi_{\mathrm{sat}}(L|M_{200})\right] . \tag{2.9}$$

The field luminosity function can therefore be estimated using a dark matter halo mass function, our measured satellite conditional luminosity functions, and an estimated contribution from central galaxies. For the mass function of dark matter halos we follow Sheth, Mo & Torman (2001) and estimate it at the mean redshift of our sample, $z = 0.03$. For the satellite conditional luminosity function, we use the global best fit given by Eq. 2.8. Since by definition the central galaxy in a host is the brightest, we

We obtain the mass function of dark matter halos from `http://hmf.icrar.org/` by Murray et al. (2013).

consider only satellite galaxies fainter than the central of their hosts. Finally, as an estimate of the luminosity function of central galaxies, we use the relation between central luminosity and halo mass given by Yang et al. (2009). Since we only consider the average contribution of central galaxies, the scatter (~ 0.15 dex) in this relation can be ignored. We thus have

$$L_c(M_{200}) = L_0 \frac{(M_{200}/M_1)^{\alpha+\beta}}{(1+M_{200}/M_1)^{\beta}}, \tag{2.10}$$

where M_1 is the characteristic halo mass so that $L_c \propto M_{200}^{\alpha+\beta}$ for $M_{200} \ll M_1$ and $L_c \propto M_{200}^{\alpha}$ for $M_{200} \gg M_1$. We use the best-fit values for these parameters provided by Yang et al. (2009) and calibrate M_1 to be consistent with the halo mass M_{200} and cosmology used in this study. The values of these parameters we use are $(\log\ L_0, \alpha, \beta, \log\ M_1) = (10.22, 0.257, 3.40, 11.21)$.

In Figure 2.6, we present the luminosity functions for satellites and central galaxy for different halo masses. In the left panel, we present the contribution from different halo mass. Our decomposition shows that the bright end of the satellite luminosity function is dominated by galaxies in massive halos ($M_{200} > 10^{13} M_{\odot}$), while the faint end is mostly contributed by galaxies in relatively small halos ($M_{200} < 10^{13} M_{\odot}$). For satellites, the bright end cutoff originates from the luminosity of the corresponding central galaxies at a given halo mass (note that satellites galaxies are assumed to always be fainter than their associated central galaxy).

In the right panel, the vertical dashed lines show the contribution of central galaxies in different halo masses. Based on Equation 2.10, we note that the absolute mag-

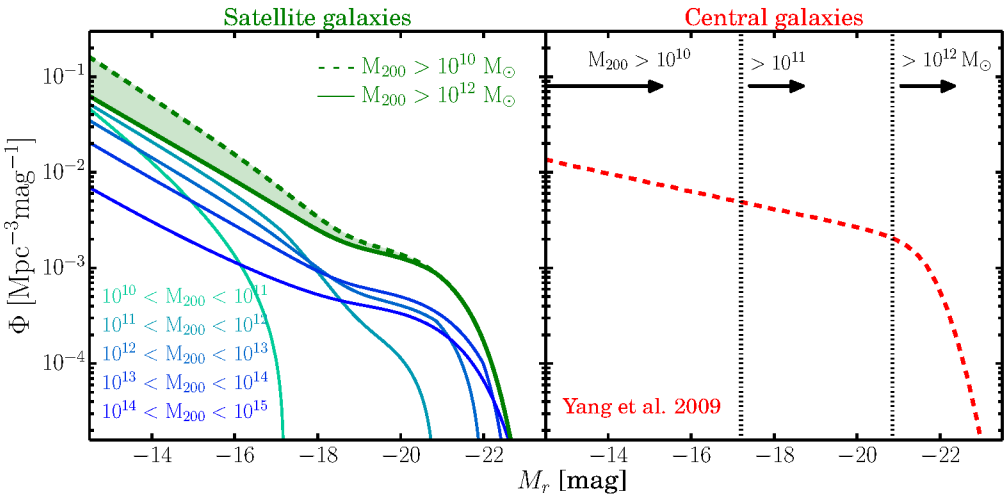

Figure 2.6: Luminosity functions corresponding to different halo masses. The left panel shows contributions from satellites derived from our global best-fit conditional luminosity functions weighted by the dark matter halo mass function. The top two curves show the cumulative contributions from halos with $M_{200} > 10^{12}$ and $10^{10}\,M_\odot$. The right panel shows the luminosity function of central galaxies according to Yang et al. (2009). The vertical lines indicate the ranges where different halo masses contribute.

nitude of central galaxies in halos with mass $10^{10}\,M_{\odot}$ is about -8, which is beyond the luminosity range of the figure. Central galaxies with absolute magnitudes brighter than -14 (the limit plotted) reside in halos with $M_{200} > 10^{10.5}\,M_{\odot}$.

Having shown the individual terms of Equation 2.9, we now present the reconstructed luminosity function and compare it to the global field luminosity function. In Fig. 2.7, the green data points are the satellite luminosity function obtained from our measured conditional luminosity functions, with the open points indicating the regions where the conditional luminosity functions may become incomplete because of the redshift distribution of our groups. The green solid and dashed lines are the satellite luminosity functions estimated from our global best-fit Schechter functions with the integration of halo mass down to 10^{12} and $10^{10}M_{\odot}$, respectively. The red dashed line shows the contribution from central galaxies, estimated from Equation 2.10 by including all halos with $10^{10}\,M_{\odot}$. The blue solid and dashed lines are the corresponding field luminosity functions calculated by adding the contribution of central galaxies (red line) to these two estimates of satellite contribution, respectively. The black triangles show the field luminosity function based on SDSS spectroscopic galaxies at $z < 0.05$ after correcting the incompleteness of the spectroscopic survey (Blanton et al., 2005).

The luminosity function obtained by combining the contribution of central and satellite galaxies is very similar to the observed luminosity function covering some 10 magnitudes. In the bright end, the luminosity function is dominated by central

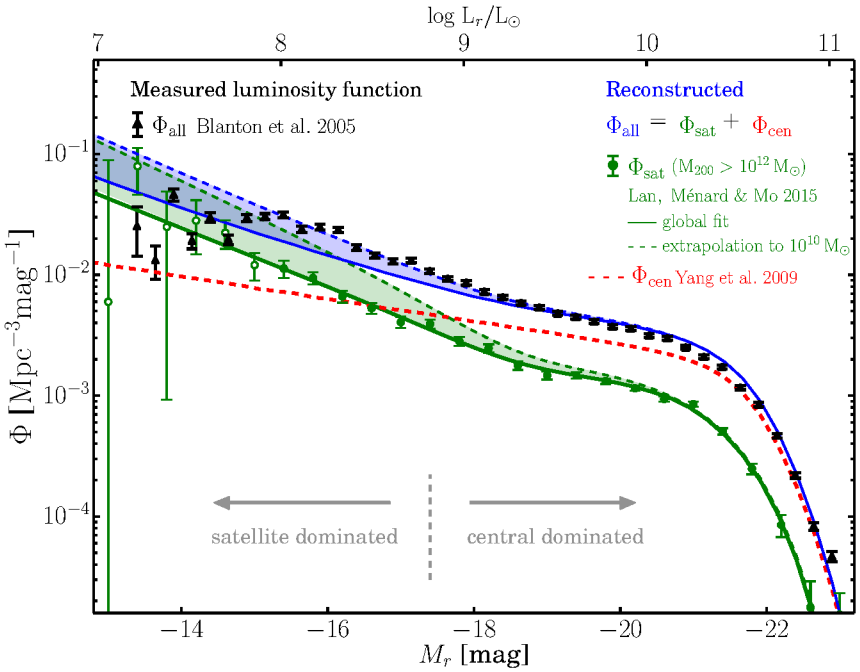

Figure 2.7: Reconstruction of the field luminosity function obtained by combining the satellite luminosity functions weighted by the dark matter halo mass function (green data points and curves) and the luminosity function of central galaxies (red curve) as estimated by Yang et al. (2009) for the same sample of halos. For comparison we show the direct measurement of the field luminosity function obtained by Blanton et al. (2005) using SDSS spectroscopic data (black data points) for a smaller volume. Taking into account halos down to masses of $10^{10} M_\odot$ leads to a remarkable agreement between the reconstructed and measured luminosity functions. We can observe that at magnitudes fainter than about -17.5, as indicated by the vertical line, the field luminosity function is dominated by the contributions of satellite galaxies.

galaxies, with some contribution from satellite galaxies in halos with $M_{200} > 10^{13}\,M_{\odot}$. This result is consistent with previous results (e.g. Cooray, 2006; Yang et al., 2009). The faint end of the luminosity function is dominated by satellite galaxies from halos with $M_{200} < 10^{13}\,M_{\odot}$. Remarkably, the composite luminosity function naturally reproduces the change of slope observed in field luminosity functions (e.g. Baldry et al., 2012; Blanton et al., 2005). Our results show that the change of slope is due to the fact that the luminosity functions of central and satellite galaxies have two distinct slopes at the faint end. Consequently, as satellites become more dominant towards fainter parts of the luminosity function, the slope of the luminosity function changes accordingly from that of central galaxies to that of satellites. The transition occurs around -18 mag where satellite galaxies start to contribute a significant fraction of the total luminosity function. This is consistent with the result of Blanton et al. (2005) who found an upturn in the slope of the luminosity function for $M_r - 5\log \mathrm{h} > -18$. However, our results demonstrate that, in order to extract meaningful physics based on the shape of the luminosity function, it is crucial to decompose the luminosity function into central and satellite populations, and into contributions from different halos.

2.5.5 Blue/red decomposition

We now study the conditional luminosity functions of blue and red satellite galaxies. To do this we use the $(u - r)$ color-magnitude demarcation suggested by Baldry

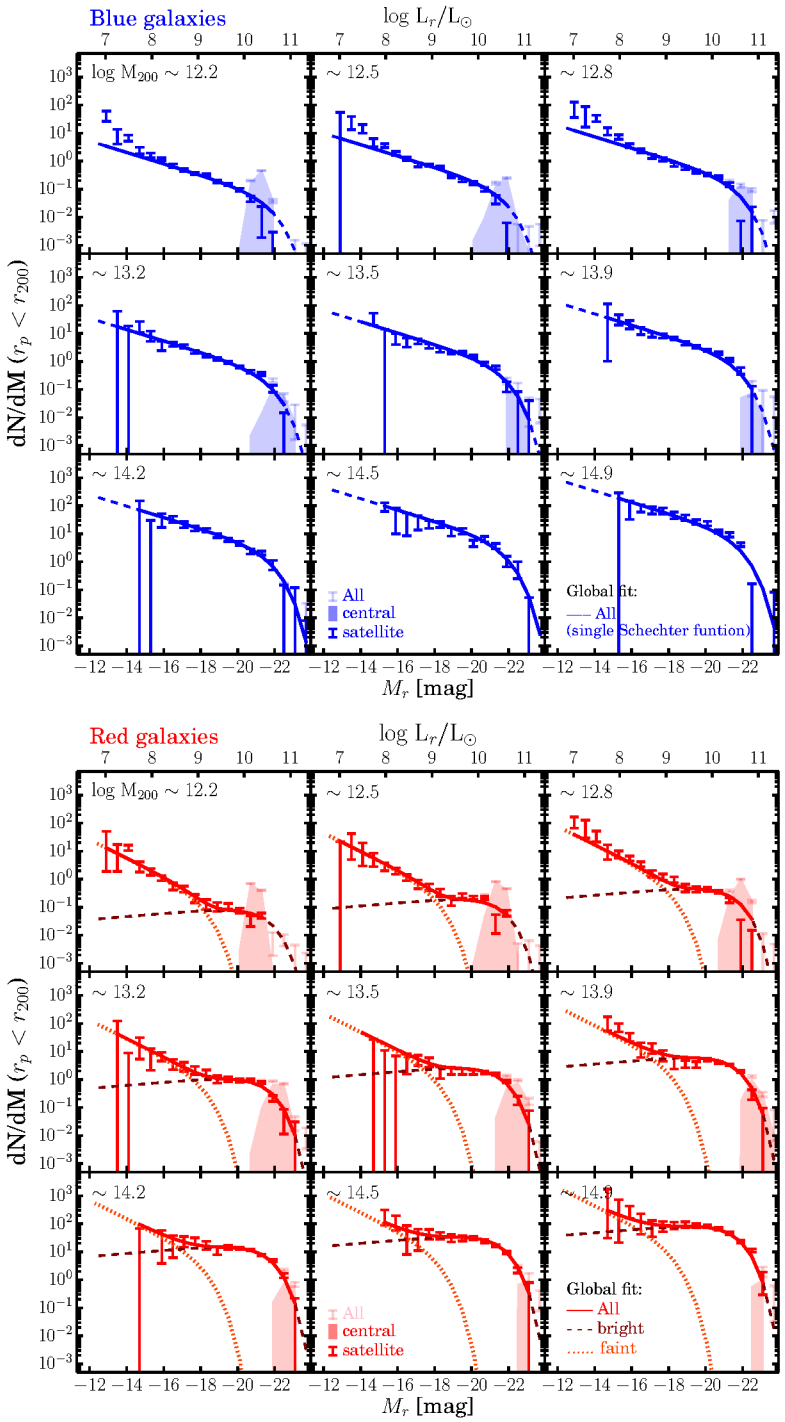

Figure 2.8: Conditional luminosity functions of blue and red galaxies. Galaxies are separated on the basis of the color-magnitude relation from Baldry et al. (2004). *Top:* The results for blue galaxies. The solid lines are the global best-fits to a single Schechter function. *Bottom:* Results for red galaxies. The global best-fits to a Schechter function in the bright end are shown with the dark red dashed lines, while those at the faint end are shown with the red dotted lines. The red solid lines are combinations of the global best-fits of both the bright and faint components. An upturn in the faint end is observed for all halo masses.

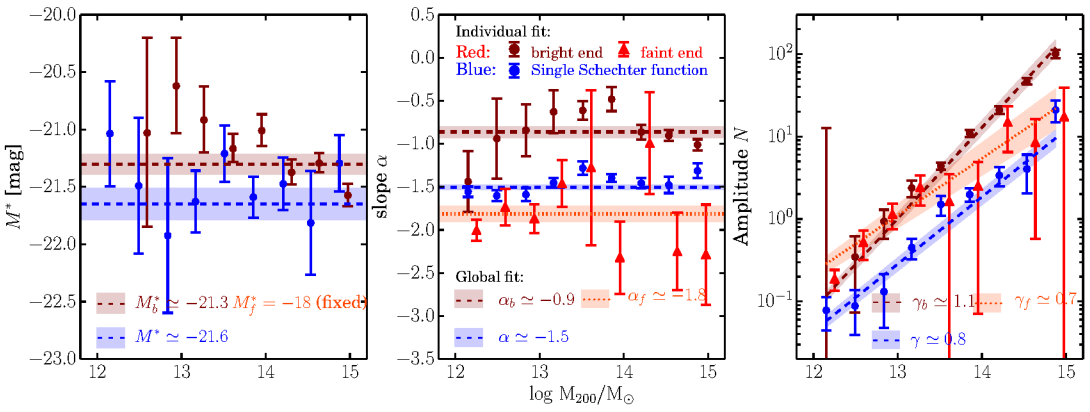

Figure 2.9: Best-fit parameters for the conditional luminosity functions of blue and red satellites described with a single and a double Schechter function. The panels show the characteristic magnitude $M^\star$, the slopes α for blue satellites and α_b and α_f for the two luminosity ranges for red satellites, and the normalizations N. In each panel, best fit parameters for individual mass bins are shown with data points and the global best-fit values (over the entire halo mass range) are shown with lines and one sigma contours.

et al. (2004) based on SDSS spectroscopic data (in Appendix A.3, we show that this choice is appropriate for our samples of galaxies in groups and clusters). We present the corresponding luminosity functions in Fig. 2.8. As done previously, we differentiate the contributions from centrals and satellites. Let us first focus on blue galaxies. Previous results (e.g. Popesso et al., 2006) suggest that a single Schechter function is capable of describing the conditional luminosity functions of blue satellites, and we therefore adopt such a model for the blue population. The blue lines in the top panel of Fig. 2.8 show the global best-fit single Schechter function. With a reduced χ^2 of 1.95, the global model of single Schechter function is found to be adequate to describe the CLFs of blue satellites. The best-fit parameters are shown by the dashed blue lines in Fig. 2.9, with errors shown by shaded regions. For comparison,

the individual best-fit parameters are shown as the blue data points. As one can see, the faint end slopes of the blue satellite luminosity functions are quite independent of halo mass. The characteristic absolute magnitudes is also roughly constant, with $M^* \approx -21.6$. The right panel shows N as a function of halo mass. This relation can be well described by a power law like that given by Eq. (2.7), with $A = 0.04$ and $\gamma = 0.82$.

For red satellites, an upturn is seen for all halo masses at the faint end, as shown in the bottom panel of Fig. 2.8. This, together with the absence of a strong upturn in the conditional luminosity functions of blue satellites, indicates that the faint end upturns of the global functions seen in Fig. 2.2 are driven by red satellite galaxies. This trend is consistent with luminosity functions of clusters measured from previous studies (e.g. Agulli et al., 2014; Barkhouse et al., 2007; Christlein & Zabludoff, 2003; Popesso et al., 2006). Blanton et al. (2005) and Moorman et al. (2015) also showed that the field luminosity function of red galaxies becomes steeper at the faint end.

To quantitatively describe the conditional luminosity functions of red satellite galaxies, we perform the same analysis as for the total population, by first characterizing the bright ends of the functions. In Figure 2.9, the dark red data points show the best-fit Schechter parameters of the bright ends for individual halo mass bins, and the dark red dashed lines show the values for the global best-fit values. The $N_{\rm b}$ - M_{200} relation is described by a power law with $A_{\rm b} = 0.08$ and $\gamma_b = 1.1$. This relation is comparable to that for the total population shown in Fig. 2.5 but steeper than that

for blue satellites. This suggests that the number of bright red satellites increases with halo mass faster than bright blue satellites, i.e. bright red satellites have the preference to live in more massive halos.

To quantify the faint components of the conditional luminosity functions of red satellites, we again first fix the Schechter function at the bright ends, using the global best-fit parameters $(M^*_{\rm b}, \alpha_{\rm b}, A_{\rm b}, \gamma_{\rm b}) = (-21.28, -0.85, 0.08, 1.11)$ obtained above. In addition, we set $M_{\rm f} = -18$ mag. The best-fit parameters for the faint components of the double Schechter function, $\alpha_{\rm f}$ and $N_{\rm f}$ for individual halo mass bins are shown by red triangles in Fig. 2.9, with the red dotted lines showing the global best-fit parameters. The global $\alpha_{\rm f}$ value for red faint galaxies is about -1.8, only slightly steeper than that of blue galaxies (for which $\alpha_{\rm f} \approx \alpha_{\rm b} \approx -1.5$ because their conditional luminosity functions can be described by a single component) and that of the total sample ($\alpha \approx -1.6$). The $N_{\rm f} \propto M^{\gamma}_{200}$ relation for red satellites has an index $\gamma_f \simeq 0.7$ which is similar to the value inferred for blue satellites. This suggests that the red-to-blue ratio is quite independent of halo mass for faint satellites, in contrast to the ratio for bright satellites. The global best-fit double Schechter functions are shown with solid red lines in the bottom panel of Figure 2.8, with the dashed and dotted lines indicating the bright and faint components, respectively.

The shape of the conditional luminosity functions of red satellites changes with halo mass because of $\gamma_{\rm b} > \gamma_{\rm f}$: the bright part becomes more dominating as the halo mass increases. In terms of 'Giant-to-dwarf' ratio, the dependence goes roughly as

$M_{200}^{0.4}$. In contrast, for blue galaxies, the shape of the conditional luminosity functions and the 'Giant-to-dwarf' ratio are almost independent of M_{200}.

2.6 Interpretation

2.6.1 The baryon content of dark matter halos

In this subsection we first use our measured conditional luminosity functions to infer the conditional stellar mass functions and then use the results to study the stellar mass contents of dark matter halos. To convert luminosity into stellar mass, one typically uses a mass-to-light relation based on galaxy color (e.g. Bell et al., 2003). This requires robust color estimates. In our case, galaxies with $r \sim 21$ in the SDSS photometric sample have typical error in the $(u-r)$ color of about 1 magnitude, mostly due to uncertainty in the u-band photometry. This error will propagate into the stellar mass estimates and can bias the stellar mass function, leading to an overestimate at the high-mass end. To reduce such bias, we estimate stellar masses using the observed mean color-magnitude relations for blue and red galaxies separately. The details of this procedure are described in Appendix A.3. As our final goal is to estimate the global baryon fractions in stars, the use of average values as opposed to full color distributions is not a severe limitation. Following Bell et al. (2003), we convert the

Our test using g-band photometry to replace u does not improve the stellar mass estimate significantly. For consistency, we will adopt the $(u-r)$ color.

observed luminosity and color into stellar mass using:

$$\log\left[\frac{M_*}{M_\odot}\right] = -0.223 + 0.299\,(u-r) - 0.4\,(M_r - 4.64) - 0.1, \qquad (2.11)$$

where $(u-r)$ is the mean color of a blue or red galaxy at a given absolute magnitude M_r. The constant, 4.64, is the r-band magnitude of the Sun in the AB system (Blanton & Roweis, 2007) and the -0.1 offset corresponds to the choice of the Kroupa initial mass function (Kroupa, 2001). Using this light-to-mass relation, we convert the global best-fit luminosity functions into the corresponding stellar mass functions. The left two panels in Figure 2.10 show the estimated conditional stellar mass functions for blue and red satellites as a function of halo mass, respectively. Since a fixed M_b^* is applied to satellite galaxies for all halo masses, a slight overestimate of the stellar mass occurs at the massive ends for small halos. As the stellar masses of central galaxies are obtained using individual observed $(u-r)$ colors, the overestimate of the stellar mass of satellites can sometimes cause the stellar mass of a satellite galaxy to exceed that of the central. The dashed lines in the left panels indicate the ranges where such situation is present. In order to estimate the total stellar mass in halos of a given halo mass, we integrate the inferred conditional stellar mass functions down to low masses. We find that using $10^7 M_\odot$, which is about the minimum stellar mass reachable by the sample used here or zero lead to similar results. The results are shown in the right panel of Fig. 2.10. The color-dashed lines show the stellar mass to halo mass ratios for blue and red satellites, respectively. The grey dashed line is the total stellar mass of satellite galaxies to halo mass ratio, while the grey solid line

is the stellar to dark matter mass ratio of central galaxies. The total ratio is shown as the black line. For halos with $M_{200} < 10^{13} M_{\odot}$, the total stellar mass is dominated by the central galaxies; in contrast, for more massive halos, it is dominated by red satellites. The contributions from red and blue satellites are comparable for halos with $M_{200} \sim 10^{12} M_{\odot}$, and the contribution from blue satellites appears to increase towards lower halo masses. Note that, although there are marked upturns in the stellar mass functions at the low-mass ends for red galaxies, the low-mass galaxies in the upturns ($M_* < 10^8 M_{\odot}$) contribute little to the total stellar mass. Our results are qualitatively consistent with estimates based on data with more limited dynamical ranges (e.g. Kravtsov et al., 2014; Leauthaud et al., 2012a,b).

2.6.2 The origin of the faint-end slope of the luminosity function

Recent progress has allowed accurate characterizations of the properties of dark matter halos as well as their sub-halos produced by the accretion and survival of progenitor halos (e.g. Giocoli et al., 2008; Jiang & van den Bosch, 2014; Sheth & Torman, 1999; Sheth, Mo & Torman, 2001; Yang et al., 2011). Since galaxies are expected to have formed at the centers of these progenitors and merged into the final halo along with their hosts (e.g. Kang et al., 2005), the statistical properties of the satellite galaxies residing in present-day groups and clusters are expected to be

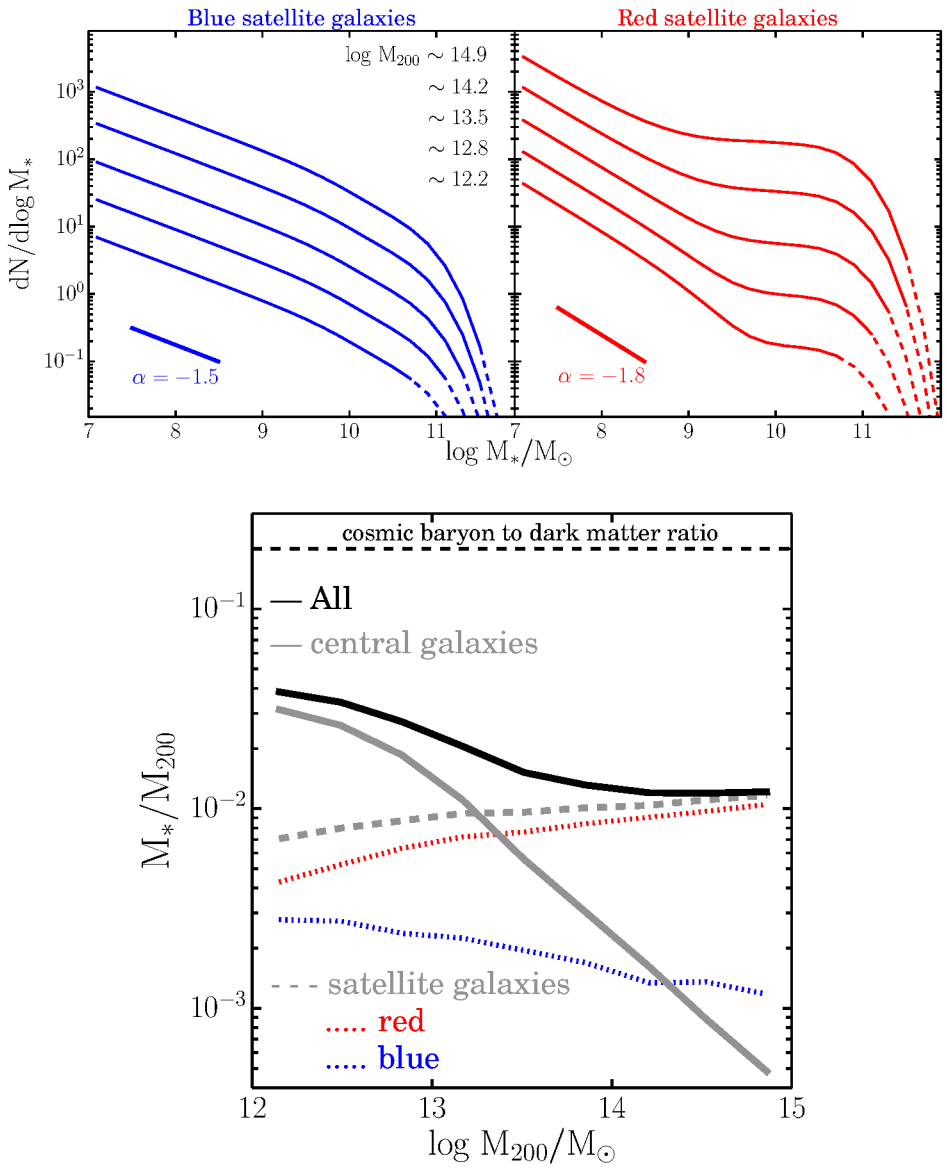

Figure 2.10: *Left*: Conditional stellar mass functions estimated for blue and red satellite galaxies. The dashed lines indicate excluded regions where satellite galaxies become theoretically brighter than centrals. *Right*: the expected stellar mass to dark matter mass ratio as a function of halo mass and galaxy types. For halos with $M_{200} > 10^{13.5}\,M_\odot$ satellites dominate. At $M_{200} \sim 10^{12}\,M_\odot$ the satellite contribution from red and blue galaxies is found to be comparable.

connected to those of the sub-halo population.

Following the idea of introducing a mapping between the luminosity and halo mass functions, one can relate the conditional luminosity function of satellites to the sub-halo mass function formally through

$$\begin{aligned}\Phi(L|M_{200}) \quad &= \int dL_a \int dm \int dz_a \, P(L|L_a, z_a, M_{200}) \times \\ &P(L_a|m, z_a) P(z_a|m, M_{200}) n_a(m|M_{200}),\end{aligned} \tag{2.12}$$

where $n_a(m|M_{200})$ is the un-evolved sub-halo mass function, $P(z_a|m, M_{200})$ describes the accretion history of a parent halo of mass M_{200}, $P(L_a|m, z_a)$ is the distribution function of initial galaxy luminosity (L_a) with respect to halo mass m and accretion redshift z_a, and $P(L|L_a, z_a, M_{200})$ is the probability for L_a to evolve into a final luminosity L (Mo, van den Bosch & White, 2010, §15.3 therein). Numerical simulations and analytical models (e.g. Giocoli et al., 2008; Jiang & van den Bosch, 2014) show that, for $m \ll M_{200}$, the un-evolved sub-halo mass function can be described by

$$n_a \propto \frac{1}{M_{200}} \Big(\frac{m}{M_{200}} \Big)^{-1-p}. \tag{2.13}$$

In the same limit and once normalized, the accretion redshift distribution, $P(z_a|m, M_{200})$, depends only weakly on the host halo mass M_{200} (e.g. Yang et al., 2011). If we make the assumptions that

- the relation between L_a and (m, z_a) is independent of M_{200},
- over a limited range of sub-halo masses, the relationship between m and galaxy

luminosity is deterministic and described by a power-law dependence,

$$L_a \propto m^{\beta}\,, \tag{2.14}$$

- the luminosities of galaxies in sub-halos do not evolve significantly so that $L \sim L_a$,

we then have

$$\Phi(L|M_{200}) \propto M_{200}^{p} L^{-p/\beta-1}\,. \tag{2.15}$$

This relation provides us with a link between the faint-end slope of the conditional luminosity function and the 'efficiency' of star formation parametrized by the index β. Considering the value of $p = 0.8$ provided by N-body simulations (e.g. Giocoli et al., 2008), we get

$$\alpha_{\rm f} \simeq -0.8/\beta - 1\,. \tag{2.16}$$

In terms of the relations given above, the measured values of the faint end slopes for satellite galaxies can be interpreted as follows.

- For blue galaxies, the observed faint end slope of the conditional luminosity function is about -1.5, implying that $\beta \approx 1.5$ (Table A2), i.e. $L \propto m^{3/2}$, and this relation holds all the way to the intermediate luminosity range at $M_r > -21$ mag.

- For red satellites with $\alpha_{\rm f} \approx -1.8$ (Table A2), implying a value of $\beta \approx 1$, i.e. $L \propto m$, which is valid for galaxies fainter than $M_r \sim -18$. For brighter objects

in the range $-18 > M_r > -21$, the conditional luminosity function is flat with $\alpha_{\rm b} \sim -1$. The scaling relation in Eq. 2.16 would then imply $\beta \gg 1$, i.e. L increases rapidly with m. However, it might also indicate that one of the simplifying assumptions breaks down in this regime. For example, the scatter in the L-m relation may not be negligible.

The different faint-end slopes, or equivalently L-m relations, for blue and red satellite galaxies suggest the existence of a dichotomy in the formation processes leading to the population of galaxies observed today. One possible interpretation is to consider a characteristic redshift z_c at which the dominant mode of galaxy formation changes. At $z > z_c$, star formation in a low halo converts a fixed fraction of its baryon mass into stars so that $L \propto m$, and such a mode of star formation may be responsible for the majority of the red satellites observed today. At lower redshifts, some processes reduce the star formation efficiency in low mass halos so that the fraction of baryon mass converted into stars in a halo is proportional to $m^{1/2}$, leading to the final scaling $L \propto m^{3/2}$.

Lu et al. (2014, 2015) reached similar conclusions by studying the redshift evolution of conditional luminosity functions. This interpretation is also consistent with the preheating model proposed by Mo & Mao (2002, 2004). In this model star formation before preheating is assumed to be in a bursting mode with a constant efficiency determined by star formation and a constant loading factor of galactic wind. After preheating, the amounts of gas that can be accreted into low mass halos are reduced

due to the raised entropy of the gas. As shown in Lu & Mo (2009), in a preheated medium, the total amount of gas that can be accreted is roughly proportional to halo mass squared, similar to what is needed to explain the faint end slope of the conditional luminosity functions for blue galaxies.

Next let us discuss why the conditional luminosity functions of red satellites have flat slopes, $\alpha \sim -1$, in the intermediate stellar mass range, 10^9 - $10^{10}\mathrm{M}_\odot$ – corresponding to the halo mass range 10^{11} - $10^{12}\mathrm{M}_\odot$ according to the relation between stellar mass and halo mass obtained for example by Lu et al. (2014, 2015). These halos have gravitational potential wells that may be deep enough so that only part of the wind material can escape. Since the escaping fraction is expected to decrease with increasing mass, $\beta > 1$ is expected, making the slope shallower than at the faint end. However, in order to get $\alpha \approx -1$, we need $\beta \to \infty$. This may indicate that the transition from a complete ejection to complete retention of galactic wind material happens over a relatively narrow halo mass range from 10^{11} - $10^{12}\mathrm{M}_\odot$.

The above discussions show that the observed luminosity functions of satellite galaxies in groups can be understood in terms of the connection between satellite galaxies and sub-halos, and that such connection contains important information about how galaxies form and evolve in dark matter halos.

2.6.3 Explaining the faint-end amplitude of the satellite luminosity function

Based on the scaling relations presented in Equation 7, we can also link the faint-end amplitudes of the conditional luminosity functions to the sub-halo mass function. Consider $N_a(m|M_{200})$, the number of satellite galaxies associated with sub-halos of mass m accreted at an earlier epoch into a host halo of mass M_{200} at the present time, we can write

$$N_a(m|M_h) \propto n_a(m|M_h) \; . \tag{2.17}$$

As discussed in the previous section, the un-evolved sub-halo mass function can be described by $n_a \propto (M_{200})^p$ with $p = 0.8$. This relation indicates that, *at the faint end*, the number of galaxies scales with host halo mass as

$$N_f(M_{200}) \propto M_{200}^{0.8} \, . \tag{2.18}$$

This is consistent with the scaling relation we found for both faint red and faint blue galaxies, $N_f \propto M_{200}^{\gamma_f}$ with $\gamma_f \sim 0.8$ (see Fig. 2.9). This indicates that the observed scaling relations may have their origins mainly in the sub-halo mass function combined with simple galaxy formation mechanisms in dark matter halos, rather than environmental effects specific to particular sets of host dark matter halos.

2.7 Summary

We have measured the luminosity functions for galaxies residing in groups and clusters with the largest possible ranges of luminosities and halo masses provided by the SDSS. Using the group catalog constructed by Yang et al. (2007) with the SDSS spectroscopic galaxy sample at $z < 0.05$, together with all photometric galaxies down to an apparent magnitude of $r \sim 21$, we can determine statistically the number counts due to galaxies physically associated with galaxy groups/clusters, and measure their luminosity functions. We have used halos with mass estimates ranging from 10^{12} to $10^{15}\mathrm{M}_{\odot}$ and measured luminosity functions from $M_r = -24$ mag down to about $M_r = -12$ mag, corresponding to luminosities spanning over four orders of magnitude, down to $L = 10^7\, L_{\odot}$.

Our results can be summarized as follows:

- The conditional luminosity functions present a characteristic magnitude, $M_r \sim -18$ mag or $L \sim 10^9\, M_{\odot}$, at which the slope of the luminosity function becomes steeper toward the fainter end. This trend is present for all halo masses. Above this luminosity scale, the luminosity functions remain flat over a few magnitudes and then decline exponentially at the bright ends, above $M_r \sim -21$ mag.

- We have shown that a double Schechter function can describe the global behavior of the data, over 3 orders of magnitude in halo mass and 4 orders of magnitude in luminosities. We have found that a set of 2×4 parameters can

reproduce more than 200 data points of measured conditional luminosity functions.

- We have shown that the luminosity functions for centrals and satellites can be combined with the halo mass function to recover the entire field luminosity function spanning 10 magnitudes (as measured by Blanton et al., 2005). This decomposition reveals that the field luminosity function is dominated by satellite galaxies at $M_r > -17$ mag, and that only halos more massive than $10^{10}\,M_\odot$ significantly contribute to the luminosity function observed above $M_r = -12$ mag.

- We have measured the conditional luminosity functions of blue and red galaxies separately as a function of halo mass. For blue galaxies, a single Schechter function provides an acceptable description of the data. In contrast, the luminosity functions of red galaxies reveal a change of slope which requires the use of a double Schechter function. These differences suggest different formation processes for red and blue galaxies.

- For blue galaxies, the observed faint end slope of the conditional luminosity function is about -1.5 all the way to the intermediate luminosity range at $M_r > -21$ mag. Using a simple model we have shown that it implies that $L \propto m^{3/2}$ for blue satellites. For red satellites we have found $\alpha_{\rm f} \approx -1.8$ for objects fainter than $M_r \sim -18$ mag, which in turn implies $L \propto m$. These

different properties can be related to differences in the formation processes of the populations of blue and red galaxies observed today.

- For both blue and red galaxies, the number of faint satellites scales with halo mass as $N_f \propto M_{200}^{0.8}$. This is consistent with the expected scaling of the number of sub-halos, indicating the direct connection between satellites and dark matter sub-halos.

Chapter 3

Probing galaxy and gas halo connections

3.1 Abstract

We explore the distribution of cool ($\sim 10^4$ K) gas around galaxies and its dependence on galaxy properties. By cross-correlating about 50,000 Mg II absorbers with millions of sources from the SDSS (optical), WISE (IR), and GALEX (UV) surveys we effectively extract about 2,000 galaxy-absorber pairs at $z \sim 0.5$ and probe relations between absorption strength and galaxy type, impact parameter and azimuthal an-

This work is published in the Astrophysical Journal, Volume 795, Issue 1, article id. 31, 13 pp. (2014)

gle. We find that cool gas traced by Mg II absorbers exists around both star-forming and passive galaxies with a similar incidence rate on scales greater than 100 kpc but each galaxy type exhibits a different behavior on smaller scales: Mg II equivalent width does not correlate with the presence of passive galaxies whereas stronger Mg II absorbers tend to be found in the vicinity of star-forming galaxies. This effect is preferentially seen along the minor axis of these galaxies, suggesting that some of the gas is associated with outflowing material. In contrast, the distribution of cool gas around passive galaxies is consistent with being isotropic on the same scales. We quantify the average excess Mg II equivalent width $\langle \delta W_0^{\rm Mg\,II} \rangle$ as a function of galaxy properties and find $\langle \delta W_0^{\rm Mg\,II} \rangle \propto SFR^{1.2}$, $sSFR^{0.5}$ and $M_*^{0.4}$ for star-forming galaxies. This work demonstrates that the dichotomy between star-forming and passive galaxies is reflected in the CGM traced by low-ionized gas. We also measure the covering fraction of Mg II absorption and find it to be about 2-10 times higher for star-forming galaxies than passive ones within 50 kpc. We estimate the amount of neutral gas in the halo of $\langle \log M_*/{\rm M}_\odot \rangle \sim 10.8$ galaxies to be a few $\times\, 10^9\,{\rm M}_\odot$ for both types of galaxies. Finally, we find that correlations between absorbers and sources detected in the UV and IR lead to physical trends consistent with those measured in the optical.

3.2 Introduction

The circum-galactic medium (CGM), tracing baryons surrounding galaxies within their own dark matter halos, plays an important role in galaxy formation and evolution. This interface between galaxies and the inter-galactic medium (IGM) is known to harbor gas flows as accretion and/or outflows but its properties and dependence on galaxy properties are poorly known. For several decades, metal absorption lines imprinted in the spectra of background sources have been the main tool to probe the gas distribution with the CGM. Studies have made use of absorber-galaxy pairs, for which the galaxy has been spectroscopically confirmed to be close to the redshift of the absorber. Various absorption lines can be used for such analyses. Among them the Mg II $\lambda\lambda 2796, 2803$ absorption lines, which trace cool gas (T $\sim 10^4$ K), have been extensively used due to their strength and visibility from the ground: $0.3 < z < 2.5$ in the optical range. From the first detection of a galaxy-Mg II absorber pair by Bergeron (1986), to the collection of a sample of about 50 systems by Steidel et al. (1994) to the latest extension to about 200 pairs by Nielsen et al. (2013a), numerous authors have attempted to find correlations between galaxy and absorber properties but the large scatter typically observed among such quantities has mostly led to non-detections (e.g. Kacprzak et al., 2011). Another line of investigation has been based on statistical analyses of large surveys (e.g. Bordoloi et al., 2011; Ménard et al., 2011; Zibetti et al., 2007). While some interesting trends have emerged from these measurements, the connection between galaxies and the distribution of cool gas in the

CGM is, after more than two decades of studies, still poorly constrained. To provide some guidance to the theoretical understanding of the CGM, additional observational constraints that can possibly reveal the physical connections between absorber and galaxy properties are needed.

In this chapter, we statistically extract about 2000 galaxy-absorber pairs using data from the Sloan Digital Sky Survey (SDSS; York et al. 2000). These pairs can then be used to measure correlations between absorber and galaxy properties such as star formation rate (SFR), stellar mass, and azimuthal angle as a function of impact parameter. We also perform our analysis to data from the GALEX and WISE surveys and provide additional support to our findings using UV and IR data. The outline of the chapter is as follows: we present the datasets and statistical estimators in Section 3.3 and the results of galaxy-absorber correlations in Section 3.4. We discuss our findings in Section 3.5 and summarize our results in Section 3.6. In this analysis, we adopt flat ΛCDM cosmology with $h = 0.7$ and $\Omega_M = 0.3$. Throughout this work we use AB magnitude system and unless stated otherwise scales are in physical units.

3.3 Data analysis

3.3.1 Datasets

We use a sample of absorber systems from the JHU-SDSS Mg II/Fe II absorption line catalog compiled by Zhu & Ménard (2013). This sample was created using an automatic algorithm to detect absorption lines in quasar spectra. The authors detected about 36,000 and 11,000 Mg II absorbers in SDSS DR7 (Schneider et al., 2010) and DR9 (Pâris et al., 2012) quasar spectra, respectively. In this analysis we focus on systems with Mg II λ2796 rest equivalent width $W_0^{\mathrm{Mg\,II}} > 0.4\,\text{Å}$ which include about 90% of the systems in the original catalog. The redshift distributions of the selected systems are shown in the bottom panel of Figure 3.1.

To select galaxies, we use catalogs from

- the SDSS DR7 photometric catalog (Abazajian et al., 2009). We select galaxies with dereddened magnitudes in the range $18 < r < 22$. With an average angular resolution of about 1.4″, SDSS allows us to probe quasar-galaxy pairs on impact parameters greater than about 10 kpc at $z = 0.5$. The magnitudes reported in this work are apparent magnitudes.

- the *Wide-field Infrared Survey Explorer* (WISE; Wright et al. 2010). We make use of the all-sky data release and select objects brighter than 20.2 in W1

`http://www.pha.jhu.edu/~gz323/jhusdss`
This absorption strength limit corresponds to conventional strong absorbers.
`http://irsa.ipac.caltech.edu/Missions/wise.html`

[3.4 μm] magnitude, which is close to the flux limit. The average angular resolution of WISE [3.4 μm] band is about 6″, corresponding to about 40 kpc at $z = 0.5$.

- the *Galaxy Evolution Explorer* (GALEX; Martin et al. 2005) all sky imaging survey (AIS). We use sources from the Data release DR4/DR5 (Morrissey et al., 2007) with NUV (1770-2730 Å) brighter than 22.5 magnitude. Similar to WISE, the average angular resolution of NUV band is about 6″.

3.3.2 Method

We measure correlations between the presence of absorbers detected in quasar spectra and photometric galaxies (without using any galaxy redshift information) detected within some angular aperture. By applying a statistical method to subtract galaxies not associated with absorbers, we are able to study correlations between galaxies and absorbers. In the following, we introduce two estimators to extract such correlations: (1) we measure the mean number of galaxies associated with selected absorbers, as a function of galaxy properties and (2) we estimate the change in absorber equivalent width as a function of proximity to a given type of galaxies.

`http://galex.stsci.edu/`

3.3.2.1 The mean number of galaxies per absorber

We statistically extract galaxies associated with absorbers by comparing the number of galaxies observed along absorber lines of sight with the number of galaxies measured at random positions in the sky. A difference is expected to arise from the presence of galaxies associated with absorbers. For each intervening absorber, the excess in the number of galaxies found within a given angular aperture is given by

$$\delta N_{\rm gal}^{{\rm abs},j} = N_{\rm gal}^{{\rm Q(abs},j)} - \langle N_{\rm gal}^{\rm Q(ref)} \rangle \ , \tag{3.1}$$

where $N_{\rm gal}^{{\rm Q(abs},j)}$ is the number of galaxies (including both associated galaxies and background/foreground galaxies) around a quasar with detected absorber j and $\langle N_{\rm gal}^{\rm Q(ref)} \rangle$ is the mean number of galaxies around reference quasars (the mean number of galaxies on the sky). We use reference quasars as opposed to random positions to also subtract a potential contribution arising from galaxies clustered with the quasars. To precisely construct a set of reference quasars, for each quasar with an absorber we randomly select four quasars with similar quasar redshift and i-band magnitude with $|\Delta z| < 0.05$ and $|\Delta i| < 0.05$ mag as reference quasars. Such selection is done without any prior knowledge of the existence of absorbers along line-of-sights. When fewer quasars satisfy these criteria, we increase the search radius in both redshift and magnitude to find matching objects. Once the reference quasars are selected, all galaxies around the absorber and reference quasars are assigned to have the same redshift of the absorber. The quantity $\delta N_{\rm gal}^{{\rm abs},j}$ can be positive or negative along a given line of

sight due to Poisson fluctuations dominated by the distributions of foreground and/or background galaxies.

By averaging the excess number of galaxies (Eq. 3.1) over an ensemble of absorbers, we obtain the mean number of galaxies associated with a Mg II absorber $\langle N_{\rm gal}^{\rm abs} \rangle$:

$$\langle N_{\rm gal}^{\rm abs} \rangle = \langle \delta N_{\rm gal}^{{\rm abs},j} \times w(\theta_j, C_j) \rangle \ , \tag{3.2}$$

where $w(\theta, C)$ is a correction function accounting for blending effects between quasars and galaxies at small angular separation when counting galaxies, θ is the angular separation between quasars and galaxies and C is the galaxy $g-i$ color. The detail of this completeness correction function $w(\theta, C)$ is presented in Appendix C.1. This correction is only important when $\theta < 4''$. We estimate the error on $\langle N_{\rm gal}^{\rm abs} \rangle$ with Poisson statistics (we note that the counts are usually dominated by the contribution from uncorrelated background objects) and we primarily focus on scales smaller than 200 kpc ($\sim 30''$ at $z = 0.5$) to probe galaxies directly associated with absorbers. This method enables us to statistically probe the properties of galaxies associated with absorbers without matching the exact galaxy-absorber pairs or assigning the properties of a certain galaxy to an absorber. It includes galaxies in all types of environments (isolated and in groups). If we assume that each absorber is connected to one galaxy, the mean number of galaxies associated with Mg II absorbers $\langle N_{\rm gal}^{\rm abs} \rangle$ represents the fraction of galaxies associated with absorbers detected within the flux limit of the survey. The quantity $\langle N_{\rm gal}^{\rm abs} \rangle$ can be measured as a function of various

parameters such as impact parameter, absorption strength, and galaxy properties (Section 3.4.1.1) to provide us with insights into the galaxy-absorber connection.

3.3.2.2 The excess MgII absorption around galaxies

We estimate the excess Mg II absorption strength with respect to the mean of the absorber sample around a given type of galaxies to investigate how the presence of a given type of galaxies influences the amount of detectable Mg II absorption. To do so we correlate the rest equivalent width of each Mg II absorber with the number of associated galaxies around it. The relative excess absorption strength around a given type of galaxies is defined as

$$\langle \delta W_0^{\mathrm{Mg\,II}} \rangle_{\mathrm{gal}} = \langle W_0^{\mathrm{Mg\,II}} \rangle_{\mathrm{gal}} - \langle W_0^{\mathrm{Mg\,II}} \rangle_{\mathrm{sample}} \,, \tag{3.3}$$

where $\langle W_0^{\mathrm{Mg\,II}} \rangle_{\mathrm{sample}}$ is the mean rest equivalent width of the absorber sample, which is equal to 1.14 Å, and $\langle W_0^{\mathrm{Mg\,II}} \rangle_{\mathrm{gal}}$ is the mean rest equivalent width around a given type of galaxies which is estimated as

$$\langle W_0^{\mathrm{Mg\,II}} \rangle_{\mathrm{gal}} = \frac{\sum_{j=1}^{\mathrm{N_{abs}}} W_{0,j}^{\mathrm{Mg\,II}}\, \delta N_{\mathrm{gal}}^{\mathrm{abs},j}\; w(\theta_j, C_j)}{\sum_{j=1}^{\mathrm{N_{abs}}} \delta N_{\mathrm{gal}}^{\mathrm{abs},j}\, w(\theta_j, C_j)} \tag{3.4}$$

with N_{abs} being the total number of absorbers, $W_{0,j}^{\mathrm{Mg\,II}}$ being the rest equivalent width of an absorber j, $\delta N_{\mathrm{gal}}^{\mathrm{abs},j}$ and $w(\theta_j, C_j)$ are given in Eq. 3.1 and 3.2. The excess rest equivalent width around galaxies $\langle \delta W_0^{\mathrm{Mg\,II}} \rangle_{\mathrm{gal}}$ can be investigated as a function of galaxy properties and impact parameter. When doing so we note that the mean

absorption strength of the absorber sample $\langle W_0^{\rm Mg\,II}\rangle_{\rm sample}$ remains the same. This will be presented in section 3.4.1.2.

3.4 Results

We first measure the mean number of galaxies per absorber $\langle N_{\rm gal}^{\rm abs}\rangle$ defined in Eq. 3.2 for the three photometric surveys presented in Section 3.3, as a function of absorber redshift. This allows us to probe the redshift range over which each survey can detect galaxies associated with the selected Mg II absorbers. In each case we use a search radius ranging from the size of the angular resolution up to a projected radius of 200 kpc. The results are presented in Figure 3.1. SDSS, WISE and GALEX all detect galaxies associated with Mg II absorbers with a wide range of redshift. At $z \sim 0.5$ about 70% of galaxies associated with our Mg II absorbers are detected in SDSS in the selected magnitude range. This fraction decreases to 30% for WISE and about 4% for GALEX. At $z > 0.8$ we find more associations between absorbers and photometric objects in WISE than SDSS, showing that the connection between galaxies and absorbers can be studied with WISE at relative high redshifts. We first study absorber-galaxy pairs found in the SDSS in Section 3.4.1.1. In Section 3.4.2 we will perform our analysis to the GALEX and WISE surveys.

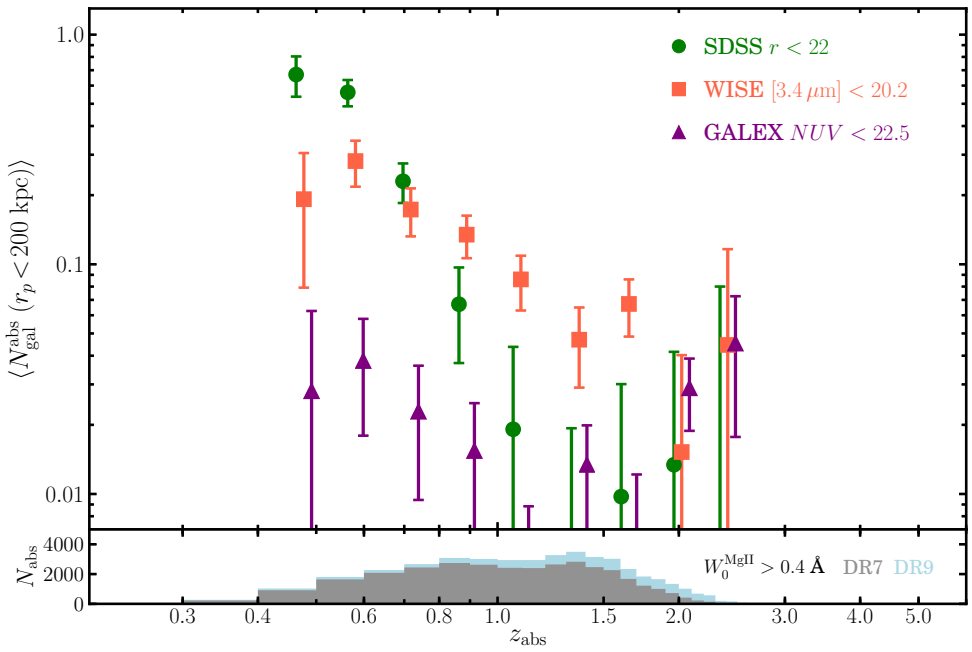

Figure 3.1: *Top*: SDSS, WISE, and GALEX mean number of galaxies per absorber as a function of $z_{\rm abs}$ from $0.4 < z_{\rm abs} < 2.8$. We search galaxies with impact parameter from 10 to 200 kpc in SDSS and 50 kpc to 200 kpc in WISE and GALEX. Three surveys all detect galaxies associated with Mg II absorbers up to $z \sim 1$. From redshift 0.4 to 0.6, SDSS detects 70% of galaxies associated with Mg II absorbers, WISE detects 30% and GALEX detects 4%. *Bottom*: The redshift distributions of DR7 and DR9 Mg II absorbers.

3.4.1 SDSS galaxy-absorber correlations

3.4.1.1 Galaxy type and spatial distribution

Focusing first on galaxies optically-selected from SDSS, we limit our analysis to the redshift range $0.4 < z < 0.6$ where the detection rate of absorber-galaxy pairs is the highest. Within 200 kpc, the mean number of SDSS galaxies around quasars with absorbers $\langle N_{\rm gal}^{\rm Q(abs)} \rangle$ is about 3.1 and the mean number of galaxies around reference quasars $\langle N_{\rm gal}^{\rm Q(ref)} \rangle$ is about 2.4. This excess corresponds to an effective number of about 2000 galaxy-absorber pairs from SDSS photometric data. This is about ten times larger than any existing galaxy-absorber pair catalog (e.g. Nielsen et al. 2013a).

Before investigating the types of galaxies connected to Mg II absorbers, it is useful to characterize the properties of galaxies in the redshift range of interest. At $z \sim 0.5$ the PRIsm MUlti-object Survey (PRIMUS; Coil et al. 2011; Cool et al. 2013) provides us with a large and complete reference galaxy sample for which both broadband colors and low-resolution spectroscopic redshifts are available. We use the first data release of the survey and measure the distribution of $g - i$ color for galaxies with redshift $0.4 < z < 0.6$ and brightness $18 < r < 22$. This distribution is well represented by the sum of two gaussians, one centered at $g - i = 1.39$ with width of 0.32 mag and the other one centered at 2.35 with a width of 0.25 mag. The relative amplitude between the blue and red is 1.4 to 1.0.

Considering the distribution of PRIMUS galaxies representative of the galaxy

`http://primus.ucsd.edu/`.

population at $z \sim 0.5$, we now investigate the properties of galaxies found to lie next to Mg II absorbers. We measure $\langle N_{\rm gal}^{\rm abs} \rangle$ as a function of Mg II absorption strength, galaxy color and impact parameter. The results are shown in Figure 3.2. The solid lines show fits of the data points using the same pair of gaussians used to characterize the color distribution of galaxies in PRIMUS (we fix the centers and the widths, as shown with the blue and red shaded regions). Our analysis shows that Mg II absorbers exist around both blue and red galaxies. Figure 3.2 shows a number of trends:

- Galaxies associated with weak absorbers (left column, $0.4 < W_0^{\rm Mg\,II} < 0.8\,$Å) are mostly found within intermediate ($50 < r_p < 120\,$kpc) and large ($120 < r_p < 200\,$kpc) impact parameters, while galaxies associated with strong absorbers (right column, $W_0^{\rm Mg\,II} > 1.5\,$Å) tend to be within small impact parameters ($10 < r_p < 50\,$kpc).

- For weak absorbers (left column), galaxies show a bimodal distribution of colors which does not strongly depend on impact parameter. In contrast, the majority of galaxies associated with strong absorbers appear to be blue galaxies with small impact parameters. Galaxies associated with intermediate absorbers (middle column, $0.8 < W_0^{\rm Mg\,II} < 1.5\,$Å) have mixed behaviors.

- At small impact parameters, our results show that the number of blue galaxies increases significantly as $W_0^{\rm Mg\,II}$ increases, showing a correlation between strong absorbers and star-forming galaxies.

To further study this correlation, we separate the galaxies according to color:

$$\text{star-forming:} \quad 0.5 < g - i < 2.0$$

$$\text{passive:} \quad 2.0 < g - i < 3.0 \; . \tag{3.5}$$

Using this separation, we show in the left panel of Figure 3.3 the mean number of star-forming and passive galaxies as a function of Mg II equivalent width for $10 < r_p < 50\,\mathrm{kpc}$. This dependence on absorber equivalent width reveals two distinct trends: over the range $0.4 < W_0^{\mathrm{Mg\,II}} < 4.5\,\text{Å}$, we observe the incidence rate of star-forming galaxies to increase by about a factor of six, indicating that stronger absorbers are more connected to star-forming galaxies within 50 kpc. On the other hand, the mean number of passive galaxies found within 50 kpc of absorbers appears to be independent on Mg II absorption strength. While our results show that Mg II absorbers live around both types of galaxies, their equivalent width appears to be primarily a function of the proximity to star-forming galaxies.

3.4.1.2 The excess absorption as a function of galaxy type

We now investigate the absorber-galaxy connection considering the excess absorption correlated with the presence of a given type of galaxies, as defined in Equation 3.3. The results are shown in the right panel of Figure 3.3 where we have used bootstrap re-sampling to estimate the errors. The measurement clearly shows that the presence of galaxies with blue $g - i$ colors within 50 kpc of Mg II absorbers increases the mean absorption equivalent width by about 0.3 Å. This amplitude reflects the *mean*

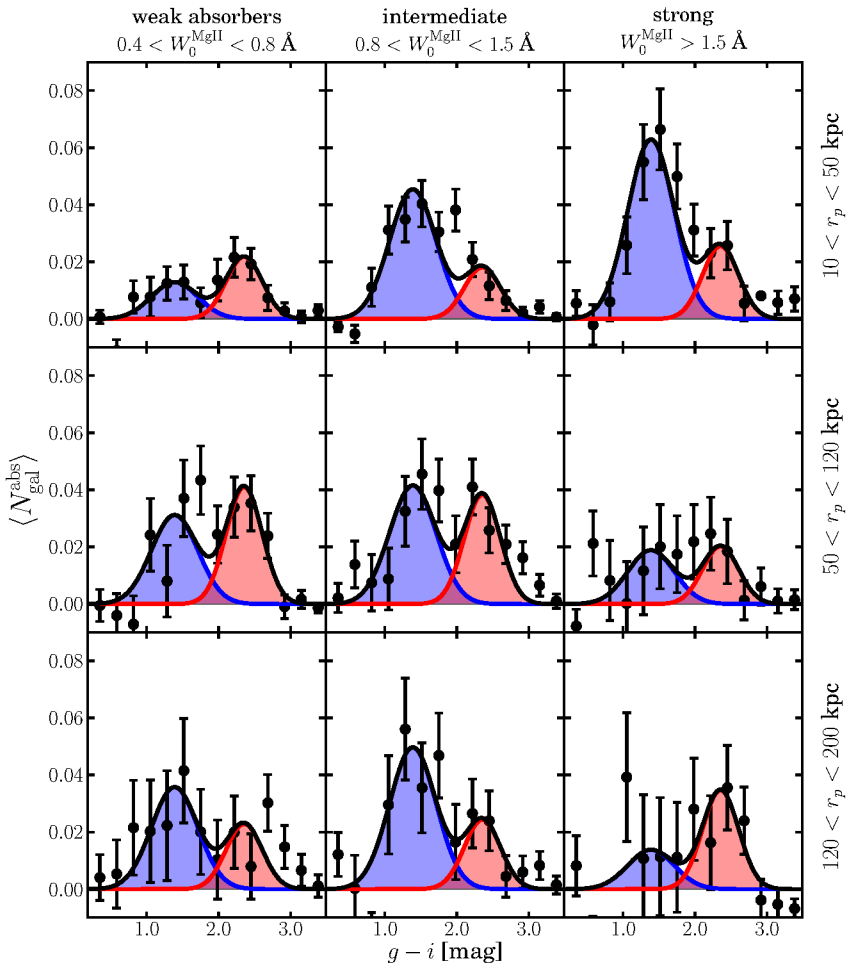

Figure 3.2: The mean number of SDSS galaxies per absorber as a function of g-i color, $W_0^{\rm Mg\,II}$, and impact parameter with $0.4 < z < 0.6$. The blue and red shaded regions are the best-fit Gaussians, for which the centers and the widths are fixed values derived from PRIMUS galaxies. The most significant trend occurs at small impact parameters where the fraction of blue galaxies increases significantly as $W_0^{\rm Mg\,II}$ increases.

correlation, i.e. averaged over all impact parameters within 50 kpc and all galaxy orientations. If a fraction f of absorbers around galaxies is responsible for this correlation, the excess equivalent width associated with these systems will intrinsically be higher by a factor $1/f$. This could be due, for example, to a preferred distribution of such systems within a given solid angle around galaxies (as will be shown in section 3.4.1.3). As shown in Figure 3.3, this excess absorption drops for galaxies with $g - i > 2.0$ and becomes consistent with zero. The transition takes place exactly at the color value separating star-forming and passive galaxies in the redshift range of interest. The dichotomy between star-forming and passive galaxies is therefore reflected in their cool, $\sim 10^4$ K gaseous halos on 50 kpc scales.

We now investigate how the excess absorption depends on physical properties of galaxies. To do so we estimate the average SFR and stellar mass of our galaxies using broadband photometry (e.g. Mostek et al., 2012), and PRIMUS galaxies at $0.4 < z < 0.6$ with estimated physical properties from Moustakas et al. (2013) as a reference. Given a galaxy with $g - i$ color and i band magnitude in our sample, we search for PRIMUS galaxies with similar observed properties $|\Delta(g - i)| < 0.05$ and $|\Delta i| < 0.05$, and use the average PRIMUS star formation rate and stellar mass values for these objects as an estimate of the parameters of the matched SDSS galaxy. When fewer than 10 PRIMUS galaxies satisfy these criteria, we increase the search radius to $|\Delta(g-i)| < 0.1$ and $|\Delta i| < 0.1$. We exclude a small subset of SDSS galaxies which do not have enough matches in PRIMUS. We note that these galaxies tend to

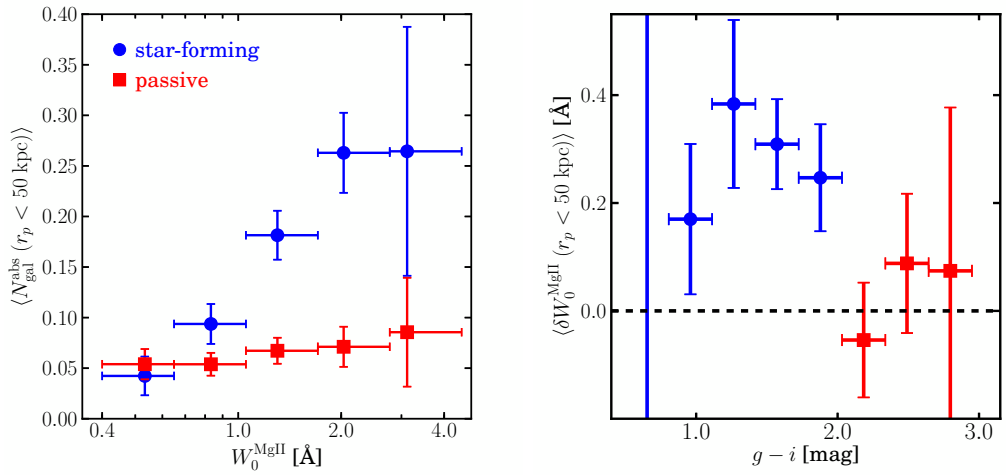

Figure 3.3: *Left:* The mean number of star-forming and passive galaxies per absorber at small impact parameter ($10 < r_p < 50$ kpc). The horizontal bars show the range of $W_0^{\rm MgII}$ values in each bin and the data point abscissae indicate the median $W_0^{\rm MgII}$ value within the bin. We apply color cuts to separate galaxies into two components: star-forming ($0.5 < g-i < 2.0$), and passive ($2.0 < g-i < 3.0$). The number of star-forming galaxies increases significantly as a function of $W_0^{\rm MgII}$. On the other hand, the number of passive galaxies remains consistent. *Right:* The excess absorption as a function of galaxy g-i color. The mean absorption strength around star-forming galaxies is about 0.3 Å higher than the mean absorption strength of the sample, while such an excess is not seen for passive galaxies. The left blue data point has value about -0.3 in y-axis which is beyond the range of our plot.

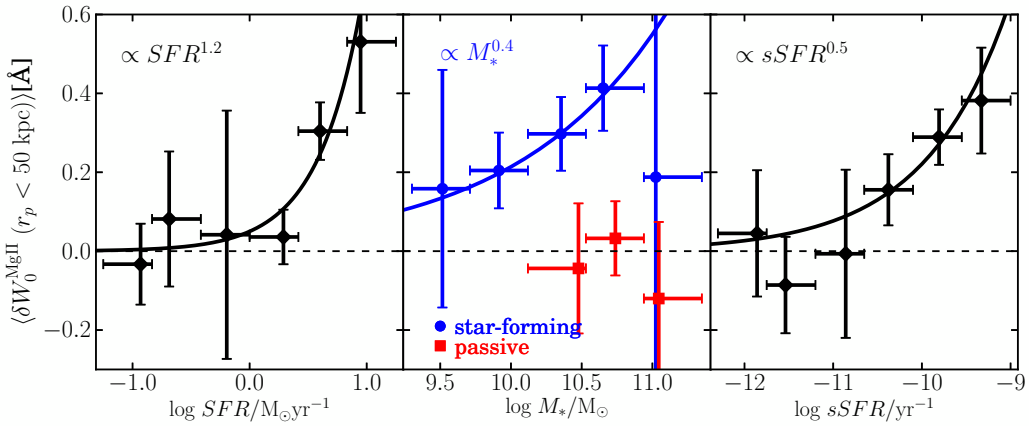

Figure 3.4: The excess absorption as a function of physical properties of galaxies. The solid lines are the best-fit power laws shown in the top-left corner of each panel. *Left*: The excess absorption as a function of SFR. *Middle*: The excess absorption as a function of stellar mass with color cuts separating star-forming and passive galaxies. *Right*: The excess absorption as a function of sSFR. The horizontal bars show the selected bin range and the data point abscissae indicate the median value within each bin.

be either too bright or too blue to be galaxies at $z \sim 0.5$. We have verified that our results are not sensitive to the exact values of the matching criteria described above.

In Figure 3.4 we show how the mean Mg II rest equivalent width changes as a function of SFR, stellar mass, and specific star formation rate (sSFR). As previously discussed, star-formation activity is correlated with excess Mg II absorption within 50 kpc. Such a correlation is detected with 5σ for SFR $\gtrsim 1\,{\rm M}_\odot/{\rm yr}$ which is roughly the value differentiating star-forming from passive galaxies. To quantify this relation we fit the observed correlation, measured over about two and half orders of magnitude in SFR: $-1.2 < \log SFR/{\rm M}_\odot{\rm yr}^{-1} < 1.2$, with a power law function and find

$$\langle \delta W_0^{\rm Mg\,II} \rangle \quad \propto \quad SFR^{\alpha} \ \text{ with } \alpha = 1.2 \pm 0.4\,. \tag{3.6}$$

This relation is shown as the solid line in the figure. The middle panel shows how the excess absorption depends on stellar mass. Here we find two distinct trends: a positive correlation between absorption strength and stellar mass for star-forming galaxies and no correlation for passive galaxies. Similarly to what was done above, we quantify this relation by fitting the observed trend with a power law function and find

$$\langle \delta W_0^{\mathrm{Mg\,II}} \rangle \quad \propto \quad M_*^{\beta} \text{ with } \beta = 0.4 \pm 0.3 \tag{3.7}$$

for star-forming galaxies and

$$\langle \delta W_0^{\mathrm{Mg\,II}} \rangle \quad \sim \quad 0 \text{ for passive galaxies.} \tag{3.8}$$

Finally we measure the correlation between absorption equivalent width and specific star formation rate sSFR $= SFR/M_*$. This is shown in the right panel of the figure. We detect a positive correlation for galaxies with $\log sSFR/\mathrm{yr}^{-1} > -11$. A power-law fit, measured over three orders of magnitude in sSFR, gives

$$\langle \delta W_0^{\mathrm{MgII}} \rangle \quad \propto \quad sSFR^{\gamma} \text{ with } \gamma = 0.5 \pm 0.2\ . \tag{3.9}$$

The above correlations are all reported for impact parameters smaller than 50 kpc. They weaken and become consistent with zero on larger scales. This is illustrated in Figure 3.5 where we show the excess absorption *within* a given radius. Note that the measurements are therefore correlated. While the excess absorption gradually decreases for star-forming galaxies as a function of impact parameter, it is always

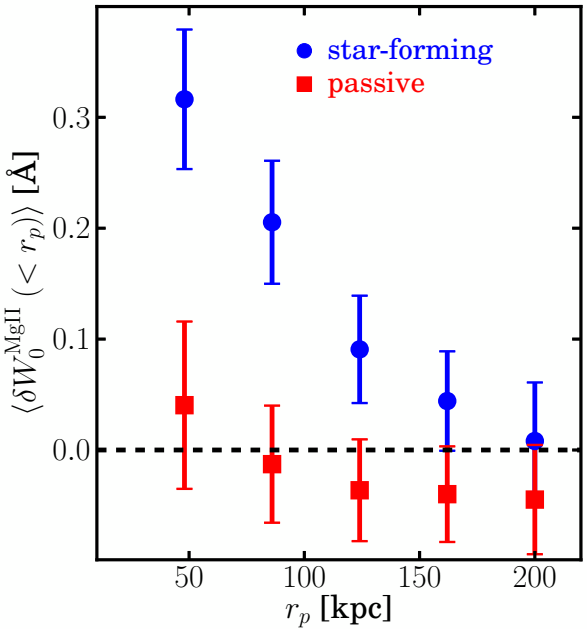

Figure 3.5: Excess absorption measured within a given impact parameter. Note that the measurements are strongly correlated. For passive galaxies, the excess absorption is consistent with zero. On the other hand, the excess absorption of star-forming galaxies decreases when averaging over a larger area. The value of data points in the x-axis represents the maximum impact parameter of each bin.

consistent with zero for passive galaxies. The above relations present a set of observational constraints for theoretical models of gas accretion and outflows around galaxies.

3.4.1.3 Azimuthal angle dependence

We now explore the azimuthal angle dependence of the correlation between galaxies and absorbers. To do so we make use of the minor-to-major axis ratio (b/a) and

the position angle of galaxies retrieved from SDSS skyserver. We select edge-on galaxies (with axis ratio $b/a < 0.55$) as a function of their azimuthal angle with respect to the quasar line-of-sight and measure the mean number of galaxies found above the background, within an aperture of 50 kpc. The results are shown in Figure 3.6.

Passive galaxies do not show any preferred orientation for the whole range of $W_0^{\rm MgII}$. On the other hand, we find that the incidence rate of star-forming galaxies around absorbers depends on absorption strength and orientation. At $W_0^{\rm MgII} > 1.5$ Å, we find about two times more systems perpendicular to the plane than along the direction of the disk. This trend is only observed at small impact parameters, i.e., with $r_p \lesssim 50$ kpc. This geometry suggests that strong absorbers might be related to outflow gas produced from star-forming galaxies. Our results are consistent with those of Bordoloi et al. (2011), Bouché et al. (2012) and Kacprzak et al. (2012), and also the stacking measurement of the Ca II absorption at $z \sim 0.1$ by Zhu & Ménard (2013).

3.4.1.4 MgII covering fraction

To fully characterize the absorber-galaxy connection one should investigate both the incidence rate of galaxies around absorbers and the incidence rate of absorbers

`http://cas.sdss.org/dr7/en/`

It should be noted that such a measurement does not require a precise position angle estimate for individual objects. It only depends on the correlation between azimuthal angle and the presence of an absorber. This is similar to the correlations measured in galaxy-galaxy lensing analyses.

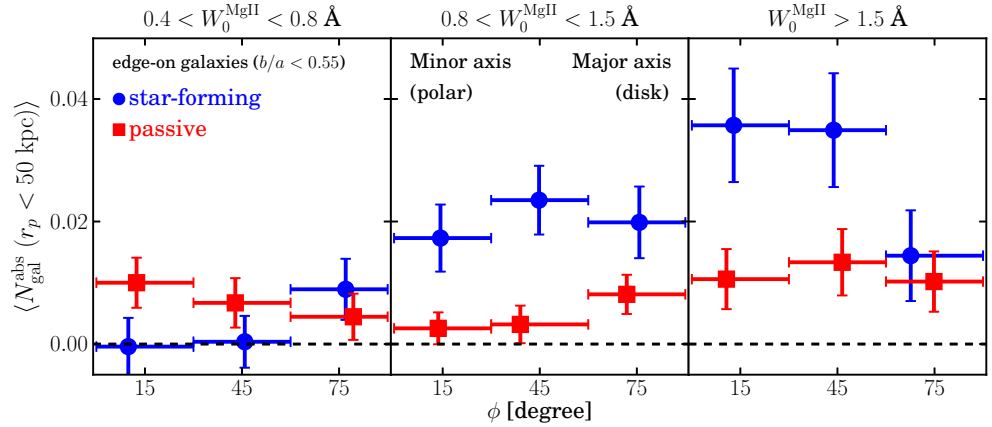

Figure 3.6: The azimuthal angle dependence of Mg II absorbers and edge-on galaxies at $10 < r_p < 50\,$kpc. We select edge-on galaxies with axis ratio $b/a < 0.55$ and separate star-forming ($0.5 < g - i < 2.0$) and passive ($2.0 < g - i < 3.0$) galaxies. A trend shows that strong absorbers preferentially exist near minor axis of star-forming galaxies. The blending correction is not applied in this analysis. The horizontal bars show the range of azimuthal angle in each bin and the data points indicate the median value within the bin.

Table 3.1: The best fit parameters of Mg II covering fraction at $r_p > 50$ kpc

	star-forming		passive galaxies	
	A(amplitude)	α (slope)	A(amplitude)	α (slope)
$0.4 < W_0^{\rm MgII} < 0.8\,$Å	0.06 ± 0.02		0.06 ± 0.01	
$0.8 < W_0^{\rm MgII} < 1.5\,$Å	0.05 ± 0.01	-1.3 ± 0.5	0.04 ± 0.01	-1.4 ± 0.3
$W_0^{\rm MgII} > 1.5\,$Å	0.02 ± 0.01		0.01 ± 0.004	
$W_0^{\rm MgII} > 1.0\,$Å	0.05 ± 0.01	-1.2 ± 0.4	0.03 ± 0.01	-1.1 ± 0.2

around galaxies, i.e. the absorption covering fraction f_c. This quantity is defined as

$$f_c \equiv N_{\rm abs}^{\rm [gal]}/N_{\rm QSO}^{\rm [gal]} \ , \tag{3.10}$$

where $N_{\rm QSO}^{\rm [gal]}$ is the total number of quasar sightlines around a given galaxy sample and $N_{\rm abs}^{\rm [gal]}$ is the total number of Mg II absorbers detected from those quasar sightlines. This quantity can also be estimated by counting galaxies around absorbers and quasars:

$$f_c = N_{\rm gal}^{\rm [abs]}/N_{\rm gal}^{\rm [QSO]} \ , \tag{3.11}$$

where $N_{\rm gal}^{\rm [abs]}$ (as estimated in Eq. 3.12) is the total number of galaxies associated with Mg II absorbers and $N_{\rm gal}^{\rm [QSO]}$ is the total number of galaxies expected in that redshift interval. We will use this approach to estimate the gas covering fraction. To do so we first calculate $N_{\rm gal}^{\rm [abs]}$, the expected total number of galaxies associated with a set of $N_{\rm abs}$ absorbers, spanning a redshift interval Δz ($0.4 < z < 0.6$). It can be written as

$$N_{\rm gal}^{\rm [abs]} = \langle N_{\rm gal}^{\rm abs} \rangle \times N_{\rm abs}^{\rm expected} \ , \tag{3.12}$$

where $\langle N_{\rm gal}^{\rm abs} \rangle$ is the mean number of galaxies per absorber defined in Equation 3.2 and $N_{\rm abs}^{\rm expected}$ is the number of expected absorbers over the selected redshift range which is given by

$$N_{\rm abs}^{\rm expected} = dN/dz(W_0^{\rm MgII}) \times \Delta z \times N_{\rm QSO} \ , \tag{3.13}$$

with $dN/dz(W_0^{\rm MgII})$ being the Mg II incidence rate with a given $W_0^{\rm MgII}$ from Zhu & Ménard (2013), Δz being the redshift interval from 0.4 to 0.6, and $N_{\rm QSO}$ being

the total number of quasars. In this analysis we only use quasars with $z > 1$ to avoid possible contributions from galaxies physically clustered with the quasars. To estimate the denominator of Eq. 3.11, i.e., the expected number of galaxies falling in the selected redshift range Δz, we use galaxy counts around the selected quasars which allow us to estimate the total number of galaxies $N^{[\mathrm{QSO}]}_{\mathrm{gal(all\ z)}}$ originating from all redshifts. We then quantify the number of galaxies expected within Δz by

$$N^{[\mathrm{QSO}]}_{\mathrm{gal}} = N^{[\mathrm{QSO}]}_{\mathrm{gal(all\ z)}} \times \eta(\Delta z) \tag{3.14}$$

where $\eta(\Delta z)$ is the fraction of galaxies with $0.4 < z < 0.6$ with respect to galaxies at all redshifts. To estimate $\eta(\Delta z)$, we select PRIMUS galaxies with robust redshift measurements and make use of the WEIGHT parameter in the catalog which accounts for the incompleteness due to target selection, fiber collision, and redshift quality (see Equation 1 in Moustakas et al. 2013). To alleviate the incompleteness of SDSS at faint end, we restrict this part of analysis to the sample of bright galaxies, i.e. selected with $i < 20.6$. The estimated $\eta(\Delta z)$ are 0.18 and 0.49 for bright star-forming and passive galaxies, respectively.

Figure 3.7 shows, for three differential bins of W_0^{MgII}, the Mg II covering fraction as a function of impact parameter for both star-forming and passive galaxies with $i < 20.6$. The average stellar mass of galaxies corresponding to the magnitude cut are $\langle \log M_*/\mathrm{M}_\odot \rangle \sim 10.6$ (star-forming) and ~ 10.9 (passive). We find that the covering fraction of strong absorbers around star-forming galaxies is more than 2 times higher than around passive galaxies at impact parameters smaller than about

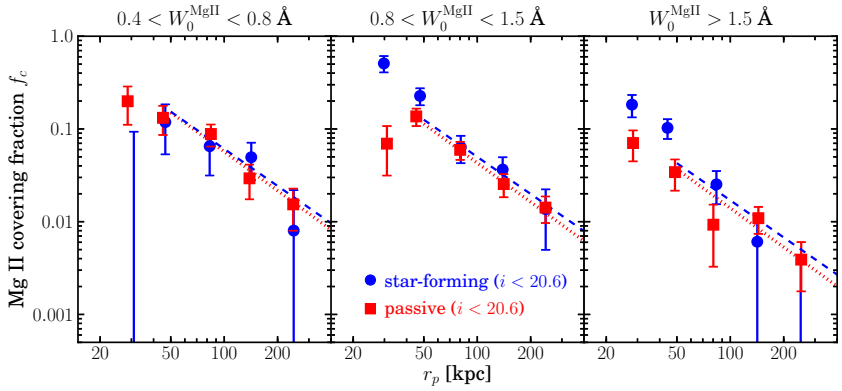

Figure 3.7: The Mg II covering fraction of bright galaxies ($i < 20.6$) as a function of $W_0^{\rm Mg\,II}$ from $20 < r_{\rm p} < 300\,$kpc. For intermediate and strong absorbers, bright star-forming galaxies ($0.5 < g - i < 2.0$) have more than 2 times higher covering fraction than bright passive galaxies ($2.0 < g - i < 3.0$) within 50 kpc, while Mg II covering fraction beyond 50 kpc around both star-forming and passive galaxies decreases with a dependence of $f_c \propto r_p^{\alpha}$ with best-fit power laws shown in blue dashed lines and red dotted lines.

50 kpc, consistent with what was found in Section 3.4.1.1. On the other hand, beyond 50 kpc, there is no significant difference between the Mg II covering fraction of two types of galaxies and the covering fraction can be described by a power-law functional form

$$f_c = \mathrm{A} \left(\frac{r_p}{100\,\mathrm{kpc}} \right)^{\alpha} \quad \text{for } r_p > 50\,\mathrm{kpc}. \tag{3.15}$$

At impact parameters greater than 50 kpc, we find similar spatial dependences for different bins of Mg II absorption strength. To characterize them we fit the covering fraction with a single slope for each type of galaxies and report the best-fit parameters shown in Table 3.1. On large scales, the Mg II covering fraction appears to be consistent with the scale dependence of galaxy-galaxy cross-correlation function,

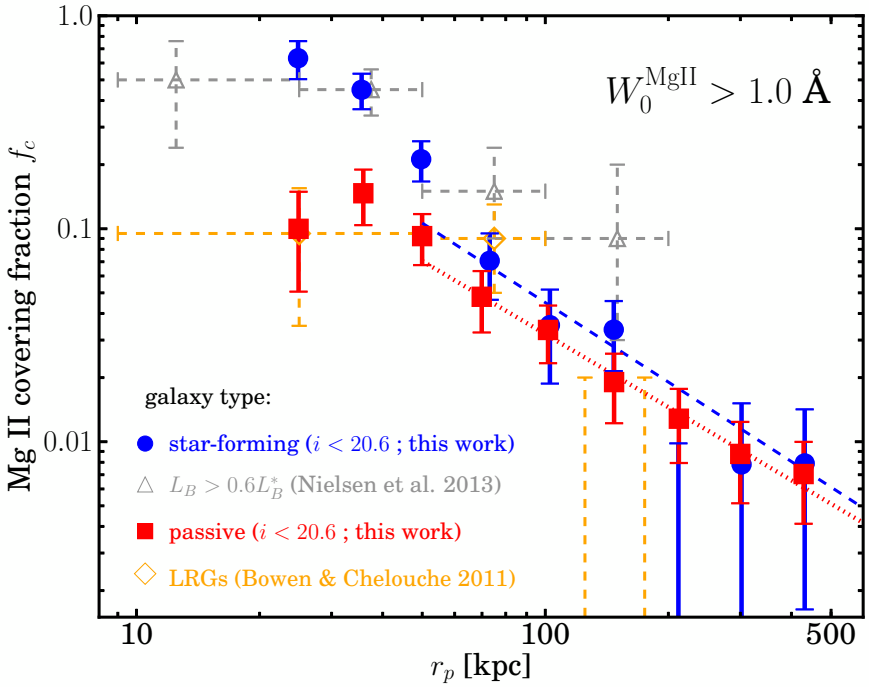

Figure 3.8: The Mg II covering fraction around galaxies compared to the previous studies. The grey triangles are the covering fraction of $L_B > 0.6L_B^*$ galaxy sample from Nielsen et al. (2013b). The covering fraction of star-forming galaxies is consistent with their measurement (although they did not observe any color dependence of covering fraction). The orange diamonds are the covering fraction of luminous red galaxies (LRGs) from Bowen & Chelouche (2011). Considering similar colors, our bright passive galaxies have consistent covering fraction as Bowen & Chelouche (2011). The blue dashed line and red dotted line are the best-fit power laws beyond 50 kpc.

suggesting that Mg II absorbers around galaxies at large scales may trace the overall mass distribution. Note that by extrapolating the power-law functions down to 20 kpc, the functions can still reproduce the covering fraction around passive galaxies while they will underestimate the covering fraction of strong absorbers around star-forming galaxies.

Figure 3.8 shows comparisons between our measurements and previous studies for

$W_0^{\mathrm{Mg\,II}} > 1.0\,\text{Å}$. In Bowen & Chelouche (2011), the authors used photo-z of LRGs to find galaxy-Mg II absorber pairs and estimated the covering fraction of Mg II absorbers around LRGs. With similar color of their LRGs, our passive galaxies have consistent Mg II covering fraction. We also compare our results to the covering fraction of high-luminosity galaxies in Nielsen et al. (2013b). Our bright star-forming galaxies have a consistent covering fraction with their measurement (although they did not observe any dependence on galaxy color). The best-fit parameters of the corresponding dependence are given in Table 3.1.

3.4.2 Correlations with WISE and GALEX

We now complement our analysis with photometric information from the WISE and GALEX datasets. The angular resolution of WISE [3.4 μm] band and GALEX NUV band is about 6″, which is 4 times that of SDSS. This restricts our number count analysis to impact parameters greater than about 50 kpc. On smaller scales, galaxies falling within the angular resolution of quasars will affect its apparent brightness by providing an extra flux contribution and/or by dimming the quasar light due to dust extinction.

3.4.2.1 Large scale correlations

We first match sources detected in GALEX and/or WISE with galaxies detected in SDSS within a radius of 6″. We find that the vast majority of WISE and/or GALEX

galaxies associated with Mg II absorbers are also detected in SDSS. The left panel of Figure 3.9 shows the $g-i$ color distribution of galaxies found in excess beyond 50 kpc of Mg II absorbers with $W_0^{\rm Mg\,II} > 0.4\,$Å. The green, orange, and purple curves show sources detected in SDSS, WISE+SDSS and GALEX+SDSS, respectively.

All the sources detected in GALEX appear to be blue in $g-i$. They are on average bluer than the blue population of galaxies selected from SDSS only. This reflects that GALEX traces UV-bright star-forming galaxies at those redshifts. We find the average NUV magnitude to be about 21.5. According to Kennicutt (1998), this corresponds to an average SFR $\sim 10\ {\rm M}_\odot\,{\rm yr}^{-1}$.

The $g-i$ color distribution of sources selected in both SDSS and WISE shows a relatively higher sensitivity to red galaxies. To better understand the relation between optical and infrared colors we select star-forming and passive galaxies from their $g-i$ colors (using the same selection cuts asin Equation 3.5) and investigate their infrared color distributions. To do so we have estimated the WISE color W1 [$3.4\,\mu$m]-W3 [$12\,\mu$m]. The result is shown in the right panel of Figure 3.9. The orange dashed line shows all WISE+SDSS galaxies distribution, blue data points represent optical-selected star-forming galaxies, and red data points represent optical-selected passive galaxies. This figure shows that, at $z \sim 0.5$ optically star-forming galaxies have red W1-W3 colors. This is because star-forming galaxies produce strong $8\,\mu$m polycyclic aromatic hydrocarbon (PAH) emission lines visible in the W3 band (e.g. Smith et al., 2007) at those redshifts. This analysis shows that the types of galaxy associated with

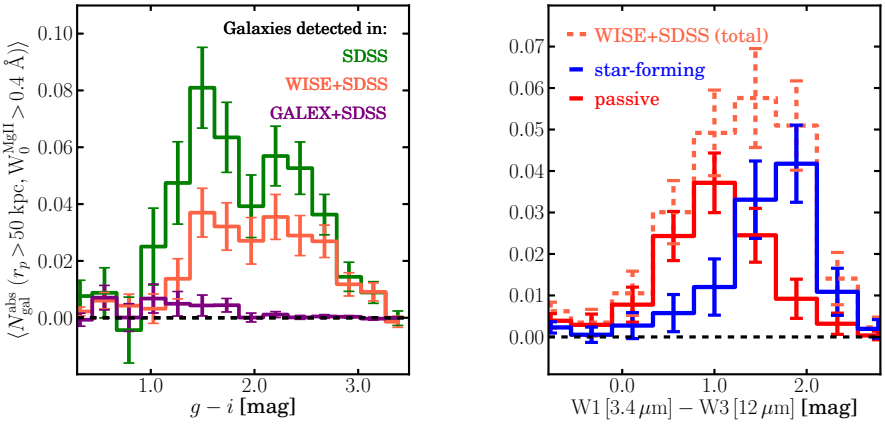

Figure 3.9: *Left:* The g-i color distributions of associated galaxies detected in SDSS, WISE and GALEX. SDSS, WISE and GALEX select different types of galaxies because of their wavelength coverages and sensitivities. With the highest sensitivity in W1 [3.4 μm] band, WISE tends to detect massive galaxies with relatively red galaxy color (in g-i). In contrast, GALEX selects young star forming galaxies bright in NUV with relatively blue galaxy color (in g-i). *Right:* The W1-W3 color distribution of galaxies associated with Mg II absorbers. Orange dashed line: the associated galaxies detected in WISE and SDSS with $W_0^{\rm Mg\,II} > 0.4$ Å from 50 kpc to 200 kpc. Blue: optical selected blue galaxies ($0.5 < g - i < 2.0$). Red: optical selected red galaxies ($2.0 < g - i < 3.0$). In general, optical-selected star-forming galaxies have more flux in W3 than W1 compared to passive galaxies due to strong 8 μm PAH emission.

absorbers found with the SDSS survey are consistent with those found at infrared wavelengths with the WISE survey.

3.4.2.2 Small scale correlations

On scales smaller than about 50 kpc, galaxies associated with absorbers fall within the angular resolution of the WISE quasar and affect its apparent brightness. In this analysis, we match our SDSS quasar sample with WISE full-sky catalog with a 6″ radius. We use WISE standard aperture magnitudes (which has a size of 8″) which

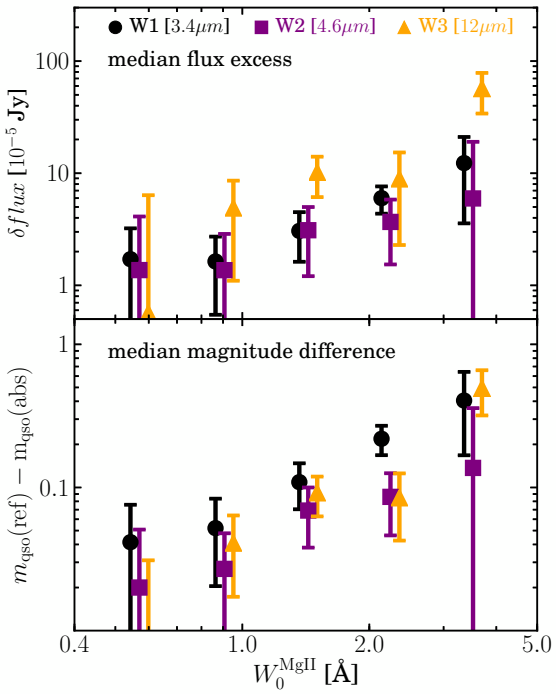

Figure 3.10: Quasar brightening effect due to the presence of a galaxy associated with a Mg II absorber. *Top:* median flux excess associated with absorbers as a function of $W_0^{\rm Mg\,II}$. The spectral energy distribution of the excess is consistent with that of a star forming galaxy with strong 8 μm PAH emission. *Bottom:* median quasar magnitude shift due to the presence of an absorber. The amplitude of this effect can be as large as 0.5 magnitude for the strongest absorbers.

contain all flux including quasars and galaxies on top of them. We limit our sample with absorber redshift from 0.4 to 0.6, calculate the median WISE flux of quasars with Mg II absorbers as a function of Mg II absorption strength, and compare them to the median WISE flux of reference quasars. The error is estimated by bootstrapping. Note that quasars with Mg II absorbers and their reference quasars have similar redshift and i-band magnitude distributions.

The top panel of Figure 3.10 shows the flux differences in WISE 1, 2, and 3 bands

as a function of $W_0^{\rm Mg\,II}$. We find that quasars with Mg II absorbers are systematically brighter than reference quasars. The flux difference increases for stronger absorbers. This is in part because stronger absorbers tend to be closer to star-forming galaxies. We note that the excess flux in W3 [12 μm] tends to be higher than that of W1 [3.4 μm]. This is consistent with the 8 μm PAH contribution discussed above.

The bottom panel of the figure shows how the IR flux contribution from the absorbing galaxies affects quasar magnitudes. The median magnitude shift induced by the presence of a Mg II absorber, $\delta m = {\rm m_{QSO}(ref)} - {\rm m_{QSO}(abs)}$, is shown as a function of $W_0^{\rm Mg\,II}$. On average, the median quasar brightness changes by about 0.1 and reaches 0.5 magnitude for the strongest absorbers. This result indicates that selecting quasars with WISE magnitudes and colors could potentially result in a selection bias for quasars with the presence of Mg II absorbers. Future surveys, for example SDSS-IV, that use WISE magnitudes and colors to select quasars should take into account such bias. In the UV, the presence of an intervening absorber gives rise to extinction effects. This was recently studied by Ménard & Fukugita (2012).

3.5 Discussion

We have shown that cool gas traced by Mg II absorption is found around both star-forming galaxies and passive galaxies, with a similar incidence rate at impact parameters greater than about 50 kpc. In contrast, at smaller impact parameters

`http://www.sdss3.org/future/`

we find that the strength of Mg II absorption depends on the level of star formation of the central galaxy. Our results are consistent with other types of observational constraints on the absorber-galaxy connection: Zibetti et al. (2007) used stacked SDSS images to measure the mean flux excess correlated with the presence of Mg II absorbers and found stronger Mg II absorbers to be preferentially associated with bluer emission. Bordoloi et al. (2011) used stacked galaxy spectra to measure the mean Mg II absorption induced by the presence of galaxies along the line-of-sight and showed that the mean Mg II absorption strength around star-forming galaxies is higher than around passive galaxies. Our analysis has, in addition, allowed us to show that while Mg II absorption is commonly found around passive galaxies, no correlation between equivalent width and galaxy properties can be detected. This is in contrast to the relation observed between star-forming galaxies and absorbers where the mean equivalent width depends on the color and/or star formation rate of the galaxies. The basic dichotomy between star-forming and passive galaxies is therefore reflected in the properties of the cool gas in the CGM. We also find that strong absorbers tend to be found along the minor axis of star-forming galaxies at small impact parameters, indicating that some of the gas detected in absorption is likely associated with outflows from galaxies. This is consistent with what was reported by Bordoloi et al. (2011), Bouché et al. (2012) and Kacprzak et al. (2012), and is similar to the measurements of Ca II absorption by Zhu & Ménard (2013). Our results also show that at impact parameters greater than 50 kpc, the covering

fraction of strong Mg II absorbers declines with scale proportionally to the overall mass distribution for both star-forming and passive galaxies. Here we note that Zhu et al. (2014) found a similar result by measuring the *total* Mg II absorption around SDSS luminous red galaxies. By measuring the mean absorption correlated with the presence of such galaxies, these authors showed that the average Mg II-to-dark matter ratio is roughly scale-independent. Strong Mg II absorbers therefore trace the overall Mg II absorption field on large scales.

3.5.1 The connection to baryons in galaxy halos

It is interesting to point out the contrast between the ubiquitous presence of cool, low-ionized gas traced by Mg II around both star-forming and passive galaxies and the distribution of highly ionized gas traced by O VI absorption which currently is only detected in the halo of star-forming galaxies (Tumlinson et al., 2011). However, despite such a difference, the two gas distributions share a common property: excess absorption is detected when the galaxy specific star formation rate $sSFR \gtrsim 10^{-11}\,\mathrm{yr}^{-1}$. This threshold is found to be the same for both phases. It appears to be a characteristic sSFR value for determining the gaseous properties of the CGM.

Similarly, we detect excess absorption when the galaxy star formation rate SFR $\gtrsim$ $1\,\mathrm{M}_{\odot}/\mathrm{yr}$ and note that this threshold is similar to the one for which blueshifted Mg II self-absorption is seen in galaxies (e.g. Bordoloi et al., 2013; Rubin et al., 2013; Weiner et al., 2009). The interpretation of such studies is usually limited by the lack of

information on the spatial scales over which the gas flow is occurring. The similarity of our results suggest that the two lines of investigation are probing the same material. If so, the gas seen as blueshifted self-absorption could extend up to scales greater than a few kpc, as typically assumed by these authors when inferring mass outflow rates from such measurements.

Having characterized the covering fraction of Mg II absorbers as a function of impact parameter, we can attempt to estimate the typical amount of cool HI gas traced by Mg II absorbers residing in the CGM of $z \sim 0.5$ galaxies. To do so we can use the empirical relation between Mg II equivalent width and median HI column density measured by Rao et al. (2006) and quantified by Ménard & Chelouche (2009): $\langle N_{\rm HI} \rangle (W_0^{\rm Mg\,II})$ (their Equation 5). Using this relation we can write

$$M_{\rm HI}^{\rm CGM}(> W_0^{\rm Mg\,II}) \;\sim\; 2\pi \int_{20\,{\rm kpc}}^{150\,{\rm kpc}} \langle N_{\rm HI} \rangle f_c(r_p)\, r_p\, {\rm d}r_p \tag{3.16}$$

where $f_c(r_p)$ is the Mg II covering fraction and r_p is the projected distance. The inner impact parameter limit (20 kpc) is selected based on the range for which we have robust covering fraction measurements, while the outer impact parameter limit (150 kpc) corresponds to the maximum impact parameter of galaxy-absorber pairs for which COS-Halos team searched. To estimate this quantity numerically we consider the cool HI gas traced by $W_0^{\rm MgII} > 1\,$Å ($\langle W_0^{\rm Mg\,II} \rangle \simeq 1.6\,$Å), corresponding to the covering fraction shown in Figure 3.8. For star-forming galaxies with $\langle \log M_*/M_\odot \rangle \sim$

10.6 we find

$$\log M_{\rm HI}^{\rm CGM}/{\rm M}_{\odot} \sim 9.5. \quad (3.17)$$

For passive galaxies with $\langle \log M_*/M_{\odot} \rangle \sim 10.9$ we find

$$\log M_{\rm HI}^{\rm CGM}/{\rm M}_{\odot} \sim 9.2. \quad (3.18)$$

The above numbers correspond only to HI gas traced by Mg II absorbers stronger than 1 Å. The estimate can be considered as a lower limit for the total amount of HI found in galaxy halos. These values imply that the ratio $M_{\rm HI}^{\rm CGM}/M_*$ is about four times lower around passive galaxies than around star-forming ones.

Recently, Werk et al. (2014) used HST-COS observations to probe the gaseous distribution around $z \sim 0.2$ $L \sim L_*$ galaxies. Using photo-ionization models they inferred that such galaxies are surrounded by cool ($T \sim 10^4$ K) gas amounting to at least $\log M_{\rm H}/M_{\odot} = 10.4$ within about 150 kpc. Most (99%) of this gas is found to be ionized, implying that the neutral component probed by the lines-of-sight of the COS-Halos survey (excluding a few damped systems) amounts to $\langle \log M_{\rm HI}/M_{\odot} \rangle \sim 8.4$. This is a factor 5-10 lower than the amount of HI probed by Mg II absorbers presented above. We note that given the low covering fraction for strong Mg II absorbers on such scales: $f_c(100\,{\rm kpc}) \sim 0.05$, most of randomly-selected lines-of-sight are not expected to intercept such clouds. Interestingly, the above numbers indicate that most of the neutral hydrogen probed by metal absorbers is located in strong Mg II absorbers, despite their low covering fraction. An additional neutral gas contribution can be

associated with pristine or low metallicity gas. Such clouds would not give rise to strong Mg II absorbers.

We also note that the spatial dependence of the covering fraction of strong Mg II absorbers (derived in Section 3.4.1.4) can be converted into a minimum value for the *mean* absorption equivalent width as a function of impact parameter. By selecting absorbers with $W_0^{\rm Mg\,II} > 1.0\,$Å with 1.6 Å mean absorption and focusing on red galaxies, we find that

$$\langle W_0^{\rm Mg\,II}\rangle(r_p) \sim 0.05 \left(\frac{r_p}{100\,{\rm kpc}}\right)^{-1.1} \text{Å}\ . \tag{3.19}$$

This dependence can be compared to the results obtained by Zhu et al. (2014). These authors measured the mean absorption around galaxies averaged over all lines-of-sight. By comparing the two sets of results, we find that absorbers with $W_0^{\rm Mg\,II} > 1\,$Å contribute to about half of the total absorption signal. By selecting systems with $W_0^{\rm Mg\,II} > 0.4\,$Å we find this contribution to be comparable to the mean absorption level. This indicates that, on large scales around galaxies, strong Mg II absorbers dominate the total Mg II absorption budget.

Our results also allow us to reveal that the covering fraction for cool gas around galaxies appears to change as a function of redshift. At $z \sim 0.5$ we found $f_c \propto r_p^{-1}$ for Mg II absorbers at impact parameters greater than about 50 kpc around both star-forming and passive galaxies. As mentioned above, this relation steepens on smaller scales around star-forming galaxies. Using a statistical analysis, Steidel et al. (2010) inferred the radial dependence of the covering fraction of cool gas around star-

forming (Lyman break) galaxies at $2 \lesssim z \lesssim 3$. They derived a significantly shallower radial dependence: $f_c \propto r_p^{-\gamma}$ with $0.2 \lesssim \gamma \lesssim 0.6$, depending on the transition. The comparison of the two analyses therefore shows that the radial dependence of metals in cool gas appears to steepen from $z \simeq 2-3$ down to $z \sim 0.5$.

Finally, if we assume that the neutral gas traced by Mg II absorbers is not only enriched in metals in the gas phase but also dusty, it implies a CGM dust mass of $M_{\rm dust}^{\rm CGM} \sim 4 \times 10^7\,{\rm M}_\odot$ using the global dust-to-gas ratio in Mg II clouds ($\sim 1/50$; Ménard & Fukugita, 2012). This value is consistent with the findings of Ménard et al. (2010) and Peek et al. (2014) who statistically mapped out the distribution of dust in galaxy halos using reddening measurements and inferred its total mass. It also implies that most of the circum-galactic dust is associated with Mg II absorbers.

3.6 Summary

We cross-correlate about 50,000 Mg II absorbers with photometric sources from SDSS, WISE, and GALEX to study the properties of cool (T$\sim 10^4\,$K) gas in the circum-galactic medium. Using the SDSS survey and focusing on the redshift range $0.4 < z < 0.6$ we statistically extracted about 2,000 galaxy-absorber pairs which have allowed us to explore the relationship between absorber and galaxy properties with an unprecedented sensitivity. Our results are summarized as follows:

- Mg II absorbers are associated with both star-forming and passive galaxies,

with a comparable incidence rate at impact parameters greater than 50 kpc. However each galaxy type exhibits a different behavior: within 50 kpc Mg II equivalent width correlates with the presence of star-forming galaxies but not with that of passive galaxies.

- The correlation between the presence of cool gas traced by strong Mg II absorbers and edge-on star-forming galaxies is stronger along the minor axis of galaxies at impact paramters reaching about 50 kpc, suggesting that some of the gas is associated with outflows, consistent with previous studies. In contrast, we find Mg II absorbers to be isotropically distributed around edge-on passive galaxies.

- We measure the average excess Mg II equivalent width $\langle \delta W_0^{\rm Mg\,II} \rangle$ as a function of galaxy properties and find $\langle \delta W_0^{\rm Mg\,II} \rangle \propto SFR^{1.2}$, $sSFR^{0.5}$ and $M_*^{0.4}$ for star-forming galaxies. These observational results can be used to constrain models of galaxy formation and feedback processes.

- We characterize the covering fraction of Mg II absorption as a function of impact parameter and find that it is about 2-10 times higher for star-forming galaxies than passive ones on scales smaller than about 50 kpc. The covering fractions appear to be comparable on scales larger than 50 kpc and follow the spatial dependence of the galaxy correlation function.

- Using GALEX and WISE we show that the UV and IR properties of the galaxies

correlated with absorbers are consistent with the dependence found with optical data. On scales smaller than 50 kpc, we show how the presence of $z \sim 0.5$ Mg II absorbers modifies the brightness and colors of quasars, consistent with dust reddening at short wavelengths and excess emission at IR wavelengths. We also show how the selection of quasars in the WISE magnitude and color space can be biased by the presence of $z \sim 0.5$ galaxies detectable through their optical metal absorption. Future surveys using WISE information to select quasars need to take into account this bias.

This work shows that the dichotomy between star-forming and passive galaxies is reflected in the CGM traced by low-ionized gas. This trend is similar to what has been reported for the distribution of higher-ionization species such as O VI (Tumlinson et al., 2011).

This statistical analysis demonstrates that absorber-galaxy correlations can be measured without prior knowledge on galaxy redshifts. This has allowed us to study the distribution of cool gas around galaxies and its dependence on galaxy properties. Our results present a set of constraints for theoretical models of galaxy growth and feedback. The analysis can be pushed towards higher redshifts as deeper imaging over large areas becomes available, such as the ongoing Hyper-Suprime Cam Survey or the planned DECam Legacy Survey.

`http://www.naoj.org/Projects/HSC`

Chapter 4

Probing mysterious unknown molecules in the Milky Way

4.1 Abstract

We use star, galaxy and quasar spectra taken by the Sloan Digital Sky Survey to map out the distribution of diffuse interstellar bands (DIBs) induced by the Milky Way. After carefully removing the intrinsic spectral energy distribution of each source, we show that by stacking thousands of spectra, it is possible to measure statistical flux fluctuations at the 10^{-3} level, detect more than 20 DIBs and measure their strength

This work is published in the Monthly Notices of the Royal Astronomical Society, Volume 452, Issue 4, p.3629-3649

as a function of position on the sky. We create a map of DIB absorption covering about 5,000 square degrees and measure correlations with various tracers of the interstellar medium: atomic and molecular hydrogen, dust and polycyclic aromatic hydrocarbons (PAHs). After recovering known correlations, we show that each DIB has a different dependence on atomic and molecular hydrogen: while they are all positively correlated with $N_{\rm HI}$, they exhibit a range of behaviors with $N_{\rm H_2}$ showing positive, negative or no correlation. We show that a simple parametrization involving only $N_{\rm HI}$ and $N_{\rm H_2}$ applied to all the DIBs is sufficient to reproduce a large collection of observational results reported in the literature: it allows us to naturally describe the relations between DIB strength and dust reddening (including the so-called skin effect), the related scatter, DIB pair-wise correlations & families, the affinity for σ/ζ-type environments and other correlations related to molecules. Our approach allows us to characterize DIB dependencies in a simple manner and provides us with a metric to characterize the similarity between different DIBs.

4.2 Introduction

The diffuse interstellar bands (DIBs) are a set of absorption features observed ubiquitously in the interstellar medium (ISM). The first features, reported by Heger (1922), were identified as interstellar in origin by Merrill (1934). In the past eighty years, the list of DIBs has increased to more than 500 features (Hobbs et al., 2008,

2009). However, despite active research, the identity of the DIB carriers remains unknown until this day. This has been one of the longest standing problems in astronomy (for a review, see Herbig, 1995). Candidates of carriers include complex carbonaceous gas-phase molecules, such as fullerenes (Foing & Ehrenfreund, 1994) and polycyclic aromatic hydrocarbons (PAHs, e.g., Salama et al., 1996). However, none of them has yet been convincingly shown to be associated with any particular DIB.

DIBs are mostly found in the optical and near-infrared, with the longest reported wavelength at $1.793\,\mu$m (Geballe et al., 2011). They display a large range in width and central depth. The narrowest lines have full width at half maximum (FWHM) less than 1 Å while the broad DIBs have FWHM reaching 30 Å. The central depth of the detected lines ranges from less than 0.1% to about 50%. To illustrate this diversity, we show a synthetic DIB absorption spectrum in Figure 4.1, using a list of detected DIBs from Jenniskens & Desert (1994) who studied high-S/N spectra of four early-type stars. This synthetic spectrum illustrates that the DIB absorption is weak: the strongest feature at $\lambda = 4430$ Å has an equivalent width of only about 0.2 Å for a dust column density corresponding to $E_{B-V} \simeq 0.1$ mag.

For several decades, the main probe of DIB absorption has been high-S/N spectra of hot stars, in which weak absorption features can be measured and blending with stellar lines is minimized. Dedicated surveys have included a few thousand stars at most (e.g., van Loon et al., 2013) or only around a hundred stars if weaker DIBs are

targeted (e.g., Friedman et al., 2011). In recent years, the availability of generic large sky surveys has allowed the study of DIBs in a statistical context. In addition to increasing the number of available lines of sight, they have also motivated the use of other strategies to detect DIBs in spectra of a wider range in spectral types. Using the Sloan Digital Sky Survey (SDSS; York et al., 2000), Yuan & Liu (2012) reported the detection of two DIBs, λ5780 & λ6283, in about $2,000$ stellar spectra (of all types) and characterized their strengths and radial velocities. Using the Radial Velocity Experiment (RAVE, Steinmetz et al., 2006), Kos et al. (2013a, 2014) detected DIB λ8620 below the noise level of individual spectra. They characterized its absorption in the composite spectra of several thousand cool stars and mapped the absorption in the sky. Using near-infrared spectra from the SDSS Apache Point Observatory Galactic Evolution Experiment (APOGEE; Majewski et al., in prep), Zasowski et al. (2015) studied the DIB absorption at $\lambda = 1.527\,\mu$m in about 100,000 stellar spectra probing a wide range of Galactic environments and mapped out its properties.

In this chapter we use all types of spectra taken by the Sloan Digital Sky Survey I, II, and III, i.e. lines of sight towards stars, galaxies and quasars, to map out the distribution of diffuse interstellar bands induced by the Milky Way. We show that, after carefully removing the intrinsic spectral energy distribution (SED) of each source, it is possible to measure statistical flux fluctuations at the 10^{-3} level. We then detect and characterize a set of 20 DIBs and study their correlations with various ISM

While performing this analysis we became aware of a similar effort by Baron et al. (2015). We decided to finish the two analyses independently and submit the two works to arxiv.org simultaneously.

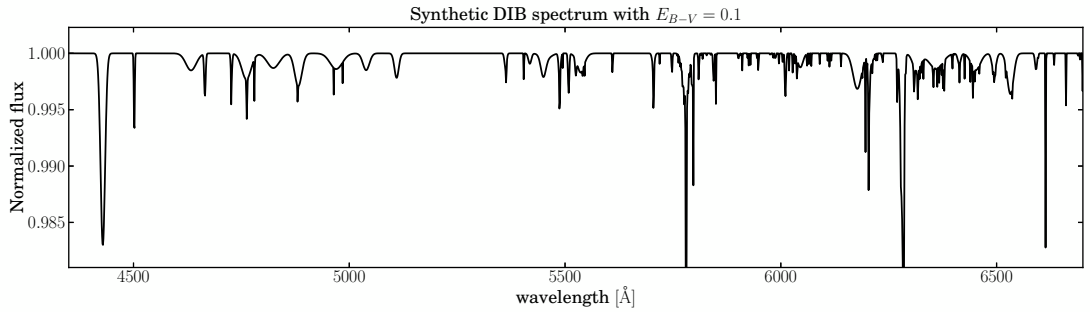

Figure 4.1: A synthetic spectrum of diffuse interstellar band absorption (from the compilation by Jenniskens & Desert 1994) representative of a line of sight with $E_{B-V} \sim$ 0.1 mag. Note the expected absorption is at the 1% level.

tracers. Throughout the chapter, we use air wavelength.

4.3 Data analysis

Our analysis makes use of optical spectra from the SDSS I, II & III surveys with spectral resolution about 2000. We explore the detectability of DIB absorption in three types of sources: quasars, galaxies and stars. In each case we create absorption spectra, normalizing the observed source spectra by an estimate of the SED intrinsic to the source. In addition, we take extra care to handle artificial residuals originating from imperfections in the calibration process of SDSS spectra. For each type of sources, we use a different strategy to estimate the corresponding absorption spectra. To avoid contamination from atmospheric emission lines, mostly due to OH and H_2O, we restrict our analysis to wavelengths below 6700 Å. We now describe the analysis procedure applied to each type of sources. The next section will then present the

characterization of absorption lines detected in the absorption spectra.

4.3.1 Creating absorption spectra

4.3.1.1 Quasars

We use the SDSS DR7 quasar catalog compiled by Schneider et al. (2010). The corresponding spectra were analyzed by Zhu & Ménard (2013) who estimated their intrinsic continuum level. They did this using a dimensionality reduction technique (non-negative matrix factorization) to obtain a basis set of "eigen spectra" which can then be used to estimate each quasar's SED. We create absorption spectra by dividing each observed spectrum by its estimated intrinsic SED. We study corresponding spectra in the observer frame.

We first create median composite absorption spectra for quasars observed in different regimes of Galactic dust reddening, derived from Schlegel et al. (1998, SFD). In the upper part of the top panel of Fig. 4.2, we show the results obtained from objects selected with $E_{B-V} < 0.02$ mag in blue and $E_{B-V} > 0.05$ mag in red. As can be seen, the composite spectra reveal systematic fluctuations at the 5×10^{-3} level. These fluctuations appear to be inconsistent with Poisson noise and their amplitude does not correlate with dust column density. These features are mainly due to systematics in the spectral calibration process in SDSS as well as sky residuals. We discuss these effects in more detail in the Appendix C.1 and identify the origin of each of these

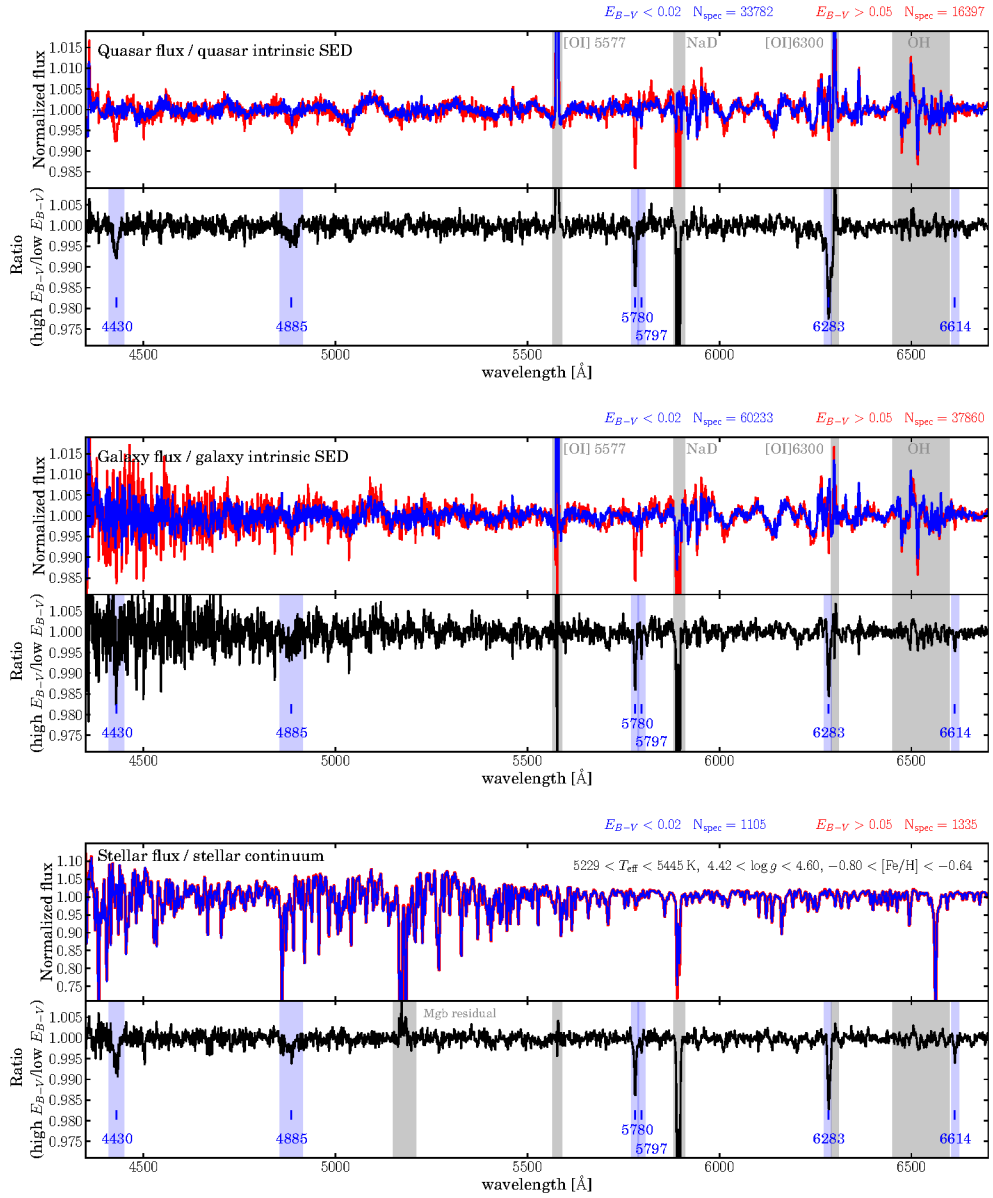

Figure 4.2: Illustration of the steps involved in the creation of composite absorption spectra using quasars (top), galaxies (middle) and stars (bottom). For each type of source the upper part of the panels shows continuum-normalized composite spectra for high and low dust reddening values and the bottom part of the panels shows their corresponding ratios. The intrinsic SED of quasars and galaxies are estimated through rest-frame template fitting. The continuum-normalized spectra show fluctuations imprinted by imperfections in the SDSS spectroscopic calibration. These features are detailed and explained in the Appendix C.1. The intrinsic SED of stars is estimated using a running median filter to characterize the large-scale continuum. Intrinsic absorption features are not modeled. Their contribution between low and high dust reddening regions is expected to be comparable and has therefore close to no effect in the final ratio estimate. Blue bands indicate known DIBs. Gray bands indicate artifacts due to the presence of strong sky emission lines.

features.

In order to overcome the limits given by the precision of the SDSS spectral calibration process we use the fact that DIB absorption is known to correlate with dust reddening (e.g., Friedman et al. 2011; Welty 2014, and Section 4.4.2.1), while spectral calibration problems are not expected to do so. We can then improve the sensitivity of our flux residuals by considering the ratio between absorption spectra at high and low dust reddening regimes. To do so we consider the composite spectrum of quasars selected from low-reddening regions with $E_{B-V} < 0.02$ mag, where DIB absorption is expected to be small, as a reference spectrum. The ratio between the composite spectrum with high Galactic reddening and the reference composite spectrum with low Galactic reddening is shown in the lower part of the top panel of Fig. 4.2. As expected this process removes features associated with the SDSS reduction process and provides us with an absorption spectrum with a scatter of order 1×10^{-3}. In the figure, the blue bands indicate the locations of several known DIBs. The number of quasar spectra used in our analysis is listed in Table 4.1.

4.3.1.2 Galaxies

We select a set of luminous red galaxies from SDSS DR7. Such galaxies have well-defined SEDs. We use estimates of their intrinsic continuum fluxes provided by Zhu et al. (2010). These authors modeled the observed galaxy spectra using single stellar population (SSP) models of Bruzual & Charlot (2003) with the Padova 1994 library

Table 4.1: Number of spectra

Source type	total number	reference sightlines $E_{B-V} < 0.02$ mag	targeted sightlines $E_{B-V} > 0.02$ mag
Quasar	105,782	33,782	72,000
Galaxy	210,726	60,233	150,493
Star	354,231	84,406	269,825

of stellar evolution tracks and the Chabrier (2003) initial mass function (IMF). We select galaxies at redshift greater than 0.2. The number of galaxy spectra used in this analysis is listed in Table 4.1.

Similarly to the procedure used for quasars, we create median composites of absorption spectra for different ranges of Galactic dust reddening. As shown in the upper part of the middle panel of Fig. 4.2 we find fluctuation patterns due to the spectroscopic calibration and sky emission/absorption features consistent with those obtained with quasar spectra. As done above, we use ratios of composite spectra to overcome these limitations, considering lines of sight with $E_{B-V}< 0.02$ mag as reference ones. The final ratio spectrum is shown in the lower part of the panel and allows us to detect a comparable set of DIBs.

`http://pleiadi.pd.astro.it/`

4.3.1.3 Stars

We use the stellar spectra collected by the SDSS SEGUE I & II surveys (Aihara et al., 2011; Yanny et al., 2009). We estimate the intrinsic SED of each star using a data-driven approach. We first remove large-scale fluctuations due to black-body emission as well as the effect of line-of-sight dust extinction using a running median filter of size 200 pixels. To create absorption spectra and normalize out the contribution due to the intrinsic SED of each source, for a given star, we search for a set of reference stars at low dust reddening with $E_{B-V}< 0.02$ mag with similar stellar parameters. To do so we make use of three parameters: effective temperature ($T_{\rm eff}$), surface gravity ($\log g$) and metallicity ([Fe/H]), estimated by the SEGUE SSPP pipeline (Lee et al., 2008; Smolinski et al., 2011) . We construct a 3-D grid spanning the full range of the corresponding values: $4200 < T_{\rm eff} < 8700\,{\rm K}$, $0.7 < \log g < 4.7$, and $-4.3 < {\rm [Fe/H]} < 0.6$, with a resolution set to be four times lower than the dispersion of the stellar parameters for reference stars in each dimension. The corresponding bin sizes are $216\,{\rm K}$ in $T_{\rm eff}$, 0.18 dex in $\log g$, and 0.16 dex in [Fe/H]. These bin sizes are about the size of the systematic error of these three parameters. The SSPP pipeline estimated the stellar parameters based on several methods. We use the parameters derived from the ANNRR method (Re Fiorentin et al., 2007) which is based on continuum-normalized spectra. We also perform our analysis by using the parameters derived from other methods and find consistent results. In addition, the SSPP pipeline also

`https://www.sdss3.org/dr10/spectro/sspp.php`

provides distances of stars. In Figure 4.5, we show the median distances of stars within each sky pixel. The typical distance of stars is about 2-3 kpc.

We match each star to a set of reference stars from the same cell in the grid of stellar parameters. To reduce the effect of outliers or problematic spectra, we discard stars located in the stellar parameter bins with less than 20 corresponding stars with $E_{B-V}<$ 0.02 mag. We also remove stars from SDSS plates with bad qualities . We then create a median composite spectrum for the reference stars. As an illustration, the upper part of the bottom panel of Fig. 4.2 shows the composite spectra of typical stars observed in SDSS at high E_{B-V} (red) and low E_{B-V} (blue) with the same stellar parameters.

To create an absorption spectrum we take the ratio between a stellar spectrum and its corresponding reference composite spectrum in the stellar rest frame. The lower part of the panel shows the ratio spectrum between the high E_{B-V} composite spectrum and low E_{B-V} reference composite spectrum. By applying our method, we are able to remove stellar absorption features effectively and detect strong DIBs, as indicated by the blue vertical bands. We note that this stellar residual spectrum makes use of only 0.5% of all the available stellar spectra, while we have used all quasar and galaxy spectra (with $E_{B-V}>$ 0.05 mag) in the top panels. This shows that most of the statistical power to map out the distribution of DIBs lies in the SDSS stellar spectra. Below we will therefore derive most of our results from stellar

`http://www.sdss3.org/dr8/algorithms/segueii/plate_table.php`

spectra. We will use quasar and galaxy spectra primarily for consistency checks. The total number of stellar spectra used in this analysis is listed in Table 4.1. More than 95% of the stars used are F, G, and K stars with $4500 < T_{\rm eff} < 7000\,{\rm K}$.

4.3.2 Detection & characterization of DIBs

We now proceed to characterize DIBs in the composite residual spectra. Detecting and characterizing absorption lines require an accurate estimate of the continuum level. To do so we use the DIB catalog compiled by Jenniskens & Desert (1994) and create mask aimed at selecting the non-absorbed pixels of our spectra. From their list, we select DIBs with equivalent widths larger than 0.05 Å at E_{B-V}= 1 mag but exclude broad DIBs with FWHM $> 25\,$Å . For those, the continuum estimation is more difficult as the width of the absorption feature is an appreciable fraction of the median filter used for continuum estimation. We then mask out the wavelength regions corresponding to these DIBs and apply a median filter with a radius of 25 pixels to eliminate small scale fluctuations in the residual spectra not accounted for in the previous steps of the analysis. Finally, we define the continuum level around each DIB by selecting wavelength regions uncontaminated by calibration errors or sky emission/absorption features.

As an illustration, Fig. 4.3 shows a composite absorption spectrum for all stars with $E_{B-V}>$ 0.1 mag. The median E_{B-V} of these stars is about 0.15 mag. This absorption spectrum shows a set of 20 DIBs detected at more than $\sim 5\sigma$, and not

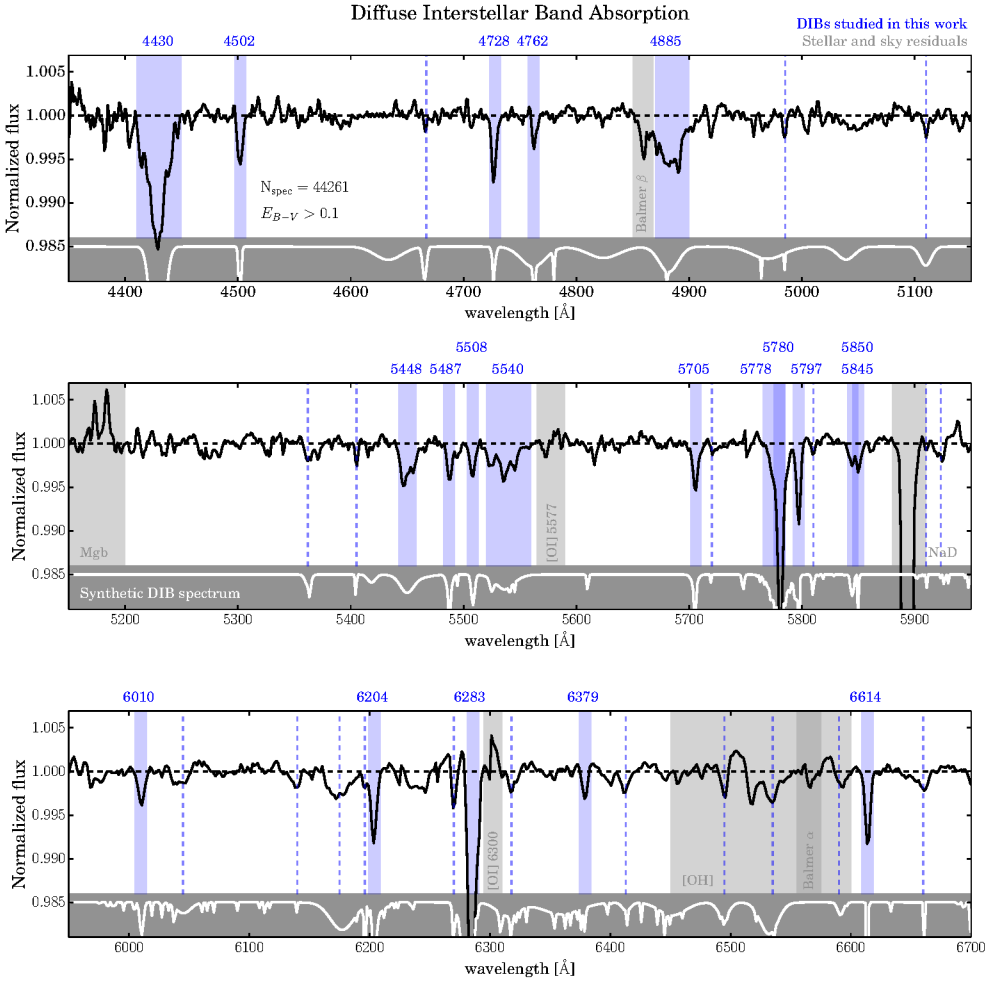

Figure 4.3: An example of final composite absorption spectrum, combining more than 40,000 stellar spectra at $E_{B-V} > 0.1$ mag. The light blue vertical bands indicate the 20 DIBs studied in this work, detected with more than 5σ and for which the line profile can be well characterized. The vertical dashed lines show weaker and/or broader DIBs detected but not used in the statistical analysis. The gray bands show residuals from sky lines and stellar absorption features. The white spectrum at the bottom shows a synthetic DIB absorption spectrum, as shown in Fig. 4.1.

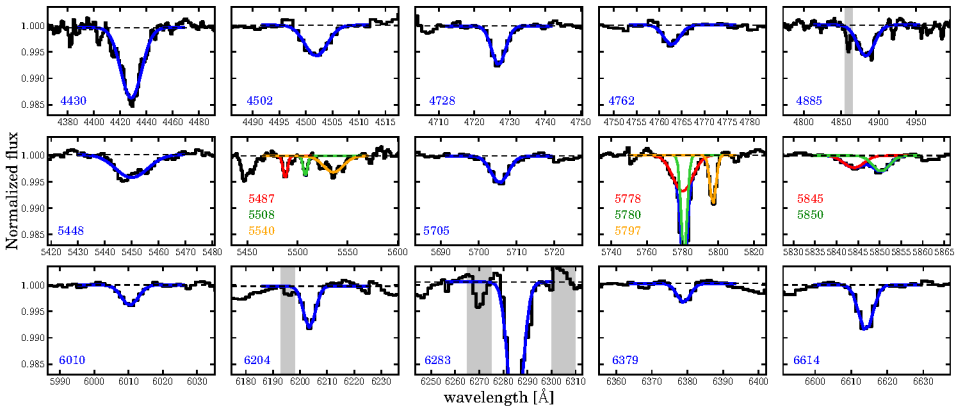

Figure 4.4: The selected set of 20 diffuse interstellar bands measured in a composite spectrum of 40,000 stellar spectra with with E_{B-V}> 0.1 mag, as shown in Figure 4.3. Gaussian fits, used to estimate line parameters, are indicated with blue lines. For wavelength regions with multiple DIBs, we fit them simultaneously with a multiple-Gaussian profile, as shown with red, green, and orange lines. The gray bands indicate regions with contamination from nearby DIBs or sky/stellar residuals, which are masked in the fitting process.

affected by residuals due to the sky lines and/or stellar absorption features. These absorption features are indicated with blue vertical bands and listed in Table 4.3. We can observe additional weak and/or broad absorption features which are consistent with known DIBs. They are indicated by blue dashed lines. Studying these weaker and broader features, however, requires a more detailed analysis to properly estimate the continuum level and the effect of possible artifacts in the flux residuals. We will therefore limit the present analysis to the above set of 20 DIBs with robust detections and characterization.

The composite spectra from quasars and galaxies allow us to robustly detect DIBs λ4430, λ4885, λ5780, λ6283, and λ6614. We note that the SDSS footprint

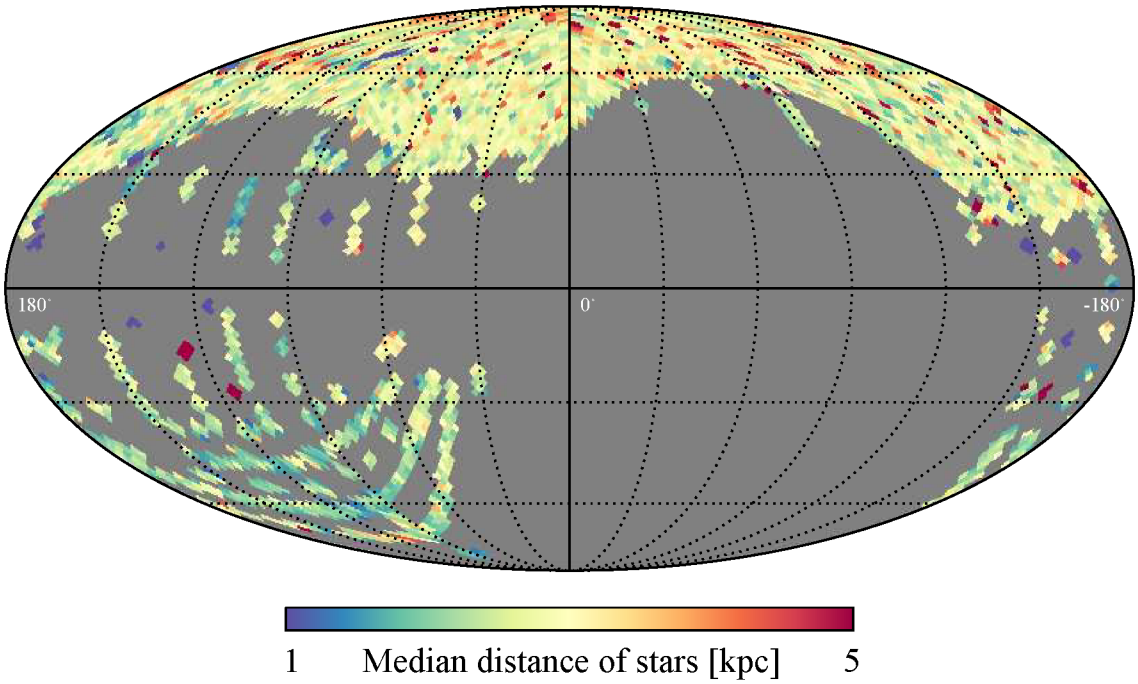

Figure 4.5: Median distance of stars in the SDSS sky coverage. The typical distance is about 2-3 kpc, depending mildly on Galactic latitude.

for extragalactic sources covers mostly high Galactic latitudes corresponding to low levels of dust reddening $E_{B-V}<$ 0.5 mag. These types of sources do not allow us to probe a wide range of ISM column densities. We therefore only use them for consistency check.

We measure the equivalent width of DIBs with a Gaussian fitting of the line profile. For cases involving blended features, we make use of Gaussian profiles with multiple components. For example, around $\lambda = 5780\,\text{Å}$, three DIBs are known to exist: λ5778 (broad), λ5780 (narrow), and λ5797 (narrow). We measure these three

DIBs simultaneously with a triple-Gaussian profile. For some broad DIB features due to blending of several weak DIBs, we fit a single Gaussian and quote them as a single DIB (e.g., λ5540). To identify DIBs potentially blended with multiple components, we compare the 20 DIBs with the DIB list compiled by Hobbs et al. (2008) with 8km/s spectral resolution and search for DIBs identified as blended. In Table 4.3, we mark those blended DIBs with star symbols. For those blended DIBs, the derived correlations in this analysis can be driven by either a single dominant DIB or a combination of multiple DIBs with different dependences. For example, DIBλ5540 in our analysis consists of three narrow DIBs with two of them, λ5541 and 5546 correlated with C_2 molecules (Thorburn et al., 2003).

We show examples of such line-profile fits in Figure 4.4. The black histogram shows the composite absorption spectrum as shown in Figure 4.3 with $E_{B-V} > 0.1$ mag and the blue lines show the best-fit single-Gaussian profiles used to estimate the absorption equivalent width. Multi-component fits are shown with red, green, and orange lines.

When measuring the equivalent widths of DIBs from high-S/N composite spectra, we fit simultaneously for the width, center, amplitude, and continuum of the absorption spectra. However, when considering lower-S/N composite spectra, which is the case for quasars, galaxies and the composite spectra of a small number of stars, we fix the width of the Gaussian profiles estimated from high-S/N composite spectra but allow other parameters to vary. From high-S/N composite spectra, we find the width

`http://dibdata.org/HD204827/`

of each DIB does not vary with the Galactic dust reddening or other variables such as sky position. Fixing the width therefore allows us to estimate the equivalent width robustly.

4.4 Results

4.4.1 The DIB absorption map

Having characterized the absorption of 20 DIBs, we can map out the strength of each band as a function of position in the sky and then use them for cross-correlation analyses with various tracers of the ISM. We first investigate the overall spatial distribution of DIB absorption. To do so we pixelize the sky under the HEALPix scheme in the Galactic coordinate system. The resolution of the map can be adjusted given the purpose of the analysis. We locate all the stars in each pixel, create a composite spectrum and measure the strength of each of the 20 selected DIBs following the procedure described in the previous section. We choose the number of HEALPix pixels along the Galactic longitude to be $N_{\rm side} = 64$, which divides the whole sphere to 49152 pixels with equal area of about one square degree. This resolution is motivated by the surface number density of the observed stars such that, in the majority of the pixels, there are enough stars for the robust absorption measurements of individual DIBs. To ensure reasonable S/N for the characterization of DIBs we only consider

`http://healpix.sourceforge.net/`

pixels with more than 5 stars. In total, we use 5929 pixels (out of 9516), covering about 5000 square degrees of the sky. The typical S/N of the composite residual spectra (normalized to unity) is about 150, and the typical error of a DIB equivalent width ranges from 20 mÅ for narrow DIBs (e.g., DIBλ4728) to 40 mÅ for broad DIBs (e.g., DIBλ4430).

We measure the absorption strength of each DIB in each pixel and create 20 such maps. To display the global DIB absorption on a map, we combine the total absorption signal from the 20 DIBs considered:

$$\mathrm{W_{all\ DIBs}} = \sum_{i=1}^{20} \mathrm{W_{DIB_i}}\,. \tag{4.1}$$

In Figure 4.6, we present the map of total DIB absorption. We note that the measurements in different pixels are quasi-independent from each other because some common reference stars are used to define the zero points. The map shows that the DIBs are more concentrated towards the disk, even though the sampling is relatively sparse at low Galactic latitude. This is expected from their known correlation with dust. We also observe smaller-scale features corresponding to known structures and clouds in the Milky Way. For example, at $l \sim 170^\circ$ and $b \sim -15^\circ$, we observe the Taurus molecular cloud (located at about 100 pc from the Sun) and its surroundings. While at $l \sim -160^\circ$ and $b \sim -20^\circ$, we are able to measure the distribution of the DIBs at the edge of the Orion molecular cloud. In the next section, we cross-correlate these maps with other all-sky surveys and study the correlations of DIBs with various ISM tracers. In Figure 4.7, we show a combined map of the 20 DIBs and an infrared

DIB detected from APOGEE spectra by Zasowski et al. (2015).

4.4.2 Dependence on ISM tracers

We now take advantage of the large-scale mapping of DIB absorption enabled by our analysis to investigate the dependence of DIBs on other ISM tracers. To do so we make use of four publicly available all-sky maps tracing metals and hydrogen:

- **Dust**: we use the map created by Schlegel et al. (1998, SFD), based on 100 μm flux from the COBE/DIRBE and IRAS/ISSA maps. It provides estimates of dust column densities in units of E_{B-V} reddening.

- **Polycyclic aromatic hydrocarbons**: we use the emission map in the 12 μm channel of the WISE all-sky survey (Wright et al., 2010), provided by Meisner & Finkbeiner (2014). The WISE 12 μm channel traces PAHs giving rise to emission between 7 and 18 μm. We note that our analysis will only make use of derivatives of this map and its overall normalization is not relevant for our purposes.

- **Neutral atomic hydrogen**: we use the Leiden/Argentine /Bonn (LAB) Galactic H I 21 cm Survey (Kalberla et al., 2005). The LAB Survey is one of the most sensitive Milky Way 21 cm survey to date, with the extensive coverage both spatially and kinematically. To compare with our integrated absorption along lines

`http://faun.rc.fas.harvard.edu/ameisner/wssa/`
`http://lambda.gsfc.nasa.gov/product/irsa/fg_LAB_HI_Survey_get.cfm`

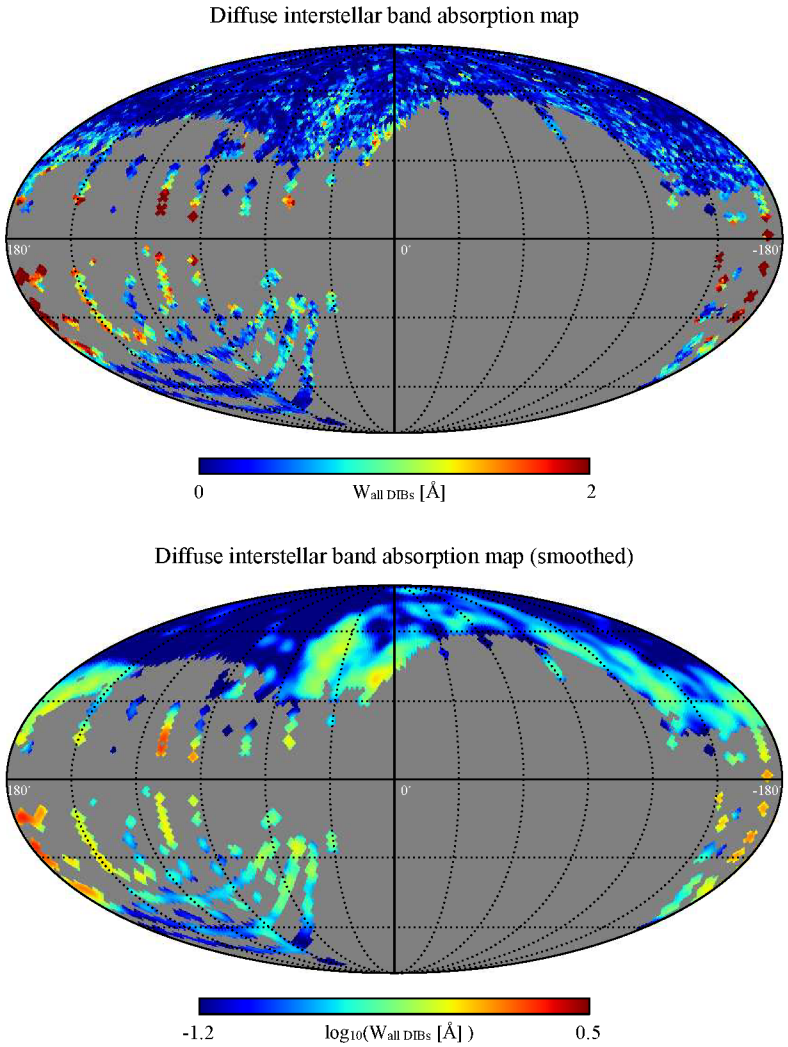

Figure 4.6: *Top:* Full sky map of DIB absorption in Galactic coordinates estimated from more than 250,000 stellar spectra. Our correlation analyses are based on about 5,000 deg^2 corresponding to regions with $E_{B-V} > 0.02$ mag. Regions with lower dust column densities, i.e. $E_{B-V} < 0.02$ mag, are used to define reference stars. The angular resolution of this map is about 1 deg^2 and the typical distance from stars to the Sun is about 2-3 kpc mildly depending on latitude. Due to variation in the density of available spectroscopic data, some of the pixels of this map are signal or noise dominated. *Bottom:* Smoothed version of the full sky map of DIB absorption, which is created by using a Gaussian kernel with FWHM=5 degrees. This map is signal dominated and reveals the morphology of DIB absorption in the sky. The map can be viewed interactively at `http://www.pha.jhu.edu/~tlan/DIB_SDSS/`.

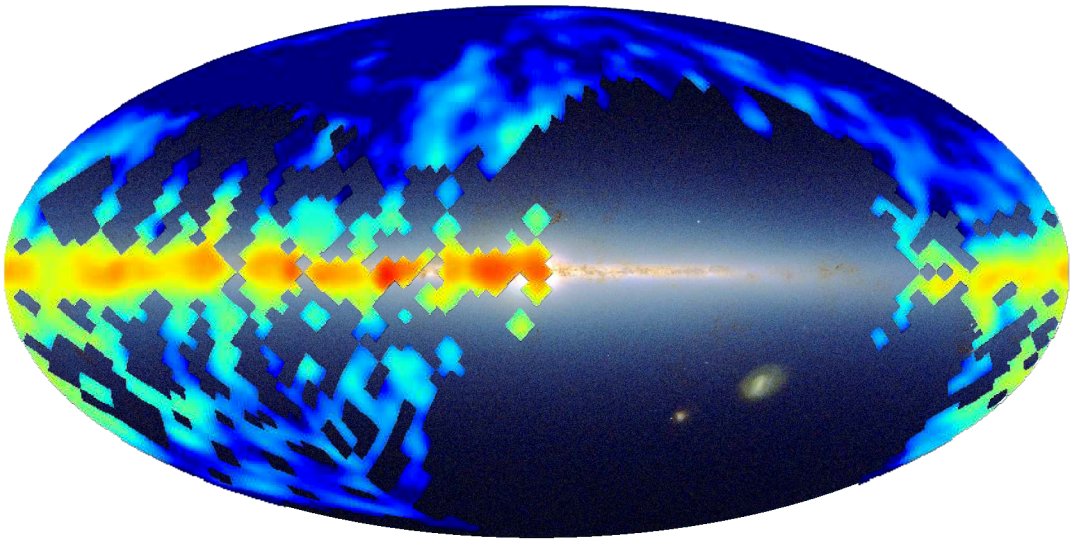

Figure 4.7: A combined map of diffuse interstellar band density from optical (Lan et al. 2015) and near-infrared (Zasowski et al. 2015) spectra.

of sight, we use their total $N_{\rm HI}$ values integrated over the full velocity range from $-450\,{\rm km\ s^{-1}}$ to $400\,{\rm km\ s^{-1}}$.

- **Molecules**: we use CO as a proxy for molecules and use $\rm CO_{1-0}$ emission map provided by PLANCK (Planck Collaboration et al., 2013). We convert the integrated line intensity $W_{\rm CO}$ to the column density of molecular hydrogen $N_{\rm H_2}$ using the CO-to-$\rm H_2$ conversion "X" factor suggested by Bolatto et al. (2013):

$$X_{\rm CO} \equiv \frac{N_{\rm H_2}}{W_{\rm CO}} = 2 \times 10^{20}\,{\rm cm^{-2}}/({\rm K\,km\ s^{-1}})\,. \tag{4.2}$$

The statistical noise level of the CO map at $\sim 15'$ resolution is about $0.45\,{\rm K\,km\ s^{-1}}$, corresponding to about $0.9 \times 10^{20}\,{\rm cm^{-2}}$.

`http://irsa.ipac.caltech.edu/data/Planck/release_1/all-sky-maps/`. We use the `Type 2` map which has a better S/N than the `Type 1` map.

We note that all the above maps are based on emission measurements which probe the ISM over substantial path lengths, including material in the Milky Way and nearby galaxies located behind the set of stars used in our absorption analysis. This effect is stronger at lower Galactic latitudes and limits our ability to measure *absolute* relations between a given absorption band and ISM column density. The tendency to overestimate ISM column densities due to the background contamination results in increasing the scattering and slightly lowering the amplitude of the measured correlation compared to the intrinsic one. However, it is important to note that this effect will affect all DIBs in the same way. Measuring changes in the correlations *between different DIBs* and ISM tracers will reveal intrinsic differences in the environmental dependencies of the corresponding DIBs.

We first investigate the total DIB absorption field and explore its global dependence on each tracer introduced above. To do so we resample all the maps to the same resolution: $N_{\rm side} = 64$ which corresponds to about one square degree. To approach the problem generically we measure the Spearman rank-order correlation coefficient between the total DIB absorption field and each tracer introduced above. We show the amplitude of the corresponding correlations, in bins of E_{B-V}, in Figure 4.8. As can be seen, we find positive correlations between DIB strength and dust, PAHs and atomic hydrogen. In contrast, the correlation coefficient with molecules appears to be negative at high dust column densities. This indicates that molecules have a different relationship with the total DIB absorption field from other tracers. We note that at

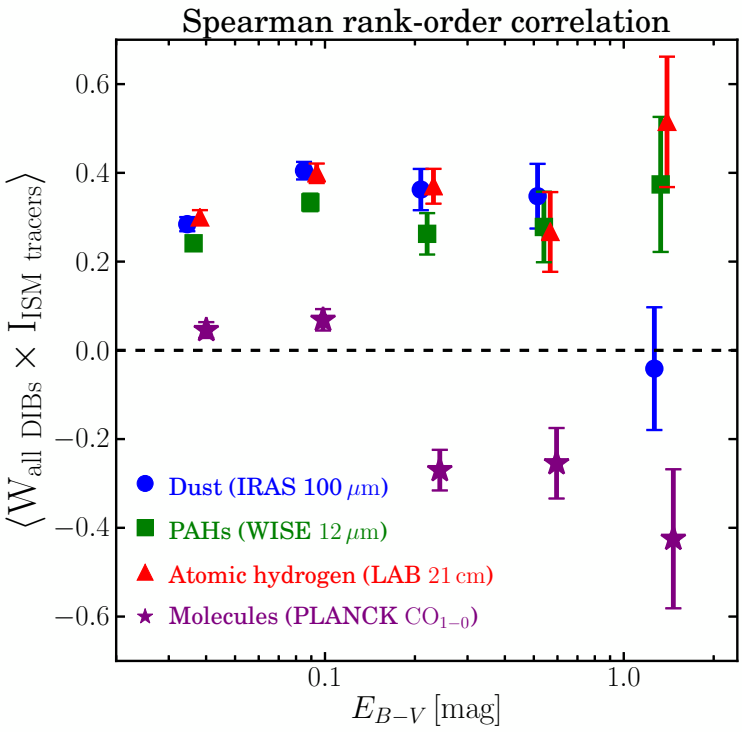

Figure 4.8: Spearman rank-order correlation coefficients between the total DIB absorption and four ISM tracers as a function of E_{B-V}. The dependence between DIBs and molecules appears to be different than that of other tracers.

low extinction, most N_{H_2} values are below the noise level, so the amplitude of the correlation coefficients in this regime mostly reflects this fact rather than the intrinsic correlation, which is unmeasurable in our data. The interpretation of the respective amplitudes needs to account for the intrinsic correlations between each ISM tracer. This is investigated in more detail below.

4.4.2.1 Dependence on dust

It has long been known that the strength of DIBs correlates with the column density of dust (Merrill et al., 1937). Our statistical approach allows us to measure these dependencies using hundreds of thousands of lines of sight. To do so we select quasar, galaxy and stellar spectra as a function of dust column density from Schlegel et al. (1998). This is done using DIB maps with a resolution comparable to that of the dust map. We combine spectra as a function of Galactic reddening to form high S/N median composite spectra and measure the equivalent widths of DIBs from the final composite spectra. The number of spectra used for the composite spectra and the corresponding S/N are shown in Table 4.2. The noise is the standard deviation of the spectra after removing outliers with 5-sigma clipping. Figure 4.9 shows the median equivalent width of the set of 20 selected DIBs in the previous section, as a function of dust reddening E_{B-V}. The blue, green and red data points are measurements from stellar, quasar, and galaxy composite spectra. The error is estimated by bootstrapping each sample and represents the error of the median equivalent width. Instead of

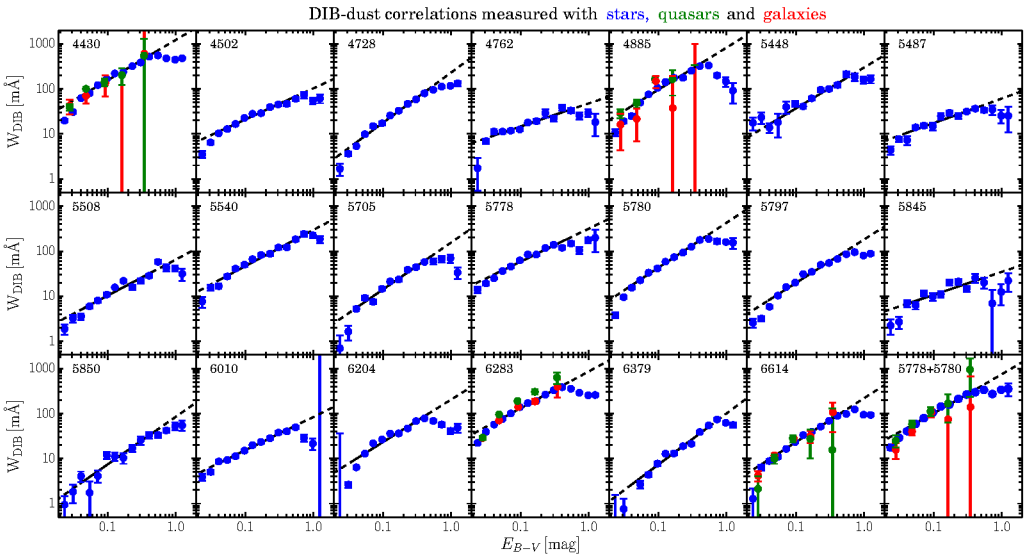

Figure 4.9: Equivalent width measurements of 20 DIBs as a function of dust reddening in units of E_{B-V}. Each equivalent width is measured from a high S/N composite spectrum combining thousands of spectra with similar Galactic reddening. Table 4.2 shows the number of spectra used in the composite spectra and the S/N of the composite spectra. The blue, green, and red data points show median values measured in star, quasar, and galaxy spectra, respectively. The solid black lines are best-fit power laws derived from measurements of stellar composite spectra at $0.04 < E_{B-V} < 0.5$ mag. The black dashed lines are extrapolations to higher and lower reddening values. We note that quasar and galaxy composite spectra do not have enough S/N to decouple DIBs λ5778 and λ5780. We therefore measure the sum of these two components and show they are consistent with measurements from stars.

constructing high S/N composite spectra, we also test the results by using the DIB equivalent width measured from each sky pixel with 1 square degree resolution and calculate the inverse variance-weighted mean DIB equivalent width as a function of E_{B-V}. Two methods yield consistent results.

We find the equivalent widths of DIBs measured from the three different types of sources to be roughly consistent with each other. We note that a perfect agreement is not expected as the extragalactic sources have a different spatial distribution over the sky. As pointed out above, composite spectra from quasars and galaxies have lower S/N ratios. In some cases they do not allow us to decompose the blending of certain lines, for example between DIB λ5778 (broad) and λ5780 (narrow). In this case we therefore compare the sum of equivalent widths of the two DIBs measured from quasar and galaxy, stellar spectra (as shown in the lower-right panel). The consistency between the measured absorption strengths of DIBs from star, quasar, and galaxy spectra shows that the methods we applied effectively remove features intrinsic to the sources and the systematics in the spectral reduction. We note that the slightly higher equivalent width for DIB λ6283 estimated from composite quasar spectra is due to contamination from a nearby sky emission line. Such effect can be observed in Fig. 4.2.

Overall, our measurements confirm strong correlations between the strength of the 20 DIBs selected and Galactic reddening. In the regime $E_{B-V}< 0.5$, the median equivalent widths of all DIBs increase with reddening and the *observed* relation

between the two quantities can be well described by a power-law function form:

$$W_{\rm DIB} = A \times (E_{B-V})^{\gamma} \, . \tag{4.3}$$

We fit our measurements with Eq. 4.3 for stellar composite spectra with $0.04 < E_{B-V} < 0.5$ mag. The lower limit is selected to eliminate the effect of the observed departure from a power-law behavior at low $\rm E_{B-V}$ values, which is due to two reasons: first we use lines of sight selected with $\rm E_{B-V} < 0.02$ mag as reference lines of sight. This prevents us from measuring the absolute level of DIB absorption as we expect the measured value of $W_{\rm DIB}$ to be zero at $\rm E_{B-V} \lesssim 0.02$ mag. Second, it is known that the SFD dust map (derived from infrared emission) is contaminated by the infrared emission of low-redshift galaxies (see Yahata et al., 2007). As a result, $\rm E_{B-V}$ values lower than 0.04 mag ($A_V < 0.1$ mag) tend to be overestimated. In the rest of our analysis we will focus primarily on the high reddening regime $\rm E_{B-V} > 0.1$ mag and these effects can be safely neglected.

The black solid lines are the best-fit power laws and the dashed lines are extrapolations towards high and low E_{B-V} values. The best-fit parameters are listed in Table 4.3 and the relative strength of each DIB, $W_{\rm DIB}/A_V$, is estimated using a value of $R_V = 3.1$ and the best-fit relations evaluated at $E_{B-V} = 0.32$ mag (or $A_V = 1$ mag). For high E_{B-V} values, we observe different behaviors in the relation between absorption strength and dust reddening. The equivalent width of certain DIBs, for example DIBs λ4502, λ4728, λ5540, and λ5850 keeps increasing with dust column density. In contrast, other DIBs depart from the trends observed at lower column

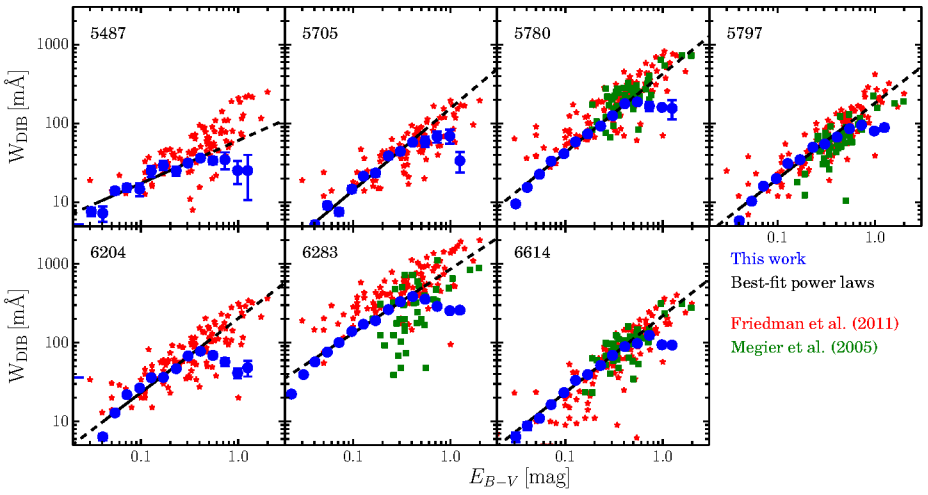

Figure 4.10: Comparison between our statistical DIB absorption measurements (blue) and results obtained from studies of individual spectra of hot stars from Friedman et al. (2011, red) and Megier et al. (2005, green). Note that the three sets of measurements involve different regions of the sky. The black line shows the best fit power law trends estimated from our measurements with $0.04 < E_{B-V} < 0.5$ mag and extrapolated at the low and high ends.

densities and flatten (e.g., λ5780) or even become weaker (e.g., λ6283). This behavior, previously reported for a number of DIBs, has been referred to as the skin effect (Herbig, 1995). We emphasize that the different trends at high E_{B-V} values reveal intrinsic differences in DIB behaviors. This relative effect is not affected by the fact that the SFD map provides estimates of dust column density across the entire Galaxy. The dust contribution associated with material located behind the set of stars is the same for *all* DIBs. To demonstrate that, we also measure the Galactic reddening based on the difference between the observed g-r color of a star and the color of the reference and perform the same analysis. The detailed comparison is discussed in Appendix C.2. We found that although the SFD map tends to overestimate the Galactic reddening along the lines of sight, the correlations derived with $0.04 < \mathrm{E_{B-V}} < 0.5$ mag and the different behaviors at high Galactic reddening from two reddening estimators are consistent.

We now compare our measurements to others results from the literature. Figure 4.10 shows the dependencies for seven DIBs studied in Friedman et al. (2011) in red and four in Megier et al. (2005) in green. These studies were based on individual high-S/N spectra of hot stars. The blue data points correspond to the statistical measurements presented above. The black lines are our best-fit power laws as shown in Fig. 4.9. As can be seen, the different sets of measurements are overall in good agreement. While our measurements only show median values, the data points from Friedman et al. (2011) and Megier et al. (2005) show measurements for individual

Table 4.2: Number of spectra used in the equivalent width vs reddening relation

$\langle E_{B-V} \rangle$ [mag]	N_{spec}	S/N
0.023	46052	3313
0.031	52501	2527
0.041	48285	2108
0.054	40407	1647
0.072	27651	1236
0.097	18913	934
0.127	15290	734
0.170	7818	599
0.230	4675	480
0.305	3476	355
0.408	2045	298
0.548	1264	283
0.728	710	268
0.982	446	255
1.247	211	239

Table 4.3: Best-fit parameters characterizing the relationships between DIBs and E(B-V) (Eq. 4.3)

λ [Å]	A	γ	W/A_V [Å/mag]
4430	1.22 ± 0.04	0.89 ± 0.02	0.44
4502	0.10 ± 0.01	0.69 ± 0.03	0.05
4728*	0.23 ± 0.01	1.11 ± 0.03	0.07
4762	0.05 ± 0.01	0.51 ± 0.06	0.03
4885*	0.82 ± 0.04	0.93 ± 0.02	0.29
5448*	0.30 ± 0.04	0.91 ± 0.07	0.11
5487	0.06 ± 0.01	0.55 ± 0.05	0.03
5508	0.07 ± 0.01	0.80 ± 0.05	0.03
5540*	0.30 ± 0.03	0.80 ± 0.04	0.12
5705*	0.16 ± 0.01	1.04 ± 0.04	0.05
5778*	0.32 ± 0.02	0.73 ± 0.03	0.14
5780	0.43 ± 0.02	1.00 ± 0.02	0.14
5797	0.18 ± 0.01	0.96 ± 0.02	0.06
5845*	0.03 ± 0.01	0.51 ± 0.08	0.02
5850	0.08 ± 0.02	1.06 ± 0.11	0.03
6010	0.08 ± 0.01	0.74 ± 0.03	0.04
6204*	0.20 ± 0.01	0.94 ± 0.02	0.07
6283*	0.86 ± 0.02	0.81 ± 0.01	0.34
6379	0.10 ± 0.01	1.13 ± 0.05	0.03
6614	0.22 ± 0.01	0.97 ± 0.02	0.07

*DIBs possibly blended with multiple weak DIBs

sightlines. We note that our estimated equivalent width for DIB λ6283 is slightly lower than that of Friedman et al. (2011). This difference could be due to that the continuum estimate is affected by the nearby strong sky line. It is interesting to note that at the high end, certain DIBs display a different behavior between the different analyses. We note that our sampling of the Galaxy differs from that of Friedman et al. (2011) and Megier et al. (2005). The SDSS spectroscopic targeting generally avoided high-extinction regions. However, a number of special fields with known molecular clouds were specifically targeted towards high Galactic reddening regions. About 50% of our lines of sight with $E_{B-V} > 0.5$ mag has molecular hydrogen fraction larger than 0.5 while the average molecular hydrogen fraction with $E_{B-V} > 0.5$ mag in the Milky way is only about 15%. Figure 4.12 also shows that most of lines of sight with $E_{B-V} > 0.5$ mag have molecular hydrogen column densities higher than the median value over the sky. Those lines of sight tend to intercept only one molecular cloud with high dust column density where the skin effect is mostly observed (Herbig, 1995). These lines of sight are therefore not directly comparable to some of the environments probed in previous studies.

4.4.2.2 Dependence on PAHs

The global rank-order correlation coefficient shown in Figure 4.8 indicates a positive correlation between the strength of the total DIB absorption field and the amount of PAHs traced by the WISE 12 μm band, similar to that found with the amount of

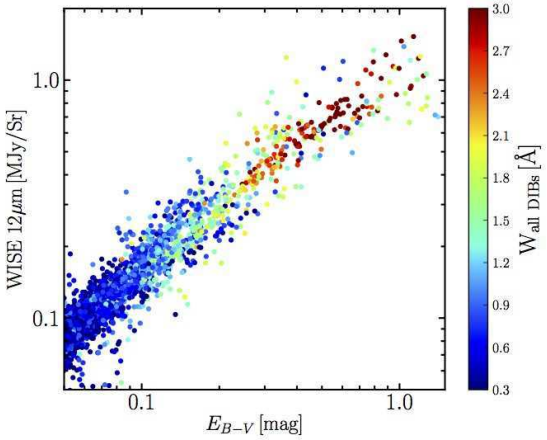

Figure 4.11: Absorption strength of the 20 selected DIBs as a function of dust reddening and PAH emission measured by WISE in its 12 μm band. We find that PAHs and dust show a tight relation which does not allow us to disentangle their respective effects on DIB absorption.

dust. In Figure 4.11 we show the relation between E_{B-V}, 12 μm flux and the total DIB absorption $W_{\rm all\,DIBs}$, where $W_{\rm all\,DIBs}$ is indicated by the color scale. First, we observe that the PAHs emission is directly proportional to that of the dust, with a scatter smaller than 0.3 dex. The strength of the total DIB absorption also appears to be roughly correlated with these two quantities. We cannot detect any vertical gradient in the DIB equivalent width at fixed E_{B-V} value. In each E_{B-V} bin the W values appear to be symmetrically distributed around the mean value. This shows that the PAHs emission does not affect the observed values of DIB equivalent width beyond the effect already due to dust column density. In other words, our analysis does not allow us to disentangle the effects of dust and PAHs on the strength of the DIB absorption.

4.4.2.3 Dependence on atomic and molecular hydrogen

We now investigate the dependence of DIB absorption on the amount of hydrogen. The global rank-order correlation coefficient shown in Figure 4.8 indicates a positive correlation between the strength of the total DIB absorption field and the amount of atomic hydrogen but a negative correlation with molecular hydrogen, traced by CO emission. As $N_{\rm HI}$ and $N_{\rm H_2}$ are not independent quantities at a fixed column density, the interpretation of the observed trends requires additional considerations. Various authors have attempted to quantify correlations between DIB strength and atomic *or* molecular hydrogen. However, it appears that only a few studies (e.g. Herbig, 1993) have investigated the dependence of DIBs on those two parameters *simultaneously*.

In Figure 4.12 we show the relationships between the total DIB absorption, $N_{\rm HI}$ and $N_{\rm H_2}$. Contrarily to what is observed with PAHs, we can see that at a fixed dust reddening value, the column density of $N_{\rm HI}$ and $N_{\rm H_2}$ affect the observed values of the total DIB equivalent width $W_{\rm all\,DIBs}$. This is illustrated in the inset of the figure which shows the DIB strength as a function of $N_{\rm HI}$ and $N_{\rm H_2}$ when selecting lines of sight with $0.2 < E_{B-V} < 0.4$ mag. We find $W_{\rm all\,DIBs}$ to increase with $N_{\rm HI}$ but decrease with $N_{\rm H_2}$. It is important to realize that E_{B-V} can be used as a proxy for the total hydrogen column density. For example, Bohlin et al. (1978) derives that, on average,

We note that recent studies (e.g. Liszt, 2014a,b) have shown that the N_{HI} and E_{B-V} relation derived from emission measurements and the relation based on $Ly\alpha$ absorption lines and reddening in background stellar spectra are not fully consistent. We caution that such an effect needs to be considered when comparing the *exact* values of derived parameters between DIBs and ISM column densities from emission-based measurements and from absorption- and reddening-based measurements. However, detailed investigations of systematics of these datasets are beyond the scope of this chapter.

$N(\mathrm{HI} + \mathrm{H_2})/E_{B-V} \simeq 5.8 \times 10^{21}\,\mathrm{atoms\,cm^{-2}\,mag^{-1}}$, over a broad range of column densities. Selecting lines of sight within a narrow range of E_{B-V} values therefore constrains the *sum* of N_{HI} and $N_{\mathrm{H_2}}$.

We now investigate, for each of our 20 DIBs, the detailed relations between absorption strength, N_{HI} and $N_{\mathrm{H_2}}$. To do so we first select regions of the sky with $\mathrm{E_{B-V}} > 0.2$ mag to focus on the regime where different DIB behaviors are observed and select three DIBs (λ4728, λ5780 and λ4885) representative of the range of correlations with molecular hydrogen. For each of them we show in Figure 4.13 the variation of their relative equivalent width as a function of both N_{HI} and $N_{\mathrm{H_2}}$. The left panel shows the 2D distribution. The arrows show the directions in which the DIB absorption strength increases. We can observe that DIB λ4728 increases towards both higher N_{HI} and $N_{\mathrm{H_2}}$, λ5780 increases only with N_{HI}, and λ4885 increases towards the lower right, where N_{HI} increases but $N_{\mathrm{H_2}}$decreases. We note that the sampling of the N_{HI}-$N_{\mathrm{H_2}}$ space is not homogeneous and some care is needed to interpret details of the data point distribution . To show more clearly the above trends, for each of the three DIBs we select a narrow $N_{\mathrm{H_2}}$ bin, $10^{20.5} < N_{\mathrm{H_2}} < 10^{21}\,\mathrm{cm^{-2}}$ corresponding to a region where N_{HI} spans a large range of value and where the sampling is not too inhomogeneous and measure the median DIB equivalent width as a function of N_{HI} , as shown in the right panel. We can see that, at fixed $N_{\mathrm{H_2}}$, these three DIBs

We note that the sky sampling provided by SDSS at high extinction, i.e. $E_{B-V} \sim 1$ mag, originates mostly from a number of special fields with known molecular clouds. As a result, our sampling of high-dust extinction regions favors high $N_{\mathrm{H_2}}$ values, a selection effect which requires some care when interpreting measured correlations.

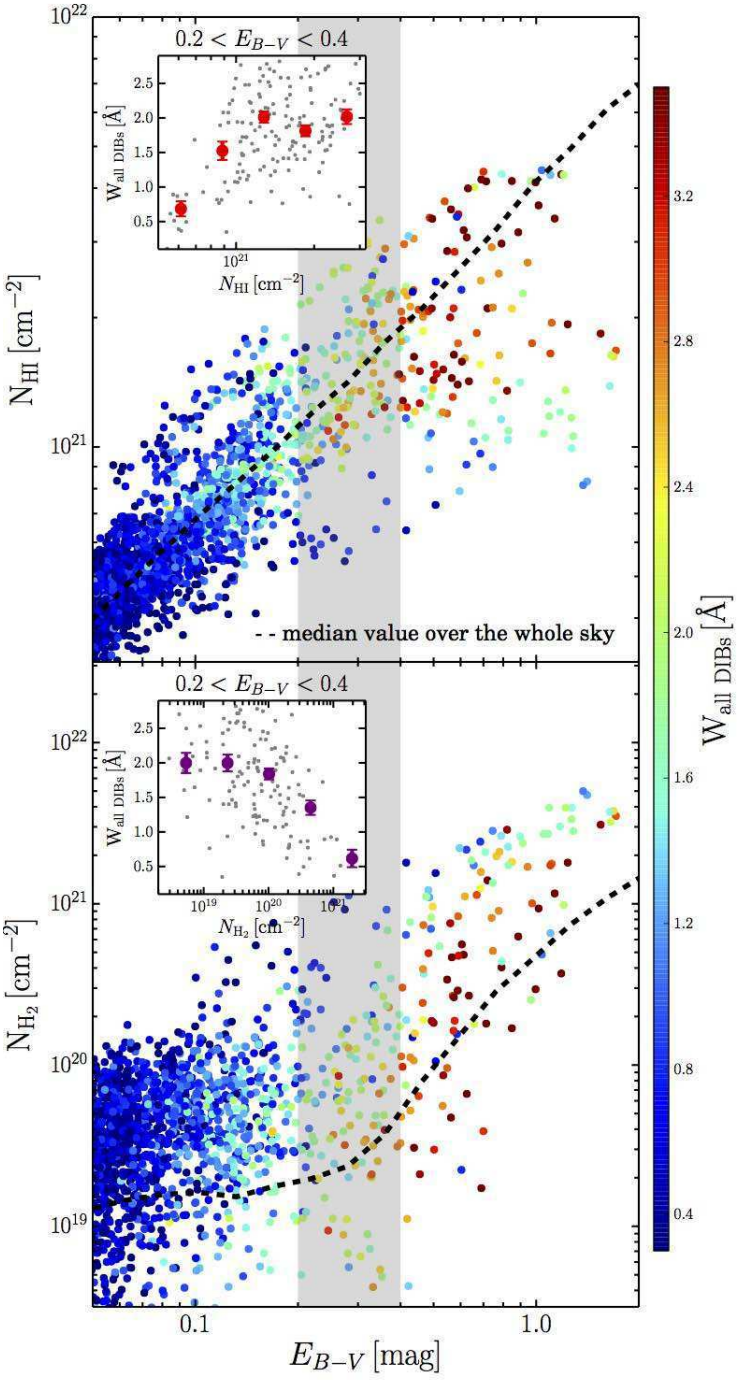

Figure 4.12: Distribution of DIB absorption strength as a function of neutral hydrogen, molecular hydrogen, and dust column densities. *Top:* projection onto the $N_{\rm HI}$-E_{B-V} plane. *Bottom:* projection onto the $N_{\rm H_2}$-E_{B-V} plane. The insets show $W_{\rm all\ DIBs}$ as a function of $N_{\rm HI}$ (top) and $N_{\rm H_2}$ (bottom) in a narrow E_{B-V} bin with $0.2 < E_{B-V} < 0.4$ mag shown in the gray regions. They illustrate the dependencies between parameters. The black dashed lines are the median of $N_{\rm H_2}$ and $\rm N_{HI}$ (including non-detection) as a function of $\rm E_{B-V}$ over the whole sky. At $\rm E_{B-V} > 0.5$ mag, most of our sightlines have $N_{\rm H_2}$ higher than the median value over the sky because of the SDSS selection. Those lines of sight tend to have lower $\rm N_{HI}$ than the median over the sky.The data points with error bars are the mean with standard errors.

all become stronger with $N_{\rm HI}$. The decline of the DIB strength at the highest $\rm N_{HI}$ is likely due to the contamination of $\rm N_{HI}$ from the background given that most of lines of sight in the bin are towards the Galactic disk with $|b| < 10$ degree. Similarly, we select a narrow $N_{\rm HI}$ bin, $10^{21} < N_{\rm HI} < 2 \times 10^{21}\,{\rm cm}^{-2}$ corresponding to a region where $N_{\rm H_2}$spans a large range of value and where the sampling is not too inhomogeneous and show the median DIB equivalent width as a function of $N_{\rm H_2}$. The different behavior as a function of molecular hydrogen is clearly seen. At a given $N_{\rm HI}$, DIB λ4728 is positively correlated with $N_{\rm H_2}$, λ5780 is not affected by the presence of molecules, while λ4885 gets weaker at higher $N_{\rm H_2}$.

We now investigate the general behavior of all of our 20 DIBs with atomic and molecular hydrogen. The previous examples, shown in Fig. 4.12 and Fig. 4.13, motivate a formulation of the overall DIB equivalent width dependence as power-law functions of $N_{\rm HI}$ and $N_{\rm H_2}$:

$$\hat{W}_{\rm DIB} = W_{21} \left(\frac{N_{\rm HI}}{10^{21}\,{\rm cm}^{-2}}\right)^{\alpha} \left(\frac{N_{\rm H_2}}{10^{21}\,{\rm cm}^{-2}}\right)^{\mu} , \tag{4.4}$$

where W_{21} is a normalization denoting the relative strength of each absorption feature (similar to W/A_V). We estimate the three parameters of the relation,W_{21}, α and μ, by minimizing a global χ^2:

$$\chi^2 = \sum_i \frac{\left(W_{{\rm DIB},i} - \hat{W}_{\rm DIB}(W_{21}, \alpha, \mu)\right)^2}{\sigma^2_{W,i}} , \tag{4.5}$$

where $W_{{\rm DIB},i}$ is the observed equivalent width of a DIB in a given pixel i of the sky map. Applying directly this χ^2 estimate to the entire dataset is subject to sampling

and selection effects. Since the vast majority of the lines of sight probe low column density values (see Fig. 4.12), a straight χ^2 evaluation would effectively only have constraining power at low $N_{\rm HI}$ values. In addition, our sampling of the high-end edge of the $N_{\rm HI}$-$N_{\rm H_2}$ space is relatively poor. To select a more homogeneously sampled space and reduce the effect of overestimating $\rm N_{HI}$ towards the Galactic disk, we only consider lines-of-sight with $E_{B-V}> 0.1\,$mag, $\rm N_{HI} < 10^{21.3}\,cm^{-2}$ and $\rm W_{CO10} > 0\,K\,km/s$. This allows us to estimate the parameters more robustly, at the cost of losing a fraction of the dataset. We note that the values of the best fit parameters α and μ are likely to be biased due to the use of emission-based column density estimations. In addition, the values also vary depending on the range of Galactic reddening and hydrogen column density values we select. However, in Appendix C.3 we show that the relative distances between pairs of the 20 (α,μ) points do not change appreciably with different selections. In the following, we will only focus on their relative distances.

The results of the fitting are presented in Table 4.4. The errors are obtained by bootstrapping the sample 200 times. Figure 4.14 shows the distribution of values as a function of α and μ. The color represents the line width of each DIB measured by Hobbs et al. (2008) with bluer indicating narrower profile. DIBs which are potentially blended with multiple components are in grey color. We find that *all* of the 20 DIBs show positive correlations with the amount of atomic hydrogen. The mean DIB equivalent widths are found to scale like $\rm N_{HI}^{\alpha}$ with $\alpha \sim 1$. In contrast, we find a range of correlations with molecular hydrogen: $W_{\rm DIB} \propto N_{\rm H_2}^{\mu}$ with $-0.2 \lesssim \mu \lesssim 0.2$. Certain

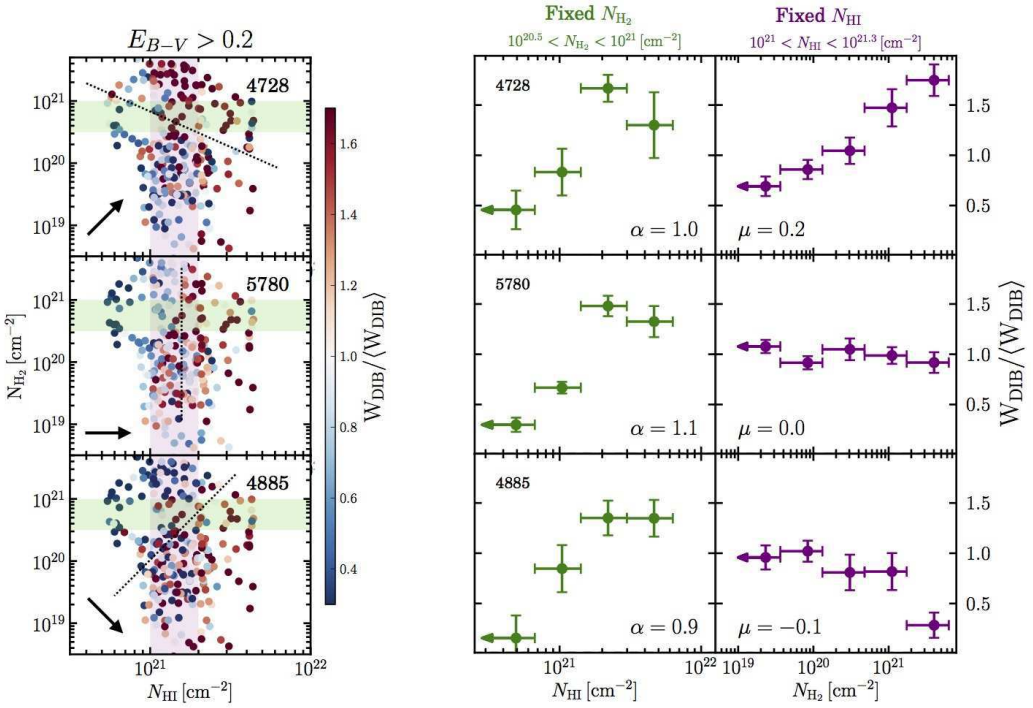

Figure 4.13: Dependencies of absorption strength as a function of $N_{\rm HI}$ and $N_{\rm H_2}$ for three DIBs: λ4728, 5780 and 4885. The left panel shows the distribution of points in the $N_{\rm HI}$, $N_{\rm H_2}$ plane for regions selected with $E_{B-V} > 0.2$. To guide the eye, the dashed lines are inserted manually to separate the blue and red data points. The arrows indicate the direction of increasing DIB strength. The green and purple bands indicate the regions at fixed $N_{\rm H_2}$ and $N_{\rm HI}$ used in the right panels respectively. *Right*: median absorption strength as a function of $N_{\rm HI}$ (at fixed $N_{\rm H_2}$) and $N_{\rm H_2}$ (at fixed $N_{\rm HI}$). The decline of the DIB strength at the highest $\rm N_{HI}$ in the middle panel is likely due to the contamination of $\rm N_{HI}$ from the background given that most of lines of sight in the bin are towards the Galactic disk with $|b| < 10$ degree.

DIBs, such as λ5780, λ4430, λ4762 have μ values consistent with zero and therefore are not sensitive to the amount of molecular hydrogen. Figure 4.15 shows the inverse variance-weighted mean of DIB equivalent widths normalized by the dependences of atomic hydrogen, $W_{\rm DIB}({\rm HI}) = {\rm W}_{21}\ \left(\frac{{\rm N_{HI}}}{10^{21}\,{\rm cm^{-2}}}\right)^{\alpha}$, as a function of molecular hydrogen column densities. DIBs are ordered by their μ values from negative (top-left) to positive (bottom-right). The figure shows that after removing the dependences on the $\rm N_{HI}$, the mean strength of DIBs with $\mu < 0$ decreases with $N_{\rm H_2}$ while those with $\mu > 0$ strengthen. We also map the relative strength of DIBs with different μ values in the sky . The map of DIBs with $\mu > 0$ appears to have more weight towards molecular clouds. We note that given the lower absolute values of μ compared to α, we expect $N_{\rm HI}$ to be the main parameter characterizing DIB equivalent widths and μ to be of secondary importance. The distribution of (α, μ) values does not appear to be correlated with the relative strength of the DIBs, W_{21}. We observe a tendency for narrower DIBs to have higher μ values, which is also shown in Welty (2014).

A number of interesting implications derive from the observed distribution of the α and μ values:

- It is expected that if two DIBs are formed from transitions between a single ground state and two different vibronic levels, their measured strengths should be perfectly correlated, with a correlation coefficient of unity. DIBs with different values of (α, μ) cannot correlate perfectly with each other and are likely

The maps of DIBs having different correlations with molecules can be viewed at `http://www.pha.jhu.edu/~tlan/DIB_SDSS/`.

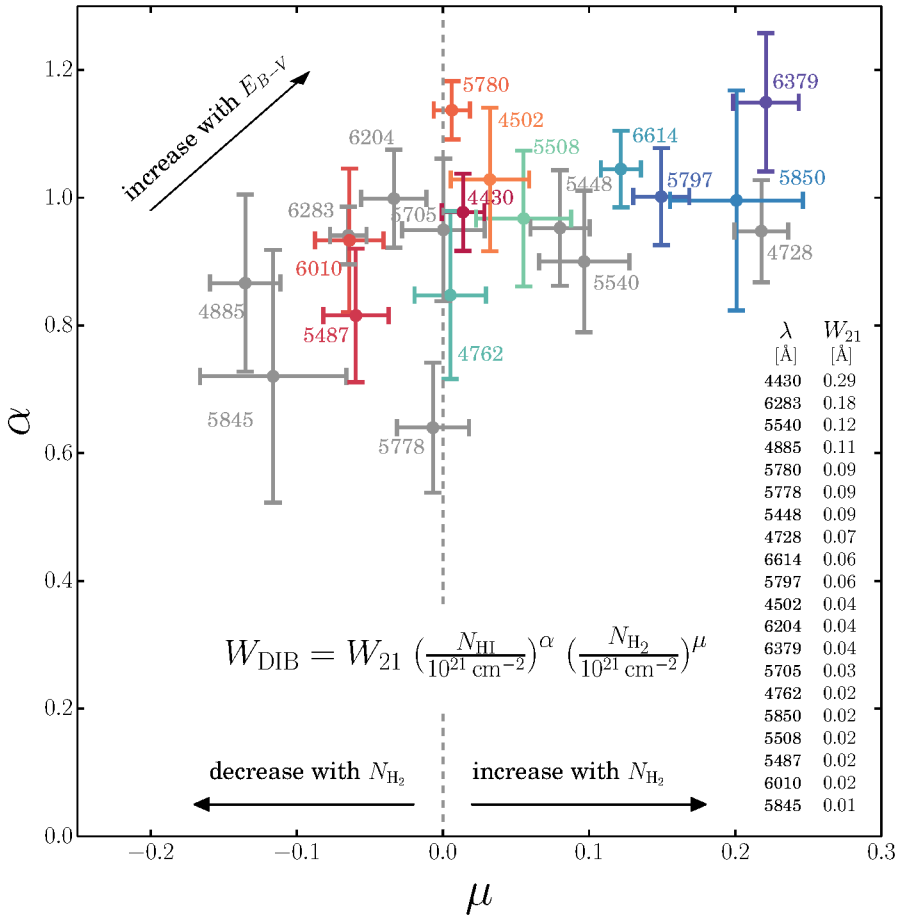

Figure 4.14: Distribution of α & μ parameters characterizing the relation between DIB equivalent width, $N_{\rm HI}$ and $N_{\rm H_2}$ according to the relation indicated in the figure (Eq.4.4). The values listed on the right show for each band the measured amplitude W_{21}. The colors indicate the line widths of DIBs from Hobbs et al. (2008). Narrow features, indicated with blue colors, appear to have preferentially positive μ values, while broad DIBs with red colors have $\mu \sim 0$. DIBs blended with multiple lines are in grey. DIBs with $\mu > 0$ are favoured in environments with higher molecular gas content.

due to different carriers. Conversely, DIBs for which measured values of (α, μ) are consistent with each other *may* originate from the same carrier or belong to a same "family". Note that this statement depends on the accuracy with which the (α, μ) parameters can be measured. Our analysis does not reveal any clustering of the points in the (α, μ) plane. Our measurements suggest a continuum in the properties of each of the 20 selected bands rather than the existence of a few families.

- At a fixed value of $N_{\rm HI}$, $N_{\rm H_2}$ can vary by orders of magnitude. The H_2 dependence, which exists when $\mu \neq 0$, explains a large fraction of the scatter observed in measurements of $W_{\rm DIB}(N_{\rm HI})$ and $W_{\rm DIB}(E_{B-V})$.

- Since dust column density is known to be roughly proportional to the *total* amount of hydrogen, we expect DIBs with high α and μ values to correlate more strongly with dust reddening. Such a trend is observed in our analysis. DIB λ6379, λ4728 and λ5850 with higher α and μ also have steeper slopes γ with E_{B-V} among the 20 DIBs shown in Table 4.3.

- DIBs with different μ values will behave differently in environments with higher molecular fraction. DIB line ratios $W(\lambda_1)/W(\lambda_2)$ will be correlated with $\mu(\lambda_1) - \mu(\lambda_2)$. This provides us with a generalization of the σ, ζ effect discussed in the literature (see Section 4.5). The distribution of points in Figure 4.14 can be used to predict that larger line ratios are expected for pairs of DIBs with greater $\Delta\mu$,

for example between DIBs λ4728 and λ4885, when probing lines of sight with higher molecular fractions.

- As shown in Eq. 4.4, the mean equivalent width of a DIB can be parametrized as a function of both $N_{\rm HI}$ and $N_{\rm H_2}$. This implies that the knowledge of the equivalent widths for two or more DIBs with different (α, μ) values can be used to infer, statistically, both $N_{\rm HI}$ and $N_{\rm H_2}$ along the corresponding lines of sight.

- We note that the dependence between $W_{\rm DIB}$ and ($N_{\rm HI}$, $N_{\rm H_2}$) can be used to predict the relation between DIB equivalent width and other ISM tracers. This can be used to explore whether additional parameters are important in describing DIB behaviors.

Here we illustrate the above point by attempting to reproduce the trends observed between $W_{\rm DIB}$ and E_{B-V}. To do so we use the same sampling of the sky as done in section 4.4.2.1. For each E_{B-V} bin we estimate the median values of $N_{\rm HI}$ and $N_{\rm H_2}$ and estimated $W_{\rm DIB}(E_{B-V})$ using Eq. 4.4 and without any knowledge of E_{B-V}. The results are shown in Figure 4.16 using red points. For comparison, we show the direct reddening-based measurements presented in section 4.4.2.1 in blue. As can be seen, our hydrogen-based formalism provides us with a good description of all the trends given by the data – for 20 DIBs over more than an order of magnitude in E_{B-V}. In particular it naturally reproduces the turnover at the high end seen for only specific DIBs, the so-called skin effect. This shows that the anti-correlation between

DIB strength and dust in the high column density regime can be explained by the correlation between hydrogen and DIBs only.

To further illustrate the meaning of the (α, μ) parameters we show expected DIB equivalent widths considering only the H I dependence. To do so we fix the amount of H_2 to a low value: $N_{H_2} = 10^{18}\,\mathrm{cm}^{-2}$. The results are shown with the green triangles. For about a third of the DIBs, we see that H_2 has virtually no effect on the relation between W and E_{B-V}. This is the case for DIBs with $\mu \sim 0$. Similarly we can see the level at which molecules can influence the trends seen for DIBs with $\mu \neq 0$.

Describing the behavior of DIBs using only H I and H_2 allows us to describe a wide range of observed properties, i.e. relations between DIBs themselves, between mean DIB strength and ISM tracers as well as some of the related scatter observed in the corresponding distributions. We note that Eq. 4.4 is not expected to reproduce all observed behaviors. Additional dimensions might be required. However our approach has shown that a large fraction of the DIB dependencies and variances could be simply explained by their relations to atomic and molecular hydrogen.

4.5 Discussion

Our analysis has allowed us to map out the strength of 20 DIBs over a large fraction of the sky and derive a simple formulation of the mean DIB equivalent width as a function of only atomic and molecular hydrogen column densities. So far, most

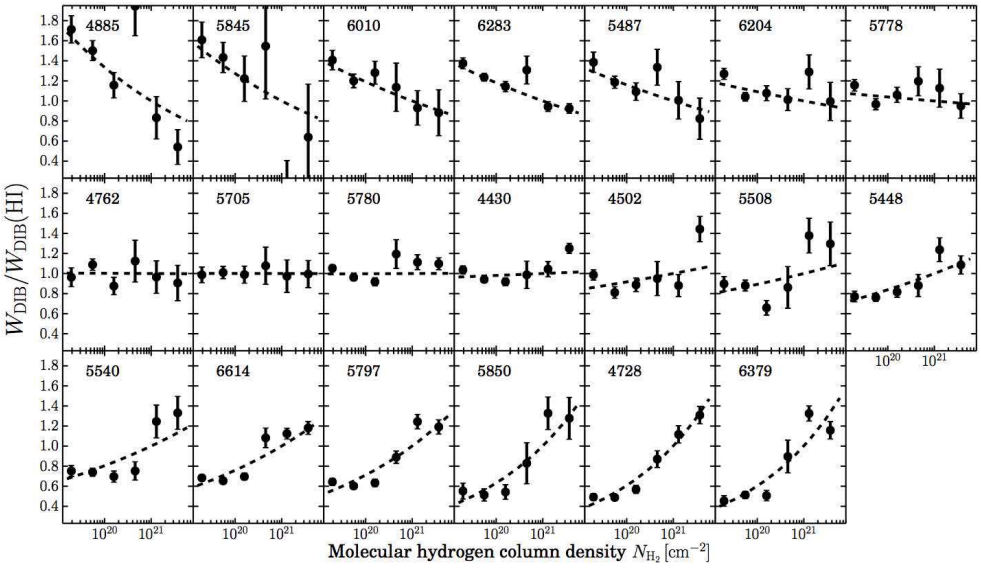

Figure 4.15: DIB equivalent widths normalized by the dependences of atomic hydrogen, $W_{\rm DIB}({\rm HI}) = {\rm W}_{21}\ \left(\frac{\rm N_{HI}}{10^{21}\,{\rm cm}^{-2}}\right)^{\alpha}$, as a function of the molecular hydrogen column densities. DIBs are ordered with their μ values from negative (top-left) to positive (bottom-right). Data points are inverse variance-weighted mean of DIB equivalent width normalized by the dependences of atomic hydrogen and the error is estimated by bootstrapping. The dashed lines are the best-fit dependences of molecular hydrogen, $\left(\frac{N_{\rm H2}}{10^{21}\,{\rm cm}^{-2}}\right)^{\mu}$.

DIB analyses focused on *projections* of the relation $W_{\rm DIB} = f(N_{\rm HI}, N_{\rm H_2})$. Only a few (Herbig, 1993; Thorburn et al., 2003) considered the simultaneous dependence on multiple ISM tracers but were based on small samples. As most interstellar quantities show a positive correlation with each other, simply due to the increase of interstellar material with distance, pairwise correlations between observables are often driven by this effect. This is most clearly reflected in the correlation of DIBs with extinction. A multi-dimensional approach is required to disentangle different effects. We now show that our results, and in particular the distribution of (α, μ) values presented in Figure 4.14 and Table 4.4, are consistent with numerous observational results reported in the literature, shed light on the origin of various correlations, and can be used to predict correlations not yet measured.

4.5.1 DIB families

Various authors have studied cross-correlations between the strength of different DIBs or compared their dependences on ISM tracers to assess whether or not they may belong to families. If not, they must arise from different carriers. Similarly, with our formalism, DIBs with different values of (α, μ) cannot correlate perfectly with each other and are likely due to different carriers.

Investigating different dependences with dust reddening, Josafatsson & Snow (1987) studied six DIBs and defined two classes: the first one consists of DIBs λ5780, λ5797 and λ5850 and the second group contains the broader bands λ5778 and λ5845.

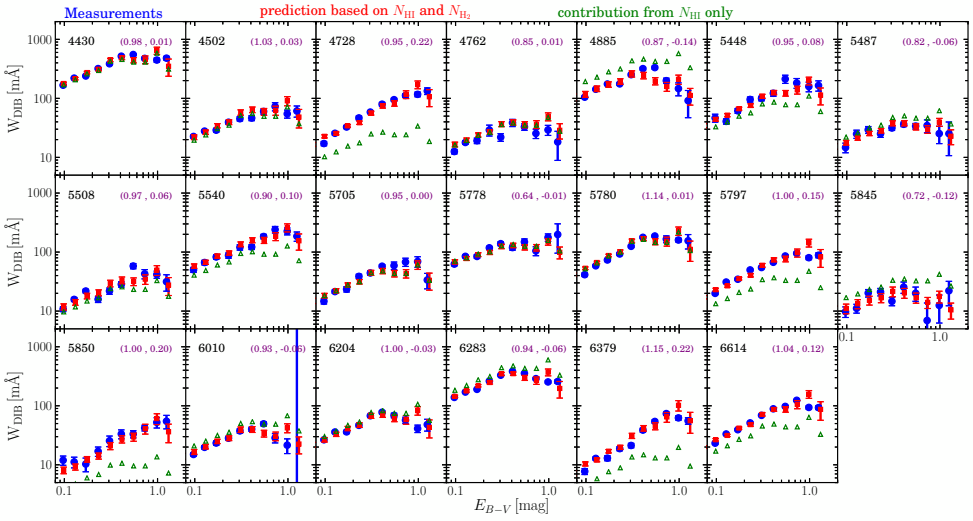

Figure 4.16: Estimates of DIB absorption as a function of dust reddening based on our formalism involving only ($N_{\rm HI}$,$N_{\rm H_2}$). The red squares are estimated equivalent widths using the median $N_{\rm HI}$ and $N_{\rm H_2}$ at each E_{B-V} bin and Equation 4.4. Blue points show direct measurements. The green triangles are shown to illustrate the effect of H_2. To do so we show the estimated absorption strength only based on measured values of $N_{\rm HI}$ and fixing $N_{\rm H_2}$= 10^{18} cm^{-2}. The purple numbers shown on the top-right corners show the (α,μ) values inferred for each DIB.

This is consistent with our decomposition. DIB λ5778 and λ5845 are on the lower left side on the figure and are expected to correlate less with dust. The (α, μ) values of the two group are inconsistent with each other.

Cami et al. (1997) studied 44 DIBs and identified two families based on the degree of correlation between bands. The first one contains DIBs λ5797, λ6379 and λ6614 and the second one the λ4502, λ5789, λ6353 and λ6792 DIBs. For the bands in common with our analysis, this decomposition is consistent: the members of the first group all live in the upper right corner of Fig.4.14 have high α and high μ values. DIB λ4502 which belongs to the second group has a lower μ value. Reported correlations between additional DIBs and these two families appear to be in agreement with our findings.

Friedman et al. (2011) investigated cross-correlations between DIB strengths for a selection of eight bands: they report that the two weakest pairwise correlations are found from (λ5797, λ6283) & (λ5797, λ5487). These two pairs correspond to the two largest distances in our (α, μ) plane. They also report that the correlation of DIBs λ5780 & λ5705 is high. We note that the (α, μ) values of these two bands are consistent with each other.

As mentioned in Cox et al. (2005) the literature indicates the existence of families of DIBs based on the shape of their profile (e.g. Cami et al., 1997; Galazutdinov et al., 2003; Josafatsson & Snow, 1987; Krelowski & Walker, 1987; Moutou et al., 1999; Porceddu et al., 1991; Wszołek & Godłowski, 2003). Members of one group

have narrow profiles with a sub-structure that is indicative of a gas-phase molecule (Ehrenfreund & Foing, 1996; Sarre et al., 1995). They are DIBs λ5797, λ5850, λ6196, λ6379 and λ6614. We note that all of them appear clustered in the upper right hand side of the α, μ plane. A second group, with DIBs λ5780, λ6283 and λ6204, has absorption features with no apparent substructure (e.g., Cami et al., 1997). We note that all these DIBs have $\mu \sim 0$.

4.5.2 Correlations with atomic hydrogen

Correlations between DIBs and the amount of atomic hydrogen have been reported for a long time (Herbig, 1993). In their analysis of eight DIBs, Friedman et al. (2011) reported cross-correlation coefficients between $N_{\rm HI}$ and eight DIBs: λ5780, λ6204, λ6283, λ6196, λ6614, λ5705, λ5797, and λ5487, in decreasing order of correlation amplitude. This trend is consistent with our results within the uncertainties: λ5780 & λ6204 have $\mu \sim 0$ while λ5797 and λ5487 have $\mu \sim 0.1$ and -0.05, respectively. The fact that DIB λ5780 has the highest degree of correlation with hydrogen is reflected by the fact that it has the highest α value and μ consistent with zero. Welty (2014) also found that DIBλ5780 tends to be weaker with higher molecular *fraction* (less atomic hydrogen) which is consistent with our results.

4.5.3 Molecules and the σ, ζ dichotomy

Krelowski & Westerlund (1988) found that the λ5797/λ5780 ratio is higher towards star ζ Oph than towards σ Sco. As molecules (e.g., CN, CH) are more abundant towards ζ Oph (Danks et al., 1984), it was suggested that clouds with higher molecular fractions cause stronger λ5797 absorption than λ5780. This result introduced the σ/ζ types of lines of sight. We can illustrate this effect with our measurements. In Figure 4.17 we show the measured ratio between the equivalent width of DIBs λ5797 and λ5780, as a function of molecular hydrogen column density for lines of sight selected within a narrow range of atomic hydrogen column densities: $10^{21} < N_{\rm HI} < 2 \times 10^{21}\,{\rm cm}^{-2}$. As clearly seen, the ratio between these two lines increases with higher molecular fraction or in other words when transitioning from σ-type lines of sight to ζ-type.

Kos & Zwitter (2013b) observed 19 DIBs and classified them into two groups according to their correlations with dust as a function of the σ/ζ type sightlines. Their classification is consistent with the results derived with our formalism. DIBs such as λ5780, λ5705 and λ6204 having values of μ consistent with zero behave similarly and are classified into the same group (Type I), while DIBs λ6614, λ5797, λ5850 and λ6379 living in the right hand side of Fig. 4.14 with $\mu \simeq 0.1 - 0.2$ are classified into the other group (Type II).

In their analysis of eight DIBs, Friedman et al. (2011) reported that DIBs λ6614 and λ5797 have the strongest correlation coefficients with H_2. This is consistent with

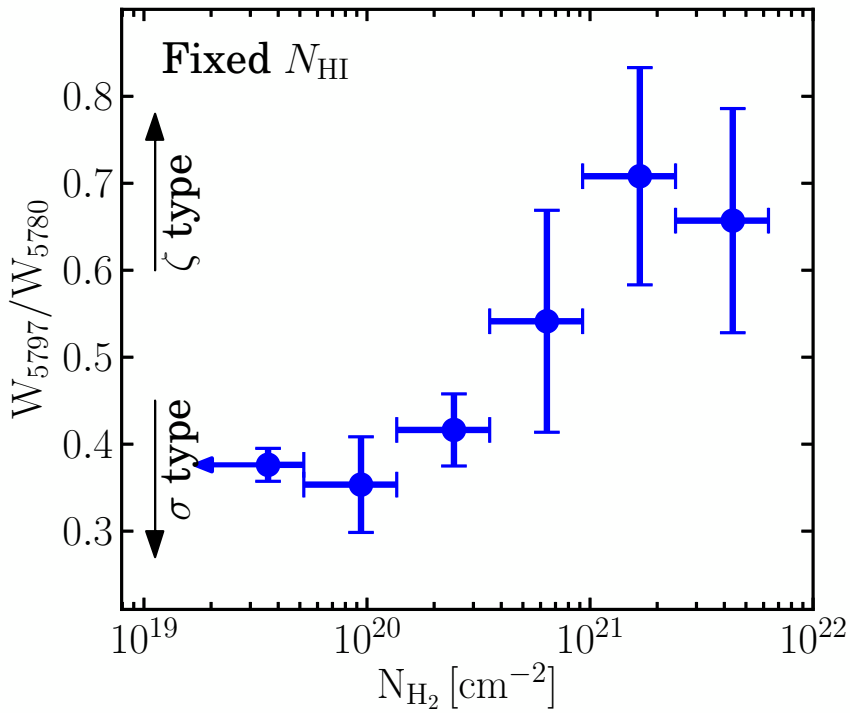

Figure 4.17: The ratio between the equivalent width of DIB λ5797 and DIBλ5780 as a function of molecular hydrogen column density for lines of sight selected with $10^{21} < N_{\rm HI} < 10^{21.3}\,{\rm cm}^{-2}$.

the fact that, among their selected DIBs, those two have the highest μ values.

Our findings are also consistent with the results of Thorburn et al. (2003). These authors found DIBλ4728 correlates with molecules traced by C_2. In addition, they found no correlation between $W(\lambda6614)/W(\lambda6196)$ and molecules (traced by C_2, CN and CH) and anti-correlation $W(\lambda6204)/W(\lambda6196)$. This is in agreement with the fact that $\mu(\lambda6614) > \mu(\lambda6204)$. Similarly, Krełowski et al. (1999) and Vos et al. (2011) showed that $W(\lambda5797)/W(\lambda5780)$ is positively correlated with $W(\mathrm{CH})/E_{B-V}$. This is in line with the fact that $\mu(\lambda5797) > \mu(\lambda5780)$. Interestingly the ratio $W(\lambda5797)/W(\lambda5780)$ does not correlate with $W(\mathrm{CH}^+)/E_{B-V}$.

4.5.4 Correlations with dust

As pointed out earlier, dust reddening can be used as a proxy for the total hydrogen column density: $N(\mathrm{HI}+\mathrm{H_2})/E_{B-V} \simeq 5.8\times10^{21}\,\mathrm{atoms\,cm^{-2}\,mag^{-1}}$ (Bohlin et al., 1978). In Figure 4.16 we have already shown that it is possible to reproduce the trends observed between the mean DIB equivalent widths and E_{B-V} by considering only the amount of H I and H_2 present along the lines-of-sight. The break seen at high E_{B-V} values, also called the skin effect (Snow & Cohen, 1974), can be quantitatively characterized by the decrease of H I in dense clouds, with higher molecular fraction. As indicated in Fig. 4.12 some of the scatter in the relation between W_{DIB} and dust reddening originates from the existence of a distribution of N_{HI} and $N_{\mathrm{H_2}}$ values along the corresponding lines of sight.

Finally, we note that our ability to reproduce a large range of observational results using the (α, μ) parametrization derived in section 4.4.2.3 indicates that the statistical measurements of DIB equivalent widths derived from the analysis of the Sloan Digital Sky Survey are robust. The existence of a relation between W_{DIB}, $N_{\rm HI}$ and $N_{\rm H_2}$ explains a large fraction of the scatter observed when considering correlations between the equivalent width of a DIB and *one* ISM tracer, such as $N_{\rm HI}$, $N_{\rm H_2}$, molecular fraction or E_{B-V}.

4.6 Summary

We have used about $500,000$ spectra of stars, quasars and galaxies taken by the Sloan Digital Sky Survey to map out the distribution of diffuse interstellar bands induced by the interstellar medium of the Milky Way. We have showed that, after carefully removing the intrinsic spectral energy distribution of each source and taking care of spectroscopic calibration effects and spectral features due to the Earth's atmostphere, it is possible to detect absorption features at the 10^{-3} level. This allows us to detect more than twenty DIBs from $\lambda = 4400$ to $6700\,\text{Å}$ and measure their strength as a function of position on the sky. Focusing on a set of twenty bands, for which we can robustly characterize the line properties, we have created a map of DIB

absorption covering about 5,000 square degrees. This map can be used to measure correlations with various tracers of the ISM: atomic & molecular hydrogen, dust and polycyclic aromatic hydrocarbons. Our findings can be summarized as follows:

- For each of the 20 selected DIBs, we have measured their mean absorption as a function of dust reddening and found results consistent with previous studies based on individual spectra of hot stars. For certain DIBs, we observe a break at high E_{B-V} value, above which the absorption strength decreases with dust column density, the so-called skin effect.

- As various ISM tracers are correlated with each other, a multi-dimensional approach is required to disentangle different effects. Investigating the dependence of DIB absorption strength on atomic and molecular hydrogen simultaneously we find that, on average, the three quantities can be described by

$$W_{\rm DIB} \propto (N_{\rm HI})^{\alpha} \, (N_{\rm H_2})^{\mu}$$

(see Eq. 4.4). For all DIBs we find $\alpha \sim 1$ but we find a range of values for μ, from -0.2 to $+0.2$, indicating that different DIBs have a different affinity to molecules. DIBs with $\mu > 0$ are favoured in environments with higher molecular gas content. This parametrization also shows that $N_{\rm HI}$ is the main parameter characterizing DIB equivalent widths. The effect of $N_{\rm H_2}$ is, in general, of secondary importance but can dominate in certain regimes for DIBs with μ values departing from zero.

- We show that the combined dependence on both $N_{\rm HI}$ and $N_{\rm H_2}$ can be used to reproduce a number of observational results. For example, one can reproduce the observed trends between DIB absorption and dust reddening (including the so-called skin effect) using only the hydrogen-based parametrization. We also note that the combined dependence on $N_{\rm HI}$ and $N_{\rm H_2}$ can largely explain the scatter observed in a number of correlations between DIBs and ISM tracers.

- We estimate the (α, μ) values for the 20 DIBs and study their distribution. DIBs with different (α, μ) values cannot correlate perfectly with each other and are likely due to different carriers. We show that the inferred (α, μ) values are consistent with numerous observational results reported in the literature and shed light on the origin of various correlations: (i) the relations between DIBs themselves; (ii) relations between DIB strength and ISM tracers; (iii) the σ/ζ dichotomy, which can be generalized to any pair of DIBs; (iv) and some of the related scatter observed in the corresponding distributions. The estimated (α, μ) values can also be used to predict correlations not yet measured. We note that the consistency with such a broad set of observational results demonstrates the robustness of the equivalent width measurements of the selected 20 DIBs from SDSS stellar spectra.

While the origin of diffuse interstellar bands is still a mystery, our approach provides us with a new view and parametrization of numerous observational results previously reported and a metric to characterize the affinity between DIBs. Our

parametrization of the strength of DIB absorption as a function of atomic and molecular hydrogen column densities might help us shed light on the physical mechanisms involved with the production and destruction of the DIB carriers.

Table 4.4: Best fit parameters characterizing the relation between DIBs and hydrogen column densities (Eq. 4.4)

λ (DIB)	α (H I)	μ (H_2)	W_{21}
[Å]	–	–	[Å]
4430	0.98 ± 0.06	$+0.01 \pm 0.01$	0.287 ± 0.012
4502	1.03 ± 0.10	$+0.03 \pm 0.03$	0.038 ± 0.003
4728*	0.95 ± 0.09	$+0.22 \pm 0.02$	0.070 ± 0.003
4762	0.85 ± 0.12	$+0.01 \pm 0.03$	0.024 ± 0.002
4885*	0.87 ± 0.13	-0.14 ± 0.03	0.111 ± 0.011
5448*	0.95 ± 0.09	$+0.08 \pm 0.02$	0.087 ± 0.005
5487	0.82 ± 0.12	-0.06 ± 0.03	0.021 ± 0.002
5508	0.97 ± 0.12	$+0.06 \pm 0.04$	0.022 ± 0.002
5540*	0.90 ± 0.11	$+0.10 \pm 0.03$	0.118 ± 0.008
5705*	0.95 ± 0.11	$+0.00 \pm 0.03$	0.028 ± 0.002
5778*	0.64 ± 0.11	-0.01 ± 0.03	0.089 ± 0.007
5780	1.14 ± 0.05	$+0.01 \pm 0.01$	0.089 ± 0.004
5797	1.00 ± 0.07	$+0.15 \pm 0.02$	0.058 ± 0.002
5845*	0.72 ± 0.21	-0.12 ± 0.05	0.010 ± 0.002
5850	1.00 ± 0.15	$+0.20 \pm 0.04$	0.024 ± 0.002
6010	0.93 ± 0.13	-0.06 ± 0.02	0.020 ± 0.002
6204*	1.00 ± 0.08	-0.03 ± 0.02	0.037 ± 0.003
6283*	0.94 ± 0.05	-0.06 ± 0.01	0.177 ± 0.007
6379	1.15 ± 0.11	$+0.22 \pm 0.02$	0.035 ± 0.002
6614	1.04 ± 0.06	$+0.12 \pm 0.01$	0.063 ± 0.002

*DIBs possibly blended with multiple weak DIBs

Chapter 5

Summary and Future Outlook

In this dissertation, I study the connection between stars, gas, and dark matter with large surveys to uncover and detail physical mechanisms that drive the evolution of galaxies. I apply cross-correlation techniques to extract new information from large sky survey data in the context of galaxy formation. With the techniques, in Chapter 2, I investigate the galaxy populations residing in dark matter halos. In Chapter 3, I examine the correlations between galaxies and gas around them. Through a new data-driven method to extract weak intervening absorption lines on top of arbitrarily complicated spectra of background sources. In Chapter 4, I map the distribution of a set of weak absorption lines, tracers of unknown molecules, in the Milky Way by applying the method to hundred thousands of SDSS spectra.

Together these studies uncover new correlations between multiple components of the Universe and shed light on physical processes driving the evolution of galaxies.

In the following sections, I will summarize the key results and discuss their future applications both from scientific and technical aspects.

5.1 Summary

In Chapter 2, I study the relationship between galaxies and dark matter halos by measuring the luminosity functions of galaxies as a function of dark matter halos, the so-called conditional luminosity functions. First, I find that there is an ubiquitous upturn at the faint-end of the luminosity functions. Second, I show that by convolving the measured conditional luminosity functions with the dark matter halo mass function, I accurately recover the field luminosity function of galaxies measured by Blanton et al. (2005) over 10 visual magnitudes. This reveals the transition at which the field luminosity function is dominated by centrals or satellites. It also shows that interpreting the faint end slope is more complicated than previously thought. Using a simple model linking satellite galaxies to dark matter halos, I demonstrate that (i) the shape of the conditional luminosity functions can be used to infer the formation history of galaxies and (ii) it is formation, rather than halo-specific environmental effect, that regulates the stellar mass of satellite galaxies in dark matter halos.

In Chapter 3, I investigate the properties of gas around galaxies, the circumgalactic medium (CGM). I find that the cool gas around star-forming galaxies and passive galaxies exhibit different behaviors, showing that the dichotomy of galaxy types is

reflected in the distribution of halo gas. Based on the spatial distribution of gas around galaxies, I show that the difference between gas content around star-forming and passive galaxies is likely associated with outflows from star-forming galaxies. Finally, I investigate the change of the cool CGM as a function of galaxy properties such as stellar mass, star formation rate, and specific star formation rate. These measurements for the first time, establish scaling relations to link the properties of galaxies and their cool CGM.

In Chapter 4, I explore diffuse interstellar bands (DIBs), a set of weak absorption lines within our Milky Way by combining a data-driven method and thousand hundreds of SDSS spectra. I study the relations between these DIBs and different ISM tracers: dust, atomic hydrogen, CO and PAHs. I show that the absorption strengths of the DIBs correlate differently with the abundance of CO molecules, suggesting that the DIB carriers have different sensitivity to the physical conditions in the ISM (e.g., radiation field). To characterize the behaviors of DIBs, I introduce an 2D parametrization involving the abundances of atomic and molecular hydrogen. I show that this parametrization is sufficient to reproduce all the trends observed between DIBs and ISM tracers in the past 25 years. This study has provided us with a new approach to investigate the behaviors of DIBs and sheds light on the physical mechanisms involved in the production and destruction of large molecules in our Milky Way.

5.2 Future outlook

Astronomical research is going through a paradigm shift. Large sky surveys have opened a new way to explore, to discover, and to characterize new phenomena. This dissertation applies novel techniques, allowing us to exploit data from large surveys. In what follows, I discuss the future applications of the techniques with ongoing and upcoming surveys.

5.2.1 Applications of cross-correlating spectroscopic and photometric datasets

The cross-correlation technique, applied in Chapter 2 and 3, allows us to extract information by combining spectroscopic and photometric datasets. Together with the technique, new sky surveys will enable us to study the redshift evolution of the connection between galaxies and their halos. As indicated by recent studies (e.g. Liang et al., 2015), the properties of gas around galaxies and their redshift evolution are important constraints on models of galaxy formation and evolution. These observational measurements will shed new light on physical processes driving the evolution of galaxies and mechanisms determining the observed properties of galaxies.

Moreover, the cross-correlation technique can be used to estimate the redshift distribution of photometric objects by using the fact that matter clusters in the Universe. This new type of redshift estimation, the clustering redshift (e.g. Ménard et al.,

2013; Rahman et al., 2015), has different properties compared to photometric redshift methods. It is free from catastrophic outliers and assumptions of the properties of training sets used in classical photometric redshift methods. This clustering-redshift method will extend photometric observations that require accurate redshift distributions of objects, e.g., weak-lensing measurements— a key scientific driver for future surveys.

Finally, the cross-correlation technique will enable us to discover unexpected and interesting signals in a dataset. For example, in Chapter 3, by cross-correlating metal absorbers with GALEX photometric objects, I found that there are some objects in GALEX correlating with metal absorbers at redshifts around 2. These objects, if confirmed, could be a new type of rare objects that are bright in rest-frame extreme ultraviolet wavelengths. They may play important roles in re-ionizing the Universe. With future surveys, more rare and interesting objects can be discovered through their cross-correlation signals with other known populations in the Universe.

5.2.2 Applications of the spectroscopic analysis

To exploit upcoming datasets provided by future spectroscopic surveys, the method, developed in Chapter 4, for studying weak intervening absorption lines on top of complicated spectra of background sources is essential. It will allow us to extract absorption-line features below the noise level of individual spectra and to characterize the properties of the mysterious absorption lines across the sky. With future sky

surveys, it is also possible to detect new unknown absorption features in optical and near-infrared wavelengths.

A recent lab experiment (Campbell et al., 2015) and an observation (Walker et al., 2015) have confirmed that four DIBs are due to C_{60}^{+} molecule. It suggests that DIBs can be absorption signatures of some large organic molecules. Once future laboratory experiments identify more carriers of the diffuse interstellar bands, the DIB map obtained in Chapter 4 will allow us for the first time to estimate the abundances of those organic molecules, which may play important roles in astro-chemistry and therefore how life forms.

In addition, the DIBs are powerful tracers of ISM properties. Based on their velocity information, we can study the kinematic structure of our ISM. It is also possible to use the DIBs as tracers of diffuse molecular clouds (dark clouds) that can not be probed through emission from CO molecules. Moreover, including the distance information of the background sources, one can build the 3D spatial distribution of gas in and around the Milky Way. Again, this is a new type of observations that can only be made possible with the combination of large spectroscopic datasets and a powerful technique to extract the signals.

5.3 Final remarks

In the past 20 years, we have made significant progress in understanding the origin of the Universe. For the next 20 years, the new challenge will be to understand the formation and evolution of galaxies. Large sky surveys will provide tremendous information about the sky and our mission will be to discover and extract signals that can not only help us to disentangle complicated interactions between multiple components of galaxies but also shed light on the physical mechanisms driving the interactions.

In this dissertation, I have showed that using statistical techniques, one can extract new relationships between the properties of stars, gas, and dark matter from sky survey datasets. This result demonstrates that combining large survey data and statistical techniques is a powerful and promising approach to study galaxy formation and evolution. Pursuing this direction, we will be able to uncover important new astronomical phenomena and make great progress in our understanding of galaxy formation and evolution.

Appendix A

Appendix for Chapter 2

A.1 Best fit parameters

The following tables list all the best-fit parameters of the measured conditional luminosity functions in various cases. The quantities listed are defined in the main text.

A.2 Reliability Tests

A.2.1 Using different background subtractions

We test our conditional luminosity functions obtained with the global and local background estimators described in Section 2.4. The results from the global (pur-

Table A.1: Best-fit parameters of conditional luminosity functions

$\log(M_{200}/\mathrm{M}_\odot)$	$\langle\log(M_{200}/\mathrm{M}_\odot)\rangle$	M_b^*	α_b	N_b	M_f^* (fixed)	α_f	N_f
$[12.00, 12.34)$	12.15	-20.23 ± 0.42	-1.08 ± 0.26	0.36 ± 0.16	-18	-1.43 ± 0.18	0.71 ± 0.20
$[12.34, 12.68)$	12.49	-20.32 ± 0.54	-0.85 ± 0.39	0.77 ± 0.39	-18	-1.72 ± 0.24	0.74 ± 0.33
$[12.68, 13.03)$	12.84	-20.63 ± 0.41	-0.92 ± 0.27	1.43 ± 0.58	-18	-1.63 ± 0.17	2.09 ± 0.60
$[13.03, 13.37)$	13.17	-21.10 ± 0.25	-0.94 ± 0.16	3.04 ± 0.79	-18	-1.58 ± 0.19	4.50 ± 1.32
$[13.37, 13.71)$	13.51	-21.18 ± 0.18	-0.85 ± 0.12	6.19 ± 1.08	-18	-1.34 ± 0.53	3.62 ± 2.56
$[13.71, 14.05)$	13.85	-21.06 ± 0.17	-0.69 ± 0.15	14.42 ± 2.24	-18	-2.36 ± 0.37	3.53 ± 2.99
$[14.05, 14.39)$	14.21	-21.54 ± 0.07	-1.04 ± 0.04	20.84 ± 1.57	-18	-2.13 ± 0.48	3.45 ± 5.18
$[14.39, 14.73)$	14.53	-21.32 ± 0.15	-1.02 ± 0.09	56.06 ± 9.74	-18	-2.38 ± 0.51	9.63 ± 9.73
$[14.73, 15.08)$	14.87	-21.45 ± 0.07	-1.05 ± 0.04	128.49 ± 10.72	-18	-2.50 ± 0.37	13.18 ± 11.97

Table A.2: Global best-fit parameters of conditional luminosity functions

Galaxy Type	M_b^*	α_b	A_b	γ_b	M_f^*	α_f	A_f	γ_f	χ_r^2
All	-21.27 ± 0.10	-1.01 ± 0.06	0.12 ± 0.02	1.06 ± 0.02	-18	-1.63 ± 0.08	0.43 ± 0.07	0.72 ± 0.05	1.72
Red	-21.28 ± 0.09	-0.85 ± 0.06	0.08 ± 0.01	1.11 ± 0.02	-18	-1.80 ± 0.09	0.23 ± 0.04	0.69 ± 0.05	1.39
Blue	-21.61 ± 0.13	-1.50 ± 0.02	0.04 ± 0.01	0.82 ± 0.01					1.95

ple) and the local (orange) background estimators are compared in Figure A.1. In addition, we also estimate the conditional luminosity functions directly from group members as identified in Yang et al. (2007) using SDSS main sample with spectroscopic redshifts. Note that the SDSS main galaxy sample is not complete (missing $\sim 10\%$ of galaxies) due to fiber collisions, which limits the minimum angular distance between two galaxies ($\theta > 55''$) (Strauss et al., 2002). The results are shown by grey bands with Poisson errors. As can be seen, these three methods yield consistent results, indicating that our results are robust.

Table A.3: Best-fit parameters of red/blue conditional luminosity functions

$\log(M_{200}/\mathrm{M}_\odot)$	$\langle\log(M_{200}/\mathrm{M}_\odot)\rangle$	M_b^*	α_b	N_b	M_f^* (fixed)	α_f	N_f
[12.00, 12.34)	12.15	-34.02 ± 123483.60	-1.44 ± 0.35	0.00 ± 13.16	-18	-2.00 ± 0.12	0.19 ± 0.05
[12.34, 12.68)	12.49	-20.70 ± 0.59	-0.77 ± 0.44	0.45 ± 0.24	-18	-1.73 ± 0.24	0.53 ± 0.23
[12.68, 13.03)	12.84	-20.62 ± 0.41	-0.84 ± 0.30	0.93 ± 0.36	-18	-1.86 ± 0.16	1.16 ± 0.39
[13.03, 13.37)	13.17	-20.92 ± 0.28	-0.63 ± 0.25	2.37 ± 0.55	-18	-1.45 ± 0.27	2.47 ± 0.98
[13.37, 13.71)	13.51	-21.17 ± 0.12	-0.61 ± 0.11	4.40 ± 0.44	-18	-1.27 ± 0.87	1.72 ± 1.89
[13.71, 14.05)	13.85	-21.01 ± 0.13	-0.48 ± 0.14	10.94 ± 1.10	-18	-2.31 ± 0.41	2.56 ± 2.39
[14.05, 14.39)	14.21	-21.37 ± 0.10	-0.86 ± 0.08	20.92 ± 2.24	-18	-1.00 ± 0.60	15.24 ± 8.69
[14.39, 14.73)	14.53	-21.29 ± 0.08	-0.90 ± 0.06	46.65 ± 4.22	-18	-2.23 ± 0.43	8.87 ± 7.86
[14.73, 15.08)	14.87	-21.57 ± 0.10	-1.01 ± 0.06	100.65 ± 11.61	-18	-2.28 ± 0.59	17.91 ± 22.59
[12.00, 12.34)	12.15	-21.04 ± 0.46	-1.56 ± 0.06	0.08 ± 0.03			
[12.34, 12.68)	12.49	-21.28 ± 0.50	-1.59 ± 0.06	0.11 ± 0.05			
[12.68, 13.03)	12.84	-21.92 ± 0.67	-1.59 ± 0.08	0.13 ± 0.08			
[13.03, 13.37)	13.17	-21.64 ± 0.25	-1.45 ± 0.05	0.44 ± 0.12			
[13.37, 13.71)	13.51	-21.21 ± 0.25	-1.28 ± 0.08	1.50 ± 0.40			
[13.71, 14.05)	13.85	-21.61 ± 0.17	-1.40 ± 0.04	1.95 ± 0.34			
[14.05, 14.39)	14.21	-21.47 ± 0.23	-1.46 ± 0.06	3.38 ± 0.88			
[14.39, 14.73)	14.53	-21.81 ± 0.45	-1.48 ± 0.09	4.03 ± 1.99			
[14.73, 15.08)	14.87	-21.29 ± 0.24	-1.31 ± 0.09	21.00 ± 6.17			

A discrepancy is observed between results from using spectroscopic galaxies and the other two methods for low-mass groups, with the spectroscopic samples giving systematically lower amplitudes. Such difference is expected since the SDSS main galaxy sample is incomplete due to fiber collisions. The incompleteness will affect small groups more significantly than massive groups because a relative larger area within r_{200} is affected by fiber collisions in smaller halos. In addition, the conditional luminosity functions obtained from the global background estimation are slightly higher

than that from the local background estimation, especially for small halos. Such difference can be explained by the fact that the global background estimator based on random positions tends to underestimate the background level because galaxies live in large-scale structure. To reduce this effect, we use the local background estimator for halos with $M_{200} < 10^{13}\,M_{\odot}$, and use the global background estimator for more massive halos.

A.2.2 Using samples at different redshifts

For consistency check, we measure the conditional luminosity functions of satellite galaxies as a function of redshift. We apply K-correction and an evolution correction with $1.62 \times z$ for the magnitude (Blanton et al., 2003). The results are shown in Figure A.2. As an illustration, only results obtained from the local background estimator are plotted. The redshift increases from left to right and the halo mass increases from top to bottom. The color bands indicate the conditional luminosity functions at $z < 0.05$. The shapes of the luminosity functions of blue satellites from $z = 0.01$ to $z = 0.2$ are consistent with each other, while the number of bright red galaxies ($-22 < M_r < -19$) tends to decrease towards higher redshift. This indicates that the fraction of blue galaxies in groups increases toward higher redshift, which is consistent with the so-called Butcher-Oemler effect (Butcher & Oemler, 1978). The consistency between the shapes of the conditional luminosity functions at different

See the animation at `http://sdss.physics.nyu.edu/vagc/`

redshifts indicates that our measurements are not subject to severe systematics and cosmic variance.

A.3 The separation of red and blue galaxies

The color-magnitude distribution of galaxies associated with groups with $M_{200} > 10^{12}\,M_{\odot}$ are shown in Figure A.3 . The overall 2D distribution is shown in the left panel with the grey scale indicating the number density of galaxies and the color distribution for each magnitude bin is shown in the right panel. We apply double Gaussian functions to characterize the blue and red sequences, and the best-fit distributions are shown as the blue and red regions in the two panels. The green lines show the color-magnitude demarcation suggested by Baldry et al. (2004) based on SDSS spectroscopic data:

$$(u-r) = 2.06 - 0.244\,\tanh\left(\frac{M_r + 20.07}{1.09}\right). \tag{A.1}$$

As can be seen, the two galaxy populations are well separated by the relation proposed by Baldry et al. (2004).

To obtain the mean color-magnitude relation for each type of galaxies, we apply the same functional form as used in Baldry et al. (2004) to fit the centers of the best-fit Gaussian distributions. The best-fit color-magnitude relations are shown with the

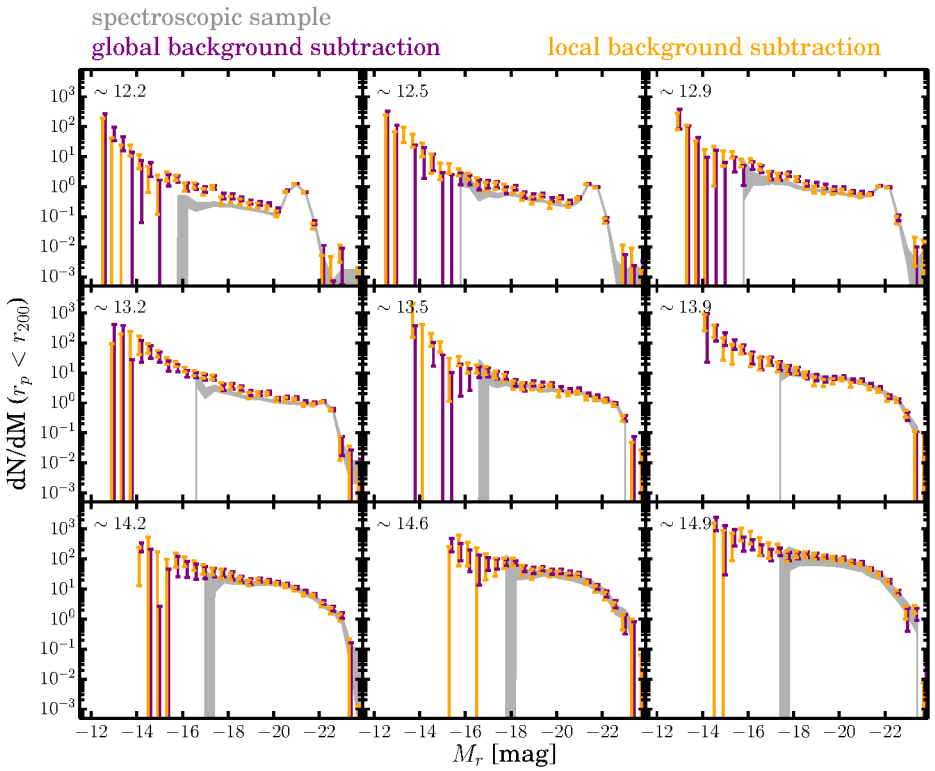

Figure A.1: Comparison between the conditional luminosity functions obtained with the global (purple) and local (orange) background estimators, and with the spectroscopic galaxy sample (grey).

red and blue dashed lines in the two panels, and the functions with parameters are

$$\text{red sequence}: u - r = 2.38 - 0.037\,(M_r + 20) - 0.108\,\tanh\left(\frac{M_r + 19.81}{0.96}\right); \quad \text{(A.2)}$$

$$\text{blue sequence}: u - r = 1.85 - 0.035\,(M_r + 20) - 0.363\,\tanh\left(\frac{M_r + 20.75}{1.12}\right). \quad \text{(A.3)}$$

As shown in the right panel, the mean relations describe the observed mean colors well. These mean color-magnitude relations are used to convert the conditional luminosity functions of red and blue galaxies into the corresponding conditional stellar mass functions in Section 2.6.1.

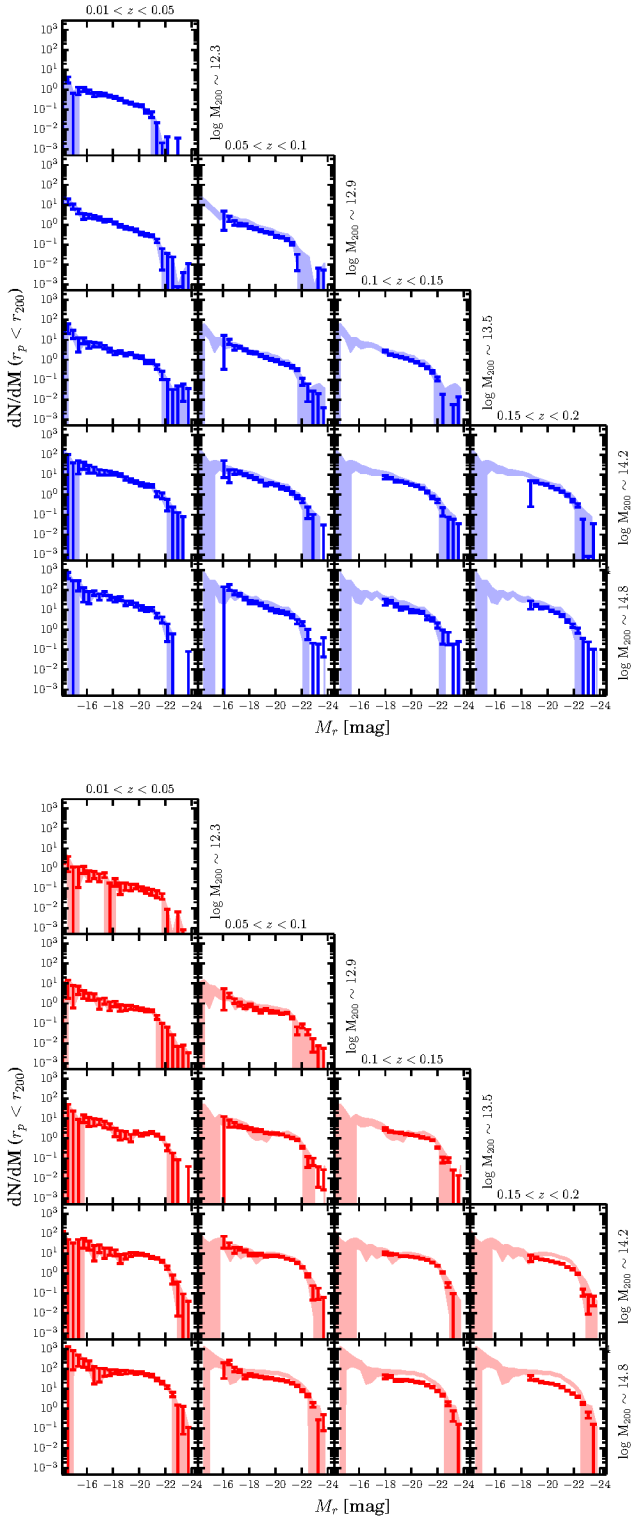

Figure A.2: Conditional luminosity functions of blue (top) and red (bottom) satellites as a function of redshift. The redshift increases from left to right columns and the halo mass increases from top to bottom. In all panels for $z > 0.05$, we include the luminosity functions at $z < 0.05$ in color bands for comparison. We only show luminosity functions for groups that are complete in the given redshift range.

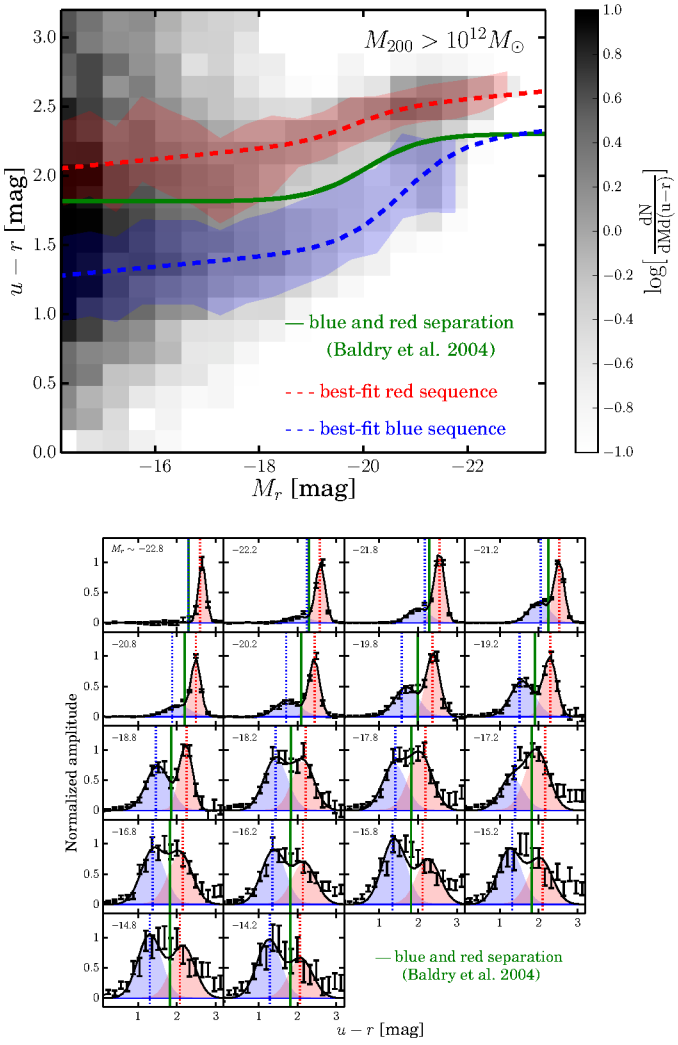

Figure A.3: Color-magnitude distribution of galaxies in halos with $M_{200} > 10^{12} M_\odot$. The overall 2D distribution is shown in the left panel with the grey scale indicating the number density of galaxies and the color distribution for each magnitude bin is shown in the right panel. Blue and red sequences described by the best-fit two Gaussian functions are indicated with the color regions and the dashed lines show the best-fit mean color-magnitude relations (Eq. A.2 and A.3). Two galaxy populations are well separated by the relation derived by Baldry et al. (2004) (Eq. A.1) indicated with the green lines.

Appendix B

Appendix for Chapter 3

B.1 Correction for blending effect

SDSS has relatively small average angular resolution (1.4″) which enables us to probe small impact parameter region. Nevertheless, the size of the angular resolution will affect the detectability of galaxies within certain scales ($< 4''$) due to the blending with quasars. To test this effect and correct for it, we estimate the surface density of galaxies around quasars. We select quasars with redshift higher than 1 to avoid the physical clustering between quasars and galaxies, and with i-band magnitudes brighter than 19.8 to have a similar magnitude limit of quasars with detected Mg II absorbers. We search galaxies around selected quasars and calculate the surface density of galaxies with a given color bin. Then, we normalize the surface densities by average surface densities calculated from large scale, $85''$ to $185''$.

The left panel of Figure B.1 shows the normalized surface densities of galaxies with three color bins around selected quasars. We find that the effect of blending extends to about $4''$, corresponding to 30 kpc at redshift 0.5. Not only that, the normalized surface densities depend on galaxy colors; red galaxies have less effect than blue galaxies. This effect could be due to the properties of galaxies within different color bins, especially the magnitude distribution and size. However, the detailed understanding of this color dependence is beyond the scope of this paper.

To correct for this blending effect, we introduce an empirical correction function, w, with a single gaussian as

$$w(\theta, C) = 1 + \frac{A}{C} \times e^{\frac{-\theta^2}{2\sigma^2}} \tag{B.1}$$

where C and θ are the $g-i$ color and angular distance between galaxies and quasars in arcsec. Given an angular distance and galaxy color, $w(\theta, C)$ represents the correct number count without the blending effect. We fit the inverse of the correction function, $1/w$, to normalized surface densities and the best-fit parameters yield to $A = 16.2 \pm 3.4$ and $\sigma = 1.65'' \pm 0.07$. The solid curves in Figure B.1 are the best-fit normalized surface densities from our formula. The width of the gaussian $\sigma \sim 1.65''$ is comparable with the size of the average angular resolution, suggesting that gaussian is a reasonable functional form to describe the blending effect.

To test the robustness of this measurement, we also randomly select about 37,500 stars from SDSS with i-band magnitude from 17 to 19.8 and apply the same analysis. The right panel of Figure B.1 shows the results from stars. The best-fit parameters

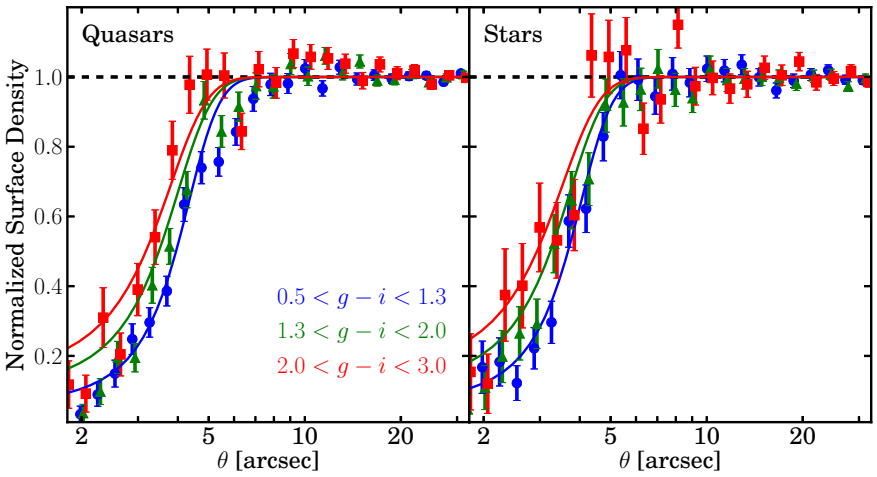

Figure B.1: The normalized surface density of galaxies around quasars and stars. *Left*: quasars. *Right*: stars. We estimate the surface densities of galaxies around quasars and stars and normalize them with the average surface densities at large scale (from 85″ to 185″). The shape can be well parametrized by the inverse of Eq. B.1 shown as the solid curves in the figure. We find that red galaxies have lesser blending effect than blue galaxies. It could be due the difference of physical properties of galaxies; nevertheless, the detailed understand is beyond the scope of this paper.

from stars are consistent with quasars within one sigma. Therefore, we confirm that the behavior is mainly driven by blending and use our parametrized correction function (Equation B.1) to correct the number count of galaxies around Mg II absorbers in Section 3.1.1, 3.1.2, and 3.1.4.

Appendix C

Appendix for Chapter 4

C.1 SDSS spectral calibration residuals

We investigate the nature of the imperfections of the spectral calibration of SDSS, as indicated by the residuals shown in Figure 4.2. To investigate the origin(s) of such features, we compare the residuals with the absorption lines caused by the atmosphere and the absorption features intrinsic to F stars that are used in the SDSS spectral calibration. In Figure C.1, we show the composite residual spectra of quasars (blue) at $E_{B-V} < 0.02$ mag within the whole wavelength range covered by SDSS. In addition, we also show a composite spectrum of the SDSS sky fibers in gray (M. Hart, private communication), i.e., fibers that were pointed to the fields where there is no detectable source in imaging, and the composite residual spectrum of luminous red galaxies by Yan (2011, red). In addition, we show a measurement of an atmospheric absorp-

tion spectrum (J. Gunn, private communication) in orange and an F-star composite spectrum from our data-driven stellar model in green.

We find that, beyond around 7000 Å, the composite residual spectra of quasars and galaxies are dominated by the residuals of atmospheric *emission* lines, mostly due to OH and H_2O. However, the atmospheric *absorption* lines, which were *not* included in the calibration, also imprint absorption features in the final residuals. The residuals of sky emission lines and absorption lines prevent us from investigating DIBs in the red end of the wavelength coverage.

At $\lambda < 7000$ Å, we find several features with no corresponding sky emission and absorption lines. To calibrate the flux of sources in each field, SDSS used standard F stars, selected based on photometric colors. The green line shows a typical F-star template normalized to unity using a running median filter. We see many of the residuals have corresponding features in the intrinsic F-star SED. Note this F-star template is a composite spectrum of F-stars observed in SDSS, so residuals due to atmospheric absorption lines are also evident.

These systematic residuals exist in all the SDSS spectra and extreme care must be taken while studying small spectral features, especially in the observer frame. In our analysis, we use sources at low-extinction regions as references and empirically remove these features.

`http://www.sdss3.org/dr8/algorithms/spectrophotometry.php`

C.2 DIB-dust correlations with different E_{B-V} estimation

We estimate the Galactic dust reddening of a star by comparing the observed g-r color of the star to the color of its reference star and convert into E_{B-V} by using the relation in Yuan et al. (2013). For individual lines of sight, the noise of E_{B-V} is dominated by the photometric error. To reduce the noise, we estimate the median E_{B-V} for each sky pixel and then combine it with the DIB map shown in Figure 4.6. Figure C.2 shows the inverse variance-weighted mean of DIB equivalent width as a function of the new reddening-based $\mathrm{E_{B-V}}$. The solid black lines are best-fit power laws (Table C.1) with new E_{B-V} measurements with $0.04 < E_{B-V} < 0.5$ mag and the dashed red lines are the best-fit power laws with SFD (Figure 4.9).

We find that the E_{B-V} from SFD tends to be overestimated which can be due to the contamination from background dust or/and systematics in the SFD map (e.g. Arce & Goodman, 1999). Nevertheless, the new derived correlations between DIB absorption strength and E_{B-V} are mostly consistent with the results based on the SFD map. In addition, the different behaviors of DIBs at high Galactic reddening, such as the decline of λ4885 and the flatten of λ5780, persist with the absorption-based E_{B-V} estimation, indicating that the different behaviors of DIBs at high Galactic reddening found in the study reflect the intrinsic correlations between DIBs and dust. It is also worth noting that the departure of power laws in the low E_{B-V} regions (< 0.05) due

to the extragalactic contamination in Figure 4.7 has largely reduced by using the reddening based E_{B-V}.

We note that the decline of some DIBs at high reddening becomes weaker with reddening based E_{B-V} due to the effects on the E_{B-V} from SFD mentioned above. However, in Figure C.3, we present the individual measurements of three DIBs with different trends and show that given a E_{B-V}, the equivalent width of DIBs correlates with the hydrogen content along the lines of sight. The color indicates the H_2 column densities along the lines of sight. For DIBλ4885 and DIBλ5780, the decline and flatten at high E_{B-V} are driven by lines of sight with high H_2 column densities. Such trends are found with both SFD and reddening based E_{B-V}, indicating that the skin effect is observed. On the other hand, using the reddening based E_{B-V}, the equivalent width of DIBλ4728 can be described by the best-fit power law from low E_{B-V} to highest E_{B-V} without noticeable deviation. The behaviors of DIBs at high E_{B-V} can be explained by the correlations between DIBs and hydrogen shown in Figure 4.14: DIBλ4885 is anti-correlated with molecular hydrogen, DIBλ5780 has no correlation with molecular hydrogen, and DIBλ4728 has a positive correlation with molecular hydrogen.

C.3 DIB-hydrogen correlations with different selections of the sample

We compare the α and μ estimation (Eq. 4.4) using two selections on the data to quantify how the results vary with the selection. The results are shown in Figure C.4. The left panel shows the result with the selection that we applied in the main analysis. The right panel shows the result with a set of stars which are more than 2 kpc away from the Sun. In addition, we also apply a latitude cut to avoid sky regions with high background contamination in N_{HI}. As can seen, the relative positions of DIBs on the α and μ plane derived from two selections are consistent with each other. In addition, we have repeated our analysis using DIBλ5780 as a N_{HI} estimate and find consistent results. Therefore, we conclude the results in the analysis are robust.

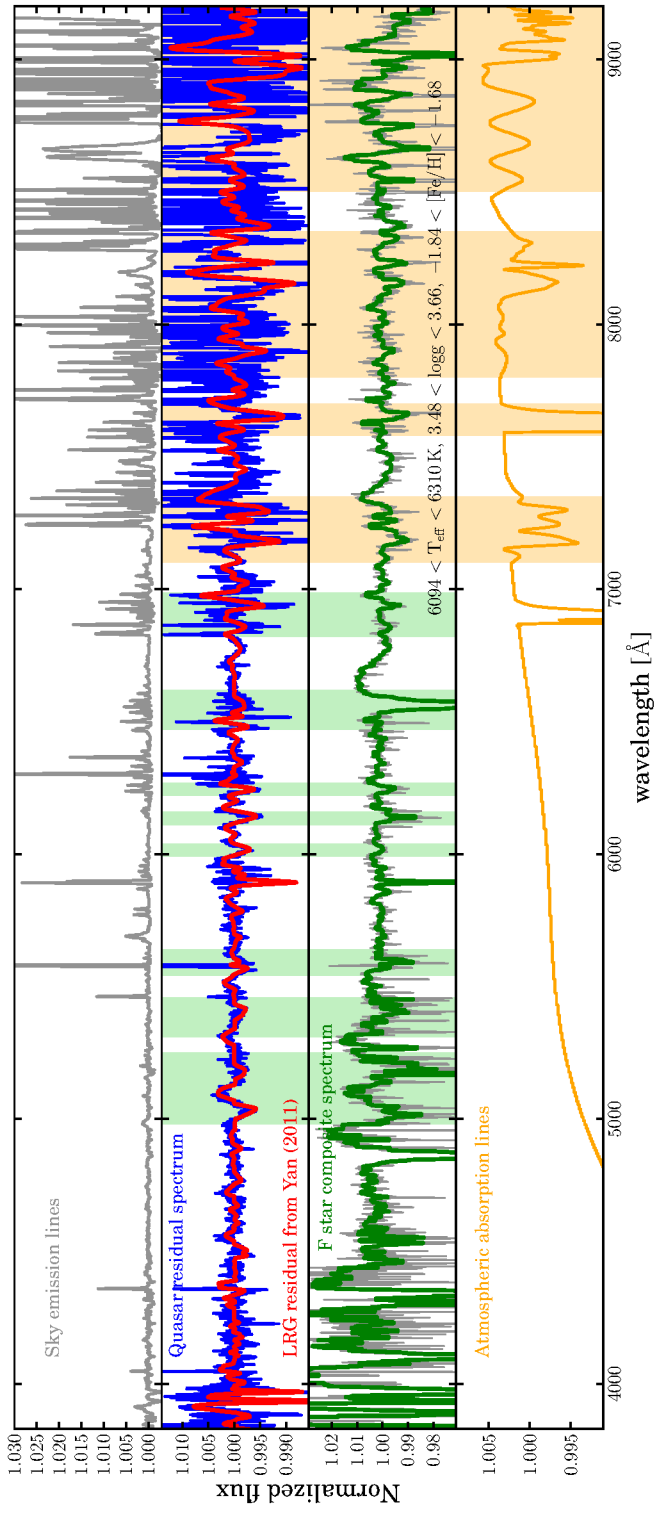

Figure C.1: Effects of atmospheric emission and absorption lines and the spectral calibration on the SDSS spectra. From top to bottom, we show the composite sky emission line spectrum (gray), the composite quasar residual spectrum (blue), the composite residual spectrum of luminous red galaxies from Yan (2011, red), a F-star composite spectrum (green), and an atmospheric absorption line spectrum (orange). We indicate features due to atmospheric absorption lines with orange vertical bands. At $\lambda < 7000$ Å, we find that many features correspond to the intrinsic stellar absorption features. Several distinct features are indicated with green vertical bands.

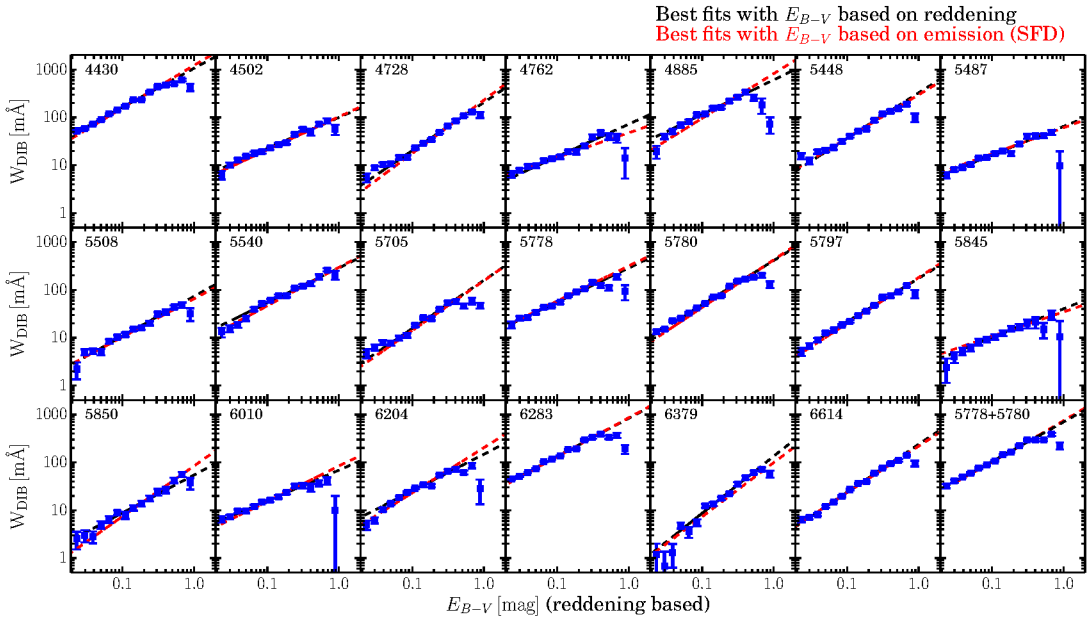

Figure C.2: Equivalent width measurements of 20 DIBs as a function of E_{B-V} based on reddening estimation. The blue data points show the inverse variance-weighted mean of DIB equivalent widths. The solid black lines are best-fit power laws derived with reddening-based E_{B-V} from 0.04 to 0.5 and the dashed black lines are extrapolation towards high and low E_{B-V}. The dashed red lines are the best-fit power laws derived with SFD E_{B-V}. The two E_{B-V} estimations yield consistent DIB-dust correlations.

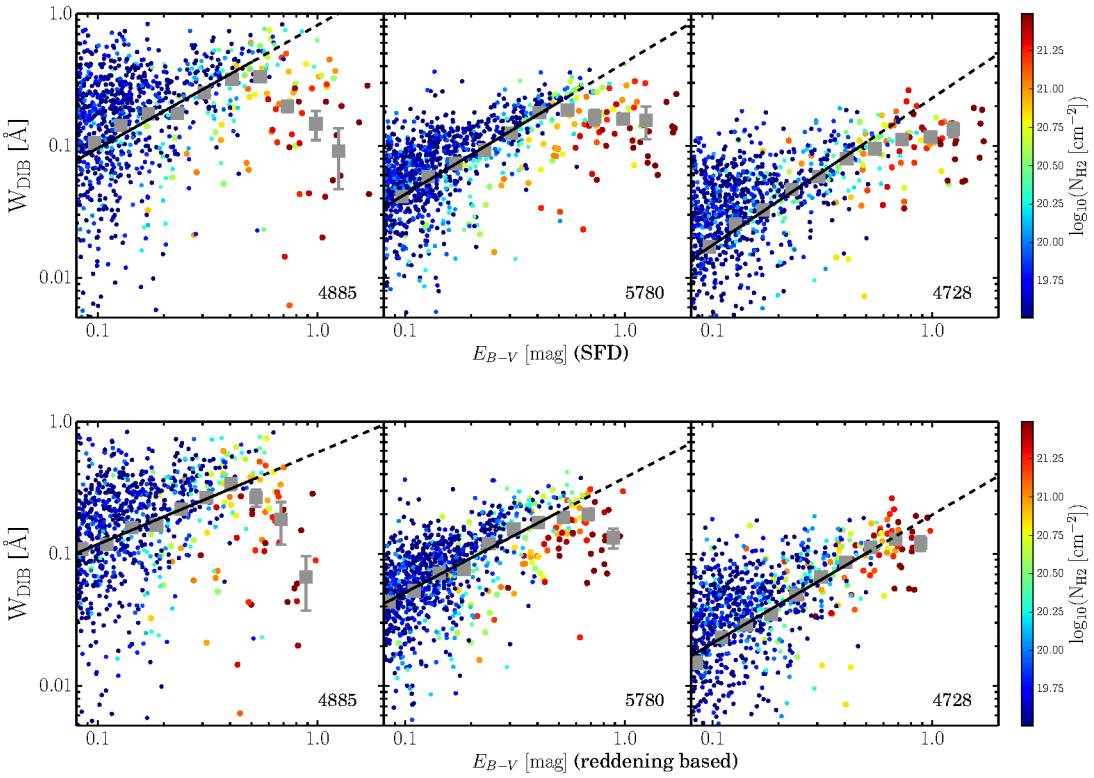

Figure C.3: Examples of individual DIB equivalent widths as a function of E_{B-V} with H_2 column densities shown in color. The grey data points and the black lines are the average values and the best-fit power laws shown in Figure 4.9 and Figure C.2. *Top*: E_{B-V} from SFD. *Bottom*: E_{B-V} based on reddening. The trends at high E_{B-V} are driven by lines of sight with high H_2 column densities, which can be explained by the correlations between hydrogen and DIBs shown in Figure 4.14.

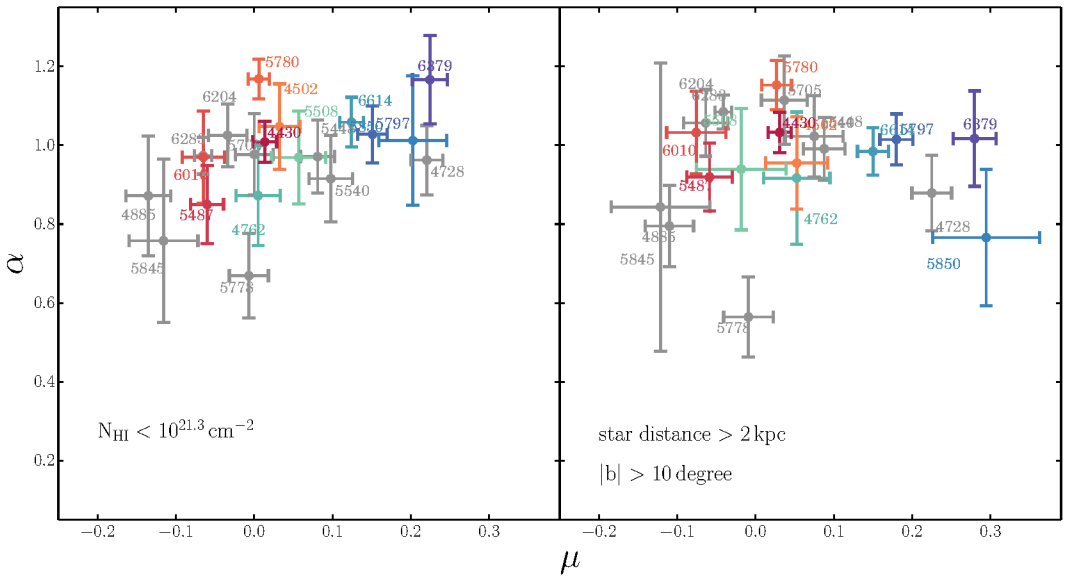

Figure C.4: Distributions of α and μ parameters derived from samples with two selections. Two samples are both selected with $E_{B-V} > 0.1$ mag and $W_{CO} > 0.0$ Kkm/s and the additional selections are shown in each panel. The left panel shows the same α and μ distribution as shown in Figure 4.12 and the right panel shows the results derived from stars which are two kpc away from the Sun and locate above the Galactic disk. The relative positions of DIBs on the α and μ plane based on two selections are consistent.

Table C.1: Best-fit parameters characterizing the relationships between DIBs and reddening-based E_{B-V}(Eq. 4.3)

λ [Å]	A	γ	W/A_V [Å/mag]
4430	1.06 ± 0.06	0.82 ± 0.03	0.42
4502	0.10 ± 0.01	0.65 ± 0.05	0.05
4728*	0.21 ± 0.02	1.01 ± 0.04	0.07
4762	0.07 ± 0.01	0.68 ± 0.05	0.03
4885*	0.61 ± 0.06	0.72 ± 0.05	0.27
5448*	0.32 ± 0.03	0.95 ± 0.04	0.11
5487	0.06 ± 0.01	0.60 ± 0.05	0.03
5508	0.07 ± 0.01	0.83 ± 0.06	0.03
5540*	0.29 ± 0.03	0.73 ± 0.04	0.13
5705*	0.16 ± 0.02	1.01 ± 0.05	0.05
5778*	0.29 ± 0.03	0.71 ± 0.04	0.13
5780	0.41 ± 0.02	0.93 ± 0.03	0.14
5797	0.17 ± 0.01	0.91 ± 0.03	0.06
5845*	0.04 ± 0.01	0.60 ± 0.09	0.02
5850	0.05 ± 0.01	0.79 ± 0.08	0.02
6010	0.07 ± 0.01	0.64 ± 0.05	0.03
6204*	0.15 ± 0.01	0.77 ± 0.04	0.06
6283*	0.84 ± 0.04	0.80 ± 0.02	0.34
6379	0.13 ± 0.01	1.19 ± 0.06	0.03
6614	0.24 ± 0.01	1.00 ± 0.03	0.08

*DIBs possibly blended with multiple weak DIBs

Bibliography

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

Abell, G. O. 1962, Problems of Extra-Galactic Research, 15, 213

Agulli, I., Aguerri, J. A. L., Sánchez-Janssen, R., et al. 2014, MNRAS, 444, L34

Aihara, H., Allende Prieto, C., An, D., et al. 2011, ApJS, 193, 29

Arce, H. G., & Goodman, A. A. 1999, ApJ, 512, L135

Angulo, R. E., Lacey, C. G., Baugh, C. M., & Frenk, C. S. 2009, MNRAS, 399, 983

Banados E., Hung L.W., de Propris R. & West M.J., 2010, ApJ, 721, L14

Barkhouse W.A., Yee H.K.C. & Lopez-Cruz O., 2007, ApJ, 671, 1471B

Baron, D., Poznanski, D., Watson, D., Yao, Y., & Prochaska, J. X. 2015, MNRAS, 447, 545

Baldry, I. K., Glazebrook, K., Brinkmann, J., et al. 2004, ApJ, 600, 681

BIBLIOGRAPHY

Baldry I.K., et al. 2012, MNRAS, 421, 621

Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D. 2003, ApJS, 149, 289

Bennett, C. L., Halpern, M., Hinshaw, G., et al. 2003, ApJS, 148, 1

Benson, A. J., Bower, R. G., Frenk, C. S., et al. 2003, ApJ, 599, 38

Bergeron, J. 1986, A&A, 155, L8

Bohlin, R. C., Savage, B. D., & Drake, J. F. 1978, ApJ, 224, 132

Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, ARA&A, 51, 207

Bordoloi, R., Lilly, S. J., Knobel, C., et al. 2011, ApJ, 743, 10

Bordoloi, R., Lilly, S. J., Hardmeier, E., et al. 2013, arXiv:1307.6553

Bouché, N., Hohensee, W., Vargas, R., et al. 2012, MNRAS, 426, 801

Boue, G., Adami, C., Durret, F., Mamon, G. A., Cayatte, V., 2008, A&A, 479, 335

Bowen, D. V., & Chelouche, D. 2011, ApJ, 727, 47

Blanton, M. R., Hogg, D. W., Bahcall, N. A., et al. 2003, ApJ, 592, 819

Blanton, M. R., Lupton, R. H., Schlegel, D. J., et al. 2005, ApJ, 631, 208

Blanton, M. R., Schlegel, D. J., Strauss, M. A., et al. 2005, AJ, 129, 2562

Blanton, M. R., & Roweis, S. 2007, AJ, 133, 734

BIBLIOGRAPHY

Blanton, M. R., & Moustakas, J. 2009, ARA&A, 47, 159

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

Butcher, H., & Oemler, A., Jr. 1978, ApJ, 219, 18

Campbell, E. K., Holz, M., Gerlich, D., & Maier, J. P. 2015, Nature, 523, 322

Cami, J., Sonnentrucker, P., Ehrenfreund, P., & Foing, B. H. 1997, A&A, 326, 822

Chabrier, G. 2003, PASP, 115, 763

Christlein, D., & Zabludoff, A. I. 2003, ApJ, 591, 764

Coil, A. L., Blanton, M. R., Burles, S. M., et al. 2011, ApJ, 741, 8

Cole, S., & Lacey, C. 1996, MNRAS, 281, 716

Cool, R. J., Moustakas, J., Blanton, M. R., et al. 2013, ApJ, 767, 118

Cooray, A., & Milosavljević, M. 2005, ApJ, 627, L89

Cooray, A. 2006, MNRAS, 365, 842

Cox, N. L. J., Kaper, L., Foing, B. H., & Ehrenfreund, P. 2005, A&A, 438, 187

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

Danks, A. C., Federman, S. R., & Lambert, D. L. 1984, A&A, 130, 62

De Lucia G., Kauffmann G., Springel V., White, S. D. M., Lanzoni B., Stoehr F., Tormen G., & Yoshida N. 2004, MNRAS, 348, 333

BIBLIOGRAPHY

de Propris, R., Pritchet, C. J., Harris, W. E., & McClure, R. D. 1995, ApJ,450, 534

Diemand, J., Kuhlen, M., & Madau, P. 2007, ApJ, 667, 859

Driver, S. P., Phillipps, S., Davies, J. I., Morgan, I., & Disney, M. J. 1994, MNRAS, 268, 393

Efstathiou, G., Ellis, R. S., & Peterson, B. A. 1988, MNRAS, 232, 431

Ehrenfreund, P., & Foing, B. H. 1996, A&A, 307, L25

Fischer, P., McKay, T. A., Sheldon, E., et al. 2000, AJ, 120, 1198

Foing, B. H., & Ehrenfreund, P. 1994, Nature, 369, 296

Friedman, S. D., York, D. G., McCall, B. J., et al. 2011, ApJ, 727, 33

Fukugita, M., & Peebles, P. J. E. 2004, ApJ, 616, 643

Galazutdinov, G. A., Musaev, F. A., Bondar, A. V., & Krełowski, J. 2003, MNRAS, 345, 365

Gao L., White S. D. M., Jenkins A., Stoehr F., Springel V., 2004, MNRAS, 355, 819

Geballe, T. R., Najarro, F., Figer, D. F., Schlegelmilch, B. W.,& de La Fuente, D. 2011, Nature, 479, 200

Giocoli C., Tormen G., van den Bosch F. C., 2008a, MNRAS, 386, 2135

BIBLIOGRAPHY

Hansen, S. M., Sheldon, E. S., Wechsler, R. H., & Koester, B. P. 2009, ApJ, 699, 1333

Harsono, D., De Propris, R.,2009, AJ, 137, 3091

Heger, M. L. 1922, Lick Observatory Bulletin, 10, 141

Herbig, G. H. 1993, ApJ, 407, 142

Herbig, G. H. 1995, ARA&A, 33, 19

Hobbs, L. M., York, D. G., Snow, T. P., et al. 2008, ApJ, 680, 1256

Hobbs, L. M., York, D. G., Thorburn, J. A., et al. 2009, ApJ, 705, 32

Holmberg, E. 1950, Meddelanden fran Lunds Astronomiska Observatorium Serie II, 128, 1

Hubble, E. 1936, ApJ, 84, 158

Jenkins L.P., Hornschemeier A.E., Mobasher B., Alexander D.M. & Bauer F.E., 2007, ApJ, 666, 846

Jenniskens, P., & Desert, F.-X. 1994, A&AS, 106, 39

Jiang, F., & van den Bosch, F. C., 2014, arXiv1403.6827

Johnston, R. 2011, A&A Rev., 19, 41

Josafatsson, K., & Snow, T. P. 1987, ApJ, 319, 436

Kacprzak, G. G., Churchill, C. W., Evans, J. L., Murphy, M. T., & Steidel, C. C. 2011, MNRAS, 416, 3118

Kacprzak, G. G., Churchill, C. W., & Nielsen, N. M. 2012, ApJ, 760, L7

Kalberla, P. M. W., Burton, W. B., Hartmann, D., et al. 2005, A&A, 440, 775

Kang, X., Jing, Y. P., Mo, H. J., Börner, G. 2005, ApJ, 631, 21

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

Kennicutt, R. C., Jr. 1998, ARA&A, 36, 189

Klypin, A., Gottlöber, S., Kravtsov, A. V., & Khokhlov, A. M. 1999, ApJ, 516, 530

Kochanek, C. S., Pahre, M. A., Falco, E. E., et al. 2001, ApJ, 560, 566

Koekemoer,A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36

Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18

Kos, J., Zwitter, T., Grebel, E. K., et al. 2013a, ApJ, 778, 86

Kos, J., & Zwitter, T. 2013b, ApJ, 774, 72

Kos, J., Zwitter, T., Wyse, R., et al. 2014, Science, 345, 791

Kravtsov, A., Berlind, A.A., Wechsler, R.H., Klypin, A., Gottlöber, S., Allgood, B., & Primack, J.R. 2004, ApJ, 609, 35

Kravtsov, A., Vikhlinin, A., & Meshscheryakov, A. 2014, arXiv:1401.7329

Krelowski, J., & Walker, G. A. H. 1987, ApJ, 312, 860

Krelowski, J., & Westerlund, B. E. 1988, A&A, 190, 339

Krełowski, J., Ehrenfreund, P., Foing, B. H., et al. 1999, A&A, 347, 235

Kroupa, P. 2001, MNRAS, 322, 231

Lan, T.-W., Ménard, B., & Zhu, G. 2014, ApJ, 795, 31

Leauthaud, A., Tinker, J., Bundy, K., et al. 2012, ApJ, 744, 159

Leauthaud, A., George, M. R., Behroozi, P. S., et al. 2012, ApJ, 746, 95

Lee, Y. S., Beers, T. C., Sivarani, T., et al. 2008, AJ, 136, 2022

Li, R., Mo, H. J., Fan, Z., Cacciato, M., van den Bosch, F. C., Yang, X., & More, S. 2009, MNRAS, 394, 1016

Li Y., Mo H. J., 2009, preprint (arXiv:0908.0301)

Liang, C. J., Kravtsov, A. V., & Agertz, O. 2015, arXiv:1507.07002

Liszt, H. 2014a, ApJ, 780, 10

Liszt, H. 2014b, ApJ, 783, 17

Loveday, J., Peterson, B. A., Efstathiou, G., & Maddox, S. J. 1992, ApJ, 390, 338

Loveday, J., Norberg, P., Baldry, I. K., et al. 2012, MNRAS, 420, 1239

Loveday, J., Norberg, P., Baldry, I. K., et al. 2015, MNRAS, 451, 1540

Lu Y., Mo H. J., 2007, MNRAS, 377, 617

Lu Z., Mo H.J., Lu Y., Katz N., Weinberg M. D., van den Bosch F. C. & Yang X., 2014, MNRAS, 439, 1294

Lu Z., Mo H.J., Lu Y., Katz N., Weinberg M. D., van den Bosch F. C. & Yang X., 2015, MNRAS, 450, 1604

S. R. Majewski et al., The Apache Point Observatory Galactic Evolution Experiment (APOGEE), Astronomical Journal, in preparation

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

Megier, A., Krełowski, J., & Weselak, T. 2005, MNRAS, 358, 563

Meisner, A. M., & Finkbeiner, D. P. 2014, ApJ, 781, 5

Ménard, B., & Chelouche, D. 2009, MNRAS, 393, 808

Ménard, B., Scranton, R., Fukugita, M., & Richards, G. 2010, MNRAS, 405, 1025

Ménard, B., Wild, V., Nestor, D., et al. 2011, MNRAS, 417, 801

Ménard, B., & Fukugita, M. 2012, ApJ, 754, 116

Ménard, B., Scranton, R., Schmidt, S., et al. 2013, arXiv:1303.4722

Merrill, P. W. 1934, PASP, 46, 206

Merrill, P. W., Sanford, R. F., Wilson, O. C., & Burwell, C. G. 1937, ApJ, 86, 274

Moutou, C., Krełowski, J., D'Hendecourt, L., & Jamroszczak, J. 1999, A&A, 351, 680

McNaught-Roberts, T., Norberg, P., Baugh, C., et al. 2014, MNRAS, 445, 2125

Milne M.L., Pritchet C.J., Poole G.B., Gwyn S.D. J., Kavelaars J.J., Harris W.E. & Hanes D.A., 2007, AJ, 133, 177

Mo H. J., Mao S., 2002, MNRAS, 333, 768

Mo H. J., Mao S., 2004, MNRAS, 353, 829

Mo H. J., van den Bosch F. C., White S. D. M., 2010, Galaxy Formation and Evolution, Cambridge University Press

Moore, B., Ghigna, S., Governato, F., Lake, G., Quinn, T., Stadel, J., & Tozzi, P. 1999, ApJ, 524, L19

Moorman, C. M., Vogeley, M. S., Hoyle, F., et al. 2015, ApJ, 810, 108

Moustakas, J., Coil, A. L., Aird, J., et al. 2013, ApJ, 767, 50

Morrissey, P., Conrow, T., Barlow, T. A., et al. 2007, ApJS, 173, 682

Moretti, A., Bettoni, D., Poggianti, B. M., et al. 2015, A&A, 581, A11

Mostek, N., Coil, A. L., Moustakas, J., Salim, S., & Weiner, B. J. 2012, ApJ, 746, 124

Murray, S. G., Power, C., & Robotham, A. S. G. 2013, Astronomy and Computing, 3, 23

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1995, MNRAS, 275, 56

Nielsen, N. M., Churchill, C. W., Kacprzak, G. G., & Murphy, M. T. 2013a, ApJ, 776, 114

Nielsen, N. M., Churchill, C. W., & Kacprzak, G. G. 2013b, ApJ, 776, 115

Norberg, P., Cole, S., Baugh, C. M., et al. 2002, MNRAS, 336, 907

Pâris, I., Petitjean, P., Aubourg, É., et al. 2012, A&A, 548, A66

Peek, J., Ménard, B., Corrales, L., in prep.

Peng, Y.-j., Lilly, S. J., Kovač, K., et al. 2010, ApJ, 721, 193

Penzias, A. A., & Wilson, R. W. 1965, ApJ, 142, 419

Popesso, P., Biviano, A., Böhringer, H., & Romaniello, M. 2006, A&A, 445, 29

Porceddu, I., Benvenuti, P., & Krelowski, J. 1991, A&A, 248, 188

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2013, arXiv:1303.5073

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2015, arXiv:1502.01589

Press, W. H., & Schechter, P. 1974, ApJ, 187, 425

Rahman, M., Mendez, A. J., Ménard, B., et al. 2015, arXiv:1512.03057

Rao, S. M., Turnshek, D. A., & Nestor, D. B. 2006, ApJ, 636, 610

Re Fiorentin, P., Bailer-Jones, C. A. L., Lee, Y. S., et al. 2007, A&A, 467, 1373

Riess, A. G.,Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009

Rines, K., & Geller, M. J.,2008, AJ, 135, 1837

Rodriguez, F., Merchán, M., & Sgró, M. A. 2015, A&A, 580, A86

Rubin, K. H. R., Prochaska, J. X., Koo, D. C., et al. 2013, arXiv:1307.1476

Salama, F., Bakes, E. L. O., Allamandola, L. J., & Tielens, A. G. G. M. 1996, ApJ, 458, 621

Sarre, P. J., Miles, J. R., Kerr, T. H., et al. 1995, MNRAS, 277, L41

Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521

Schechter, P. 1976, ApJ, 203, 297

Schneider, D. P., Richards, G. T., Hall, P. B., et al. 2010, AJ, 139, 2360

Schneider, D. P., Richards, G. T., Hall, P. B., et al. 2010, AJ, 139, 2360

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Scranton, R., Connolly, A. J., Nichol, R. C., et al. 2003, arXiv:astro-ph/0307335

Sheth R. K., & Tormen G. 1999, MNRAS, 308, 119

Sheth R. K., Mo H. J., & Tormen G. 2001, MNRAS, 323, 1

Smith, J. D. T., Draine, B. T., Dale, D. A., et al. 2007, ApJ, 656, 770

Smolinski, J. P., Lee, Y. S., Beers, T. C., et al. 2011, AJ, 141, 89

Snow, T. P., Jr., & Cohen, J. G. 1974, ApJ, 194, 313

Somerville, R. S., & Primack, J. R. 1999, MNRAS, 310, 1087

Steidel, C. C., Dickinson, M., & Persson, S. E. 1994, ApJ, 437, L75

Steidel, C. C., Erb, D. K., Shapley, A. E., et al. 2010, ApJ, 717, 289

Steinmetz, M., Zwitter, T., Siebert, A., et al. 2006, AJ, 132, 1645

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

Tal, T., Dekel, A., Oesch, P., et al. 2014, ApJ, 789, 164

Thorburn, J. A., Hobbs, L. M., McCall, B. J., et al. 2003, ApJ, 584, 339

Trayford, J. W., Theuns, T., Bower, R. G., et al. 2015, MNRAS, 452, 2879

Tumlinson, J., Thom, C., Werk, J. K., et al. 2011, Science, 334, 948

Vale, A., & Ostriker, J. P. 2004, MNRAS, 353, 189

van den Bosch, F. C., Yang, X., & Mo, H. J. 2003, MNRAS, 340, 771

van den Bosch F. C., Tormen G., Giocoli C., 2005, MNRAS, 359, 1029

van Loon, J. T., Bailey, M., Tatton, B. L., et al. 2013, A&A, 550, A108

Vogelsberger, M., Genel, S., Springel, V., et al. 2014, MNRAS, 444, 1518

Vos, D. A. I., Cox, N. L. J., Kaper, L., Spaans, M., & Ehrenfreund, P. 2011, A&A, 533, A129

Walker, G. A. H., Bohlender, D. A., Maier, J. P., & Campbell, E. K. 2015, ApJ, 812, L8

Wegner G., 2011, MNRAS, 413, 1333

Weiner, B. J., Coil, A. L., Prochaska, J. X., et al. 2009, ApJ, 692, 187

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

Welty, D. E. 2014, IAU Symposium, 297, 153

Werk, J. K., Prochaska, J. X., Tumlinson, J., et al. 2014, arXiv:1403.0947

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

Wilson, G., Smail, I., Ellis, R. S., & Couch, W. J. 1997, MNRAS, 284, 915

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868

Wszołek, B., & Godłowski, W. 2003, MNRAS, 338, 990

Yahata, K., Yonehara, A., Suto, Y., et al. 2007, PASJ, 59, 205

Yan, R. 2011, AJ, 142, 153

Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, AJ, 137, 4377

Yang X., Mo H. J., van den Bosch F. C., 2003, MNRAS, 339, 1057

Yang X., Mo H. J., van den Bosch F. C., Jing Y. P.,2005, MNRAS, 356, 1293

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

Yang, X., Mo, H. J., & van den Bosch, F. C. 2009, ApJ, 695, 900

Yang, X., Mo, H. J., Zhang, Y. C., & van den Bosch, F. C. 2011, ApJ, 741, 13

Yang, X., Mo, H. J., van den Bosch, F. C., Zhang, Y., Han, J.,2012, ApJ, 752, 41

Yuan, H. B., & Liu, X. W. 2012, MNRAS, 425, 1763

Yuan, H. B., Liu, X. W., & Xiang, M. S. 2013, MNRAS, 430, 2188

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

Zasowski, G., Ménard, B., Bizyaev, D., et al. 2015, ApJ, 798, 35

Zhu, G., Blanton, M. R., & Moustakas, J. 2010, ApJ, 722, 491

Zhu, G., & Ménard, B. 2013, ApJ, 770, 130

Zhu, G., & Ménard, B. 2013, ApJ, 773, 16

Zhu, G., Ménard, B., Bizyaev, D., et al. 2014, MNRAS, 377

Zibetti, S., Ménard, B., Nestor, D. B., et al. 2007, ApJ, 658, 161

Vita

Ting-Wen Lan received his Bachelor degree in physics at National Taiwan University in 2009, and joined the Ph.D. program in the physics and astronomy department at the Johns Hopkins University in 2011.

At JHU, he has been working with Professor Brice Ménard on various projects from exploring galaxy properties in dark matter halos, studying the connections between gas and galaxies, to probing mysterious molecules in the Milky Way. After receiving his doctoral degree at JHU, he will continue to explore the Universe as a Kavli fellow at the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU) in Tokyo.