October 2019

Galaxies and Gas in Dark Matter Halos and the Cosmic Web

Seunghwan Lim

GALAXIES AND GAS IN DARK MATTER HALOS AND THE COSMIC WEB

A Dissertation Presented

by

SEUNGHWAN LIM

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2019

Astronomy

GALAXIES AND GAS IN DARK MATTER HALOS AND THE COSMIC WEB

A Dissertation Presented

by

SEUNGHWAN LIM

Approved as to style and content by:

Houjun Mo, Chair

Daniela Calzetti, Member

Alexandra Pope, Member

Lorenzo Sorbo, Member

Daniela Calzetti, Department Chair
Astronomy

DEDICATION

To my parents Sang-Hyuk and Soo Ahn;

my wife Hyeji;

and my sons, K Jisoon, Jeremiah Jian and David Jion.

Success is going from failure to failure without loss of enthusiasm.

If you are going through hell, keep going.

- Winston Churchill

ACKNOWLEDGMENTS

I simply thank every one that I encountered and every thing that I experienced or occured to me during the six years of my PhD, my marriage, my first living in a foreign country, and my raising three kids that began and proceeded all at the same period. It wasn't always very beautiful, but gave me lessons and kept me growing to become a more mature person. I, in particular, am grateful for the environment of the department and the academic field that has the highest density of wise and intelligent people that I ever experienced in my life, from whom and which I learned a lot of lessons and insights.

I would like to thank Houjun Mo, my advisor for the second year project and thesis. I learned a lot from him in many different aspects. He is the smartest person that I ever had a long-term relationship with. I particularly learned a lot from the way that he thinks, and how his approach differs for different matters. I feel truly fortunate that I met and worked with him during this critical period of my life.

I thank Daniela Calzetti, Houjun Mo, Alexandra Pope, and Lorenzo Sorbo, for having accepted to be on my thesis committee and having provided helpful suggestions and comments. I also thank Jounghun Lee, my advisor for master thesis at the Seoul National University, who suggested and enabled my coming to U.S. from the very beginning.

I am grateful that I also have had many chances to interact or collaborate with the scientists outside the UMass, including Huiyuan Wang, Xiaohu Yang, Mark Vogelsberger, Eiichiro Komatsu, Klaus Dolag, David Barnes, Colin Hill, and many others.

I also thank every one that helped me through some difficult times that I had, including my parents, Eiichiro Komatsu, Houjun Mo, Dooseok Jung, Haesoon Choi,

Hansung Gim, Grant Wilson, Rob Gutermuth, and many others that I just cannot list all because of the limited space, for their time to listen to me, care about me, or provide useful advices. Wisdom and advice from Eiichiro Komatsu were particularly helpful and I am still heavily relying on the conversation that we had at a biergarten after the Sunday baseball whenever the same issue repeats. I truly appreciate it. I also miss so much the baseball friends.

I extend a thank you to those who acknowledged the works included in this dissertation. I thank Eiichiro Komatsu, Houjun Mo, and Mark Vogelsberger for writing recommendation letters for me. I enjoyed talking with the scientists all over the world during the interviews conducted for the fellowships and positions that I was shorlisted for. I appreciate people at the Kavli IPMU, University of Zurich, and the Canadian Institute for Theoretical Astrophysics and University of British Columbia, for having offered me a great opportunity.

Lastly, I thank my family and friends that were not included above, for their help and support.

Without any one of the help and people mentioned so far, it might have been not possible to end up in this dissertation.

(As a father of three kids including twins, all born in the middle of my PhD, and also as the only one in the household that speaks their language living in a foreign country, I wish a good luck for many other fathers and husbands in a similar situation and ask for warm words as well as practical helps to them from the society and community, which are often times, if not always, found to be lacking. Just showing dads and husbands one tenth of the interest that you show to mothers and wives would bring them back the smile they lost. So please do so when you find one around you, because I believe dads lives also matter.)

ABSTRACT

GALAXIES AND GAS IN DARK MATTER HALOS AND THE COSMIC WEB

SEPTEMBER 2019

SEUNGHWAN LIM

B.Sc., POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY

M.Sc., SEOUL NATIONAL UNIVERSITY

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Houjun Mo

In the current understanding of galaxy formation, galaxies are believed to form and evolve in dark matter halos. The dark matter halos are collapsed objects that form and grow via gravitational instability of small initial density fluctuation in the cosmic field. They are not only the hosts of galaxies but are tracers of the cosmic web of the Universe. They are thus crucial components for understanding how galaxies form and evolve within the cosmic web. This dissertation is a systematic investigation of the galaxies and gas in the dark matter halos and the cosmic web, using observation data of large galaxy surveys such as the Sloan Digital Sky Survey (SDSS), and of the Cosmic Microwave Background (CMB) survey such as the *Planck*, together with simulations and modellings for comparison and interpretation. Specifically, we have identified dark matter halos in the low-redshift Universe, and constructed the largest

to date all-sky group catalog. We also investigated correlations between many of the galaxy and halo properties, particularly finding, for the first time, an observational proxy of halo age. Then, I developed a series of novel approaches to maximize the detection of the Sunyaev-Zel'dovich effect (SZE), and from it explored the gas properties in halos to find that the gas mass fraction even in Milky Way-size halos is about the cosmic mean fraction. The analysis identified the baryons in a warm-hot medium on halo scales. We also show that the thermal SZE can be used to constrain the mean relationship between thermal energy of IGM gas and local total matter density for the first time. We support the reliability of our methods and results with tests where the methods are applied to the mock CMB maps constructed from simulations. Finally, we present comparisons with simulations of the SZE, and discuss the implications for its constraining power of galaxy formation models.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

PROLOGUE

Since galaxies form and evolve in and with dark matter halos and the cosmic web, it is required, not optional, to understand galaxy evolution within the context of the large-scale environments. Galaxies with the same properties but in different environments follow totally different paths of evolution. Saying anything about galaxies but not about their environments is almost, if not completely, meaningless. The key to understanding of such environmental effects is dark matter halos. They are not only the hosts of galaxies but are tracers of the large-scale structure. They naturally separate environments into small scales (within halos) and large scales (filaments, sheets, etc), making them also observationally important.

The small initial fluctuation in the cosmic density field grows by gravitational instability in the cosmic web and collapses to form objects called dark matter halos. While dark matters collapse and accrete, baryons, the 'ordinary' matter, also fall together into the potential well of dark matter halos. The shock created by the infall of matter heats up the baryons, which were initially 'cold gas', to form the hot gaseous halo. Then the radiation via recombinations and line transitions of the gas cools down the gas and condense it into disky structures and stars, thus galaxies. The feedback from stars such as stellar winds and supernova explosion (SNe), and from active galactic nuclei at the center of galaxies, however, introduce a tweak in galaxy evolution, by heating or ejecting the gas into surroundings, regulating further star formation. The accurate mechanisms and impact of the feedback on galaxies and environment are the key question yet to be answered. Another component dictating the galaxy evolution path is environmental regulation of star formation such as tidal stripping (stripping via interaction between galaxies and potential well of dark matter

halos; e.g. Merritt 1984), ram pressure stripping (stripping via interaction between galaxies and gas in dark matter halos; e.g. Gunn & Gott 1972), and galaxy harassment (quenching via quick and frequent, weak encounters between galaxies; e.g. Moore et al. 1996). The merger between galaxies and between halos also plays an important role in galaxy evolution and growth of the cosmic structures.

This dissertation is a coherent, comprehensive investigation of the galaxy evolution and gas properties in dark matter halos and the cosmic web. Here, using well-defined dark matter halos from observation, we examine correlations between the properties of galaxy, halo, and environment, and its implications about galaxy evolution from high-z to the present day. We use the dark matter halos as a tracer of the large-scale structures to define environment via the reconstruction of density field, velocity field, and tidal field. Then we construct the largest to date all-sky catalog of groups and associated dark matter halos in the low-redshift Universe, by developing an improved halo-based group finder and applying that to four large redshift galaxy surveys. The catalog provides an accurate association of galaxies with dark matter halos as well as an unbiased estimate of halo properties such as mass. Using the updated group catalogs, we cross-correlate the galaxy systems with the Sunyaev-Zel'dovich effect (SZE) from the temperature maps of the *Planck* CMB observation, to infer the properties of gas in halos. We also use the large-scale density field to infer the gas properties in the cosmic web from the SZE. Our results clearly demonstrate that the SZE is a promising tool to probe the gas properties and constrain galaxy formation models.

Because galaxy formation and evolution are governed by a number of physical processes, each of which becomes effective at different characteristic scales, it is essential to have a well-defined group catalog that covers a wide span of mass range. It is also very important to have an accurate mass estimate of those halos since many of the galaxy properties depend heavily on the halo mass. In Chapter 1, we improve the halo-based group finder developed by Yang et al. (2005, 2007) by using a better mass

proxy and by extending to about an order of magnitude lower halo mass. From tests using realistic mock catalogs, we show that our group finder, on average, identifies $\sim 94\%$ of the true member galaxies with an unbiased halo mass estimate with a typical uncertainty of ~ 0.2 dex. We apply our group finder to four large redshift galaxy surveys, the 2MRS, 6dF, SDSS, and 2dF, to construct and publicize the largest to date all-sky group catalogs of the low-redshift Universe.

Most of the previous studies with attempt of linking galaxies to dark matter halos relied only on using the halo mass while many studies have suggested that galaxy properties are also determined by the other halo properties than mass. Among those halo properties, it is well known that the halo age affects the galaxy properties significantly. In Chapter 2, we propose the ratio between mass of central galaxy and host halo, $f_c \equiv M_{*,cen}/M_h$, as an observational proxy of halo age. The motiviation behind the idea is from simulations where the ratio increases with time (e.g. Wang et al. 2011) because the mass in halos is gravitationally attracted to and thus gradually accreted onto central galaxies. Using the group catalogs of Yang et al. (2007), we investigate the correlations between the galaxy properties and f_c. We find that galaxies with higher f_c at same stellar mass are found to be redder, more quenched, and smaller in size, except for $f_c < 0.02$ where the trend is reversed because of the down-sizing effect. These results clearly demonstrate the validity of f_c as the observational proxy of halo age.

Another way of studying galaxy evolution is to investigate the galaxy number statistics. Specifically, the number of galaxies with given stellar mass, called the stellar mass function or SMF, are one of the simplest measure from observation that can constrain galaxy formation models. In Chapter 3, we use the conditional stellar mass function (CSMF) of galaxy systems, measured by Lan et al. (2016), to test galaxy formation models and compare their predictions about the star formation history in halos of different mass. We show that the CSMF provides additional constraints

compared to the total SMF, and that the observation data clearly prefer a model with a characteristic redshift at which star formation in low-mass halos changes behavior.

Yet another non-mass property that may greatly affect galaxy evolution is the flow of mass accretion along the cosmic web, which may be reflected in the alignment of galaxies in position, shape, and angular momentum with the large-scale structures. Previous studies that investigated these alignments often times reported conflicting findings (e.g. Holmberg 1969, Zaritsky et al. 1997, Yang et al. 2006), mainly due to the small size of samples. In Chapter 4, using the group catalog of Yang et al. (2007) and the tidal field reconstructed by Wang et al. (2012) within the volume spanned by the SDSS Data Release 7, we examine various alignments between galaxies and the large-scale structures. Our results imply that galaxy formation occurs via the coherent, anisotropic flow of mass accretion along the cosmic web, and that the large-scale tidal field is a good tracer of the cosmic web.

While so far we have focused on stars, investigating gas is a more direct way to study the evolution of galaxies and the cosmic web, and their interaction. With the advent of large CMB surveys, the Sunyaev-Zel'dovich effect (SZE) provides a new, promising avenue to probe the gas properties. The SZE refers to the change in the CMB spectrum due to the scattering of the CMB photons with free electrons that they encounter. It has, in particular, advantages over other popular tools used to study the gas, including the quasar absorption line studies and X-ray observations. In Chapter 5, we use the thermal SZE, which measures the integrated gas pressure along line-of-sight (LOS), from the *Planck* CMB survey to infer the thermal energy of gas in halos of different mass. To overcome the large beam size of the *Planck* as well as the low signal-to-noise of the SZE, we employ the matched filter (MF) to extract the signal. Our analysis also takes into account the projection effects of halos along same LOS. Our results indicate that the thermal energy of halo gas does not scale with halo mass in a self-similar way, but is found to be much lower in low-

mass halos, instead. We then, in Chapter 6, cross-correlate the *Planck* temperature maps and our group catalog to extract the kinematic SZE (kSZE) signals from halos. From the integrated flux, we infer the gas mass fraction within halos of different mass. We find that almost all baryons expected from cosmology are detected, but the gas temperature is lower than the virial temperature when combined with the tSZE results. We continue to probe the gas in Chapter 7, but now the gas outside halos, by extending a similar method used in Chapter 5 and 6, using the reconstructed density and tidal field by Wang et al. (2012, 2016). We constrain the thermal energy of intergalactic medium (IGM) in different environments. We find that the IGM in regions of higher tidal field has a higher thermal energy at a given density. This may be owing to higher cumulative energy provided via feedback due to the earlier formation of structures in those regions. In Chapter 8, we compare our results with the predictions from simulations, to demonstrate that the SZE is a promising tool to provide critical constraints on galaxy formation models.

PART I: GALAXIES IN THE COSMIC WEB

CHAPTER 1

GALAXY GROUPS IN THE LOW-REDSHIFT UNIVERSE[1]

We apply a halo-based group finder to four large redshift surveys, the 2MRS, 6dFGS, SDSS and 2dFGRS, to construct group catalogs in the low-redshift Universe. The group finder is based on that of Yang et al. but with an improved halo mass assignment so that it can be applied uniformly to various redshift surveys of galaxies. Halo masses are assigned to groups according to proxies based on the stellar mass/luminosity of member galaxies. The performances of the group finder in grouping galaxies according to common halos and in halo mass assignments are tested using realistic mock samples constructed from hydrodynamical simulations and empirical models of galaxy occupation in dark matter halos. Our group finder finds $\sim 94\%$ of the correct true member galaxies for $90-95\%$ of the groups in the mock samples; the halo masses assigned by the group finder are un-biased with respect to the true halo masses, and have a typical uncertainty of $\sim 0.2\,\mathrm{dex}$. The properties of group catalogs constructed from the observational samples are described and compared with other similar catalogs in the literature.

1.1 Introduction

Grouping galaxies observed in a galaxy catalog into systems (clusters and groups) is a practice of long history. In the early attempts, clusters of galaxies were identified

[1] THE CONTENTS OF THIS CHAPTER ARE PUBLISHED IN LIM ET AL. 2017, MNRAS, 470, 2982.

based on optical photometric data, using the local density contrast of galaxies in the sky as a proxy of spatial density and using distance estimates that are based on galaxy magnitudes. For example, Abell (1958) constructed a catalog of about 2,700 clusters from the POSS plates using local galaxy surface number densities. A similar selection was used by Abell et al. (1989) to construct a catalog of 1,600 clusters from the UKST plates. Zwicky et al. (1961-1968) identified 9, 133 clusters in the northern celestial hemisphere using the POSS plates, and adopting a galaxy number density criterion that is relative to the immediate neighborhood. Because these catalogs are constructed from photographic plates and no redshift information is available for individual galaxies, they suffer severely in in-homogeneity, incompleteness, and projection effects.

With the advent of large redshift surveys in 1980s, a lot of efforts were made to select galaxy clusters/groups on the basis of closeness of galaxies in redshift space. Although differing in details, many of these investigations have adopted the so-called friends-of-friends (FoF) method, which identifies galaxy systems as member galaxies that are linked by some adopted linkage criteria. For example, Postman & Geller (1984) identified galaxy groups from the CfA redshift survey (Huchra et al. 1983) by applying the FoF algorithm, developed by Huchra & Geller (1982), which uses two linking criteria, one on projected separation and the other on redshift difference, to link galaxies. With modifications, the FoF algorithm has been applied to various redshifts surveys of galaxies, including the Two Degree Field Galaxy Redshift Survey (2dFGRS; e.g. Eke et al. 2004), the Two Micron All Sky Redshift Survey (2MRS; e.g. Crook et al. 2007), and the Sloan Digital Sky Survey (SDSS; e.g. Goto 2005, Berlind et al. 2006). Lavaux & Hudson (2011) applied the FoF group finder to their own compilation combining the 2MRS, SDSS and Six Degree Field Galaxy Survey (6dFGS).

As high density regions in the galaxy distribution, clusters and groups of galaxies have been widely used to study the environmental dependence of the galaxy population and its evolution. For example, Dressler (1980) found that the morphology of a galaxy is correlated with the local density of galaxies in that the fraction of elliptical galaxies is higher in regions of higher density. Butcher & Oemler (1978, 1984) studied the galaxy populations in rich and compact clusters at redshifts of ~ 0.4 and found that the ratio of blue galaxies is higher than that in nearby clusters of similar richness and morphology, implying a strong recent evolution in galaxy color. Galaxy systems have also been assumed to be associated with dark matter halos. In the 1930s, Zwicky studied the motion of galaxies within the Coma Cluster and found that the total mass of the cluster estimated using the virial theorem is more than 100 times higher than that estimated from the total luminosity of member galaxies. This is the first evidence for the presence of a large amount of non-luminous (dark) matter in clusters of galaxies. While the result was not widely accepted at the time, subsequent observations based on galaxy velocity dispersion, X-ray emission and gravitational lensing effects have provided indisputable evidence that galaxy clusters and groups are all associated with massive dark matter halos. Indeed, even isolated galaxies are also found to be embedded in massive halos, as inferred from their rotation curves and velocity dispersion of stars.

Theoretically, the current ΛCDM model predicts that all galaxies form and evolve in dark matter halos. These halos are virialized clumps of dark matter that form in the cosmic density field through gravitational instability (see Mo et al. 2010, for a review). Therefore, galaxy systems, if selected properly so as to represent halos, can be used to study how galaxies form and evolve in dark matter halos. Furthermore, since dark matter halos are simple but biased tracers of the underlying mass density field (e.g. Mo & White 1996), galaxy systems so selected can also be used to study the structure and evolution of the mass density field in the universe. In particular, as shown in

Wang et al. (2009), a well-defined group sample can be used to reconstruct the cosmic density field, which, in turn, can be used to reconstruct the initial conditions from which the observed structures form and evolve (Wang et al. 2013, 2014, 2016).

A key in using galaxy systems as a proxy of the dark halo population is a group finder that can group galaxies according to common dark matter halos. The widely adopted FoF algorithm is not optimal for the purpose. More recently, Yang et al. (2005) (Y05 hereafter) developed a halo-based group finder, which identify groups based on dark matter halo properties, such as mass and velocity dispersion, expected from the CDM cosmogony. This halo-based group finder has been extensively tested using mock galaxies from simulations and found to perform much better than the traditional FoF algorithm, particularly in identifying poor systems. The group finder of Y05 has been applied to redshift surveys such as the 2dFGRS (e.g. Yang et al. 2005), the SDSS (e.g. Weinmann et al. 2006, Yang et al. 2007), and the 2MRS (e.g. Lu et al. 2016). Similarly, Duarte & Mamon (2015) adopted an iterative group membership assignment algorithm but in a probabilistic way using galaxy distribution statistics extracted from N-body simulation. An important step in the halo-based group finder is the use of a halo-mass proxy to assign halo masses to tentative groups/clusters in the grouping process. Y05 suggested the use of the ranking of the total luminosity of galaxies that have luminosities above a certain value as the proxy of halo mass, and this mass proxy was adopted in the SDSS and 2dFGRS group catalogs mentioned above. However, this mass proxy may not suitable for shallower surveys, such as the 2MRS, where many systems contain only a small number of galaxies. In order to overcome this limitation, Lu et al. (2016) (L16 hereafter) proposed a "GAP correction" method, in which the luminosity/stellar mass of the most luminous/massive member is combined with the "GAP" to form a mass proxy, where "GAP" is defined to be the difference in luminosity/stellar mass between the most luminous/massive member and n-th most luminous/massive member. Using a 2MRS mock sample, L16

found that the GAP method yields a typical dispersion of ~ 0.3 dex in the estimated masses for galaxy systems of a given true halo mass.

In this paper, we modify the group finders of Y05 and L16, paying particular attention to the extension of the methods to poor systems, such as groups containing one member or a small number of members, in a uniform way. We use mock samples constructed from numerical simulations and an empirical model to calibrate the halo mass proxies and to test the performances of the group finder under different sample selections. As we will see below, our modified group finder not only gives more accurate halo mass estimates for groups than the original group finders, it also enables us to uniformly extend the group samples to systems with halo masses that are about an order of magnitude lower than in the existing group catalogs. We apply our group finder to a number of redshift surveys in the local universe, including the 2MRS, the 6dFGS, the updated release of SDSS and the 2dFGRS. As mentioned above, group catalogs have been constructed from some of these catalogs with various group finders. Our goal here is to extend, update, and add values to, these catalogs by providing group samples that are uniformly selected from improved data using improved methods.

The outline of this paper is as followings. In Section 1.2, we describe the observational data to which we apply our group finder, and the simulation that we use to calibrate and test the group finder. Section 1.3 explains in detail the group finder and how to test and calibrate a variety of halo-mass proxies using mock galaxies. In Section 1.4, we apply the group finder to the mocks of the same sample selection as the observational data, and assess its performance by comparing halo masses, membership assignments, and global completeness between the constructed mock groups and the simulations. In Section 1.5, we apply the group finder to real observations, construct our group catalogs, describe their basic properties and how to use them,

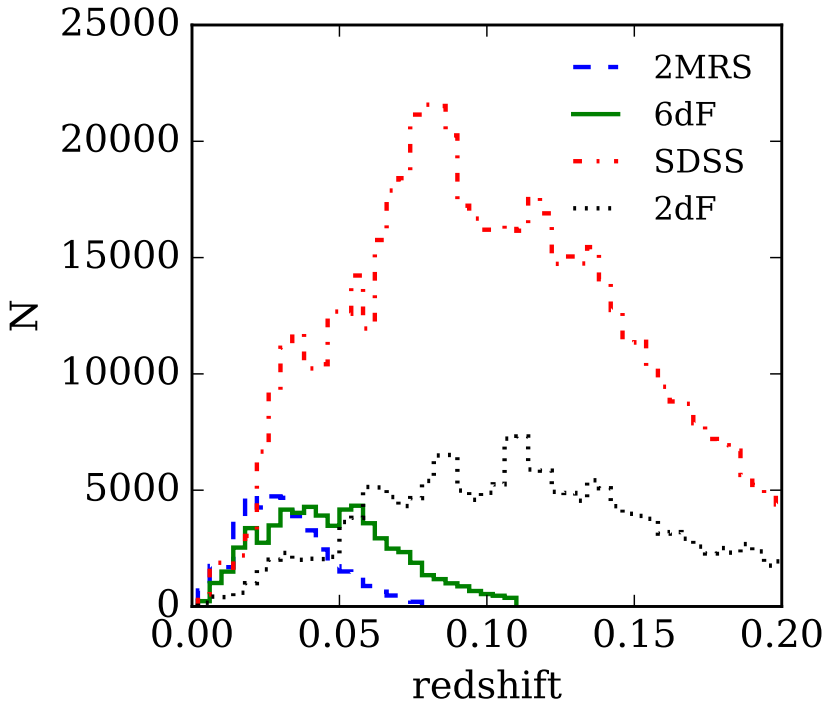

Figure 1.1. The redshift distributions of galaxies in the four samples we use to identify galaxy groups. The bin size of the histograms is $\Delta z = 0.004$.

and make comparisons with other catalogs in the literature. Finally, we summarize our results and discuss applications of the catalogs in Section 2.6.

1.2 Observational data and mock samples

In this section, we describe in detail the galaxy samples we use to construct our group catalogs. Since our goal is to provide well-defined group catalogs in the local universe, we decide to use all major redshift surveys at low redshift ($z <\sim 0.2$) that are publicly available. A brief summary of our sample selections is given in Table 1.1. The redshift distributions of these samples are shown in Figure 1.1, and their sky coverages are plotted in Figure 1.2.

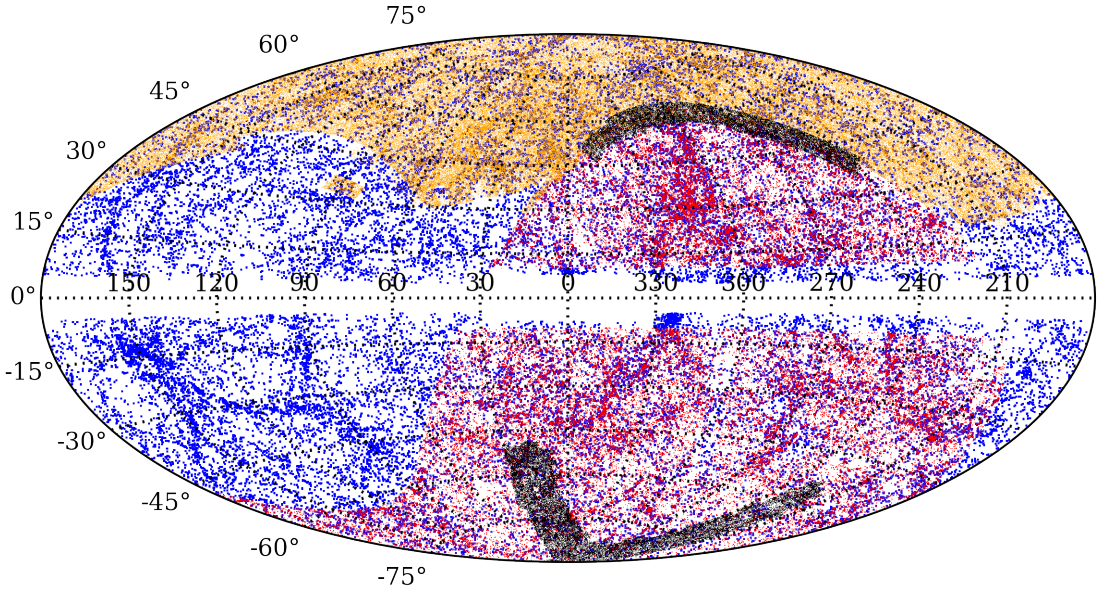

Figure 1.2. The galaxy distributions in Galactic coordinates (Aitoff projection) of the 2MRS (blue), 6dFGS (red), SDSS (orange), and 2dFGRS (black) samples.

Table 1.1. A summary of galaxy samples.

Sample	Sky Coverage (%)	Depth[a] (z)	Magnitude limit (mag)	No. of Galaxies[b]
2MRS	91%	0.08	$K_s \leq 11.75$	$43,249$ $(44,310)$
6dFGS	40%	0.11	$K_{s,tot} \leq 12.5$	$62,987$ $(73,386)$
SDSS	21%	0.2	$r \leq 17.77$	$586,025$ $(600,458)$
2dFGRS	3.3%	0.2	$b_J \leq 19.45$	$180,967$ $(189,101)$

Notes.
a. Upper limit of redshift in CMB rest-frame.
b. The numbers in parentheses are for the extended (the catalogs with '+') catalogs.

1.2.1 The 2MRS catalog

Our first galaxy sample is selected from the 2MASS Redshift Survey (2MRS; Huchra et al. 2012), which is based on the Two Micron All Sky Survey (2MASS;

Skrutskie et al. 2006). 2MASS covers $\sim 91\%$ of the entire sky in the near-infrared J, H, and K_s bands. Because of the reduced dust extinction in the NIR, 2MASS is an almost uniform survey down to the magnitude limit $K_s \leq 13.5$, except in the region within $\pm 5^\circ$ of the Galactic plane (the "zone of avoidance" or ZoA). The extended source catalog (the 2MASS XSC) contains $\sim 10^6$ objects. The 2MRS attempted to obtain redshifts, either from its own observation or from other data bases, for $45,086$ sources of the 2MASS XSC that meet the following criteria:

1. $K_s \leq 11.75$ mag and detected at H,
2. $E(B-V) \leq 1$,
3. $|b| \geq 5^\circ$ for $30^\circ \leq l \leq 330^\circ$; $|b| \geq 8^\circ$ otherwise,

where b is the Galactic latitude, and $E(B-V)$ is the extinction based on the dust map of Schlegel et al. (1998). As shown in Huchra et al. (2012), to the magnitude limit $K_s = 11.75$ the completeness does not change significantly within the region specified by criterion (iii). Of the $45,086$ sources, 2MRS rejected a small fraction that is of galactic origin, only partially detected, or not clearly detected due to contamination. This leaves a total of $44,599$ galaxies. The details about the selection can be found in Huchra et al. (2012) and its appendix. For the $44,599$ galaxies, 2MRS eventually obtained redshifts for $43,533$ systems, achieving a completeness of about 97.6%. These include $11,000$ galaxies measured by the 2MRS team, $7,069$ galaxies with redshifts from SDSS, $11,763$ from 6dFGS DR3, $12,952$ from the NASA Extragalactic Database (NED), and 749 from J. Huchra's personal compilation (ZCAT). For objects with redshifts from more than one source, the preference was given in the order of 2MRS, SDSS, 6dF, NED, and ZCAT.

For the $1,066$ galaxies that do not have redshifts from the 2MRS, we either adopt redshifts of their nearest neighbors or use those given by the 2MASS Photometric Redshift catalog (2MPZ; Bilicki et al. 2014). The 2MPZ uses the optical, NIR, and

mid-IR photometry from SuperCOSMOS, 2MASS, and WISE respectively, to obtain photometric redshifts for about 1 million galaxies, by employing an artificial neural network approach trained with several redshift surveys. The photometric redshifts obtained have a typical error of 12%. We first assign redshifts of the nearest neighbors (z_{NN}) to all galaxies without the spectroscopic redshift. Then, if a galaxy also has a redshift from the 2MPZ (z_{pho}) and z_{pho} differs from z_{NN} by more than 12%, we assign z_{pho} as the redshift of the galaxy instead of z_{NN}.

With all these, we obtain redshifts for 288 additional galaxies from their nearest neighbors, and redshifts for 778 galaxies from the 2MPZ, thus assigning redshifts to all the $44,599$ galaxies. Because of the uncertainties in the nearest-neighbor and photometric redshifts, we will provide two separate catalogs: the first is constructed from the sample of galaxies that all have 2MRS redshifts; the second uses all galaxies that have 2MRS redshifts, z_{pho}, or z_{NN}. The latter will have a flag that shows source of redshifts for each galaxy, as well as the separation to the nearest neighbor for galaxies with z_{NN} so that a user can decide an uncertainty that may be allowed to suit his/her scientific goal. For convenience, we refer to the 1st catalog as 2MRS and the 2nd as 2MRS+.

Our sample also contains a number of refinements appropriate for our purpose. First, we correct all the redshifts (radial velocities) of galaxies to the CMB rest-frame. To do this, we assume that the heliocenter is moving with a velocity of 368 km/s towards $(l, b) = (263.85°, 48.25°)$ with respect to the CMB (Bennett et al. 2003). Second, we only use galaxies with corrected redshifts $z \leq 0.08$, which eliminates about 1% of the galaxies from the sample. The final numbers of galaxies are then $43,249$ and $44,310$ for 2MRS and 2MRS+, respectively. As an example, the redshift and sky distributions of the galaxies in 2MRS are shown in Figures 1.1 and 1.2, respectively, and a brief summary of the samples is given in Table 1.1.

We use the extinction-corrected K_s isophotal magnitudes from the 2MRS. The extinction-correction accounts for dust extinctions of the Milky Way relying on the dust map by Schlegel et al. (1998). We use WMAP9 cosmology to convert apparent magnitudes to absolute magnitudes in a bandpass Q as followings:

$$M_Q = m_Q + \Delta m_Q - DM(z) - \mathcal{K}_Q(z) - \mathcal{E}_Q(z) - \mathcal{S}_Q(z) \tag{1.1}$$

where Δm_Q is the zero-point correction from the survey photometric system to the Vega system (or the AB system for surveys introduced later in this section that use the AB system), which is 0.017 for the 2MASS K_s-band filter (Cohen et al. 2003), $DM(z)$ is the distance modulus at z, $\mathcal{K}_Q(z)$ and $\mathcal{E}_Q(z)$ are the K- and evolution-corrections at redshift z, respectively, and $\mathcal{S}_Q(z)$ corrects for the effect of decreasing aperture size within which flux is integrated with increasing redshift due to dimming of surface brightness. The term $\mathcal{S}_Q(z)$ is not needed when using extrapolated total magnitudes. We follow Lavaux & Hudson (2011) to model $\mathcal{K}_Q(z)$, $\mathcal{E}_Q(z)$, and $\mathcal{S}_Q(z)$, and correct the values of M_Q of individual galaxies to redshift $z = 0.1$. For nearby galaxies that have negative recession velocities in the CMB rest-frame (a total of 25 galaxies in the 2MRS catalog), we adopt distances from 'EDD distances' available at the Extragalactic Distance Database (EDD; Tully et al. 2009) to calculate the absolute magnitudes. These distances, however, are not used in identifying galaxy systems via the group finder, as our group finder works in redshift space. In the cases where we do not find matches from the EDD (a total of 5 galaxies in the 2MRS catalog), we assign the distances of their nearest neighbors that have EDD distances available. Later in §1.5 where we construct group catalogs, we estimate stellar mass using the mean relation between stellar mass and K_s-band luminosity from the simulation described in §1.2.5.

1.2.2 The 6dFGS catalog

Our second sample is selected from the 6dF Galaxy Survey (6dFGS; Jones et al. 2004, 2005). Specifically we use the 6dFGS Data Release 3 (6dFGS DR3; Jones et al. 2009), the final redshift release of the survey. As 2MRS, the 6dFGS is based mainly on the K_s-selected 2MASS, but is deeper, with a magnitude limit of $K_{s,tot} = 12.65$ mag, where $K_{s,tot}$ is the total magnitude from the 2MASS. Note that the magnitudes are corrected for foreground dust extinction, as mentioned above. As shown in McIntosh et al. (2006), these magnitudes are robust against uncertainties in surface brightness. The survey has a sky coverage of $\sim 41\%$ in the southern hemisphere.

According to Jones et al. (2009), the final 6dFGS catalog contains $126,754$ unique redshifts from their own observations, 563 redshifts from the SDSS, $5,210$ redshifts from the 2dFGRS, and $9,042$ redshifts from the ZCAT. For their own observations, the catalog contains only the spectra with quality parameters $Q = 3$ and $Q = 4$, which are appropriate for scientific analysis according to Jones et al. (2009). Redshifts with $Q = 3$ and $Q = 4$ have typical uncertainties of 55 km/s and 45 km/s, respectively (see Jones et al. 2009, for details). However, the 6dFGS has poorer coverage in some regions, such as those toward the Large Magellanic Cloud (LMC) and the South Pole. This can affect the performance of our group finder. To reduce this effect we select a shallower sample, using the 2MASS Extended Source Catalog (2MASS XSC) with a flux limit of $K_{s,tot} = 12.5$ mag as an input catalog. The 2MASS XSC with this flux limit contains $75,098$ entries. For galaxies that have spectroscopic redshifts from the 6dFGS catalog, we assign the 6dF redshift. We also find redshifts for $1,533$ galaxies from the 2M++ galaxy redshift catalogue (2M++; Lavaux & Hudson 2011), which are originally from the NED. For galaxies without spectroscopic redshifts available, we assign redshifts of their nearest neighbors or from the 2MPZ in the same way as for the 2MRS described above. Of all the 2MASS XSC galaxies, $62,929$ have redshifts from the 6dFGS DR3, $1,533$ redshifts from the 2M++, $3,354$ redshifts from the nearest neighbor, and $7,282$

redshifts from the 2MPZ. This, of course, corresponds to 100% redshift completeness. In the end, we will provide two separate catalogs: one constructed using only galaxies with spectroscopic redshifts, and the other using all galaxies, including the ones with nearest-neighbor and 2MPZ redshifts. For convenience, we refer to the 1st catalog as 6dFGS, and the 2nd catalog as 6dFGS+.

For our analysis, we correct all the radial velocities (redshifts) to the CMB rest-frame, as we did for the 2MRS, and we only use galaxies with corrected redshift $z \leq 0.11$. This leaves $62,987$ and $73,386$ galaxies in our final 6dFGS and 6dFGS+ samples, respectively. The redshift distribution and a summary of the final samples are given in Figure 1.1 and Table 1.1, respectively.

The absolute magnitudes of individual galaxies are again calculated using equation (1). The same K-, E-, and surface brightness corrections as those for the 2MRS are used to correct the K_s-band magnitudes to $z = 0.1$. For galaxies with negative recession velocity, we again use the EDD distances to compute their luminosities. We approximate stellar mass using the mean relation between stellar mass and K_s-band luminosity obtained from the simulation described in §1.2.5.

1.2.3 The SDSS catalog

Our third sample is selected from the Sloan Digital Sky Survey Data Release 13 (SDSS DR13; Albareti et al. 2016). DR13 is the first data release of the fourth phase of the Sloan Digital Sky Survey (SDSS-IV) and is built upon prior releases. It includes updated data for the SDSS Legacy Survey, which is a magnitude limited redshift survey completed in SDSS-II, as well as objects from the Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013), the selection of which barely overlaps with that of the legacy survey. The main part of the SDSS Legacy Survey was already released in DR7, and remained more or less steady through DR12. Significant changes were made to photometric calibration in the DR13, including updated zero

points and flat-fields in the g, r, i, and z bands from the hypercalibration procedure of Finkbeiner et al. (2016). These affect all photometric quantities of the galaxies in the Legacy Survey. In addition to the updated photometry calibration, another significant improvement in DR13 relative to, for example, DR7 is that some of the fiber-collision galaxies in DR7 have their redshifts measured in DR13. The SDSS spectrograph used for the Legacy Survey did not allow two fibers to be positioned within 55 arcsec, and so no spectroscopic measurement was available for galaxies that have close neighbors within the fiber separation, the so-called 'fiber-collision' galaxies. Many of the 'fiber-collision' galaxies ($\sim 60\%$, e.g. Guo et al. 2015) have been measured spectroscopically in the later data releases through DR13. The Legacy Survey covers approximately $\sim 23\%$ of the sky, and is complete to an extinction-corrected Petrosian magnitude of 17.77 mag in the r-band.

From the full photometric catalog of DR13, we select all objects that are in the Legacy Survey region and identified as galaxies (type = 3) brighter than the r-band magnitude limit of 17.77. We take the photometric quantities only from the primary observation in the cases where an object was observed multiple times (mode = 1). We also get rid of galaxies in the Southern Galactic Cap, as its narrow angular boundary makes our group finder unreliable for many systems close to the boundary. Note that these selections may include some of the BOSS galaxies that pass the selection criteria. The selections leave a total of $638,191$ entries, of which $16,251$ galaxies do not have redshifts for reasons such as fiber-collisions, broken or unplugged fibers, bad spectra, or poor fit to models. Of the $621,940$ galaxies that have redshifts, $20,780$ ($\sim 3.3\%$) are BOSS galaxies in the Legacy region.

For the $16,251$ galaxies without SDSS redshifts, we find redshifts from other sources: the 2dFGRS, 6dFGS, the Korea Institute for Advanced Study Value-Added Galaxy Catalog (KIAS VAGC; Choi et al. 2010), a complementary galaxy sample in the LAMOST Survey (Luo et al. 2015, Shen et al. 2016), the nearest neighbors, or

the 2MPZ, to achieve 100% redshift completeness. For galaxies that have redshifts available from more than one sources, preferences are given in the order given above, i.e. from 2dFGRS to 2MPZ. As a result, 294, 29, 168, 227, 13,548, and 1,985 additional redshifts are obtained from 2dFGRS, 6dFGS, KIAS VAGC, LAMOST, the nearest neighbors, and 2MPZ, respectively. In the following, we will construct two different kinds of group catalogs from the SDSS data, one using only galaxies that have spectroscopic redshifts, and the other using all galaxies including the ones with estimated redshifts from nearest neighbors and from 2MPZ. For brevity, we refer to the 1st catalog as the SDSS and the 2nd as the SDSS+. We covert all the recession velocities (redshifts) to the CMB rest-frame, and restrict our samples to $z \leq 0.2$. This leaves a total of 586,025 and 600,458 galaxies as our final SDSS and SDSS+ samples, respectively.

We compute the absolute magnitudes in the r-band of the sample according to equation (1), using the WMAP9 cosmology, with K- and evolution- corrections to $z = 0.1$ following Poggianti (1997). We also calculate the $(g - r)$ color, corrected to $z = 0.1$, for each galaxy. From the DR13 photometric catalog, we adopt the cmodel magnitude to calculate the flux, and the model magnitudes to compute the color, following the recommendations of the SDSS team. The zero-point offset between the DR13 magnitude and the AB magnitude is practically zero for g- and r- bands within the error of 0.01 mag. Galaxies with colors outside the 3σ of the color distribution at a given luminosity are assigned the median color. For galaxies with negative redshifts, their luminosities are obtained from their EDD distances. Finally, we estimate the stellar masses of individual galaxies from their r-band absolute magnitudes and $(g-r)$ colors, following the formula of Bell et al. (2003):

$$\log M_* = -0.306 + 1.097(g - r) + 0.4(4.67 - M_r), \tag{1.2}$$

where 4.67 is the absolute magnitude of the Sun in the r-band.

1.2.4 The 2dFGRS catalog

Finally, we also select a sample from the 2dF Galaxy Redshift Survey (2dFGRS; Colless et al. 2001). The 2dFGRS provides redshifts for about $250,000$ galaxies, measured with the Two-degree Field (2dF) multifibre spectrograph on the Anglo-Austrailian Telescope, down to a magnitude limit of $b_J = 19.45$ after Galactic extinction correction. The survey consists of two strips in the northern and southern Galactic hemispheres (the northern and southern Galactic caps, respectively), and 99 'random' fields of 2° each over and around the southern Galactic cap. The full survey covers about $2,000\mathrm{deg}^2$ with a median redshift of $z \sim 0.11$. Because the random fields are not contiguous and our group finder can be affected severely at the edges of these field, we use only the two Galactic caps for our purpose. The final sky coverage of our sample is about 3.5%.

The quality of a spectrum is characterized by a quality parameter, $Q = 1$ - 5, with a higher value of Q indicating higher quality. From the final release spectroscopic catalog, we use only galaxies with $Q \geq 3$, for which the redshifts are 98.4 per cent reliable, with a typical uncertainty of 85 km/s (Colless et al. 2001). Of all the $245,591$ galaxies from the 2dFGRS catalog, $12,340$ systems do not have spectroscopic redshifts from the survey. For these galaxies, we find matches and assign redshifts from the SDSS DR13, the 6dFGS, the nearest neighbors, and the 2MPZ. In the cases where a galaxy has redshifts from more than one of these sources, the priority is given, in the order of decreasing priority, to the SDSS, the 6dFGS, the nearest neighbor redshift, and the 2MPZ. As a result, we have $233,251$ redshifts from the 2dFGRS, 322 from the SDSS, 43 from the 6dFGS, $11,852$ from the nearest neighbors, and 123 from the 2MPZ. Again, we will provide two catalogs, one using only galaxies with spectroscopic redshifts, and the other using all galaxies. We refer to these two catalogs as the 2dFGRS and 2dFGRS+ samples, respectively.

Redshifts are corrected to the CMB rest-frame, and we limit our sample to $z \leq 0.2$. The final samples contain $180,967$ and $189,101$ galaxies for the 2dFGRS and 2dFGRS+, respectively. When selecting galaxy groups from these samples, we adopt the survey masks provided in the 2dFGRS website [2]. For reference, the redshift distribution of 2dFGRS is shown in Figure 1.1.

We use equation (1) to convert the observed apparent magnitudes to absolute magnitudes, assuming WMAP9 cosmology. To do this, we first make K- and E-corrections to $z = 0.1$ following the method given in Poggianti (1997). The stellar masses of individual galaxies are obtained from their b_J-band absolute magnitudes and $b_J - R$ colors using the approximation of Bell et al. (2003):

$$\log M_* = -0.976 + 1.111(b_J - R) + 0.4(5.48 - M_{b_j}), \tag{1.3}$$

where 5.48 is the absolute magnitude of the Sun in the b_J-band. For a small number of galaxies that have colors outside the 3σ range of the $b_J - R$ distribution, and for a total of 276 galaxies without the R-band photometry, each of them has been assigned a $(b_J - R)$ color that is equal to the median value given by the galaxies which have b_J luminosities similar to the galaxy in question and have $b_J - R$ colors. Here again, the EDD distances have been used to convert the observed flux to the luminosity for galaxies with negative redshifts.

1.2.5 Mock samples used to test methods

The quality of the group samples to be constructed depends on the performance of the group finder used to identify the groups from the observational data. To test the performance of our group finder (to be described in §1.3), we use mock

[2] http://www.2dfgrs.net/

samples constructed from a hydrodynamical simulation of galaxy formation, where information about dark matter halos and their galaxy memberships are all known.

The hydrodynamical simulation used here is the Evolution and Assembly of GaLaxies and their Environments (EAGLE; Schaye et al. 2015, Crain et al. 2015, McAlpine et al. 2015). EAGLE follows the evolution of gas, stars, dark matter, and massive black holes in a cosmological context, implementing physical models for gas cooling, star formation, stellar and AGN feedback. Sub-grid processes, in particular feedback processes, are modeled with simple parametric forms, with model parameters tuned to match observations, such as the stellar mass function and stellar mass - black hole mass relation at $z \sim 0$, as detailed in Crain et al. (2015). The simulation starts from $z = 127$ and adopts the *Planck* cosmology with $(\Omega_m, \Omega_\Lambda, h) = (0.307, 0.693, 0.678)$ (Planck 2014). This cosmological model is not exactly the same as the WMAP9 cosmology we adopt in this paper. However, since the purpose here is to test our group finder and halo mass proxies (see below), this difference in cosmology should not be a concern, as long as the analysis is done in a self-consistent way. EAGLE provides a set of simulations assuming different sets of model parameters and different box sizes. Here we use the simulation with the largest box size of $100\mathrm{Mpc}^3$, their fiducial simulation. The simulation contains about $11,500$ dark matter halos with masses above $10^{11}\mathrm{M}_\odot$, and $\sim 10,000$ galaxies with masses comparable to or above that of the Milky Way. EAGLE adopted the Chabrier (2003) IMF and the spectral synthesis model of Bruzual & Charlot (2003) to get luminosities and stellar masses of individual galaxies from their star formation histories, and these are used in our analysis as well. A few galaxies are found to have extremely low halo masses for their stellar masses. These extreme outliers are excluded from our analysis.

We construct realistic mock catalogs of galaxies from EAGLE. Since the original simulation box, 100 Mpc, is smaller than the volumes of our samples, we stack the duplicates of the original box side by side as many times as is required to cover the

volume of the sample in question. A location is chosen for the observer in the stack, and apparent magnitudes of individual galaxies are calculated from their luminosities and their distances to the observer. The same sample selections as those for the observational samples, as detailed earlier in this section, are applied to construct the mock samples. Specifically, we choose galaxies in the simulation box that are in the same sky regions as the observational samples, as well as apply the apparent magnitude and redshift limits of each survey to eliminate faint galaxies from the mock samples. Finally, we also apply the same masks, if any, as provided for the observational samples by each survey. Since all the galaxies in the simulation are linked to dark matter halos, we can use these mock catalogs to quantify the accuracy of our methods.

As an independent check, we have also constructed mock samples using an empirical model of galaxy formation. The details of these mock samples are given in Appendix A.

1.3 The halo-based group finder

1.3.1 The basic algorithm

The method adopted here is similar to the 'halo-based' group finder developed by Yang et al. (2005) (Y05 hereafter). This group finder makes use of physical properties of dark matter halos expected from the current cold dark matter (CDM) cosmogony, such as halo mass, virial radius and velocity dispersion, in assigning galaxies into groups. The group finder has been tested extensively using mock galaxies, and is found to be more effective than the traditional Friends-of-Friends (FoF) algorithm in grouping galaxies according to common halos, and particularly in dealing with poor groups associated with small halos. This allows the identification of systems over a wide range of masses. However, in the original group finder of Y05, halo masses assigned to galaxy groups are based on the ranking order of the total luminosity of

member galaxies (or the sum of the luminosities of member galaxies above a certain luminosity limit). It becomes inaccurate for groups that contain only a small number of members, and is not appropriate for shallow surveys where a large number of the identified groups contain only one or a small number of relatively bright galaxies. In order to overcome this limitation, we make some modifications to the group finder of Y05, in particular in the assignments of masses to galaxy groups. Specifically, for systems containing more than one member galaxy, we adopt a modified version of the 'GAP' model developed by Lu et al. (2016). For systems containing only one member, we use halo mass proxies that are calibrated by realistic mock catalogs. As we will show below, these modifications not only provide more accurate halo mass estimates, but also allow us to reach to systems with lower halo masses in a uniform way. The detailed steps of the group finder are as followings:

Step 1. *Assign preliminary halo mass to every galaxy.*

While the group finder of Y05 starts by linking galaxies using the FoF algorithm with a small linking length to identify preliminary group centers, we start by treating all galaxies as isolated galaxies associated with distinct tentative dark matter halos with preliminary halo masses computed according to the halo mass proxies described in §1.3.2. We have checked that this leads to no significant differences in membership, mass, and the number of final groups in comparison to that of Y05.

Step 2. *Membership assignment using halo properties.*

For all groups identified at each iteration, we compute the size and the line-of-sight velocity dispersion, which are used to determine which galaxies should be assigned to a certain group,

$$\frac{r_{180}}{\mathrm{Mpc}} = 1.33\ h^{-1}\left(\frac{M_h}{10^{14}h^{-1}M_\odot}\right)^{1/3}(1+z_{\mathrm{group}})^{-1}$$

$$\frac{\sigma}{\mathrm{km\ s^{-1}}} = 418\left(\frac{M_h}{10^{14}h^{-1}M_\odot}\right)^{0.3367}, \tag{1.4}$$

where z_{group} is the redshift of the group in question, and r_{180} is the radius of the halo, within which the mean mass density is 180 times the mean density of the universe at the given redshift. The numbers used are appropriate for the WMAP9 cosmology (e.g. Lu et al. 2016). Next, we assume that the phase-space distribution of galaxies in dark matter halos follows that of dark matter particles and that the group center is the same as the halo center. The number density contrast of galaxies at the redshift of z_{group} can then be expressed as

$$P_M(R,\Delta z) = \frac{H_0}{c}\frac{\Sigma(R)}{\bar{\rho}}p(\Delta z) \tag{1.5}$$

where R is the projected distance, c is the speed of light, $\bar{\rho}$ is the mean density of the Universe, $\Sigma(R)$ is the surface density, and $\Delta z = z - z_{\mathrm{group}}$. We assume that the redshift distribution of galaxies within a halo, $p(\Delta z)$, has the Gaussian form,

$$p(\Delta z) = \frac{c}{\sqrt{2\pi}\sigma(1+z_{\mathrm{group}})}\exp\left(\frac{-c^2\Delta z^2}{2\sigma^2(1+z_{\mathrm{group}})^2}\right) \tag{1.6}$$

where σ is the line-of-sight velocity dispersion. Furthermore, halos are assumed to follow a spherical NFW density profile, so that the surface density $\Sigma(R)$ can be written as

$$\Sigma(R) = 2r_s\bar{\delta}\bar{\rho}f(R/r_s) \tag{1.7}$$

where r_s is the scale radius, and

$$f(x) = \begin{cases} \frac{1}{x^2-1}\left[1 - \frac{\ln\frac{1+\sqrt{1-x^2}}{x}}{\sqrt{1-x^2}}\right], & \text{if } x < 1 \\ \frac{1}{3}, & \text{if } x = 1 \\ \frac{1}{x^2-1}\left[1 - \frac{\mathrm{atan}\sqrt{x^2-1}}{\sqrt{x^2-1}}\right], & \text{if } x > 1 \end{cases}$$

$$\bar{\delta} = \frac{180}{3}\frac{c_{180}^3}{\ln(1+c_{180}) - c_{180}/(1+c_{180})} \tag{1.8}$$

with the concentration, $c_{180} = r_{180}/r_s$, given by the model of Zhao et al. (2009). Finally, we calculate $P_M(R, \Delta z)$ for each of all the galaxy-group pairs. If the value of P_M is above a certain background value, P_B, an association between the galaxy and the group is assumed. If a galaxy is associated with more than one group according to this criterion, the galaxy is assigned to the group with the largest $P_M(R, \Delta z)$. As demonstrated in Y05 using realistic mock samples, a compromise between the completeness and contamination can be achieved with $P_B \sim 10$, and the performance of the group finder is not very sensitive to the exact value of P_B. Note that $P_B \sim 10$ is also in agreement with theoretical expectations for dark matter halos (see the discussion in section 3.2 of Y05 for details). We therefore adopt $P_B = 10$ throughout this paper.

After the membership of a group is determined, we define the stellar mass-weighted center of member galaxies as the group center, if stellar masses are available. Otherwise, we use luminosity-weighted center as the group center.

Step 3. *Rank groups according to halo mass proxies.*

Table 1.2. The best-fit parameters for the GAP correction from the EAGLE.

Sample	α_1 ($\times 10^5$)	α_2 ($\times 10^6$)	β	γ	δ_1	δ_2	δ_3
2MRS	-2.3	-1.1	6.6	5.1	0.28	0.13	0.035
6dFGS	-1.1	-3.9	4.8	6.1	0.13	-0.15	0.037
SDSS	-3.6	-6.0	5.3	5.2	0.11	-0.039	0.0014
2dFGRS	-3.6	-5.9	5.5	5.7	0.12	-0.046	0.0035

As described in §1.3.2.1 and §1.3.2.2, in the beginning of each iteration, tentative halo masses are assigned to groups identified in the previous step by ranking groups according to a mass proxy. In short, for tentative groups containing only one member galaxy in the previous step, we use the galaxy stellar mass (luminosity) - halo mass relation obtained from a hydrodynamical simulation to assign the preliminary halo mass, as described in §1.3.2.2. For tentative groups that contain more than one member at a given iteration, we use the 'GAP correction' method of Lu et al. (2016), modified with our own re-calibrations (see below).

Step 4. *Group mass update and iteration.*

To assign masses to groups, we use abundance matching between the mass function of the preliminary groups and an adopted theoretical halo mass function. A new halo mass, $M_{\rm halo}$, is assigned to a group to replace the preliminary group mass, $M_{\rm pre}$, according to

$$N(> M_{\rm halo}) = N(> M_{\rm pre})$$

where N is the cumulative number density of groups (halos) more massive than $M_{\rm pre}$ ($M_{\rm halo}$). We use the theoretical halo mass function of Sheth et al. (2001) for this. Note that for flux-limted samples, halos of a given mass are complete only to a certain redshift. The abundance matching used to assign halo mass is applied only for groups in samples that are complete (see §1.4.2). For groups residing in volumes within which the samples are not complete, we use the mean relation between the halo mass and the mass proxy from the last iteration to assign halo masses to them. Once group masses are updated, we iterate Steps 2 through 4 until convergence in group membership is achieved.

1.3.2 Halo mass proxies of galaxy groups

As mentioned in the previous section, our group finder relies on the reliability of the halo mass model for groups. Here we test different halo mass proxies by comparing their predictions with the results obtained from the hydrodynamical simulation. We have also made similar tests using a mock sample of galaxies constructed by applying the empirical model of Lu et al. (2015) to the simulated halos. The results obtained from the empirical model are very similar to those obtained from the hydrodynamical simulation, and are presented in Appendix A.

1.3.2.1 Halo mass proxies for groups containing more than one galaxy

In the original paper presenting the halo-based group finder, Y05 uses the sum of the luminosities of member galaxies down to some luminosity limit as a proxy of group masses. However, this proxy may not be appropriate for a shallow galaxy survey where many groups have only a small number of members. Because of this, Lu et al. (2016) (L16 hereafter) suggested the use of a combination of the luminosity/stellar mass of the central galaxies, and the luminosity/stellar mass GAP (the difference in luminosity/stellar mass between the central galaxy and the n-th brightest galaxy) as

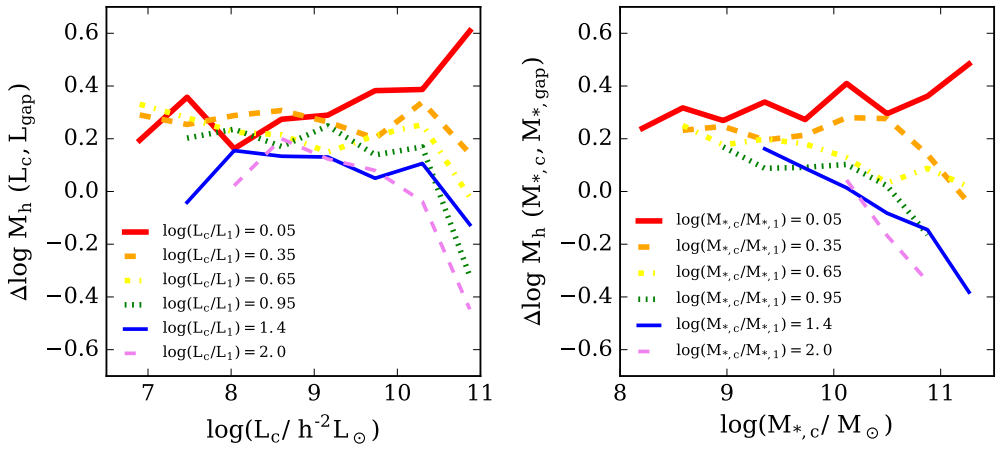

Figure 1.3. Dependence of halo mass on the luminosity or stellar mass of the central galaxy and the GAP parameter (defined to be the difference in luminosity or stellar mass between the central galaxy and the n-th brightest satellite; see text for detailed definition) as given by EAGLE. The left panel is the result based on the K-band luminosity, while the right panel shows result based on stellar mass. For clarity, only the GAP correction using the brightest satellite (i.e. $n = 1$) is shown.

a group mass proxy. Based on mock galaxy samples, L16 came up with the following model for the halo mass,

$$\log M_h(L_c, L_{\rm gap}) = \log M_h(L_c) + \Delta \log M_h(L_c, L_{\rm gap}) \tag{1.9}$$

where L_c is the luminosity of the central galaxy, $M_h(L_c)$ is the mean halo mass at a given L_c, and $L_{\rm gap} = L_c/L_n$ with L_n the luminosity of the n-th brightest satellite. The masses and luminosities here are in units of $\mathrm{M}_\odot/h$ and $\mathrm{L}_\odot/h^2$, respectively. Using mock samples constructed for the 2MRS, L16 found that their group masses are consistent with true halo masses obtained from the simulation used in their mock samples, and the best result is achieved with $n = 4$. For groups with less than four satellites, L16 used the faintest satellite in a group for the GAP correction. The basic motivation behind the GAP correction is that groups of the same central galaxy

luminosity but with more contribution from satellites should possess more massive halos.

Here we adopt the same idea of the GAP correction, but use our own functional form for $\Delta \log M_h(L_c, L_{\rm gap})$:

$$\Delta \log M_h(L_c, L_{\rm gap}) = \alpha(L_{\rm gap}) \times (\log L_c - \beta)^\gamma + \delta(L_{\rm gap})$$

with

$$\begin{aligned} \alpha(L_{\rm gap}) &= \alpha_1 + \alpha_2 \log(L_{\rm gap}) \\ \delta(L_{\rm gap}) &= \delta_1 + \delta_2 \log(L_{\rm gap}) + \delta_3 [\log(L_{\rm gap})]^2 \end{aligned} \tag{1.10}$$

where the free parameters α_1, α_2, β, γ, δ_1, δ_2, and δ_3 are constants, and the masses and luminosities are again in solar units. We use mock catalogs constructed from EAGLE to calibrate these free parameters. For example, from the 2MRS mock catalog we obtain $(\alpha_1, \alpha_2, \beta, \gamma, \delta_1, \delta_2, \delta_3) = (-2.3\times10^{-5},\ -1.1\times10^{-6},\ 6.6,\ 5.1,\ 0.28,\ 0.13,\ 0.035)$ for $L_{\rm gap} = L_c/L_2$. The values of these parameters for other cases are given in Table 1.2.

As shown later in this section, the use of stellar masses gives better halo proxies than the use of luminosities. Thus, halo masses based on stellar masses are preferred to those based on luminosities whenever stellar masses are available. The halo mass proxy using stellar mass is modeled in the same way as that given above, except with L_c and $L_{\rm gap}$ replaced by $M_{*,c}$ and $M_{*,\rm gap}$, respectively. Figure 1.3 shows the relations given by equations (9) and (10) using galaxy luminosities or galaxy stellar masses.

1.3.2.2 Halo mass proxies for groups containing one galaxy

Next we consider systems that contain only one member galaxy. Here we present the best proxy for such systems for each catalog based on tests with a number of proxies. Note that the GAP correction from the previous section is not applicable for

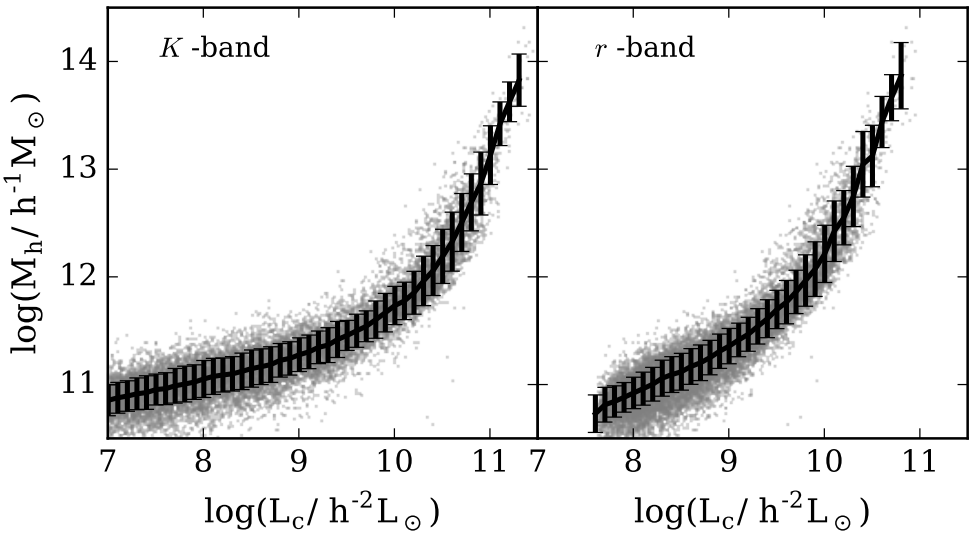

Figure 1.4. The relation between the K-band (left) and the r-band (right) luminosity of central galaxies and the halo mass, as obtained from the EAGLE simulation. The gray points are individual systems, while the black line and bars show the median and scatter of the relation, respectively.

isolated galaxies, as by definition there is no observed satellite in systems containing only one member. In L16, it was assumed that each isolated galaxy has, with 50% chance, one potential satellite galaxy with $K_s = 11.75$, which is the magnitude limit of 2MRS catalog. The average of the corresponding GAP-corrected group mass and $M_h(L_c)$ were used as the halo mass proxy in L16, if the GAP-correction $\Delta \log M_h$ is larger than 0.5. Such a prescription sometimes leads to too high or too low a halo mass for a given L_c according to our test with the mock samples used here. Because of this, here we attempt to revise the proxy so that it is more reliable for groups with only one member.

1.3.2.2.1 Proxy-L: Galaxy luminosity Our first halo mass proxy is based on the luminosities of galaxies. To do this, we first obtain the luminosity - halo mass relation of central galaxies from EAGLE. Figure 1.4 shows such relations in the K-band and r-band. The distribution of the halo masses at a given luminosity is roughly

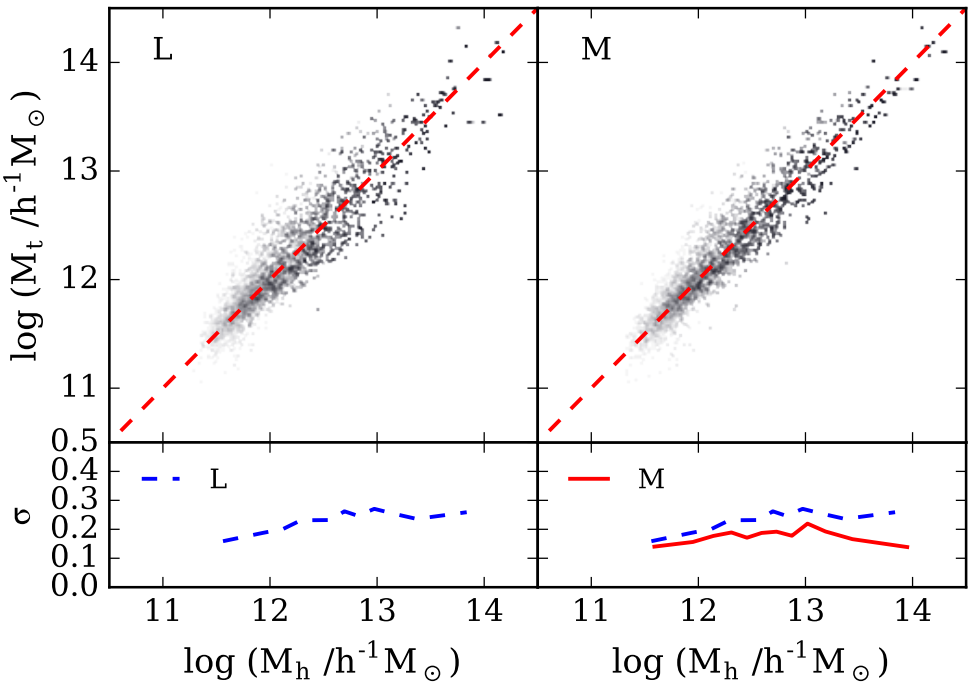

Figure 1.5. The correlations of the halo masses given by a mass proxy for groups containing a single member galaxy (horizontal axis) with the true halo mass (vertical axis), obtained from the 2MRS mock sample constructed with the EAGLE simulation. The results shown use proxies based on the K-band luminosity (L; left), and stellar mass (M; right). The red straight line in each big panel shows a perfect correlation, while the curves in the smaller panels show the scatter in the correlation.

log-normal. In the K-band, which will be used for both 2MRS and 6dFGS, the mean relation can be well described by

$$\log M_h = 10.789 + 2.109 \times 10^{-4} \exp\left(\log L_c / 1.184\right) \tag{1.11}$$

and the typical width is about 0.2 dex. The units of M_h and L_c are in $\mathrm{M}_\odot/h$ and $\mathrm{L}_\odot/h^2$, respectively. In the r-band, which will be used for SDSS and 2dFGRS, the mean relation is given by

$$\log M_h = 10.595 + 4.370 \times 10^{-4} \exp\left(\log L_c / 1.214\right) \tag{1.12}$$

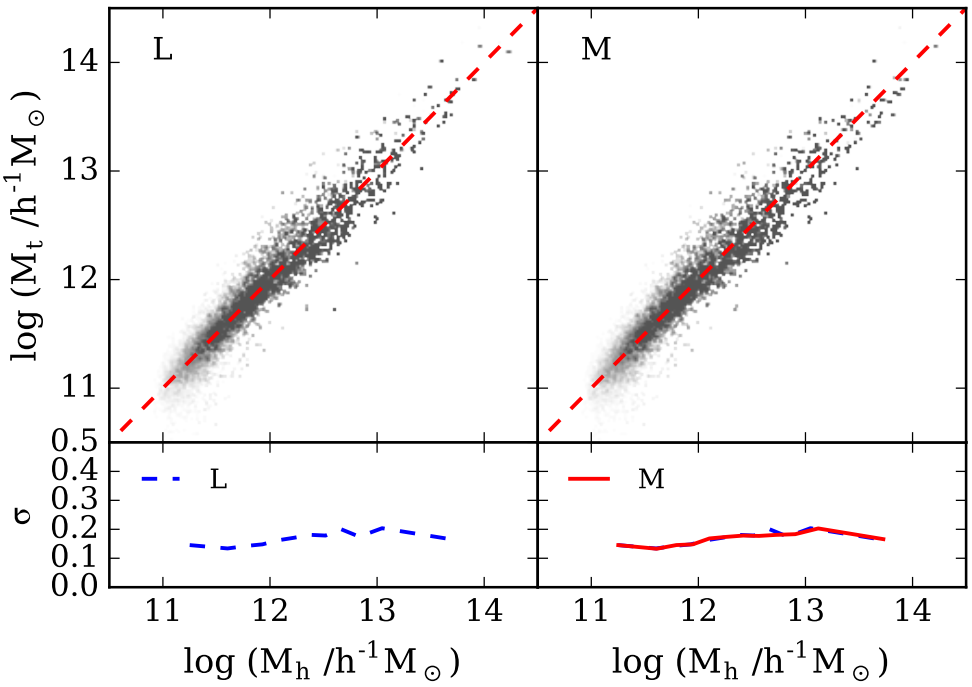

Figure 1.6. The same comparison as in Figure 1.5 but here for the SDSS mock sample.

and the width of the log-normal distribution is about 0.22 dex. We assign the mean halo mass at a given luminosity as the tentative halo mass to each galaxy. We also tested generating a random mass at given luminosity around the mean halo mass and using it as the mass proxy, and found that the resulting scatter between the true halo mass and final halo mass from the group finder is larger by ~ 0.1 dex than that given by using the mean relation.

Figure 1.5 compares the group masses given by the group finder with true halo mass from EAGLE for the 2MRS mock sample. The overall agreement between true mass and group mass using this proxy is found to have scatter of $0.2-0.25$ dex. Note that Figure 1.5 only includes isolated galaxies, which are expected to have larger scatter than groups of more than one member for which the GAP correction will be used.

1.3.2.2.2 Proxy-M: Galaxy stellar mass We also test a halo mass proxy based on stellar mass of the central galaxy. To do this, we use the mean relation between the halo mass and stellar mass of isolated galaxies from EAGLE to assign preliminary halo mass. Figure 1.5 shows the comparison of the resulting final group mass with true halo mass. It is clear that the scatter in the group mass is significantly reduced, by ~ 0.05 dex or more, relative to Proxy-L, suggesting that stellar mass is a better halo mass proxy for isolated galaxies. In real observations, however, stellar mass estimates introduce additional uncertainties. Thus we provide catalogs based on both Proxy-L and Proxy-M in §1.5, where we construct and present our group catalogs.

Figure 1.6 shows the same comparisons between the true and estimated halo masses for the SDSS mock sample. The two mass proxies, Proxy-L and Proxy-M are used in the same way as described above for the 2MRS mock, except that the parameters in the mass models are obtained for the SDSS r magnitude. We see that for the SDSS mock catalog, the two mass proxies give very similar scatter in the halo mass, $\sim 0.15-0.2$ dex. This is different from 2MRS, for which Proxy-M appears to be significantly more accurate than Proxy-L. This may be due to the fact that isolated galaxies in the SDSS mock are dominated by low-mass galaxies (because of its fainter magnitude limit) for which the galaxy color does not depend systematically on halo mass. We also found the same level of scatter, with perhaps a slight increase at the massive end, in tests based on the mock samples constructed from the empirical model, where uncertainties such as that in the stellar mass measurements, are taken into account.

1.3.2.2.3 Other proxies tested We have tested a number of other quantities available from the EAGLE such as velocity dispersion and metallicity of galaxies, as well as halo formation time and local density of galaxies. While some of these quantities are also found to be strongly correlated with halo mass according to the simulations, we found that halo mass proxies based on these quantities are not as

accurate as those given by stellar mass and luminosity. We have also tested using combinations of stellar mass/luminosity and one of these additional quantities as halo mass proxies and found that none of them makes significant improvement in the halo mass estimate, at least according to the simulation we use here. Note also that these additional quantities are usually not available from actual observations, making them less useful in practice.

Motivated by L16 who used a 'GAP limit' as a second parameter in the halo mass proxy for isolated galaxies, we have also made tests with the use of some measurements of the 'GAP limit'. The GAP limit, as defined in L16, is the GAP correction described in the previous section but using $G_{\rm gap,lim} = L_c/L_{\rm lim}$ instead of $G_{\rm gap} = L_c/L_n$, where $L_{\rm lim}$ is the luminosity that corresponds to the observational magnitude limit at the redshift of the galaxy in question. Thus, an isolated galaxy with smaller $G_{\rm gap,lim}$ should have, on average, a more massive satellite that is not observed due to the magnitude limit. The 'GAP limit' is an attempt to take such an effect into account. However, our test showed that using 'GAP limit' does not lead to further improvement in the final halo mass.

Given all these test results, we use the stellar mass when it is available, and use luminosity otherwise, as the halo mass proxy for isolated galaxies. However, given the observational uncertainties in stellar mass estimates. we will provide two catalogs for each data set: a catalog constructed based on Proxy-L and a catalog based on Proxy-M.

1.4 Testing the group finder with mock samples

Before we apply the group finder to observational samples, we test its performances by applying it to realistic mock samples described in §1.2.5, and analyzing the accuracy of group masses, and the completeness, contamination and purity of group memberships that are expected from each of the observational samples.

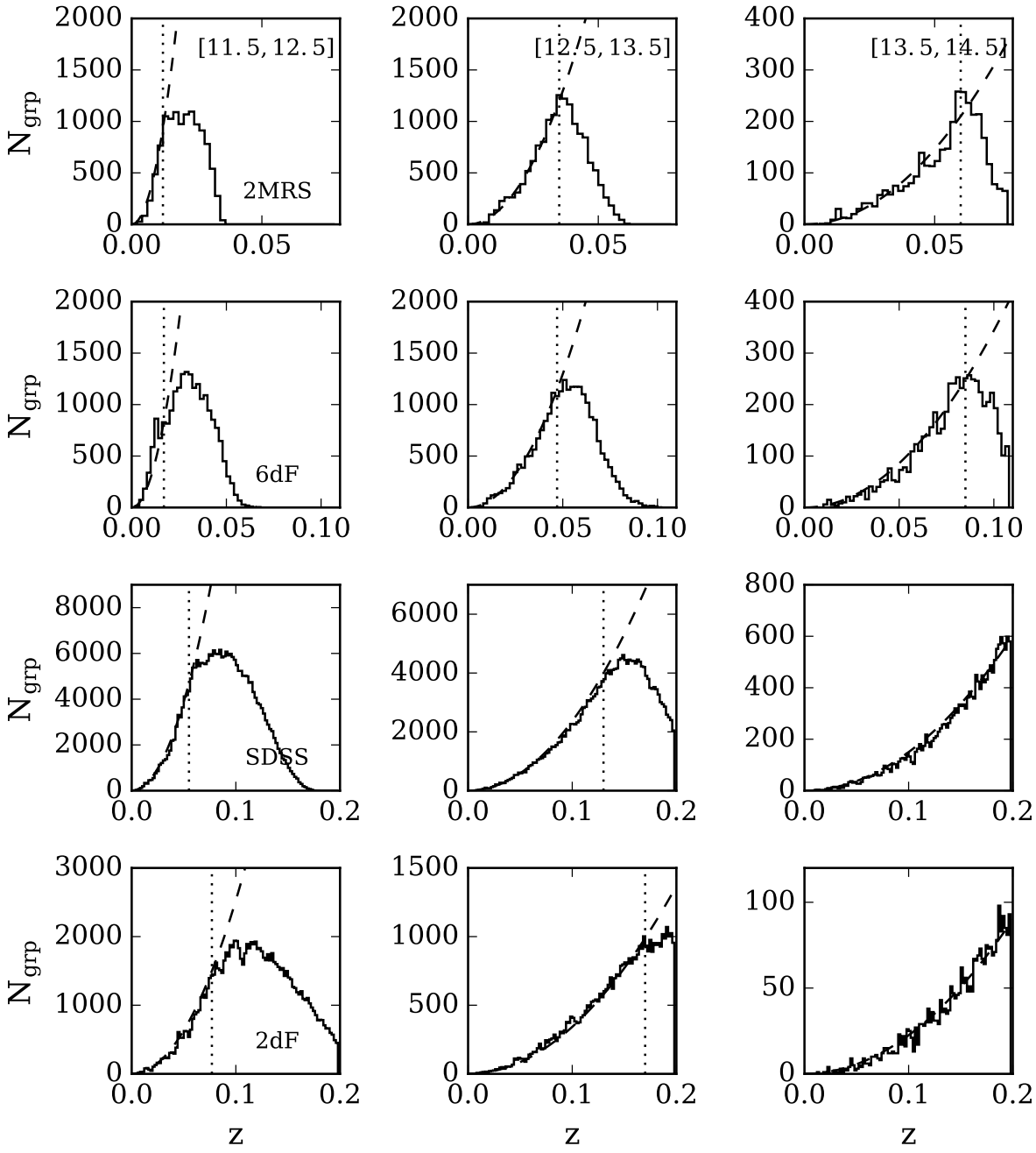

Figure 1.7. The number of halos as a function of redshift from the EAGLE simulation in several mass bins as indicated in the upper panels for each mock sample. The vertical dotted lines show the redshift limits to which the samples are complete for a given halo mass.

Table 1.3. A summary of group catalogs constructed from the mock samples.[a]

Mock	Total groups	$N^{b} = 1$	$N = 2$	$N = 3$	$N \geq 4$
2MRS	30, 124 (29, 462)	25, 160 (24, 305)	3, 128 (3, 310)	881 (845)	955 (1002)
6dFGS	49, 796 (48, 496)	41, 623 (39, 938)	5, 119 (5, 427)	1, 505 (1, 477)	1, 549 (1, 654)
SDSS	473, 303 (454, 474)	397, 300 (377, 280)	49, 166 (51, 832)	13, 937 (14, 587)	14, 846 (15, 847)
2dFGRS	163, 413 (156, 825)	134, 996 (126, 670)	17, 960 (18, 548)	5, 108 (5, 740)	5, 349 (5, 867)

Notes.
a. The numbers in parentheses are from the simulation.
b. The number of member galaxies in a group.

1.4.1 Applying the group finder to the mock samples

As we have shown, stellar mass is theoretically a better proxy of group mass than luminosity (see §1.3) but only when observational uncertainties in stellar mass estimate are negligible. Therefore, we present catalogs that use both luminosity and stellar mass as mass proxies. We use the GAP correction in luminosity (or stellar mass) for groups with more than one member, and Proxy-L (or Proxy-M) for isolated galaxies.

1.4.2 Group mass estimates

In the end of each iteration of the group finder, we finalize group masses using abundance matching. The abundance matching is applied only to volumes within which groups of a given mass are complete. As the surveys and the mock samples are flux-limited, halos of a given mass are only complete to a certain redshift. Figure 1.7 shows the number of halos as a function of redshift for each mock sample from the simulation. As one can see, in each case the number of groups first follows well the expectation of a constant density indicated by the dashed curve in each panel, and starts to go below the expectation at some redshift as incompleteness becomes severe. We can therefore define a limiting redshift, within which the group sample in question is approximately complete. The limiting redshift, $z_{\rm lim}$, is indicated as the dot vertical line in each panel, and Figure 1.8 shows the value of $z_{\rm lim}$ as a function of halo/group mass for the four mock samples corresponding to the four observational samples. These relations can all be well described by a power law, $(1 + z_{\rm lim}) \propto M_{\rm h}^{\zeta}$, as shown in Figure 1.8. The limiting redshifts obtained in this way are used to define complete samples for abundance matching. For groups that are outside the limiting redshift, we use the mean relation between halo mass and the mass proxy to assign halo masses to them.

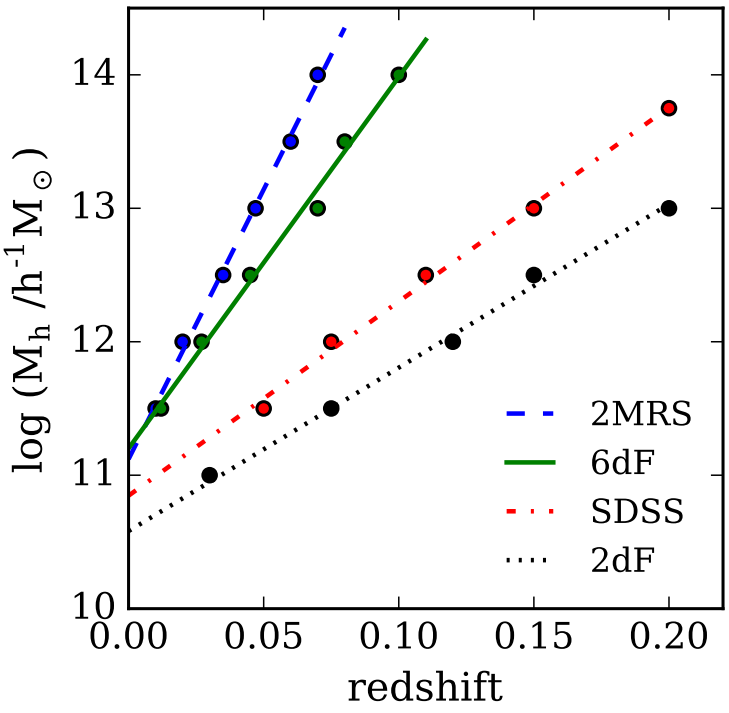

Figure 1.8. Halo mass that is complete as a function of redshift for the mock samples of the simulation (circles), and linear fits to it (lines). We use the linear relation for abundance matching to assign halo masses to the mock groups.

Figure 1.9 compares the true halo masses from the simulation with the final group masses obtained by our group finder using the K_s-band (r-band, for the SDSS and 2dFGRS) luminosity as the mass proxy. It is clear that the group finder performs quite well in assigning correct masses to groups over the whole range of halo mass for various samples. No significant bias is seen in the assigned mass for any of the samples. The horizontal stretching of the data points appearing at the massive end is due to the stacking of the simulation box and the small number of massive halos in the original simulation box. The true halo masses are exactly the same for some of the halos that are the duplicates of the same halo in the original simulation, but the group masses assigned to them can be different because they are located at different redshifts in the mock sample. For a given true mass, the typical scatter in the assigned mass is

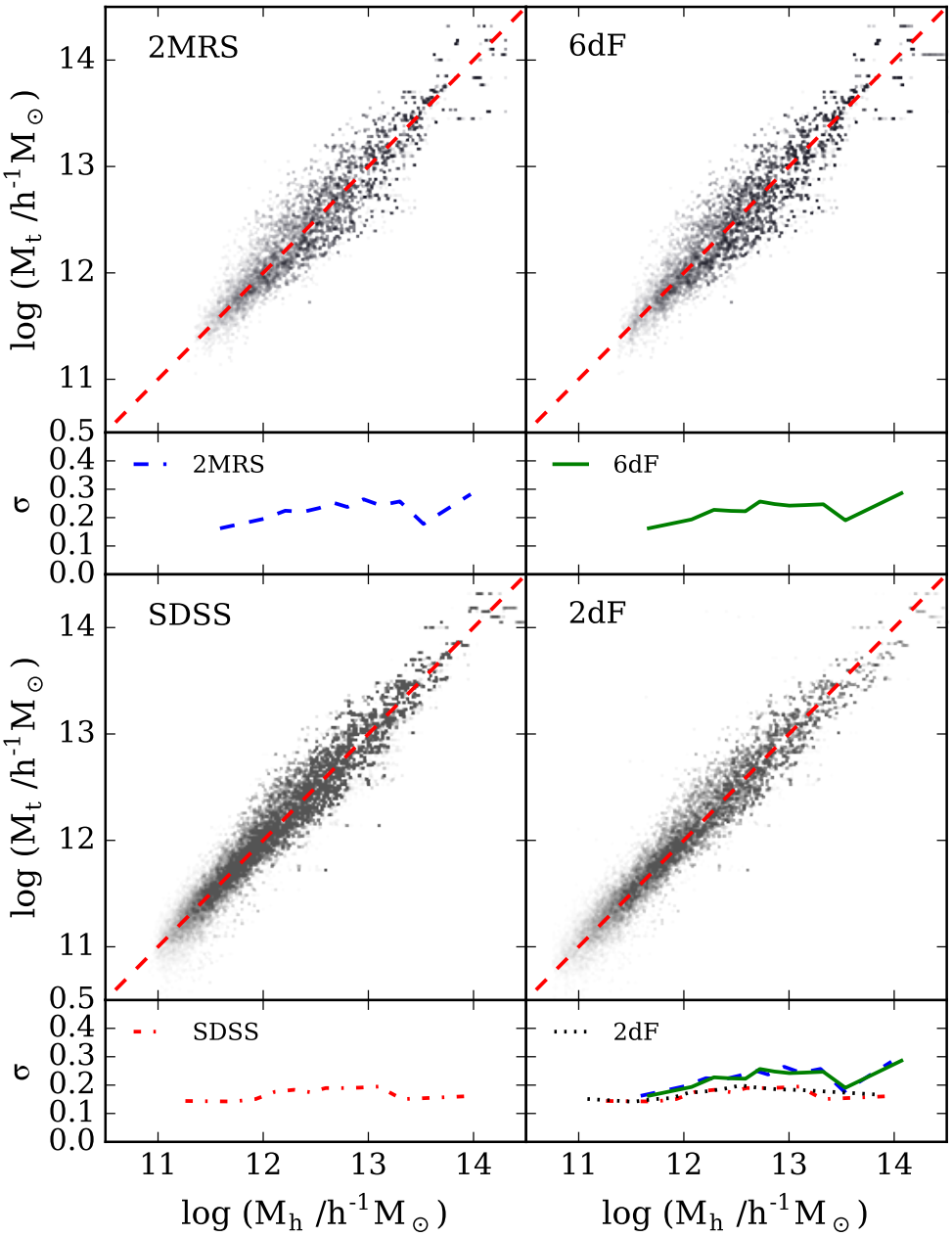

Figure 1.9. Comparison between true halo mass (vertical axis) and group mass identified by our group finder (horizontal axis) using luminosity as the proxy of halo mass for the mock samples of 2MRS, 6dFGS, SDSS, and 2dFGRS constructed from the EAGLE simulation (see text for the sample selections). The small rectangular panels plot the scatter of true halo mass at given group mass.

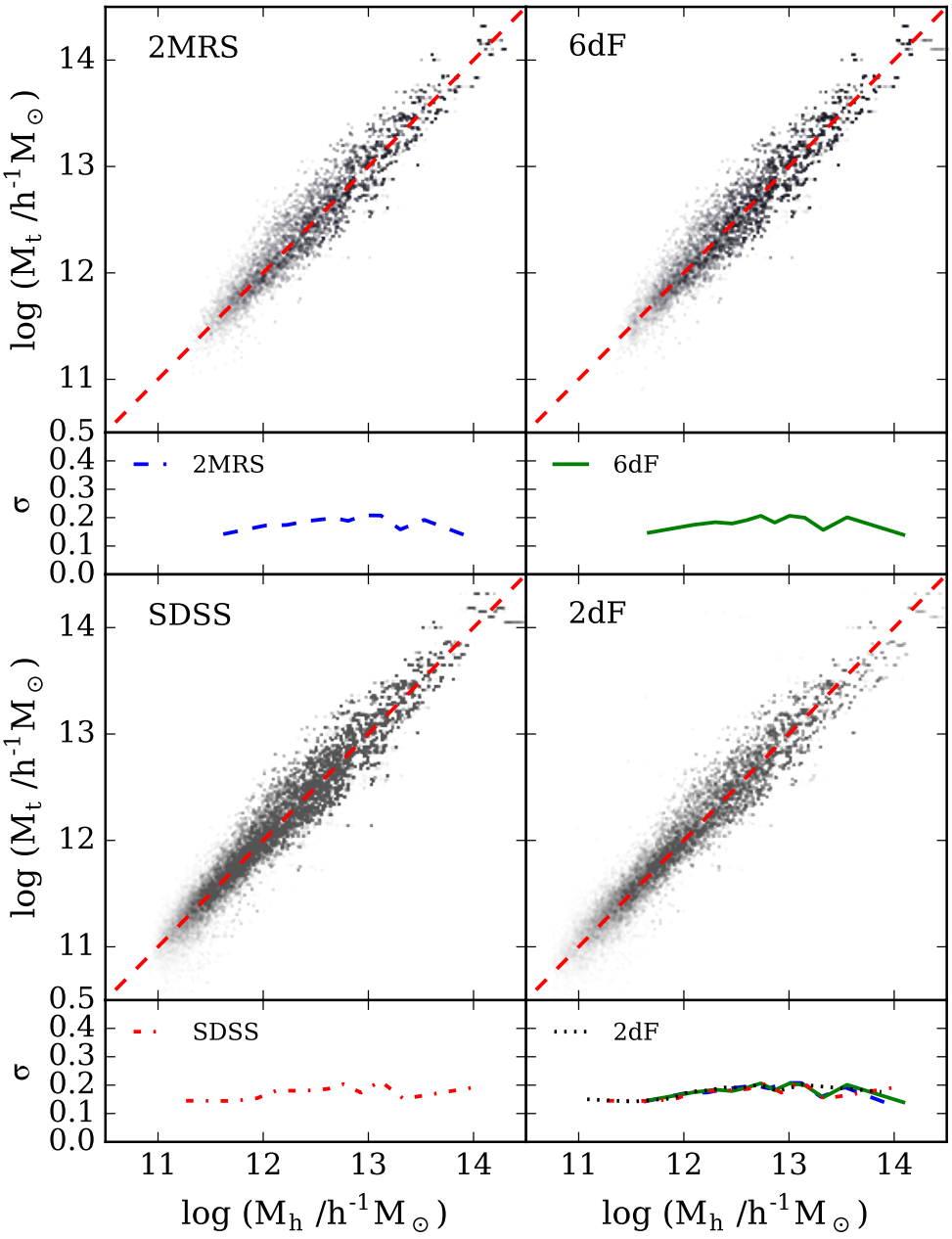

Figure 1.10. Same comparison as Figure 1.9 but using stellar mass as the proxy of halo mass.

~ 0.2 dex. The scatter is larger for the 2MRS and 6dFGS mock samples, reflecting the less tight K_s-luminosity vs. halo-mass relation than the r-luminosity vs. halo-mass relation in the simulation. Table 1.3 compares the total number of groups (halos)

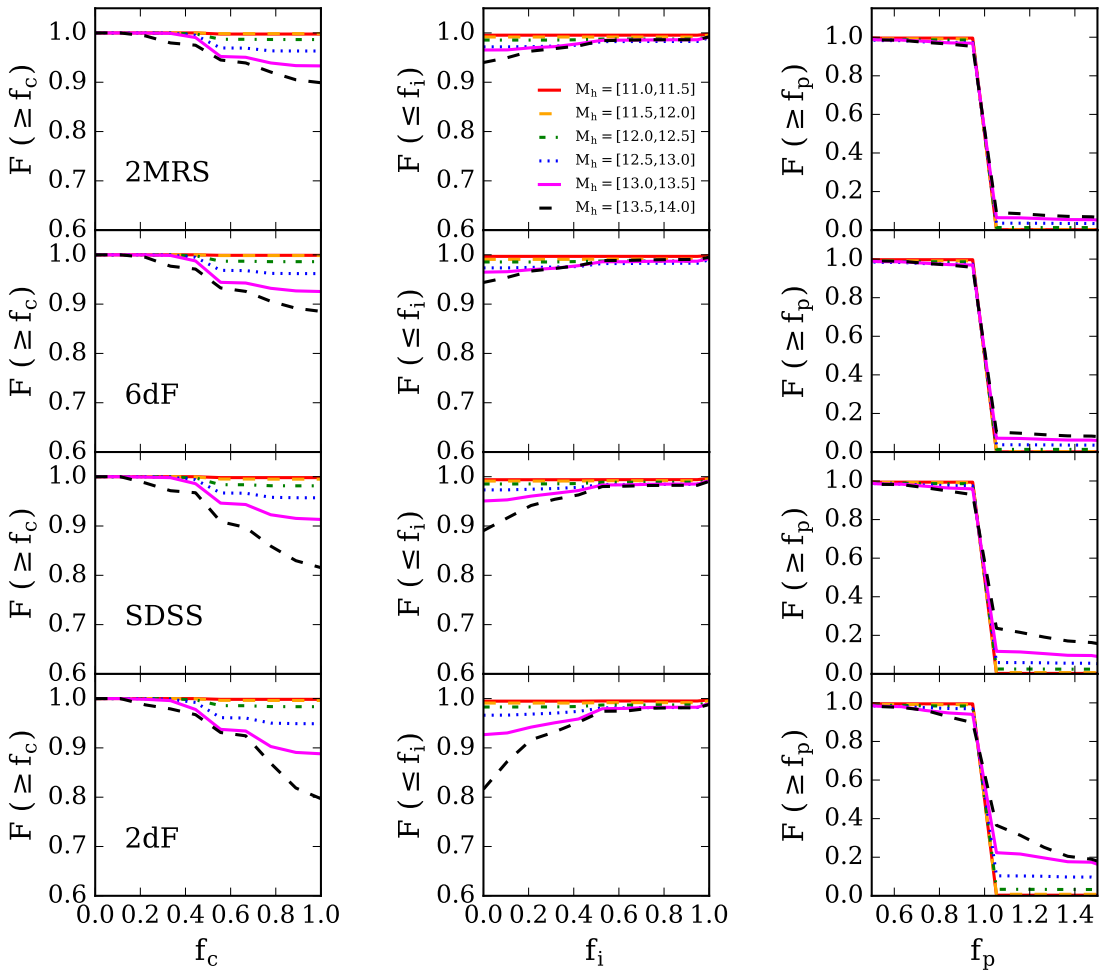

Figure 1.11. Membership assignments by the group finder applied to the mock samples, in terms of the completeness (left), contamination (middle), and purity (right). The vertical axis plots the cumulative fraction of the groups identified via the group finder, and the different lines are for halos of different masses as indicated.

and the number of groups (halos) of given richness between the mock group catalogs and the original simulation.

Figure 1.10 shows the same comparison of halo mass but obtained using stellar mass as the mass proxy for all surveys. It is seen that, unlike in Figure 1.9, the scatter in halo mass is almost identical for all surveys, and that stellar mass performs as a better mass proxy than the K_s-band luminosity by ~ 0.05 dex or more, as seen earlier

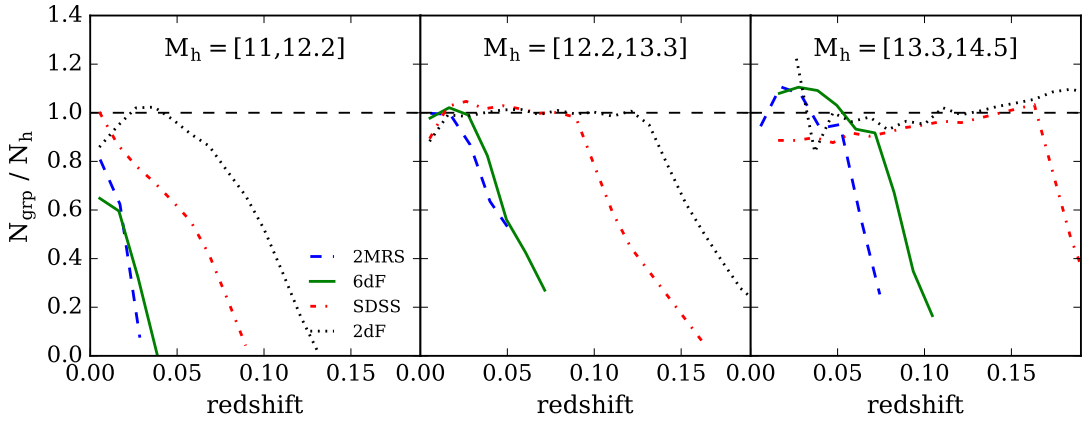

Figure 1.12. Global completeness, defined as the number of groups identified relative to the number of true halos, for the mock samples as a function of redshift for halos of different masses as indicated.

in Figure 1.5, while the r-band luminosity is an as good proxy as the stellar mass for the deeper surveys.

1.4.3 Completeness, contamination, and purity

In addition to group masses, comparisons are also made between the membership assignment by the group finder and the true membership given by the simulation. To do this, we first assume that each mock group identified corresponds to the simulation halo that is associated with the brightest member of the mock group. As a quantitative assessment of the membership assignment, we follow Yang et al. (2007) and define the following quantities,

- Completeness: $f_c \equiv N_s/N_t$;
- Contamination: $f_i \equiv N_i/N_t$;
- Purity: $f_p \equiv N_t/N_g$

where N_t is the total number of member galaxies of each halo from the simulation, N_s is the number of member galaxies of the corresponding mock group that are

true members of the simulation halo (thus $N_s \leq N_t$), N_i is the number of member galaxies of the mock group that are not true members of the simulation halo, and $N_g = N_i + N_s$ is the total number of members of the mock group. For a perfect group finder, $N_s = N_t = N_g$ and $N_i = 0$, and so $f_c = f_p = 1$ and $f_i = 0$.

Figure 1.11 shows the completeness, contamination, and purity for the mock groups of different masses. The 2MRS and 6dFGS mock samples appear to have better membership assignments than the deeper SDSS and 2dFGRS mock samples. This happens because, in the two shallower samples, larger fractions of groups have a single member galaxy, which by definition have perfect completeness and zero contamination. For the 2MRS and 6dFGS mocks, $\sim 90\%$ of all groups have completeness $\sim 100\%$, being lower for more massive halos. For the SDSS and 2dFGRS, about 85% (95%) of the groups have completeness $\geq 95\%$ ($\sim 70\%$). On the other hand, about 95% and 90% of the groups have zero contamination for the shallower two and deeper two surveys, respectively. Overall $80 - 90\%$ of the groups have purity between 0.95 and 1.05, indicating that there is only a 5% difference in the total number of members between the true and selected memberships. We also check the global completeness of the identified groups as a function of redshift, and the results are shown in Figure 1.12. As expected, it declines beyond the redshift to which halos of a given mass is complete.

1.4.4 Comparison with other group finders

As mentioned above, our group finder is built upon the group finders of Y05 and L16, but there are differences in details, especially in the halo mass proxies. Here we compare the performance of our group finder with respect to the earlier group finders by applying them to the same mock samples.

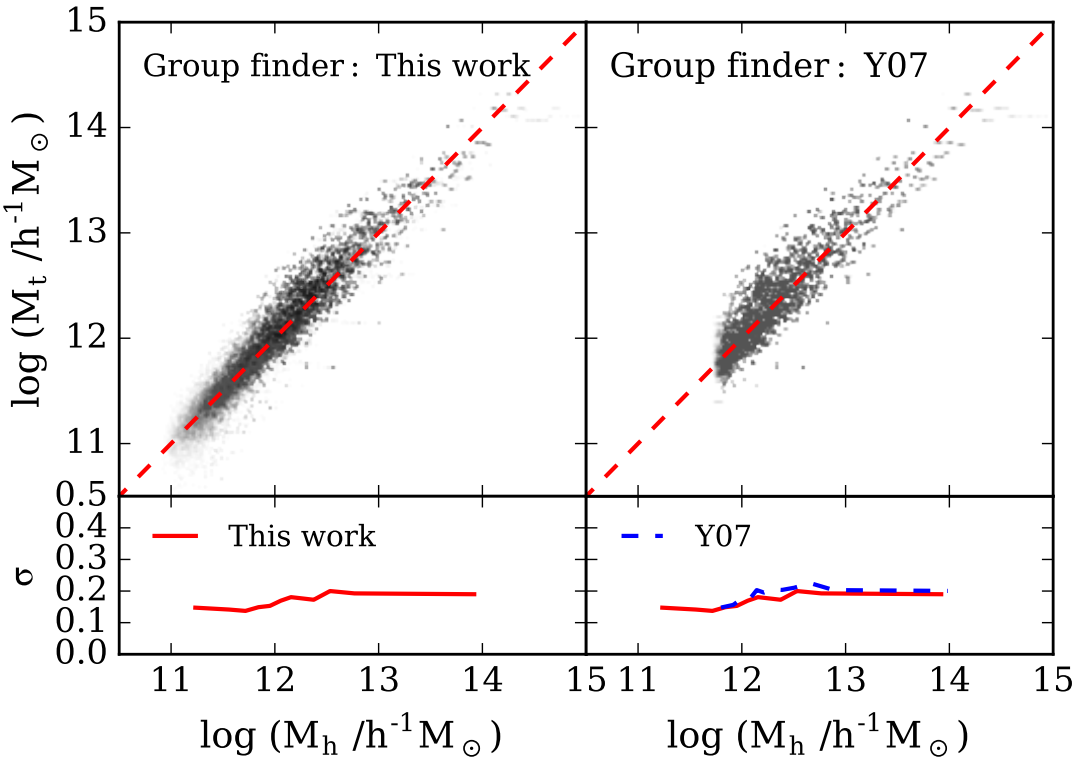

Figure 1.13. Comparison between the true halo mass of the EAGLE simulation and the group mass identified by our group finder (left) and by the group finder of Yang et al. (2007) (Y07; right) for the SDSS mock samples restricted to $z \leq 0.09$. The lower panels plot the scatter of the true halo masses at a given group mass.

1.4.4.1 Comparison with Yang et al.

In Yang et al. (2007) (Y07), the total group luminosity (group stellar mass) of member galaxies brighter than $M_r = -19.5 + 5\log(h)$ in the r-band was used as the proxy of the halo mass. These group luminosity and stellar mass will be denoted as $L_{19.5}$ and $M_{*,19.5}$, respectively. For galaxies at redshifts where the survey limit corresponds to an absolute magnitude brighter than the limit, Y07 used the observed luminosity function to account for the contribution to $L_{19.5}$ from the missing galaxies due to the magnitude limit. However, as found in Y07, this correction introduces uncertainties in the group masses. This is not surprising given that most groups identified have only a few member galaxies even for the SDSS, and an extrapola-

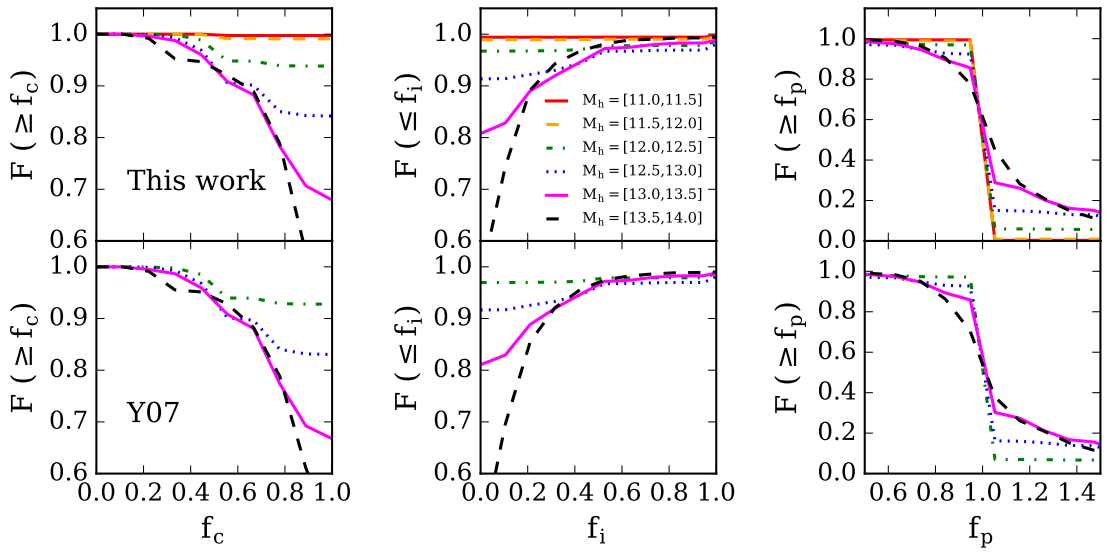

Figure 1.14. Comparison of membership assignment in terms of the completeness (left), contamination (middle), and purity (right) between our group finder (upper) and the group finder of Yang et al. (2007) (Y07; lower) for the same mock samples as Figure 1.13. The vertical axis plot the cumulative fraction of the groups, and the different lines are for haloes of different masses as indicated.

tion according to an average luminosity function is not expected to give an accurate estimate of $L_{19.5}$ ($M_{*,19.5}$) for individual groups. Y07 found that the uncertainty introduced by this is larger than that introduced by the group finder itself, and is comparable to the intrinsic scatter in the true halo mass at a given $L_{19.5}$ (or $M_{*,19.5}$). As a more demanding test of our group finder against that of Y07, we restrict the mock sample to $z \leq\sim 0.09$, the redshift limit to which the selection is complete to $M_r = -19.5+5\log(h)$ so that no extrapolation is needed in the group mass proxy used in Y07. The mock sample here is that constructed for the SDSS from the EAGLE simulation, as described in §1.2.5.

Figure 1.13 shows the group masses obtained from our group finder and the Y07 group finder with respect to the true halo masses of the simulation. When applying the Y07 group finder, the ranking of groups in $M_{*,19.5}$ is used to assign group masses, while our group finder uses the halo mass proxy (stellar mass based) as described

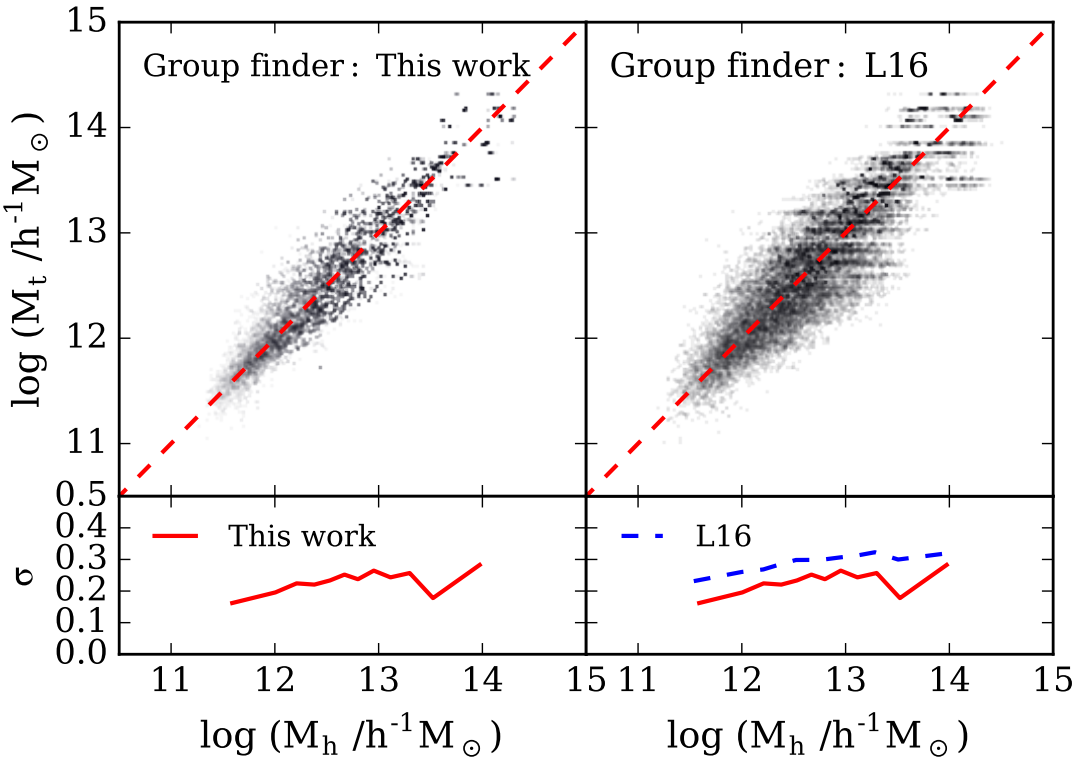

Figure 1.15. Comparison between the true halo mass of the EAGLE simulation and the group mass identified by our group finder (left) and by the group finder of Lu et al. (2016) (L16; right) for the 2MRS mock samples. The lower panels plot the scatter of the true halo masses at a given group mass.

in §1.3.2. As one can see, our group finder matches the true halo masses with an accuracy slightly higher than that of Y07, with scatter typically of $0.15 - 0.2\,\mathrm{dex}$. This indicates that Proxy-M and the GAP correction work as well as using $M_{*,19.5}$ to assign halo masses to groups. However, had we included groups at $z > 0.09$, where extrapolation is needed in Y07's group mass proxy, the scatter given by Y07 would become $0.25 - 0.3\,\mathrm{dex}$ while that given by our group finder remains at the level of $0.2\,\mathrm{dex}$. In addition, our group finder performs equally well even for halos with masses as low as $10^{11}h^{-1}\mathrm{M}_{\odot}$, about an order of magnitude lower than that reached by Y07. Many of these low mass halos contain only galaxies with M_r fainter than $-19.5 + 5\log h$, which are not assigned halo masses in the original Y07 method.

However, while the group finder of Y07 itself does not include these low-mass groups in the SDSS group catalog, halo mass assignment can be extended to lower masses by using a relation between halo mass and central galaxy, as given in, e.g., Yang et al. (2012). The number of groups identified by our group finder and the Y07 group finder are $180,835$ and $184,833$, of which $35,376$ and $32,343$ have more than one member, respectively. These are very close to the true number of halos of $177,013$, of which $35,439$ have more than one member.

Figure 1.14 shows the comparison of the two group finders in group completeness, contamination, and purity. For both of the group finders, the completeness decreases and the contamination increases with increasing halo masses, as we have seen in §1.4.3. The two group finders perform almost equally well in membership assignments.

1.4.4.2 Comparison with Lu et al.

L16 developed and calibrated their group finder with their 2MRS mock samples constructed from an empirical conditional luminosity function model (see L16 for details). Here we apply our group finder and that of L16 to our own 2MRS mock sample, and make comparisons in their performances. For both of the group finders, we adopt the functional forms given by equation (9) and equation (10), and the corresponding best parameters for the GAP correction for groups of more than one member. Otherwise we follow the methodology of L16 as closely as possible to reproduce their group finder. The major difference between the two group finders is in the prescription for isolated galaxies. While our group finder uses Proxy-L to assign halo masses, we follow the prescription of L16 for their group finder. There are a total of $29,464$ true halos, of which $5,158$ have more than one member, and the number of groups identified by our group finder (by L16) are $30,118$ ($29,968$), of which $4,980$ ($4,522$) have more than one member. Furthermore, our group finder identifies 879

groups with three members and 364 with four members. The corresponding numbers by L16 are 888 and 362.

The group masses obtained by the two group finders are compared to the true halo masses in Figure 1.15. One can see that our group finder reduces the overall scatter by ~ 0.1 dex relative to that given by L16. As the two group finders work in a similar way for groups of more than one member, the improvement in our group finder is mainly due to a better mass proxy for groups containing only one member. Note also that the mass proxy used by L16 is calibrated with a mock catalog constructed from the observed conditional luminosity functions in the r band and scaled to the K band using abundance matching, while our mass proxies are calibrated with the EAGLE simulation. Part of the difference may also be due to the different calibrations. The scatter we obtain here for the L16 group finder is very similar to that obtained in the original L16 paper from a completely different mock sample, suggesting that the test results are not particularly sensitive to the mock samples adopted for the test. This is also demonstrated in Appendix A, where it is shown that our group finder performs equally well for an independent mock sample constructed from an empirical model of galaxy formation.

1.5 The group catalogs

In this section, we present the group catalogs we construct by applying the group finder, as described in §1.3, to the observational samples described in §1.2. As mentioned earlier, we provide four catalogs for each observation sample:

1. a catalog constructed with galaxies that have spectroscopic redshifts, using Proxy-L to estimate halo masses;

2. a catalog constructed with galaxies that have spectroscopic redshifts, using Proxy-M to estimate halo masses;

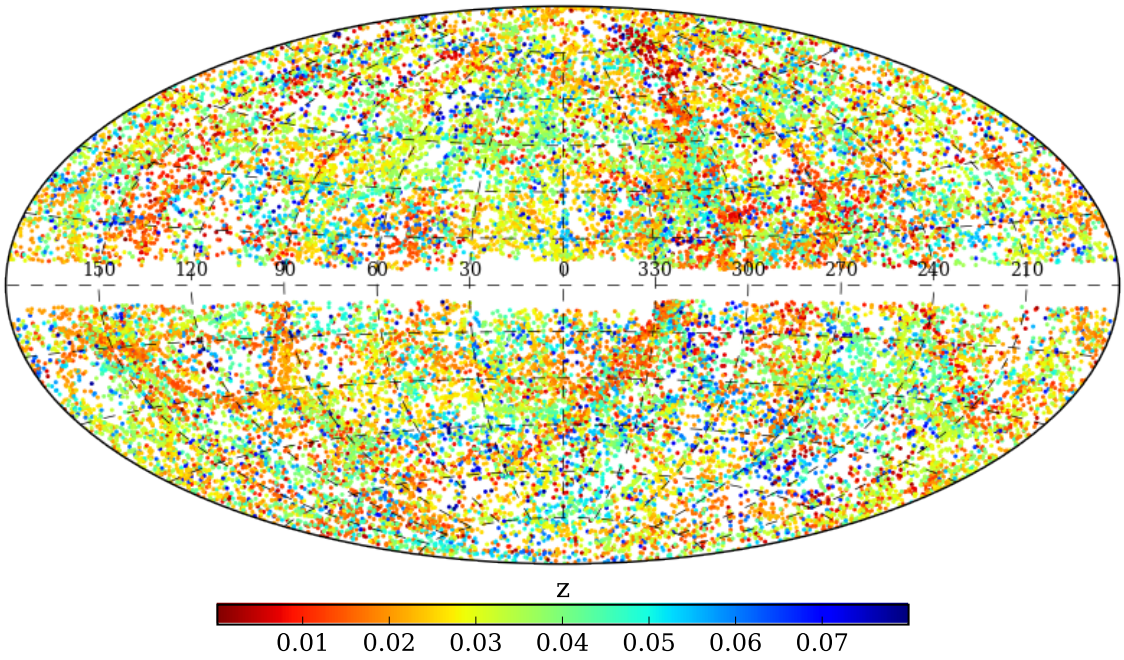

Figure 1.16. The group distributions in Galactic coordinates (Aitoff projection) of the 2MRS group catalog.

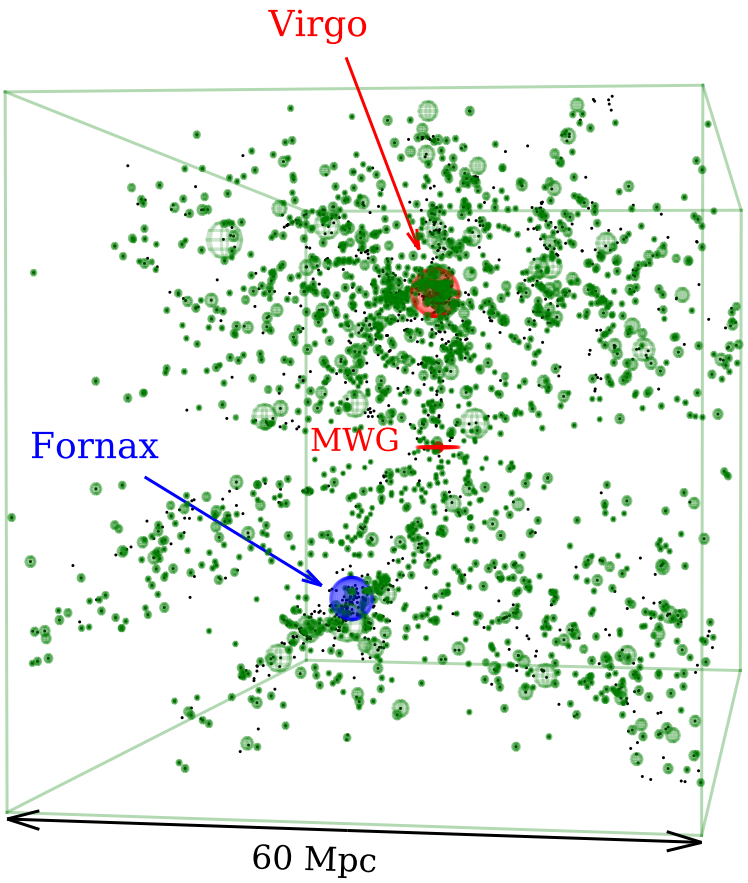

Figure 1.17. Three dimensional distribution of the 2MRS galaxies (black dots) and groups identified by the group finder (wire-framed green spheres with radii of r_{180}) in the local Universe with the Milky Way at the center.

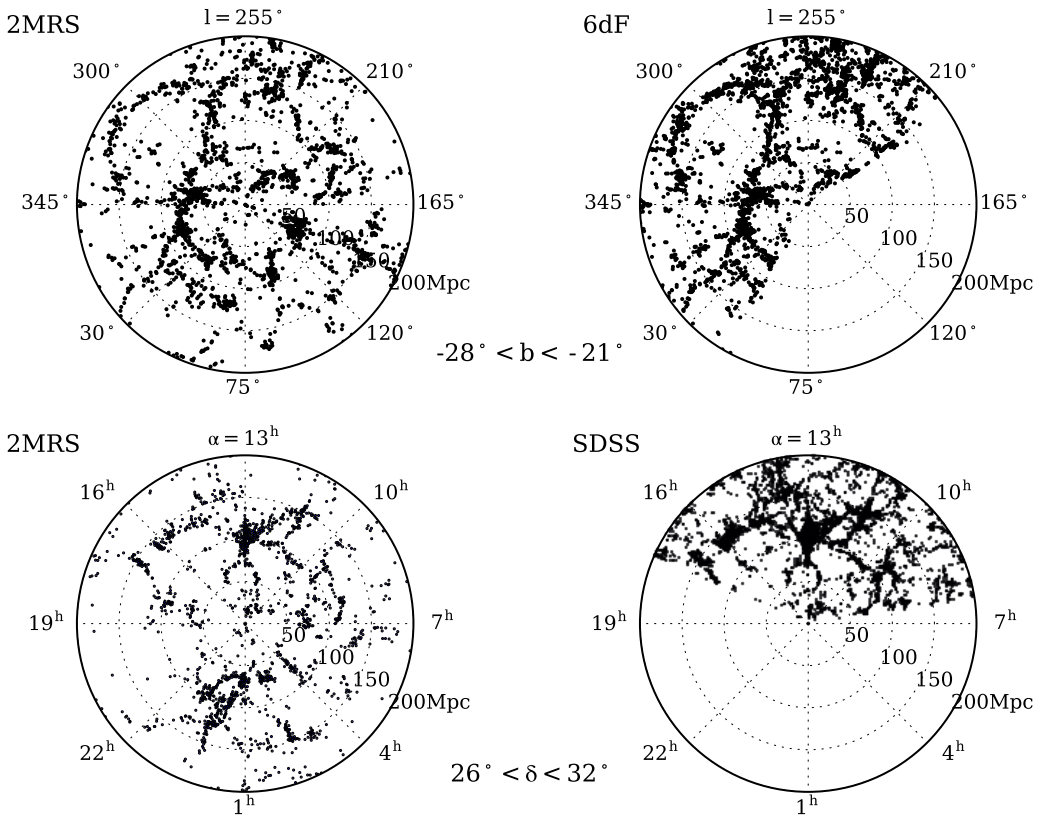

Figure 1.18. Comparison of the distribution of the groups constructed by the group finder in slices between different surveys. The upper and the lower panels show different slices as indicated.

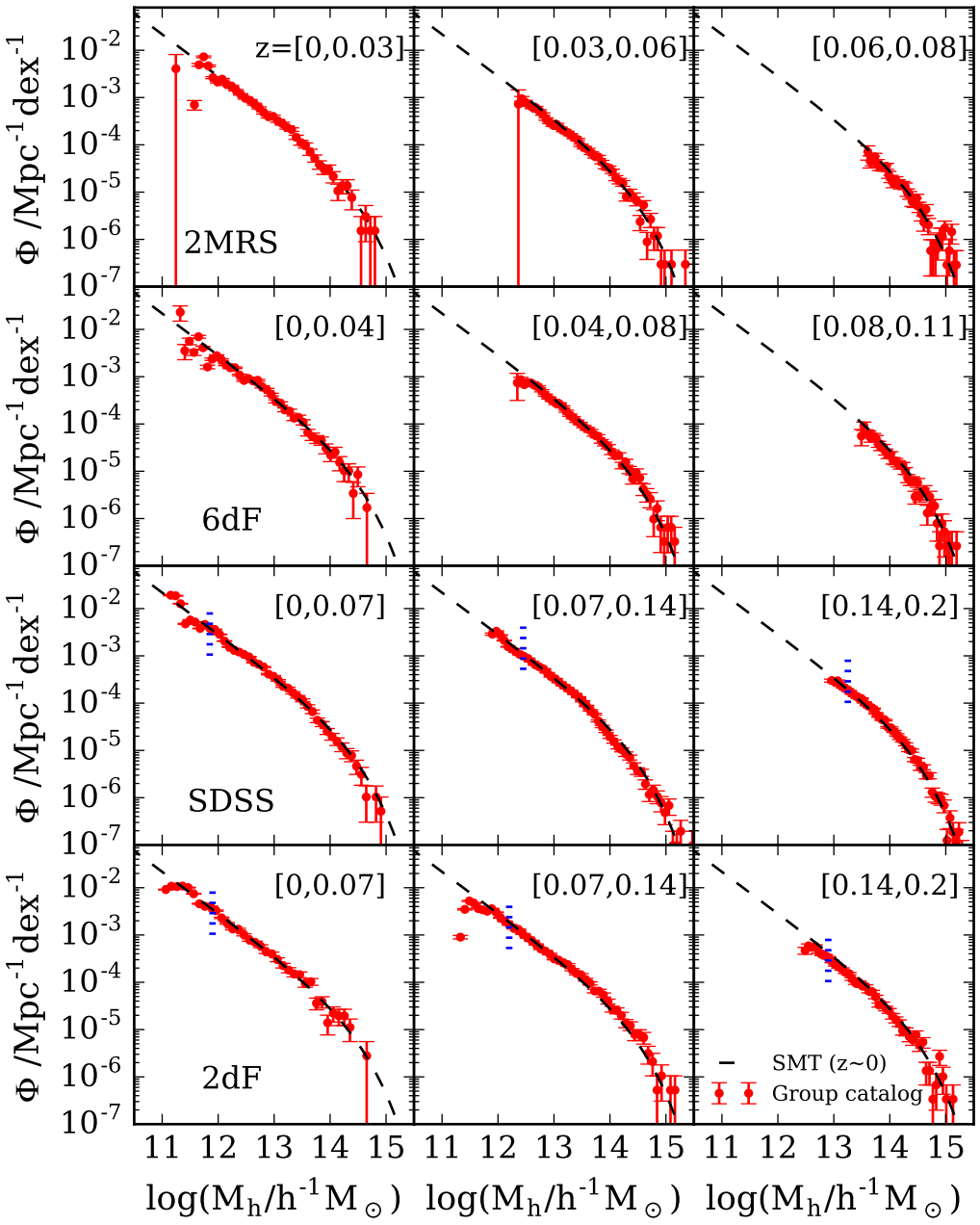

Figure 1.19. Halo mass functions of galaxy systems constructed based on the four surveys. The dashed lines are the theoretical mass function by Sheth et al. (2001), which we used for abundance matching for the group finder. The dotted short ticks in the lower two panels indicate the lower limits of the halo masses of the group catalogs by Yang et al. (2007) and Yang et al. (2005), respectively.

Table 1.4. A summary of group catalogs.

Catalog	Total galaxies	Total groups[a]	Total groups with reliable mass	$N^{b}=1$	$N\geq 2$	$M_h \geq 10^{14}\mathrm{M}_\odot/h$	$M_h \geq 10^{13}\mathrm{M}_\odot/h$
2MRS(L)	43,249	30,937	18,650	13,311	5,339	982	6,836
2MRS(M)	43,249	31,752	19,224	13,913	5,311	1,016	7,156
2MRS+(L)	44,310	31,804	18,731	13,275	5,456	984	6,877
2MRS+(M)	44,310	32,693	19,307	13,923	5,384	1,014	7,211
6dFGS(L)	62,987	46,676	17,907	11,126	6,781	1,004	6,919
6dFGS(M)	62,987	47,176	18,555	11,789	6,766	1,045	7,291
6dFGS+(L)	73,386	59,515	21,481	14,168	7,313	1,154	8,030
6dFGS+(M)	73,386	59,512	22,223	15,278	6,945	1,191	8,459
SDSS(L)	586,025	446,495	165,538	112,444	53,094	3,757	39,565
SDSS(M)	586,025	421,715	167,638	105,979	61,659	3,780	43,880
SDSS+(L)	600,458	453,927	164,694	107,066	57,528	3,712	39,464
SDSS+(M)	600,458	426,932	166,999	101,518	65,481	3,760	43,649
2dFGRS(L)	180,967	144,965	77,423	62,101	15,322	606	8,526
2dFGRS(M)	180,967	145,756	77,365	61,309	16,056	632	9,116
2dFGRS+(L)	189,101	147,757	77,861	59,606	18,255	634	8,553
2dFGRS+(M)	189,101	148,290	77,757	58,909	18,848	638	9,099

Notes.
a. This includes groups without reliable halo mass assigned because halo mass is not complete at given redshift.
b. The number of member galaxies in a group. This excludes groups without reliable halo mass assignment.

3. a catalog constructed with all galaxies, using Proxy-L to estimate halo masses;

4. and a catalog constructed with all galaxies, using Proxy-M to estimate halo masses.

For convenience we will use the name of the galaxy sample together with the halo mass proxy adopted to refer to a group catalog. For example, the catalogs constructed from the SDSS survey are referred to as SDSS(L), SDSS(M), SDSS+(L), and SDSS+(M), respectively. For brevity, the following presentations are mainly based on catalogs (i) and (ii), unless stated otherwise.

1.5.1 The catalogs and their basic properties

For SDSS and 2dFGRS, Proxy-L uses the r- and R-band luminosities, respectively, and Proxiy-M uses the stellar masses of galaxies as described in §1.2.3 and §1.2.4. For 2MRS and 6dFGS, Proxy-L is based on the K_s-band luminosities of galaxies. The Proxy-M for these two samples are based on the stellar masses obtained from the mean relation between the K_s-band luminosity and stellar mass from the EAGLE simulation. We use the same calibrations of the halo mass proxies as described in §1.3.2. Our tests show that it is not necessary to re-calibrate the mass proxies for individual samples, as the outcomes with and without such a re-calibration converge in the end. This is expected, because our group finder uses the mass proxies only to rank group masses, and the halo masses are re-adjusted at the end of each iteration using abundance matching. In the tables and figures shown in this section, we exclude groups that are not assigned halo masses by abundance matching because of sample incompleteness at the given halo mass and redshift (see §1.4.2). In the catalogs, however, we include these groups (with a flag), and assign them masses according to the mean relation between the halo mass and the mass proxy obtained from the last iteration of the group finder.

The distribution in the sky of the 2MRS groups selected by our group finder is shown in Figure 1.16, and Figure 1.17 shows the three dimensional distributions of galaxies and groups in the local Universe from the 2MRS catalogs. Also, the distributions of groups from different surveys in the same slices are compared in Figure 1.18.

Figure 1.19 compares the mass function of the halos from the group catalogs with the theoretical model of Sheth et al. (2001), which was used for abundance matching for the group finder. The good agreement between the observational data and the theoretical model is largely by design. However, the plots do show the halo-mass and redshift ranges covered by different samples, as well as the statistical uncertainties in the number densities of groups.

Table 1.4 lists the total number of groups, as well as the number of groups of given richness and halo masses selected from different samples. Figure 1.20 shows in more detail the distributions of groups with respect to richness (number of member galaxies), halo mass, and redshift. Note, again, that these distributions are obtained from groups for which halo masses are complete at a given redshift, as shown in Figure 1.8. It is seen that the results from (L) and (M) catalogs are consistent with each other within the Poisson uncertainties. For comparison, we also show the results for the +(L) catalogs [the results for the +(M) catalogs are very similar] as the small dots. As one can see, results from the extended (+) catalogs are consistent with the non-extended catalogs, except for the 6dFGS which has the poorest completeness in spectroscopic redshifts. Also, some massive clusters in the catalogs have only one galaxy particularly for 2MRS and 6dFGS, because of their shallow depths that make satellites not observable.

Figure 1.21 compares halo masses (based on Proxy-M) for individual groups cross-identified between the group catalogs. While we do not present the comparison for the 2dFGRS because the number of such groups is small, we did check that the

mean relation and scatter for the 2dFGRS are similar to those for the SDSS. We used the tolerances of angular separation less than 10 arcsec and $|\Delta z| \leq 10^{-3}$ for the cross-identifications. One can see that there is a very tight correlation in halo masses between the 2MRS and the 6dFGS group catalogs, while a larger dispersion of $0.2-0.3$ dex is found between the 2MRS and SDSS, mainly because of the differences in the stellar mass estimates.

1.5.2 Comparison with other catalogs

Here we compare our catalogs with a number of other catalogs in the literature, including the 2MRS catalogs of L16 and Tully (2015) (T15), the SDSS catalog of Y07, and the 2dFGRS catalog of Y05.

1.5.2.1 Comparison of the 2MRS group catalog with L16

As mentioned earlier, L16 built a group catalog based on the same sample of the 2MRS galaxies as ours using a similar methodology. Figure 1.20 shows the comparison between the two catalogs in the number of groups as functions of richness, mass, and redshift, and it is clear that the two catalogs are in good agreements. We also checked the mass of individual haloes for groups that are cross-identified between the two catalogs, and found that our mass assignments are in general agreement with those of L16 with a typical dispersion of ~ 0.25 dex between them.

As described earlier, our group finder is different from that of L16 in two ways. The first is that we re-calibrated the gap-based mass model of L16, so that the mass assigned to a group may be different from that of L16 even if it has the same membership of galaxies. Second, L16 used a 'Gap limit' prescription to assign masses for groups containing only one member (see §1.3.2.2), while our group finder does not. We believe that these are the sources of the dispersions and discrepancies found in the comparison of halo mass between the two catalogs.

1.5.2.2 Comparison of the 2MRS group catalog with T15

Tully (2015) (T15) constructed a 2MRS group catalog using an empirical relation between halo mass (and the corresponding size and velocity dispersion) and a characteristic group luminosity to assign galaxies into groups. Figure 1.20 shows the comparison of our catalog with theirs in the number of groups as functions of richness, mass, and redshift. As T15 stated that their group catalog is less reliable outside the recession velocity range between $3,000$ and $10,000$km s^{-1}, we make comparisons only for the groups within the velocity range. One can see that T15 contains more massive clusters than our catalogs, while the richness and redshift distributions are in better agreements. T15 compared the mass function of their groups with a theoretical halo mass function and found that, although the shape of their group mass function is similar to that of the theoretical function, the normalization is about a factor of 4.6 higher. As mentioned above, T15 used an empirical model for their group masses, which is different from the mass proxies used in our group finder. Furthermore, their definition of halo masses is also different from ours. All these produce the differences seen between the two catalogs.

1.5.2.3 Comparison of the SDSS group catalog with Yang et al.

Y07 built a group catalog of the SDSS DR7 galaxies. As we described earlier, their group finder is similar to ours in that it uses halo mass and velocity dispersion of groups identified to update galaxy memberships at each iteration until its iteration reaches convergence in the membership assignments. The main difference is that it uses a summed stellar mass (or luminosity) of member galaxies brighter than $M_r = -19.5+5\log h$ as a halo mass proxy. However, as the group catalogs are dominated by groups containing one galaxy (see Table 1.4), it results in no significant net difference except that our group catalog extends to lower mass. Figure 1.20 compares the SDSS group catalog of Y07 with that given by our group finder. The Y07 catalog contains

smaller number of low-mass systems at relatively low redshifts, mainly because of the magnitude limit of $M_r = -19.5 + 5\log h$ adopted by Y07 in their halo mass proxy, which is brighter than the observational flux limit at $z <\sim 0.09$.

1.5.2.4 Comparison of the 2dFGRS group catalog with Yang et al.

Y05 also constructed a group catalog of the 2dFGRS galaxies. Again, their group finder is similar to ours but differs in that Y05 uses a summed stellar mass of member galaxies brighter than $M_{b_J} = -18 + 5\log h$ as a halo mass proxy instead of stellar mass (luminosity) of central galaxy and the n-th most massive (brightest) galaxy, which our group finder uses. The sample selection for the 2dFGRS is almost identical to our sample selection. Figure 1.20 shows comparisons between the two catalogs in the number of groups of given richness, mass, and redshift. The lower number of groups in Y05 in low-mass end and low-redshift is again because of the limit of $M_{b_J} = -18 + 5\log h$ used by Y07 for the halo mass proxy, which is brighter than the flux limit in the observation at $z <\sim 0.12$. Otherwise, the agreement between the two catalogs is reasonably good.

1.5.3 Contents of the catalogs

The group catalogs list the properties of groups, while the galaxy samples present not only the properties of galaxies but also their links to groups. Object indexes are also provided for galaxies so that one can identify them from the original galaxy catalogs. As mentioned above, there are four group catalogs for each galaxy sample, and so in total we provide 16 group catalogs, as summarized in Table 1.4. Tables 1.5 and 1.6 show the structures of the catalogs we provide, using the 2MRS as an example. In what follows we explain the different columns in more detail.

1.5.3.1 The group catalogs

The following items are provided for individual groups.

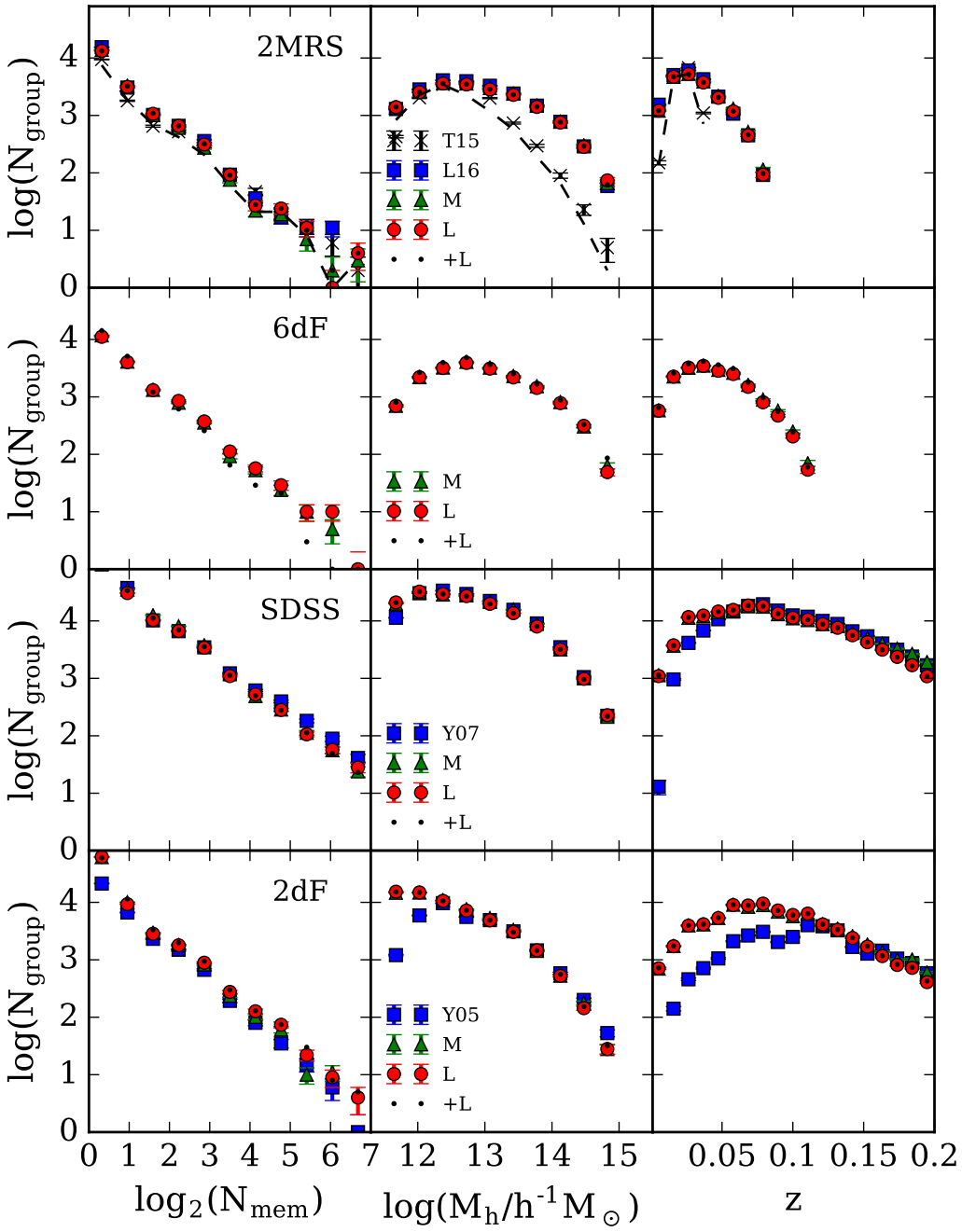

Figure 1.20. The number of groups as a function of the number of members, halo mass, and redshift for L (circle) and M (triangle) catalogs for each survey, compared with Tully (2015) (T15; cross) and Lu et al. (2016) (L16; square) for the 2MRS, with Yang et al. (2007) (Y07; square) for the SDSS, and with Yang et al. (2005) (Y05; square) for the 2dFGRS. The results from the +(L) catalogs (dots) are also shown for comparison. The T15 results (crosses) should be compared with the dashed lines, which are obtained by using only groups with recession velocities between $3,000$ and $10,000$km s^{-1}, within which the T15 sample is complete. The comparison is only made for groups that halo mass is complete at a given redshift. The error bars shown represent Poisson errors.

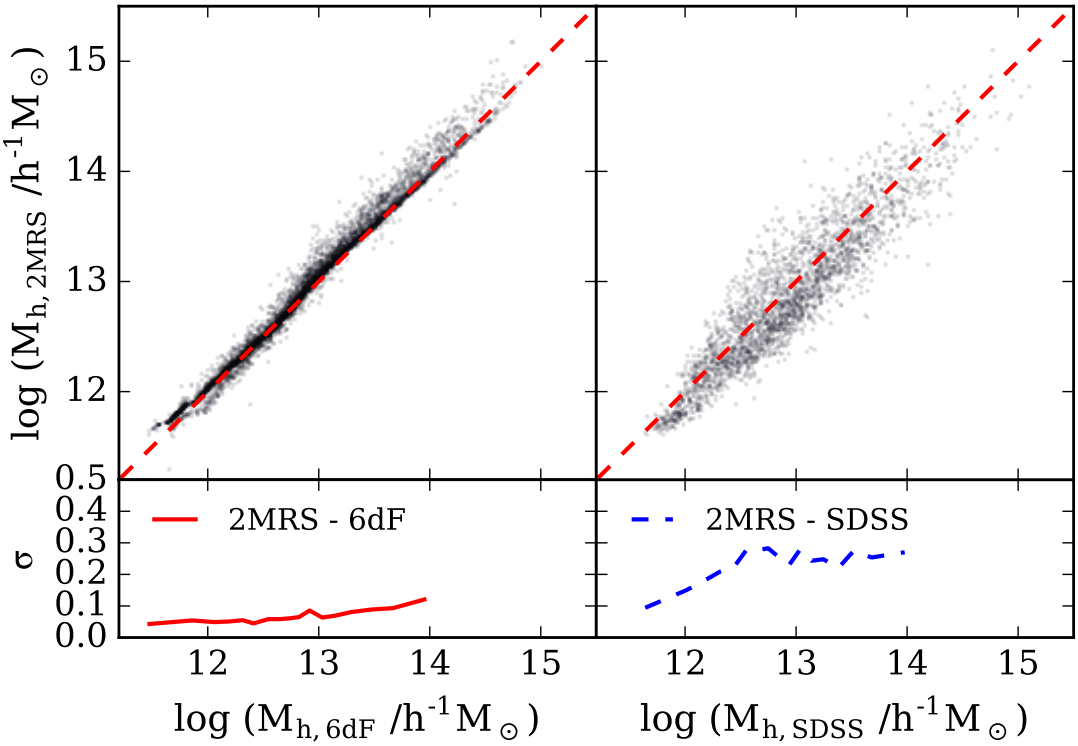

Figure 1.21. Comparison of halo masses (based on Proxy-M) for individual groups cross-identified between different group catalogs. We used the tolerances of ≤ 10 arcsec and $|\Delta z| \leq 10^{-3}$ for the cross-identification. The lower panels plot the scatter of the 2MRS halo masses at a given mass from the other catalogs.

Table 1.5. The 2MRS group catalog.[a]

(1) group ID	(2) cen ID	(3) ra (deg)	(4) dec (deg)	(5) z	(6) $\log(M_h/h^{-1}\mathrm{M}_\odot)$	(7) N_{mem}	(8) f_{edge}	(9) i-o	(10) known as
1	15	187.74899	12.20402	0.00362	14.290	109	1.00	1	Virgo
2	486	243.52157	−60.79748	0.01663	14.366	106	1.00	1	Norma
3	530	49.47894	41.53527	0.01748	14.297	92	1.00	1	Perseus
4	672	194.81308	27.97206	0.02476	14.639	88	1.00	1	Coma
5	386	192.33072	−41.18290	0.01437	14.342	62	1.00	1	Centaurus
6	1094	258.01088	−23.30095	0.03030	14.758	52	1.00	1	Ophiuchus
7	18	53.13952	−35.56552	0.00492	14.097	48	1.00	1	Fornax
8	257	210.67554	−33.83448	0.01527	14.534	47	1.00	1	
9	608	17.31215	32.68145	0.01609	14.042	42	1.00	1	
10	432	207.35991	−30.43195	0.01628	14.220	41	1.00	1	

Notes.
a. The full catalog is available at `http://gax.sjtu.edu.cn/data/Group.html`.

Column (1) group ID: an unique ID of a group within a given group catalog;

Column (2) cen ID: galaxy ID of the central galaxy of a group in the corresponding galaxy sample;

Column (3) ra (in degrees): right ascension (J2000) of the luminosity-weighted (for catalogs using Proxy-L) or mass-weighted (for catalogs using Proxy-M) group center;

Column (4) dec (in degrees): declination (J2000) of the group center;

Column (5) z: redshift of group center in the CMB rest-frame;

Column (6) $\log(M_h/h^{-1}\mathrm{M}_\odot)$: 10-based logarithm of the halo mass of a group in units of $h^{-1}\mathrm{M}_\odot$;

Column (7) $N_{\rm mem}$: number of member galaxies in a group;

Column (8) $f_{\rm edge}$: the volume fraction that is not cut out from the halo of a group (assumed to be spherical) by the survey boundary or mask;

Column (9) i-o: A flag that indicates whether a group is inside or outside the region of completeness for a given halo mass. For a group inside the completeness region (value = 1), mass is obtained directly from the abundance matching. For a group that is outside the completeness region (value = 0), mass is estimated using the relation between the halo mass and its proxy from the last iteration of the group finder.

Column (10) known as: conventional name of a system, identified only for well-known massive clusters.

1.5.3.2 The galaxy catalogs

The following items are provided for individual galaxies.

Table 1.6. The 2MRS galaxy catalog.[a]

(1) galaxy ID	(3) group ID	(4) ra	(5) dec (deg)	(6) l (deg)	(7) b (deg)	(8) $z_{\rm CMB}$ (deg)	(9) $z_{\rm EDD}$	(10) $z_{\rm comp}$	(11) $z_{\rm src}$	(12) $\rm dist_{NN}$ (deg)	(13) $\log L$ ($h^{-2}\rm L_\odot$)	(14) $\log M_*$ ($h^{-2}\rm M_\odot$)
1	25128	10.68	41.26	121.17	−21.57	−0.00194	0.00018	1	0	0	10.644	10.286
2	1116	11.88	−25.28	97.36	−87.96	−0.00013	0.00086	1	0	0	10.771	10.432
3	345	148.88	69.06	142.09	40.90	0.00016	0.00085	1	0	0	10.749	10.406
4	299	201.36	−43.01	309.51	19.41	0.00267	0.00083	1	0	0	10.69	10.339
5	299	196.36	−49.46	305.27	13.34	0.00268	0.0	1	0	0	11.497	11.413
6	30717	23.46	30.65	133.61	−31.33	−0.00152	0.00021	1	0	0	9.439	9.201
7	345	148.96	69.67	141.40	40.56	0.00094	0.00082	1	0	0	10.394	10.02
8	29389	56.70	68.09	138.17	10.57	−0.0002	0.00053	1	0	0	10.12	9.754
9	5266	204.25	−29.86	314.58	31.97	0.00263	0.00115	1	0	0	10.691	10.34
10	12365	189.99	−11.62	298.46	51.14	0.00455	0.00228	1	0	0	11.146	10.906

Notes.
a. The full catalog is available at `http://gax.sjtu.edu.cn/data/Group.html`.

Column (1) galaxy ID: unique ID of galaxies within each sample. This can be used to match galaxies across the galaxy and group catalogs;

Column (2) survey ID: ID of galaxies from the original survey data release. This can be used to match galaxies across our catalogs and the original surveys;

Column (3) group ID: ID of the group of which a galaxy is a member;

Column (4) ra (in degrees): right ascension (J2000);

Column (5) dec (in degrees): declination (J2000);

Column (6) l (in degrees): Galactic longitude;

Column (7) b (in degrees): Galactic latitude;

Column (8) $z_{\rm CMB}$: redshift in the CMB rest-frame. This is used for the group finder;

Column (9) $z_{\rm EDD}$: redshift for nearby galaxies based on the EDD distances. Otherwise equals to 0. This is only used for converting apparent magnitude to luminosity;

Column (10) $z_{\rm comp}$: redshift completeness along the direction on the sky where a galaxy lies;

Column (11) $z_{\rm src}$: a numerical value indicating the source of $z_{\rm CMB}$. As the sources vary for different samples, please refer to the individual catalogs for more detailed descriptions;

Column (12) $\rm dist_{NN}$: angular separation to the nearest neighbor (deg) for galaxies that $z_{\rm src}$ is the nearest neighbor. Otherwise equals to 0.

Column (13) $\log(L/h^{-2}{\rm L}_\odot)$: 10-based logarithm of the luminosity in units of $h^{-2}{\rm L}_\odot$. Luminosities are in the K_s-band for 2MRS and 6dFGS, in the r-band for SDSS, and in the R-band for 2dFGRS. K- and evolutionary

corrections to $z = 0.1$ are made following Lavaux & Hudson (2011) (for 2MRS and 6dFGS) and Poggianti (1997) (for SDSS and 2dFGRS). All quantities are calculated with the assumption of the WMAP9 cosmology;

Column (14) $\log(M_*/h^{-2}\mathrm{M}_\odot)$: 10-based logarithm of the stellar mass in units of $h^{-2}\mathrm{M}_\odot$. Please refer to the relevant sections for how the stellar masses are estimated in different samples;

Column (15) color: provided only for SDSS $(g-r)$ and 2dFGRS $(b_J - R)$.

1.6 Summary

In this paper, we have constructed group catalogs from four large redshift surveys in the low-z universe: the 2MRS, 6dFGS, SDSS, and 2dFGRS. The groups are identified with a halo-based group finder that is based on the group finders developed in Y05, Y07 and L16 but has improved halo mass assignments that can be applied uniformly to various observations. The group finder uses stellar mass or luminosity of central galaxies combined with the luminosity/stellar mass gap between the central galaxy and the n-th brightest/most massive satellite as halo mass proxies. It assigns galaxies into groups using halo properties, such as halo size and velocity dispersion, and iterates with updated halo properties until the membership converges. We use an abundance matching technique to assign final halo masses to individual groups selected. For groups that are not assigned mass by abundance matching, due to the fact that they are outside the redshift limit within which groups of a given mass is complete, halo masses are assigned based on the mean relation between halo mass and its proxy obtained from the last iteration of the group finder.

We have used realistic mock galaxy samples constructed from a hydrodynamical simulation (EAGLE) to test the performance and to calibrate our group finder, and

used another set of mock samples constructed from an empirical model of galaxy formation as an independent check. The tests showed that our group finder can find $\sim 95\%$ of the 'true' member galaxies for about 95% (85%) of the groups for the 2MRS and 6dFGS samples (for the SDSS and 2dFGRS samples), with better membership assignment for lower mass halos. The tests on mock samples also showed that the halo masses of individual groups estimated by the group finder are consistent with the true halo masses, with scatter of ~ 0.2 dex. The scatter in the estimated mass - true mass relation obtained here for the SDSS sample is similar to Y07, but it extends uniformly to halo masses that are about 0.7 dex lower.

We have constructed group catalogs by applying our group finder to the real redshift surveys of galaxies. From each survey, two samples of galaxies are constructed, one using only galaxies with spectroscopic redshifts, and the other using all galaxies, including the ones with redshifts estimated from nearest neighbors or from photometry (photometric redshifts). For each galaxy sample, two group catalogs are constructed, one using the luminosity-based halo mass proxy (Proxy-L) and the other using the stellar mass-based halo mass proxy (Proxy-M). Thus, we provide a total of 16 group catalogs, four different sets of catalogs for each of the four surveys. A summary of the all the group catalogs and how to use them are presented in §1.5. We have also described some of the basic properties of the group catalogs, such as the distributions in richness, redshift, and mass. Comparisons are made with other similar catalogs in the literature.

It should be noted that the group catalogs constructed are cosmology dependent, and we have adopted WMAP9 cosmology in the present paper. This dependence comes from both the properties of dark matter halos (halo size and velocity dispersion as functions of halo mass) adopted in grouping galaxies into common halos, and the halo mass function used in abundance matching. However, as demonstrated in Y07, the grouping of galaxies into groups is not sensitive to the cosmological model, unless

the adopted model is very different from that favored by current observations. The cosmology dependence in the halo mass assignments is also not a significant problem, as it is straightforward to convert the masses to other cosmologies with abundance matching.

CHAPTER 2

AN OBSERVATIONAL PROXY OF HALO ASSEMBLY TIME AND ITS CORRELATION WITH GALAXY PROPERTIES[1]

We show that the ratio between the stellar mass of central galaxy and the mass of its host halo, $f_c \equiv M_{*,c}/M_{\rm h}$, can be used as an observable proxy of halo assembly time, in that galaxy groups with higher f_c assembled their masses earlier. Using SDSS groups of Yang et al., we study how f_c correlates with galaxy properties such as color, star formation rate, metallicity, bulge to disk ratio, and size. Central galaxies of a given stellar mass in groups with $f_c > 0.02$ tend to be redder in color, more quenched in star formation, smaller in size, and more bulge dominated, as f_c increases. The trends in color and star formation appear to reverse at $f_c < 0.02$, reflecting a downsizing effect that galaxies in massive halos formed their stars earlier although the host halos themselves assembled later (lower f_c). No such reversal is seen in the size of elliptical galaxies, suggesting that their assembly follows halo growth more closely than their star formation. Satellite galaxies of a given stellar mass in groups of a given halo mass tend to be redder in color, more quenched in star formation and smaller in size as f_c increases. For a given stellar mass, satellites also tend to be smaller than centrals. The trends are stronger for lower mass groups. For groups more massive than $\sim 10^{13}{\rm M}_\odot$, a weak reversed trend is seen in color and star formation. The observed trends in star formation are qualitatively reproduced by an empirical model

[1] THE CONTENTS OF THIS CHAPTER ARE PUBLISHED IN LIM ET AL. 2016, MNRAS, 455, 499.

based on halo age abundance matching, but not by a semi-analytical model tested here.

2.1 Introduction

In the current standard ΛCDM model, dark matter halos form through gravitational instability - induced hierarchical clustering, and galaxies are believed to form at the centers of dark matter halos through cooling and condensation of baryonic gas (e.g. Mo et al. 2010, for a review). The formation and evolution of galaxies are, therefore, expected to be closely linked to the assembly history of their host halos. There have been continuous efforts to establish the connections between galaxies of different properties and dark matter halos using empirical models, such as halo occupation distribution (HOD) (e.g. Jing et al. 1998, Peacock & Smith 2000, Seljak 2000, Soccimarro et al. 2001, Berlind & Weinberg 2002, Zheng et al. 2007, Leauthaud et al. 2012, Watson et al. 2012), conditional luminosity function (CLF) (Yang et al. 2003, van den Bosch et al. 2007), and halo abundance matching (HAM) (Mo et al. 1999, Kravtsov et al. 2004, Vale & Ostriker 2004, 2006, Conroy & Wechsler 2009, Guo et al. 2010, Neistein et al. 2010, Watson et al. 2012, Kravtsov 2013). The CLF and HOD models assign galaxies into dark matter halos predicted by a given cosmology, so that the predicted galaxy population matches the observed luminosity (stellar mass) functions and spatial clustering properties of galaxies. The HAM approach, on the other hand, populates galaxies into halos and sub-halos, assuming that there is a roughly monotonic correspondence between the ranking orders of the luminosities (or stellar masses) of galaxies and those of the masses of dark matter halos.

Most of the studies based on these approaches have so far focused on using the mass of halos to link galaxies with halos, thus implicitly assuming that galaxy properties are determined by halo mass alone. In reality, however, other properties of halos, such as assembly history, spin, and shape, may also play an important role in

galaxy formation and evolution. These halo properties, therefore, should also be used in understanding the relationships between galaxies and halos.

In this paper, we investigate how the properties of galaxies of a given stellar mass are correlated with the assembly time of their host halos. To this end, we propose an observational proxy of halo assembly time motivated by the results of Wang et al. (2011). Using high-resolution N-body simulations, Wang et al. investigated a large number of halos properties, such as formation time, substructure fraction, spin and shape, and their correlations among themselves and with large scale environments. Most of these halo properties are, unfortunately, not directly observable, and so it is difficult to test directly their effects on galaxy formation with observational data. One exception is the sub-structure fraction, which is defined as $f_s = 1 - (M_{\rm main}/M_{\rm h})$, where $M_{\rm h}$ is the mass of the halo, and $M_{\rm main}$ is the mass of the main sub-halo located at the center of the host halo. This quantity is found to be correlated tightly with many other halo properties, in particular the formation time, spin and shape. More importantly, this quantity may be estimated from observations. Indeed, with a well-defined galaxy system, such as a galaxy group selected with the halo-based group finder of Yang et al. (2005), a good proxy of $M_{\rm main}$ is $M_{*,c}$, the stellar mass of the central galaxy in a group according to halo-galaxy abundance matching, and $M_{\rm h}$ can be estimated from the total stellar mass of the group, as demonstrated in Yang et al. (2005, 2007). Thus, one can use $f_c \equiv M_{*,c}/M_{\rm h}$ as an observational proxy of the assembly time of the host halo of the group, and study how galaxy properties change with f_c. The goal of the present paper is to use this proxy to study the correlations between galaxy properties and the assembly time of their host groups (halos).

This paper is organized as follows. In Section 2.2, we describe the observational galaxy catalogs from which our galaxy and group samples are selected. In Section 2.3 we demonstrate how f_c can be used as a reliable proxy of halo assembly time. Detailed analyses of the correlations between galaxy properties and f_c of their host groups are

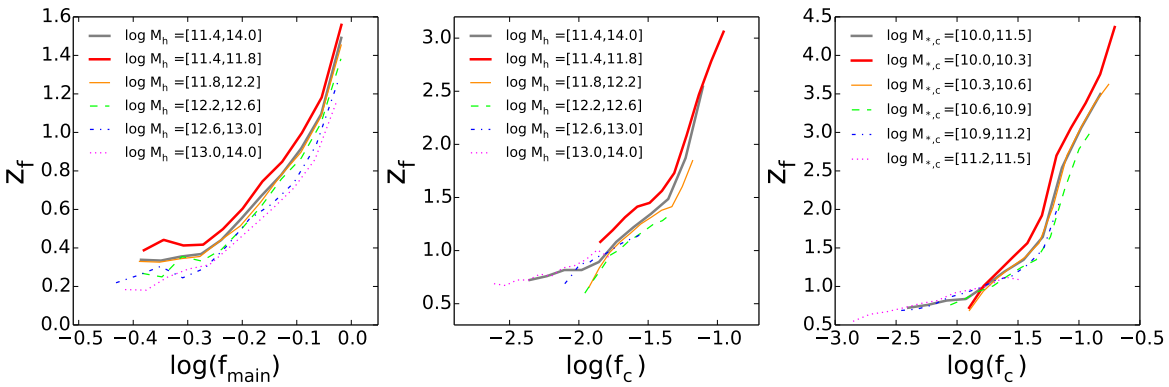

Figure 2.1. A demonstration how f_c can be used as a proxy of halo assembly time. *Left* : The correlation between half-mass assembly time z_f and $f_{\rm main} = M_{\rm main}/M_{\rm h}$ (median) obtained from N-body simulations, based on data published in W11, where $M_{\rm main}$ is the mass of the most massive sub-halo in each host halo. Results are shown for halos in five mass ranges, as indicated. For comparison, the result for the total halo sample is shown as the gray line. *Middle* : The correlation between z_f and $f_c \equiv M_{*,c}/M_{\rm h}$ (median), where $M_{*,c}$ is the stellar mass of the central (most massive) galaxy, obtained from the mock galaxy catalog of Hearin & Watson (2013), constructed using an age abundance matching model combined with halos from the Bolshoi N-body simulation. Different curves denote different host halo mass bins, as indicated. The results for the total sample is shown as the gray line. *Right* : The same as the middle panel, except that different curves show different stellar mass bins of central galaxies, as indicated. Here again the result for the total sample is shown as the gray line for comparison.

presented in Section 2.4, and a preliminary comparison of our results with models is made in Section 2.5. Finally, in Section 2.6, we summarize our main conclusions.

2.2 Observational data

2.2.1 SDSS galaxies

The galaxy samples used in this paper are obtained from the Sloan Digital Sky Survey (SDSS). Specifically, the galaxy catalog, as described in Wang et al. (2012) (W12 hereafter) and publicly available at `http://gax.shao.ac.cn/data/Group.html`, is constructed from the New York University Value-Added Galaxy Catalogue (NYU-VAGC; Blanton et al. 2005), which is based on SDSS Data Release 7 (SDSS DR7;

Abazajian et al. 2009), but updated with a set of improvements over the original pipeline. From this catalog, we select all galaxies in the Main Galaxy Sample with extinction-corrected apparent r-band magnitude brighter than 17.72, with redshifts in the range $0.01 \leq z \leq 0.20$, and with redshift completeness $C_z > 0.7$. This leaves 639,359 galaxies in total, with a sky coverage of 7,748 deg^2. Of these, 599,301 galaxies have redshifts from the SDSS DR7, 2,450 galaxies with redshifts from the 2dFGRS (Colless et al. 2001), 819 with redshifts from the Korea Institute for Advanced Study Value-Added Galaxy Catalogue (KIAS-VAGC; Choi et al. 2010), 36,759 galaxies with redshifts from their nearest neighbors (since they do not have spectroscopic redshift measurements due to fiber collisions), and 30 galaxies with redshifts from ROSAT X-ray clusters. We exclude galaxies with assigned redshifts that have ${}^{0.1}M_r - 5\log h \leq -22.5$ to prevent fiber-collided galaxies with real redshifts much lower than the nearest neighbors so that their luminosities are vastly over-estimated. The catalog also contains, for each galaxy, the $(g-r)$ and other colors, which are all $K+E$-corrected to $z = 0.1$. In the following, this catalog will be referred to as the SDSS DR7 catalog to distinguish it from other catalogs we use in our study.

For all galaxies, we adopt stellar masses (M_*) from the data release of Brinchmann et al. (2004), available at `http://www.mpa-garching.mpg.de/SDSS/DR7/`. The data release also provides star formation rates (SFRs), and specific star formation rates (sSFRs, defined to be SFR divided by M_*). The SFRs are obtained by fitting the SDSS spectra with a spectral synthesis model. Specifically, Hα luminosities are used for star forming galaxies and the D4000 breaks are used for galaxies without significant emission lines. Gas phase metallicities [for example, oxygen abundance, in terms of log(O/H)] are also available for a fraction of the galaxies, as described in Tremonti et al. (2004). In total, about 6% of the galaxies in the SDSS DR7 catalog are missing in the Brinchmann et al. data release, most of which are fiber-collided galaxies missing spectra. The number of galaxies for which a given quantity is actually available varies

from quantity to quantity. For example, gas phase metallicity is available only for emission line galaxies.

2.2.2 Disk-bulge decomposition

We also make use of the results of Simard et al. (2011) obtained from bulge-disk decompositions of galaxies, which fit each galaxy image with the sum of a pure exponential disk and a de Vaucouleurs bulge using GIM2D. The code returns parameters such as the total flux, the bulge to total ratio B/T, the bulge half-light radius R_{50} and the disk scale length $R_{\rm disk}$. In this paper we use the results based on the r-band images. About 92% of our SDSS DR7 galaxies can be cross identified in Simard et al.'s data base.

2.2.3 Information from the Galaxy Zoo

The Galaxy Zoo is a project in which volunteers are asked to classify images of over 900,000 SDSS DR7 galaxies into six morphological categories. The Galaxy Zoo 2 (GZ2 hereafter; Willett et al. 2013), the successor of the original Galaxy Zoo, continued the spirit of the original project but asking volunteers much more detailed morphological questions such as the number of spiral arms, tightness of the arms, etc. To enable such detailed questions, GZ2 uses a subsample of the brightest 25% of the resolved galaxies in the SDSS North Galactic Cap region within the redshift range of $0.0005 < z < 0.25$ along with a few more selection criteria (see Willett et al. 2013). This leaves a grand total of 245,609 SDSS DR7 galaxies.

The SDSS metadata for GZ2 (available at `http://data.galaxyzoo.org/`) adds a series of useful information for SDSS DR7 galaxies, in particular, morphological classifications made by volunteers' votes. Whenever 'ellipticals' or 'spirals' are seen in our following analyses, the classification is based on GZ2. Out of all galaxies cross-matched between SDSS DR7 and GZ2, 97,785 are ellipticals and 135,634 are spirals.

2.2.4 SDSS groups

Given that galaxy groups are defined as galaxies that reside in the same dark matter halo, galaxy groups can be used to directly probe the connections between galaxies and their host halos. Yang et al. (2005, 2007) have developed a halo-based group finder optimized for grouping galaxies in common dark matter halos. The performance of this group finder has been tested extensively using mock galaxy redshift surveys constructed from CLF models (Yang et al. 2003, van den Bosch et al. 2003, Yang et al. 2004) and from a semi-analytical model (Kang et al. 2005). It was found that this group finder is more successful than the traditional friends-of-friends (FoF) algorithm in grouping galaxies into their common dark matter haloes (see Yang et al. 2007, (Y07 hereafter)). The group finder performs consistently even for very poor systems such as isolated galaxies in small mass haloes, which enables its suitability to probe the galaxy-halo connection over a wide range of different haloes.

In the present paper, we use the DR7 group catalog, publicly available at `http://gax.shao.ac.cn/data/Group.html` to associate galaxies with groups. This catalog is made basically by applying exactly the same group finder of Y07 to SDSS DR7 galaxies. The details of the group finder is described in Y07. *WMAP*5 cosmology was used to calculate distances from redshifts and to assign halo masses to selected groups. We adopt the group catalog 'modelC', which uses model magnitudes rather than Petrosian magnitudes. For each group in the group catalog, the fraction, f_{edge}, of each group's volume that falls inside of the SDSS DR7 survey volume is given. Only groups with $f_{edge} \geq 0.6$ are used here, which removes about 1.6% of all groups.

The group halo masses $M_{\rm h}$ in the catalog are estimated using the ranking of groups either in the combined luminosity ($L_{19.5}$) or in the combined stellar mass ($M_{\rm stellar}$) of all member galaxies with ${}^{0.1}M_r - 5\log h \leq -19.5$. The conversion from $L_{19.5}$ or $M_{\rm stellar}$ to $M_{\rm h}$ is made by adopting the halo mass function of Tinker et al. (2008) and the method of abundance matching assuming one-to-one correspondence

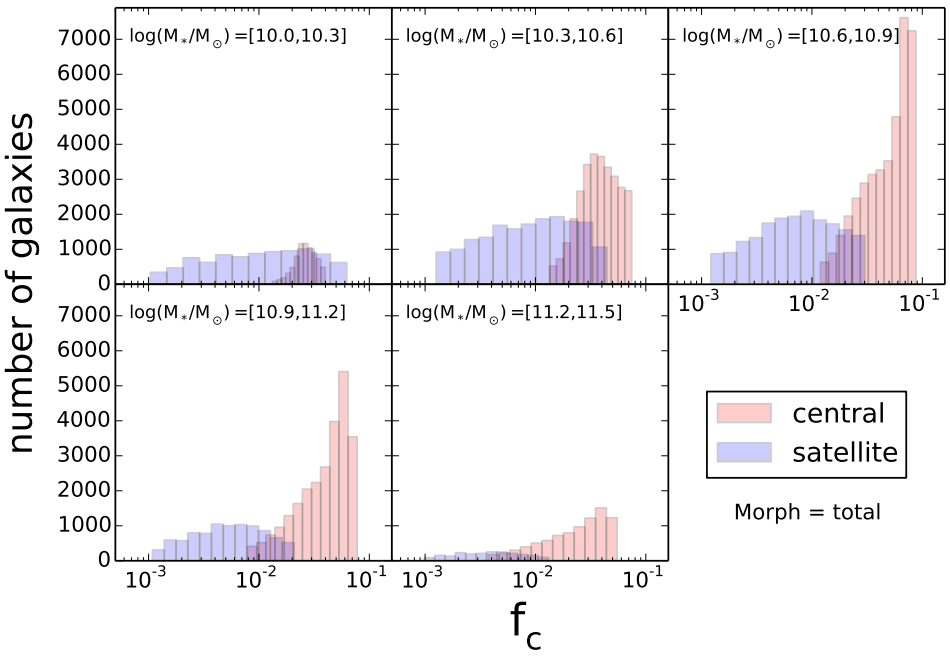

Figure 2.2. The number distribution of galaxies in f_c, with each panel corresponding to different stellar mass bins, as indicated on the top of each panel, for centrals (red) and satellites (blue).

between $L_{19.5}$ ($M_{\rm stellar}$) and $M_{\rm h}$. As shown in Y07, while both $L_{19.5}$ and $M_{\rm stellar}$ are tightly correlated with $M_{\rm h}$, the $M_{\rm h}$ - $M_{\rm stellar}$ relation is slightly tighter, with a typical dispersion of ≈ 0.2 dex in $M_{\rm h}$ for a given $M_{\rm stellar}$ over the halo mass range considered here. We therefore use $M_{\rm h}$ based on $M_{\rm stellar}$, although our tests showed that using $L_{19.5}$ does not change any of our results. For very small groups, no masses are assigned, and they are excluded from our analysis.

The identification of central galaxies for each group is also provided in two different ways: the brightest galaxy or the most massive galaxy in terms of stellar mass. In this study, we choose the latter as the definition of centrals.

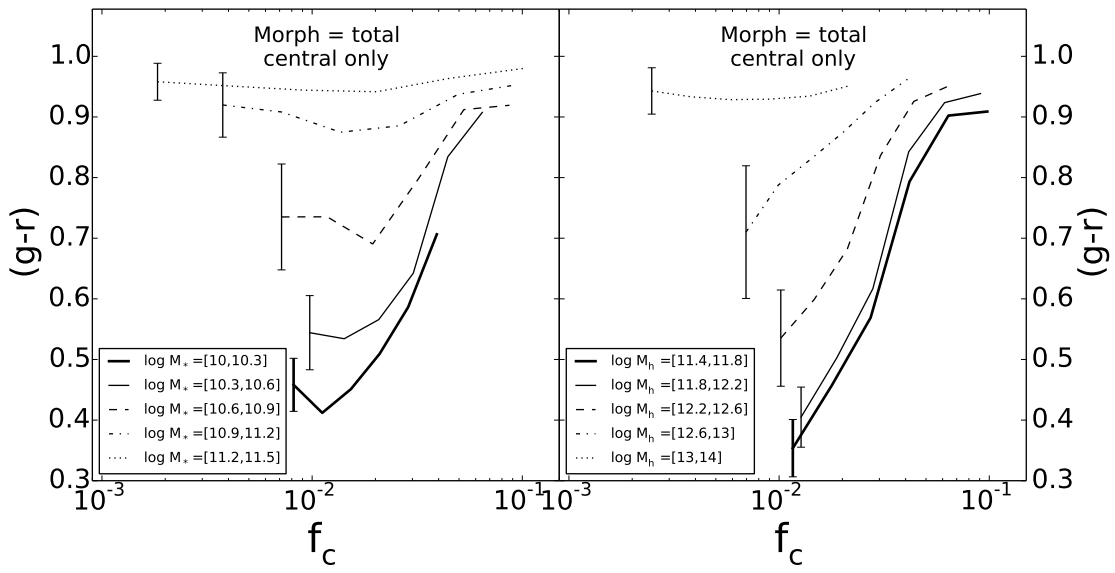

Figure 2.3. The correlation between $(g-r)$ color, $K+E$ corrected to $z=0.1$, and f_c, for centrals. The curves plot the median values in f_c bins. The error bars on the leftmost sides are 'typical' $[16\%, 84\%]$ ranges for each mass bin. In the left panel different curves refer to galaxies in different stellar mass bins, as indicated, while in the right panel different curves are for galaxies residing in halos in different halo mass bins.

2.3 An observational proxy of halo assembly time

As mentioned in the introduction, Wang et al. (2011) (hereafter W11) explored the correlations among various halo properties using dark matter halos identified from high-resolution N-body simulations. One of the most important properties of a halo is its formation time, z_f, which is defined to be the redshift at which the main progenitor of the halo has first assembled half of its final mass. This formation time is believed to have significant impact on the properties of the galaxies the halo hosts, such as galaxy age, color, star formation rate (SFR), etc. Unfortunately, z_f itself is not directly observable, and so it is not possible to examine the correlation between z_f of a halo and the properties of the galaxies the halo host. However, as shown in figure 1 of W11, and reproduced in the left panel of Figure 2.1, the halo formation time z_f shows a tight correlation with the sub-structure fraction, $f_s = 1 - f_{\rm main}$ with

$f_{\rm main} = (M_{\rm main}/M_{\rm h})$, where $M_{\rm main}$ is the mass of the main sub-halo at the center of each host halo, quite independent of the mass of the host halo. This suggests that $f_{\rm main}$ can be used as a proxy of z_f. Since $M_{\rm h}$ can be estimated for halos using halo abundance matching, as described in the previous section, and $M_{\rm main}$ can be estimated using sub-halo abundance matching, we can define an 'observable' quantity,

$$f_c \equiv \frac{M_{*,c}}{M_{\rm h}}\,, \tag{2.1}$$

as a proxy of z_f. Here $M_{*,c}$ is the stellar mass of the central galaxy obtained from the rank of $M_{\rm main}$. If there were no scatter in the halo-galaxy abundance matching, so that there is a one-to-one relation between galaxy stellar mass and sub-halo mass, $M_{*,c}$ would be a perfectly faithful indicator of $M_{\rm main}$. By definition $M_{*,c}$ would also be the stellar mass of the most massive galaxy in a group because the main sub-halo is the most massive one among all sub-halos. In reality, however, the halo mass - galaxy mass relation may not be one-to-one. Given this and that f_s is not perfectly correlated with z_f, f_c defined above can only be used as a proxy of z_f. As an illustration, the middle and right panels of Fig. 2.1 shows the correlation between z_f and f_c obtained from the HAM model of Hearin & Watson (2013) applied to dark matter halos in a high-resolution N-body simulation. As one can see, there is a tight correlation between f_c and z_f both for halos of a given mass (middle panel) and for centrals of a given stellar mass (right panel). In particular, the z_f - f_c relation does not seem to depend strongly on halo mass or on galaxy mass, although massive systems extend further towards the low-f_c end because of the fact that $M_{*.c}$ only increases slowly with halo mass at the massive end (e.g. Yang et al. 2012). All these validate the use of f_c as an observational proxy of z_f.

In what follows we will examine how galaxy properties are correlated with f_c, and use the results to understand the connection between galaxy properties and halo assembly histories as represented by the formation redshift z_f. For reference, we show

the distribution of galaxies in f_c for the entire SDSS DR7 sample in Figure 2.2. Each panel corresponds to a given stellar mass bin, as indicated in individual panels, and results are shown for both centrals and satellites. As expected, the centrals, defined to be the most massive ones in groups, have on average a higher f_c value than satellites, since groups with lower f_c tend to have more satellites in them.

2.4 Correlation of galaxy properties with f_c

This section examines the correlation of galaxy intrinsic properties with the value of f_c of the host group in which the galaxy resides. Results will be shown separately for central and satellite. While our presentation includes all groups, our test using only groups with more than one member galaxies brighter than ${}^{0.1}M_r - 5\log h = -19.5$ gives qualitatively similar results.

2.4.1 Central galaxies

2.4.1.1 Color and star formation

Figure 2.3 shows the correlation between the $(g-r)$ color of central galaxies and f_c of their host groups. In the left panel, results are shown separately for galaxies in five different stellar mass ranges, as indicated in the inner panel, while in the right panel results are shown separately for five different halo mass bins. The lines are the median values within narrow f_c bins, while the bars present the typical $[16\%, 84\%]$ range of the distribution in the corresponding halo mass or stellar mass range.

As one can see from the right panel, for a given halo mass, the $(g-r)$ color depends strongly on f_c, with centrals in halos with higher f_c being redder, except for the most massive halos, where the centrals are all equally red. Note that for halos with masses below $10^{12.6}\mathrm{M}_\odot$, the dependence of color on halo mass is not strong for a given f_c. By definition, for a given halo mass, f_c is directly proportional to $M_{*,c}$, and it is well known that the intrinsic properties of galaxies depend strongly on their stellar mass.

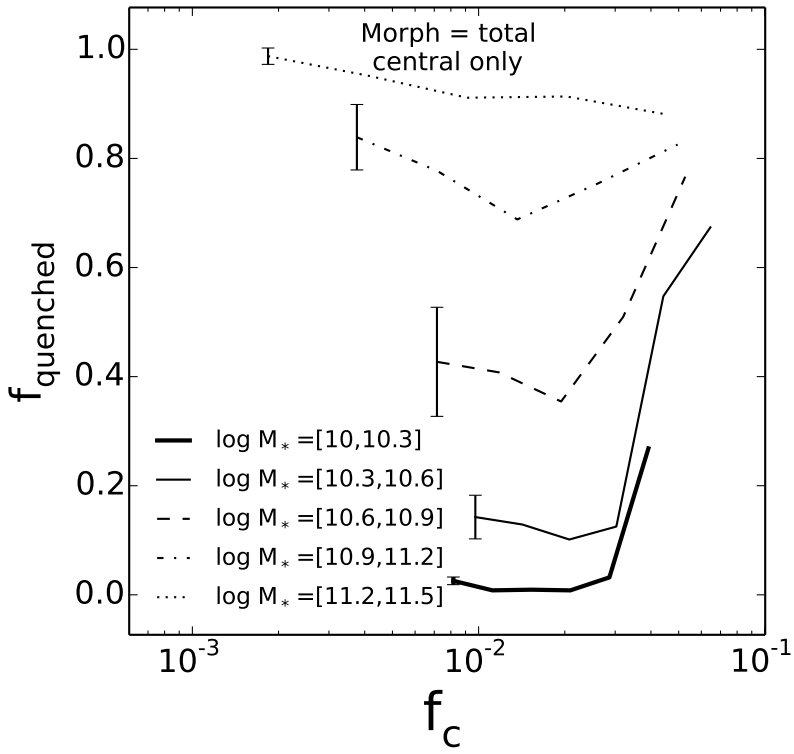

Figure 2.4. The correlation between the fraction of quenched galaxies and f_c for central galaxies. Quenched galaxies are defined to be the ones with star formation rate lower than the devision line defined by equation (2.2). The curves plot the quenched fractions in f_c bins. Different curves refer to galaxies in different stellar mass bins, as indicated. The error bars here are 'typical' 1-σ dispersions among 100 bootstrap re-sampling.

Thus, the strong dependence of color on f_c for a given halo mass bin see here is not surprising. However, given that f_c is strongly correlated with halo formation time (see Fig. 2.1), our results suggest that halo formation time may play an important role in determining the color of the central galaxies. This is demonstrated more clearly in the left panel of Figure 2.3, where the $(g-r)$ color is shown as a function of f_c for centrals of fixed stellar mass. As one can see, massive galaxies are more or less all red, independent of f_c, while for galaxies with $M_* < 10^{11}\mathrm{M}_\odot$, their colors depend strongly on f_c. There seems to be a characteristic value $f_c \sim 0.01$ - 0.02, below and above which the color shows the opposite trends with f_c. At the high f_c end, galaxies

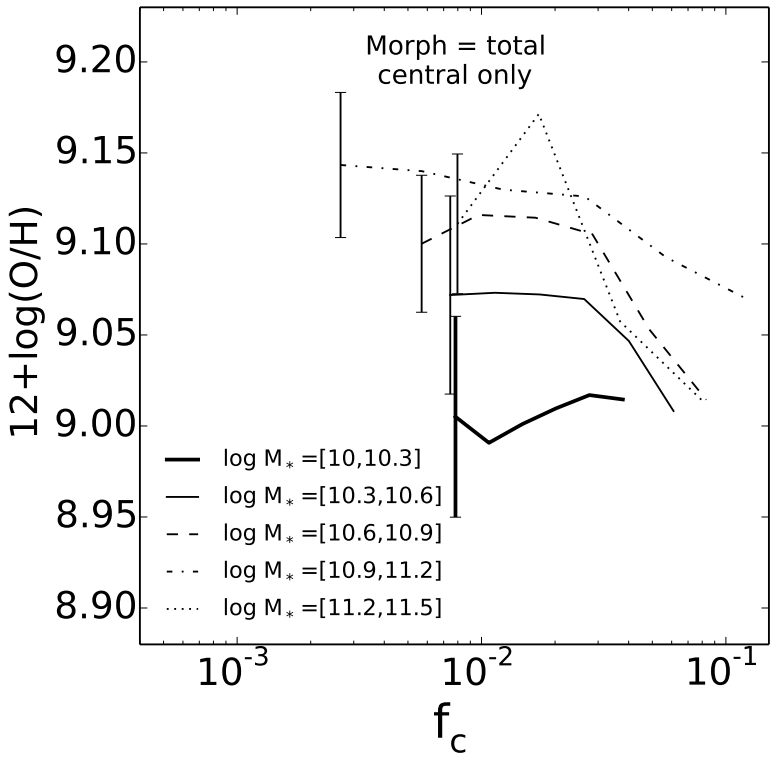

Figure 2.5. The correlation between gas phase oxygen abundance, $12 + \log(\mathrm{O/H})$, and f_c for central galaxies. The curves plot the median values in f_c bins. The error bars on the leftmost sides are 'typical' $[16\%, 84\%]$ ranges for each mass bin. Different curves refer to galaxies in different stellar mass bins, as indicated.

become increasingly redder as f_c increases, which may be produced by the fact that groups with higher f_c on average assembled their halos earlier. In contrast, galaxies in groups with $f_c < 0.02$ seem to have a reversed, albeit weak, trend between color and f_c. Note that for a given central stellar mass, lower f_c corresponds to higher halo mass. The reversed trend at low f_c reflects a 'down-sizing' effect of massive halos, in that centrals in massive halos formed their stars earlier than in low mass halos (e.g. Lu et al. 2015), although the massive host halos themselves assembled (half of their masses) later (lower f_c). This is consistent with the fact that *in situ* star formation in massive halos is quenched once their masses reached a few times $10^{12}\mathrm{M}_\odot$

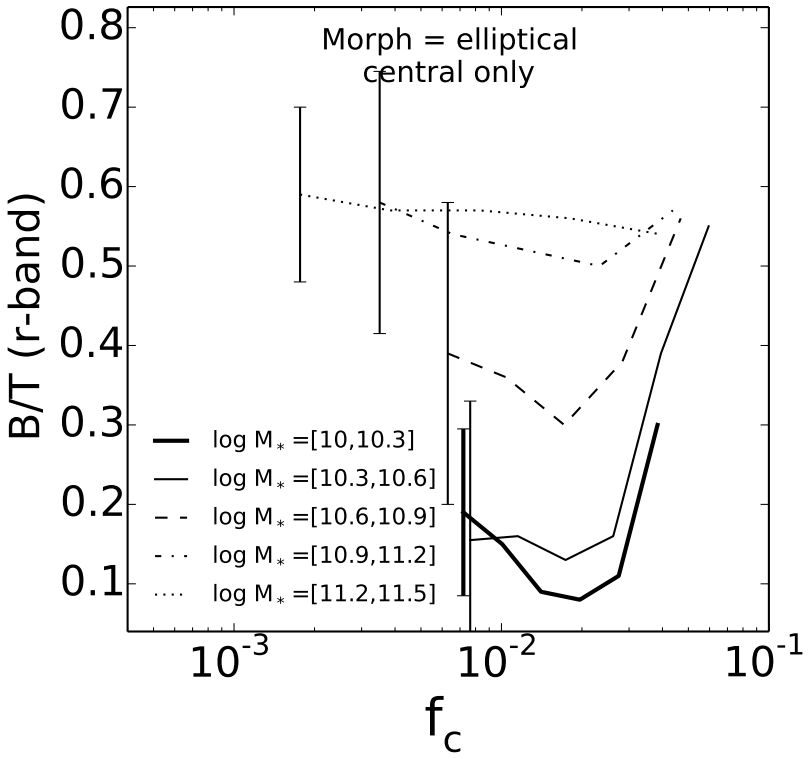

Figure 2.6. The correlation between the bulge to total ratio (in r-band), B/T, and f_c for central galaxies. The curves plot the median values in f_c bins. The error bars on the leftmost sides are 'typical' $[16\%, 84\%]$ ranges for each mass bin. Different curves refer to galaxies in different stellar mass bins, as indicated.

(e.g. figure 14 of Lu et al. 2014b), and, for high mass halos, more massive ones on average assembled a fixed amount of mass earlier (Li et al. 2008).

Figure 2.4 shows the quenched fraction of centrals as a function of f_c. Because for a given halo mass, f_c and stellar mass is strongly degenerated for centrals, here and in the following we only show results for centrals divided into different stellar mass bins but not divided further according to halo mass. For galaxies in each stellar mass bin, we separate them into quenched and star forming sub-populations using the definition of Moustakas et al. (2013),

$$\begin{aligned} \log\left(\frac{\mathrm{SFR}}{\mathrm{M}_\odot \mathrm{yr}^{-1}}\right) &= -0.49 + 0.65\log\left(\frac{M_*}{10^{10}\mathrm{M}_\odot}\right) \\ &\quad +1.07(z-0.1)\,. \end{aligned} \tag{2.2}$$

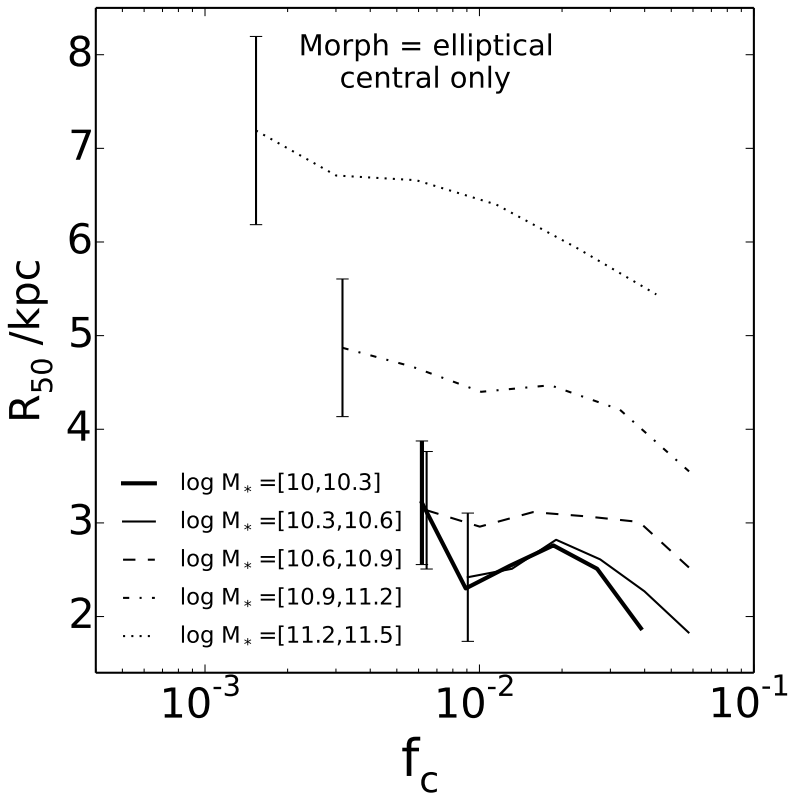

Figure 2.7. The correlation between the half-light radius (r-band) R_{50} and f_c for central ellipticals. The curves plot the median values in f_c bins. The error bars on the leftmost sides are 'typical' $[16\%, 84\%]$ ranges for each mass bin. Different curves refer to galaxies in different stellar mass bins, as indicated.

For a given M_*, galaxies with star formation rate (SFR) above the value given by the above equation are defined to be star forming, and those with SFR below the value are defined to be quenched. Given that the the specific star formation rate (sSFR, defined as the ratio between SFR and M_*) of a galaxy is closely related to its color, it is not surprising that the general trends seen in this plot are similar to those shown in Fig. 2.3. Low-mass centrals are dominated by star forming galaxies in halos of low f_c but become dominated by quenched galaxies at the high end of f_c. A reversal of trend is again seen at $f_c \sim 0.02$.

Finally, let us look at the gas phase metallicity of galaxies, which is shown as a function of f_c in Figure 2.5. The gas phase metallicity estimates are available only

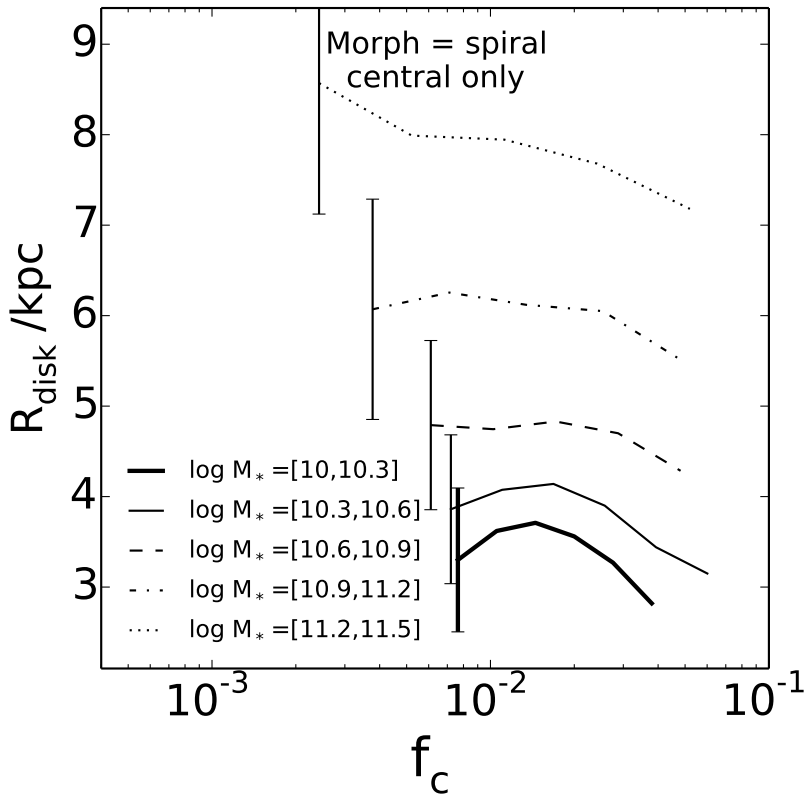

Figure 2.8. The correlation between disk scale length (r-band) $R_{\rm disk}$ and f_c for central spirals. The curves plot the median values in f_c bins. The error bars on the leftmost sides are 'typical' $[16\%, 84\%]$ ranges for each mass bin. Different curves refer to galaxies in different stellar mass bins, as indicated.

for a limited fraction of galaxies, mostly star forming ones. The result for the highest stellar mass bin is quite noisy because here only a small fraction of galaxies are star forming. For a given stellar mass, there is a clear trend that the gas phase metallicity decreases with increasing f_c. For centrals with $M_* > 10^{10.3}{\rm M}_\odot$, the decrease with f_c is quite rapid, by almost 0.1 dex. This decrease is comparable to the scatter in the gas phase metallicity - stellar mass relation obtained by Tremonti et al. (2004), suggesting that the scatter may be dominated by the variance in halo assembly, with galaxies formed in older dark matter halos tend to have lower gas-phase metallicities.

2.4.1.2 Structure and size

The bulge to total ratio, B/T, as described in the data section, is plotted against f_c in Figure 2.6. There are a number of interesting trends. Overall, the B/T increases with stellar mass, simply owing to the fact that earlier type galaxies are on average more massive. For massive galaxies with M_* higher than about $10^{11}\mathrm{M}_\odot$, the B/T ratio on average decreases with f_c. For galaxies with lower stellar masses, the trend changes at $f_c \sim 0.02$. While the B/T ratio decreases with increasing f_c at the low f_c end, it increases with f_c rapidly at $f_c > 0.02$.

As mentioned above, for a given central mass $M_{*,\mathrm{c}}$, a lower f_c on average corresponds to a higher halo mass M_h. Since a higher halo mass on average corresponds to a higher group richness, the decline of B/T with f_c may, therefore, be understood in terms of the morphology-density relation found by Dressler (1980) that early-type galaxies (higher B/T) are preferentially found in high density environments, while late-type galaxies are more likely to be found in poor groups and in the lower density fields. The increase of B/T with decreasing f_c at the low f_c end shown in Fig. 2.6 follows such a morphology-density relation. However, our results also contain new information, in that the morphology-density relation is present even for centrals of a given stellar mass.

The strong increase of B/T with increasing f_c seen for low-mass central galaxies with $M_\star < 10^{11}\mathrm{M}_\odot$ at $f_c > 0.02$ runs against the morphology-density relation. Since larger f_c means an earlier assembly time, as shown in the last section, the trend of B/T with f_c indicates an dependence on halo assembly time, in that central galaxies in older halos tend to have higher B/T. In the current CDM paradigm of structure formation, the formation of halos of a given mass at earlier time is on average more dominated by major mergers and older halos are on average more compact (e.g. Li et al. 2007, Zhao et al. 2009). If the bulge components are formed through major mergers or through secular evolutions of the disk components, their formation is expected

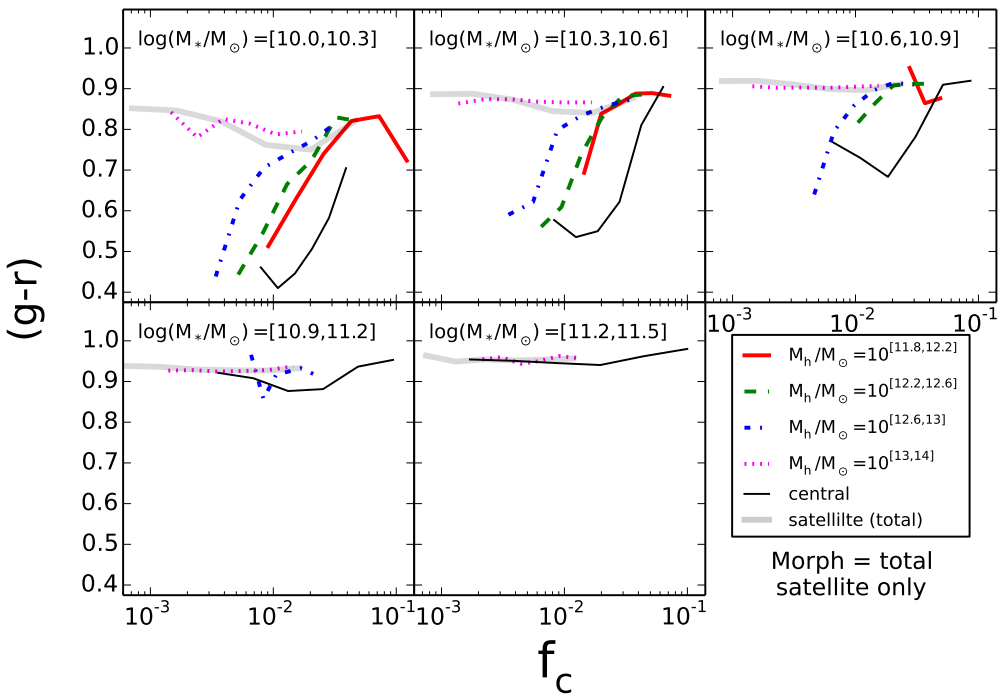

Figure 2.9. The correlation between $(g-r)$ color and f_c for satellite galaxies (thick lines). Individual panels show the medians in f_c bins for satellites of different stellar masses. Within each panel, satellites are divided into four subsamples according to the masses of their host halos, as denoted in the legend. The result for the total satellite sample in a given stellar mass bin is shown as the translucent thicker line in each panel. For comparison medians for centrals shown in Fig. 2.3 are re-plotted here as the thin solid lines.

to be promoted by both major mergers and a compact structure of dark matter halos. The positive correlation between B/T and f_c obtained here may follow directly from such formation. The reversal of the trend at $f_c < 0.02$ is also consistent with such interpretation, because central galaxies in massive halos actually have earlier formation due to the down-sizing effect described above.

Figures 2.7 and 2.8 show how the sizes of central galaxies of a given stellar mass correlate with f_c. Results are shown separately for the half-light radius (R_{50}) of ellipticals and the disk scale-length ($R_{\rm disk}$) of spiral galaxies. Here the morphological separation is made according to the visual classification from GZ2, and the sizes are taken from the r-band bulge-disk decompositions of Simard et al. (2011). For both

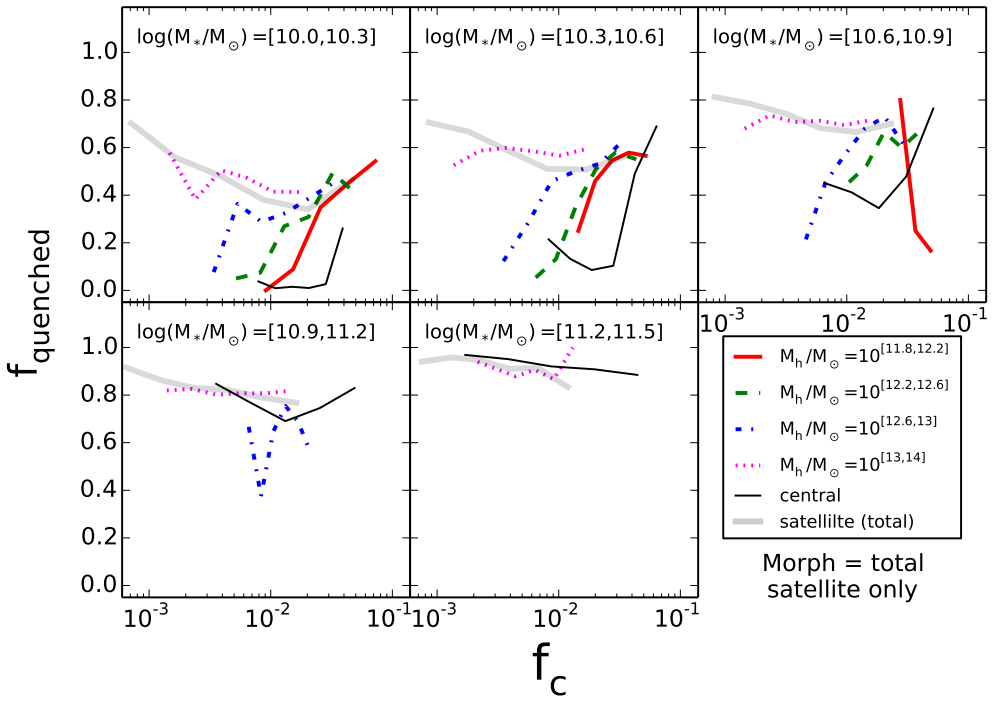

Figure 2.10. The correlation between the quenched fraction and f_c for satellite galaxies (thick lines). Individual panels show the medians in f_c bins for satellites of different stellar masses. Within each panel, satellites are divided into four subsamples according to the masses of their host halos, as denoted in the legend. The result for the total satellite sample in a given stellar mass bin is shown as the translucent thicker line in each panel. For comparison medians for centrals shown in Fig. 2.4 are re-plotted here as the thin solid lines.

ellipticals and spirals, more massive galaxies are larger, as expected. For a given stellar mass, the sizes of centrals decrease with f_c at $f_c > 0.02$. This is consistent with the interpretation that halos formed earlier on average are smaller. However, unlike star formation, there is no strong reversal of trend at $f_c < 0.02$, in particular for massive galaxies. For elliptical galaxies, this may be due to the fact that the assembly of the stellar component follows halo assembly more closely than star formation (e.g. figure 5 in Lu et al. 2015). For spiral galaxies, this result may indicate that disks can continue to accrete cold gas from halos as the halos grow, even in relatively massive systems.

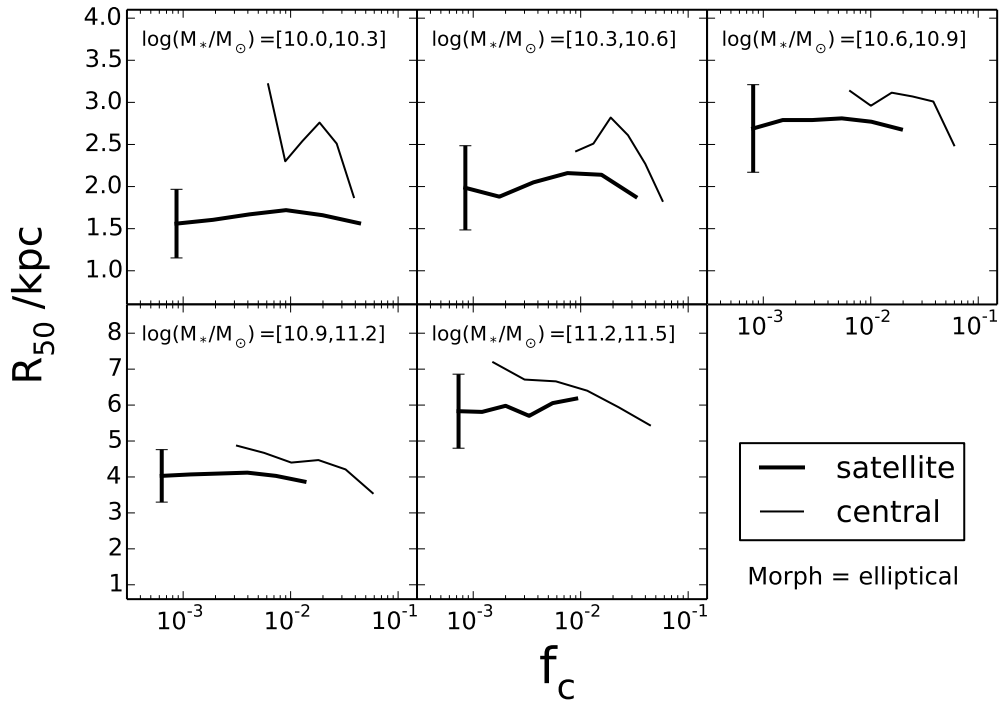

Figure 2.11. The correlation between R_{50} and f_c for satellite ellipticals. The thick curves are the medians in f_c bins, while the 'typical' $[16\%, 84\%]$ ranges are indicated by the bars on the leftmost sides. For comparison, results for central ellipticals shown in Fig. 2.7 are re-plotted here as thin lines. Different panels show the results in different stellar mass bins, as indicated.

2.4.2 Satellite galaxies

Figure 2.9 shows the $(g-r)$ color of satellite galaxies as a function of f_c of their host groups. The five panels show the results of galaxies in five stellar mass bins, as indicated. For each stellar mass bin, results are shown separately for galaxies in groups of four different halo mass bins, as indicated in the small panel. For given f_c and halo mass, more massive galaxies on average are redder. For the most massive galaxies with $M_* > 10^{11}\mathrm{M}_\odot$, which are only found in massive halos, their $(g-r)$ colors are all red, quite independent of f_c. For satellites with lower stellar masses ($M_* < 10^{11}\mathrm{M}_\odot$), there is a marked trend that the $(g-r)$ color becomes increasingly redder as f_c increases. The trend is weaker for groups with higher halo masses, and becomes almost totally flat for halo masses above $\sim 10^{13}\mathrm{M}_\odot$ (the magenta dotted

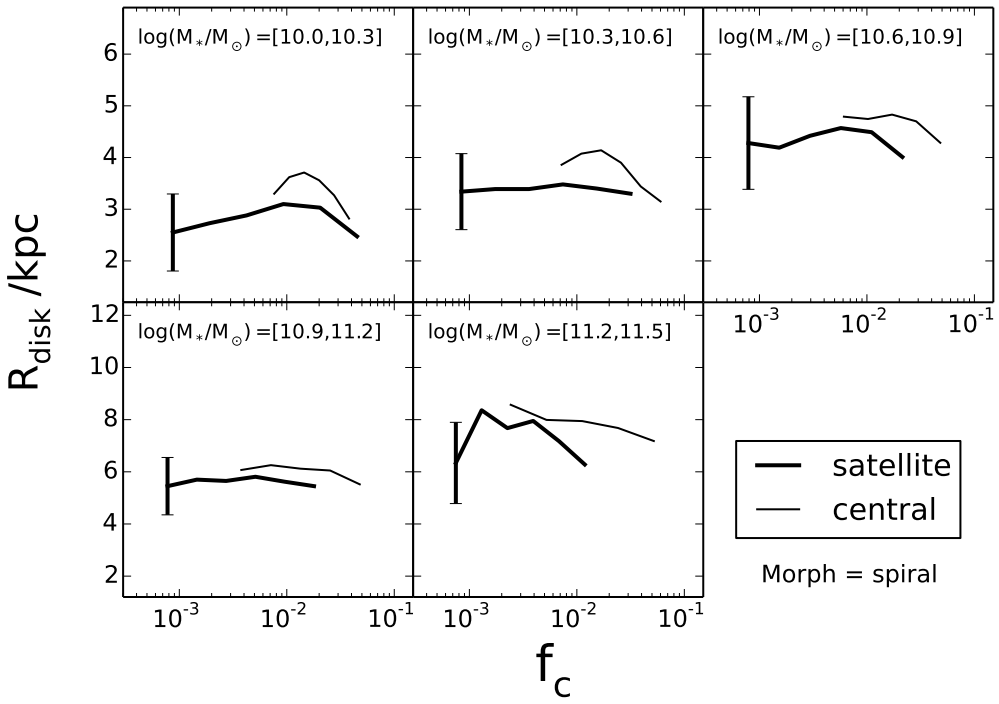

Figure 2.12. The correlation between $R_{\rm disk}$ and f_c for satellite spirals. The thick curves are the medians in f_c bins, while the 'typical' $[16\%, 84\%]$ ranges are indicated by the bars on the leftmost sides. For comparison, results for central spirals shown in Fig. 2.8 are re-plotted here as thin lines. Different panels show the results in different stellar mass bins, as indicated.

curve in each panel). We do not see a reversal in the trend in any ranges of f_c as seen in central galaxies shown in Fig. 2.3 (reproduced here as the black solid curves for comparison), because here results are shown separately for groups in different halo mass bins. If we consider all satellites of a given stellar mass regardless of their host halo mass, then we get the results as shown by the thick shaded line in each panel. Here we do see a change of trend at $f_c \sim 0.02$, which is similar to, albeit weaker than that for central galaxies. Clearly, satellites at the low-f_c end are dominated by the ones in massive groups. The reversed trend at $f_c < 0.02$, is consistent with the fact that galaxies in massive halos actually have earlier formation due to the down-sizing effect described earlier.

Figure 2.10 shows the quenched fraction of satellites as a function of f_c of their host groups. The format of this figure is exactly the same as Fig. 2.9, and the quenched fraction is again determined by using equation (2.2). The trends shown here are very similar to those in Fig. 2.9, again because the $(g-r)$ color is closely correlated with the sSFR used to separate quenched from star-forming galaxies.

Finally let us look at the sizes of galaxies. Here we consider ellipticals and spirals separately. Our tests showed that the dependence of size on halo mass is weak for satellites and the current samples are too small to give significant results for the halo-mass dependence. Thus, we only divide galaxies into stellar mass bins, but not further into halo mass bins. Figures 2.11 and 2.12 show how the sizes of satellite galaxies of a given stellar mass correlate with f_c (thick solid curves). Here results are shown separately for the half-light radius (R_{50}) of ellipticals and the disk scale-length ($R_{\rm disk}$) of spiral galaxies, both taken from the r-band bulge-disk decompositions of Simard et al. (2011). For both ellipticals and spirals, the trend with f_c is rather weak, although for low-mass galaxies the size seems to decrease as one moves away from $f_c \sim 0.01$ toward both the low and high ends of f_c. This trend suggests that galaxies of a given stellar mass on average have smaller sizes if formed earlier.

Compared with central galaxies of the same stellar mass (shown by the thin curves), satellites are smaller. This is true for both spirals and ellipticals, and the difference is larger for lower mass galaxies. It is interesting to note that the average sizes of satellites are comparable to those of centrals with the highest f_c, which indicates that sub-halos which host satellites may have as early formation as the oldest halos of similar masses that host centrals.

Weinmann et al. (2009) found that, at fixed stellar mass, late-type satellite galaxies have smaller radii than late-type central galaxies. Our results confirm theirs. However, Weinmann et al. (2009) found no difference in size for early-type galaxies, while Fig. 2.11 shows clearly that such difference also exists for ellipticals, particularly

for ellipticals with low stellar masses. The discrepancy may arise from the difference in the separation of early versus late types. While Weinmann et al. (2009) used the concentration parameter, defined as the ratio between R_{90} (radius within which 90% of the total light is included) and R_{50}, we use morphological classifications from GZ2. Weinmann et al. (2009) interpreted their finding as owing to the fading of stellar disks due to the aging of stars. However, it is unclear if such an interpretation can also explain the systematic change of disk size of central galaxies with f_c. Passive stellar evolution alone is also difficult to explain the difference between centrals and satellites for elliptical galaxies. Based on our results, the more likely reason is that halos formed earlier are more compact, and that the difference in sizes between centrals and satellites is due to differences in formation time, just as centrals in halos of different f_c.

2.5 Comparison with models

In order to explore the implications of our findings, we make comparisons of our results with some theoretical models. Since our results are derived from galaxy groups selected from a redshift catalogue, a detailed comparison between our observational results with theoretical models requires the construction of theoretical mock catalogs that take into account all observational selection effects. This is beyond the scope of this paper, and we will come back to this in a forthcoming paper. In this paper, we use halo occupations of galaxies predicted directly by models, ignoring all observational selection effects. As a demonstration, we use two specific models: the empirical age abundance matching (AAM) model published in Hearin & Watson (2013), Hearin et al. (2014), and the semi-analytical model (SAM) as described in Lu et al. (2014a).

While traditional abundance matching techniques only exploit the correlation between luminosity of galaxies and mass of their host haloes to assign galaxies in haloes from simulations, the AAM connects galaxies to haloes as a function of both color

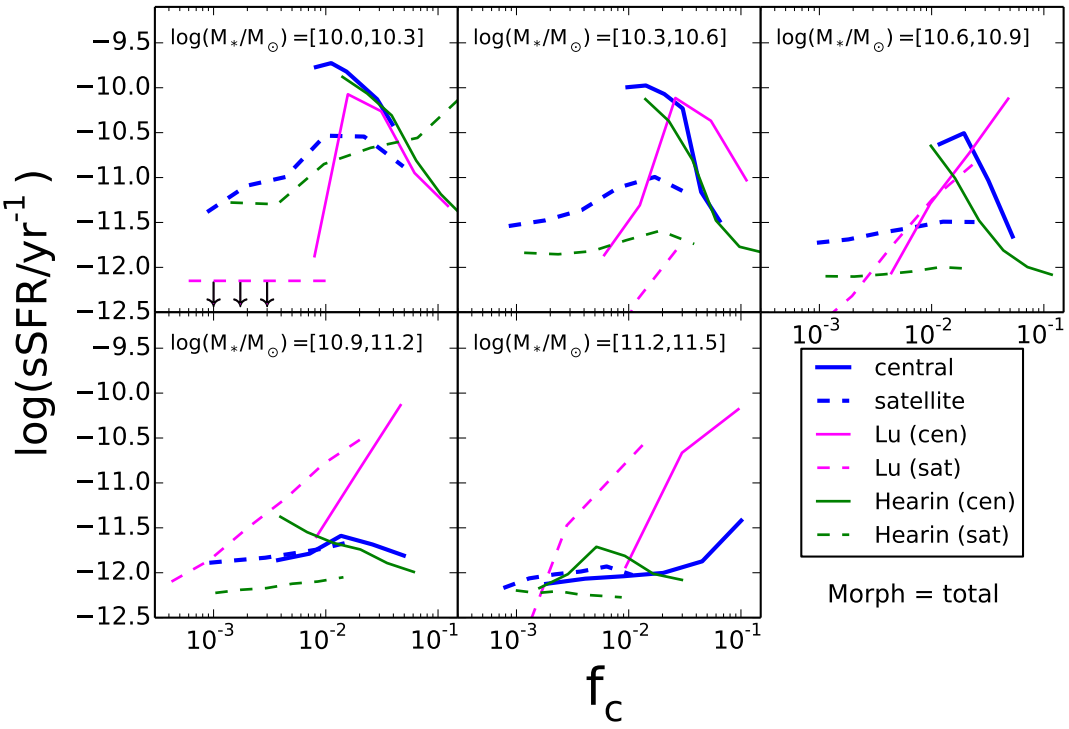

Figure 2.13. The observed correlation between f_c and the specific star formation rate, sSFRs (blue lines: solid for centrals and dashed for satellites) in comparison to the predictions of the semi-analytical model (SAM) of Lu et al. (2014a) (thin magenta) and the age abundance matching model of Hearin & Watson (2013) (thin green). Note that the sSFR of satellites in the first panel for the SAM are too low to show, and are represented by a horizontal line with down pointing arrows.

and luminosity. Specifically, it assigns stellar masses to galaxies according to the mass ranking of their host halos, and assign colors to galaxies of a given stellar mass according to the formation time ranking of their halos. The SAM approach, on the other hand, attempts to model physical processes using simplified receipts parameterized in simple functional forms. A SAM generally contains a large number of free parameters. Lu et al. (2014a) used a Monte Carlo Markov Chain method to infer their model parameters from observational constraints such as luminosity functions of galaxies at different redshifts. The Lu et al. SAM contains many of the same components as other SAMs. In particular it assumes a strong star formation feedback and an efficient gas stripping to prevent too much star formation in dark matter halos.

Since none of the models provides reliable predictions for the structural properties of galaxies, here we focus only on the star formation properties as represented by the specific star formation rate (sSFR) of galaxies. Figure 2.13 shows the sSFR as a function of f_c as predicted by the models of Hearin & Watson (2013) and Lu et al. (2014a). As in the observation, we identify the most massive galaxy in a halo to be the central galaxy, and use the ratio between $M_{c*,c}$ and the halo mass to define f_c. Here results are shown separately for centrals and satellites in five different stellar mass bins. For comparison, our observational results are included in each panel. As one can see, the AAM model reproduces the observational trends qualitatively. In particular, the rapid decreases of sSFR with increasing f_c for central galaxies in the low stellar mass bins are well reproduced. The trends for satellite galaxies are also well produced, although the predicted sSFR are systematically lower than the observational results. This discrepancy should not be taken too seriously, as the satellite population in observational groups may be contaminated by centrals that on average have higher SFR than the satellites of the same mass. As mentioned above, such contaminations can only be taken into account properly by applying the same group finder to the mock catalog constructed from the AAM model.

In contrast, the predictions of the SAM are very different from the observational results. The model predicts too much quenching of star formation in low mass satellites, while the star formation rates in centrals, particularly in groups with high f_c, are over-predicted by more than an order of magnitude. The SAM also fails to catch the overall trends in the observation, even qualitatively. These results suggest that the halo assembly plays an important role in regulating star formation, and the underlying physical processes are still poorly captured in the SAM considered here. It is clearly interesting to compare our results with other SAMs and simulation results, not only in sSFR, but also in other properties, such as size, B/T, and metallicity, to explore the implications of our results.

2.6 Summary

We have showed that the ratio, $f_c \equiv M_{*,c}/M_{\rm h}$, can be used as a reliable observational proxy of halo assembly time, with higher f_c for halos that assembled earlier. This use was motivated by the results of W11, who used N-body simulations to show that there is a tight correlation between $M_{\rm main}/M_{\rm h}$ ($M_{\rm main}$ being the main sub-halo mass) and halo half-mass assembly redshift (z_f), combined with (sub)halo abundance matching. We used the SDSS groups by Yang et al. to investigate how galaxy properties are correlated with the assembly times of their host halos.

Central galaxies of a given stellar mass with higher f_c are found to be redder and more quenched in star formation while $f_c > 0.02$. This implies that star formation in centrals in this regime is dictated by their halo assembly history. A reversed albeit weak trend is seen for centrals with $f_c < 0.02$, which reflects the down-sizing effect that a more massive halo on average reaches the mass of most efficient *in situ* star formation, $\sim 10^{12}{\rm M}_{\odot}$, earlier. Similar trends with f_c are found for the bulge to total ratio, B/T: central galaxies hosted by older halos tend to have higher B/T ratios. We suggest that this is because older halos are more compact and their formation is more dominated by major mergers. For a given stellar mass, the sizes of central galaxies are also correlated with f_c for both ellipticals (in terms of the half-light radius, R_{50}) and spirals (in terms of the disk scale-length, $R_{\rm disk}$), with centrals hosted by older halos being smaller. This trend is again consistent with the fact that halos of a given mass are more compact at higher redshifts.

We have also analyzed how the intrinsic properties of satellite galaxies change with the value of f_c of their host halos. Here we found that, for a given stellar mass, satellites residing in older halos are redder and more quenched, and this trend is stronger for lower mass halos. Satellites also appear smaller than centrals of the same mass, and this is true for both ellipticals and spirals. These results can again be explained by the fact that halos that assembled earlier are more compact. As for

centrals, a weak down-sizing effect in the quenching of star formation is also seen for satellites hosted by massive halos with $f_c < 0.02$. These results, together with those found for the centrals, demonstrate clearly that halo assembly plays an important role in determining the properties of galaxies the halos host.

We present our preliminary comparisons of our observational results with the predictions by the AAM model of Hearin & Watson (2013) and by the SAM of Lu et al. (2014a). The AAM model reproduces well the general trends in the observational data, while the SAM fails to do so. The SAM predicts too many quenched low-mass satellites and too small fraction of quenched high-mass galaxies. These imply that halo assembly history is another important factor in addition to halo mass that can affect star formation in galaxies, and such effects have yet to be properly modeled in the SAM. In this context, the observational results obtained here are expected to provide stringent constraints on theoretical models of galaxy formation and evolution. We will come back to detailed comparison between our observational results and model predictions in a forthcoming paper.

CHAPTER 3

TESTING GALAXY FORMATION MODELS WITH GALAXY STELLAR MASS FUNCTIONS[1]

We compare predictions of a number of empirical models and numerical simulations of galaxy formation to the conditional stellar mass functions (CSMF) of galaxies in groups of different masses obtained recently by Lan et al. to test how well different models accommodate the data. The observational data clearly prefer a model in which star formation in low-mass halos changes behavior at a characteristic redshift $z_c \sim 2$. There is also tentative evidence that this characteristic redshift depends on environment, becoming $z_c \sim 4$ in regions that eventually evolve into rich clusters of galaxies. The constrained model is used to understand how galaxies form and evolve in dark matter halos, and to make predictions for other statistical properties of the galaxy population, such as the stellar mass functions of galaxies at high z, the star formation and stellar mass assembly histories in dark matter halos. A comparison of our model predictions with those of other empirical models shows that different models can make vastly different predictions, even though all of them are tuned to match the observed stellar mass functions of galaxies.

3.1 Introduction

In the current paradigm of structure formation within the Λ cold dark matter (ΛCDM) framework, initial small fluctuations in the cosmic density field are amplified

[1] THE CONTENTS OF THIS CHAPTER ARE PUBLISHED IN LIM ET AL. 2017, MNRAS, 464, 3256.

by gravitational instability, eventually forming highly nonlinear structures called dark matter halos (see Mo et al. (2010) for a review). Galaxies then form at the centers of the gravitational potential wells of the dark matter halos by radiative cooling and condensations of baryonic gas (e.g. White & Rees 1978, Fall & Efstathiou 1980, Mo et al. 1998). In order to reproduce the observed stellar mass function of galaxies in the CDM scenario, however, star formation in dark matter halos has to be inefficient (e.g. Yang et al. 2003), and various feedback processes have been proposed to suppress the star formation efficiency in dark matter halos.

In this framework, therefore, galaxy formation and evolution are governed by a number of physical processes which, in turn, are characterized by a number of characteristic scales. First, cosmological N-body simulations have shown that the assembly histories of dark matter halos in general consist of two distinctive phases: an earlier phase of fast mass acquisition during which the potential well of a halo deepens rapidly with time, and a later phase of slow accretion, with a time scale longer than the Hubble time (e.g. Zhao et al. 2003). Zhao et al. (2009) found that the two phases are separated at a time when a halo obtains about $\sim 4\%$ of its final mass (see also van den Bosch et al. 2014). Second, hydrodynamical simulations have demonstrated that radiative cooling is effective in halos with masses smaller than $M_h \sim 6 \times 10^{11}\mathrm{M}_\odot$, so that the accretion rate of cold gas into galaxies is determined by the halo mass accretion rate, independent of radiative cooing (e.g. Kereš et al. 2005, 2009). Above this mass scale, on the other hand, radiative cooling is ineffective, so that the cold gas accretion is delayed by the cooling time scale. For massive halos with masses above $10^{13}\mathrm{M}_\odot$, a significant fraction of the baryonic gas is expected to be in the hot halo in the absence of a heating source. Third, supernova feedback from star formation is believed to be effective for halos with masses below $\sim 10^{11}\mathrm{M}_\odot$ (e.g. Dekel & Silk 1986, Somerville et al. 2008, Lu et al. 2012). Finally AGN feedback from accreting super-massive black holes has been proposed as a mechanism to suppress

star formation in massive halos, with masses above $M_h \sim 10^{13}\mathrm{M}_\odot$ (e.g. Ferrarese & Merritt 2000, McConnell et al. 2011).

A number of approaches have been adopted to explore the physical processes that govern galaxy formation and evolution, and to facilitate comparisons between theory and observation. The first is hydrodynamical simulation that includes both dark matter and baryonic components (e.g. Dubois et al. 2014, Khandai et al. 2015, Vogelsberger et al. 2014, Schaye et al. 2015). However, due to limited resolution and subgrid implementations of some key processes, the results obtained from such simulations are still questionable, even though they can match some observational data (e.g. Governato et al. 2004, 2010, Okamoto et al. 2005, Guedes et al. 2011). Furthermore, high resolution hydrodynamical simulations are computationally expensive, which prohibits the explorations of a large parameter space. Because of this, an alternative approach, the semi-analytic model (SAM) of galaxy formation, has been developed (e.g. White & Frenk 1991, Kauffmann et al. 1999, Kang et al. 2005, Bower et al. 2006, Croton et al. 2006, Somerville et al. 2008, Guo et al. 2011, Lu et al. 2011). The SAM approach combines halo merger histories, obtained either from dark-matter only simulations or from analytical models, with gas and star formation processes using parametrized functions that describe the underlying physical processes. This approach is computationally inexpensive, allowing one to investigate a large set of different models. However, since all the physical processes are approximated with simple empirical functions, the reliability and accuracy of this approach needs to be checked. More recently, a third approach has been adopted to understand how galaxies form and evolve in the cosmic density field. The goal of this approach is to establish the connections between galaxies and dark matter halos through an empirical approach, using observational data as constraints. Models developed along this line include the halo occupation distribution (HOD; e.g. Jing et al. 1998, Peacock & Smith 2000, White 2001, Berlind & Weinberg 2002, Bullock et al. 2002, Zehavi et

al. 2004, 2011), the conditional luminosity function (CLF; Yang et al. 2003, 2012, van den Bosch et al. 2003), the halo abundance matching model (HAM; Kravtsov et al. 2004, Vale & Ostriker 2004, 2006, Conroy et al. 2006, Behroozi et al. 2010, Guo et al. 2010, Moster et al. 2010, Reddick et al. 2013), and the halo-based empirical model (Lu et al. 2014, 2015).

To a certain degree, both the SAM and empirical approaches are methods to summarize observational data in terms of model parameters characterizing the galaxy-halo connections. Much progress has been made recently in this area. Using the CLF model and constraints of the observed luminosity function and correlation function of galaxies, Yang et al. (2003) found a characteristic halo mass scale, $\sim 10^{12}\mathrm{M}_{\odot}$, in the relationship between galaxy luminosity/stellar mass and halo mass relation, suggesting that star formation efficiency declines rapidly toward both the higher and lower mass ends. With the use of galaxy groups selected from the 2dF (Yang et al. 2005) and SDSS (Yang et al. 2007), Yang et al. (2005) found a similar mass scale from the observed galaxy luminosity/stellar mass - halo mass relations obtained directly from galaxy groups. In particular, Yang et al. (2005) suggested the existence of another characteristic mass scale, $\sim 10^{11}\mathrm{M}_{\odot}$, where the galaxy luminosity-halo mass relation may change its behavior. Similar results have since been obtained at higher z with the use of the observed luminosity/stellar mass functions of galaxies. In particular, the presence of the mass scale at $\sim 10^{12}\mathrm{M}_{\odot}$ seems to extend to higher z without showing strong evolution (Moster et al. 2010, Behroozi et al. 2013, Guo et al. 2010, Yang et al. 2012).

More recently, Lu et al. (2014, 2015) developed a halo-based empirical model to follow the star formation and stellar mass assembly histories of galaxies in dark matter halos. In particular, they used the observed conditional luminosity functions of cluster galaxies obtained by Popesso et al. (2006) as an constraint in addition to the field stellar mass functions at different redshifts. They found that the observational data

require two additional characteristic scales, a characteristic redshift, $z \sim 2-3$, and a corresponding mass scale at $10^{11}\mathrm{M}_{\odot}$, below which star formation changes behavior at the characteristic redshift. These results clearly demonstrate that the observed conditional luminosity/stellar mass function of galaxies in clusters can provide important information about galaxy formation and evolution at high redshift. However, since clusters of galaxies only contain a small fraction of the total galaxy population, the results may be affected by some environmental effects that are specific only to clusters of galaxies.

Using the galaxy groups of Yang et al. (2007) combined with galaxies in the SDSS photometric catalogue, Lan et al. (2016) have recently measured the conditional luminosity/stellar mass functions (hereafter CSMFs) that cover four orders of magnitude in galaxy luminosity, and three orders of magnitude in halo mass, from $\sim 10^{12}$ to $10^{15}\mathrm{M}_{\odot}$. They found a characteristic luminosity scale, $L \sim 10^{9}L_{\odot}$, below which the slope of the CSMF becomes systematically steeper, and that this trend is present for all halo masses. This ubiquitous faint-end upturn suggests that it is formation, rather than cluster-specific environmental effect, that plays the dominating role in regulating the stellar masses of faint satellites. Clearly, these observational results will provide new constraints on models.

This paper consists of two parts. First, we use the new CSMFs to update the empirical model of Lu et al. (2014, 2015) and show that there is only marginal difference between the original model and the updated model. Second, we compare model predictions from empirical models and numerical simulations to the CSMFs of Lan et al. to test how well different models accommodate the new data. We will show that, among all the models considered, only the Lu et al. (2014, 2015) model can match the observational data reasonably well. Also, we present predictions of the different models for other statistical properties of the galaxy population.

Table 3.1. A list of the models and the simulations.

Model / Simulation	Reference
Y12	Yang et al. (2012)
M13	Moster et al. (2013)
B13	Behroozi et al. (2013)
L15	Lu et al. (2015)
L15-U	this work
Illustris	Vogelsberger et al. (2014)
EAGLE	Schaye et al. (2015)

The organization of this paper is as follows. Section 3.2 describes the empirical models to be tested and two recent numerical simulations, Illustris (Vogelsberger et al. 2014) and EAGLE (Schaye et al. 2015), to be compared. In Section 3.3, we describe the observational data that are used in our analysis, and present comparisons of the empirical models and the simulations with them. In Section 3.4, we present a more detailed comparison of the model predictions in star formation rate, stellar mass - halo mass relation, mass assembly history, and stellar mass function for high redshifts. Finally, we summarize and discuss our results in Section 3.5.

3.2 Models

In this paper we select a number of popular empirical models and two recent hydrodynamical simulations to test against observational data. Here we describe these models and simulations briefly. Table 1 lists the models and the simulations that we test. Readers are referred to the original papers for details.

3.2.1 Empirical models

One of the simplest way to link galaxies to their dark matter halo/subhalo population is to use halo abundance matching (e.g. Mo et al. 1999). This approach assumes a monotonic relation between halo mass and galaxy stellar mass. Satellite galaxies observed at a given redshift were all once central galaxies before they were accreted onto larger halos. Since satellites are expected to evolve differently due to environmental effects such as tidal stripping and ram pressure stripping, many abundance matching models apply a monotonic relation between galaxy stellar mass and halo mass at the time when a halo first became a subhalo, instead of at the time of observation. Most of previous investigations make the assumption that the halo mass - galaxy mass relation is independent of when a sub-halo is accreted into its host (e.g. Vale & Ostriker 2004, 2006, Conroy et al. 2006, Behroozi et al. 2010, Guo et al. 2010, Moster et al. 2010). With this assumption, at a given redshift, halos of a given mass are therefore always linked to galaxies of the same stellar mass. However, it was found that applying this method to different redshifts actually leads to different stellar mass - halo mass relation (e.g. Conroy et al. 2006), suggesting that the method implemented in this way is not self-consistent. As an improvement, models have been developed in which the galaxy-halo relation is allowed to depend on both halo mass (defined e.g. at the time when a halo first becomes a sub-halo) and the time when a halo becomes a sub-halo. We test four models in this category, by Yang et al. (2012), Moster et al. (2013), Behroozi et al. (2013), and Lu et al. (2015), respectively.

3.2.1.1 Yang et al. model

The model of Yang et al. (2012) (Y12, hereafter) takes the same functional form as that proposed in Yang et al. (2003) for the halo mass - galaxy luminosity/stellar mass relation:

$$\frac{M_*}{M_h} = N \left[\left(\frac{M_h}{M_1} \right)^{-\beta} + \left(\frac{M_h}{M_1} \right)^{\gamma} \right]^{-1}. \tag{3.1}$$

This is basically a double power law specified by two asymptotic slopes, β and γ, describing the low- and high-mass end behaviors, respectively, and by a characteristic mass scale M_1 where the transition between the two power laws occurs, and with N being an overall amplitude. The four free parameters were assumed to be redshift dependent and the dependencies were modeled by simple functions. The above relation was used to assign stellar masses to halos at different redshifts. They adopted the halo mass function of Sheth et al. (2001) to model the halo population. For sub-halos, the model of Yang et al. (2011) was used to follow both the mass function and the distribution in the accretion time (the time when a halo first becomes a subhalo). A stellar mass is assigned to a sub-halo at the time of accretion according to its mass at that time using equation (3.1). The subsequent evolution of the satellite associated with a sub-halo was followed according to its orbit determined through a dynamical friction model. The model parameters were then obtained by fitting the model predictions to the observed stellar mass functions (SMFs) of galaxies from $z = 0$ to 4, and the correlation function of $z \sim 0$ galaxies as a function of galaxy luminosity/stellar mass.

3.2.1.2 Moster et al. model

Moster et al. (2013) (M13) adopted a similar double power-law for the stellar mass - halo mass relation as described by equation (3.1), and simple functional forms to describe the redshift dependencies of the model parameters. They applied the relation to halos and sub-halos obtained from N-body simulations. Individual halos and sub-halos are matched and traced across different snapshots (i.e. different redshifts), so that merger trees are generated to track their evolutions. Galaxies hosted at the centers of halos and sub-halos were referred to as centrals and satellites, respectively. For centrals, the stellar masses were given by the stellar mass - halo mass relation

using the redshift and halo mass at the snapshot in question. For satellites, the stellar masses were obtained by applying the stellar mass - halo mass relation at the redshift when their halos first became sub-halos using their halo masses at this redshift, as in Y12. The stellar mass of a satellite was assumed to remain unchanged in the subsequent evolution. Some uncertainties in the stellar mass - halo mass relation were taken into account. Model parameters characterizing the stellar mass - halo mass relation were then tuned to match a set of observed SMFs from $z = 0$ to 4.

3.2.1.3 Berhoozi et al. model

The approach adopted by Behroozi et al. (2013) (B13) was similar to those of Y12 and M13, but the stellar mass - halo mass relation assumed was more complicated and was designed in part to reproduce the observed SMFs at the faint ends. Here again, halo merger trees extracted from N-body simulations were used to trace the formation of dark matter halos. As in Y12 and M13, they applied their stellar mass - halo mass relation to 'infall' mass at the time of accretion to assign stellar masses to subhalos. Subsequent stellar mass loss of satellites after their accretion into their host halos was also taken into account. Finally, they used the observed SMFs at $z = 0$ - 8, as well as the cosmic star formation rates and specific star formation rates, to constrain their model parameters.

3.2.1.4 Lu et al. model

The Lu et al. (2014, 2015) model (hereafter L15) was based on the star formation rate (SFR) - halo mass relation as a function of redshift:

$$\dot{M}_*(M_h, z) = \varepsilon \frac{f_b M_h}{\tau} (x+1)^\alpha \Big(\frac{x+R}{x+1}\Big)^\beta \Big(\frac{x}{x+R}\Big)^\gamma \tag{3.2}$$

where $f_b = \Omega_{b,0}/\Omega_{m,0}$, $\tau = [10H_0(1+z)^{3/2}]^{-1}$ approximates the dynamical time of halos, $x \equiv M_h/M_c$, with M_c being a characteristic mass scale and R is parameter

of $0 \leq R \leq 1$. Thus, $\dot{M}_*/M_h \propto M_h^{\{\alpha,\beta,\gamma\}}$ for $\{M_h \gg M_c,\ RM_c < M_h < M_c$, and $M_h \ll RM_c\}$, respectively. This relation is applied only to central galaxies. After a galaxy becomes a satellite, Lu et al. assumed that it moves on an orbit determined by its initial energy and orbital angular momentum together with dynamical friction. A satellite galaxy is assumed to merge with the central galaxy once it sinks to the center of the halo. At this time, it adds a fraction (treated as a free parameter, f_{sc}) of its mass to the central galaxy, and the rest is assumed to become halo stars. The SFRs in satellites were modeled with a simple exponential model,

$$\dot{M}_{*,sat} \propto \exp\Big[-\frac{t - t_{acc}}{\tau_s}\Big] \tag{3.3}$$

where t_{acc} is the time when the galaxy becomes a satellite, and $\tau_s = \tau_{s,0} \exp\big[-M_*/M_{*,c}\big]$ is adopted to reflect halo mass dependence of the time scale, with $\tau_{s,0}$ and $M_{*,c}$ being free parameters. The stellar mass in a galaxy is then obtained by integrating the SFR over time, taking into account mass loss due to stellar evolution. Lu et al. used halo merger trees generated with the algorithm developed by Parkinson et al. (2008), which is based on the extended Press-Schechter formalism calibrated with N-body simulations.

Lu et al. adjusted both their functional forms and free parameters to match the SMFs at $0 < z < 4$ and the CSMFs of galaxies in clusters of galaxies as given by Popesso et al. (2006). They found that the model assuming all the parameters to be independent of redshift is not able to match the observed SMFs at high redshift. They therefore extended their model by allowing α to change with redshift as $\alpha = \alpha_0(1+z)^{\alpha'}$. This model was referred to as Model II in Lu et al.. Model II was found to be able to describe all the stellar mass functions (SMFs) at both low and high redshifts, but fails to match the faint-end upturn in the CSMF of cluster galaxies. Because of this, Lu et al. extended their model once more by allowing the parameter γ, which dictates the SFR in low-mass halos, to depend on redshift:

$$\begin{aligned} \gamma &= \gamma_a && \text{if } z \le z_c \\ &= (\gamma_a - \gamma_b)\Big(\frac{1+z}{1+z_c}\Big)^{\gamma'} && \text{if } z > z_c \end{aligned}$$

In this model, referred to as Model III by Lu et al., $\gamma \to \gamma_b$ at $z \gg z_c$, and the free parameter, γ', controls how rapidly the transition to γ_b occurs above the characteristic redshift z_c. This Model III was found to be able to fit both the SMFs at different redshifts and the CSMF of cluster galaxies simultaneously.

3.2.1.5 Updating the parameters of the L15 model

Instead of using the model parameters of Lu et al., we use only the observed CSMFs as constraints to update the model parameters. We use the MULTINEST method developed by Feroz et al. (2009), which makes use of the nested sampling algorithm of Skilling et al. (2006), to compute the posterior distribution of the model parameters. The MULTINEST is found to yield practically the same results as the traditional MCMC method but with ~ 10 times smaller number of likelihood calculations for the problem concerned here. The reader is referred to the original papers for details.

Table 2 compares the updated parameters with the model parameters of Lu et al. The average star formation rates predicted with the updated parameters at various redshifts are very similar to those obtained by L15, as shown in Figure 3.1. We also found that the differences in the two parameter sets result only in marginal changes in the CSMFs in that the updated model (hereafter L15-U) predicts slightly flatter slopes at the faint-ends for massive haloes. This is owing to the fact that the Lan et al. CLFs have shallower faint-end slopes for massive halos than the cluster galaxy luminosity function used by L15. The marginal difference between the two parameter sets demonstrates that the low-z CSMFs alone can constrain models in a similar way as the field SMFs at different redshifts. Furthermore, as we will see in §3.4), they also contain information about the low-mass end of the SMF at high z, where direct

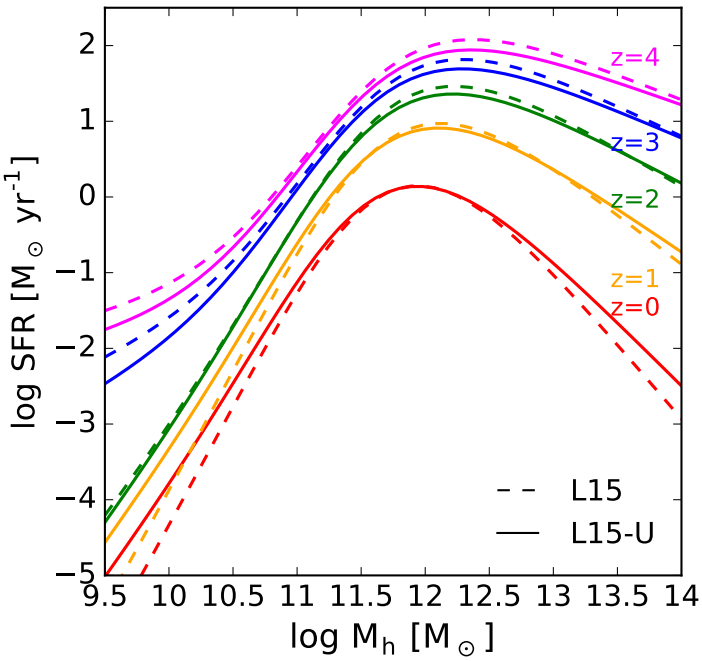

Figure 3.1. The average star formation rate of central galaxies as a function of halo mass at different redshifts as predicted by the original L15 model (dashed lines) and the L15-U model (solid lines).

observations are still uncertain. We use L15-U to present results throughout this paper.

3.2.1.6 Need for a more extended model family?

As mentioned above, L15 assumed the characteristic redshift, z_c, the redshift at which the SFR in low mass progenitors changes behavior, to be independent of the host halo mass. However, it is plausible that z_c depends on the host halo mass, because structure formation, and presumably star formation, are expected to occur earlier in regions that correspond to higher mass halos at the present day. Motivated by this, we test a more extended model family in which the characteristic redshift changes with host halo mass at $z = 0$, $M_h(0)$:

Table 3.2. A list of the model parameters. The medians and the standard deviations are presented.

parameters	L15	L15-U
α_0	-3.0 ± 1.0	-2.7 ± 0.79
α'	-0.36 ± 0.16	-0.32 ± 0.21
β	3.7 ± 0.73	3.5 ± 1.0
γ_a	2.0 ± 0.55	1.3 ± 0.69
γ_b	-0.84 ± 0.14	-1.1 ± 0.21
γ'	-4.4 ± 0.52	-3.1 ± 0.88
z_c	1.8 ± 0.31	2.0 ± 0.28
$\log_{10} M_c$	1.6 ± 0.15	1.6 ± 0.11
$\log_{10} R$	-0.86 ± 0.18	-0.92 ± 0.20
ε	0.20 ± 0.29	0.050 ± 0.11
$\log_{10} H_0 \tau_{s,0}$	-0.90 ± 0.16	-0.85 ± 0.11
$\log_{10} M_{*,c}$	0.34 ± 0.28	0.18 ± 0.19
f_{sc}	0.44 ± 0.22	0.52 ± 0.15

$$(1 + z_c) = (1 + z_{c,0}) \Big(\frac{M_h(0)}{10^{12} \mathrm{M}_\odot} \Big)^{\zeta} \tag{3.4}$$

where ζ controls the halo mass dependence of z_c, and $z_{c,0}$ is z_c for halos of $M_h(0) = 10^{12} \mathrm{M}_\odot$. We use the same CSMFs as used in the earlier subsection to constrain model parameters.

To test if such an extension is necessary, we use the Bayes factor,

$$K = \frac{P(D|M_1)}{P(D|M_2)} = \frac{\int P(D|\theta_1, M_1) P(\theta_1|M_1) d\theta_1}{\int P(D|\theta_2, M_2) P(\theta_2|M_2) d\theta_2}, \tag{3.5}$$

where D is a given data set, M_1 and M_2 are two different models, and θ_1 and θ_2 are the parameter space of the models. This factor quantifies the preference of a given

data set for one model family over the other. As it integrates over all parameter space of each of the model families, it naturally penalizes over-fitting.

When all the data points of the CSMFs are used as constraints, the Bayes factor between the extended model (the one including ζ) and the original L15 parametrization is given by $2\ln K \approx 56$, which indicates a strong need for having ζ statistically. The median value of $\zeta \approx 0.064$ thus obtained implies that the characteristic redshift z_c is $z \approx 3.8$ for halos of $M_h(0) = 10^{15}\mathrm{M}_\odot$, in comparison to $z_c \approx 2.1$ for halos with $M_h(0) = 10^{12}\mathrm{M}_\odot$. This increase of z_c with host halo mass leads to flatter faint-end slopes for massive halos, giving better matches to the faint-ends of the CSMFs for both low-mass and high-mass halos.

It is worth noting, however, that the uncertainties in the stellar mass estimates may change the CSMFs in both the lowest and highest mass ends, where the slopes of the CSMF are steep. As a test, we use only the CSMFs in the range $M_* = [10^8, 10^{11}]\mathrm{M}_\odot$ as the observational constraints. In this case, the models with or without ζ are almost equally favored in terms of the Bayes factor. Given these, we conclude that the original form of the L15 model can still accommodate the new CSMFs, and that the current data are still too uncertain to determine if a more extended model family is required.

3.2.1.7 Model implementations

We implement the empirical models described above to the dark matter halo population. We use the algorithm developed by Parkinson et al. (2008) to generate halo merger trees and to follow the build-up of dark matter halos. As mentioned above, this algorithm is based on the extended Press-Schechter formalism calibrated with results from N-body simulations. As shown in Jiang & van den Bosch (2014), the predictions of this algorithm match accurately many properties of halo merger

trees obtained directly from simulations, including halo mass assembly history, halo merger rate, and sub-halo mass functions.

The empirical models described above also take into account some uncertainties in the observational data and in the model assumptions, such as the intrinsic scatter in the stellar mass - halo mass relation, uncertainties in the stellar population synthesis and dust models, Eddington bias, and errors in redshift measurements. Unfortunately, how these uncertainties change as a function of redshift is poorly established. They are treated differently in different models. M13 adopted constant scatter in the stellar mass - halo mass relation and in the stellar mass estimate, while B13 parametrized the uncertainties as functions of redshift and treated them as a new set of free parameters to be determined in their model fitting. The treatment by Y12 lies in between. In our implementations, we follow each individual model as close as possible.

We use WMAP7 cosmology to obtain the halo mass function, to construct halo merger trees, and to estimate distances from redshifts. We adopt the Chabrier (2003) IMF, the stellar population synthesis model of Bruzual & Charlot (2003) to account for stellar mass loss and to obtain stellar mass function from observations. These assumptions are the same as adopted in the original models, except for Y12 where a Kroupa (2001) IMF was adopted. We correct the stellar masses of Y12 model by a factor of $\sim 1.4/1.7$ to match the IMF we adopt.

3.2.2 Hydrodynamical simulations

We also test the predictions from two recent high-resolution, cosmological hydrodynamical simulations. The first is Illustris simulation (Nelson et al. 2015), which follows 1820^3 particles for each of the gas and dark matter components in a total volume of $(106.5\,\mathrm{Mpc})^3$, assuming WMAP9 cosmology ($\{\Omega_m, \Omega_\Lambda, h\} = \{0.273, 0.727, 0.704\}$). The other components the simulation traces are stars, stellar wind particles, and super-massive black holes. The simulation starts from $z = 127$ and includes physical

processes such as radiative cooling, star formation, and various feedback processes. The free parameters in their model were constrained by using the star formation efficiency obtained from separate simulations that are more accurate in resolving small-scale structures. In our analysis, we use Illustris-1, their flagship simulation that has the highest mass resolution ($1.6 \times 10^6 M_\odot$ and $6.3 \times 10^6 M_\odot$ for baryon and dark matter, respectively). To match the set of observations adopted here for model testing, we use the snapshot at $z = 0.03$, which contains a total of $7,647,219$ groups identified by the FoF algorithm. In the simulation, galaxies are defined according to the spatial distribution of stars and stellar wind particles, and the brightness profile fit to them. The simulation assumes the Chabrier (2003) IMF and the stellar population synthesis model of Bruzual & Charlot (2003). As the cosmological parameters of WMAP9 are similar to those of WMAP7, the difference in cosmology is ignored in our analysis. We bin their stellar masses to obtain the stellar mass function (SMF).

Another simulation we use is the Evolution and Assembly of GaLaxies and their Environments (EAGLE; Schaye et al. 2015). EAGLE traces the evolution of gas, stars, dark matter, and massive black holes, and implements physically motivated models for gas cooling, star formation law, stellar and AGN feedback. The free parameters of the feedback models were tuned to match the SMF and black hole mass - stellar mass relation at $z \sim 0$. The simulation starts from $z = 127$ and adopts cosmological parameters from *Planck*: $(\Omega_m, \Omega_\Lambda, h) = (0.307, 0.693, 0.678)$ (Planck 2014). We use their simulation of the largest volume of $(100\mathrm{Mpc})^3$ for our analysis. It contains $\sim 10,000$ galaxies with stellar masses similar to or above that of the Milky Way. Unfortunately, recalibrating their result to account for different cosmology is not trivial, since the impact of changing the parameters to the mass function is highly non-linear in principle. However, the other uncertainties that enter the models or the data must overpower the change in cosmology. We thus do not attempt any recalibration of the simulation results to account for the difference in cosmology. The

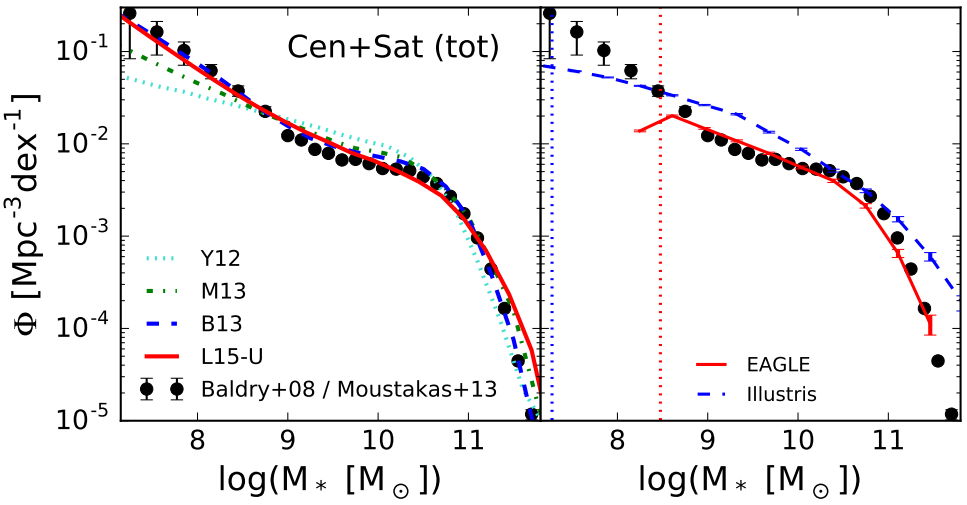

Figure 3.2. The observed stellar mass function of galaxies (data points) in comparison with the predictions of individual empirical models (left) and hydrodynamical simulations (right), as indicated in the panels. The vertical lines in the right panel show the resolution limits of the two simulations, as given in the original papers describing the simulations. The Poisson errors are presented for the simulations.

Chabrier (2003) IMF and the spectral synthesis model of Bruzual & Charlot (2003) were assumed. We bin their stellar masses to get the stellar mass function.

3.3 Comparisons with observational data

3.3.1 The field stellar mass function of galaxies at $z \sim 0$

We use the local ($z \approx 0.1$) SMF obtained from the combination of the results obtained by Baldry et al. (2008) and Moustakas et al. (2013). The data for stellar masses below $M_* \approx 10^9 \mathrm{M}_\odot$ is from Baldry et al., while the data at larger stellar masses is from Moustakas et al.. Here we briefly summarize the methodologies with which the SMFs were computed, and refer the reader to their original papers for details.

Baldry et al. used the New York University Value-Added Galaxy Catalogue (NYU-VAGC; Blanton et al. 2005), which includes $49,968$ galaxies at $z < 0.05$, to construct the local SMF. They adopted the stellar mass estimates from Kauff-

mann et al. (2003), Gallazzi et al. (2005), and Panter et al. (2007). In the data set we use, the stellar mass estimates are corrected to a Chabrier (2003) IMF.

Moustakas et al. estimated the local SMF using galaxies cross-identified between the Sloan Digital Sky Survey Data Release 7 (SDSS DR7; Abazajian et al. 2009) and the Galaxy Evolution Explorer (GALEX; Martin et al. 2011) Deep Imaging Survey. This results in $\sim 170,000$ galaxies with a total sky coverage of 2505 deg^2. Near-infrared photometry of these galaxies was obtained from the Two Micron All Sky Survey (2MASS; Skrutskie et al. 2006) and the Wide-field Infrared Survey Explorer (WISE; Wright et al. 2010). The photometry in a total of 12 bands (near-UV and far-UV of the GALEX, *ugriz* bands from SDSS model magnitudes, JHK_s magnitudes from the 2MASS, and the integrated photometry at 3.4 and 4.6μm from the WISE All-Sky Data Release) was used to infer the galaxy stellar masses from spectral energy distribution (SED) modelling. More specifically, Moustakas et al. used the Flexible Stellar Population Synthesis model of Conroy et al. (2009), a Chabrier (2003) initial mass function (IMF), exponentially declining star formation histories (SFHs), and the dust attenuation curve of Charlot & Fall (2000), to model the SEDs of individual galaxies. The SMF obtained by Moustakas et al. (2013) is in good agreement with some previous measurements, such as those of Cole et al. (2001), Li & White (2009), and Baldry et al. (2012). See their Appendix B for detailed analyses how variations in the IMF, SFH, spectral synthesis model, and dust attenuation can affect the SMF obtained.

The left panel of Figure 3.2 compares the predictions of the empirical models with the observed local SMF described above. As one can see, the prediction of the Y12 model is too flat in the low-mass end to match the upturn seen in the observation. This discrepancy owes partly to the simple functional form (a double power-law) they adopted for the stellar mass-halo mass relation, and partly to the SMFs that they used as observational constraints. In fact, Y12 found that the two sets of SMFs at

high redshifts that they adopted led to significant differences in the inferred values of model parameters. The results used here are the predictions of 'SMF2' referred in the original paper. The model of M13 also predicts a shallower faint-end slope than the observational data. Similar to Y12, M13 also adopted a simple double power-law form for the stellar mass - halo mass relation, and used a local SMF that has shallower faint-end slope than the one adopted here to constrain their parameters. In contrast, the prediction of B13 matches well the observed SMF, even in the faint end. B13 adopted a rather flexible functional form for the stellar mass - halo mass relation, which is probably required to match the faint-end upturn in the SMF. In addition B13 adopted the combined SMF of Baldry et al. (2008) and Moustakas et al. (2013) as one of their observational constraints, and so the good match between the model prediction and the data is not surprising. The prediction of the L15 model also matches well the observational data. Note that L15 used the SMF of Baldry et al. (2012) as an observational constraint. Their SMF extends only to $10^{8.5}M_{\odot}$ and so the faint-end upturn is not well represented. The faint-end upturn predicted by L15 is largely due to the CSMF of galaxies in rich clusters, as given by Popesso et al. (2006), they adopted to constrain their model.

The right panel of Figure 3.2 compares the numerical simulation results with the observational data. Illustris simulation produces too many galaxies in the intermediate mass range as well as in the massive end, but too few low-mass galaxies. The overall shape of the predicted SMF is very different from that of the observed SMF. On the other hand, the prediction of EAGLE simulation matches the observational data reasonably well above the resolution limit. This may not be very surprising, because the free parameters in EAGLE simulation were tuned to match local observations. Unfortunately, the relatively poor mass resolution does not allow us to investigate whether a faint-end upturn is predicted in the simulation.

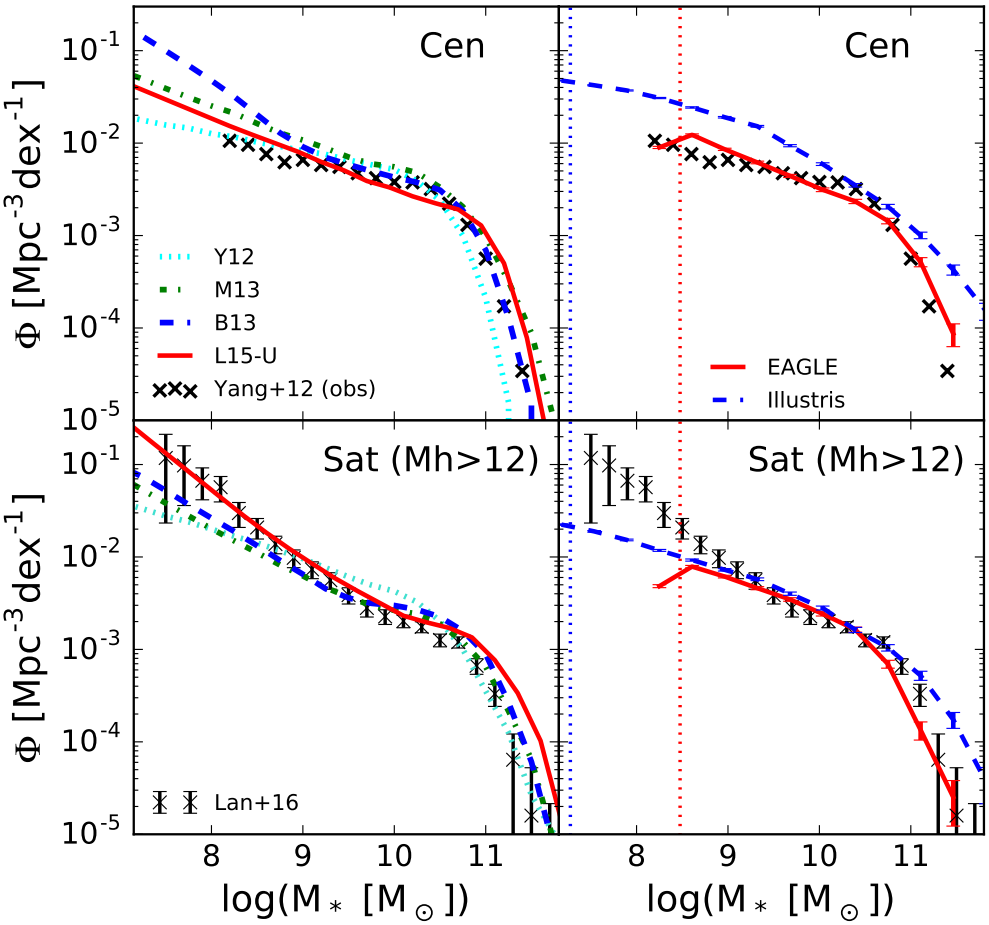

Figure 3.3. The observed stellar mass function of central (data points in the upper two panels) and satellite (data points in the lower two panels) galaxies, in comparison with the predictions by individual empirical models (left panels) and gas simulations (right panels), as indicated. The completeness in stellar mass from the observation of centrals is not guaranteed for $M_* < 10^8 \mathrm{M}\odot$. The vertical lines in the right panels show the resolution limits of the two simulations, as given in the original papers describing the simulations.

3.3.2 Central and satellite galaxies

Using the group memberships provided by Yang et al. (2007) group catalog (see next subsection for more details), we can separate galaxies into two populations, centrals and satellites. A central galaxy is defined to be the most massive member in a group, while all other members in a group are called satellites. The CSMFs can

then be estimated separately for the centrals and satellites. Formerly the total SMF can be expressed in terms of these conditional functions as

$$\Phi_{\rm tot}(M_*) = \int_{M_{h,min}}^{\infty} dM_h\ n(M_h) \times \{\Phi_{\rm cen}(M_*|M_h) + \Phi_{\rm sat}(M_*|M_h)\}\,, \tag{3.6}$$

where $\Phi_{\rm cen}(M_*|M_h)$ and $\Phi_{\rm sat}(M_*|M_h)$ are the CSMFs of the centrals and satellites, respectively, in halos of mass M_h, while $n(M_h)$ is the halo mass function, which is the number density of halos of masses between M_h and $M_h + dM_h$. In Lan et al. (2016), the CSMFs are given only for satellites in groups with halo masses above $10^{12}{\rm M}_\odot$ (see the next subsection for details). The satellite SMF used here is obtained directly from their measurements by summing up the CSMFs of such halos. For central galaxies, we use the results obtained by Yang et al. (2012) from their group catalog. Since the group catalog is based on the SDSS spectroscopic data, the central SMF was measured only for galaxies above $10^8{\rm M}_\odot$ (see table 6 in their paper).

The data points in Figure 3.3 show the SMFs for central and satellite galaxies, respectively. Separating galaxies into centrals and satellites provides more information about the galaxy population than the total SMF alone, and Figure 3.2 and Figure 3.3 demonstrate this point clearly. For instance, although the empirical model by B13 (see §3.2) matches well the faint-end upturn in the observed total SMF, this match is now revealed as due to an excess in the SMF of central galaxies combined with a deficit in the SMF of satellite galaxies. The M13 model has similar problems; it under-estimates the number of satellite galaxies at the low-mass end even more strongly than B13. The Y12 model matches the central SMF reasonably well, but it fails to reproduce the strong upturn in the low-mass end seen in the observed SMF of satellite galaxies. Overall, the L15 model can match both the observed central and satellite SMFs, although some discrepancies in details can still be seen. This match is not trivial, because these observations were not used as constraints in L15.

The comparisons of the two gas simulations with the observational results are shown in the right two panels of Figure 3.3. Here we see that the EAGLE simulation matches the observational data reasonably well above its mass resolution limit. Illustris simulation matches the SMF of satellites only in the intermediate mass range; it over-predicts the central SMF over almost the entire mass range, except at the knee of the SMF.

3.3.3 The conditional stellar mass functions of galaxies in groups

We use the CSMFs obtained by Lan et al. (2016) as our main data set to compare with models. Here we summarize briefly their methodology and results. Lan et al. used galaxy samples from the NYU-VAGC, which is based on the Sloan Digital Sky Survey Data Release 7 (SDSS DR7; Abazajian et al. 2009). A K-correction was applied using the model of Blanton et al. (2003). In order to associate galaxies with clusters/groups of galaxies, they adopted the group catalog of Yang et al. (2007), which was constructed by applying the halo-based group finder developed by Yang et al. (2005) to the SDSS DR7. The group finder assigns galaxies into halos using certain criteria in phase space, and galaxies residing in a common halo are considered to be members of the same group. More specifically, a tentative halo mass is assigned to a tentative group based on the galaxies that have already been assigned to the group, assuming a monotonic relation between the total stellar mass of all assigned members with $M_r < -19.5$ and halo mass. The tentative mass is then used to estimate the virial radius and velocity dispersion of the halo, which in turn are used to update the group membership. The procedure is iterated until both group memberships and halo masses converge for all groups. Yang et al. (2007) used mock catalogs constructed from N-body simulations to show that the dark matter halo masses estimated in this way are consistent with those directly obtained from the simulations, with scatter of $\sim$ 0.3 dex over three orders of magnitude in halo masses that cover the mass

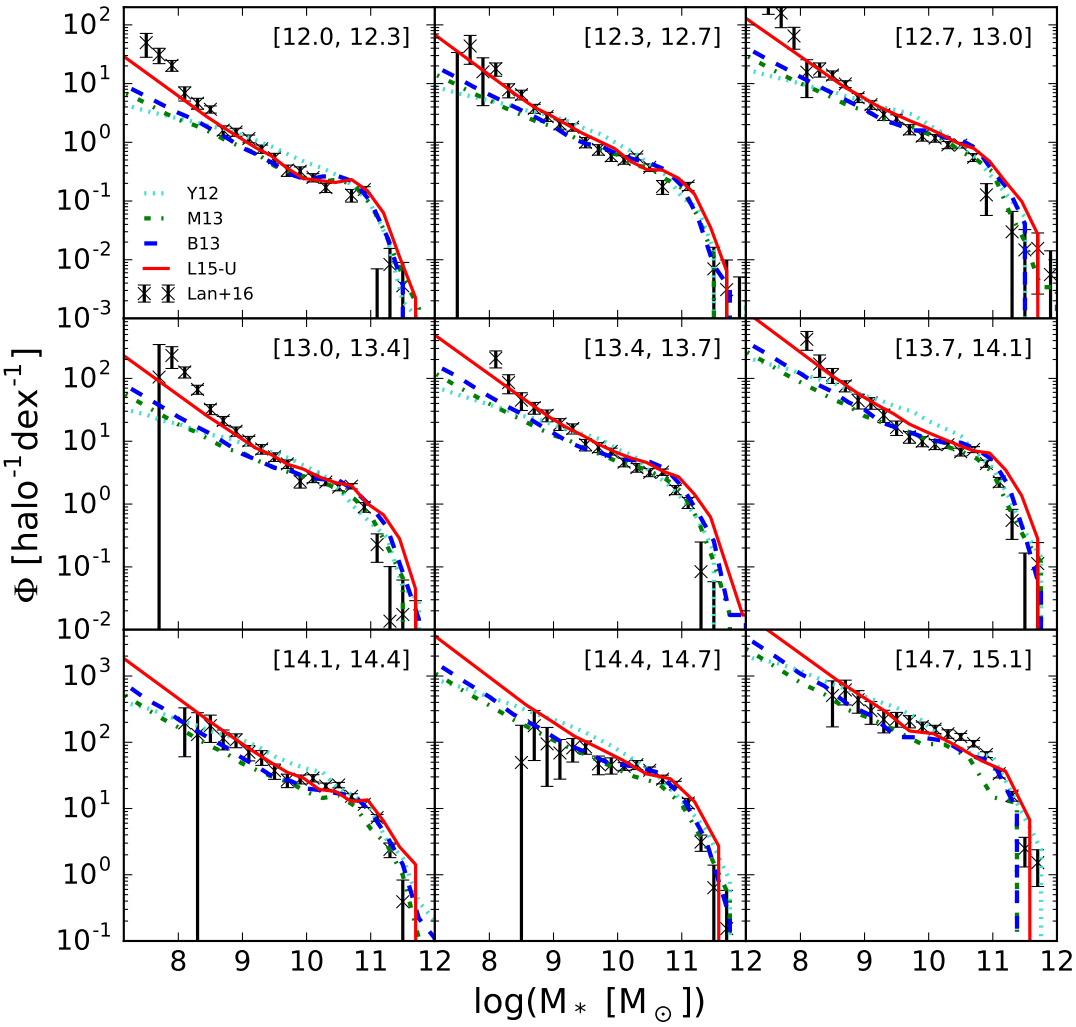

Figure 3.4. The observed (data points) and predicted (lines) conditional stellar mass functions of galaxies in groups of different halo masses, as indicated in individual panels.

range relevant to our analyses. When we compare the CSMFs obtained from models with the observational results, an uncertainty of ~ 0.3 dex is included in the model predictions. The halo masses used here are M_{200}, the total mass enclosed by a radius, r_{200}, within which the average density is 200 times the mean density of the universe.

Lan et al. used only groups at $z < 0.05$, where halos with masses of $M_{200} > 10^{12}\mathrm{M}_{\odot}$ are complete. To limit the uncertainty in redshifts due to peculiar velocities, they also eliminated groups at $z < 0.01$. With the groups and their positions identified

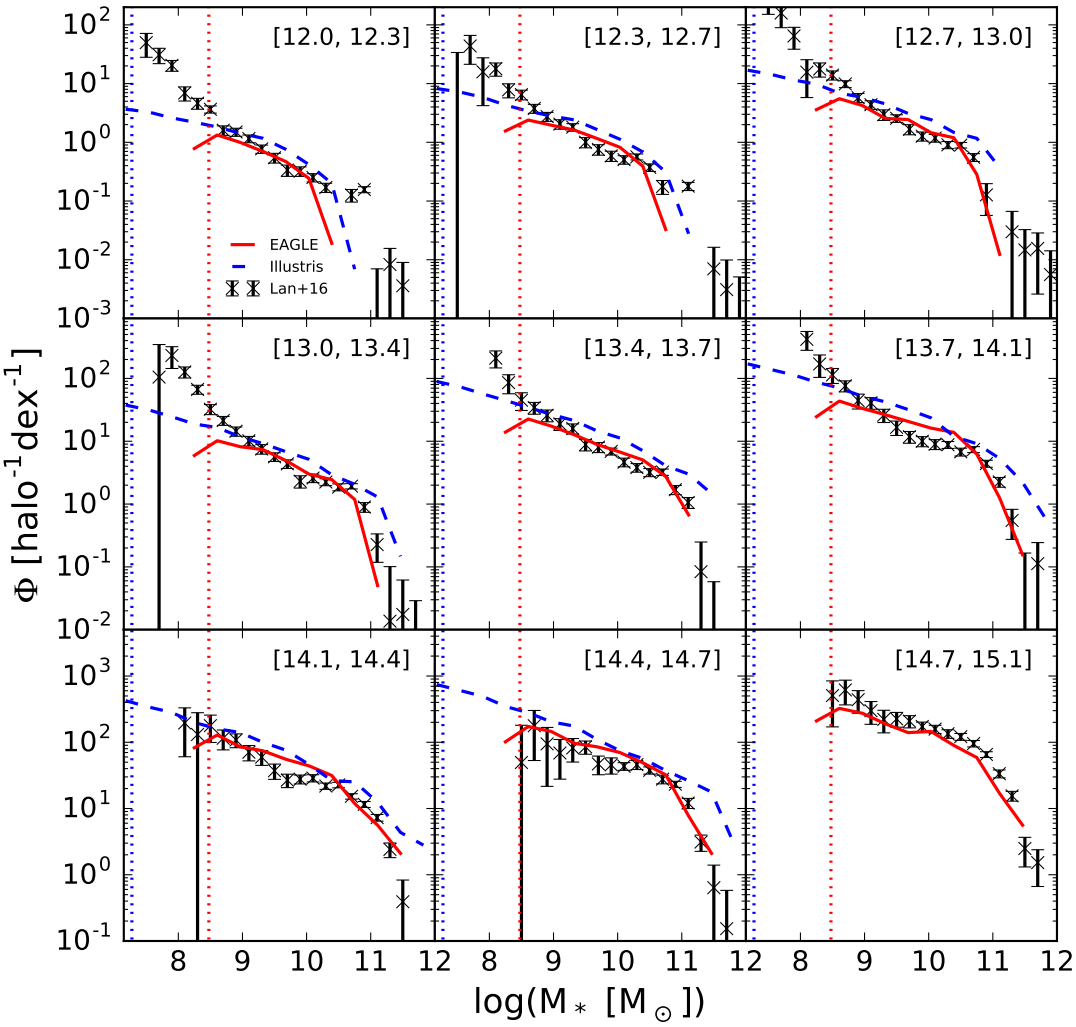

Figure 3.5. The comparison between the observed conditional stellar mass functions of galaxies (data points) with the results of EAGLE and Illustris simulations (lines), for groups of different halos masses, as indicated in each panel. The two vertical lines indicate the mass limits of the two simulations.

in the SDSS DR7 survey area, Lan et al. estimated the excess of galaxy number in each luminosity bin within a projected distance of r_{200} of each group. The conditional luminosity function (CLF) of galaxies is then obtained by averaging galaxy counts within all groups of a given halo mass, with subtractions of the background and projection effects due to clustering on large scales (see Lan et al. 2016, for the detail). Lan et al. applied this method to the photometric sample of SDSS DR7, down to

a r-band model magnitude of 21. This corresponds to $M_r \approx -12$ (or $L \approx 10^7 L_\odot$) and $M_r \approx -14$ at $z = 0.01$ and 0.05, respectively. Since the number density of more massive halos is smaller, Lan et al. was able to estimate the CLF down to $M_r \sim -12$ for low-mass halos ($M_{200} \sim 10^{12}\mathrm{M}_\odot$), but only to $M_r \sim -14$ for massive halos.

To convert their CLFs into the corresponding CSMFs, we use a mass-to-light relation based on galaxy colors and luminosities (e.g. Bell et al. 2003) to obtain the stellar masses of galaxies. However, the uncertainty in the observed galaxy colors, especially for faint galaxies, may bias the stellar mass estimates and, therefore, the stellar mass functions. To reduce such bias, we first separate galaxies into blue and red populations by using the u-r color separation suggested by Baldry et al. (2004) [see their equation (11)]. We then use the observed luminosity of a galaxy and the mean u-r color for the galaxy population at that luminosity, instead of the observed color of the galaxy, to estimate the stellar mass. The mean u-r color-luminosity relations for the blue and red populations are derived in Lan et al. (2016) [their equations (C2) and (C3)] based on the same data set. We have made tests either by using the observed galaxy color or by artificially introducing some uncertainties in the galaxy color, and found that all these do not lead to any qualitative change of our results. Note again, as described in §3.2, our model predictions for the stellar masses of individual galaxies also include some uncertainties in the stellar mass estimates to mimic the uncertainties in the observational stellar masses. Lan et al. adopted a Kroupa (2001) IMF for the CSMFs.

With the estimated stellar masses of individual galaxies, we measure the CSMFs using the same method Lan et al. did for the CLFs. The stellar mass functions are measured down to the limiting stellar masses at which both the stellar masses of blue and red galaxies derived from the flux limit photometric sample ($r < 21$) are complete. In addition, the limiting stellar mass bins are selected to ensure that

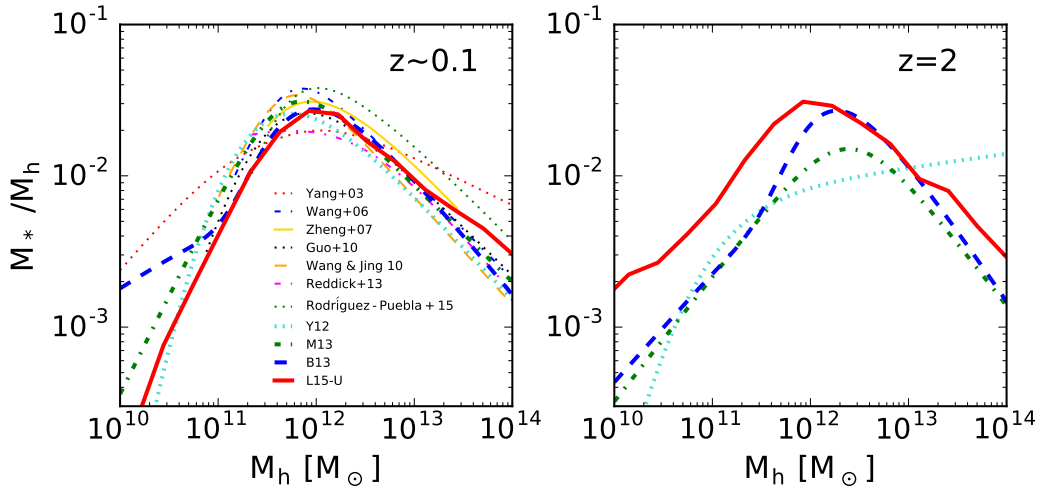

Figure 3.6. The average stellar mass - halo mass relation for central galaxies from the empirical models considered in this paper, along with the results in the literature from recent studies that adopted empirical approaches such as halo abundance matching, conditional luminosity function and halo occupation distribution, for local Universe (left) and $z = 2$ (right).

they each contain at least five groups. We bootstrap the group catalog 200 times to estimate the errors in the derived CSMFs.

Figure 3.4 compares the CSMFs to those predicted by the empirical models. The predictions by the M13 and B13 models are qualitatively similar, with B13 predicting more low-mass galaxies. Both models under-predict the CSMFs at the low stellar mass ends, and the under-prediction is more significant for groups of lower halo masses. Only for massive clusters are the predictions consistent with the observational data. The predictions of Y12 are too shallow in the low mass end; the model systematically under-predicts the CSMF at the low mass end and over-predicts that in the intermediate mass range. In particular, Y12 does not predict any upturn seen in the data.

The L15 model matches the overall behaviors of the CSMFs over the entire halo mass range. It also matches the CSMFs in detail for most of the halo mass bins. However, the low-mass upturn it predicts for more massive halos may be too steep,

especially for the two most massive samples. As mentioned above, the L15 model used the composite CLF of galaxies in rich clusters given by Popesso et al. (2006) as one of the constraints on their model. The faint-end upturn in this composite CLF is significantly steeper than that of Lan et al. used here. The over-prediction is therefore due to the observational data which the model was tuned to match with. The model seems to under-predict the CSMF at the low-mass end in two mass bins: the lowest mass bin of $\log(M_h/\mathrm{M}_\odot) = [12.01, 12.34]$, and the intermediate mass bin of $\log(M_h/\mathrm{M}_\odot) = [13.03, 13.37]$. It is unclear if these discrepancies are due to random fluctuations in the data, or indicate that the L15 model has to be modified to accommodate the data. We will come back to this in the following section.

Figure 3.5 compares the observational data with the two gas simulations. Illustris simulation mismatches the observation over a wide range of stellar masses for almost all the halo mass bins. Overall, the simulation significantly under-predicts the CSMFs at the faint ends, and over-predicts them in both the massive and intermediate mass ranges. EAGLE simulation appears to be in better agreement with the observation, except that it does not reproduce sufficient number of massive galaxies in low-mass halos. Unfortunately, its mass resolution prevents us from probing its behavior at the faint end.

3.4 Model predictions

In this section, we compare all the empirical models in their predictions for the stellar mass - halo mass relation, the star formation rates and stellar masses in halos of different masses at different redshifts, and for the SMFs of high-redshift galaxies.

3.4.1 Stellar mass - halo mass relation

Figure 3.6 shows the stellar mass - halo mass relation predicted by our updated model, in comparison with the predictions of the other three empirical models con-

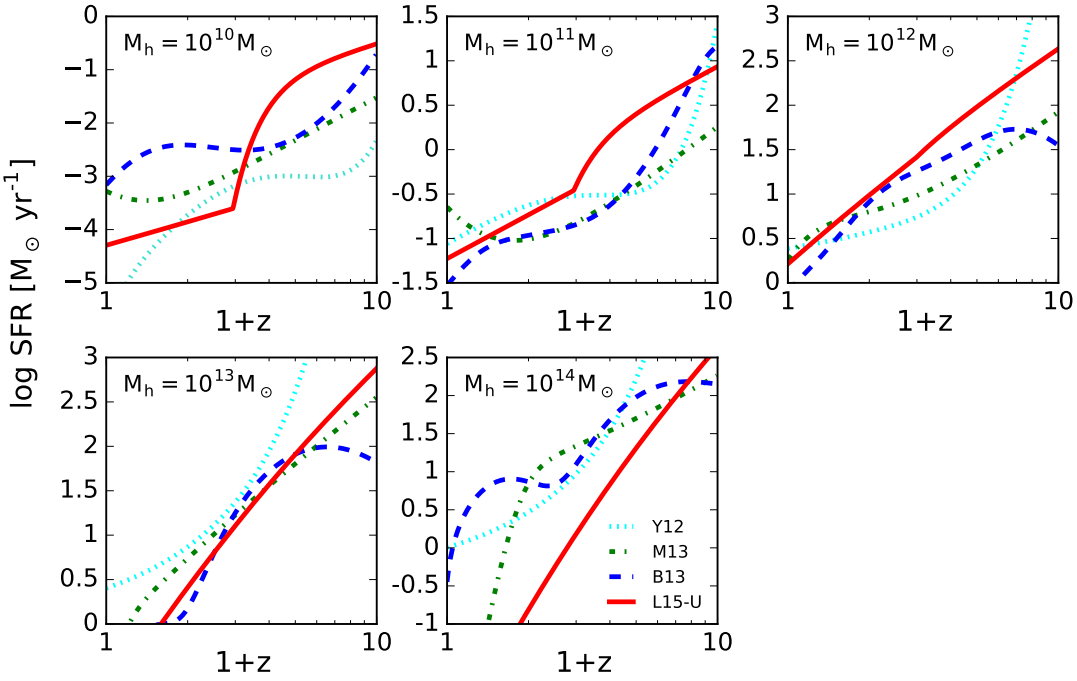

Figure 3.7. The average star formation rate of central galaxies as a function of redshift for halos of different masses, as predicted by various empirical models, as indicated.

sidered here and the results from the literature. Different models made different assumptions for conversions from luminosity to stellar mass, for prescription of scatter in the relation, and cosmological models. They also employed different observations as constraints for the models. Given all these differences, it is remarkable that the predictions of most models are consistent with each other within ~ 0.2 dex at $z \sim 0.1$ for a large range of halo masses. All models predict a characteristic mass scale, $M_h \sim 10^{12}$M$\odot$, at which the stellar mass to halo mass ratio peaks. Among the more recent results, B13 is an exception in that it predicts a strong upturn at the low-mass end. The earlier result of Yang et al. (2003) was obtained by using their luminosity - halo mass relation together with the assumption of a constant stellar mass to luminosity ratio, $M/L = 1.8$ M$_\odot/L_\odot$ (in the b_J band of 2dFGRS which they used to constrain their model).

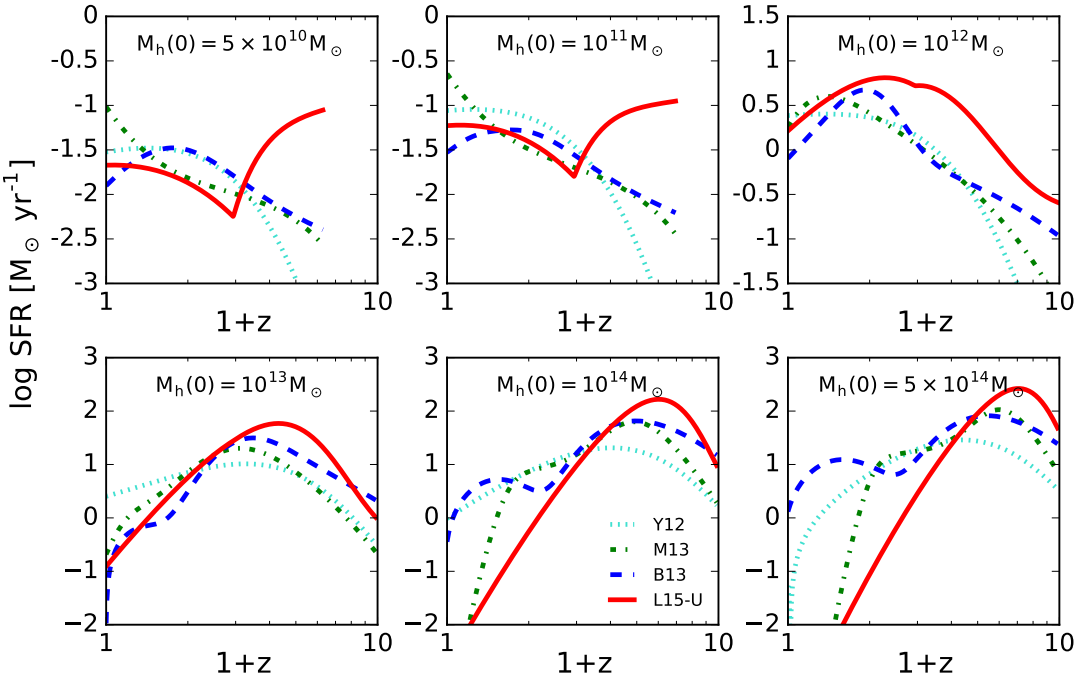

Figure 3.8. The average star formation rate of central galaxies as a function of redshift for halos of different present-day masses, predicted by the empirical models, as indicated.

At higher redshift, however, the predictions by different empirical models differ significantly. In particular, the update of L15, L15-U, predicts a much higher star formation efficiency for low mass halos at high redshift, because of the boost of star formation rate at $z > z_c$ in low mass halos to match the upturns in the CSMFs.

3.4.2 Star formation histories in dark matter halos

Figure 3.7 compares the empirical models in terms of their predictions for the average star formation rate (SFR) of central galaxies in halos at different redshifts. Some models predict complicated star formation histories that are clearly due to over-fitting of the observational data. The L15-U model predicts much higher SFRs at $z \geq 2$ in low-mass halos than other models, which is clearly a consequence of the strong upturns at the faint-ends of the CSMFs used to constrain the model. The existence

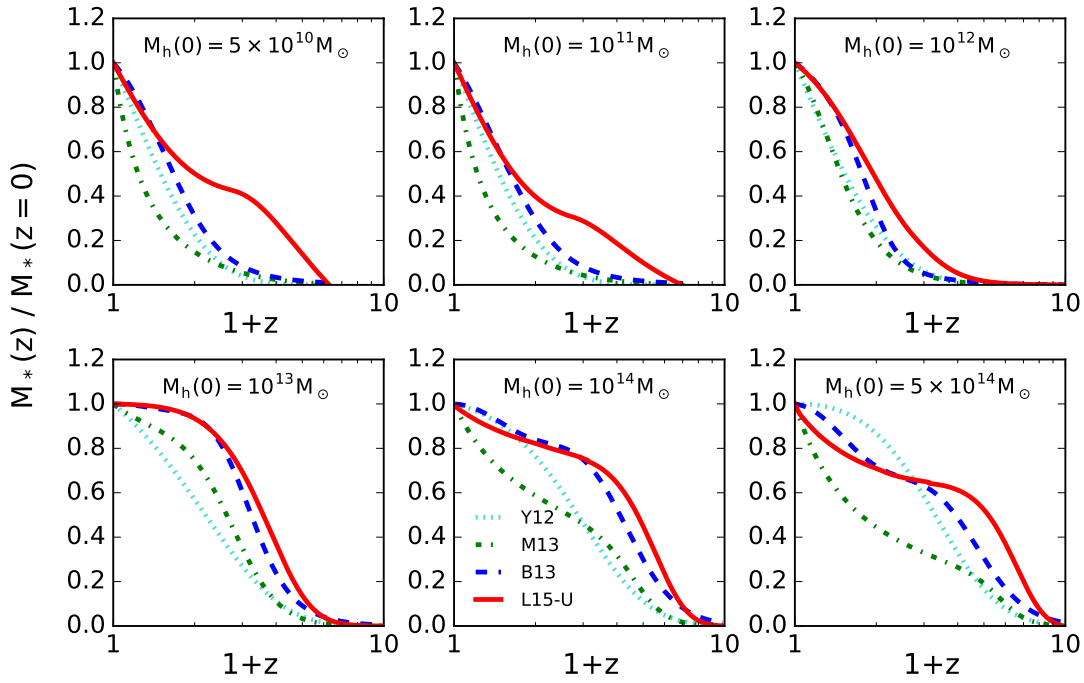

Figure 3.9. The average stellar mass assembly history of central galaxies as a function of redshift for halos of different present-day masses, predicted by various empirical models as indicated.

of a characteristic redshift, $z \sim 2$, is clearly seen in halos with $M_h < 10^{12}\,\mathrm{M}_\odot$, and its physical implications will be discussed later.

Figure 3.8 shows the model predictions for the average star formation histories of central galaxies in halos of different present-day masses. The predictions of different models are very different. In particular, for present-day dwarf galaxies that reside in halos of $M_h(0) < 10^{11}\mathrm{M}_\odot$, L15-U predicts a very active star formation episode at $z > 2$. In contrast, most of the stars in such halos are formed at $z < 2$ in all other models. This difference has other observational consequences. Indeed, as discussed in Lu et al. (2014) and Lu et al. (2015), the early starburst in low-mass halos predicted by L15 is consistent with the observations that a significant fraction of old stellar population exists in local dwarf galaxies (e.g. Weisz et al. 2011) and that the star formation rate function at the low-rate end is very steep at $z > 4$ (e.g. Smit

et al. 2012). For Milky-Way sized halos, the star formation history predicted by L15-U is broader than those predicted by the other three models. For massive halos with $M_h(0) \geq 10^{14}\mathrm{M}_\odot$, L15-U predicts a decline of the SFR with decreasing redshift starting from relatively high redshifts, in contrast to the predictions of B13 and Y12 that the star formation rates remain relatively high all the way to the present time, and to the prediction of M13 that a rapid decline only occurs at $z < 1$.

All these results demonstrate that different empirical models can make vastly different predictions for the star formation histories for present-day galaxies, even though all the models are tuned to match the observed SMFs.

3.4.3 Stellar mass assembly histories

Figure 3.9 shows the average stellar mass assembly histories for the central galaxies in halos of different present-day masses predicted by different models. The model predictions take into account *in situ* star formation, accretion of satellites, and stellar mass loss due to stellar evolution. Again, for low-mass halos, where the increase of stellar mass is dominated by *in situ* star formation (Lu et al. 2015), L15-U is distinct from the other models in that about half of their stellar mass at the present was already in place by $z \sim 2$ via star formation (see Figure 3.8).

For Milky-Way sized halos, however, the differences between the model predictions are milder. All the models predict that about half of stellar mass was in place by $z \sim 1$. There is a significant difference between L15-U and other models at high z. For example, L15-U predicts that about 15% of the final stellar mass was assembled by $z \sim 2$, while less than 10% was predicted by the other models.

For central galaxies in present-day massive halos with $M_h(0) > 10^{14}\mathrm{M}_\odot$, the predictions of different models again become very different. M13 predicts a much later assembly for these galaxies than any other models. The predictions of B13 look similar to L15-U, but the increase in stellar mass with time is due to different reasons.

While L15-U predicts that the increase at $z < 2$ is dominated by accretion of stars from satellites, B13 predicts that a significant fraction of the increase at $z < 2$ is actually due to *in situ* star formation (see Figure 3.8). This difference is again due to the boost of star formation in low-mass halos at high z in the L15-U model. The increased amount of stars formed in progenitors at high z makes the accretion of stars more important in the growth of stellar mass in a massive galaxy, and the fraction of stars formed in situ has to be decreased proportionally in order to match the final stellar mass of the galaxy. The results demonstrate the importance of properly modeling the star formation in low-mass progenitors at high z in order to understand the star formation and stellar mass assembly histories of massive galaxies at the present day.

3.4.4 Stellar mass functions of high-redshift galaxies

Figure 3.10 shows the predictions of the empirical models for the stellar mass functions of galaxies at a number of redshifts. The predictions of B13 and M13 are similar in both slopes and amplitudes at the low mass ends, but B13 predicts many more massive galaxies than M13, particularly at high redshifts. Y12 predicts significantly flatter slopes at the low-mass ends, and more galaxies in the intermediate mass range, than the other three models. The stellar mass functions predicted by the L15-U model match the predictions of B13 at $M_* > 10^{10}\mathrm{M}_\odot$, but are significantly steeper at the low-mass ends.

We select some observational SMFs of high redshift galaxies from the literature to compare with the model predictions. Specifically, we use the SMFs at $1.3 < z < 3.5$ given by Pérez-González et al. (2008) and Marchesini et al. (2009). Pérez-González et al. used a sample combining data in three different fields with a total area of 664 arcmin^2 that have a total $\sim 28,000$ systems selected with the $3.6-4.5\mu m$ photometry of *Spitzer Space Telescope* (Werner et al. 2004). The sample is complete

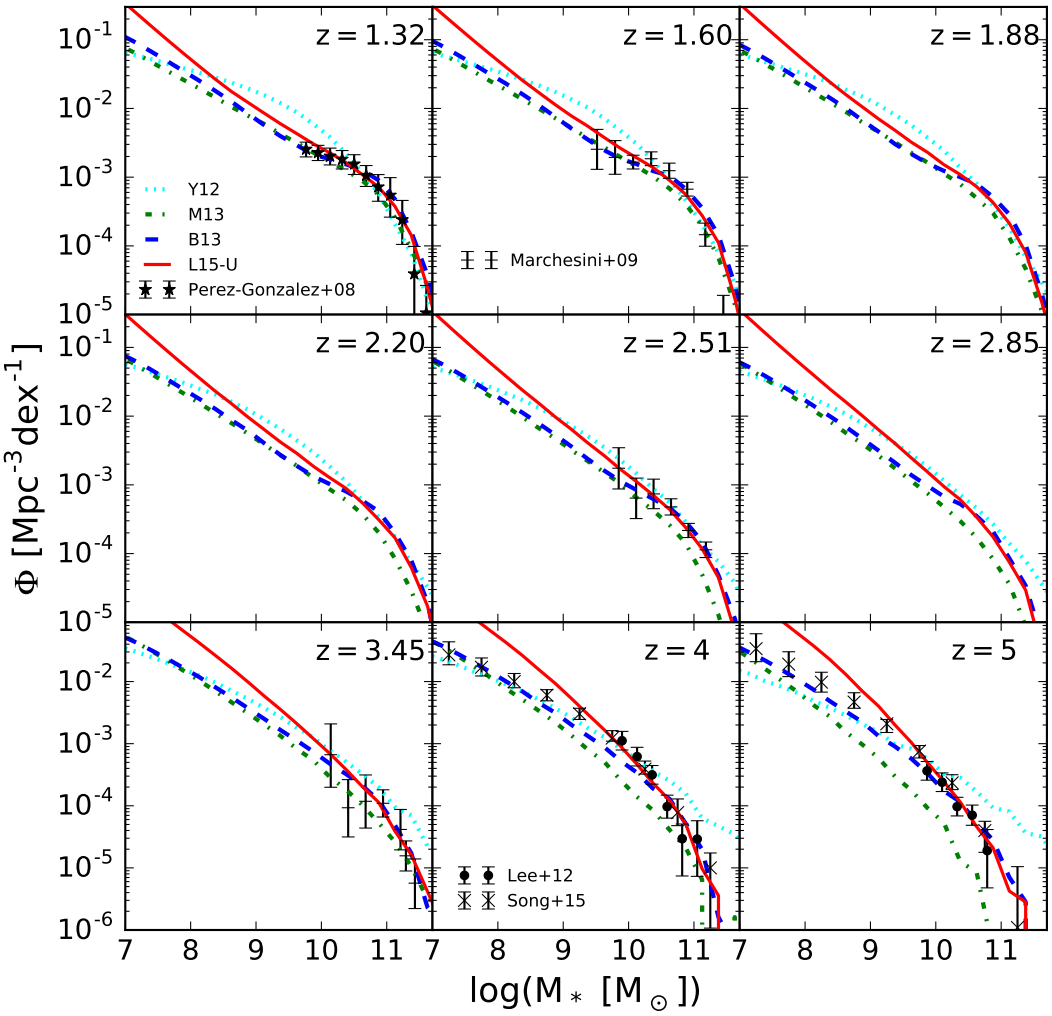

Figure 3.10. The model predictions for the field stellar mass functions at high redshifts (solid) in comparison with observations.

down to $M_* = 10^{10}\mathrm{M}_\odot$. Marchesini et al. combined data from the deep NIR MUSYC, the ultra-deep FIRES, and the GOODS-CDFS surveys to derive the SMFs from the optical to MIR broad bands photometry. Pérez-González et al. assumed a Salpeter (1955) IMF while Marchesini et al. adopted a pseudo-Kroupa (2001) IMF. It is known that the stellar mass estimated using a Salpeter IMF is roughly a factor of 1.4 higher than that given by a pseudo-Kroupa or Chabrier IMF, and we correct all the stellar masses to the IMF we adopt here. For the SMFs at even higher redshifts, $z = 4 - 5$,

we use the results by Lee et al. (2012) and Song et al. (2016). Both B13 and L15-U match the observational data well at $M_* > 10^9 M_\odot$, while the other two models match the high-z data poorly at the high-mass end. The prediction of the L15-U is significantly steeper than the observational results given by Song et al. at $z > 4$. If the high-z SMFs are as shallow as those given by Song et al., then there may be a tension between the observed CSMFs at low z and the observed SMFs at high z, at least within the model family represented by the halo-based empirical model of L15.

3.5 Summary and discussion

Galaxy formation and evolution within the current cosmological frame are controlled by a number of physical processes, many of which are still poorly understood from first principles. In the absence of a proper understanding of these processes, halo-based empirical models provide a useful way to establish the link between galaxies and CDM halos purely on the basis of observations and the current cosmology. In this paper we use a variety of galaxy stellar mass functions to test a number of popular empirical models. In particular, we focus on using the conditional stellar mass functions (CSMFs) of galaxies in galaxy groups as obtained by Lan et al. (2016) to test the models. We find that the CSMFs predicted by different models can be very different, even though they are all tuned to match the observed stellar mass function of the total galaxy population. This clearly demonstrates the power of the CSMFs in constraining models. Since the CSMFs are measured from observations in the nearby Universe, the samples that can be used are larger, and the stellar mass functions can be measured to the low-mass ends. As the galaxies that reside in present-day galaxy systems, such as clusters and groups of galaxies, are expected to have formed at various redshifts, the CSMFs in groups/halos of different masses carry important information about galaxy formation in dark matter halos at different redshifts.

The CSMFs are then used as constraints to update the original model by Lu et al. (2014, 2015). The model parameters obtained here are very similar to those obtained in the original paper which uses a completely different set of observational constraints, demonstrating that the different data sets are consistent with each other. The observational constraints clearly prefer a model in which star formation in low-mass halos changes behavior at a characteristic redshift $z_c \sim 2$. There is also a tentative evidence that this characteristic redshift depends on environments, becoming $z_c \sim 4$ in regions that eventually evolve into rich clusters of galaxies. However, given the uncertainties of the current observed CSMFs in the low-mass ends, this environmental dependence of z_c needs to be confirmed with better data.

We compare the predictions of a number of popular halo-based empirical models and two numerical simulations of galaxy formation. We find that the two numerical simulations fail to match the observational data one way or another. The empirical models by Yang et al. (2012) and Moster et al. (2013) fail to reproduce the faint-end upturn of the field SMFs from observations. The model by Behroozi et al. (2013) reproduces the faint-end upturn, but it is a combined result of over-prediction for central galaxies and under-prediction for satellites at the faint-end. In contrast, the model by Lu et al. (2014, 2015) matches reasonably well the CSMFs in halos of different masses. The Lu et al. model predicts a much higher star formation efficiency than the other models for low-mass halos at redshifts higher than a characteristic redshift after which the star formation is suppressed.

We use our constrained model to make predictions for a number of statistical properties of the galaxy population. These include the stellar mass functions of galaxies at high z, the stellar mass - halo mass relations at different redshifts, and the star formation and stellar mass assembly histories of galaxies in dark matter halos of different masses. A comparison of our model predictions with those of other empirical models shows that different models can make vastly different predictions

for these properties, even though all of them are tuned to match the observed stellar mass functions of galaxies. In particular, our constrained model predicts a much higher *in situ* star formation rate at $z \geq 2$ for present-day dwarf galaxies than the other models. As a result, such galaxies have about 40% of their current-day stellar mass already in place by $z \sim 2$. Because of this boosted star formation in low-mass halos at high z, the role of accretion of stars from satellite galaxies, relative to *in situ* star formation, in the build up of massive galaxies is more important in our model than in the other models.

One of the main predictions of our constrained model is the existence of a characteristic redshift that separates an active star formation phase from a subdued star formation phase in low-mass halos. This change in star formation mode is likely related to the feedback processes that regulate star formation. As discussed in Lu et al. (2014, 2015), energy feedback from stars and AGNs associated with active star formation and super-massive black hole accretion at high redshift may preheat the gas media around dark matter halos and suppress gas accretion and star formation at lower redshift (Mo & Mao 2002, 2004). Based on plausible assumptions about the star formation histories of the universe and the density of the intergalactic medium, the pre-heating is expected to occur around $z = 2 - 3$, and the specific entropy of the preheated gas is $\sim 10\mathrm{KeVcm}^2$, which is important in affecting star formation in low-mass halos, because of their relatively shallow gravitational potential wells, but has no significant effects on halos with masses above $\sim 10^{12}\,\mathrm{M}_\odot$ (e.g. Lu & Mo 2007). This preheating may also explain why the cold gas mass function at $z \sim 0$ is shallow (Mo et al. 2005). In such a scenario, the pre-heating is expected to occur earlier in regions occupied by present-day massive halos, because intensive star formation and AGN activity are expected to occur earlier in higher density regions where gravitational collapse is more accelerated. Our tentative finding of the positive dependence of the characteristic redshift on halo mass is in agreement with such an expectation,

but better observational data are needed in order to examine such dependence in more detail.

CHAPTER 4

LARGE-SCALE TIDAL FIELD, A KEY TO UNDERSTANDING ALIGNMENTS OF GALAXIES

Using the SDSS DR7 galaxy and group catalogs by Yang et al. and the large scale tidal field constructed by Wang et al., we explore various alignments of galaxies with their large scale structures characterized by the tidal field. We find that both orientation and spatial distribution of galaxies tend to preferentially align with the tidal field. The orientation alignment is found stronger for centrals (the brightests in groups) due to weaker non-linear effects on them, which weakens such alignment. In massive halos, the spatial distribution of satellites in outer regions of halos is more aligned with the tidal field, but less aligned with the orientation of centrals than those in inner regions. The alignment between the orientation and the spatial distribution of galaxies is also found as a combined result of their alignments with the tidal field. Lastly, we probe alignment between the tidal fields of group pairs as a function of separation and find that it extends out to $80 - 100h^{-1}$Mpc while the alignment of groups in sheets continues until even larger scales. Similar alignment of the position angles between central galaxy pairs also is detected but at much weaker strength. All these results are consistent with a scheme on the galaxy formation that galaxies form via coherent but anisotropic mass inflow along the large scale structures, and that the large scale tidal field is a reliable measure to quantify and characterize the large scale structures.

4.1 Introduction

In the standard paradigm, galaxies form in dark matter halos by gravitational collapse of the cosmic density field (e.g., Mo et al. 2010). Being formed via flow of mass accretion, properties of halos such as shape, angular momentum, and formation time are expected to be affected by their environments. Therefore, being formed inside dark matter halos, properties of galaxies may also reflect such environmental effects, which can give us important implications on galaxy formation processes. Especially, given that galaxy formation is a dynamical process via mass flow, alignments of galaxies have been vigorously studied and considered as one of the key properties characterizing galaxy formation.

The alignment of galaxies has been analyzed in many different ways. Spatial distribution of satellites was first studied and has been studied long to yield mutually conflicting results. Holmberg (1969) analyzed the spatial distribution of satellites using 218 satellites around 58 prominent spirals and claimed their preferential alignments along minor axis of the spirals. The study, however, was not only restricted by small number of samples and maximum projected distance of $\simeq$ 50kpc, but disputed by subsequent studies that failed to confirm such alignment (Hawley & Peebles 1975, Sharp et al. 1979, MacGillivray et al. 1982). A while later, Zaritsky et al. (1997) failed to observe such alignment within the scale of 200kpc, but reported the alignment on larger scales of $300 - 500$kpc. Sales & Lambas (2004) used $\simeq$ 3000 satellites from the 2dFGRS observation and claimed the same alignment on similar physical scales regardless of morphology of the centrals. However, while such alignment has been even claimed for the inner ($\leq$ 250kpc) region of the Milky Way (e.g., Majewski 1994, Kroupa et al. 2005), Brainerd (2005) and Agustsson & Brainerd (2006b) found the opposite trend, i.e. alignment along major axis by analyzing SDSS galaxies with a very similar method and similar number of samples to Sales et al. Yang et al. (2006) used about an order of magnitude more central-satellite pairs of 24728 from

the SDSS Data Release 2 and detected alignments of satellites along major axis of central galaxies at a highly convincing level. Yang et al. added that the conflicting results from the previous studies may be due to small number statistics or confusion on the definition of the position angles. Their results are in agreement with subsequent studies (e.g., Azzaro et al. 2007, Faltenbacher et al. 2007, 2009).

There have been studies on various alignments with different approaches other than the spatial distribution of satellites. Li et al. (2013) reported alignments between galaxies and the large scale structures that the galaxies reside in. A recent study by Smargon et al. (2012) confirmed such alignment among the SDSS clusters of galaxies as well. Radial alignment was also consistently found in that satellites tend to orient toward their centrals (Agustsson & Brainerd 2006a, Donoso et al. 2006, Faltenbacher et al. 2007, Okumura et al. 2009, Rong et al. 2015). The correlations between such alignments and density or galaxy properties have also been explored (e.g., Mandelbaum et al. 2006, Hirata et al. 2007, Blazek et al. 2011, Joachimi et al. 2011).

The alignments may be owing to that the collapse of the cosmic density field occurs in a coherent but yet highly anisotropic way. The quantity that best describes such collapse processes is the tidal fields. These fields are coherent on large scale but strongly anisotropic, as demonstrated in Zel'dovich approximation. The observed alignments thus may be a manifestation of galaxy alignment with the tidal field. In this context, Zhang et al. (2013) used the tidal field constructed using the method of Wang et al. (2012) to characterize the large scale structures and study the alignments between galaxy shapes and the large scale structures. They found that the strength of the alignment is stronger for more massive, redder galaxies, and galaxies in more massive halos.

In this paper, using the galaxy samples from the Sloan Digital Sky Survey Data Release 7 (SDSS DR7; Abazajian et al. 2009), we examine various alignments among

the tidal field, orientation and distribution of galaxies, with an aim to understand how the alignments may be interpreted in terms of the tidal field. We adopt as the orientation of galaxies their position angles computed by Simard et al. (2011) through the GIM2D bulge+disc decomposition. We use the tidal field constructed based on the SDSS DR7 galaxies by Wang et al. (2011).

This paper is organized as follows. In Section 4.2, we describe the observational data as well as position angles and the tidal field we use. We show our main results on different types of alignments in Section 4.3. In Section 4.4, we summarize our main conclusions.

4.2 Observational data

4.2.1 Galaxy samples

We use the New York University Value-Added Galaxy Catalogue (NYU-VAGC; Blanton et al. 2005), which is based on the SDSS DR7 but with improvements over the pipeline. From the catalog, we only select galaxies with extinction-corrected Petrosian r-band magnitude brighter than 18, redshifts $0.01 \leq z \leq 0.20$, and redshift completeness $C_z > 0.7$. We also exclude galaxies with ${}^{0.1}M_r - 5\log h \leq -22.5$, to prevent fiber-collided galaxies from being assigned much lower redshifts than their actual redshifts and thus the luminosities largely overestimated. The resulting sample consists of 639,359 galaxies. We refer to this sample as DR7 galaxies throughout this paper.

4.2.2 Galaxy morphology and position angles

We study the alignment of galaxies with the large scale environments. The orientation of a galaxy in the sky is represented by the position angle of its major axis. It is measured counterclockwise from north to east to be between 0 and 180 degree. We adopt position angles from Simard et al. (2011) where they used the GIM2D software

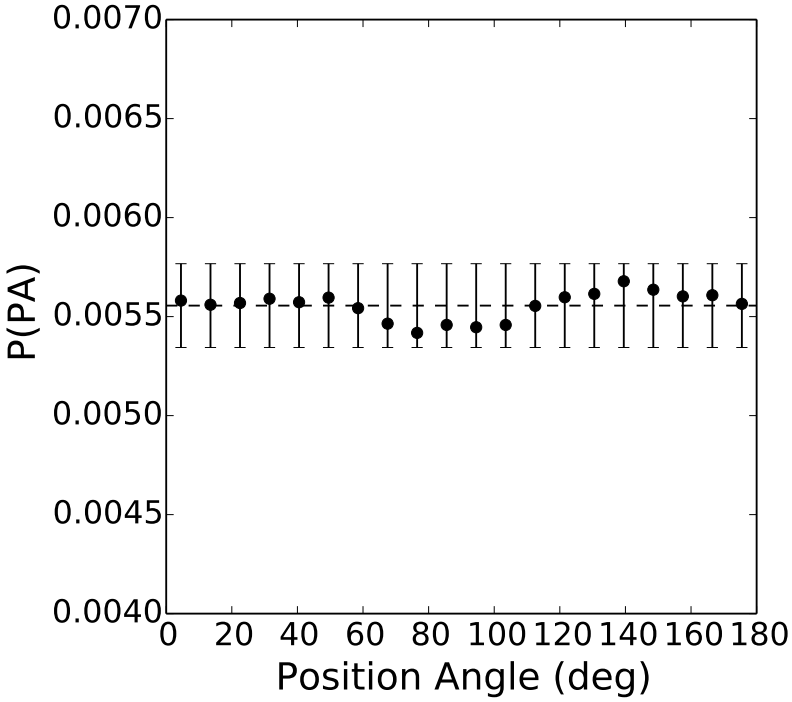

Figure 4.1. Distribution of the position angles (PA) of the DR7 galaxies. They are measured counterclockwise from north to east to be between 0 and 180 degree. The horizontal line is the expected random distribution with the error bars presenting Poisson errors.

to perform two dimensional PSF-convolved bulge+disc decompositions in the g and r bands on a sample of $1,123,718$ galaxies from the Legacy area of the SDSS DR7. It fits each galaxy image with a combination of a pure exponential disc and a de Vaucouleurs bulge. The resulting structural parameters include total flux, bulge to total ratio B/T, sizes, position angle, and many other parameters. The work by Simard et al. is by far the largest catalog of bulge+disc structural parameters estimated of the SDSS DR7 galaxies. The photometric errors reconstructed with their resulting fitting parameters are below 0.1 mag down to bulge and disc magnitudes of $g \simeq 19$ and $r \simeq 18.5$. About 90% of the DR7 galaxies are cross-identified with the catalog by Simard et al. Figure 4.1 shows that the distribution of the position angles for the DR7 samples thus obtained is unbiased.

As using the position angles as orientation of galaxies is only valid for elliptical galaxies, we limit ourselves only to elliptical galaxies as our sample. We classify galaxies of $(B/T)_r \geq 0.6$ as ellipticals and use only these throughout this paper unless stated otherwise. We checked that changing the selection criteria to $(B/T)_r \geq 0.5$ or 0.7 does not change our results qualitatively.

4.2.3 SDSS group samples

Given that galaxy groups are defined as galaxies that reside in common dark matter halo, galaxy groups can be used to directly probe the connections between galaxies and their host halos. To this end, this study uses the SDSS DR7 group catalog constructed by Yang et al. (2005, 2007) where their halo-based group finder has been applied to DR7 galaxies to group them into common dark matter halos. The performance of their halo-based group finder has been tested extensively using mock galaxies constructed from CLF models (Yang et al. 2003, van den Bosch et al. 2003, Yang et al. 2004) and a semi-analytic model (Kang et al. 2005), and was found better than friends-of-friends (FoF) method in assigning galaxies to dark matter halos (Yang et al. 2007). The group finder also performs consistently even for very poor systems such as isolated galaxies in small mass halos, which enables its suitability to probe the galaxy-halo association over a wide range of different halos. This DR7 group catalog thus obtained (publicly available at `http://gax.shao.ac.cn/data/Group.html`) is used for this study. WMAP7 cosmology was adopted for calculating distances and assigning halo masses to groups in the catalog.

The catalog provides halo mass estimates using either ranking of groups in the total luminosity ($L_{19.5}$) or total stellar mass ($M_{stellar}$) of all member galaxies with ${}^{0.1}M_r - 5\log h \leq -19.5$ in each group. Specifically, a monotonic one-to-one correspondence is assumed between $L_{19.5}$ or $M_{stellar}$ and the halo mass function of Tinker et

al. (2008) to assign halo mass to each group. This study uses halo masses estimated from $L_{19.5}$ throughout.

The identification of centrals of each group in the catalog is also provided twofold : brightest or most massive ones. Throughout this study, we define centrals to be the brightest ones, and center of each group to be position of the central galaxy.

4.2.4 Tidal fields

We use halo tidal field for this study, which is different from the conventionally defined mass tidal field in that it excludes self-gravity of halos, thus is better at focusing on the environmental effect. It is computed as follows. Total tidal force on surface of each halo exerted by all the other halos of mass above a certain threshold is first calculated. It, then, is normalized by the halo's self-gravity so that it measures only environmental tidal effect by surrounding halos on the halo in question. The surface on which the tidal force is calculated is defined to be the shell at the halo's radius, and the threshold mass of $10^{12}h^{-1}\mathrm{M}_{\odot}$ was used.

Following the methods above, one can construct the tidal tensor and its eigenvectors and eigenvalues t_1, t_2 and t_3. Then by the definition, the direction of t_1 is approximately along stretch of materials in large scale environments at the position of each halo, while t_3 is along compression of materials and t_2 is to be perpendicular to both t_1 and t_3. Unlike the mass tidal field, $t_1 + t_2 + t_3 = 0$ always. And, if, for example, only t_1 is positive while the others are negative, then as the material stretches along t_1, but compresses along the other two, the large scale structure is identified as 'filament'.

As already indicated, the main difference between the halo and the mass tidal field is that the former excludes the self-gravity while the latter includes. However, except that, they have much similarity and show tight correlations in their alignments, as

have been shown by Wang et al. (2011). We refer the readers to Wang et al. (2011) for more details on how to compute both tidal fields and how they differ or correlate.

4.2.5 Morphological classification of large scale structures

One of our results includes an attempt to study alignments of groups in different morphological large scale structures. To this end, we classify the morphology based on the eigenvalues of the conventional mass tidal field tensor, defined as,

$$T_{ij} = \partial_i \partial_j \phi \tag{4.1}$$

where ϕ is the peculiar potential. When we define the eigenvalues of the tensor T_1, T_2 and T_3 with $T_1 > T_2 > T_3$, as the tensor measures basically the 2nd-derivatives of the gravitational potential, T_1 is along the most compressed direction of large scale mass distribution, and each T being positive (negative) means the compression (stretch) of materials along that direction. Therefore, if only T_3 is negative for a certain halo, it is in a filamentary structure, while if only T_1 is positive then it is identified as a sheet. If all three eigenvalues are positive, it is a cluster. Only groups of $M_{\rm h} \geq 10^{12} h^{-1} {\rm M}_{\odot}$ were used to construct the tidal field, and we use the smoothing length scale of $10^{13} h^{-1} {\rm M}_{\odot}$ as this scale is known to best trace the visual classification of large scale structures (Hahn et al. 2007). The more details of the calculation of the mass tidal field used for this study can be found in Wang et al. (2012).

4.3 Analyses and results

4.3.1 Alignments of group galaxies

4.3.1.1 Alignments of group galaxies with tidal fields

Here we analyze alignments between orientations of galaxies and the tidal fields. We project the three dimensional tidal vectors onto two dimensional celestial sphere. Then the angles between the projected tidal vectors and the position angles of member

Table 4.1. Group and galaxy samples.

$\log M_{\rm h}(h^{-1}{\rm M}_\odot)$	$N_{\rm grp}$	$N_{\rm cen}$	$N_{\rm 2nd+3rd}$	$N_{\rm 4th+5th}$	$N_{\rm 6th+7th}$	$N_{\rm sat}$
≥ 12.5	32662	11056	9115	3348	2009	23643
$[12.5, 13]$	23657	7738	3686	363	83	4185
$[13, 13.5]$	6747	2485	3531	1286	418	5582
$[13.5, 14]$	1894	712	1578	1417	1128	7029
$[14, 14.5]$	334	120	315	379	376	6830

Column 2 contains the number of groups in each halo mass bin as indicated in column1. Column 3 provides the total number of the brightest galaxies in groups. Columns 4-6 list combined counts of the 2nd and the 3rd, the 4th and the 5th, and the 6th and the 7th brightest galaxies in groups, respectively. Column 7 shows the total number of satellites in each halo mass bin. While all groups with the tidal field value available are counted in column 2, columns 3-7 provide only the number of galaxies of $B/T \geq 0.6$.

galaxies are measured to be within [0, 90] degrees. Thus its zero value means, in principle, perfect alignment of a galaxy with its large scale tidal field, while noise is introduced inevitably since we are using projected ones for both the direction galaxies and the tidal vectors.

Figure 4.2 shows the alignments with the tidal field vectors, t_1, t_2, and t_3, separately, as a function of the brightness ranking orders of galaxies in their groups. It is seen that galaxies are aligned along t_1, thus stretches of large-scale structures, while perpendicular to t_3, the direction of mass compression, whether they are centrals or satellites. However, there are also some clear differences in strength of the alignment. In particular, central galaxies show much stronger alignment than satellites, and it tends to monotonically decrease as it goes down to the lower-ranked galaxies in brightness. It is worth noting that the second and the third brightest combined show much weaker alignment relative to the centrals. It has been often pointed out that there are much uncertainties in defining centrals observationally and it is neither meaningful nor accurate to strictly distinguish centrals from the others. The significantly stronger alignments shown for the centrals, however, may indicate that

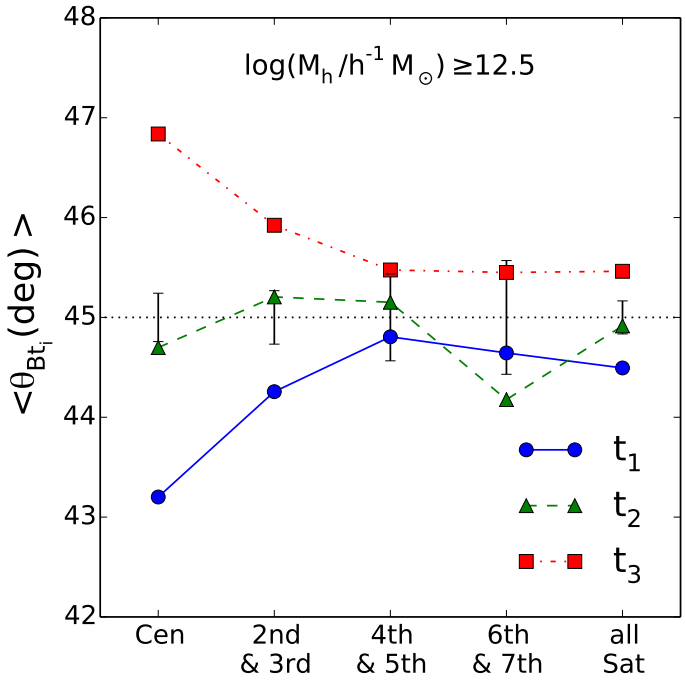

Figure 4.2. Alignment between the orientation of galaxies and the tidal fields at their locations. The bins are for the brightest (i.e. central), the 2nd and the 3rd brightest integrated, the 4th and the 5th brightest integrated, the 6th and the 7th brightest integrated, and total satellites. The error bars are 1σ scatters with 100 random realizations of the PAs for each galaxy, which average around 45 degree.

the centrals are indeed in distinct positions in their groups. We checked that there is nearly no dependence of the alignment on distance from centrals (Figure 4.5), and thus the dependence on the brightness rank is not a reflection of distance dependence. In principle, the monotonic decrease of the alignments with the brightness rank may be a result of mass dependence, as low-mass groups have only a few satellites and thus do not contribute much to, for example, the 6th or the 7th brightest. Figure 4.3, however, shows that the dependence on the brightness rank is still found when halos of narrow mass bins are separately considered.

Another thing that is consistently found in both Figure 4.2 and 4.3 is that satellites show much weaker alignments than centrals. There are two possible reasons. The

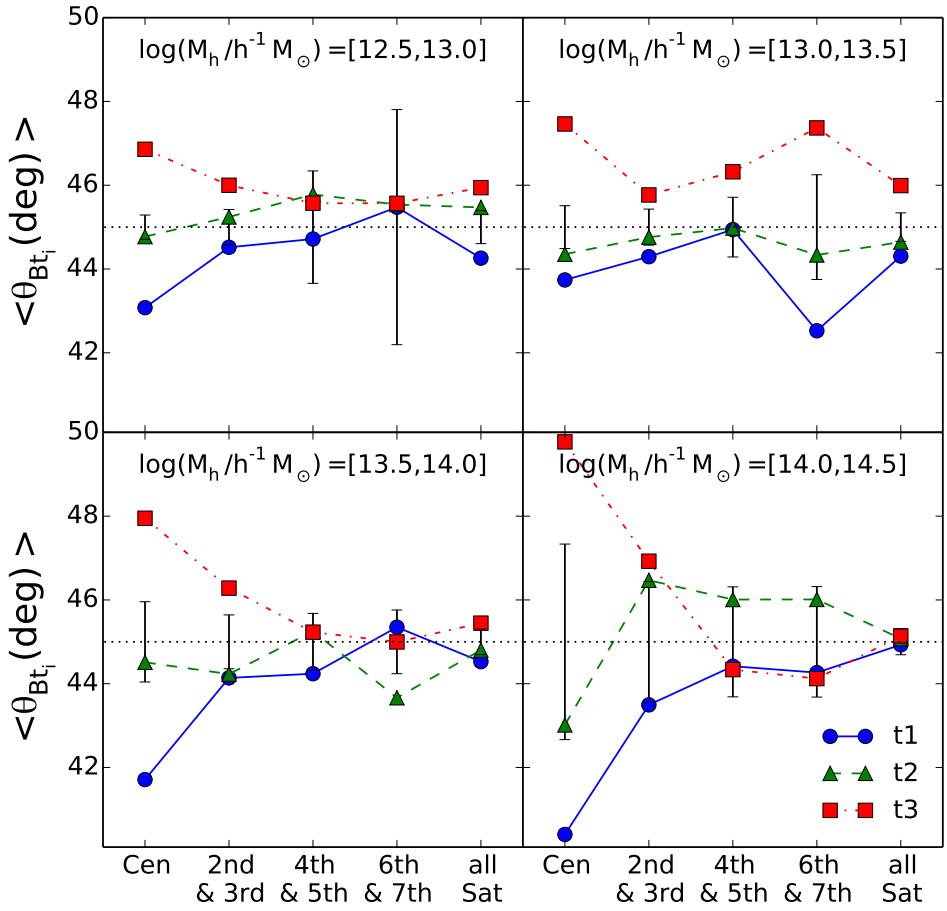

Figure 4.3. Same as Figure 3, but binned into halo mass ranges, as indicated in each panel.

alignments found in Figure 4.2 is believed to reflect that the accretion of mass occurs via mass flow into halos along the surrounding large scale structures and the tidal fields trace the large scale structures. However, during the accretion, satellites are more likely affected by non-linear effects than centrals, which can perturb and deviate such a flow of accretion from an uniform one, suppressing the alignment. Another reason is that while satellites orbit after their accretions, their position angles may evolve with time, changing depending on their initial dynamical conditions.

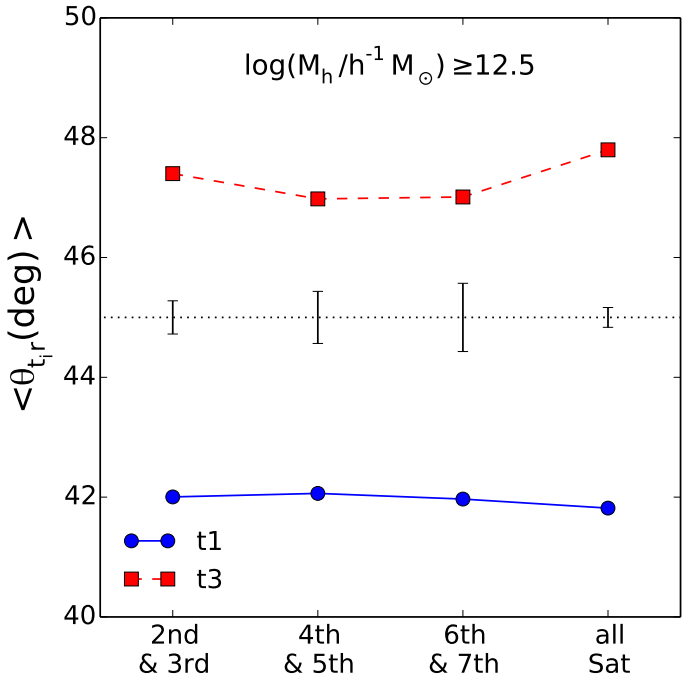

Figure 4.4. Alignment between the projected position vector of satellites and the tidal field. The position vector is defined as a vector from the brightest galaxy in each group to each galaxy. The brightness ranking orders of galaxies are presented similarly to Figure 4.2. The error bars are 1σ scatters with 100 random realizations of the projected position vectors for each galaxy.

4.3.1.2 Alignments between tidal fields and group shape

We also investigate alignment between the tidal field and the spatial distribution of satellites. In each group, we draw a vector from the brightest member galaxy, which we define as the center of the group, to each member galaxy as its position vector. Then we project both the position vectors and the tidal field onto two dimensional celestial sphere, and measure the subtending angle. Figure 4.4 shows the average of the angles thus obtained as a function of the brightness ranking order for t_1 and t_3 separately, in which we can clearly see that galaxies are preferentially distributed along the large scale structure that they are embedded in as predicted, but the alignment shows no dependence on the brightness rank.

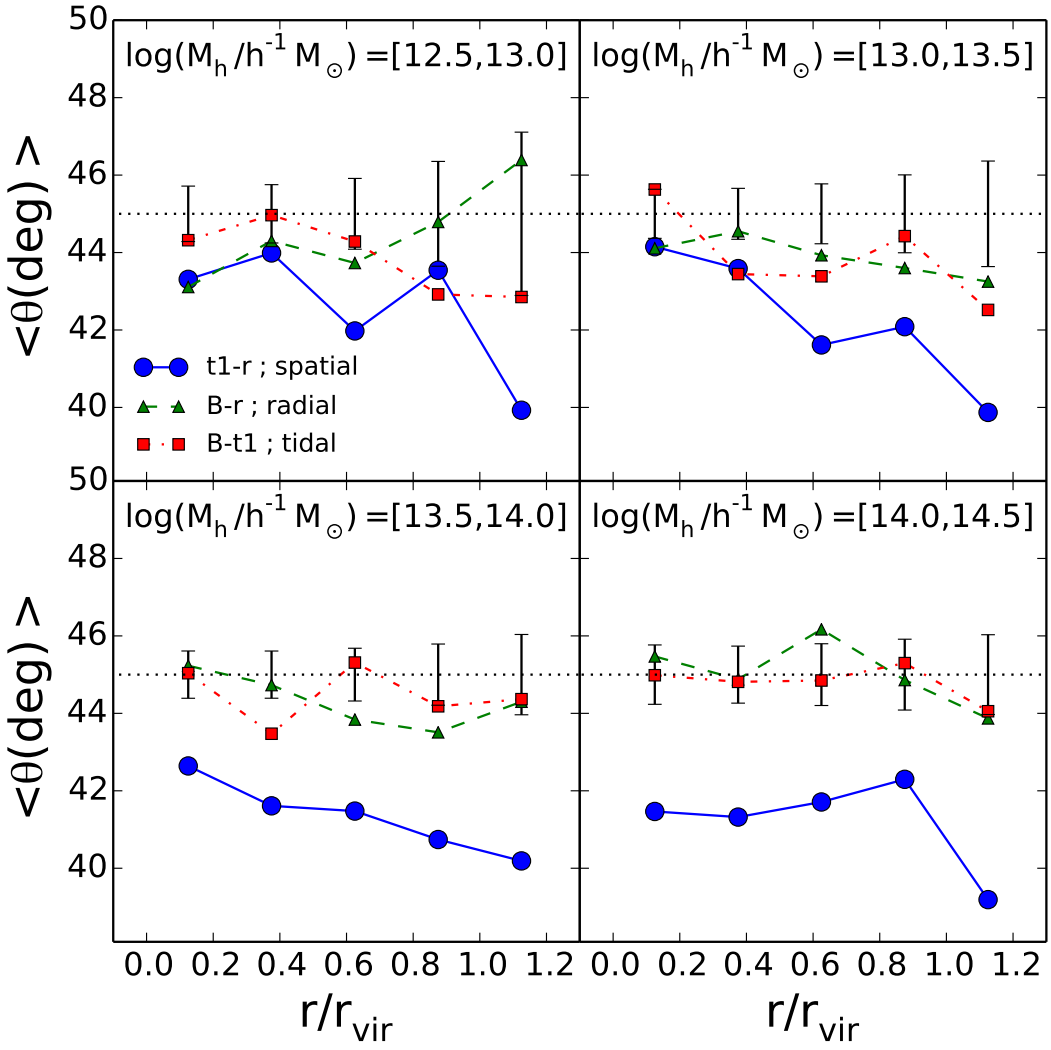

Figure 4.5. Same alignment as in Figure 4.4, but as a function of projected distance of satellites from centrals divided by virial radii (blue circles). Tidal and radial alignments are also shown by red rectangles and green triangles. Of the three different alignments, only the spatial alignment shows strong dependence on the distance. The error bars are 1σ scatters with 100 random realizations of the relevant angles for each galaxy.

In Figure 4.4, however, we examine the spatial alignment as a function of the projected distance of each member galaxy relative to the virial radius of their hosts, and clearly see a strong dependence that galaxies at outer region of groups are more

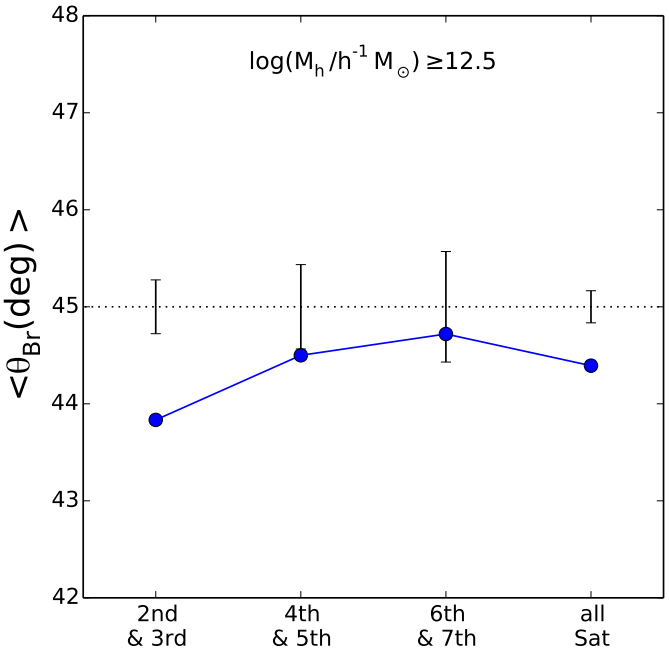

Figure 4.6. Alignment between the orientation and the projected position vector of satellites. The position vector is defined as a vector from the brightest galaxy in each group to each galaxy. The brightness ranking orders of galaxies are presented similarly to Figure 4.2. The error bars are 1σ scatters with 100 random realizations of the PAs for each galaxy.

aligned with the t_1 vectors. This may reflect that the tidal fields are calculated on the surface of halos, and that, in inner regions, the self-gravity of the host halo plays a role, reducing the relative influence of the large scale tidal fields.

4.3.1.3 Radial alignments

In this section, we examine radial alignment, i.e. alignment between the position angles and the position vectors. The position vectors are defined in same way as in the previous subsection. Then, we measure angles between the position angles and the projected position vectors onto the celestial sphere. Figure 4.6 shows that there is such radial alignment. This can be understood by combining the results from the two previous subsections. In other words, the galaxies are spatially distributed along

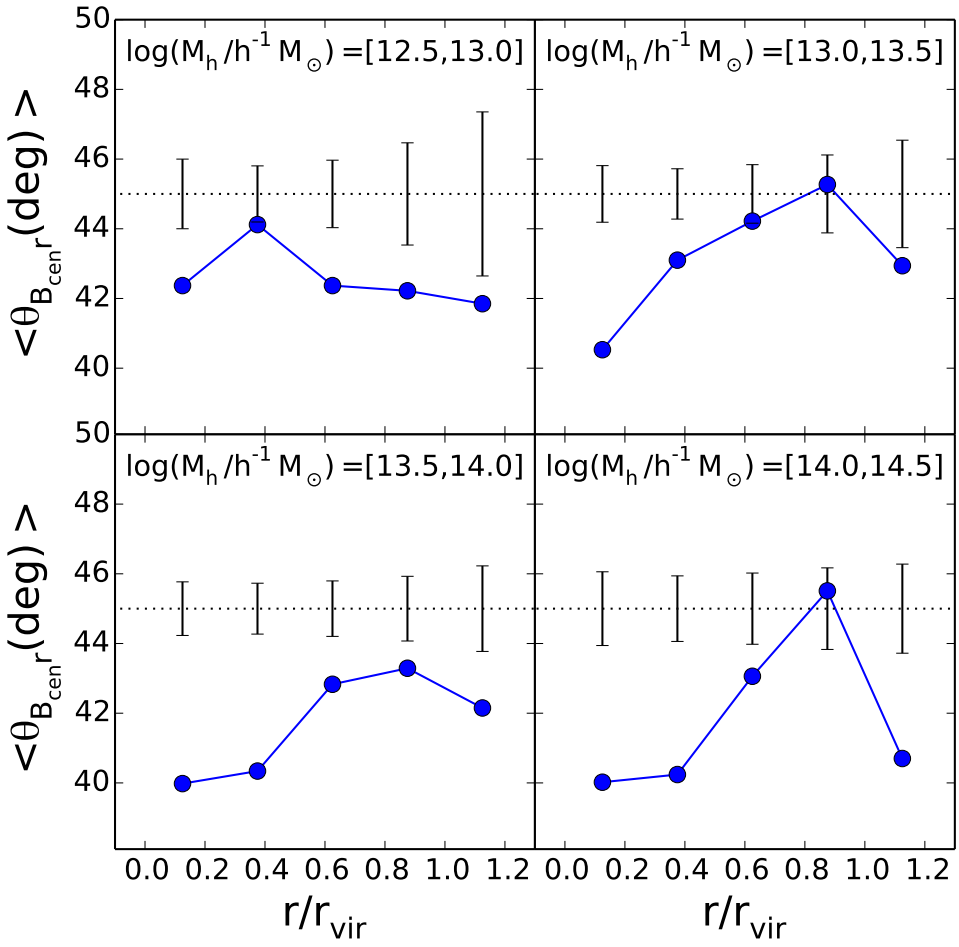

Figure 4.7. Alignment between the projected position vector of satellites and the orientation of centrals as a function of projected distance of satellites from centrals divided by virial radii. The error bars are 1σ scatters with 100 random realizations of the projected position vectors for each galaxy.

the tidal field, and orientation of galaxies is also aligned with the tidal field, thus there is also an alignment between orientation and spatial distribution of galaxies. For spirals, however, we found no radial alignment.

This result is consistent with the findings by Rong et al. (2015), where the radial alignment of the position angles has been claimed both using Monte Carlo simulation and from observation on the cluster Abell 2744. As there is no clear dependence

of the radial alignment on distance from centrals as shown in Figure 4.5, the radial alignment should be interpreted as alignment by the large scale tidal field rather than alignment by the self-gravity. The weaker alignment of galaxies of the lower brightness rank can be due to stronger non-linear effects. According to a recent study by Shi et al. (2015), halos from N-body simulations tend to be very weakly aligned or even anti-aligned with the large scale tidal field when they are under strong tidal field, opposed to the positive alignment they show under a weak up to moderate tidal field. Given that the galaxies of the lower brightness rank mainly reside in massive halos which form more recently when the tidal field is stronger, the weaker alignment of those may reflect the finding by Shi et al. (2015).

4.3.1.4 Alignments between group shape and central galaxies

As briefed in the introduction, there have been a series of conflicting studies on alignment between distribution of satellites and orientation of centrals until recently. With about 30000 central-satellite elliptical pairs, we confirm that distribution of satellites is aligned with orientation of their centrals as shown in Figure 4.7, which is consistent with other recent studies. The alignment is found stronger for galaxies in inner region in intermediate to massive halos. Galaxies in low mass halos, however, show no significant distance dependence. Interestingly, the galaxies right outside the virial radius in massive halos appear to be as much aligned as galaxies in inner to intermediate distances, despite that, due to large error bars, more statistics are needed to show any clear trend.

4.3.2 Large scale alignments

4.3.2.1 Tidal field alignments on large scales

Here we turn our focus onto the alignment between the tidal fields at the location of groups. We measure the angle between the projected tidal vectors for each group-group pair. Figure 4.8 shows the average angle as a function of three dimensional

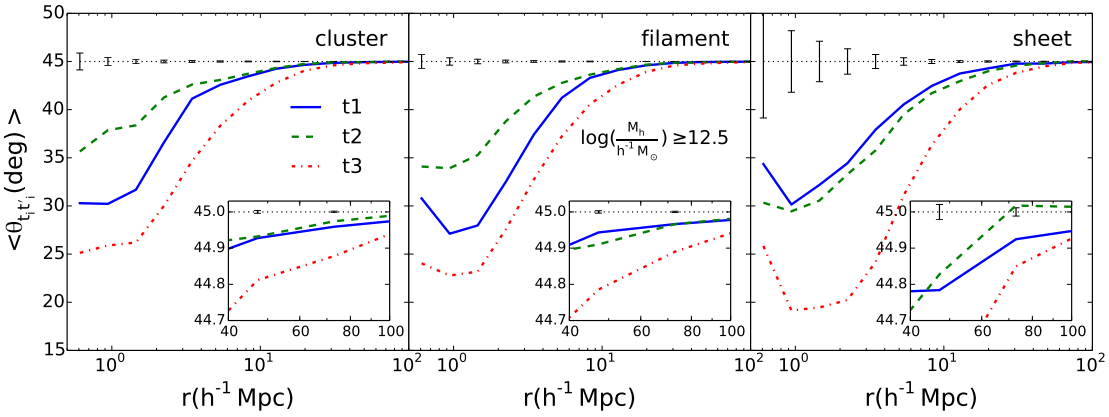

Figure 4.8. Alignment between the tidal fields of group pairs of $M_h \geq 10^{12.5} h^{-1}\mathrm{M}_\odot$ as a function of separation, separately for groups in clusters (left-), filaments (middle-), and sheets (right-panel). The morphologies that galaxies reside in are classified according to the signs of the mass tidal field, as described in the section 4.2. The error bars are 1σ scatters with 100 random realizations of the tidal field vectors for each group. The lower-right panels provide zoom-ins of the largest scale regime.

distance between pairs, separately for groups in cluster, filament, and sheet structure. In sheets, t_3 is the eigenvector defined with the least ambiguity, while the directions of t_1 and t_2 can be perturbed in sheets. We thus see t_3 is most aligned in sheets in the figure. On the other hand, t_1 is best defined in filamentary structure due to its morphology and that is the reason why the alignment of t_1 is found most strongly in filaments than in the other structures. It is expected that t_2 is difficult to be unambiguously defined due to how it is determined, which is the reason t_2 shows generally weaker alignments than t_1 and t_3. Also, for all three t's, groups in sheets show largest scales that the alignment extends out to, as expected, as filaments and clusters are identified where sheets are intersected, and thus sheets are of the largest scale among those. It is clearly seen that the large scale alignments extend as far as up to $80 - 100h^{-1}$Mpc,

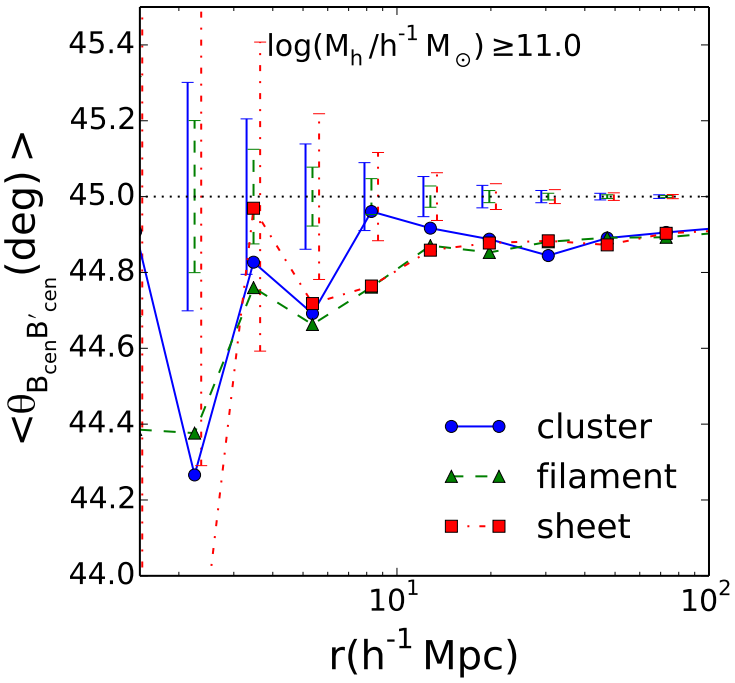

Figure 4.9. Alignment between the position angles of central galaxy pairs as a function of separation, separately for groups in clusters, filaments, and sheets. The morphologies that galaxies reside in are classified according to signs of the mass tidal field, as described in the section 4.2. The error bars are 1σ scatters with 100 random realizations of the tidal field vectors for each central galaxy. Note that three error bars assigned at each radial bin for those in cluster, filament, and sheet, are dislocated slightly for visual clarity.

4.3.2.2 Galaxy-galaxy alignments on large scales

We also probe the alignment of galaxy pairs as a function of distance. Figure 4.9 plots the average difference of the position angles of the pairs, and shows a certain degree of such alignment despite the strength is much weaker than that of t_1. In addition, possibly being affected by smaller number of samples than the tidal field case, there is no clear sign of its dependence on the morphology of large scale structure. It is worth mentioning that alignment extended up to large scales in Figure 4.9 is partly due to not perfectly uniform distribution of the position angles of the elliptical centrals in our sample, despite of its uniform distribution shown in Figure

4.1 for centrals and satellites combined. Therefore, even though the clear qualitative dependence on distance is shown, the exact numbers should be taken carefully.

4.4 Summary

Using the SDSS DR7 samples and the large scale tidal field constructed by Wang et al. (2011, 2012), we investigated a variety of alignments of ellipticals with the tidal field. The tidal field, more specifically, was calculated by applying the group finder developed by Yang et al. (2005, 2007) to the DR7 galaxies and using the distribution of the halos thus obtained to compute the tidal fields exerted on each halo. This way the halo tidal field is different from the mass tidal field in that it excludes the self-gravity of halo in question and that it is normalized by the self-gravity. Thus, the halo tidal field is better by design to examine the environmental effects on the alignments of galaxies. As orientation of galaxies, we used the position angle estimates from the GIM2D two-dimensional bulge+disc decomposition by Simard et al. (2011), which is based on the SDSS g and r band pass images. The distribution of the position angles is found to be very uniform (Figure 4.1), and thus does not bias our results.

Figure 4.2 and Figure 4.3 show that galaxies tend to preferentially orient along the stretch of their surrounding mass distribution. The alignment is found stronger for galaxies of the higher brightness ranking orders in their groups. Interestingly, the centrals, defined to be the brightest galaxy in each group, show much stronger alignment than satellites. This may imply that the centrals are indeed in distinct positions in their groups. The dependence on the brightness rank may be owing to non-linear effects, as non-linear effects are expected to be larger for satellites and disturb such alignment.

The spatial distribution of galaxies is also found to be aligned with the tidal field. Unlike the orientation, no clear dependence on the brightness rank is found. Instead, there is a radial dependence in that galaxies at outer region of groups are more aligned

with the tidal field. This, however, could be due to that the tidal field used here is calculated on the surface of each halo, and also neglect the self-gravity of halos, which will increase going toward inner regions.

We found radial alignment in that the shapes of satellites tend to preferentially orient toward the brightest galaxy in their groups. This can be understood as a combined result of both the orientation and the spatial distribution alignments of satellites with the tidal field : Galaxies are distributed along the tidal field, their orientations are also along the tidal field, and thus their orientations are aligned with their spatial distributions. This is also consistent with the recent finding by Rong et al. (2015) where the radial alignment has been claimed both analytically and observationally. The weaker alignment of galaxies of the lower brightness rank may reflect that the tidal field and thus non-linear effects, which disturb the alignment, are stronger for those as they formed recently on average.

We also revisited alignment between orientation of centrals and spatial distribution of satellites with $\simeq 30000$ central-satellite pairs, and detected the positive alignment, which is consistent with other recent studies. The alignment strength is increasing with decreasing distance from centrals for massive halos, while such trend is not found for low mass halos. Satellites outside the virial radius is more aligned than those in outer region, while more samples are required to further study this region.

We also probed alignment between the tidal fields of group pairs as a function of distance, separately for groups in different morphologies of their surrounding large scale structures. We confirmed that such large scale alignments are extended out to $80-100h^{-1}$Mpc. In terms of morphologies, the alignments of groups in sheets extend out most as expected, as they are of the largest scales when considering the process of collapse and the structure formation. As best characterizing the orientation of sheets, t_3 is found to be most aligned in sheets. Similarly, t_1, which well characterizes filaments, is most strongly aligned among groups in filaments.

Similarly, alignment of orientations is also found among central galaxy pairs, though much more weakly in strength, as shown in Figure 4.9. While the study of such alignment suffers from poor statistics of small number samples, no clear dependence on the morphology of the large scale structures is found.

All these results of various alignments imply that galaxy formation occurs via the coherent but highly anisotropic mass inflow and accretion along the large scale structure. The results are also in good agreements with previous studies where alignments with the large scale structure have been probed in various forms. This agreement indicates that the large scale tidal field indeed characterizes well the collapse process of the cosmic density field toward the formation of large scale structures.

PART II: PROBING THE GAS COMPONENT OF THE COSMIC WEB WITH THE SZ EFFECTS

CHAPTER 5

GAS CONTENTS OF GALAXY GROUPS FROM THERMAL SUNYAEV-ZEL'DOVICH EFFECTS[1]

A matched filter technique is applied to the *Planck* all-sky Compton y-parameter map to measure the thermal Sunyaev-Zel'dovich (tSZ) effect produced by galaxy groups of different halo masses selected from large redshift surveys in the low-z Universe. Reliable halo mass estimates are available for all the groups, which allows us to bin groups of similar halo masses to investigate how the tSZ effect depends on halo mass over a large mass range. Filters are simultaneously matched for all groups to minimize projection effects. We find that the integrated y-parameter and the hot gas content it implies are consistent with the predictions of the universal pressure profile model only for massive groups above $10^{14}\,M_{\odot}$, but much lower than the model prediction for low-mass groups. The halo mass dependence found is in good agreement with the predictions of a set of simulations that include strong AGN feedback, but simulations including only supernova feedback significantly over predict the hot gas contents in galaxy groups. Our results suggest that hot gas in galaxy groups is either effectively ejected or in phases much below the virial temperatures of the host halos.

5.1 Introduction

In the current paradigm of galaxy formation, galaxies are thought to form and evolve within dark matter halos (see Mo et al. 2010, for a review). During the

[1] THE CONTENTS OF THIS CHAPTER ARE PUBLISHED IN LIM ET AL. 2018, MNRAS, 854, 181.

formation of dark matter halos, the cosmic gas component first moves along with the dark matter, then gets shock-heated as the halos collapse, and eventually forms hot gaseous halos. In an adiabatic case, the resulting distribution of the hot gas component follows roughly that of dark matter, and the amount of the hot gas per dark matter is roughly a constant, about the universal baryon fraction of the universe. In reality, however, a myriad of other physical processes, such as radiative cooling, star formation, feedback from supernovae (SNe) and active galactic nuclei (AGN), etc, can change the hot gas content of the halos. Indeed, the hot gas fractions in low-mass halos are found to be lower than the universal baryon fraction both in observations (e.g. David et al. 2006, Gastaldello et al. 2007, Pratt et al. 2009, Sun et al. 2009) and in numerical simulations (e.g. McCarthy et al. 2010, Battaglia et al. 2013, Le Brun et al. 2014). Even in massive systems, such as rich clusters of galaxies where the total hot gas is found to be closer to the universal value, the distribution of the hot gas is found to be different from that of dark matter (e.g. Arnaud et al. 2010, Battaglia et al. 2012). However, current observational results are still uncertain, particularly for low-mass systems, and many competing theoretical models have been proposed to describe the formation and structure of gaseous halos. Clearly, an accurate determination of the hot gas content in dark matter halos is crucial for understanding galaxy formation and evolution in a way complimentary to the information provided by stars and cold gas.

The thermal Sunyaev-Zel'dovich effect (tSZ hereafter; Sunyaev & Zeldovich 1972) provides a promising avenue to probe the hot gas in halos. As the CMB photons pass through galaxy systems, such as clusters and groups of galaxies (collectively referred to as groups of galaxies), they are scattered by the hot electrons by the inverse Compton process, producing a net energy gain in the photon gas and changing the CMB temperatures in the directions to the groups. Thus, studying the cross-correlation of the imprints of tSZ effect on the CMB with galaxy groups allows one to

probe the hot gas components in halos associated with galaxy systems. Compared to X-ray observations, the tSZ effect is less sensitive to the hot gas density, thus making it possible to explore the hot gas in the outskirts of halos and also in low-mass halos where the gas density is expected to be low.

However, extracting the tSZ signal reliably from CMB observations is not easy. First, the signal to be detected is usually comparable to or lower than the primary CMB, and there are also contaminations, such as Galactic emissions, dust, and point sources. As a result, individual detection and analysis of the tSZ effect are currently only possible for rich clusters of galaxies (e.g. Planck Collaboration V 2013). For low-mass groups, stacking of many systems is required to increase the signal-to-noise. Second, the beam sizes of current instruments are usually insufficient to resolve low-mass systems, so that assumptions about the spatial distribution of the hot gas are required. Finally, the signal from low-mass systems can be contaminated by the projections of larger halos along the same line-of-sights, and such contamination is not straightforward to eliminate.

Recently, Planck Collaboration XI (2013) used the all-sky *Planck* multi-frequency temperature maps and a sample of locally brightest galaxies as tracers of dark matter halos to investigate the tSZ effects produced by galaxy systems with halo masses down to $\sim 4 \times 10^{12}\,\mathrm{M}_{\odot}$. Remarkably, their results show that the universal pressure profile (UPP) model, in which the hot gas fraction relative to halo mass is independent of halo mass, matches their data well. This finding is in conflict with the results obtained from X-ray observations and hydrodynamic simulations where a much lower fraction is found for hot gas in low-mass systems. Using a hydrodynamic simulation, Le Brun et al. (2015) that the universal pressure profile of Arnaud et al. (2010) adopted by Planck Collaboration XI (2013) in their matched filter method may lead to over-estimations of the integrated tSZ signal within R_{500} for low-mass systems, although they found that the flux within $5R_{500}$ is robust to the change in the adopted profile.

However, Greco et al. (2015) showed that adopting another popular pressure profile of Battaglia et al. (2012), instead of that of Arnaud et al. (2010), leads to differences that are well within the observational uncertainties, and so the high gas fraction found by Planck Collaboration XI (2013) cannot be explained by the adopted profile. Ma et al. (2015) cross-correlated the *Planck* tSZ map with gravitational lensing map from CFHTLenS survey and found that the prediction of UPP model is 20% higher than the data. Vikram et al. (2017) cross-correlated the *Planck* tSZ map with the group catalog of Yang et al. (2007) and found that the two-halo terms dominate the tSZ signal for systems of $M_{200} \leq 10^{13}\, h^{-1}\mathrm{M}_{\odot}$, indicating that projection effect is an important issue.

In this paper, we extract the tSZ signal from the *Planck* all-sky Compton parameter map for galaxy systems of different halo masses, using the group catalog of Lim et al. (2017a). The catalog is constructed for four large redshift surveys with the use of the halo-based group finder of Yang et al. (2005, 2007). This provides the largest sample of galaxy groups in the low-z universe to study the tSZ effects over a large range of galaxy systems. In particular, reliable halo mass estimates are provided for all groups, so that we can bin groups of similar masses to investigate how the tSZ effect depends on halo mass. We employ the matched filter technique (Haehnelt & Tegmark 1996, Herranz et al. 2002, Melin et al. 2005, 2006) to extract the tSZ signal from the *Planck* map. In particular, we simultaneously match the filters to all galaxy systems in the catalog, so that projection effects produced by halos along the line-of-sights are properly taken care of.

The outline of this paper is as followings. We describe the observational data used in our analysis in Section 5.2, and our method to extract the tSZ signal in Section 5.3. We present our main results as well as comparisons with results from earlier studies and from numerical simulations in Section 5.4. Finally, we summarize and conclude in Section 5.5.

We adopt the cosmological parameters from the *Planck* observation (Planck Collaboration XIII 2016) throughout this paper unless specified otherwise.

5.2 Observational data

5.2.1 The *Planck* y-map

The *Planck* (Tauber et al. 2010, Planck Collaboration I 2011), a space mission to measure the CMB anisotropy, is an all-sky observation in nine frequency bands ranging from 30 to 857 GHz, with angular resolutions from 31 to 5 arcmin. For our analysis of the thermal Sunyaev-Zel'dovich (tSZ) effects, we use the *Planck* NILC (Needlet Independent Linear Combination; Remazeilles et al. 2011) all-sky tSZ Compton parameter map (Planck Collaboration XXII 2016), also referred to as the NILC y-map, which is part of the publicly released *Planck* 2015 data[2]. The map is constructed from the full mission data set, using a combination of different frequency maps to remove the primary CMB fluctuations and to minimize contamination from foreground sources. For more details of the y-map construction, the readers are referred to the original paper cited above. To limit the Galactic foreground contamination, which is mainly due to thermal dust emissions, we mask the brightest 40% of the sky by applying the corresponding mask provided in the *Planck* 2015 data release. For contamination from extra-galactic sources, such as radio and infrared galaxies, we apply the mask provided in the same data release for point sources.

5.2.2 Galaxy groups

In order to determine the tSZ signals from halos associated with different galaxy systems, we need a well-defined group catalog that provides reliable information for both the positions and halo masses of the galaxy systems in the universe. Furthermore, since the tSZ signals are typically weak for individual groups, and since it is

[2] `https://pla.esac.esa.int`

necessary to stack many systems to increase the signal to noise ratio, a well-defined group catalog is also needed to interpret the stacking results. In this paper we use the group catalogs given in Lim et al. (2017a), which uses four redshift catalogs of galaxies (2MRS, 6dF, SDSS, and 2dF) to achieve an almost all-sky (91%) coverage and the best depth reachable by these galaxy catalogs in each region of the sky. Groups are identified with the adaptive halo-based group finder of Yang et al. (2005, 2007) with some modifications (see Lim et al. (2017a) for the detail). Tests with realistic mock galaxy catalogs show that the halo masses assigned by the group finder match well the true masses, with typical scatter of $0.2 - 0.3$ dex. The catalogs provide two different halo mass estimates based either on the luminosities and stellar masses of member galaxies, and we use the masses based on the stellar masses. We combine 2MRS, 6dF, and SDSS to construct our sample of groups with $\log M_{500}/\mathrm{M}_{\odot} \geq 12$ for the tSZ analysis. For sky regions covered by more than one catalog, the preference is given in the order of SDSS, 6dF, and 2MRS. The sample contains a total of $471,696$ galaxy systems (groups), of which $3,851$ have $\log M_{500}/\mathrm{M}_{\odot} \geq 14$, $112,494$ have $13 \leq \log M_{500}/\mathrm{M}_{\odot} \leq 14$, and $240,747$ have $12 \leq \log M_{500}/\mathrm{M}_{\odot} \leq 13$. Following conventions in previous SZ effect analyses, we define a halo by a radius R_{500}, within which the mean density is 500 times the critical density at the redshift in question. The mass, M_{500}, used above is the halo mass within R_{500}. The halo masses and radii provided in the group catalogs are M_{200} and R_{200}, respectively. To convert these quantities to the corresponding M_{500} and R_{500}, we assume NFW profiles (Navarro et al. 1997) and concentration parameters as given by Neto et al. (2007).

5.3 Method and analysis

5.3.1 The matched filter technique

Detecting the SZ signals physically related with a galaxy system is not trivial, as other effects, such as the primary CMB anisotropies, Galactic foreground, and other

sources, can all contaminate the signals we want to obtain (see Section 5.4). Using a simple aperture photometry to extract the signals may thus lead to large uncertainties in the extracted signals (see e.g. Melin et al. 2006). To limit source confusions and background contamination, we employ the matched filter (MF) technique, first proposed for SZ analyses by Haehnelt & Tegmark (1996), which is designed to maximize the signal-to-noise for a SZ source by imposing prior knowledge of the signals given the noise power spectra. For the case considered here, this means to optimally extract the tSZ signals from groups of galaxies, under the constraint of the power spectrum of the noise of the *Planck* maps. In practise, we closely follow Melin et al. (2005, 2006), who presented an extended and general formalism to extract signals from SZ surveys using the multi-filtering technique of Herranz et al. (2002). Such a MF technique has been applied in many recent analyses of the SZ effects in different surveys (e.g. Planck Collaboration V 2013, Planck Collaboration XI 2013, Planck Collaboration LIII 2017, Li et al. 2014, Le Brun et al. 2015).

In the MF approach, the Fourier transform of the filter that maximizes the signal-to-noise is given by:

$$\hat{F}(\boldsymbol{k}) = \Big[\int \frac{|\hat{\tau}(\boldsymbol{k'})\hat{B}(\boldsymbol{k'})|^2}{P(k')} \frac{\mathrm{d}^2 k'}{(2\pi)^2} \Big]^{-1} \frac{\hat{\tau}(\boldsymbol{k})\hat{B}(\boldsymbol{k})}{P(k)} \tag{5.1}$$

where $\hat{\tau}(\boldsymbol{k})$ is the Fourier transform of the assumed spatial profile of groups, $\hat{B}(\boldsymbol{k})$ is the Fourier transform of a Gaussian beam function that mimics the convolution in *Planck* observation with the FWHM of 5 arcmin, and $P(k)$ is the noise power spectra. As the NILC y-map used here is already cleaned of the primary CMB anisotropies, $P(k) = P_{\rm noise}$, where $P_{\rm noise}$ is the power spectrum of the noise map for the *Planck* y-map, as provided in the data release. The choice of the spatial filter function is not straightforward, and it can affect the integrated signals extracted. Indeed, using hydrodynamic simulations, Le Brun et al. (2015) found that the extracted tSZ signals can change significantly depending on the filter shape adopted. In our analysis, we

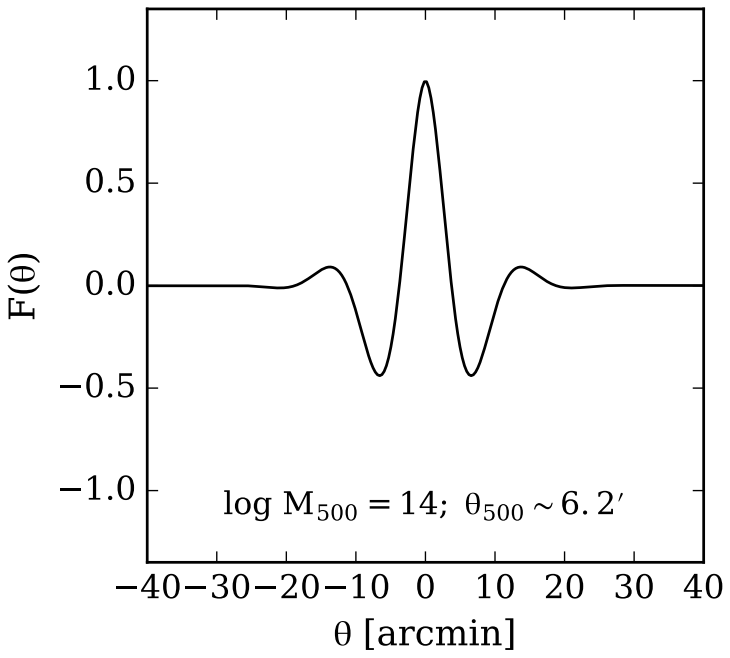

Figure 5.1. An example of the matched filter constructed for the tSZ analysis. Here a universal profile of Arnaud et al. (2010) is adopted as a spatial filter for a group of $\log M_{500}/\mathrm{M}_{\odot} = 14$ and an angular radius $\theta_{500} \sim 6.2$ arcmin.

adopt the universal pressure profile (UPP) given in Arnaud et al. (2010, hereafter A10), the form of which can be written as

$$P(r) = A[E(z)]^{8/3}\mathcal{P}(r/R_{500}, M_{500})\,, \tag{5.2}$$

where $E(z) \equiv H(z)/H_0$, $\mathcal{P}$ specifies the shape of the profile, and A is an overall amplitude (see A10 for details). This profile was derived from a combination of X-ray observations of XMM-Newton REXCESS cluster sample (Böhringer et al. 2007) at $r \leq R_{500}$ and hydrodynamic simulations at larger radii. As a test, we have also used the spatial filter adopted in Le Brun et al. (2015) but did not find any significant changes in our results.

Figure 5.1 shows an example of the constructed filters, which assumes a group with $\log(M_{500}/\mathrm{M}_\odot) = 14$ and an angular radius $\theta_{500} \sim 6.2'$, and the universal pressure profile given above.

5.3.2 Extracting the tSZ signal

Theoretically, the tSZ signal is characterized by a Compton y-parameter,

$$y \equiv \frac{\sigma_\mathrm{T}}{m_\mathrm{e}c^2} \int P_\mathrm{e}\, \mathrm{d}l, \tag{5.3}$$

where σ_T is the Thompson cross-section, m_e the rest-mass of electron, c the speed of light, $P_\mathrm{e} = n_\mathrm{e} k_\mathrm{B} T_\mathrm{e}$ the electron pressure, and the integration is over the line-of-sight to the observer.

The filters described above are then put at the group centers and 'matched' to the y-map to yield an estimate of the tSZ flux within R_{500}, Y_{500} defined by

$$d_\mathrm{A}(z)^2 Y_{500} = \frac{\sigma_\mathrm{T}}{m_\mathrm{e}c^2} \int_{R_{500}} P_\mathrm{e}\, \mathrm{d}V, \tag{5.4}$$

where $d_\mathrm{A}(z)$ is the angular diameter distance to a group at redshift z. Since Y_{500} depends mainly on halo mass at a given z and evolves with z as $E^{2/3}(z)$, at a fixed halo mass, it is useful to define a new quantity,

$$\tilde{Y}_{500} \equiv Y_{500} E^{-2/3}(z) \Big(\frac{d_\mathrm{A}(z)}{500\mathrm{Mpc}} \Big)^2, \tag{5.5}$$

which is expected to be a function of only halo mass scaled to $z = 0$, if the intrinsic tSZ flux is indeed only a function of mass.

To extract the SZ signals associated with galaxy groups from the observed y-map, a matched filter is put at each of all the groups in our group sample according to its halo mass and redshift. We then tune simultaneously the amplitudes of the filters

for individual M_{500} bins as listed in Table 5.1, assuming the amplitudes for all the groups in a given M_{500} bin to be the same. The overall best match between the matched filters and the observed y-map is sought on the basis of the sum of the χ^2 over all the pixels covered by the filters. The simultaneous matching of individual groups allows us to take into account the line-of-sight contributions from other halos, which is important, particularly for low-mass systems (see next section). For most of our analyses, we truncate the filter at $3\,\theta_{500}$ and only use fluxes within it to estimate the integrated flux, where $\theta_{500} = R_{500}/d_{\rm A}$. Note that $R_{500} \sim 0.5\,R_{200}$. As shown later in Section 5.4, truncation of the filters at $10\,\theta_{500}$ leads to little change in our results. Finally, the mean flux within R_{500} for groups in a given mass bin is estimated from the assumed spatial profile together with the amplitude obtained from the best match. The fluxes within R_{500} for individual groups in a M_{500} bin are also estimated by fixing the amplitudes of the matched filters of other mass bins to their best fitting values, while tuning the amplitudes of the filters for individual groups in the M_{500} bin in question to achieve the best match.

5.4 Results

5.4.1 The M_{500}-$\tilde{Y}_{500}$ relation and the hot gas content

Figure 5.2(a) shows our main result for the tSZ flux-halo mass relation. The values obtained from our fiducial sample are also given in Table 5.1 for reference. In all our analyses, we use only groups at $z > 0.03$, to avoid the domination by a small number of nearby groups each covering a large number of pixels. The error-bars shown are each the 1σ dispersion among individual groups in the corresponding mass bin. For comparison, we also divide the total sample into three independent sub-samples that have about the same amount of sky coverage, and apply the MF method to them separately. The results are shown by the green crosses in the figure. As one can see,

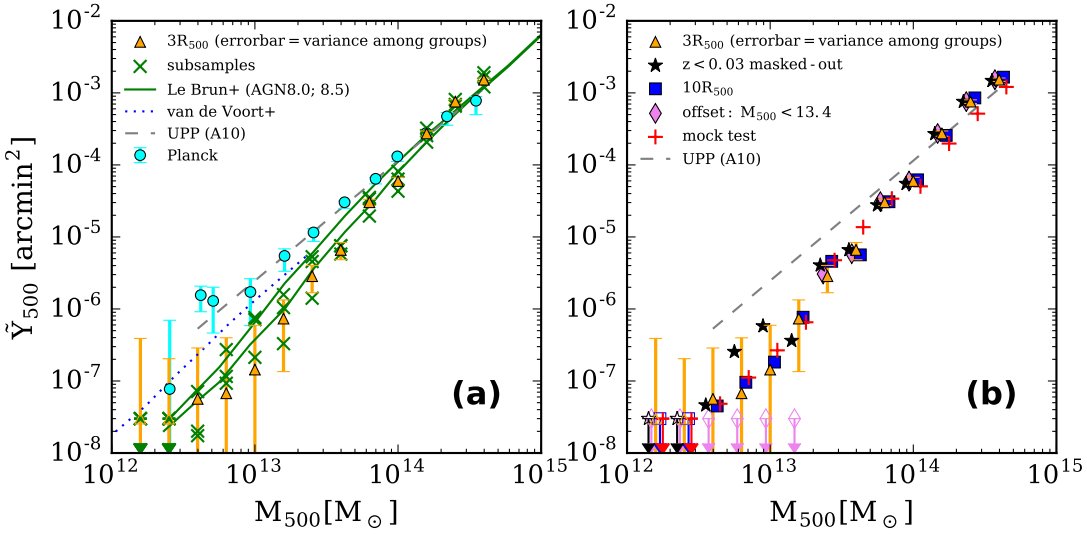

Figure 5.2. Panel (a): The M_{500}-$\tilde{Y}_{500}$ relations obtained by applying the matched filters, with shapes following that of Arnaud et al. (2010) (A10) truncated at $3R_{500}$ (triangles). The green crosses show the results obtained from three independent sub-samples of total sample. The results are compared with those from Planck Collaboration XI (2013) (cyan dot), the universal pressure profile of A10 (UPP; dashed), Le Brun et al. (2015) (the two solid lines, with the upper one for their AGN8.0 and the lower one for AGN8.5), and van de Voort et al. (2016) (dotted line). The error bars indicate the variance among the signals from individual systems in each mass bin. The unfilled symbols with downward arrows are used for cases where the tSZ fluxes are negative. Panel (b): The M_{500}-$\tilde{Y}_{500}$ relations obtained by masking out the pixels covered by groups at $z < 0.03$ (stars) and by truncating the matched filters at $10R_{500}$ (squares). The case where the filters for groups with $\log(M_{500}/\mathrm{M}_{\odot}) < 13.4$ are offset by $3\theta_{500}$ is shown by diamonds. The red crosses plot the difference between the results from the co-added map and the added component (see text for details of this test). The data points for some tests are shifted by up to 0.05 dex horizontally for clarity. For comparison, the triangles with error bars and the dashed line in panel (a) are repeated in this panel.

the scatter among these sub-samples is comparable to the variance among individual groups, indicating that the uncertainties are dominated by systematic effects.

Comparing the result obtained from the fiducial case, where the matched filters are truncated at $3R_{500}$ with that obtained by using a larger truncation radius, $10R_{500}$,

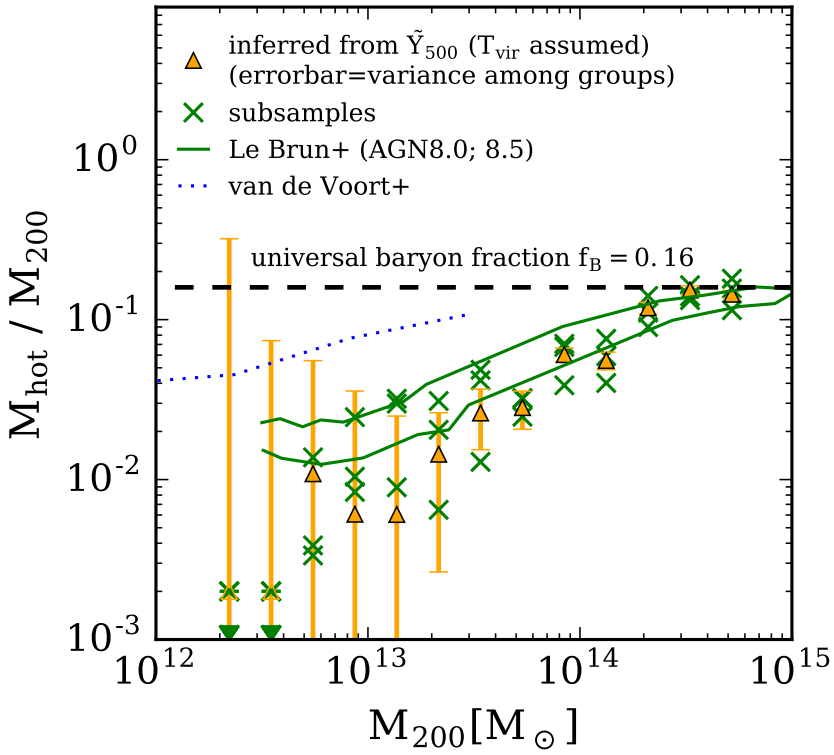

Figure 5.3. Hot gas mass fraction with respect to halo mass within r_{200} as a function of halo mass inferred from $\tilde{Y}_{500}$ by assuming the virial temperature, compared with that from Le Brun et al. (2015) (solid; the upper for AGN8.0 and the lower for AGN8.5) and van de Voort et al. (2016) (dotted), which are based on hydrodynamic simulations. The gas content is lower than the universal baryon fraction of $f_B = 0.16$ (dashed) in low-mass systems by a factor of up to ~ 10. Note that the gas mass ratio estimated is inversely proportional to a temperature assumed. The error bars indicate the variance of the signals among individual systems in each mass bin. The crosses show the results obtained from three independent sub-samples. The unfilled symbols are used for those with the tSZ flux below zero.

as shown in Figure 5.2(b), shows that $3R_{300}$ is sufficient to cover the signals from individual groups.

Because the hot gas halos of groups are extended and the observational beam size is relatively large, projection effects may contaminate the signals, particularly for small groups, even though our simultaneous matching of filters is supposed to eliminate

Table 5.1. The M_{500}-$\tilde{Y}_{500}$ relation[a].

$\log M_{500}/\mathrm{M}_\odot$	$\tilde{Y}_{500}$ $[10^{-6}$ arcm$^2]$	Variance[b] $[10^{-6}$ arcm$^2]$	No. of systems
12.3	-0.0600	0.360	40,689
12.5	-0.0480	0.174	41,848
12.7	0.0564	0.231	40,521
12.9	0.0675	0.330	37,344
13.1	0.144	0.450	32,063
13.3	0.735	0.600	25,744
13.5	2.85	1.17	19,020
13.7	6.60	1.77	12,500
13.9	30.3	3.00	6,203
14.1	59.7	7.50	2,163
14.3	273	22.8	484
14.5	756	45.3	195
14.7	1520	83.7	71

Notes.
a. These data are presented in Fig. 5.2 by triangles.
b. These are the 1σ dispersion among individual systems in each mass bin.

such effects. We thus carry out a number of further tests to examine any possible residuals due to projection effects. In our first test, we remove all pixels covered by the groups at $z < 0.03$, and the results are shown by stars in Fig. 5.2(b). As one can see, the signals for low-mass systems are changed relative to the fiducial case, suggesting that our results for low mass groups may still be affected by projection effects. In the second test, the filters for groups below a halo mass limit are shifted by a given amount with respect to the group centers. Thus, if the signals extracted for these groups were not associated with them but produced by diffuse electrons associated with larger structures, such a shift would not change the signals obtained

for these groups. As an example, the diamonds in Fig. 5.2(b) show the result in which filters for groups with $\log M_{500}/\mathrm{M}_{\odot} < 13.4$ are shifted randomly by $3R_{500}$, while the filters for more massive groups are still located at the group centers. As one can see, while no significant change is seen for groups with $\log M_{500}/\mathrm{M}_{\odot} > 13.4$, the signals for the lower mass groups are reduced. This suggests that the signals detected in the matched filters are associated with these low-mass groups. Finally, we make another test by adding to each group an artificial y-parameter profile which is given by the observed mean profile corresponding to its mass and redshift. The artificial y-parameter profile has the same shape as given by the adopted pressure profile and its amplitude is determined from a broken power-law fit to the observed mean amplitude as a function of halo mass. The test is intended to examine the robustness of the results against a change in the signal-to-noise. As the artificial signals added are the mean values at given masses, the test changes the signal-to-noise of the composite y-map in each pixel compared to that of the original map, which changes the weights given to individual pixels when matching filters. The matched filter technique is then applied to the sum of this artificial map with the original map. The original signals are well recovered by the differences between the results obtained from the co-added map and the added component, as shown by the crosses in Fig. 5.2(b), demonstrating that our method can extract the signals we put in reliably.

Based on the test results presented above, we conclude that the results for groups with masses above $10^{13.5}\mathrm{M}_{\odot}$ are stable. For groups of lower masses, however, significant variations are still present from sample to sample. We have also applied the same methods to different maps, such as the *Planck* MILCA (Modified Internal Linear Combination Algorithm; Hurier et al. 2013) all-sky tSZ Compton parameter map (Planck Collaboration XXII 2016), which is known to have different degree of dust contamination, and *Planck* multi-frequency temperature maps at 100 and 143 GHz, in which dust effects are expected to be smaller than in other bands. We found that the

results do not change significantly, especially for groups above $10^{13.5}\mathrm{M}_{\odot}$, indicating that our results are robust against residual dust emissions from galaxies.

Assuming virial temperatures, we estimate the hot gas contents of galaxy groups within R_{200} from the integrated fluxes of $\tilde{Y}_{500}$. Here the NFW profile and the hot gas profile of A10 are used to convert quantities to the corresponding ones within R_{200}, and the virial temperature is defined as

$$T_{\rm vir} = \frac{\mu m_{\rm p} G M_{200}}{2 k_{\rm B} R_{200}} \tag{5.6}$$

where μ is the mean molecular weight, $m_{\rm p}$ the proton mass, and $k_{\rm B}$ the Boltzmann constant. The results obtained from the fiducial sample and the three sub-samples are shown in Fig. 5.3. Here we see that the inferred hot gas contents of low-mass groups within R_{200} are lower than the universal baryon fraction, shown by the horizontal line, by a factor of ~ 10. Even for groups with $M_{200} \sim 10^{14}\mathrm{M}_{\odot}$, the hot gas fraction is only about a half; only in the most massive groups (clusters) is the fraction close to unity. Of course, the gas fraction could be much higher if the gas temperature is much lower than the virial temperature, and measurements of tSZ effect alone cannot break the density-temperature degeneracy. In any case, the implied low density and/or temperature of the halo gas in low-mass groups have important implications for theories of galaxy formation, as to be discussed in the following.

5.4.2 Comparisons with earlier results and theoretical models

Planck Collaboration XI (2013) (PCXI hereafter) used the same *Planck* data and a similar matched filter approach to extract the tSZ signals around locally brightest galaxies (LBGs) selected from the SDSS survey. An isolation criterion is adopted so that each LBG is the dominating one (in terms of luminosity) in its neighborhood, probably representing the central galaxy of a halo. Based on the mean relation between the stellar masses of central galaxies and the halo masses obtained from the

semi-analytic galaxy formation model of Guo et al. (2013), a halo mass is assigned to each of the LBGs. The $\tilde{Y}_{500}$-M_{500} relation obtained by PCXI is plotted in Fig. 5.2(a) as circles, and matches well the expectation of the UPP model of A10, shown by the dashed line. As one can see, our results are in good agreement with that of PCXI only for massive groups with $M_{500} > 10^{14}\mathrm{M}_{\odot}$, but the amplitudes we obtain for groups of lower masses are much lower. Indeed, our $\tilde{Y}_{500}$-M_{500} relation is very different from that given by the UPP model. We suspect that there are two factors that may cause the difference between our and PCXI results.

First, we simultaneously match all groups in our sample, which takes into account the projection effects by larger halos along the line-of-sights of low-mass groups, while PCXI matches individual filters separately. In a test where we first subtracted the local flat backgrounds averaged over annulus between $[2R_{200}, 3R_{200}]$ around each group, and then matched individual filters and stacked the signals for groups of similar masses, we found that we can roughly recover the results of PCXI for low-mass halos, despite of the differences in other details between our method and theirs. This indicates that the contamination by other groups is not flat, and that it is important to match the filters to all groups simultaneously in order to correct for such projection effects. In principle, there could be residual projection effects from halos not included in the catalogs either because they are located outside the redshift ranges of the catalogs or because they are too faint to be included in the group catalogs. The contribution from halos beyond the redshift range is expected to be uncorrelated with the groups at lower z, and so it increases the noise level but does not bias the average signals obtained for these groups. Groups that are located within the sample volume but missing because they fall below the observational limits have halo masses $M_{500} < 10^{13}\mathrm{M}_{\odot}$. Our tests by matching filters only to groups with higher masses showed that the signals obtained for groups above $10^{13}\mathrm{M}_{\odot}$ are not affected significantly by excluding lower mass groups in the filter matching. This demonstrates

that the projection effects produced by halos below the mass completeness limit do not have a significant impact on our results.

Second, PCXI uses the mean relation between the central galaxy mass and halo mass to estimate halo mass, while our halo masses are estimated from our halo-based group finder. Given that the central galaxy mass increases only slowly with halo mass at $M_{200} > 10^{13}\mathrm{M}_{\odot}$ (Yang et al. 2003), and the relation has significant amounts of scatter (e.g. Yang et al. 2008), binning based on central galaxy mass may mix halos of very different masses.

Greco et al. (2015) used a LBG sample similar to that used by PCXI, together with the *Planck* temperature aperture photometries, instead of the matched filter, to extract tSZ signals associated with the LBGs. They found that their results are consistent with the UPP model within the uncertainties of the data. It is unclear if the difference between their results and ours is produced by the different mass proxies used to bin the data or by the different methods used to extract the tSZ signals. In a forthcoming paper, we will address this issue by examining how different methods adopted in the literature affect the extracted SZ signals. Vikram et al. (2017) examined the cross-correlation between groups in the catalog of Yang et al. (2007) and the *Planck* y-map, and found that two-halo terms dominate the signals around halos of $M_{200} \leq 10^{13-13.5}\,h^{-1}\mathrm{M}_{\odot}$. This is in qualitative agreement with our finding that the stacked signals for low-mass groups are dominated by projection effects. Taking account of the projection effects based on Vikram et al. (2017), Hill et al. (2017) found some evidence for a broken power-law relation between M_{500} and $\tilde{Y}_{500}$, which is in qualitative agreement with our results.

We also compare our results with results from two hydrodynamic simulations. The first is that presented in Le Brun et al. (2015), who used the cosmo-OWLS suite of cosmological simulations (Le Brun et al. 2014), an extension of the OverWhelmingly Large Simulations (OWLS; Schaye et al. 2010), to model the tSZ effects. The simu-

lation has a box size of $400h^{-1}$ Mpc on a side, and assumes cosmological parameters either from the *WMAP*7 or the *Planck*. Their fiducial runs include both stellar and AGN feedbacks. In Figs. 5.2(a) and 5.3, the predictions of two of their models are plotted as the two solid curves. The upper curves correspond to their AGN feedback model AGN8.0, which assumes that accreting black holes heat their surrounding gas to a temperature $\Delta T_{\rm heat} = 10^8\,{\rm K}$, while the lower curves are for their AGN8.5, which assumes $\Delta T_{\rm heat} = 3 \times 10^8\,{\rm K}$. Clearly, our results are in a good agreement with their results, particularly from that of the AGN8.5 run.

van de Voort et al. (2016) used a suite of cosmological zoom-in simulations from the Feedback In Realistic Environments (FIRE; Hopkins et al. 2014, Faucher-Giguère et al. 2015, Feldmann et al. 2016) project, to study the tSZ effects around halos with $M_{500} = 10^{10}$–$10^{13}\,{\rm M}_\odot$. Sixteen and thirty six zoom-in simulations were run to $z = 0$ and $z \sim 2$, respectively. In Figs. 5.2(a) and 5.3, we use straight lines to roughly represent their low-z results. Here the universal profile of A10 is used to convert the predictions, which are integrated quantities within projected radius, to quantities within spheres needed in the comparison. It is seen that the predicted tSZ signals are much stronger than both our results and the simulations of Le Brun et al. (2015). We note, however, that the simulations used by van de Voort et al. (2016) do not include AGN feedback, which may be important for the halo mass range concerned here.

5.5 Summary and conclusion

In this paper, we use the measurements of the thermal Sunyaev-Zel'dovich (tSZ) effect from the *Planck* NILC all-sky Compton parameter map, together with the group catalogs of Lim et al. (2017a) to investigate the hot gas contents of galaxy groups. The catalogs contain a large number of uniformly selected groups with reliable halo mass estimates, which allows us to bin groups of similar halo masses to investigate

the dependence of the tSZ effect on halo mass over a large mass range. We adopt the matched filter approach (Haehnelt & Tegmark 1996, Herranz et al. 2002, Melin et al. 2005, 2006), which optimizes the signal-to-noise ratio by imposing prior knowledges of the expected signals, to extract the tSZ signals produced by galaxy groups from the map. We jointly match the filters to all groups to minimize projection effects.

We test the robustness of our method by retaining or eliminating pixels covered by local galaxy systems, by truncating the matched filters at different radii, by shifting the filters for low-mass groups, and by adding artificial signals to the observational map. We find that our method performs well in these tests. We also found that the background fluctuations around low-mass systems are significantly affected by projections of massive halos. Such a projection effect can lead to overestimation of the tSZ signals associated with low-mass groups if filters are not matched simultaneously to all groups.

We find that the integrated y-parameter and the hot gas content it implies are consistent with the predictions of the UPP model only for massive groups with masses above $10^{14}\,M_{\odot}$, but much lower, by a factor of ~ 10, than the model prediction for low-mass groups. Our results are in conflict with the findings from some previous studies (e.g. Planck Collaboration XI 2013, Greco et al. 2015), which reported that their data are in agreement with the predictions of UPP model. The disagreement likely comes from the different treatments of projection effects and the different halo mass models used in these studies. The halo mass dependence we find is in good agreement with the predictions of a set of hydro simulations presented in Le Brun et al. (2015) that include strong AGN feedback, but the simulations of van de Voort et al. (2016), which include only supernova feedback, over-predict the hot gas contents in galaxy groups by a factor of 5 to 10.

Since the integrated y-parameter is a measure of the thermal energy content of the hot halo gas, our results indicate that this energy content in low-mass groups

is much lower than that expected from the universal baryon fraction in a hot halo at the virial temperature. This has important implications for galaxy formation and evolution. Since the total baryon fraction of stars and cold gas in galaxy groups and clusters is found to be well below the universal baryon fraction (e.g. Fukugita & Peebles 20040, it has been speculated that the missing baryons may be in hot defused halos. However, if the low energy content found here is due to a low gas content in the hot phase, then hot gas halos cannot account for the missing baryons. Alternatively, baryons originally associated with galaxy groups may be heated and ejected by some processes. The agreement of our results with the predictions of the simulation results of Le Brun et al. (2015) suggests that strong AGN feedback may be able to provide such a process and to accommodate the observational results. Yet another possibility is that a large fraction of baryons may be in phases with temperatures much lower than the virial temperatures of the groups. In this case, the low thermal energy contents observed in low-mass halos are produced by the low gas temperature rather than by a reduced amount of gas. To distinguish the different possibilities, it is crucial to estimate the total mass in the warm-hot phase, so as to obtain a complete inventory of the baryons in low-mass halos. This can be done either through quasar absorption studies (e.g. Werk et al. 2014), or by investigating the kinetic SZ effect of galaxy groups (e.g. Hernández-Monteagudo et al. 2015, Planck Collaboration XXXVII 2016, Hill et al. 2016, Schaan et al. 2016, Lim et al. 2017b), which depends on the electron density but not the temperature of the halo gas.

CHAPTER 6

THE DETECTION OF MISSING BARYONS IN GALAXY HALOS WITH KINETIC SUNYAEV-ZEL'DOVICH EFFECT[1]

We present the detection of the kinetic Sunyaev-Zel'dovich effect (kSZE) signals from groups of galaxies as a function of halo mass down to $\log(M_{500}/\mathrm{M}_{\odot}) \sim 12.3$, using the *Planck* CMB maps and stacking about $40,000$ galaxy systems with known positions, halo masses, and peculiar velocities. A multi-frequency matched filter technique is employed to maximize the signal-to-noise, and the filter matching is done simultaneously for different groups to take care of projection effects of nearby halos. The total kSZE flux within halos estimated from the amplitudes of the matched filters implies that the gas fraction in halos is about the universal baryon fraction, even in low-mass halos, indicating that the 'missing baryons' are found. Various tests performed show that our results are robust against systematic effects, such as contamination by infrared/radio sources and background variations, beam-size effects and contributions from halo exteriors. Combined with the thermal Sunyaev-Zel'dovich effect, our results indicate that the 'missing baryons' associated with galaxy groups are contained in warm-hot media with temperatures between 10^5 and 10^6 K.

6.1 Introduction

According to the current scenario of galaxy formation, galaxies form and evolve in dark matter halos (see Mo et al. 2010 for a review). As a dark matter halo forms

[1] THE CONTENTS OF THIS CHAPTER ARE ORIGINALLY FROM LIM ET AL. 2017 (ARXIV:1712.08619).

in the cosmic density field, the cold gas associated with it falls into its potential well and gets shock-heated, eventually forming a hot gaseous halo, with a temperature roughly equal to the virial temperature of the halo. However, various processes, such as radiative cooling, star formation, and feedback from supernovae and active galactic nuclei (AGN), can affect the evolution of galaxies and the properties of the gaseous halos, so that the distribution of baryons may be very different from that of the dark matter. Indeed, observations have shown that both the hot gas fraction and the total baryon fraction in present-day galaxy systems are much lower than the universal baryon fraction, especially in low-mass systems (e.g. David et al. 2006, Gastaldello et al. 2007, Pratt et al. 2009, Sun et al. 2009). Even for massive clusters of galaxies, the distribution of the gas is found to be different from that of the dark matter, although the total amount of the hot gas is found to be close to the universal fraction (e.g. Arnaud et al. 2010, Battaglia et al. 2012). It has been suggested that a significant portion of the 'missing baryons' may be in the form of diffuse warm-hot intergalactic media (WHIM), with temperature in the range of 10^5 - 10^7 K, within and/or around dark matter halos (e.g. Cen & Ostriker 1999, Dave et al. 1999, 2001, Smith et al. 2011), but the detection of the WHIM from observation has so far been uncertain (e.g. Bregmann 2007).

The Sunyaev-Zel'dovich effect (SZE; Sunyaev & Zel'dovich 1972) offers a promising way to probe the WHIM. As the cosmic microwave background (CMB) photons pass through galaxy systems, such as clusters and groups of galaxies, they are scattered by the free electrons in these systems. The effect produced on the CMB by the thermal motion of electrons is referred to as the thermal SZE (tSZE), while that produced by the bulk motion of electrons is called the kinetic SZE (kSZE). Thus, cross-correlating galaxy systems (clusters and groups of galaxies, collectively referred to as galaxy groups hereafter) and their SZE in the CMB provides a promising way to probe the WHIM associated with dark matter halos. X-ray observations can also

be used to study the hot gas, but are effective only in probing dense and hot regions, such as the central parts of galaxy systems, and not sensitive to the diffuse WHIM in which the bulk of the missing baryons are expected to reside.

Great efforts have been made to measure the tSZE from observations and to use it to constrain the gas associated with galaxy systems. Planck Collaboration XI (2013) used the *Planck* multi-frequency CMB temperature maps and dark matter halos identified based on isolation criteria, to study the tSZE down to a halo mass $\sim 4 \times 10^{12} \mathrm{M}_{\odot}$. Remarkably, they found that the hot gas fraction in halos is independent of halo mass, as expected from the simple self-similar model. Similarly, Greco et al. (2015) used the locally brightest galaxies to represent dark matter halos to extract the tSZE and found that their results are consistent with the self-similar model. In a recent paper, Lim et al. (2018a) used a large sample of galaxy groups (Lim et al. 2017b) to extract the tSZE associated with galaxy systems from the *Planck* Compton parameter map (Planck Collaboration XXII 2016). By stacking about half a million galaxy systems, they were able to obtain the tSZE as a function of halo mass down to $\log(M_{500}/\mathrm{M}_{\odot}) \sim 12$, where M_{500} is the halo mass enclosed by a radius in which the mean mass density is 500 times the critical density. They found that the thermal contents of the gas in low-mass halos are much lower than that expected from the cosmic mean baryon fraction and the virial temperature of halos, in contrast to the results obtained by Planck Collaboration XI (2013) and Greco et al. (2015).

Detecting the kSZE signals from CMB observations is not a trivial task. First, the signals are weak; even for massive clusters of galaxies, the kSZE amplitude is an order of magnitude lower than the tSZE and two orders of magnitude smaller than the primary CMB fluctuation. As such, stacking a large number of similar galaxy systems is needed to detect the effect. Second, since the kSZE is directly proportional to the radial peculiar velocity of the galaxy system, and since the peculiar velocities of different systems have a symmetric distribution around zero, stacking individual

systems without using the peculiar velocity information leads to cancellation rather than enhancement of signals. Third, the large beam sizes of current CMB experiments require assumptions of the locations and gas profiles of the galaxy systems to be stacked, in order to extract the kSZE they produce. Finally, since the observed effects are projected on the sky, signals from low-mass systems may be contaminated by projections of the more massive systems along the same line-of-sight.

The detection of kSZE has so far been made only for cluster-size systems (e.g. Kashlinsky et al. 2010) and from statistical measurements based on, e.g., the pair-wise and cross correlation methods (Hand et al. 2012, Hernandez-Monteagudo et al. 2015, Planck Collaboration XXXVII 2016, Hill et al. 2016, Schaan et al. 2016, Soergel et al. 2016, De Bernardis et al. 2017). Using peculiar velocity fields reconstructed from galaxy distributions and the aperture photometry, Hernandez-Monteagudo et al. (2015), Planck Collaboration XXXVII (2016), Schaan et al. (2016) found that a significant fraction of baryons may be associated with the large-scale structure traced by galaxies. A similar conclusion was reached by Hill et al. (2016) by cross-correlating galaxies with CMB maps. The signals measured in these investigations are the averages over individual galaxies in the galaxy samples used, including effects both confined to galaxy halos and unbound over large scales. These results, therefore, constrain the total amount of free electrons associated with the large-scale structure traced by galaxies, but cannot be interpreted directly in terms of baryon fractions in halos of different masses. Thus, the missing baryon problem on halo scales, which has important implications for galaxy formation in dark matter halos, was unresolved.

In this paper, we investigate the kSZE from halos of different mass, using group catalog and the *Planck* temperature maps, by extending the same methods as in Lim et al. (2018) to kSZE. As described below, our analysis differs from earlier studies in that the halo-based group catalog with reliable halo mass allows us to probe the kSZE and baryon fractions in halos of different masses, and in that the simultaneous

matching of filters to the CMB maps takes into account the line-of-sight contamination by projection effects. In addition, the combination of the kSZE measurements here with the tSZE measurements obtained in Lim et al. (2018) not only allows us to obtain the total mass, but also the effective temperature of the WHIM associated with galaxy systems.

The structure of the paper is as follows. We describe the observational data for our analysis in Section 6.2, and our method to extract the kSZE in Section 6.3. We present our main results and inferences from combining kSZE with tSZE in Section 6.4. Finally, we summarize and conclude in Section 6.5.

6.2 Observational data

6.2.1 The *Planck* CMB map

The *Planck* observation (Tauber et al. 2010, Planck Collaboration I 2011) measures the all-sky CMB anisotropy in nine frequency bands from 30 to 857 GHz, with angular resolutions ranging from 31 to 5 arcmin. In our analysis for the kSZE, we use the 100, 143, and 217 GHz channel maps from the *Planck* 2015 data release[2]. To minimize Galactic contamination, the brightest 40% of the sky is masked using the masks provided in the data release. We also mask known radio and infrared point sources using the corresponding masks. From the reduced maps, subtractions are made of the tSZE,

$$\left(\frac{\Delta T}{T_{\rm CMB}}\right)_{\rm tSZ} = g(x)y \equiv g(x)\frac{\sigma_{\rm T}}{m_{\rm e}c^2}\int P_{\rm e}{\rm d}l, \tag{6.1}$$

where $T_{\rm CMB} = 2.7255\,{\rm K}$, y is the Compton parameter, $g(x) = x\coth(x/2) - 4$ is the conversion factor at a given $x \equiv h\nu/(k_{\rm B}T_{\rm CMB})$, $\sigma_{\rm T}$ is the Thompson cross-section, c is

[2] `https://pla.esac.esa.int`

the speed of light, $m_{\rm e}$ is the electron rest-mass, and $P_{\rm e} = n_{\rm e} k_{\rm B} T_{\rm e}$ is the electron pressure with $n_{\rm e}$ and $T_{\rm e}$ being the free electron density and temperature, respectively. The electron pressure is integrated over the path length, dl, along the line-of-sight (LOS). We adopt the Compton parameter y from the *Planck* NILC (Needlet Independent Linear Combination; Remazeilles et al. 2011) all-sky y-map (Planck Collaboration XXII 2016), which is constructed from the full mission data set of the *Planck*, using a combination of different frequency maps to minimize the primary CMB fluctuations and contamination from foreground sources. Integrating over the *Planck* bands gives the conversion factor $g(x)T_{\rm CMB} = -4.031$, -2.785, and $0.187\,{\rm K}$ for the 100, 143, and 217 GHz maps, respectively. As a test, we also applied the same analysis to the *Planck* MILCA (Modified Internal Linear Combination Algorithm; Hurier et al. 2013) y-map, which is known to have a different level of dust contamination, and found no significant changes in our results, indicating that our results are robust against residual dust emissions from galaxies. The readers are referred to the original papers for more details about the constructions of the y-maps. Finally, each of the resulting maps is subtracted by a constant to zero the average background. The final maps still contain components other than the kSZE, such as the primary CMB and instrumental noise. These are included in our multi-frequency matched filter (MMF), to be described in §6.3.

6.2.2 Galaxy groups

To extract the SZ signals associated with galaxy groups requires a well-defined group catalog. In our analysis, we use the group catalog of Yang et al. (2007), which is constructed from the Sloan Digital Sky Survey Data Release 7 (SDSS DR7; Abazajian et al. 2009) with the use of the halo-based group finder developed in Yang et al. (2005). All the groups in the original catalog have accurate estimates of halo masses, spatial positions, and peculiar velocities. Halo masses of the groups

are estimated from abundance-matching based on the ranking of their characteristic luminosities. Tests using realistic mock catalogs show that the halo masses given by the group finder match well the true halo masses, with typical scatter of $0.2-0.3$ dex. Following conventions in SZE studies, we define a halo by a radius, R_{500}, within which the mean density is 500 times the critical density at the redshift of the halo. The corresponding halo mass is M_{500}. The masses provided in the group catalog, M_{200}, are converted to M_{500} by assuming the NFW profile (Navarro et al. 1997) and a model for the halo concentration parameter (Neto et al. 2007). These properties of individual groups are used in a MMF technique (see below) to optimize the signal-to-noise and to extract kSZE from galaxy groups over a large range of masses. They are also used to interpret the kSZE in terms of the total amount of ionized gas associated with these groups. We adopt the radial peculiar velocities, v_r, reconstructed for the same sample of groups by Wang et al. (2012). Tests with mock catalogs show that the errors in the reconstructed peculiar velocities have a symmetric distribution around zero, with dispersion of about $90\,\mathrm{km\,s^{-1}}$. Our final sample contains all groups with $z \leq 0.12$, within which groups with $M_{200} > 10^{12.5}\,h^{-1}\mathrm{M}_{\odot}$ are complete (Yang et al. 2007).

6.3 Method and analysis

6.3.1 The multi-frequency matched filter technique

Because the SZE signals are typically lower than various other sources (contamination), such as the primary CMB anisotropy, Galactic foreground and cosmic infrared background, it is essential to optimize the signal-to-noise in order to extract the signals reliably. Using a simple aperture photometry may lead to large uncertainties (see e.g. Melin et al. 2006). Here we employ the multi-frequency matched filter (MMF) technique (Haehnelt & Tegmark 1996, Herranz et al. 2002, Melin et al. 2005, 2006), which is designed to minimize source confusions and background contamination, and

to maximize the signal-to-noise by imposing priors on the signals and the noise power spectra. A general formalism of applying the MMF to SZE measurements has been developed by Melin et al. (2006), and we adopt it for our analysis. In the MMF formalism, the Fourier transform of the filter that maximizes the signal-to-noise is given by

$$\begin{aligned}\hat{F}_I(\boldsymbol{k}) &= \left[\int \frac{\mathrm{d}^2k'}{(2\pi)^2}|\hat{\tau}(\boldsymbol{k'})|^2\hat{\boldsymbol{B}}^*(\boldsymbol{k'})\cdot\boldsymbol{P}^{-1}(k')\cdot\hat{\boldsymbol{B}}(\boldsymbol{k'})\right]^{-1} \\ &\quad\times\hat{\tau}(\boldsymbol{k})\, j_{\nu,I}\left\{\boldsymbol{P}^{-1}(k)\cdot\hat{\boldsymbol{B}}(\boldsymbol{k})\right\}_I \end{aligned} \tag{6.2}$$

where $\hat{F}_I(\boldsymbol{k})$ is the Fourier transform of the filter for each of the three frequencies, 'I', $\hat{\tau}(\boldsymbol{k})$ is the Fourier transform of the projection of the assumed spatial profile of the gas distribution, the elements of $\hat{\boldsymbol{B}}(\boldsymbol{k})$ are the Fourier transform of the Gaussian beam function that mimics the convolution of the *Planck* observation in in individual frequency bands, and $\boldsymbol{P}(k)$ is the noise power spectra. For the kSZE analysis, $P_{IJ}(k) = P_{\rm CMB}\,\hat{B}_I\hat{B}^*_J + P_{{\rm noise},I}\delta_{IJ}$ where $P_{\rm CMB}$ is the CMB spectrum and $P_{{\rm noise},I}$ the power spectra of the *Planck* noise map at frequency 'I', as given by the *Planck* data release. The quantity $j_{\nu,I}$ in the above equation is a frequency-dependent conversion factor between the differential temperature of the CMB and the filter at the frequency 'I'; $\boldsymbol{j}_\nu = g(\boldsymbol{x})$ for tSZE, and is frequency-independent for kSZE. The choice of the spatial profile is not straightforward. For our analysis, we adopt an empirical β-profile,

$$n_{\rm e}(r) = n_{\rm e,0}[1 + (r/r_{\rm c})^2]^{-3\beta/2}, \tag{6.3}$$

where $r_{\rm c} = r_{\rm vir}/c$ is the core radius of a group with concentration c, and $\beta = 0.86$ is the best-fit value obtained from South Pole Telescope (SPT) cluster profiles (Plagge et al. 2010). Note that $r_{\rm vir}$ and c are determined by the halo mass and redshift of the galaxy group in question. In principle, the integrated signals extracted can

depend on the assumed filter shape. However, because of the beam size of the *Planck* instrument, which is comparable to halo radius for a significant fraction of the groups in our sample, the results do not depend strongly on the specific choice of the spatial filter (see §6.4.2).

6.3.2 Extracting the kSZE signal

The CMB spectrum is distorted when CMB photons interact with free electrons that are moving collectively. In this kinetic Sunyaev-Zel'dovich effect (kSZE), temperature change is characterized by a dimensionless parameter,

$$k \equiv \left(\frac{\Delta T}{T_{\rm CMB}}\right)_{\rm kSZ} = -\frac{\sigma_{\rm T}}{c}\int n_{\rm e}(\boldsymbol{v}\cdot\hat{\boldsymbol{r}})\mathrm{d}l, \tag{6.4}$$

where $\boldsymbol{v}$ is the velocity of bulk motion, and $\hat{\boldsymbol{r}}$ is the unit vector along a LOS. Assuming that electrons are moving together with the galaxy system containing them, which is justified by the fact that the correlation length of the peculiar velocity field is much larger than a halo size (e.g. Hand et al. 2012, Hernandez-Monteagudo et al. 2015), we have

$$k = -\frac{v_r}{c}\tau, \qquad \tau(R) = \sigma_{\rm T}\int n_{\rm e}(\sqrt{R^2+l^2})\mathrm{d}l \tag{6.5}$$

where v_r is the CMB rest-frame peculiar velocity of the galaxy system along the LOS. Note that the Fourier transform of $\tau(R)$ gives $\hat{\tau}(\boldsymbol{k})$ in equation (6.2).

To extract the kSZE of a group from the *Planck* data, we use the matched filter, $\{F_i\}_I$, defined in equation (6.2) with redshift, virial radius and concentration appropriate for the group, 'i', in question, at the frequency 'I' to estimate the signal within the filter,

$$\{A_{D,i}\}_I = \int \{F_i(\boldsymbol{\theta})\}_I \{D_i(\boldsymbol{\theta})\}_I \mathrm{d}^2\theta\,, \tag{6.6}$$

where $\boldsymbol{\theta}$ is the projected position relative to the group center in the sky, and $\{D_i(\boldsymbol{\theta})\}_I$ is the data map around the group. The uncertainty of the estimate is given by

$$\sigma_i = \left[\int \frac{\mathrm{d}^2 k'}{(2\pi)^2} |\hat{\tau}_i(\boldsymbol{k'})|^2 \hat{\boldsymbol{B}}^*(\boldsymbol{k'}) \cdot \boldsymbol{P}^{-1}(k') \cdot \hat{\boldsymbol{B}}(\boldsymbol{k'}) \right]^{-1/2} , \tag{6.7}$$

as shown in Haehnelt & Tegmark (1996) and Melin et al. (2006).

The values of $\{A_{D,i}\}_I$ obtained this way can be affected by projection effects by other groups along the same line of sight. Because of this, we do not use these values to represent the signal produced by individual groups and to obtain the average quantities for groups of a given mass. Instead, we construct model maps that take into account the projection effects to compare with data. As described above, the kSZE signal expected from a given group is determined by its peculiar velocity and its gas density profile, which is modeled by equation (6.3). In our analysis, we use the reconstructed peculiar velocities described in §6.2.2. The virial radius, $r_{\rm vir}$, and concentration, c, of a group are given by its halo mass, as described above. To model the total ionized mass associated with a halo, we assume that the amplitude of the profile, $n_{\rm e,0}$, depends only on halo mass. We specify the halo mass dependence by the values of $n_{\rm e,0}$ at $\log(M_{500}/\mathrm{M}_\odot) = 12.3$, 12.7, 13.1, 13.5, 13.9, and 14.3, together with linear interpolations in $\log(n_{\rm e,0})$ - $\log(M_{500})$ space to predict the profile amplitude for any given M_{500}. Thus, the model is completely specified by a set of six model parameters that give the profile amplitudes at the six values of M_{500}. The numbers of groups in the six mass bins are listed in Table 6.1, together with the median values of M_{200}. For a given set of model parameters, denoted collectively by Θ, we first generate a theoretical map by convolving the center of each group with the 2-d profile appropriate for its halo mass, redshift and peculiar velocity. We then add to the map the primary CMB anisotropy using $P_{\rm CMB}$. Finally, we convolve the resulting map with the beam function, and add the instrument noise using $P_{\rm noise}$ to obtain a model map. Note that we have three model maps, containing the same CMB but corresponding to

the beam sizes and noise of the three channel data we use. Note also that the model maps include the projection effects generated by the superposition of the profiles of halos along a LOS, because they are constructed with the positions of all groups.

To extract signals from the model maps, we put a MMF at the center of each of all the groups in the model map, again according to the halo mass and redshift of the group, to extract the corresponding model signal,

$$\{A_{M,i}\}_I = \int \{F_i(\boldsymbol{\theta})\}_I \{M_i(\boldsymbol{\theta})\}_I \mathrm{d}^2\theta, \tag{6.8}$$

where $\{M_i(\boldsymbol{\theta})\}_I$ stands for the model map at frequency 'I' around group 'i'. With these measurements, we compute a χ^2-based likelihood function,

$$\mathcal{L}(\Theta|A_D) \propto \exp(-\chi^2/2), \tag{6.9}$$

where

$$\chi^2 = \sum_i \frac{|\boldsymbol{A}_{D,i} - \boldsymbol{A}_{M,i}(\Theta)|^2}{\sigma_i^2}\,. \tag{6.10}$$

To efficiently explore the parameter space, we make use of the MULTINEST method developed in Feroz et al. (2009), which implements the nested sampling algorithm developed in Skilling et al. (2006). The posterior distribution of the model parameters, i.e. the values of $n_{\mathrm{e},0}$ at the six halo masses, is used to make inferences from the data.

6.4 Results

6.4.1 The $\tilde{K}_{500}$-M_{500} relation

The amplitude of the density profile obtained from the matched filters, together with the assumed spatial profile shape, can be used to estimate the total number of electrons within R_{500} for a group of a given mass,

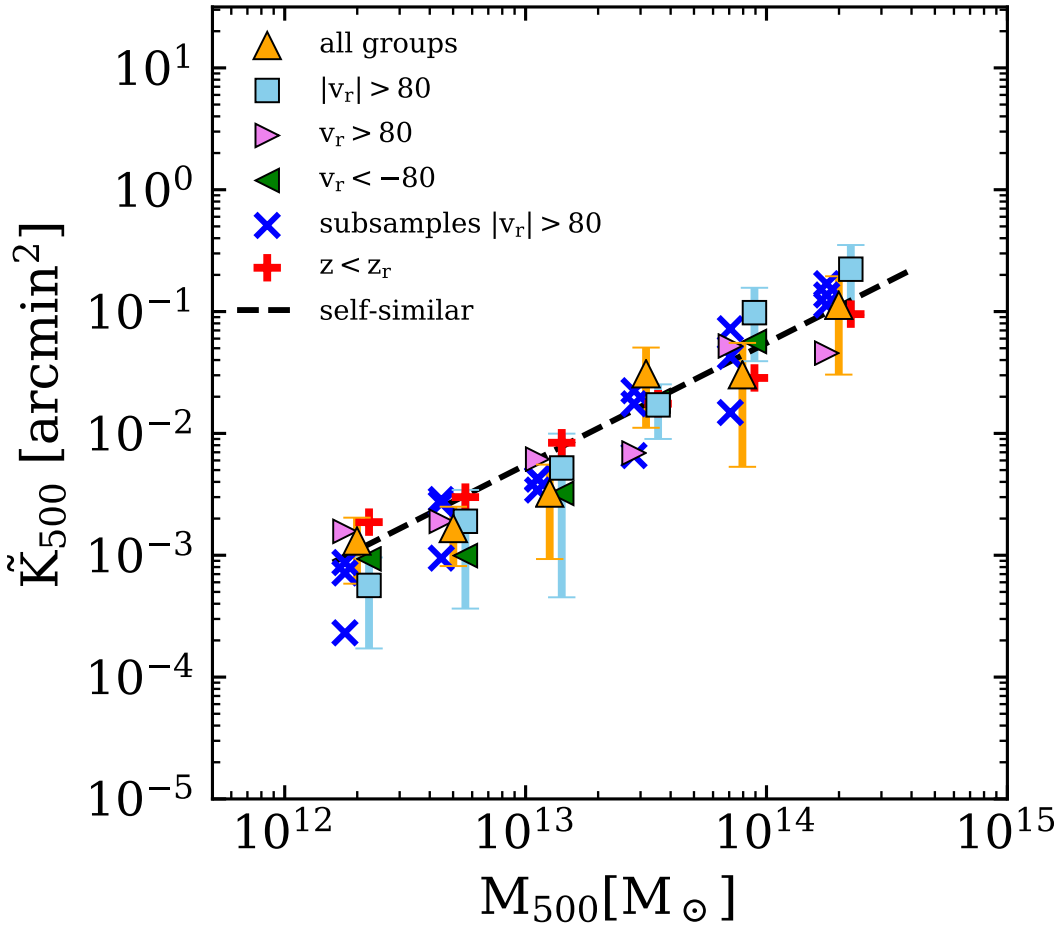

Figure 6.1. The $\tilde{K}_{500}$-M_{500} relations obtained from eight samples: all groups (yellow); groups with $|v_r| > 80\,\mathrm{km\ s^{-1}}$ (skyblue); groups with $v_r > 80\,\mathrm{km\ s^{-1}}$ (violet); groups with $v_r < -80\,\mathrm{km\ s^{-1}}$ (green); groups with $|v_r| > 80\,\mathrm{km\ s^{-1}}$ in three parts of the sky (blue), and groups with $z < z_\mathrm{r}$ (red). The data points for some samples are shifted by 0.05 dex horizontally for clarity. The dashed line shows the 'self-similar' model, $N_{\mathrm{e},500} = [(1 + f_\mathrm{H})/2m_\mathrm{p}] \cdot f_\mathrm{B} M_{500}$, where $N_{\mathrm{e},500}$ is the total number of electrons within R_{500}, $f_\mathrm{H} = 0.76$ the hydrogen mass fraction, m_p the proton mass, and $f_\mathrm{B} = \Omega_\mathrm{B}/\Omega_m = 0.16$ the cosmic baryon fraction. All the data points are based on the medians of the corresponding posterior distributions given by the MULTINEST sampler. Error bars, plotted for two samples, indicate the 68 percentile ranges of the corresponding posterior distributions.

$$d_{\rm A}(z)^2 K_{500} \equiv \sigma_{\rm T} \int_{R_{500}} n_{\rm e} {\rm d}V \,. \tag{6.11}$$

where $d_{\rm A}(z)$ is the angular diameter distance of the group in question. We also define the following quantity, analogous to the convention used in the tSZE studies,

$$\tilde{K}_{500} \equiv K_{500} (d_{\rm A}(z)/500{\rm Mpc})^2 \,. \tag{6.12}$$

We emphasize, however, that we are not measuring $\tilde{K}_{500}$ directly from the data, but using it to represent the amplitude of the density profile, $n_{\rm e}$. Once $\tilde{K}_{500}$ is obtained, we can also obtain corresponding quantities within other choices of radius, such as $\tilde{K}_{200}$ within R_{200}. The results obtained from the entire sample are shown as yellow triangles in Fig. 6.1. The data points represent the median values obtained from the posterior distribution, while the error bars are the 68 percentile range. The dashed line in Fig. 6.1 shows the 'self-similar' model prediction in which the total number of electrons within R_{500} is,

$$N_{\rm e,500} = [(1 + f_{\rm H})/2m_{\rm p}] \cdot f_{\rm B} M_{500} \tag{6.13}$$

with $f_{\rm H} = 0.76$ the hydrogen mass fraction, $m_{\rm p}$ the proton mass, and $f_{\rm B} = \Omega_{\rm B}/\Omega_m = 0.16$ the universal baryon fraction. Our data points follow well the self-similar model, indicating that the total ionized gas fractions in halos of different masses are comparable to the universal baryon fraction (see §6.4.3 for the details).

Fig. 6.1 shows that the $\tilde{K}_{500}$ - M_{500} relation is approximately a power law. This motivates another way to extract the $\tilde{K}_{500}$ associated with groups. Here we assume that

$$\tilde{K}_{500} = A \times (M_{500}/10^{13.5}{\rm M}_{\odot})^{\alpha} \tag{6.14}$$

and use the matched filter method to constrain the values of α and A. The result for the entire sample is shown in Fig. 6.2, with a yellow triangle. As comparison, each of the small dots shows the result of a random sample. Each random sample is constructed by shifting and rotating the group sample by some random amounts relative to the *Planck* maps before applying the matched filter method. In this case, the relative positions of individual groups and their spatial clustering are preserved in the random samples, but the cross correlation between the groups and the kSZE signals is destroyed. The distribution of the 200 random realizations is around $(A, \alpha) = (0, 0)$, as expected from a zeroed mean background. The symmetry relative to $(0, 0)$ is because the peculiar velocities have a roughly symmetric distribution around zero. The dipolar pattern of the random samples in (A, α) space, that positive (negative) values of A tend to correspond to positive (negative) values of α, indicates the presence of residual background fluctuation on scales larger then individual groups. In this case, the flux associated with a group in a random sample due to background fluctuation is proportional to the angular size of the group, with approximately the same probability to be positive or negative. Thus, the distribution of random groups in the $\tilde{K}_{500}$-halo mass plane is expected to have a wedge-like pattern symmetric with respect to the halo-mass axis. Since we have more lower-mass systems, the averages of the $\tilde{K}_{500}$ in the lower-mass bins are closer to zero and have lower random fluctuations. This has the effect of increasing the opening angle of the wedge within which the average $\tilde{K}_{500}$ versus halo mass relations from different random samples are confined. The assembly of straight lines, each covering the whole halo mass range in such a wedge, thus tend to show the dipolar pattern seen.

We have also used random samples in which each group in the observational sample is assigned a random position in the *Planck* sky, and found very similar results. The results indicate that the detection of the $\tilde{K}_{500}$ is very significant relative to the random

samples, as we will quantify below after the uncertainties in the measurements are tested.

6.4.2 Tests of uncertainties

A number of contaminating effects can affect our measurements. Here we present the analyses we have carried out to test the reliability of our results against the contamination. One source of uncertainty comes from the errors in the peculiar velocities of groups adopted in our analysis. As shown in detail by Wang et al. (2012) using realistic mock samples, the errors in the reconstructed peculiar velocities have a symmetric distribution around zero. Since the kSZE of a group is directly proportional to its peculiar velocity, the uncertainty in the peculiar velocities is not expected to produce any bias in our results, but will contribute to noise in the measurements. The typical uncertainty in the radial component, v_r, is between 70 and 90 km s^{-1} and so the signals from groups with peculiar velocities lower than this are all dominated by noise. To reduce the effect of this uncertainty, we separate the total sample into two sub-samples, according to whether $|v_r|$ is smaller or larger than 80 km s^{-1}. The ratio in the total number of groups between these two sub-samples is about 1/2. The model parameters, i.e. $n_{\rm e,0}$ at the six halo masses, are then tuned independently for the two sub-samples to obtain a best match to the data. The blue squares with error bars in Fig. 6.1 show the results obtained for the $|v_r| > 80$ km s^{-1} sample. As we can see, the results are not affected significantly by excluding low $|v_r|$ systems.

Another source of contamination is from fluctuations in the background and foreground, such as the primordial CMB, signals produced by background sources that are not included in our sample, and residual Galactic foreground. As discussed above, our MMF is designed to minimize the background contamination by including the background fluctuations in the noise spectrum [see equation (6.2)]. Furthermore, if the background/foreground fluctuations are not correlated with the groups in our

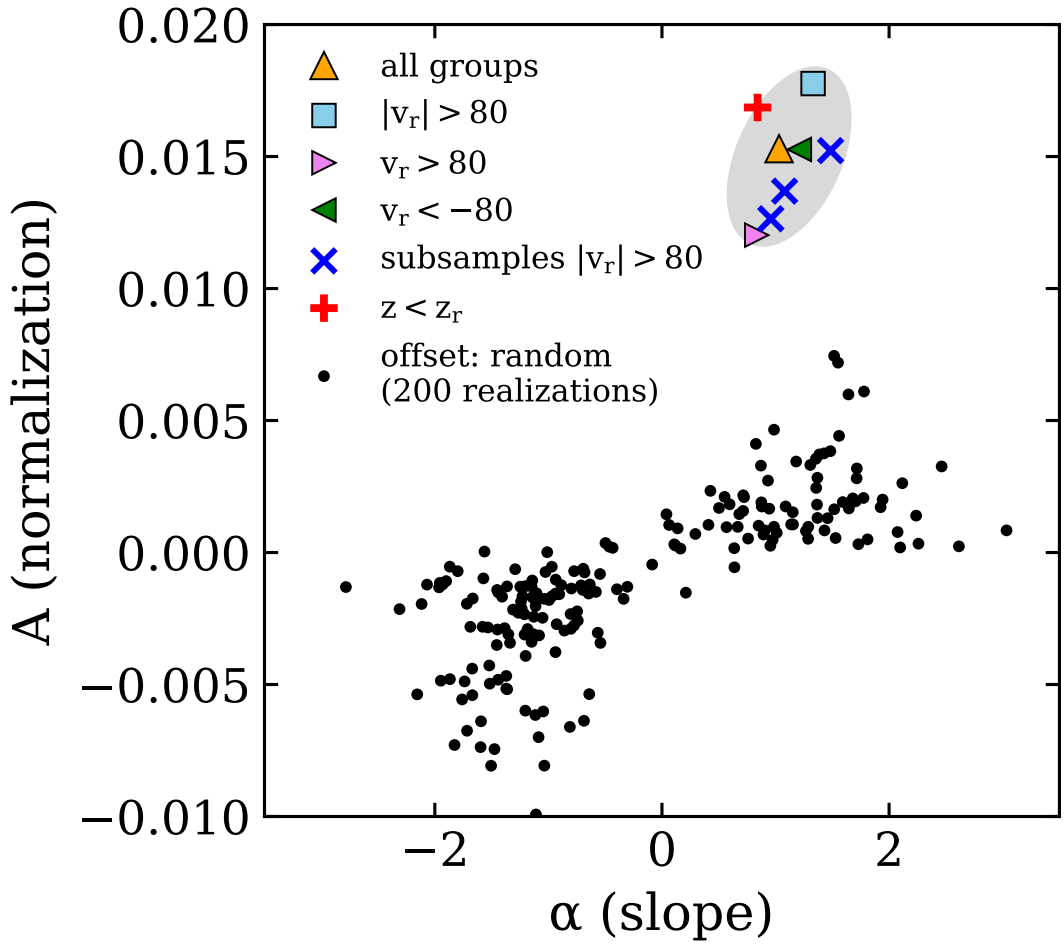

Figure 6.2. The results obtained from the power-law model, $\tilde{K}_{500} = A \times (M_{500}/10^{13.5}\mathrm{M}_{\odot})^{\alpha}$, for the eight observational samples (coloured symbols), and for the 200 random realizations, in which the total group sample is shifted and rotated by some random amounts (black dots). The shaded ellipse covers the area occupied by different observational samples, as indicated in the panel.

sample, then the contamination is not expected to lead to any bias in our results, but can increase the noise in our measurements. To test this, we divide the total sample into three sub-samples, each containing groups in a $\sim 1/3$ portion of the sky coverage, and repeat the procedures to obtain the $\tilde{K}_{500}$ for groups with $|v_r| > 80\,\mathrm{km\ s^{-1}}$ in each of the three sub-samples. The results of the three sub-samples, shown as the blue crosses in Fig. 6.1, are similar for all the mass bins. In addition, as shown in our test using random samples (Fig. 6.2), any residual background/foreground fluctuations are well below the signals we detect. These tests show that this type of contamination does not have a significant impact on our results.

Yet another source of contamination comes from the emissions of the groups in the observational wavebands, such as radio and infrared emissions and the tSZE. Although we have attempted to subtract the tSZE from the observational data, some residual may still exist. One unique property of the kSZE is that two similar groups with opposite peculiar velocities produce temperature fluctuations with opposite signs, in contrast to the contaminating emissions mentioned above, which should be independent of the sign of the peculiar velocity. To check that the signals we detect are indeed produced by the kSZE, we divide the $|v_r| > 80\,\mathrm{km\ s^{-1}}$ sample into two sub-samples, one with $v_r > 80\,\mathrm{km\ s^{-1}}$ and the other with $v_r < -80\,\mathrm{km\ s^{-1}}$, and tune the model parameters independently for the two sub-samples, along with the $|v_r| < 80\,\mathrm{km\ s^{-1}}$ sample, to achieve the best match to the data. The results obtained for these two sub-samples are shown in Fig. 6.1 as the violet and green triangles, respectively. The two give consistent $\tilde{K}_{500}$ in all the mass bins. The fact that the sub-samples of opposite peculiar velocities give similar $\tilde{K}_{500}$ - M_{500} relations [i.e. opposite signals in k defined in equation (6.4)] suggests that the contamination by emission sources does not change our results significantly, and that the signals we detect are the kSZE.

Table 6.1. The $\tilde{K}_{200}$-M_{200} relation.

Bin no.	1	2	3	4	5	6
$\log M_{200}\,[\mathrm{M}_\odot]$	12.4	12.8	13.2	13.6	14.0	14.4
number of systems	23,997	9,795	3,780	1,287	346	58
$\tilde{K}_{200}\,[10^{-2}\mathrm{arcmin}^2]$[a]	0.18	0.36	0.86	3.1	8.9	20
$\sigma(\tilde{K}_{200})\,[10^{-2}\mathrm{arcmin}^2]$[b]	0.091	0.13	0.30	1.3	4.5	9.4
$\delta(\tilde{K}_{200})\,[10^{-2}\mathrm{arcmin}^2]$[c]	0.053	0.13	0.24	1.1	3.1	5.7
z_r[d]	0.057	0.078	0.10	> 0.12	> 0.12	> 0.12
$\mathrm{cov}(i,1)$[e]	$(0.039)^2$					
$\mathrm{cov}(i,2)$	-0.00061	$(0.15)^2$				
$\mathrm{cov}(i,3)$	-0.0017	-0.0046	$(0.48)^2$			
$\mathrm{cov}(i,4)$	-0.0052	-0.013	-0.035	$(0.81)^2$		
$\mathrm{cov}(i,5)$	-0.051	-0.12	-0.34	-1.0	$(5.9)^2$	
$\mathrm{cov}(i,6)$	-0.095	-0.22	-0.62	-1.9	-19	$(13)^2$

Notes.
a. The average of the median values of the posterior distribution from the eight samples used for our test of systematic effects.
b. The variance of the median values of the posterior distribution among the eight samples.
c. The variance divided by $\sqrt{6}$ among the six independent sub-samples, each containing groups with $|v_r| > 80\,\mathrm{km\ s^{-1}}$ in a $\sim 1/6$ portion of the sky, to represent the statistical uncertainty in the estimates.
d. The redshift where the median θ_{200} of halos equals the *Planck* beam size of 5 arcmin.
e. $\mathrm{cov}(i,j)$ is the covariance of $\tilde{K}_{200}$ between i-th and j-th bins, calculated for the sample of groups with $|v_r| > 80\,\mathrm{km\ s^{-1}}$, according to the bin number assignment in the first column of the table. The values are in units of $10^{-4}\mathrm{arcmin}^4$.

However, there is still the possibility that a fraction of the signals we detect actually comes from the gas outside the virial radii of halos that moves together with the halos, given that the typical correlation length of the peculiar velocity field is large and that projection effects are enhanced by the large beam size of the *Planck*. This contamination comes in two ways: the contamination by the gas outside halos

that lies perpendicular to the LOS to groups, and that lies along the LOS to groups. To check the contamination of the first kind, we make a test in which the halos in each mass bin are divided roughly into two sub-samples: the 'resolved' one and the 'unresolved' one. Specifically, we define a redshift, $z_{\rm r}$, at which the median θ_{200} of the halos in a given mass bin is equal to the *Planck* beam size, 5 arcmin. The values of $z_{\rm r}$ for individual mass bins are listed in Table 6.1. The model parameters for the two sub-samples, $z < z_{\rm r}$ ('resolved') and of $z > z_{\rm r}$ ('unresolved'), are then tuned independently to match the observational data. The results so obtained for the 'resolved' sub-samples are shown as the red crosses in Fig. 6.1. Note that the results for the massive mass bins obtained here are not identical to those obtained from the total sample, even though all groups are resolved in these mass bin. This is because the model parameters for these groups are re-adjusted so that they, together with those for lower mass groups, produce the best match to the data (see Table 6.1 for the covariance between different mass bins). Clearly, the results obtained from the 'resolved' sample are in good agreement with those obtained from the total sample within the uncertainty.

As an independent test, we have used numerical simulations to check this directly. For the simulated halos, we estimate the LOS momentum of dark matter particles within a cylinder of a given length $l_{\rm cyl}$ and compare it with that within R_{200}. We choose dark matter particles rather than gas particles for this test because the gas distribution may depend on particular feedback models implemented in simulations. As one can see from Fig. B.1 in Appendix B, on average, the momentum of particles outside R_{200} is about 20% or less of that inside R_{200}. Note that the projected density profile used in the MMF is obtained from the LOS integration of the three dimensional profile that extends beyond R_{200} [see equation (6.5)], but suppresses the contribution from large distances, which reduces the contamination by mass outside halos, as shown in Appendix B. These test results, therefore, indicate that the LOS contamination by

large-scale coherent motion in the cosmic density field does not affect our results by a large amount.

As a further, more comprehensive test, we apply our method to the realistic kSZE mock maps for the *Planck*, constructed based on the Magneticum simulation (Dolag et al. 2016). A brief summary of the simulation and the details of the map construction are given in Appendix B. As seen in Fig. B.2, our method recovers the 'true' flux directly measured from the simulation within the uncertainty.

Finally, our MMF uses equation (6.3) to model the gas profile, while the true gas profiles may be different. To examine how our results are affected by the assumed profile, we have made tests by increasing the values of r_c by a constant factor, μ. We found that our results do not change significantly as long as $\mu < 1.5$. When μ becomes larger than 1.5, the values of $\tilde{K}_{500}$ obtained start to decrease, particularly for low-mass groups. The test shows that our results are insensitive to the gas density profile, as may be expected from the low angular resolution of the *Planck* data. However, this also indicates that the current data are not able to provide significant constraints on the details of the gas profiles around halos.

To summarize, all the eight samples we have analyzed above, namely all groups, groups with $|v_r| > 80\,\mathrm{km\ s^{-1}}$, groups with $v_r > 80\,\mathrm{km\ s^{-1}}$, groups with $v_r < -80\,\mathrm{km\ s^{-1}}$, groups with $|v_r| > 80\,\mathrm{km\ s^{-1}}$ in three parts of the sky, and groups with $z < z_\mathrm{r}$, give consistent results (as summarized in Fig. 6.1), demonstrating that our detection of the kSZE is reliable. For reference we list, in Table 6.1, the averages and variances of $\tilde{K}_{200}$ for individual halo mass bins obtained from these 8 samples. Since these samples are not independent, the variances listed can only provide a measure of systematic effects we have tested, but cannot be used to represent the statistical uncertainties in the estimates of the averages of $\tilde{K}_{200}$. To get some insight into the statistical uncertainties in our estimates, we divide the groups with $|v_r| > 80\,\mathrm{km\ s^{-1}}$ into six independent sub-samples, each containing groups in a $\sim 1/6$ portion of the sky coverage,

and obtain $\tilde{K}_{200}$ versus halo mass for each of the six sub-samples. The uncertainties in the estimates of $\tilde{K}_{200}$ obtained from these six independent sub-samples are also listed in Table 6.1. The sample-to-sample variations among the eight samples are only modestly larger than the statistical uncertainties obtained from the six independent sub-samples. Table 6.1 also lists the covariance matrix of $\tilde{K}_{200}$ among the six mass bins for groups with $|v_r| > 80\,\mathrm{km\ s^{-1}}$, obtained by drawing from the posterior distribution and then by calculating,

$$\mathbf{cov} = \left\langle \left(\tilde{\boldsymbol{K}}_{200} - \langle\tilde{\boldsymbol{K}}_{200}\rangle\right)\left(\tilde{\boldsymbol{K}}_{200} - \langle\tilde{\boldsymbol{K}}_{200}\rangle\right)^{\mathrm{T}}\right\rangle \tag{6.15}$$

where '$\langle\cdots\rangle$' denotes the mean. As expected from the nature of our simultaneous matching, the off-diagonal elements are all negative. To estimate the joint signal-to-noise from the six mass bins, we use the covariance matrix to define a multivariate normal distribution function, and calculate the cumulative probability of null detection. The probability of our measurements thus calculated corresponds to a significance level of about 6.3σ.

The results with the power-law model [see equation 8.8] for all the eight samples we have analyzed above are shown as the colour points in Fig. 6.2. The data points are clustered in a region $(A, \alpha) \in (0.015 \pm 0.0031, 1.12 \pm 0.55)$. The dispersion among the eight samples is comparable to that of A at a given α obtained from the 200 random samples, indicating that the errors in the estimates are dominated by fluctuations in the background and foreground. With this dispersion to represent the uncertainty in the results, our detection is at a level of more than 5σ, consistent with the estimate from the binned data given above.

6.4.3 The gas fraction and temperature

Fig. 8.2 shows the gas fraction within R_{200} obtained from the best-fit parameters and the assumed gas profile. The data points are the averages of and the dispersion

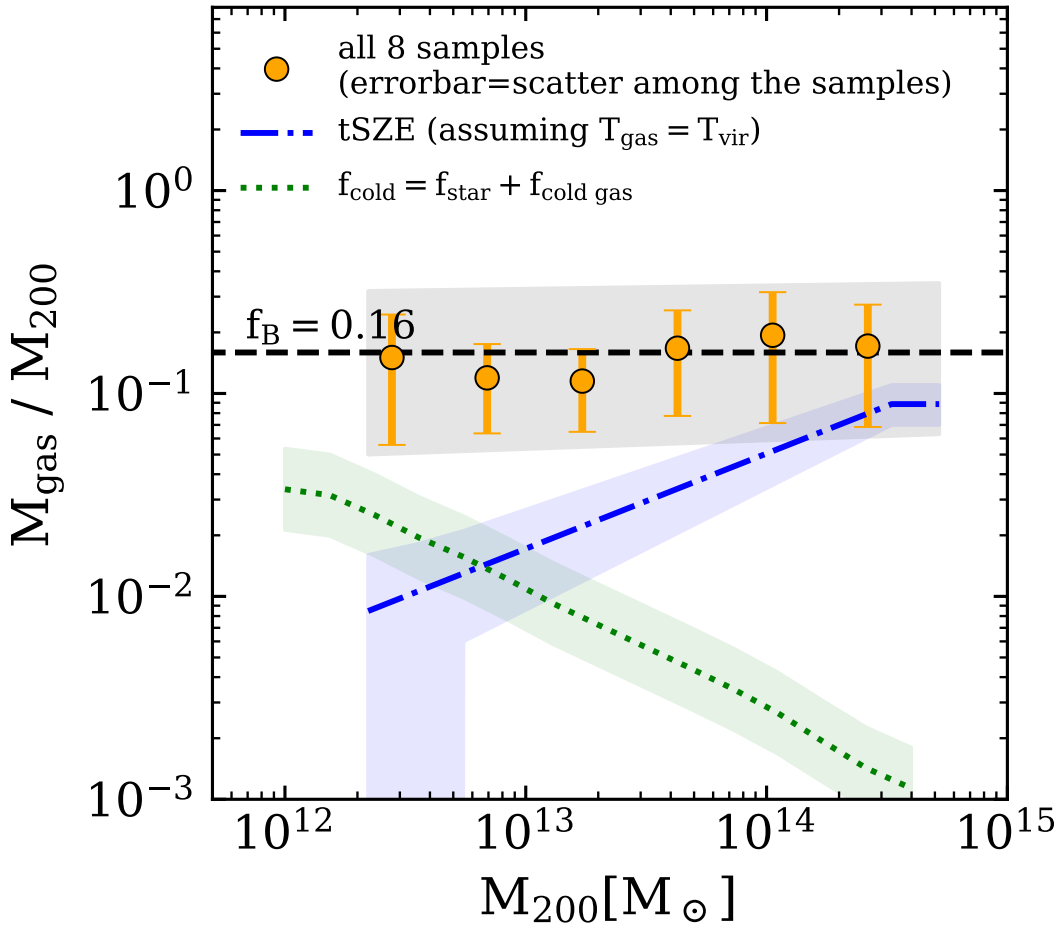

Figure 6.3. The ratio between gas mass and halo mass within R_{200} as a function of halo mass inferred from the observed $\tilde{K}_{500}$ - M_{500} relations. Data points and error bars are the averages of, and the dispersion among, all the eight samples, respectively. The shaded band is based on the ellipse shown in Fig. 6.2. The dashed line shows the universal baryon fraction of $f_B = 0.16$. The dot-dashed line represents the gas mass fraction inferred from the tSZE by Lim et al. (2018) assuming the gas to be at the virial temperature, with the shaded band indicating the typical uncertainties in the data. The dotted line shows the mass fraction in stars Lim et al. (2017a) and cold gas (Popping et al. 2014), with the shaded band indicating the typical uncertainties in the data.

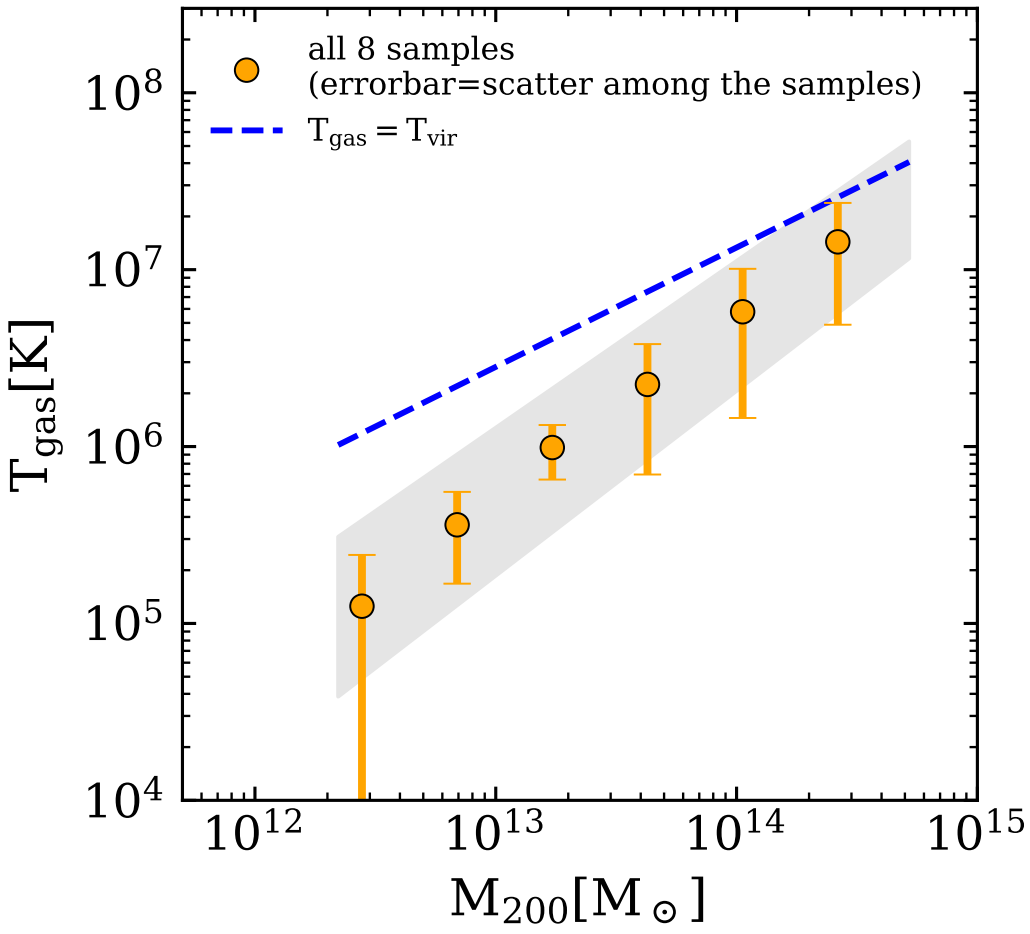

Figure 6.4. The effective gas temperature obtained by dividing the electron thermal energy obtained from the tSZE measurement by the total number of electrons obtained from the kSZE. Data points and error bars show the averages of, and the dispersion among, all the eight samples, respectively. The shaded band is based on the ellipse shown in Fig. 6.2. The dashed line shows the virial temperature as a function of halo mass.

among the eight samples (see Table 6.1 for the averages and the dispersion for each mass bins), while the shaded band contains the predictions of the power laws enclosed by the ellipse in Fig. 6.2. The inferred gas fraction is consistent with the universal baryon fraction, and much higher than the baryon fraction in stars and cold gas (shown as the dotted line).

Recently, Lim et al. (2018) applied a MMF method similar to the one used here to the *Planck* y-map to measure tSZE produced by galaxy groups using the nearly all-sky group catalog of Lim et al. (2017b), which was constructed by applying the halo-based group finder to four large redshift surveys. The gas fraction inferred from the tSZE assuming the virial temperature, $T_{\rm vir} = \mu m_{\rm p} G M_{200}/2k_{\rm B}R_{200}$, with $\mu = 0.59$ the mean molecular weight, is shown in Fig. 8.2. This fraction is much lower than that given by our kSZE data except for the most massive groups, indicating that the average temperature of the gas responsible for the kSZE in lower-mass groups is much lower than the virial temperature. The effective temperature, estimated by combining the gas mass obtained from the kSZE and the thermal energy content given by the tSZE is shown as a function of halo mass in Fig. 8.3. The derived effective temperature is about 10^5-10^6K for halos with $M_{200} \leq 10^{13.5}{\rm M}_\odot$, and much lower than the corresponding virial temperatures. This relatively low temperature has its origin from the relatively low pressure measured in Lim et al. (2018), and we refer the reader to that paper for a detailed discussion about the comparison with other tSZE measurements.

The temperature obtained here can be compared with that obtained from X-ray observations. As shown in Pratt et al. (2007), the gas temperature in clusters of galaxies appears to decline in the outer parts. Since the effective temperature inferred from SZE is sensitive to the low-density gas in the outer parts, this may partly explain the lower temperature found here than that inferred from X-ray observations. In addition, as shown in Wang et al. (2014), for a given halo mass, the scatter in the

X-ray luminosity is very large; at $M_{200} \sim 10^{13.5-14}\mathrm{M}_{\odot}$, the scatter is more than one order of magnitude (see their figure 7). As our SZE measurement is the average over all systems of a given halo mass, the thermal content inferred from it is expected to be lower than that obtained from X-ray selected samples (e.g. Planck Collaboration X 2011).

6.5 Summary and conclusion

We have examined the kinetic Sunyaev-Zel'dovich effect (kSZE) from gas in dark matter halos associated with galaxy groups as a function of halo mass down to $\log(M_{500}/\mathrm{M}_{\odot}) \sim 12.3$. Our analysis uses the stacking of about $40,000$ galaxy groups to extract the kSZE from the *Planck* temperature maps in three different frequency bands, and employs the multi-frequency matched filter (MMF) technique to maximize the signal-to-noise ratio. The MMF are matched simultaneously for individual groups so as to minimize projection effects of halos along the same LOS. Accurate reconstructed peculiar velocities of the groups are used so that we can convert reliably the observed kSZE to the amounts of ionized gas associated with galaxy groups. A number of tests are made to examine the uncertainties in our results, from errors in the reconstructed peculiar velocities, from residual background/foreground fluctuations, from contamination by the tSZE and emissions from galaxy groups, from the large beam size of *Planck* observation, from contamination by large-scale coherent motion, and from the gas density profile adopted in the MMF. We found that our results are robust against these potential sources of uncertainties. In a forthcoming paper, we will check in detail the significance of these uncertainties by applying the methods in our analysis and in the literature to the SZE light-cone maps generated from hydrodynamic simulations.

The strength of the kSZE as a function of halo mass is found to be consistent with the 'self-similar' model, in which the baryon fraction is independent of halo mass,

suggesting that the 'missing baryons' on halo scales are found. Combined with the tSZE measured for galaxy groups by Lim et al. (2018), our results indicate that the gas temperatures in low-mass halos are much lower than the corresponding halo virial temperatures. This suggests that it is the low temperature of the gas, not the total amount of baryons, that is responsible for the low thermal energy contents in low-mass halos found in tSZE and X-ray observations. Our results, therefore, provide direct support to the hypothesis that the missing baryons in galaxy groups are contained in the WHIM with temperatures between $100,000$ and one million Kelvin.

Our results also demonstrate the potential of using SZE to study both the circumgalactic media (CGM) and the galaxy formation processes that produce them. Such studies have advantages over absorption line studies, in that they are not limited to a small number of lines of sight, and that gas metallicity and ionization states are not needed to obtain the total gas mass. In the future, when high-resolution SZE data are available, the same analysis as carried out here can be used to constrain not only the total amount of ionized gas associated with galaxy groups (dark matter halos), but also to investigate the density and temperature profiles of the gas around them. One may also use galaxy groups with different star formation and/or AGN activities to study how the ionized gas distribution is affected by these activities. Clearly, the synergy between the SZE and observations of galaxy systems in other wavebands should be exploited in the future to provide detailed information both about the WHIM and about the galaxy formation processes that produce them.

CHAPTER 7

EXPLORING THE THERMAL ENERGY CONTENTS OF THE INTERGALACTIC MEDIUM WITH THE SUNYAEV-ZEL'DOVICH EFFECT[1]

We examine the thermal energy contents of the intergalactic medium (IGM) over three orders of magnitude in both mass density and gas temperature using thermal Sunyaev-Zel'dovich effect (tSZE). The analysis is based on *Planck* tSZE map and the cosmic density field, reconstructed in the SDSS DR7 volume and sampled on a grid of cubic cells of $(1h^{-1}\mathrm{Mpc})^3$, together with a matched filter technique employed to maximize the signal-to-noise. Our results show that the pressure - density relation of the IGM is roughly a power law given by an adiabatic equation of state, with an indication of steepening at densities higher than about 10 times the mean density of the universe. The implied average gas temperature is $\sim 10^4\,\mathrm{K}$ in regions of mean density, $\rho_\mathrm{m} \sim \overline{\rho}_\mathrm{m}$, increasing to about $10^5\,\mathrm{K}$ for $\rho_\mathrm{m} \sim 10\,\overline{\rho}_\mathrm{m}$, and to $> 10^6\,\mathrm{K}$ for $\rho_\mathrm{m} \sim 100\,\overline{\rho}_\mathrm{m}$. At a given density, the thermal energy content of the IGM is also found to be higher in regions of stronger tidal fields, likely due to shock heating by the formation of large scale structure and/or feedback from galaxies and AGNs. A comparison of the results with hydrodynamic simulations suggests that the current data can already provide interesting constraints on galaxy formation.

---

[1] THE CONTENTS OF THIS CHAPTER ARE PUBLISHED IN LIM ET AL. 2018, MNRAS, 480, 4017.

7.1 Introduction

The baryon mass that has been identified by current observations in galaxies and galaxy systems in the $z \sim 0$ universe accounts for about 2% of the critical density (e.g. Fukugita et al. 1998, Fukugita & Peebles 2004), which is about one third of the cosmic baryon abundance predicted by the primordial nucleosynthesis (e.g. Olive et al. 2000) and required to explain the fluctuations in the cosmic microwave background radiation (CMB) (e.g. Hinshaw et al. 2013, Planck Collaboration XIII 2016). Thus, a large fraction of baryons must be contained in the intergalactic medium. Indeed, observations of the Lyα forest at low-z suggest that about $30 - 40\%$ of the cosmic baryons may be hidden in the forest (e.g. Danforth & Shull 2008), but the results are still very uncertain, because the sample of low-z forest is small and because of uncertainties in the ionization correction that is needed to obtain the total hydrogen density.

Cosmological simulations have shown that the cosmic baryons in the low-z universe can exist in a variety of forms. In addition to stars and cold gas that are associated with galaxies, some baryons are predicted to be contained in the hot gaseous halos that are produced by the collapse of dark matter halos and feedback of galaxy formation. Furthermore, many gas simulations demonstrated that a large fraction of the baryons in the low-z universe actually reside in a diffuse warm-hot intergalactic medium (WHIM) (e.g. Cen & Ostriker 1999, Daé et al. 1999, 2001), with temperature in the range between 10^5 and 10^7 K. This medium is generated by a combination of feedback from galaxy formation and shocks accompanying the formation of the large-scale structure (e.g. Sunyaev & Zeldovich 1972, Nath & Silk 2001, Furlanetto & Loeb 2004, Rasera & Teyssier 2006).

Clearly, a comprehensive investigation of all gas components is required to have a complete understanding of how galaxies and larger scale structure form and evolve in the universe. In particular, the study of the WHIM, which is potentially the

dominating gas component in the low-z universe but yet poorly understood so far, is important not only for obtaining a complete census of the cosmic baryons, but also for understanding how galaxies and structure formation interact with the gas component in the universe. Indeed, observations of Lyα absorption systems have revealed that the intergalactic medium (IGM) at $z \sim 0$ is significantly enriched in metal, indicating that a significant fraction of baryons may have been ejected from galaxies and moved to the IGM by dynamical processes, such as winds driven by stellar and AGN feedback or by ram-pressure stripping (e.g. Aguirre et al. 2001). In addition, the similarity in the metallicities of the intra-cluster media (ICM) observed for different clusters suggests that the formation of the ICM is perhaps dominated by inflow of the gas from the large scale cosmic web (e.g. Werner et al. 2013, Ettori et al. 2015, McDonald et al. 2016, Mantz et al. 2017). Both of these demonstrate the importance of the interaction between galaxies and gas in shaping the gas media we observe. However, the detection of the WHIM in observations is challenging. At a temperature of $T \sim 10^5$-10^7 K, the gas is almost completely ionized, making it difficult to detect in absorption. The low density of the WHIM also makes it difficult to detect in UV and X-ray emission.

With the advent of large surveys of the cosmic microwave background (CMB), a promising new avenue is opened up. As the CMB photons travel to reach us, they are scattered by free electrons associated with hot gas via Compton scattering. This produces a change in the energy distribution of the CMB photons, an effect referred to as the thermal Sunyaev-Zel'dovich effect (tSZE; Sunyaev & Zeldovich 1972). The tSZE is a measure of the projected electron pressure along the line of sight, and so it provides a way to probe the thermal energy content of the ionized gas in the universe. Compared to X-ray emissions, the tSZE is more sensitive to diffuse gas, making it more suitable for probing relatively low-density media, such as WHIM. It

also complements absorption studies of highly ionized gas, as ionization correction and gas metallicity are not needed to obtain the total gas content.

Great amounts of effort have been made to measure the tSZE from observational data. Planck Collaboration XI (2013) used the *Planck* multi-frequency CMB temperature maps and a sample of locally brightest galaxies to study the tSZE produced by galaxy systems down to a halo mass $\sim 4 \times 10^{12}\,\mathrm{M}_{\odot}$. Remarkably, they found that their results are consistent with the self-similar model in which the hot gas fraction in halos is independent of halo mass. A similar conclusion was also reached by Greco et al. (2015) using a similar method. Tanimura et al. (2017) and de Graaff et al. (2017) reported the detection of warm-hot gas in cosmic filaments by cross-correlating filaments identified by galaxy pairs with the *Planck* tSZE map. By cross correlating the *Planck* tSZE map with the mass density map obtained from the gravitational lensing data of the CFHTLenS survey, Van Waerbeke et al. (2014), Ma et al. (2015), Hojjati et al. (2015) found significant correlations between the gas and dark matter distributions. Finally, Hill & Spergel (2014) reported a significant detection of cross-correlation between the *Planck* CMB lensing potential map and the *Planck* tSZE map. All these indicate that tSZE provides a powerful way to investigate the gas distribution in the universe.

Recently, Lim et al. (2018a, b) used the group catalogs of Yang et al. (2007) and Lim et al. (2017) to extract, from the *Planck* CMB observation, both the tSZE and kinematic SZ effects (kSZE) of the warm-hot gas associated with halos of different masses. They employed the matched filter technique (Haehnelt & Tegmark 1996, Herranz et al. 2002, Melin et al. 2005, 2006) to maximize the signal-to-noise ratio. In particular, they matched the filter simultaneously to all galaxy groups to constrain the corresponding signals jointly as a function of group mass, so as to minimize projection effects for groups lying closely in the sky. Combining the tSZE and kSZE, they found that the total amount of the baryons associated with dark matter halos

are consistent with the universal baryon fraction, even in low-mass halos, but that the gas temperature is much lower than the virial temperature in low-mass halos. Various tests performed demonstrate that the method is very powerful in extracting the SZE signals effectively and reliably.

In this paper, we study the thermal contents of the IGM at $z \sim 0$, extending the analyses of Lim et al. (2018a, b) by going beyond scales of dark matter halos. Specifically, we constrain the pressure of the gas in different environments, using the total density field reconstructed by Wang et al. (2009, 2016) in the Sloan Digital Sky Survey Data Release 7 (SDSS DR7) volume and using the *Planck* tSZE map as the observational constraint. As we will see, our analysis is able to provide constraints on the thermal energy contents of the WHIM in regions covering almost three orders of magnitude in mass density. In particular, our results are shown to be capable of providing constraints on models of galaxy formation. The outline of this paper is as follows. We describe the data used in our analysis in Section 7.2, and our method to constrain the gas pressure in Section 7.3. We present our main results and comparisons with results from numerical simulations in Section 7.4. We summarize and conclude in Section 7.5.

7.2 Data

7.2.1 The *Planck* y-map

The temperature change in the CMB spectrum by the tSZE is given by

$$\left(\frac{\Delta T}{T_{\rm CMB}}\right) = g(x)y \equiv g(x)\frac{\sigma_{\rm T}}{m_{\rm e}c^2}\int P_{\rm e}{\rm d}l, \tag{7.1}$$

where y is the Compton parameter, $g(x) = x\coth(x/2) - 4$ is the conversion factor at a given $x \equiv h\nu/(k_{\rm B}T_{\rm CMB})$, $T_{\rm CMB} = 2.7255\,{\rm K}$, $\sigma_{\rm T}$ is the Thompson cross-section, $m_{\rm e}$ is the electron rest-mass, c is the speed of light, $P_{\rm e} = n_{\rm e}k_{\rm B}T_{\rm e}$ is the electron pressure with

$n_{\rm e}$ the free electron density, and dl is the path length along each line-of-sight (LOS). The *Planck* survey (Tauber et al. 2010, Planck Collaboration I 2011) is an all-sky observation of the CMB in nine frequency bands from 30 to 857 GHz, with angular resolutions ranging between 5 and 31 arcmin. We use the *Planck* MILCA (Modified Internal Linear Combination Algorithm; Hurier et al. 2013) all-sky tSZ Compton parameter map (Planck Collaboration XXII 2016), also known as the MILCA y-map, which is part of the *Planck* 2015 data release [2]. The MILCA y-map is constructed from the full mission data set of the *Planck*, using a combination of different frequency maps to minimize the primary CMB fluctuations and the contamination from foreground sources. The details about the y-map construction can be found in the original papers. We mask the brightest 40% of the sky to limit the Galactic foreground contamination, by using the corresponding mask provided in the *Planck* 2015 data release. We also apply the mask for point sources, provided in the same data release, to reduce the contamination from radio and infrared sources.

As a test, we have also applied the same analysis to the NILC (Needlet Independent Linear Combination; Remazeilles et al. 2011) y-map (Planck Collaboration XXII 2016), which treats dust contamination differently than the MILCA, as well as to the three low HFI channel (100, 143 and 217 GHz) maps, which are less contaminated by dust emission. In both cases, no significant change is found in our results.

7.2.2 The dark matter density field in the SDSS DR7 volume

Another set of data that we use for our analysis is the reconstructed cosmic density field in Sloan Digital Sky Survey Data Release 7 (SDSS DR7; Abazajian et al. 2009) volume given by Wang et al. (2016) (W16). The reconstruction uses galaxy groups selected with the halo-based group finder (Yang et al. 2005, 2007) to represent dark matter halos. Extensive tests using mock galaxy redshift surveys constructed from the

[2] `https://pla.esac.esa.int`

conditional luminosity function (CLF) model (e.g. Yang et al. 2003, van den Bosch et al. 2003) and semi-analytical models (Kang et al. 2005) revealed that this group finder is very successful in grouping galaxies into their common dark matter halos (see Yang et al. 2007). By partitioning the SDSS volume into domains associated with individual groups, and by modelling the mass distribution in each domain using profiles calibrated with N-body simulations, W16 reconstructed the real space density field within the entire SDSS DR7 volume. W16 used groups of halo masses $M_{\rm h} \geq 10^{12}h^{-1}{\rm M}_{\odot}$ which are complete for $z \leq 0.12$. The reconstruction was restricted to the contiguous region of the SDSS DR7 in the Northern Galactic Cap of $\sim 7,000\,{\rm deg}^2$. While W16 provides the density field smoothed on various scales, we use the density field smoothed on $1\,h^{-1}$Mpc for our analysis.

7.2.3 Hydrodynamic simulations for comparison

To compare with our results, we use two recent cosmological hydrodynamic simulations. The first is Illustris (Nelson et al. 2015), which follows the evolution of the simulated Universe in a box of $(106.5\,{\rm Mpc})^3$, implementing physical processes such as radiative cooling, star formation, and various feedback processes. The free parameters in sub-grid models were constrained by observations as well as simulations of higher resolutions. We use Illustris-1, their flagship run, which has the highest mass resolution, with gas particle mass of $1.6 \times 10^6\,{\rm M}_{\odot}$. The simulation adopted WMAP9 cosmology, with $\Omega_{\rm m} = 0.273$, $\Omega_{\Lambda} = 0.727$, $h = 0.704$, and a baryon fraction, $\Omega_{\rm B}/\Omega_{\rm m} = 0.165$ (Hinshaw et al. 2013).

Another simulation we use is the Evolution and Assembly of GaLaxies and their Environments (EAGLE; Schayet et al. 2015, Crain et al. 2015, McAlpine et al. 2015). EAGLE implemented sub-grid physics models for cooling, star formation, and stellar and AGN feedback to evolve different types of particles, such as gas, stars, dark matter, and black holes. The parameters for these models are tuned to match

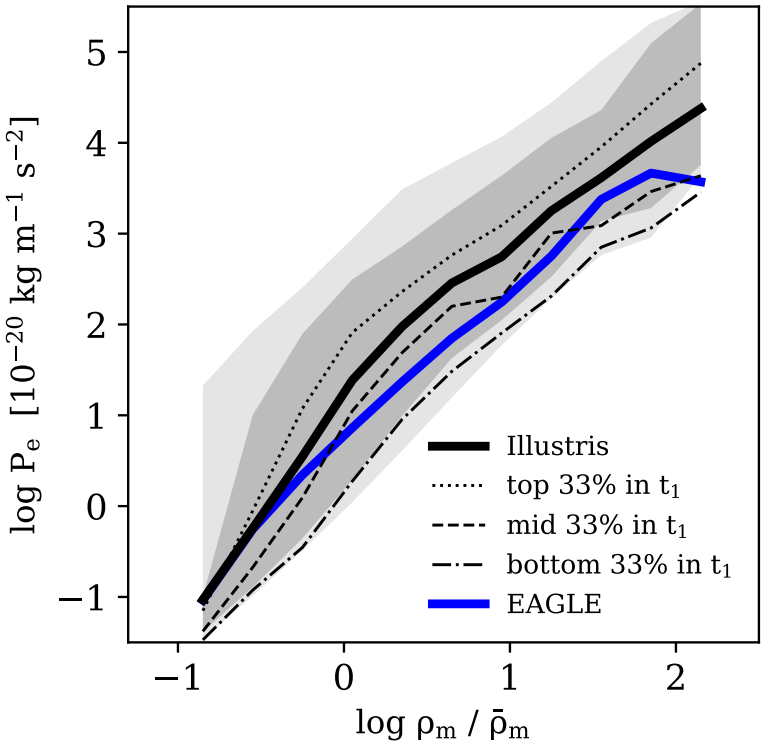

Figure 7.1. The relations between the electron pressure and mass density for grid-cells of $1\,(h^{-1}\mathrm{Mpc})^3$ obtained from Illustris and EAGLE. The solid lines show the mean relations, and the gray bands show the 1σ and 2σ dispersion obtained from Illustris. The thin lines show the mean relations obtained for the three sub-samples of the grid-cells according to the ranking in the tidal field strength, t_1, at given density, obtained from Illustris.

a set of observations, such as the stellar mass function and stellar mass - black hole mass relation at $z \sim 0$. EAGLE provides a number of simulations assuming different model parameters and box sizes. We present results based on the 'RefL0100N1504' run, their fiducial simulation, which has the largest box size, $(100\,\mathrm{Mpc})^3$, among all the EAGLE runs. Our tests showed that the results used for our comparisons were not significantly affected if other simulations were used. EAGLE assumes the *Planck* cosmology, $(\Omega_\mathrm{m}, \Omega_\Lambda, h) = (0.307, 0.693, 0.678)$ (Planck Collaboration XIII 2016).

Fig. 7.1 shows the pressure - density relation averaged over grid-cells of $1\,(h^{-1}\mathrm{Mpc})^3$ in the simulations. The cell size chosen here is to match the size adopted in our anal-

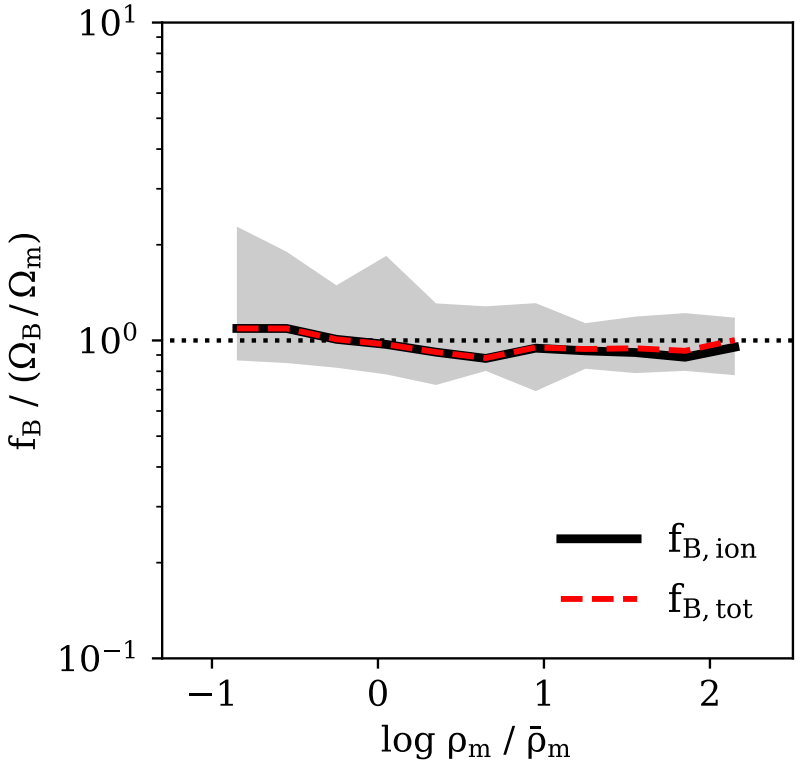

Figure 7.2. The ionized gas mass fraction, normalized by the cosmic mean baryon fraction, in grid-cells as a function of mass density as obtained from Illustris. The thick line shows the mean relation, while the shaded region indicates the 1σ dispersion. For comparison the dashed line shows the average total baryon fraction as a function of mass density.

ysis. The mean relations, presented by the thick solid lines, have slopes roughly consistent with the adiabatic equation of state, $P_e \sim \rho^{5/3}$ for both simulations. However, the average pressure at $\rho_m > \overline{\rho}_m$ predicted by Illustris is about two to three times as high as that predicted by EAGLE, presumably because of cosmic variance and of the differences in the feedback models adopted in the two simulations. The scatter in pressure at a given density (the gray bands showing the 1σ and 2σ dispersion given by Illustris) is also very large. This large scatter motivates us to explore the possibility of a second parameter, in addition to density, that can affect the gas pressure (see §7.2.4).

On the scale of a grid cell, the distribution of baryons is expected to approximately follow that of the dark matter. Since most of the baryons are expected to be in the state of diffuse ionized gas, the total amount of ionized gas in a grid cell is roughly that given by the universal baryon fraction. Fig. 7.2 shows that this is indeed the case, at least in simulations. This result will be used to interpret the tSZE obtained from the observational data.

7.2.4 The large-scale tidal field

We have tested a number of quantities available from the simulation to see how they affect the gas pressure. These include stellar mass, star formation rate, black hole mass, mean velocity, velocity dispersion, the fluctuation in dark matter density, distance to nearest halos, fraction of mass contained in halos, and local tidal field, all estimated for individual grid cells that are used to sample the density field. We found that, among these quantities, the large-scale tidal field is the most significant second parameter that can change the gas pressure on top of the dependence on local dark matter density.

In our analysis, we tested two definitions of large-scale tidal field, one based on dark matter halos and the other based on the mass density field. We find that both definitions of the tidal field give practically the same results in our analysis [see Wang et al. (2011) for details of how the two tidal fields are correlated]. In what follows, we only present results using the halo-based tidal field. The estimate of the halo-based tidal field is based on the halo tidal force along a direction $\mathbf{t}$ exerted on the surface of a sphere of diameter $1\,h^{-1}$Mpc that approximates a grid-cell, normalized by the self-gravity of the matter inside the grid-cell,

$$f(\mathbf{t}) = \frac{\sum_i GM_i R_{\rm g}(1+3\cos 2\theta_i)/r_i^3}{2GM_{\rm g}/R_{\rm g}^2} \tag{7.2}$$

where the summation is over all the halos, M_i is the mass of halo i, $R_{\rm g} = 0.5h^{-1}$Mpc is the radius of the sphere that approximates the grid-cell, $M_{\rm g}$ is the mass enclosed within the grid-cell in question, r_i is the separation between the center of the grid-cell and the i-th halo, and θ_i is the angle between $\mathbf{t}$ and $\mathbf{r}_i$ (e.g. Wang et al. 2011). So defined, the tidal field measures the total tidal force exerted on a grid-cell normalized by its self-gravity. The ellipsoid of the local tidal field is then used to compute the eigenvalues t_1, t_2, and t_3 ($t_1 \geq t_2 \geq t_3$) of the halo tidal tensor. We use t_1 to describe the strength of the halo tidal field at any given grid-cell of $1\,(h^{-1}\mathrm{Mpc})^3$.

Fig. 7.1 shows that grid-cells with higher t_1 have higher mean pressure at a given density in the Illustris simulation, and the result for the EAGLE simulation is qualitatively the same. Apparently, the formation of large scale structure and/or intense feedback in regions of strong tidal field can heat the IGM on large scales.

7.3 Methods

7.3.1 The matched filter technique

Extracting the tSZE signals reliably requires to optimize the signal-to-noise, since the signal is generally more than an order of magnitude lower than other sources, such as the primary CMB anisotropy, Galactic foreground, and cosmic infrared background. Thus, using a simple aperture photometry can lead to large uncertainties in the results (e.g. Melin et al. 2006). Here, we employ the matched filter (MF) technique (Haehnelt & Tegmark 1996, Herranz et al. 2002, Melin et al. 2005, 2006), which is designed to minimize source confusions and contamination, and to maximize the signal-to-noise by imposing prior knowledge of the signals and the noise power spectra. In the MF technique, the Fourier transform of the filter that optimizes the signal-to-noise is given by

$$\hat{F}(\boldsymbol{k}) = \left[\int \frac{|\hat{\tau}(\boldsymbol{k'})\hat{B}(\boldsymbol{k'})|^2}{P(k')} \frac{\mathrm{d}^2 k'}{(2\pi)^2} \right]^{-1} \frac{\hat{\tau}(\boldsymbol{k})\hat{B}(\boldsymbol{k})}{P(k)} \tag{7.3}$$

where, in our application, $\hat{\tau}(\boldsymbol{k})$ is the Fourier transform of the projected electron pressure, $\hat{B}(\boldsymbol{k})$ is the Fourier transform of the Gaussian beam function that mimics the convolution in the *Planck* observation with the FWHM of 5 arcmin, and $P(k)$ is the noise power spectra. Because the MILCA y-map is already cleaned of the primary CMB anisotropy, $P(k) = P_{\rm noise}$ where $P_{\rm noise}$ is the power spectrum of the noise map for the MILCA y-map, as provided in the *Planck* data release. Both the shape and amplitude of the projected electron pressure profile, $\hat{\tau}(\boldsymbol{k})$, are simultaneously constrained by matching the filters to all pixels, as to be described in details below.

7.3.2 The pressure - density relation

We assign a value of electron pressure to each of the grid-cells by assuming a simple double power-law relation between the reconstructed matter density field and the pressure:

$$P_{\rm e} = \begin{cases} A \times (\rho_{\rm m}/\rho_{\rm m,0})^{\alpha_1}, & \text{if } \rho_{\rm m} \leq \rho_{\rm m,0} \\ A \times (\rho_{\rm m}/\rho_{\rm m,0})^{\alpha_2}, & \text{if } \rho_{\rm m} > \rho_{\rm m,0}. \end{cases} \tag{7.4}$$

The pressure field is smoothed with a Gaussian kernel of radius $1\,h^{-1}$Mpc, and is integrated along each line of sight to obtain the predicted y-parameters for all the pixels in in the *Planck* y-map. The predicted y-parameter profile at each pixel is used in equation (8.3) to obtain the corresponding filter, and the filters for all the pixels are matched to the *Planck* map. Finally, we perform the Monte Carlo Markov Chain (MCMC) to constrain the parameters of the double power-law relation so as to yield the best overall match between the filters and the observed y-map, based on the sum of the χ^2 over all the pixels. We assume a constant background contribution from free electrons outside the volume in which the density reconstruction was made, and the background level is treated as a free parameter to be constrained with the MCMC. This is expected to be valid as long as the structures lying beyond the boundary of

Table 7.1. The medians and 68% ranges (95% ranges, in parentheses) of the posterior distribution of the parameters in the double power-law model. The values are obtained from the marginalized distributions of the parameters.

Samples	A^{a}	$\rho_{\mathrm{m},0} \,/\, \overline{\rho}_{\mathrm{m}}$	α_1	α_2	$P_{\mathrm{e}}(\rho_{\mathrm{m}} = \overline{\rho}_{\mathrm{m}})$ [b]
all cells	$20^{+42\,(65)}_{-9.5\,(13)}$	$3.0^{+0.90\,(1.6)}_{-0.31\,(0.48)}$	$1.7^{+0.12\,(0.27)}_{-0.41\,(0.57)}$	$2.2^{+0.18\,(0.26)}_{-0.15\,(0.22)}$	$2.6^{+1.7\,(4.0)}_{-1.5\,(1.8)}$
high t_1	$25^{+31\,(70)}_{-7.3\,(10)}$	$1.1^{+2.8\,(3.2)}_{-0.35\,(0.57)}$	$1.9^{+0.27\,(0.38)}_{-0.63\,(0.77)}$	$1.8^{+0.20\,(0.29)}_{-0.31\,(0.42)}$	$20^{+23\,(61)}_{-5.8\,(10)}$
mid t_1	$20^{+33\,(82)}_{-9.0\,(11)}$	$2.1^{+3.5\,(4.3)}_{-0.89\,(1.6)}$	$1.7^{+0.58\,(0.72)}_{-0.39\,(0.51)}$	$1.9^{+0.29\,(0.38)}_{-0.08\,(0.17)}$	$5.8^{+2.8\,(4.2)}_{-2.9\,(3.7)}$
low t_1	$2.6^{+3.9\,(8.9)}_{-1.1\,(1.4)}$	$1.4^{+2.7\,(3.4)}_{-0.71\,(0.97)}$	$1.8^{+0.36\,(0.49)}_{-0.45\,(0.63)}$	$2.2^{+0.19\,(0.24)}_{-0.41\,(0.61)}$	$1.3^{+1.2\,(2.7)}_{-0.60\,(0.80)}$

Notes.
a. The values are in units of $10^{-20}\,\mathrm{kg\,m^{-1}\,s^{-2}}$.
b. The pressure at the mean density obtained from the posterior prediction of the double power-law model, in units of $10^{-20}\,\mathrm{kg\,m^{-1}\,s^{-2}}$.

the reconstruction volume are not correlated with the structures within the volume. In this case, the fluctuations of the background do not lead to bias, but increase noise. Since the cross-correlation of the tSZ signals with the residual contamination by the Cosmic Infrared Background (CIB) is found to be insignificant (Planck Collaboration XXII 2016), the assumption of a constant background is expected to be able to remove any residual CIB contamination in a statistical sense.

7.4 Results

7.4.1 The pressure - density relation

Fig. 8.4 shows the pressure - density relation obtained from the data. The orange band shows the 1σ scatter based on the uncertainties of the constrained parameters given by the MCMC, representing the dispersion in the relation. For reference, we list, in Table 7.1, the medians, 68, and 95 percentile ranges obtained from the marginalization of the posterior distributions of the parameters. Since the overall relation is roughly a single power law, the two parameters in the double power law model, A and $\rho_{\mathrm{m},0}$, are highly degenerate. To describe the overall amplitude of the relation, we

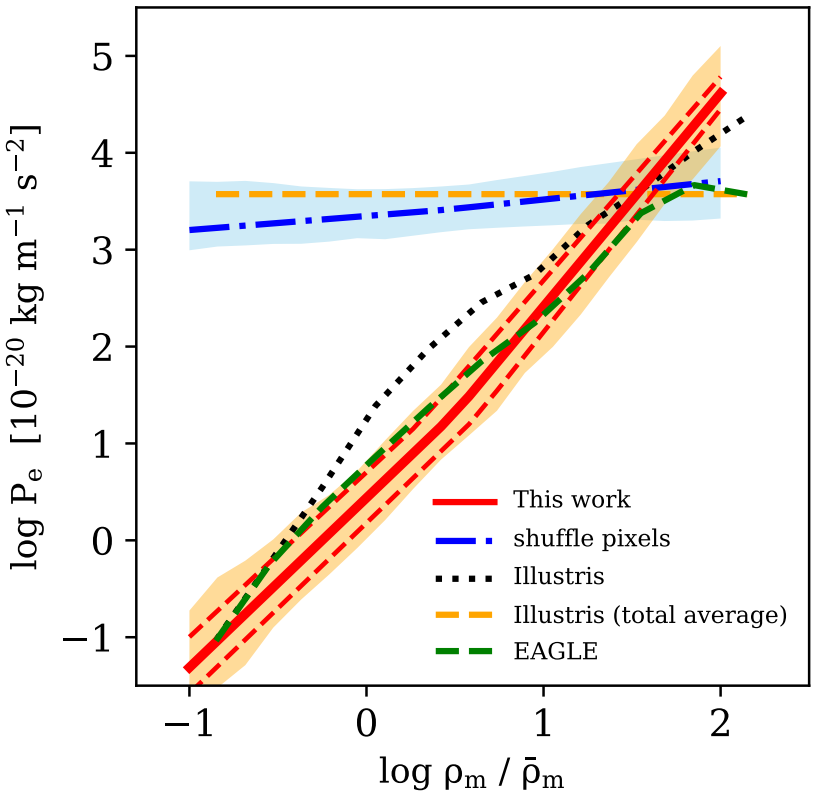

Figure 7.3. The pressure - density relation, compared with the simulations. The red line shows the mean relation obtained with our method, with the orange band showing the 1σ dispersion estimated from the uncertainties in the constrained parameters, as given by the MCMC sample. The red dashed lines are the relations constrained separately for the two nearly equal-sized sub-samples of the grid-cells according to the sky region, one including the Sloan Great Wall (Gott et al. 2005), and the other not. The black and green lines show the mean relations from Illustris and EAGLE, respectively. Also, the mean relation obtained from shuffling the grid-cells (dot-dashed), with the band showing the relations from 50 realizations, is compared with that obtained from Illustris (orange dashed).

therefore also list, in the last column of the table, the median value, as well as the 68 and 95 percentile ranges, of the pressure measured at the mean density.

The slope of the relation for low-density region is close to that expected from an adiabatic equation of state, $P_e \sim \rho^{5/3}$, but the relation is slightly steeper for dense regions. This indicates that gas in the dense regions may be heated by additional

sources, such as supernova and AGN feedback from galaxies residing in these regions, and shocks associated with the formation of large scale structure. As a null test, we randomly shuffle the grid-cells and apply the same method to constrain the relation. The mean relation, plotted as the thick blue line, and the relations from a total of 50 realizations, plotted as the sky-blue band, are all flat, as expected from random fluctuations. The mean pressure obtained from shuffling the grid-cells is determined by the distribution of the cells in density combined with the pressure - density relation. This demonstrates that our method is able to detect the true relation between the gas and dark matter distribution. Note that the scatter in the constrained pressure - density relation is similar to that among the random samples, indicating that the uncertainty in the constrained relation is dominated by fluctuations of the background/foreground. Finally, we divide the grid-cells into two nearly equal-sized sub-samples according to the sky region they belong to, one including the Sloan Great Wall (Gott et al. 2005) and the other not, and constrain the relation separately for the two sub-samples. The mean relations obtained for the two sub-samples, shown by the red dashed lines, are within the 1σ scatter band obtained for the entire sample, indicating that cosmic variance does not affect our results significantly.

We find that using a different smoothing scale, e.g. $2\,h^{-1}$Mpc, leads to no significant change in our results. Choosing an even larger smoothing scale leads to bigger uncertainties in the constrained relations, because of the decreased number of grid-cells. Note that the uncertainty in the reconstructed peculiar velocities is about $100\,\mathrm{km\,s^{-1}}$ (see W16), which corresponds to $\sim 1\,h^{-1}$Mpc in real-space positions. Thus, choosing a smoothing scale smaller than $\sim 1\,h^{-1}$Mpc may not be appropriate.

7.4.2 The temperature - density relation

Assuming that the ionized gas mass fraction with respect to the total mass within the grid-cells approximately equals the cosmic mean baryon fraction, as is motivated

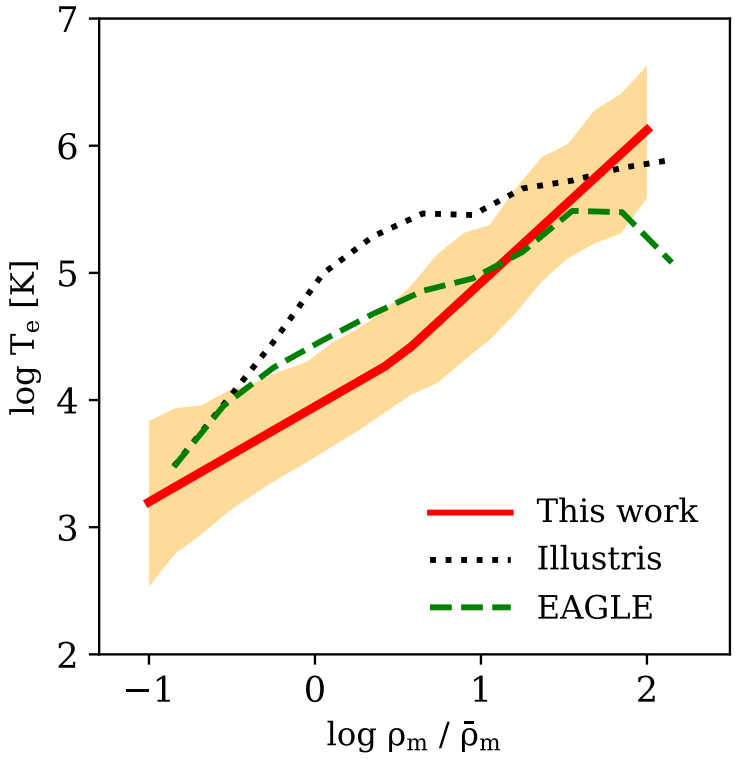

Figure 7.4. The temperature - density relation derived assuming $\rho_{\rm B,ion}/\rho_{\rm m} \sim \Omega_{\rm B}/\Omega_{\rm m} = 0.16$ and $n_{\rm e} = \rho_{\rm B,ion} \cdot [(1 + f_{\rm H})/2m_{\rm p}]$, where $\rho_{\rm B,ion}$ is the ionized gas density, $f_{\rm H} = 0.76$ is the hydrogen mass fraction and $m_{\rm p}$ is the proton mass. The red line shows the mean relation obtained with our method, with the orange band showing the 1σ dispersion estimated from the uncertainties in the constrained parameters. The dotted and dashed lines show the mean relations from Illustris and EAGLE, respectively.

by the simulation results shown in Fig. 7.2, one can convert the observed pressure - density relation to a relation between gas temperature and mass density. Specifically, we assume $\rho_{\rm B,ion}/\rho_{\rm m} \sim \Omega_{\rm B}/\Omega_{\rm m} = 0.16$ and $n_{\rm e} = \rho_{\rm B,ion} \cdot [(1 + f_{\rm H})/2m_{\rm p}]$, where $\rho_{\rm B,ion}$ is the ionized gas density, $f_{\rm H} = 0.76$ is the hydrogen mass fraction and $m_{\rm p}$ is the proton mass, to obtain the electron temperature. The mean temperature - density relation thus derived for the entire observational sample is shown in Fig. 7.4 by the solid line, along with the 1σ dispersion estimated from the uncertainties in the constrained parameters. As one can see, the average temperature is about 10^4 K in regions of

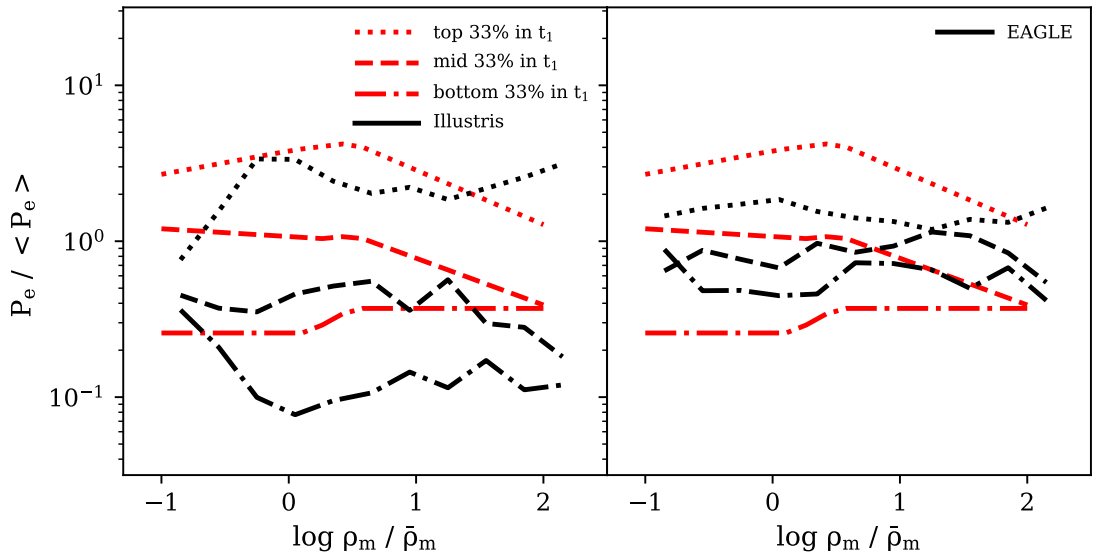

Figure 7.5. The pressure - density relations obtained for the three sub-samples of the grid-cells according to their ranking in the tidal field strength, t_1, at given density. The red lines in both panels show the mean relations obtained from the data, which are compared with the results of Illustris in the left panel and that of EAGLE in the right panel. For each case, the results are normalized by the corresponding mean relation shown in Fig. 8.4.

mean density, $\rho_{\rm m} \sim \overline{\rho}_{\rm m}$, increasing to $\sim 10^5\,{\rm K}$ for $\rho_{\rm m} \sim 10\,\overline{\rho}_{\rm m}$, the typical density for cosmic filaments and sheets (e.g. Shen et al. 2006), and to $> 10^6\,{\rm K}$ for $\rho_{\rm m} \sim 100\,\overline{\rho}_{\rm m}$, the typical density of dark matter halos.

7.4.3 Dependence on local tidal field

To examine how the pressure - density relation depends on large-scale environment, we divide the grid-cells into three sub-samples, each containing a third of the total number of cells at the density in question, according to the ranking in the halo tidal field strength, t_1. We then constrain the pressure - density relations for the three sub-samples jointly, assuming different sets of the parameters for each of the sub-samples The results are shown in Fig. 8.5, with the pressure - density relation for each sample normalized by the mean relation shown in Fig. 8.4. For reference, we also

list, in Table 7.1, the posterior model parameters and the predictions of the pressure at the mean density. It is clear that, for a given density, the electron pressure is higher in regions of stronger tidal field. As shown in Table 7.1, the pressures at the mean density obtained from the three samples, which measure the overall amplitudes of the corresponding pressure-density relations, are significantly different. However, the shapes of the relations are not very different among the sub-samples. This result is consistent with that obtained from gas simulations, as we will see in §7.4.4. The two breaking points in the relations obtained from the data arise because the relations for the sub-samples have different values of $\rho_{\rm m,0}$ from that for the total sample.

As mentioned in §7.2.4, we have tested a number of environmental quantities other than the tidal field as the second parameter that affects the thermal contents of the IGM. Among them, the distance of a grid-cell to the nearest massive halo is found to be nearly as a good indicator as the tidal field both from the simulation and from the application of our method to the observation. This is not surprising, as the local tidal field is strongly correlated with the presence of massive structures nearby.

7.4.4 Comparisons with simulations

In this subsection, we make comparisons of our results with gas simulations. Fig. 8.4 shows that the pressure - density relation obtained from the data roughly matches those given by the simulations. The result obtained from shuffling the grid-cells is also found to be very similar to the total average pressure from the simulations, as indicated by the horizontal dashed line in Fig. 8.4. In more detail, the observed pressure - density relation matches well with that given by EAGLE in both amplitude and shape, but Illustris significantly over-predicts the gas pressure in the intermediate density range from $\sim \overline{\rho}_{\rm m}$ to $\sim 10\,\overline{\rho}_{\rm m}$. As mentioned earlier, the difference between Illustris and EAGLE is likely caused by the different implementations of feedback processes adopted in the two simulations. Indeed, as shown in Vogelsberger et al.

(2014), the strong AGN feedback adopted in Illustris can heat a significant fraction of the IGM at large distances from over-dense regions, which may explain the high pressure seen in the intermediate density range. Our results, however, suggest that such strong feedback may not be favored by the data.

The dotted and dashed lines in Fig. 7.4 are the average temperatures obtained directly from Illustris and EAGLE, respectively. Thus, if the ionized baryon fraction is approximately equal to the universal baryon fraction on scales of $\sim 1\,h^{-1}$Mpc in the real universe, as is assumed in deriving the temperature - density relation from the data, then the average IGM temperature predicted by Illustris in the density range $(1-10)\overline{\rho}_{\rm m}$ is too high. The prediction of EAGLE is consistent with the observation, given the uncertainty in the data. This again shows that the data are already capable of providing interesting constraints on models of galaxy formation.

Finally, in Fig. 8.5, we compare observation and simulation results in their dependence on the strength of local tidal field. Here the simulation results are normalized by the corresponding mean relations shown in Fig. 8.4. The simulation results are consistent with the observational data in that the gas pressure at a given mass density is higher in regions of stronger tidal field. We have checked that the average ionized gas fraction is quite independent of the tidal strength in the simulations, and so the higher pressure in stronger tidal field is due to higher gas temperature rather than higher gas density. The dependence on the tidal field predicted by Illustris is much stronger than that by EAGLE. If shock heating by gravitational collapse has similar effects in both simulations, the difference in the tidal field dependence should then be a result of the different prescriptions of feedback used in the two simulations. The feedback effects on the IGM are expected to depend strongly on the local tidal field. For example, a cell with relatively low density but high t_1 must have some massive structures nearby to produce the strong tidal field. Such massive structures are also where strong stellar/AGN feedback is produced. In fact, as can be seen from

figure 4 in Vogelsberger et al. (2014), feedback effects on the IGM are clearly more important in the neighborhoods of more massive structures, where the tidal field is also stronger. It is interesting to note that, in both simulations, the strongest tidal field dependence occurs in regions with $\rho_{\rm m} \sim \overline{\rho}_{\rm m}$, although the signal is rather weak, indicating that the gas temperatures in such regions may be affected the most by feedback from nearby structures. The tidal field dependence obtained from the data is weaker than that predicted by Illustris but stronger than that by EAGLE. We have estimated the velocity dispersion of dark matter particles, σ, in individual grid cells in the simulations, and examined the average of σ^2 for cells of different $\rho_{\rm m}$ in regions of different tidal strengths. We found that the dependence of σ^2 on the tidal strength in the intermediate density range, $\rho_{\rm m} \sim (1-10)\overline{\rho}_{\rm m}$, is weaker than that of gas temperature, both in the observation and particularly in Illustris. If we take σ^2 as a measure of heating by gravitational collapse, then non-gravitational processes, such as stellar and AGN feedback, must have played an important role in heating the IGM in the intermediate density range.

7.5 Summary and discussion

In this paper we examine the thermal energy contents of the IGM over three orders of magnitude in both mass density and gas temperature using thermal Sunyaev-Zel'dovich effect (tSZE). Our results are based on *Planck* tSZE map and the cosmic density field, reconstructed in the SDSS DR7 volume and sampled on a grid of cubic cells of $(1h^{-1}{\rm Mpc})^3$, together with a matched filter technique employed to maximize the signal-to-noise.

Our results obtained by matching all the grid cells show that the pressure - density relation of the IGM is roughly a power law given by an adiabatic equation of state, with some indication of a steepening at densities higher than about 10 times the mean density of the universe. The result from shuffling the grid-cells shows a nearly

zero slope for the pressure-density relation, demonstrating that the relation obtained by our method indeed captures the thermal properties of the gas that produces the observed tSZE.

Using the simulation result that the ionized gas mass fractions within individual grid cells are about equal to the universal baryon fraction, we convert the pressure - density relation to a temperature - density relation. The result shows that the average temperature is about $10^4\,\mathrm{K}$ in regions of mean density, $\rho_\mathrm{m} \sim \overline{\rho}_\mathrm{m}$, increasing to about $10^5\,\mathrm{K}$ for $\rho_\mathrm{m} \sim 10\,\overline{\rho}_\mathrm{m}$, the typical density for cosmic filaments and sheets, and to $> 10^6\,\mathrm{K}$ for $\rho_\mathrm{m} \sim 100\,\overline{\rho}_\mathrm{m}$, the typical density of virialized dark matter halos.

The thermal energy content of the IGM is also found to be higher in regions of stronger tidal fields. By dividing grid cells into three equal-sized sub-samples according to the local tidal field strength, we find that the average gas temperature in the sub-sample of highest tidal field is a factor of 10 higher than that in the lowest tidal field sub-sample. Such an increase of temperature in intermediate density regions is stronger by a factor of two than that expected from the increase of average velocity dispersion of dark matter in simulations, suggesting that feedback from galaxy formation may be responsible for the increase in gas temperature.

We compare our results with those obtained from two hydrodynamic simulations, Illustris and EAGLE. While the simulations can reproduce the general trends observed in the observation, such as the increases of gas pressure and temperature with dark matter density and the strength of local tidal field, there are significant discrepancies between the two simulations, as well as between the simulations and our observational results. Within the uncertainties of the data, the predictions of EAGLE are consistent with the data. However, Illustris predicts significantly higher gas pressure and temperature in the intermediate density range, $\rho_\mathrm{m} \sim (1-10)\overline{\rho}_\mathrm{m}$, than both the observation and EAGLE. The dependence on tidal field strength predicted by Illustris is also too strong in comparison with the observational data and EAGLE.

We suspect that these differences are produced by the strong AGN feedback adopted in Illustris that can heat the IGM at large distances from massive structures.

Our results clearly demonstrate the promise of using SZE, combined with reconstructed density field, to study both the IGM and the galaxy formation processes that produce them. This approach is complementary to absorption line studies, in that it is not constrained by a limited number of lines of sight, and that corrections for metallicity and ionization effects are not needed to obtain the total gas mass. It also complements X-ray observations, in that it is more sensitive to the diffuse warm-hot gas that is expected to dominate the IGM. In the future, when high-resolution SZE data are available, the same approach as developed here can be used to study not only the detailed distribution and state of the IGM, but also to investigate how the IGM is related to and affected by galaxies and AGNs in the cosmic web.

CHAPTER 8

GAS CONTENTS OF THE CGM AND IGM AS CONSTRAINTS ON GALAXY FORMATION MODELS

8.1 Introduction

The current understanding of galaxy formation is significantly limited due to the complexity of the physical processes responsible for the interchange of mass and energy between galaxies and their surroundings (e.g. Mo et al. 2010). Gas and dark matter falls into the potential wells of haloes, collapsed objects formed via gravity from the initial matter fluctuations present in the early Universe. As gas passes the though the virial shock it is heated, but it eventually cools to form stars. Observations, however, have accumulated evidence that there is an outflow 'wind' that heats and returns the gas from galaxies to the circumgalactic medium (CGM) and intergalactic medium (IGM), possibly driven by the stellar and AGN feedback (e.g. Steidel et al. 2010, Jones et al. 2012, Martin et al. 2012, Newman et al. 2012, Rubin et al. 2014, Heckman et al. 2015, Chisholm et al. 2016, Tumlinson et al. 2017). The warm-hot gas returned to the surrounding medium is mixed with the cold gas newly accreted on to haloes from the cosmic web. It is still an open question whether the wind velocities are fast enough to return the gas from haloes to the IGM, or whether a substantial portion of the gas recycles back on to galaxies (Stocke et al. 2013, Werk et al. 2014, 2016, Borthakur et al. 2016, Prochaska et al. 2017, Rudie et al. 2019). Recent 'zoom-in' simulations predict that the dominate source of gas in Milky Way-like haloes is the cosmic accretion on to haloes in mass, and even the gas in the outflow returns back to haloes and is recycled. Therefore, the baryon fraction

in such haloes is close to the cosmic mean fraction (e.g. Angles-Alcazar et al. 2017, Hafen et al. 2018). Some cosmological simulations, however, found a large fraction of baryons 'missing' in haloes, completely ejected out of haloes by the feedback (e.g. Vogelsberger et al. 2014, Schaye et al. 2015). As a result, a systematic investigation of the CGM and IGM properties is required to understand the galaxy evolution.

With the advent of large CMB surveys, the Sunyaev-Zel'dovich effect (SZE; Sunyaev & Zel'dovich 1972) provides a promising way to probe the CGM and IGM gas properties. The scattering of the CMB photons with the free electrons on their paths from the last scattering surface to us changes the CMB spectrum, which is called the Sunyaev-Zel'dovich effect. The SZE produced by the thermal motions of electrons is referred to as the thermal SZE (tSZE), while that produced by the bulk motions of electrons is called the kinetic SZE (kSZE). Recent studies have demonstrated that the tSZE and kSZE from observations can be used to constrain the properties of the ionized gas in the CGM and IGM (Hand et al. 2012, Planck Collaboration XI 2013, Van Waerbeke et al. 2014, Hojjati et al. 2015, Ma et al. 2015, Hill et al. 2016, 2018, Lim et al. 2017b, 2018a, 2018b, de Graaff et al. 2019, Tanimura et al. 2019). Using the SZE to trace the gas has advantages that it can probe relatively low-density regions such as outskirt of haloes and the IGM compared to X-ray observations, and that the derived gas properties do not depend on the gas metallicity and ionization states unlike absorption line studies toward quasars.

Cosmological hydrodynamic simulations offer a theoretical framework to study the physical processes involved in the galaxy evolution, and the resulting properties of the CGM and IGM (e.g. Keres et al. 2005, 2009; Faucher-Giguere et al. 2011; Somerville & Dave 2015; Oppenheimer 2018). Large simulations that trace the evolution of the matter in a box with a few hundred Mpc scale (e.g. Vogelsberger et al. 2013, 2014; Crain et al. 2015; Schaye et al. 2015; Dolag et al. 2016; McCarthy et al. 2017; Pillepich et al. 2018), and high-resolution 'zoom-in' simulations that focus on

the evolution of individual haloes (e.g. Hopkins et al. 2014, 2018; Muratov et al. 2015, 2017; van de Voort et al. 2016, 2019; Angles-Alcazar et al. 2017; Hafen et al. 2018), incorporate the relevant physics to reproduce the properties of galaxies and gas from observations reasonably well (e.g. Faucher-Giguere et al. 2010, 2015, 2016; Faucher-Giguere & Keres 2011; Hummels et al. 2013; Nelson et al. 2018). However, simulations that are known to reproduce a few key observations have employed widely different physical models. In order to break the degeneracy between the models, we need to test the models with more detailed observational constraints.

In this paper, using the *Planck* observations and four large hydrodynamic simulations, we demonstrate that the SZE is a useful tool to investigate the gas properties of haloes and large-scale structures, and provides stringent constraints for simulation models. We use the cross-correlations between the SZE signals, dark matter haloes and large-scale environments as constraints. Our analysis closely follows that presented in Lim et al. (2017b, 2018a, 2018b). We employ the simultaneous matched filter approach (Lim et al. 2017b) to increase signal-to-noise of the signal as well as to disentangle the projection effects of haloes along same line-of-sight (LOS). With extensive tests, (Lim et al. 2017b) found that the results are robust against systematic effects such as the residual foreground/background fluctuations, the uncertainties in the reconstructed velocity field, the beam size of the *Planck* survey, and residual dust emission from galaxies. We assume the *Planck* cosmology (Planck Collaboration XIII 2016) throughout the paper, and scale the results accordingly.

This paper is organized as follows. We present the observational data and the simulations used for our analysis in Sec 8.2. The method used in our analysis is described in Sec 8.3. We report our results in Sec 8.4. Finally, we discuss our results and summarize our conclusions in Sec 8.5.

8.2 Data

8.2.1 The *Planck* CMB map

The *Planck* (Tauber et al. 2010; Planck Collaboration I 2011) measured the all-sky CMB anisotropy in nine frequency bands from 30 to 857 GHz with angular resolution ranging between 5 and 31 arcmin. For our analysis of the tSZE, we make use of the *Planck* NILC (Needlet Independent Linear Combination; Remazeilles et al. 2011) all-sky tSZ Compton parameter map (Planck Collaboration XXII 2016), also referred to as the NILC y-map, from the *Planck* 2015 data release. The differential temperature of the CMB spectrum caused by the tSZE is characterized by a dimensionless parameter,

$$\left(\frac{\Delta T}{T_{\rm CMB}}\right)_{\rm tSZE} = g(x)y \equiv g(x)\frac{\sigma_{\rm T}}{m_{\rm e}c^2}\int P_{\rm e}{\rm d}l, \tag{8.1}$$

where $T_{\rm CMB} = 2.7255\,{\rm K}$, y is the Compton parameter, $g(x) = x\coth(x/2) - 4$ is the conversion factor at a given $x \equiv h\nu/(k_{\rm B}T_{\rm CMB})$, $\sigma_{\rm T}$ is the Thompson cross-section, $m_{\rm e}$ is the electron rest-mass, c is the speed of light, $P_{\rm e} = n_{\rm e}k_{\rm B}T_{\rm e}$ is the electron pressure with $n_{\rm e}$ and $T_{\rm e}$ the number density and temperature, respectively, of the free electron, and finally dl is the path length along given line-of-sight (LOS). The final map is constructed by weighting the different frequency maps to remove components other than the tSZE signal, such as the primary CMB fluctuations and foreground sources. As a test, we also use the *Planck* MILCA (Modified Internal Linear Combination Algorithm; Hurier et al. 2013), which is known to have a different degree of dust contamination than the NILC, and found no significant change in our results.

The temperature change in the CMB spectrum by the kSZE is given by

$$\left(\frac{\Delta T}{T_{\rm CMB}}\right)_{\rm kSZE} = -\frac{\sigma_{\rm T}}{c}\int n_{\rm e}(\boldsymbol{v}\cdot\hat{\boldsymbol{r}}){\rm d}l, \tag{8.2}$$

where $\boldsymbol{v}$ is the velocity of the free electrons in bulk motion, and $\hat{\boldsymbol{r}}$ is the unit vector along the LOS. For our analysis of the kSZE, we use the *Planck* CMB temperature

maps at 100, 143, and 217 GHz frequency bands provided in the *Planck* 2015 data release. From the three frequency maps, we subtract the tSZE using the y-map and equation 8.1, with the conversion factor $g(x)T_{\rm CMB} = -4.031$, -2.785, and 0.187 integrated over the 100, 143, and 217 GHz bands, respectively. Again, we found no significant change in our results between when using the NILC and MILCA y-maps for the tSZE subtraction.

For both the tSZE and kSZE analysis, we minimize Galactic contamination by masking out the brightest 40% of the sky using the masks provided in the data release. We also mask radio and infrared point sources using the relevant masks provided in the same data release to limit the contamination.

8.2.2 Hydrodynamic simulations

For our analysis, we use a number of state-of-the-art cosmological gas simulations including Illustris, IllustrisTNG, EAGLE, and Magneticum. These simulations adopt different numerical techniques, cosmological models, and different implementations of physical processes, to trace the evolution of the simulated Universe. All these simulations identify haloes using a friends-of-friends (FoF; Huchra & Geller 1982, Davis et al. 1985) algorithm.

8.2.2.1 Illustris

The first simulation is Illustris (Vogelsberger et al. 2014a, Vogeslberger et al. 2014b, Genel et al. 2014, Sijacki et al. 2015), which was run with the moving-mesh code AREPO (Springel 2010), assuming WMAP9 cosmology with $h = 0.704$, $\Omega_m = 0.273$, and $\Omega_\Lambda = 0.727$ (Hinshaw et al. 2013). The traced components include gas cells, dark matter particles, stars and stellar wind particles, and super-massive black holes. Sub-grid models are employed for the physical processes such as cooling (Katz et al. 1996, Wiersma et al. 2009), star formation (Springel & Hernquist 2003) with a Chabrier (2003) initial mass function, stellar feedback (Vogelsberger et

al. 2013, Torrey et al. 2014), and AGN feedback (Springel et al. 2005, Sijacki et al. 2007). For the detailed implementation, we refer the reader to Vogelsberger et al. (2013). The free parameters in the models were constrained using observations as well as simulations of higher resolutions. In this paper, we use Illustris-1, their fiducial run, that has a box size of $L = 75\,h^{-1}\,\mathrm{Mpc}$ and contains $2 \times (1820)^3$ gas and dark matter particles. The target baryon mass and dark matter particle mass are $m_{\mathrm{baryon}} = 1.6 \times 10^6 \mathrm{M}_\odot$ and $m_{\mathrm{DM}} = 6.3 \times 10^6 \mathrm{M}_\odot$, respectively. The gravitational softening lengths for the dark matter particles and gas cells particles are 1.4 kpc and 0.7 ckpc, respectively.

8.2.2.2 IllustrisTNG

The IllustrisTNG (TNG; Marinacci et al. 2018, Naiman et al. 2018, Nelson et al. 2018, Pillepich et al. 2018, Springel 2018), the successor of the Illustris, is a hydrodynamic simulation run with AREPO code assuming the cosmological model given by Planck Collaboration XIII (2016) with $\sigma_8 = 0.816$, $h = 0.677$, $\Omega_m = 0.309$, and $\Omega_b = 0.0486$. It is sampled with $(2500)^3$ dark matter particles and $(2500)^3$ gas particles in a periodic box of $(205\,h^{-1}\,\mathrm{Mpc})^3$. The physical models used in TNG build upon those developed for the original Illustris, but with a lot of improvements including the prescriptions for magnetic field (Pakmor et al. 2011) and the black-hole-driven feedback (Weinberger et al. 2017). One of the major changes in IllustrisTNG is a new black-hole-driven kinetic feedback at low accretion rates (radio-mode), compared to the original Illustris where thermal energy is injected into surrounding gas in the form of 'bubbles'. The target baryon mass and dark matter particle mass are $m_{\mathrm{baryon}} = 1.1 \times 10^7 \mathrm{M}_\odot$ and $m_{\mathrm{DM}} = 5.9 \times 10^7 \mathrm{M}_\odot$, respectively. The $z = 0$ Plummer equivalent gravitational softening of the collisionless component, and the minimum comoving value of the adaptive gas gravitational softening are 1.5 kpc and 0.37 ckpc, respectively.

8.2.2.3 EAGLE

The Evolution and Assembly of GaLaxies and their Environments (EAGLE; Schaye et al. 2015, Crain et al. 2015, McAlpine et al. 2015), run with a modified version of GADGET-3 Smoothed Particle Hydrodynamics (SPH) code (Springel 2005) tracks the evolution of gas, stars, dark matter, and massive black holes in the simulated Universe, by implementing sub-grid models for cooling (Wiersma et al. 2009), star formation (Schaye & Dalla Vecchia 2008), stellar and AGN feedbacks (Booth & Schaye 2009, Rosas-Guevara et al. 2016). The models are parameterized and the parameters are tuned to match observations including the stellar mass function and stellar mass-black hole mass relation at $z \sim 0$. The simulation assumes the *Planck* cosmology (Planck Collaboration XVI 2014). Throughout this paper, we use the simulation run of the largest box of $(100\,\mathrm{Mpc})^3$ sampled by $2\times(1504)^3$ particles. The initial baryonic particle mass and dark matter particle mass are $1.8\times10^6\mathrm{M}_\odot$ and $9.7\times10^6\mathrm{M}_\odot$, respectively. The comoving Plummer-equivalent gravitational softening and the maximum physical softening length are roughly 2.7 kpc and 0.70 ckpc, respectively.

8.2.2.4 Magneticum

The Magneticum simulations (e.g. Dolag et al. 2016) are a set of cosmological hydrodynamic simulations of various volumes and resolutions, performed with an improved version of GADGET-3. The simulations adopted a WMAP7 flat ΛCDM cosmology with $\sigma_8 = 0.809$, $h = 0.704$, $\Omega_m = 0.272$, and $\Omega_b = 0.0456$ (Komatsu et al. 2011). The simulations include a variety of physical processes such as cooling and star formation (Springel & Hernquist 2003), black holes and AGN feedback (Fabjan et al. 2010, Hirschmann et al. 2014), and thermal conduction (Dolag et al. 2004). The results presented in this paper are produced from a particular run that has a box size of $L = 352\,h^{-1}\,\mathrm{Mpc}$ and is sampled by $2\times(1584)^3$ particles. The dark matter particle mass and gas particle mass are $6.9\times10^8\mathrm{M}_\odot$ and $1.4\times10^8\mathrm{M}_\odot$, respectively.

The softening lengths are 3.8 kpc commonly for the dark matter particle and gas particle.

8.3 Methods

8.3.1 The matched filter technique

Extracting the SZE signal is not a trivial task because the signal is very weak; for massive clusters of galaxies, the kSZE amplitude is about an order of magnitude lower than that of the tSZE and two orders of magnitude smaller than other sources such as the primary CMB fluctuation, Galactic foreground, and cosmic infrared background. As such, we employ the matched filter (MF) technique (Haehnelt & Tegmark 1996, Herranz et al. 2002, Melin et al. 2005, 2006) in our analysis, which is designed to maximize the signal-to-noise and minimize source contamination by imposing prior knowledge of the signals and the noise power spectra. In the MF technique, the Fourier transform of the filter that optimizes the signal-to-noise is given by

$$\begin{aligned} \hat{F}_I(\boldsymbol{k}) &= \left[\int \frac{\mathrm{d}^2 k'}{(2\pi)^2} |\hat{\tau}(\boldsymbol{k'})|^2 \hat{\boldsymbol{B}}^*(\boldsymbol{k'}) \cdot \boldsymbol{P}^{-1}(k') \cdot \hat{\boldsymbol{B}}(\boldsymbol{k'})\right]^{-1} \\ &\quad \times \hat{\tau}(\boldsymbol{k})\, j_{\nu,I} \left\{\boldsymbol{P}^{-1}(k) \cdot \hat{\boldsymbol{B}}(\boldsymbol{k})\right\}_I \end{aligned} \tag{8.3}$$

where $\hat{F}_I(\boldsymbol{k})$ is the Fourier transform of the filter for each of the three frequencies, 'I', $\hat{\tau}(\boldsymbol{k})$ is, in our application, the Fourier transform of the projected electron pressure for tSZE analysis, and of the electron column density for kSZE analysis, the elements of $\hat{\boldsymbol{B}}(\boldsymbol{k})$ are the Fourier transform of the Gaussian beam function that mimics the convolution of the *Planck* observation in each frequency band, and $\boldsymbol{P}(k)$ is the noise power spectra. For the tSZE analysis, because the y-maps are already cleaned of the primary CMB fluctuation, $P_{IJ}(k) = P_{\mathrm{noise},I}\delta_{IJ}$ where $P_{\mathrm{noise},I}$ is the power spectra of the *Planck* noise map in frequency band 'I', as provided in the date release. For the kSZE analysis, $P_{IJ}(k) = P_{\mathrm{CMB}}\, \hat{B}_I \hat{B}_J^* + P_{\mathrm{noise},I}\delta_{IJ}$ where P_{CMB} is the CMB spectrum.

The $j_{\nu,I}$ is a frequency-dependent conversion factor between the differential temperature of the CMB and the filter; $\boldsymbol{j}_\nu = g(\boldsymbol{x})$ for tSZE, and is frequency-independent for kSZE.

8.3.2 Extracting the tSZE flux of haloes

To associate the SZE signals with haloes, we use the catalogs given in Lim et al. (2017a). The catalogs were constructed for four redshift surveys including 2MRS, 6dF, SDSS, and 2dF, by applying an adaptive halo-based group finder that was initially developed by Yang et al. (2005, 2007) and improved by Lim et al. (2017a). The group finder determines the membership and mass of groups using stellar mass or luminosity of centrals and satellites as a halo mass proxy, and by using, in turn, the resulting halo properties to update the galaxy membership, iteratively until membership converges. Using realistic mock catalogs constructed based on simulations and empirical models, they showed that the halo masses determined by the group finder is unbiased to the true halo masses with a typical scatter of $0.2 - 0.3$ dex. For our analysis, we use the combined catalogs of the 2MRS, 6dF, and SDSS. The resulting catalog contains $471,696$ haloes with $\log M_{500}/\mathrm{M}_\odot \geq 12$, of which $3,851$ haloes are $\log M_{500}/\mathrm{M}_\odot \geq 14$ and $240,747$ haloes are $12 \leq \log M_{500}/\mathrm{M}_\odot \leq 13$. We define the halo by a radius R_{500}, within which the mean density is 500 times the critical density at a given redshift, with M_{500}, a mass enclosed within R_{500}.

We then cross-correlate the *Planck* y-map with the haloes in the catalog, using information such as the position, redshift, and mass, to extract the tSZE signal associated with the haloes. For the extraction of the flux, we refer the reader to Lim et al. (2017b, 2018a) (hereafter L17b and L18a, respectively) for details, and here we provide only a brief summary of it. Note that the method used in L18a changed significantly from L17b, and that we follow the new method presented in L17b for the extraction of the flux in this paper. The first step is the construction of model maps.

We adopt the universal pressure profile (UPP) from Arnaud et al. (2010), which was derived from X-ray observations for $r \leq R_{500}$ combined with simulations for larger radii. Then we employ a model that is fully specified by the amplitude of the UPP at $\log M_{500}/\mathrm{M}_{\odot} = 12.3, 12.5, 12.7, ...,$ and 14.7. Haloes of any mass are assigned the filter amplitude by linear interpolation in the log-log space. The model map is constructed by placing at the position of each galaxy groups the model profiles with the shape and amplitude appropriate for each halo's size, mass, and redshift. Then the map is convolved with the beam function, and the instrument noise is added using $P_{\rm noise}$ to make the model map realistic. Then, at the center of each halo 'i' in each frequency band 'I', we apply the matched filter, $\{F_i\}_I$,

$$\begin{aligned} \{A_{D,i}\}_I &= \int \{F_i(\boldsymbol{\theta})\}_I \{D_i(\boldsymbol{\theta})\}_I \mathrm{d}^2\theta \\ \{A_{M,i}\}_I &= \int \{F_i(\boldsymbol{\theta})\}_I \{M_i(\boldsymbol{\theta})\}_I \mathrm{d}^2\theta \,, \end{aligned} \tag{8.4}$$

where $\boldsymbol{\theta}$ is the projected position of a given pixel relative to the center of the group, and $\{D_i(\boldsymbol{\theta})\}_I$ and $\{M_i(\boldsymbol{\theta})\}_I$ are the data map and the model map around the group, respectively. With the set of $\{A_i\}_I$'s from the data and the model map, we constrain the filter amplitudes by minimizing the χ^2,

$$\chi^2 = \sum_i \frac{|\boldsymbol{A}_{D,i} - \boldsymbol{A}_{M,i}(\Theta)|^2}{\sigma_i^2} \tag{8.5}$$

with the uncertainty,

$$\sigma_i = \left[\int \frac{\mathrm{d}^2 k'}{(2\pi)^2} |\hat{\tau}_i(\boldsymbol{k'})|^2 \hat{\boldsymbol{B}}^*(\boldsymbol{k'}) \cdot \boldsymbol{P}^{-1}(k') \cdot \hat{\boldsymbol{B}}(\boldsymbol{k'}) \right]^{-1/2} \tag{8.6}$$

as given in Haehnelt & Tegmark (1996) and Melin et al. (2006). To explore the parameter space, we use the MULTINEST (Feroz et al. 2009) that implements the nested sampling developed in Skilling et al. (2006).

8.3.3 Extracting the kSZE flux of haloes

For the kSZE flux, we adopt the result of L17b, which was obtained from a very similar analysis to that for the tSZE described above except for some differences in the details. Here we brifely describe the method and refer the reader to the original paper for details. The kSZE also depends on the radial component of the peculiar velocity (see equation 8.2). Therefore, we make use of the velocity field reconstructed by Wang et al. (2012). Tests using mock catalogs show that the velocity field reconstructed by this method is unbiased with a typical dispersion of $\sim 90\,\mathrm{km\,s^{-1}}$. The reconstructed velocity field is available only for a subset of haloes with $z \leq 0.12$. We also use the haloes from the catalog by Yang et al. (2007) as our sample for the kSZE analysis. For the profile shape, we adopt a β-profile,

$$n_{\rm e}(r) = n_{\rm e,0}[1 + (r/r_{\rm c})^2]^{-3\beta/2}, \tag{8.7}$$

where $r_{\rm c} = r_{\rm vir}/c$ is the core radius of a group with concentration c, and $\beta = 0.86$ is the best-fit value obtained from South Pole Telescope (SPT) cluster profiles (Plagge et al. 2010). Due to the lower signal-to-noise of the kSZE than the tSZE, the model is specified by the amplitudes at $\log M_{500}/\mathrm{M}_\odot = 12.3, 12.7, 13.1, 13.5, 13.9$, and 14.3. The model maps are constructed by adding the CMB fluctuation, convolved with the beam function and adding the instrument noise at each of the three frequencies 100, 143, and 217 GHz. For the kSZE, $\{D_i(\boldsymbol{\theta})\}_I$ and $\{M_i(\boldsymbol{\theta})\}_I$ in equation 8.4 are the differential temperatures.

8.3.4 The pressure - density relation of IGM

The impact of feedback is believed not to be confined within haloes, but also leaves its imprint in the intergalactic medium (IGM). For a better understanding of feedback, it is helpful to investigate the SZE signal of the IGM. Here we adopt the analysis and result of Lim et al. (2018b) as follows.

Using the *Planck* y-map, we aim to constrain the average relation between the thermal energy and matter density of the IGM. To do this, we use another set of data, which is the reconstructed cosmic density field in the Sloan Digital Sky Survey Data Release 7 (SDSS DR7; Abazajian et al. 2009) volume given in Wang et al. (2016). The reconstruction uses the DR7 haloes from Yang et al. (2007) as a proxy of the density field. We use the density field smoothed on $1\,h^{-1}$Mpc. To each of those $1\,h^{-1}$Mpc grid-cells, we assign a value of electron pressure by assuming a double power-law relation between the pressure and the density field,

$$P_{\rm e} = \begin{cases} A \times (\rho_{\rm m}/\rho_{\rm m,0})^{\alpha_1}, & \text{if } \rho_{\rm m} \leq \rho_{\rm m,0} \\ A \times (\rho_{\rm m}/\rho_{\rm m,0})^{\alpha_2}, & \text{if } \rho_{\rm m} > \rho_{\rm m,0}. \end{cases} \tag{8.8}$$

The resulting pressure field is smoothed with a Gaussian kernel with the radius of $1\,h^{-1}$Mpc, and is integrated along each LOSs to obtain the model map of y. The shape of the y-profile in each pixel is used as a filter shape for the MF. Similar to the SZE flux analysis, the model parameters in equation 8.8 are constrained by minimizing the χ^2 over all pixels. We assume a constant background from gas outside the SDSS DR7 volume, which is treated as another free parameter to be constrained.

We also investigate the dependence of the pressure - density relation on the large-scale tidal field. This is motivated by the findings based on simulations that, among various quantities such as stellar mass, star formation rate, black hole mass, velocity dispersion, etc, the large-scale tidal field is the most crucial second parameter that affects the gas pressure (Lim et al. 2018b). Following Wang et al. (2011), the halo-based tidal field for each grid-cell is estimated as the halo tidal force exerted on the surface of a sphere along a direction $\boldsymbol{t}$, normalized by the self-gravity of the sphere,

$$f(\boldsymbol{t}) = \frac{\sum_i GM_i R_{\rm g}(1+3\cos 2\theta_i)/r_i^3}{2GM_{\rm g}/R_{\rm g}^2} \tag{8.9}$$

where the summation is over all the haloes, M_i is the mass of halo i, $R_{\rm g} = 0.5h^{-1}$Mpc is the radius of the sphere that approximates the grid-cell, $M_{\rm g}$ is the mass enclosed within the grid-cell in question, r_i is the separation between the center of the grid-cell and the halo 'i', and θ_i is the angle between $\boldsymbol{t}$ and $\boldsymbol{r}_i$. The tidal field satisfies $t_1 + t_2 + t_3 = 0$ where t_1, t_2, and t_3 ($t_1 \geq t_2 \geq t_3$) are the eigenvalues of the tidal field tensor. Wang et al. (2011) showed that t_1 represents well the magnitude of the tidal field. We thus use t_1 to characterize the tidal field strength of the grid-cells.

8.4 Results

8.4.1 The thermal energy of halo gas

Using the assumed profile shape, i.e. UPP, and its amplitudes determined from the χ^2 minimization, we can infer the integrated tSZE flux within R_{500}, Y_{500}, defined by,

$$d_{\rm A}(z)^2 Y_{500} \equiv \frac{\sigma_{\rm T}}{m_{\rm e}c^2} \int_{R_{500}} P_{\rm e}\, {\rm d}V, \tag{8.10}$$

where $d_{\rm A}(z)$ is the angular diameter distance to a group at given redshift. At a fixed halo mass, Y_{500} evolves with redshift as $E^{2/3}(z)$. Thus we follow conventions to define a new, redshift-independent quantity,

$$\tilde{Y}_{500} \equiv Y_{500} E^{-2/3}(z) \Big(\frac{d_{\rm A}(z)}{500{\rm Mpc}} \Big)^2, \tag{8.11}$$

which is expected to be a function of only halo mass scaled to $z = 0$, provided that the intrinsic tSZE flux is indeed a function of only mass. The $\tilde{Y}_{500}$ thus inferred from the medians and 68 percentile ranges of the posterior distribution of the model parameters are shown by the yellow triangles and the error bars, respectively, in Fig. 8.1. L18a demonstrated that the results are indeed dominated by the tSZE signals rather than

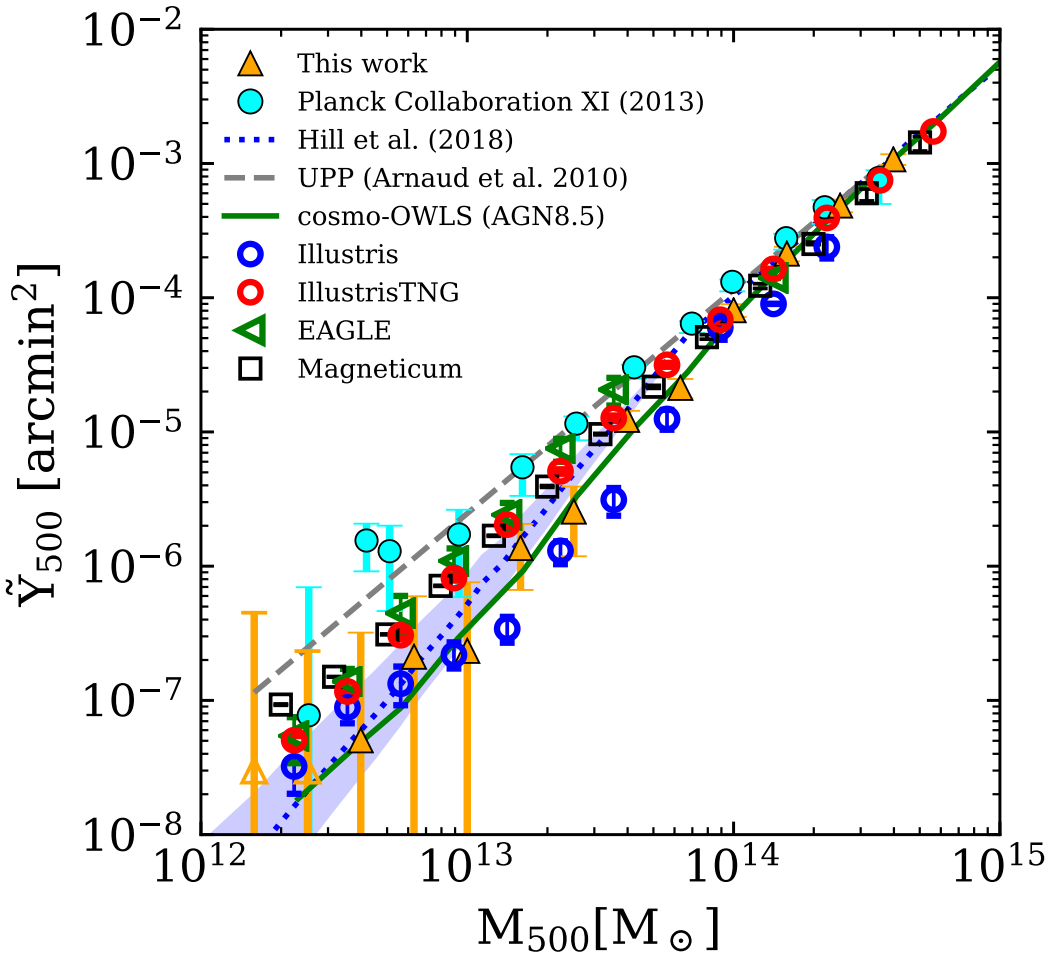

Figure 8.1. Comparison of the tSZE flux from gas within R_{500} of haloes, $\tilde{Y}_{500}$, between several observations (this work (triangle), Planck Collaboration XI (2013) (cyan dot), Hill et al. (2018) (dotted), Arnaud et al. (2010) (dashed)), and simulations (cosmo-OWLS; Le Brun et al. (2015) (green solid), Illustris (blue dot), IllustrisTNG (red dot), EAGLE (triangle), Magneticum (square)). The error bars for the tSZE result obtained with the presented method represent the 68 percentile ranges of the posterior distribution. The error bars for the simulations are obtained from 200 bootstrap samples. The unfilled symbols represent the values below zero.

the residual backgrounds, and are robust against truncation of the filters at different radii, mass incompleteness of the halo catalog, moderate variations in the filter shape, and fluctuations of the background. They suspect that the results are insensitive to the different profile shapes assumed due to the relatively large beam size of the *Planck* survey. They also showed that the projection effects of larger haloes along same LOSs are significant for the signals from low-mass haloes.

We compare our results with other observational studies by Planck Collaboration XI (2013) (PCXI hereafter; cyan dots) and Hill et al. (2018) (blue dotted line). Using the *Planck* temperature maps, and locally brightest galaxies as a tracer of haloes, PCXI found that the tSZE flux follows the self-similar case (dashed line) that the hot gas mass fraction relative to halo mass is independent of halo mass. This is in conflict with predictions from simulations and X-ray observations that a fraction of gas is expelled out of the potential well in low-mass systems. L18a suspect that this is due to the fact that PCXI did not fully take into account the projection effects of other haloes. PCXI tested both aperture photometry and matched filter to extract the flux, but in both cases they assumed flat local backgrounds to subtract. However, as demonstrated in Vikram et al. (2017), the two-halo terms dominate the tSZE signals around haloes of $M_{200} \leq 10^{13-13.5}\,h^{-1}\mathrm{M}_{\odot}$, thus even a very small deviation from flat backgrounds can significantly change the estimation of the tSZE flux for those haloes. L18a confirmed that they recover the PCXI results when assuming flat backgrounds, implying that the local background indeed changes with distance from halo centers due to the clustering of haloes. Taking into account the projection effects based on Vikram et al. (2017), Hill et al. (2018) also found some evidence for a deviation of the relation from the self-similar case. As seen in Fig. 8.1, their results are consistent with our results even for low-mass systems. Hill et al. used the *Planck* map and the group catalog by Yang et al. (2007), which is very similar to the data set that we use here.

Fig. 8.1 show that all the simulations considered in our analysis predict a certain degree of deviations from the self-similar case, reflecting the effect of feedbacks in low-mass systems. The error bars are obtained from 200 bootstrap samples. The level of the deviation, however, differs significantly among different simulations, indicating that the gas contents in haloes are significantly affected by models adopted to address physical processes. In other words, the relation can be used to constrain the physics of the feedback process. For example, Illustris, which is known to have a model with much stronger AGN feedback implemented than the other simulations, predicts a much lower electron pressure than the other simulations, and more than an order of magnitude lower gas mass than the self-similar case for haloes with $M_{500} \sim 10^{13}\,\mathrm{M}_{\odot}$. Those haloes are the systems that are believed to be most affected by the AGN feedback. The prediction from Illustris is vastly different even from that by TNG. Since TNG differs significantly in the model for AGN feedback relative to Illustris while retaining from Illustris many of the simulation techniques and sub-grid models, the difference in the tSZE flux is dominated by the AGN feedback. The differences between the TNG, EAGLE, and Magneticum are insignificant despite of differences in details of the models for the physical processes as well as the resolution of the simulations.

8.4.2 The mass and temperature of halo gas

Using the amplitudes of the β-profile, i.e. $n_{\mathrm{e},0}$ in equation 8.7, determined as described in section 8.3.3, we infer the total hot gas mass within R_{500},

$$M_{\mathrm{gas}} = N_{\mathrm{e},500} \cdot \frac{2m_{\mathrm{p}}}{1 + f_{\mathrm{H}}}, \tag{8.12}$$

where $N_{\mathrm{e},500} = \int_{R_{500}} n_{\mathrm{e}} \mathrm{d}V$, $f_{\mathrm{H}} = 0.76$ is the hydrogen mass fraction, and m_{p} is the proton mass. Fig. 8.2 shows the averages (the yellow circles) and the dispersions (the error bars) of the gas mass fraction from the eight samples described in L17b. The

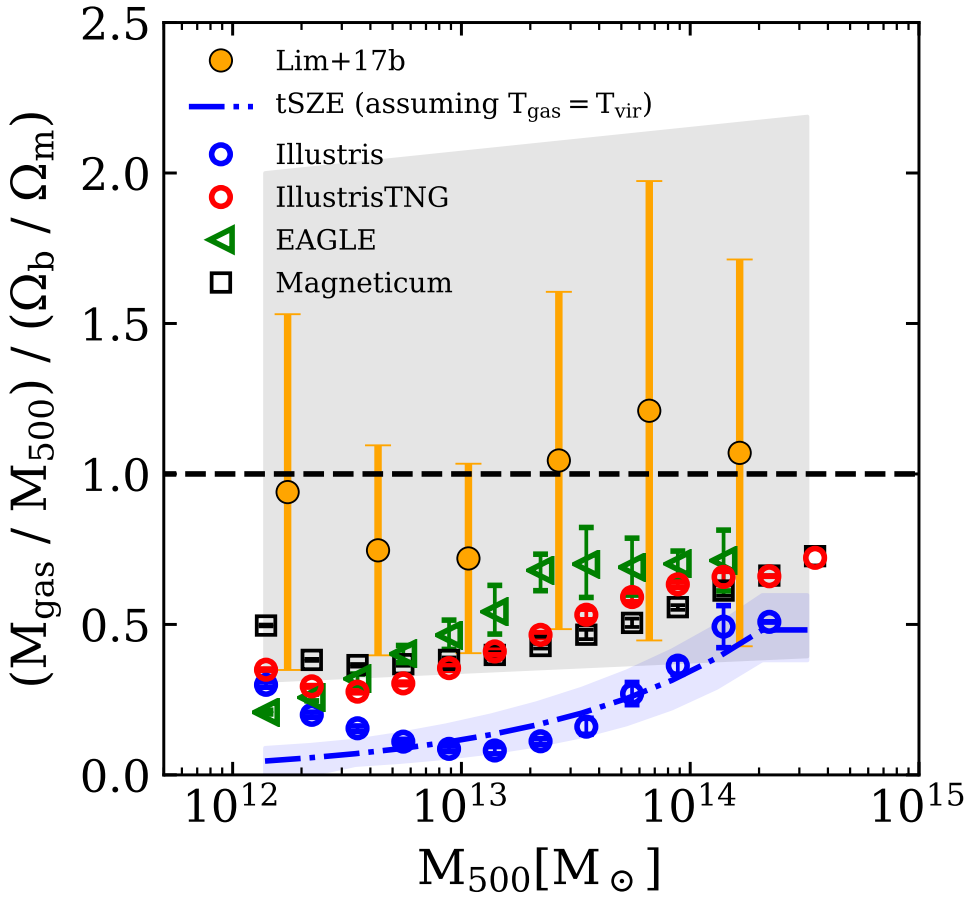

Figure 8.2. The gas mass fraction within haloes inferred from the kSZE analysis (filled circle), compared with that from the simulations including the Illustris (blue dot), IllustrisTNG (red dot), EAGLE (triangle), and Magneticum (square). The error bars for the observational result represent the dispersion of the estimate among the eight samples as described in Lim et al. (2017b). The shaded band spans the same dispersion among the eight samples but obtained with a power-low model. The error bars for the simulations are obtained from 200 bootstrap samples. The dashed line shows the cosmic baryon fraction. The dot-dashed line indicates the gas mass fraction inferred from the tSZE analysis by assuming $T_{gas} = T_{vir}$, with the band showing the uncertainty.

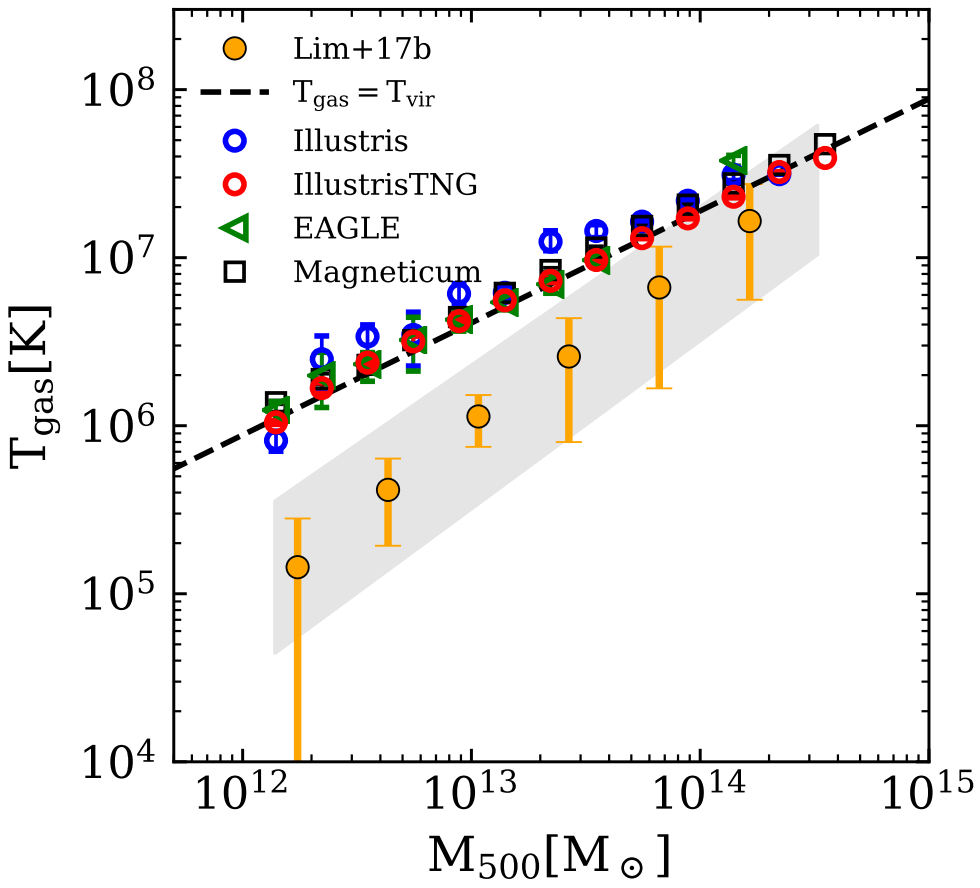

Figure 8.3. The effective gas temperature within haloes obtained by dividing the tSZE flux with the kSZE flux (filled circle), compared with that from the simulations including the Illustris (blue dot), IllustrisTNG (red dot), EAGLE (triangle), and Magneticum (square). The error bars for the observational result represent the dispersion of the estimate among the eight samples as described in Lim et al. (2017b). The shaded band spans the same dispersion among the eight samples but obtained with a power-low model. The error bars for the simulations are obtained from 200 bootstrap samples. The dashed line indicates the virial temperature, $T_{\rm vir} = \mu m_{\rm p} G M_{500}/2k_{\rm B}R_{500}$, with $\mu = 0.59$ the mean molecular weight.

shaded band spans the dispersion among the eight samples, inferred from a power-law model (see L17b for details). As seen in the figure, the inferred gas fraction is consistent with the cosmic baryon fraction (black dashed), thus there is no missing baryons in haloes. L17b showed that the results are robust against the uncertainties in the reconstructed velocity field, residual background/foreground fluctuations, the large beam size of *Planck* and residual tSZE signal and dust emission from galaxy groups. Using simulations, L17b also quantified the contamination by the projection of gas outside haloes along LOS. They found that the projection effect is, on average, capped at about 20% of the total flux estimated, but is sensitive to the definition of haloes, the fluctuation of matter distribution in large-scale structures, and the large-scale velocity field coherence predicted from the linear theory. The projection effect is also expected to strongly depend on the baryonic processes such as the stellar and AGN feedback, with the higher projection effect from the stronger feedback.

We compare the gas fraction from the kSZE analysis with that inferred from the tSZE results by assuming that the gas is at the virial temperature, $T_{\rm vir} = \mu m_{\rm p} G M_{500}/2k_{\rm B}R_{500}$ where $\mu = 0.59$ is the mean molecular weight, which is shown by the blue dot-dashed line with the band showing the errors in the estimate. The inferred gas fraction from the tSZE is significantly lower particularly in low-mass systems than that from the kSZE. This implies that the effective temperature of the gas is much lower than the virial temperature.

We also compare the results with the predictions from the simulations. The error bars for the simulations are obtained from 200 bootstrap samples. Overall, the simulations predict lower gas fractions than the observational results across the whole mass range considered here, up to by a factor of ~ 4 at the low-mass end of $M_{500} \sim 10^{12}{\rm M}_{\odot}$. That also means that the simulations predict $20-40\%$ lower baryon fraction in low-mass systems relative to the cosmic fraction. Illustris has a much higher fraction of gas expelled out of the halo potential than the other simulations possibly due to the

stronger AGN feedback implemented. Recent 'zoom-in' simulations focusing on individual haloes with a much higher resolution, however, find almost all baryons expected by cosmology within the virial radius even for lower-mass haloes of $M_h \sim 10^{12}\,\mathrm{M}_\odot$, indicating that the galactic wind is not effective in expelling gas completely out of haloes (e.g. Angles-Alcazar et al. 2017, Hafen et al. 2018).

We also infer the effective temperature, $T_{\rm eff}$, obtained by dividing the total tSZE flux with the total kSZE flux from the observational result, which is shown in Fig. 8.3. The temperature is found to be much lower than that from X-ray observations. As discussed in L17b, this may be due to a different weighting to average the temperature as well as due to a selection bias in the sample for X-ray observations. Recent studies using absorption lines toward quasars found that the CGM gas may be much cooler than the virial temperature (e.g. Werk et al 2014, 2016). For comparison, we estimate $T_{\rm eff}$ from the simulations by dividing the total sum within R_{500} of the electron pressure with that of the number of free electrons. As seen in Fig. 8.3, all the simulations predict that the effective temperature of the gas is about the virial temperature, which is up to an order of magnitude higher than that inferred from the SZE analysis. It is interesting to note that the different feedback models implemented in simulations predict different fraction of gas expelled out of haloes but predict same temperature of gas in haloes.

8.4.3 The pressure - density relation of IGM

Fig. 8.4 shows the pressure - density relation obtained following the method described in section 8.3.4 by the red solid line with the band showing the 1σ scatter based on the posterior distribution. The median values of the parameters are $\{\rho_{\rm m,0}/\,\overline{\rho}_{\rm m},\, \alpha_1,\, \alpha_2\} = \{3.0,\, 1.7,\, 2.2\}$. We refer the reader to Lim et al. (2018b) for more details of the posterior distribution of the parameters. As one can see from the median values of the parameters, the relation closely follows that of an adiabatic

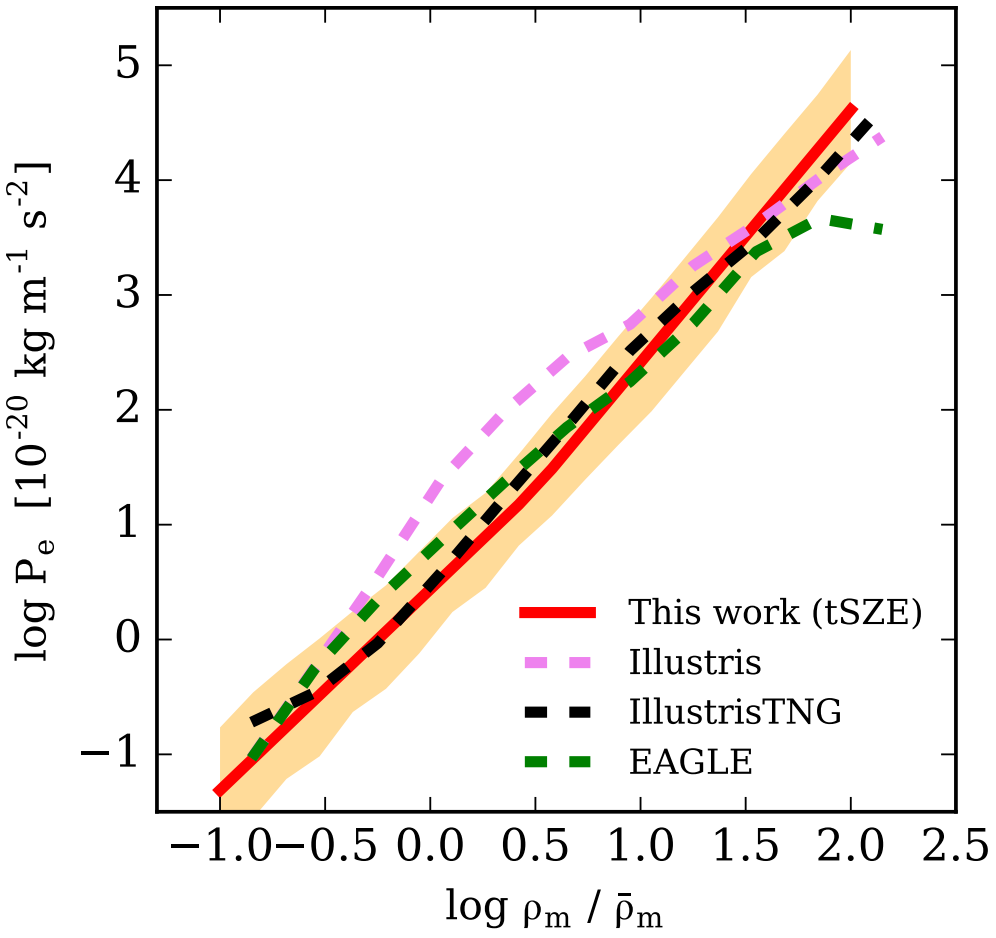

Figure 8.4. The pressure - density relation of IGM, compared with that from the simulations. The red line shows the mean relation obtained with our method, with the orange band showing the 1σ dispersion estimated from the uncertainties in the constrained parameters. The violet, black, and green lines show the mean relation from the Illustris, IllustrisTNG, and EAGLE, respectively.

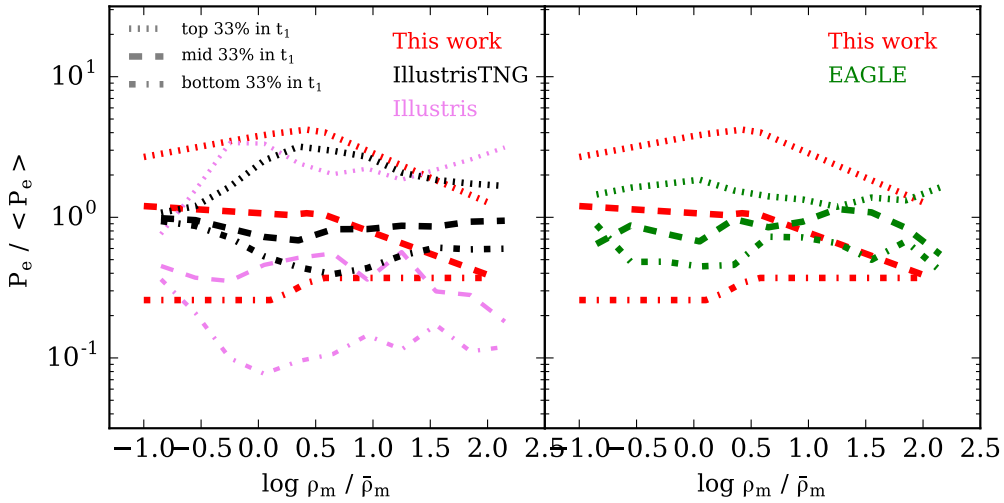

Figure 8.5. The pressure - density relation obtained for the three sub-samples of the grid-cells according to their ranking in the tidal field strength, t_1, at given density. The red lines show the mean relations obtained from the observation data, which are compared with the results from the IllustrisTNG and Illustris (left panel) and EAGLE (right panel). For each case, the results are normalized by the mean relation from the whole sample of the grid-cells shown in Fig. 8.4.

equation of state, $P_{\rm e} \sim \rho^{5/3}$, but with a steeper slope in dense regions. This may be due to heating sources available in dense regions because of star formation such as the AGN feedback.

In the figure, we compare the results with the predictions from the simulations. As one can see, the match between the observational result and the TNG or EAGLE is remarkable. Illustris predicts a much higher thermal energy in regions with $1 \leq \rho_{\rm m}/\overline{\rho}_{\rm m} \leq 10$, again, possibly due to the stronger radio-mode AGN feedback implemented in the simulation.

We also convert the relations onto the temperature - density space, by assuming that each grid-cells has the average ionized gas mass fraction equal to the cosmic baryon fraction, as is motivated by the simulation results (see Lim et al. 2018b). The temperature thus estimated is lower than 10^4K in regions with $\rho_{\rm m} \leq \overline{\rho}_{\rm m}$, and increases with the density, up to 10^6K in regions with $\rho_{\rm m} \geq 100\,\overline{\rho}_{\rm m}$, the regions of haloes. The

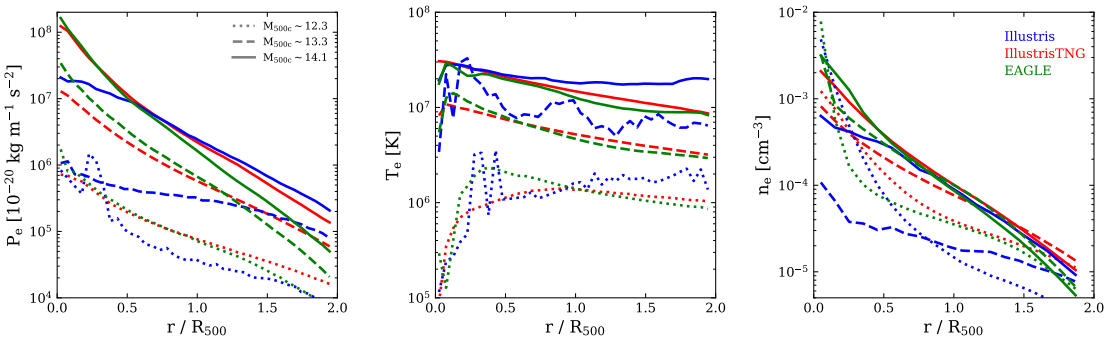

Figure 8.6. Comparison of the pressure (left), temperature (middle), and electron number density (right) profiles of gas associated with haloes of different mass between the simulations. Different line styles were used for the profiles in haloes of different mass as indicated.

TNG predicts higher gas temperature in regions with $3 \leq \rho_{\rm m}/\,\overline{\rho}_{\rm m} \leq 10$, the regions of cosmic filaments and sheets, while the EAGLE predicts higher temperature in less dense regions with $0.3 \leq \rho_{\rm m}/\,\overline{\rho}_{\rm m} \leq 3$.

Now we sub-sample the grid-cells into three according to the tidal field strength, t_1, each sub-samples containing a third of the total number of cells at a given density, and constrain the pressure - density relation for the three sub-samples jointly. We also compute t_1 for the grid-cells in the simulations as described in section 8.3.4, and divide the cells into three equal-sized sub-samples according to the ranking in t_1. The average relations between the pressure and matter density thus estimated from the observation and the simulations are shown in Fig. 8.5. One can see that the predictions from the TNG and EAGLE are clearly different, breaking the degeneracy between the two simulations that have been seen in the gas contents of haloes and the total average pressure of IGM at given density.

8.4.4 The profiles of halo gas properties

We also investigate the profiles of gas properties including the pressure, temperature, and number density of the free electrons, in the proximity of galaxy haloes from

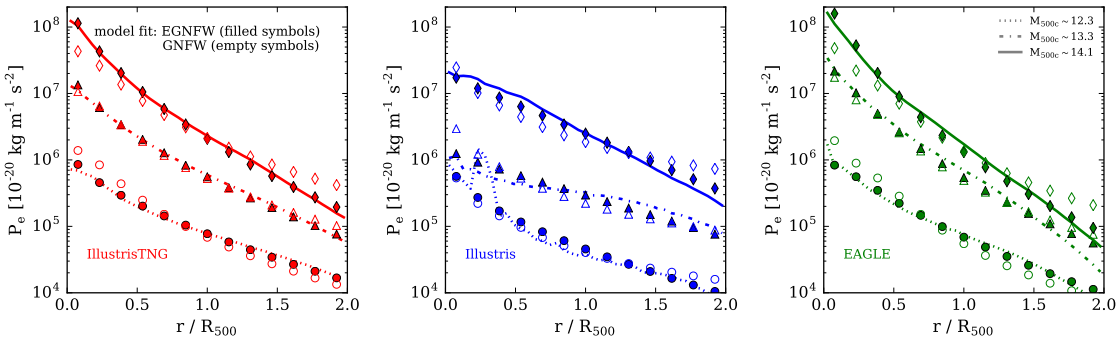

Figure 8.7. The pressure profile of halo gas from the simulations (lines with different styles) fitted with the generalized NFW (GNFW; empty symbols) and extended GNFW (EGNFW; filled symbols) models. For clarity, the results are shown only for three halo mass ranges as indicated in the right-panel.

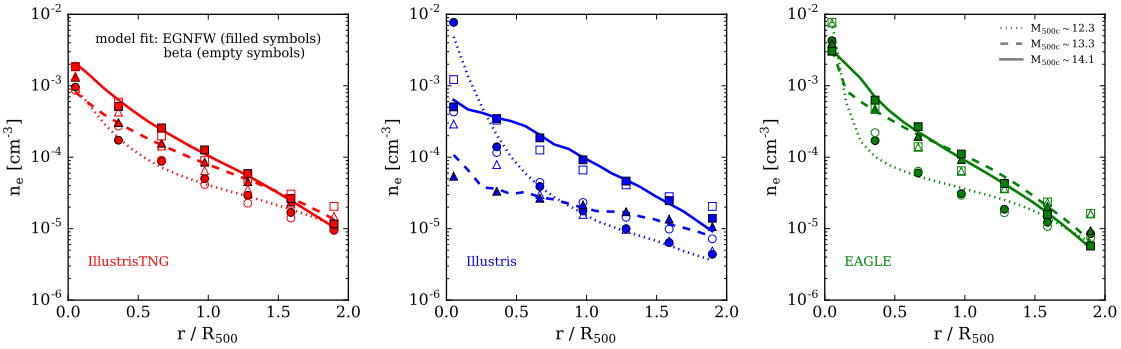

Figure 8.8. The electron density profile of halo gas from the simulations (lines with different styles) fitted with the β-profile (empty symbols) and EGNFW (filled symbols) model. For clarity, the results are shown only for three halo mass ranges as indicated in the right-panel.

the different simulations. We take into account only the free electrons gravitationally bound to the haloes. The median profiles are shown in Fig. 8.6 for a few halo mass bins. For massive haloes, Illustris predicts shallower profiles of the pressure and temperature compared to the other simulations, with lack of gas at halo center, implying that the stronger AGN feedback implemented in Illustris removes the gas from halo center to outskirt. That also explains the higher temperature in the outer region from Illustris. TNG also shows shallower profiles than EAGLE with slightly lower electron number density at $r \leq 0.5R_{500}$. For intermediate-mass haloes, the same trends as seen for the massive haloes are found but enhanced much stronger. Illustris predicts more than an order of magnitude lower density in the inner region, and much less smooth temperature profile. For low-mass haloes, the difference among the simulations at $r \leq 0.5R_{500}$ are much weaker. It is interesting to note, however, that the gas in the outskirt of haloes is much less dense but is at a higher temperature in Illustris than in the TNG and EAGLE. The temperature of the gas is also much more stochastic in Illustris compared to the other two simulations.

The functional form of the UPP is a generalized NFW (GNFW) model which was first introduced by Nagai et al. (2007),

$$\frac{P(r)}{P_0} = \frac{1}{(c_{500}\, r/R_{500})^{\gamma}[1 + (c_{500}\, r/R_{500})^{\alpha}]^{(\beta-\gamma)/\alpha}}, \tag{8.13}$$

where P_0 is a normalization parameter, and the parameters γ, α, and β describe the slopes at $r \ll r_s$, at $r \sim r_s$, and at $r \gg r_s$, respectively, with $r_s = R_{500}/c_{500}$. It was shown that the pressure profile of X-ray groups and clusters from observations is well described by the GNFW model (e.g. Arnaud et al. 2010, Plagge et al. 2010, Sun et al. 2011) but only out to $r \sim R_{500}$ in most cases. Using a combination of X-ray observation and simulations, Arnaud et al. (2010) found that the best parameter values are $\{\alpha, \beta, \gamma, c_{500}\} = \{1.05, 5.49, 0.308, 1.18\}$.

Table 8.1. The pressure profile fitting.

Profiles	α	β	γ	$c_{500,0}$	δ
Illustris	1.17	29.8	0.131	0.0972	0.180
IllustrisTNG	1.08	6.28	0.331	0.842	0.197
EAGLE	0.821	6.74	0.398	1.26	0.212

Le Brun et al. (2015; LB15 hereafter), however, found that the pressure distribution predicted from simulations is not well described by the GNFW for haloes of different mass in general, but requires an additional parameter via a mass dependence of c_{500} instead,

$$c_{500} = c_{500,0}\,(M_{500}/10^{14}\mathrm{M}_{\odot})^{\delta}\,. \tag{8.14}$$

As seen in the equation, this extended GNFW (EGNFW, hereafter) predicts that the concentration parameter may be a function of mass, which is a reasonable expectation. LB15 indeed found a non-zero mass dependence with $\delta \sim 0.263$, indicating that the gas is more concentrated in massive haloes than in lower-mass haloes. This may be because the gas at the center is blown out to the outskirt by galactic feedbacks more effectively in lower-mass haloes, resulting in a shallower profile. We fit the profiles from the TNG, Illustris, and EAGLE with both the GNFW and EGNFW models, shown by the empty and filled symbols, respectively in Fig. 8.7. The best-fit parameters for the EGNFW model are listed in Table 8.1. Unlike LB15, the normalization of the profile, P_0, is treated as a free parameter, the value of which is independently determined for haloes of different because our main interest is the profile shape. It is clearly seen that the GNFW model, in which the shape is independent of mass, is insufficient to describe the profiles accurately because the profiles from haloes of different mass have different shapes thus are not matched only with normalizations.

Table 8.2. The density profile fitting.

Profiles	β	γ_0	ϵ	$c_{500,0}$	δ
Illustris	4.03	0.0498	-0.953	0.990	0.595
IllustrisTNG	11.0	0.602	-0.104	0.446	0.104
EAGLE	13.3	0.754	-0.195	0.454	0.354

Our results also show that gas distribution is flatter in lower-mass haloes, which confirms the finding from previous studies. We find, however, a significant degeneracy between the parameters as well as correlations. We will explore this in more details in a forthcoming paper (Lim et al. in prep) where we will apply the EGNFW functional form as a filter shape for the matched filter to the CMB maps to constrain the model parameters, thus the pressure profile, from observations as well as extract the tSZE signal from halo gas.

We also fit the profile of electron number density with the β-profile. As seen in Fig. 8.8, the shape of the profile strongly depends on mass, thus cannot be well described by a mass-independent profile including the β-profile. Instead, we apply the EGNFW in another form,

$$n_{\rm e}(r) \propto \frac{1}{(c_{500}\, r/R_{500})^{\gamma}[1+(c_{500}\, r/R_{500})^2]^{(\beta-\gamma)/2}}, \tag{8.15}$$

where

$$\begin{aligned} c_{500} &= c_{500,0}\,(M_{500}/10^{14}{\rm M}_{\odot})^{\delta} \\ \gamma &= \gamma_0\,(M_{500}/10^{14}{\rm M}_{\odot})^{\epsilon}\,. \end{aligned} \tag{8.16}$$

The best-fit values of the parameters for the EGNFW model are listed in Table 8.2.

8.5 Summary and discussion

In this paper, we compared the gas contents of the circumgalactic medium (CGM) and intergalactic medium (IGM) between observation and simulations, to explore the possibility of using the gas contents to constrain models for physics of galaxy evolution. Specifically, we used the recent Sunyaev-Zel'dovich effect (SZE) results obtained by following Lim et al. (2017b, 2018a, 2018b) as our main observation data, together with four state-of-art cosmological gas simulations. Briefly, the observation data were obtained based on *Planck* CMB temperature maps, by cross-correlating the CMB maps with galaxy groups identified by a group finder, and with reconstructed large-scale environments such as density field, tidal field, and velocity field. We adopted the matched filter (MF) technique to minimize contamination and maximize the signal-to-noise. The filters are jointly implemented to automatically take into account the two-halo terms.

For the tSZE signal from haloes, all observational results and predictions from the simulations considered in our analysis except for that by PCXI indicate a certain degree of deviations from the self-similar case, implying the effect of feedbacks on galaxy systems with $M_{500} \leq 10^{13}\,\mathrm{M}_{\odot}$. We suspect that the assumption of flat background adopted by PCXI led to their finding that the tSZE flux follows the self-similar case even for low-mass systems with $M_{500} \sim 10^{12}\,\mathrm{M}_{\odot}$. Different simulations make different predictions because of variations in physical models implemented in the simulations. The Illustris, in particular, predicts a significantly lower thermal energy of gas in haloes with $M_{500} \sim 10^{13-13.5}\,\mathrm{M}_{\odot}$. This is believed to be due to the stronger AGN feedback adopted in the simulation. The predictions from IllustrisTNG and EAGLE are remarkably similar to each other even down to $M_{500} \sim 10^{12}\,\mathrm{M}_{\odot}$.

The kSZE signal from the observation data infers that the gas fraction in haloes is almost equal to the cosmic baryon fraction even in the low-mass systems, unlike the predictions from simulations where that in haloes is only $20 - 40\%$ of the cosmic

mean fraction. There can be still, however, the residual contamination by gas along the LOSs between haloes not taken into account by the method presented here. Such contamination is found to be about 20% according to simulations but is expected to strongly depend on the baryon physics. The effective temperature from the observation, obtained by dividing the tSZE flux with the kSZE flux, is also found to be much higher than that from the simulations. Notably, even the simulations with completely different predictions about the gas mass fraction produce very similar predictions that the gas is at the virial temperature.

We then investigated the pressure - density relation of the IGM. The overall slope of the relation closely follows that of an adiabatic equation of state, but with some evidence for a steeper slope in the regions of haloes and large-scale structures. Our result matches well the predictions from simulations except for the Illustris that predicts a much higher thermal energy of IGM in the regions of intermediate mean density. The pressure - density relation also depends on the strength of large-scale tidal field. This may reflect that where the tidal field is strong, the structure formation occurs early that the heating sources such as stellar and AGN feedback become available early. We show that the tidal field dependence of the relation can help break degeneracy between models that produce very similar predictions elsewhere.

Finally, we probed the radial profiles of gas properties in haloes from different simulations. The simulations with stronger AGN feedback tend to predict shallower profiles of gas pressure and temperature with lack of gas at the halo center. The difference between the predictions is most prominent in intermediate-mass systems with $M_{500} \sim 10^{13-13.5}\,\mathrm{M}_{\odot}$. We show that the profiles from the different simulations are not described well by a generalized NFW (GNFW) profile. We demonstrate that a mass-dependent model describe the profiles well, instead.

Our results clearly demonstrate the power of using the Sunyaev-Zel'dovich effect to infer the gas contents of the CGM and IGM, and its potential to constrain galaxy

formation models. In particular, because most state-of-art simulations have their model parameters tuned to match the galaxy properties averaged over a broad range of galaxy population from observation, it is essential to disentangle the average galaxy properties with the properties of their host haloes and large-scale structures, in order to break the degeneracy between models. With future CMB surveys such as the CMB-S4, Simons observatory, and Toltec, it is also expected that the SZE will allow us to directly probe the gas profiles of low-mass systems down to $M_{500} \sim 10^{12}\,\mathrm{M}_{\odot}$ by stacking. The same approach as developed here can be applied to the future surveys to provide stringent constraints on the impact of galactic feedbacks.

CHAPTER 9

CONCLUSION AND FUTURE WORK

9.1 Conclusion

In this dissertation, we studied the relations between galaxies, dark matter halos, gas, and cosmic web in a comprehensive, systematic way. To do this, we developed an improved halo-based group finder, and applied it to large redshift surveys including 2MRS, 6dF, SDSS DR13, and 2dF to construct the group catalogs. Using the well-defined dark matter halo catalog and cross-correlating it with galaxy properties, we found a proxy of halo age that can be applied to observation. We also demonstrated that the conditional stellar mass function provides additional, tighter constraints on galaxy formation models than the total stellar mass function. Using a parametrized model, we found that observations of the conditional stellar mass function in the low-redshift Universe requires a higher star formation at high redshift in low-mass halos. We then probed the alignments between galaxies and cosmic web by using the tidal field as a tracer of the large-scale structures, finding that both the orientation and position of galaxies are aligned with the morphology of cosmic web. Then, we cross-correlated the galaxy systems with the Sunyaev-Zel'dovich effect (SZE) signal from the *Planck* CMB observation, to constrain the gas properties as a function of halo mass. The results indicate that the gas in low-mass halos deviates from the self-similar case due to its much lower temperature than the virial temperature, while being as abundant as the cosmic baryon fraction. We also cross-correlated the reconstructed large-scale properties with the SZE signal, to constrain the properties of gas in different environment. Our results indicate that the gas in IGM closely

follows the adiabatic equation of state, while the gas energy is higher in the regions of stronger tidal field. This may be because the structures form earlier in those regions, resulting in a more accumulated energy provided by feedbacks. Finally, we compare our results from the SZE analysis with the predictions from the state-of-art simulations, demostrating that the SZE is a useful tool to constrain the galaxy formation models.

9.2 Future work

9.2.1 Testing the SZE flux extraction methods

Due to the low signal-to-noise of the SZE signals as well as the large statistical and systematic uncertainties, the SZE results from the current data are sensitive to the flux extraction methods, and thus consistent with a wide range of galaxy formation models depending on which method to be used. Given the higher sensitivity and better resolution of the upcoming CMB surveys such as the Toltec, CMB-S4, and Simons Observatory, it is crucial to have an accurate extraction method to provide the tighter constraints for the galaxy models. As briefly mentioned in this dissertation, we are currently testing the performance of popular SZE flux extraction methods in literature including our method, by applying them to the realstic mock CMB maps constructed based on simulations (Lim et al. in prep; also see Fig. 9.1). In the paper, we also test the impact of group finder such as a miss-centering and the mass uncertainty of halos on the flux calculated from the mock maps of different resolutions.

9.2.2 Direct probe of the SZE profile

The relatively large beam size of the current CMB observations required assumptions on the profiles to interpret the results. The up-to-arcsecond resolution of the future CMB surveys, however, will allow a direct probe of the SZE profile down to $10^{12}M_{\odot}$ halos by stacking many of similar systems (see Fig. 9.2), significantly reduc-

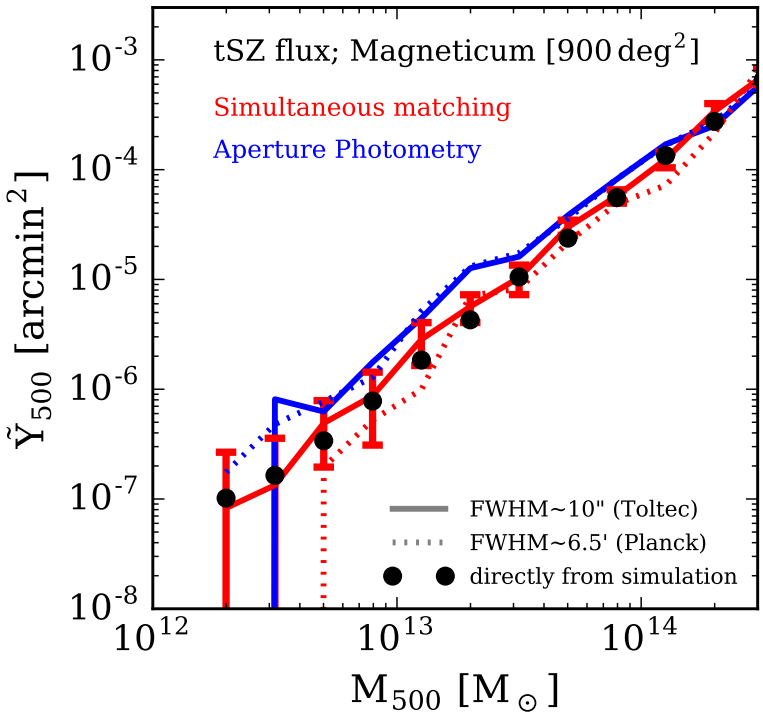

Figure 9.1. Test of the flux extraction methods for the tSZE. Our simultaneous matching method applied to the tSZE light-cone constructed from the Magneticum simulation, shown by the red lines, recovers the true flux (black dots) directly from the simulation, in particular, when it is applied to a survey of a high angular resolution (solid lines) while aperture photometry (blue) over-predicts the flux.

ing the systematic uncertainties of the current SZE studies. In addition, the directly probed SZE profile itself will provide additional constraint on galaxy formation models such as the impact of AGN feedbacks.

9.2.3 Extension to high-z

Most of the works presented in this dissertation has been done so far only in the low-redshift Universe of $z \leq\sim 0.2$. As some of our results indicate, however, relying only on the low-redshift observations results in a degeneracy and large uncertainties to tightly constrain the galaxy formation models. It is essential, therefore, to extend

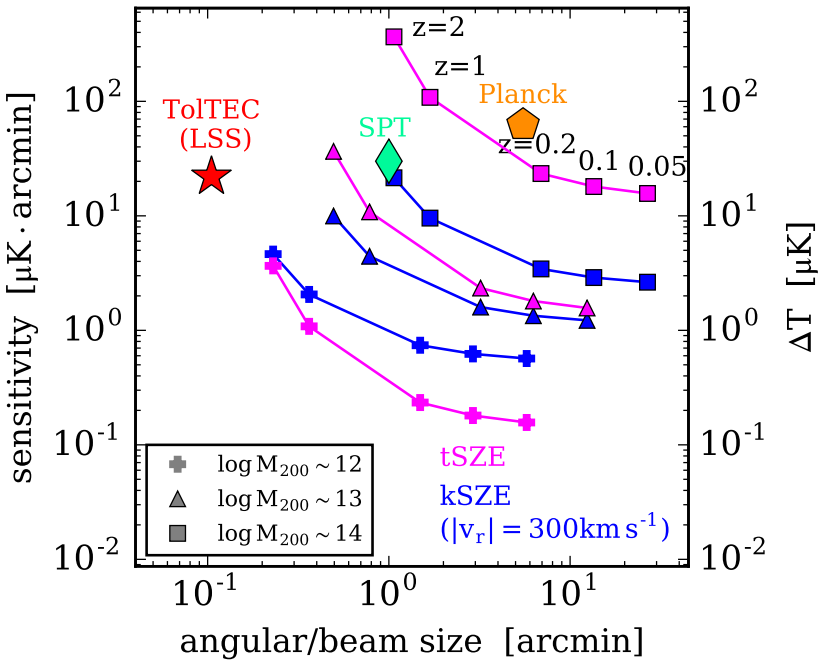

Figure 9.2. The beam size and sensitivity of the CMB surveys compared with the angular size and mean SZE signal (the vertical axis on the right) for halos of different mass as indicated in the legend at different redshifts. The kSZE signal is calculated for $|v_r| = 300\,\mathrm{km\,s^{-1}}$.

the works to the higher redshift up to $z \sim 2$. This will become feasible very soon, fortunately, with the upcoming galaxy surveys such as the Euclid, DESI and Subaru PFS survey, which will scan more than hundreds of thousands galaxies at the high redshift with an accurate redshift determination, together with the upcoming CMB surveys. Preliminary results show that our group finder presented in this dissertation can also identify well the groups at the high redshift with accurate halo mass estimates (Fig. 9.3). Our predictions about the stellar mass function at the high redshift can be also tested to provide a further constraints on the galaxy models. The cross-correlation of the halos identified from the galaxy surveys with the SZE signals from the future CMB surveys will allow us to study the evolution of gas properties during the past $\sim$ 10 billion years in our Universe, which will significantly improve our

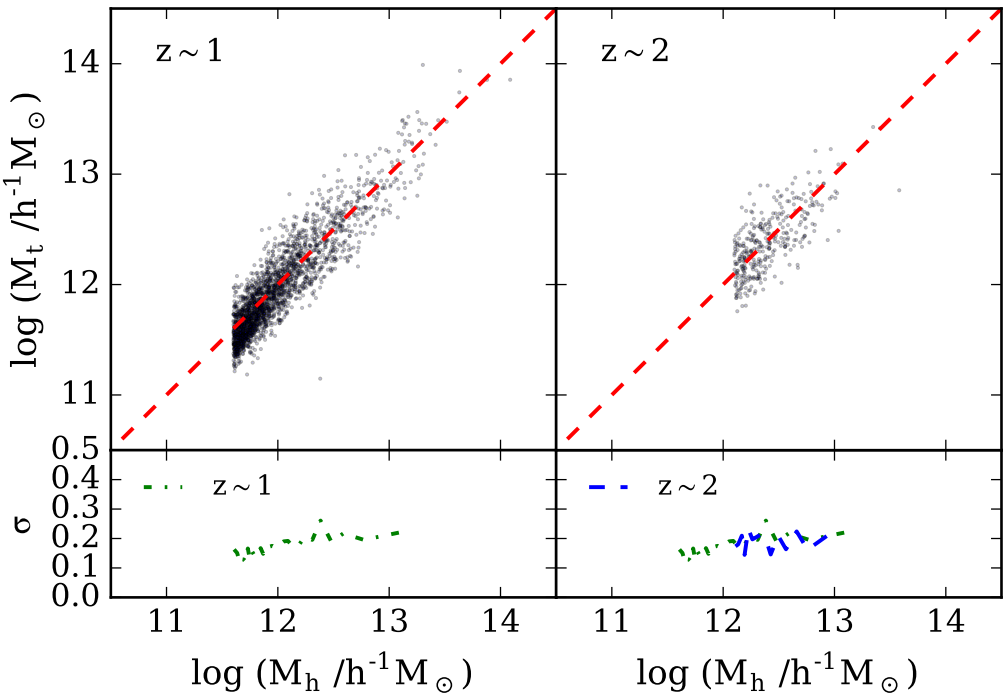

Figure 9.3. The group finder from this dissertation applied to the mock catalogs at high redshift of $z = 1$ (left) and $z = 2$ (right). The group mass estimated by the group finder (horizontal axis) is unbiased with respect to the true halo mass (vertical) with a scatter of 0.2 dex (lower panels), comparable to that for the low-redshift mock samples.

understanding of galaxy evolution, galactic feedbacks and their connection to the cosmic web.

APPENDIX A

TESTING THE GROUP FINDER WITH MOCK SAMPLES CONSTRUCTED USING AN EMPIRICAL MODEL

To further check the consistency of our group finder, which is calibrated with galaxies in the EAGLE simulation, we have applied it to another set of mock samples constructed using the empirical model of galaxy formation described in Lu et al. (2015), which is based on Lu et al. (2014). The tests we have made are the same as those with the EAGLE mock samples presented in §1.4, using exactly the same methods described in §1.3. To do this, we first applied the empirical model to the merger trees extracted from the EAGLE to assign stellar masses to galaxies. As an example, Figure A.1 compares the true halo masses from the EAGLE simulation with the final group masses obtained by applying our group finder to the mock samples thus constructed. The scatter in the halo mass comparison is around 0.2 dex, very similar to what was found in Figure 1.10 except for the very massive end which shows slightly larger scatter. This demonstrates that the performance of our group finder is not sensitive to the details of how galaxies form in dark matter halos, as represented by the differences between the EAGLE and the empirical model of Lu et al. (2015).

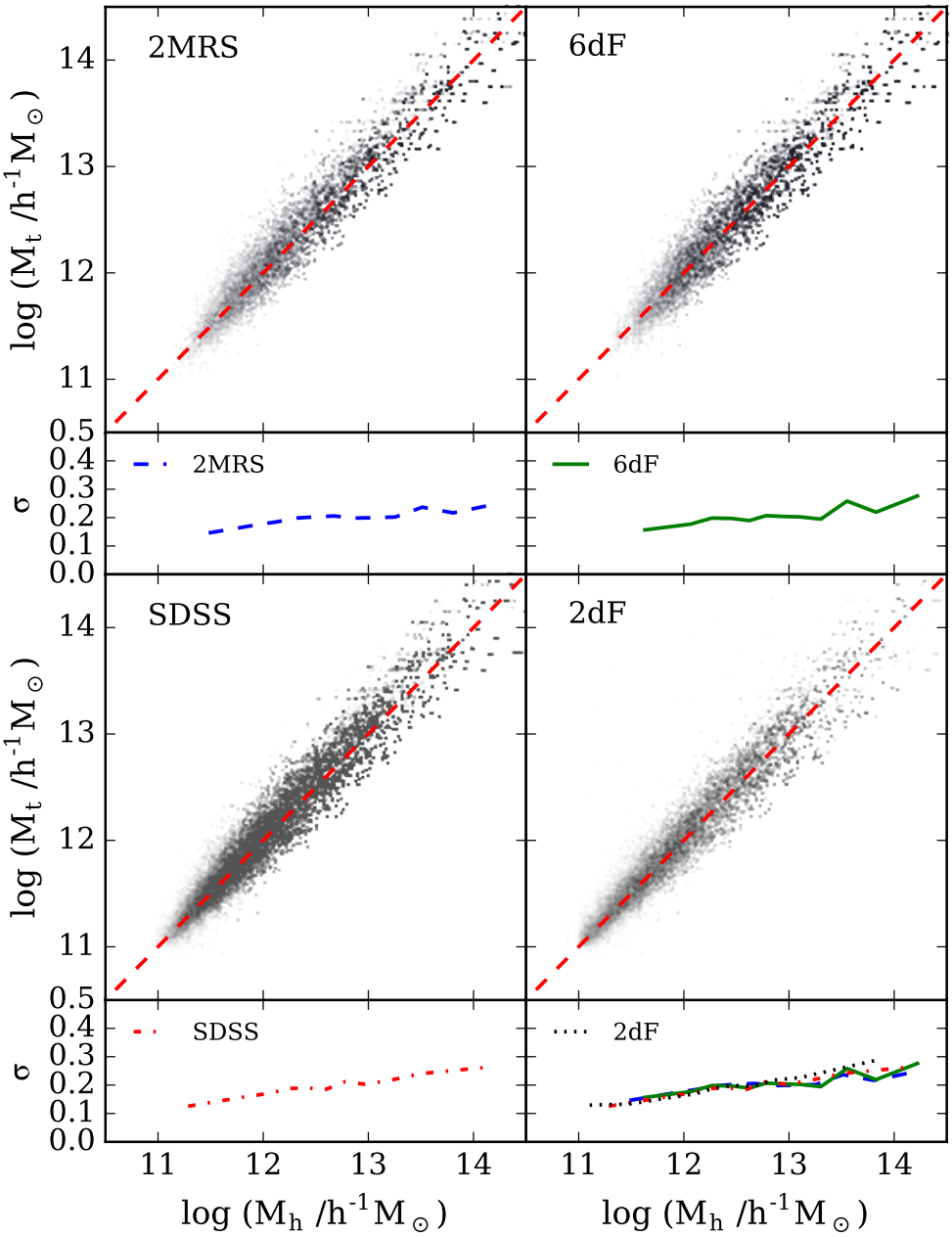

Figure A.1. Comparison between the true halo mass (vertical axis) and the group mass identified by our group finder (horizontal axis) using stellar mass as the proxy of halo mass for the mock samples of 2MRS, 6dFGS, SDSS, and 2dFGRS. Here mock samples are constructed by applying the empirical model of Lu et al. (2015) to the halo merger trees extracted from the EAGLE simulation. The small rectangular panels plot the scatter of the true halo masses at given group mass.

APPENDIX B

CONTAMINATION BY PROJECTION EFFECTS

We use the ELUCID (Wang et al. 2016) simulation to check the contamination to the kSZE by projection effects. ELUCID is a N-body simulation run with GADGET-2 (Springel 05), using $(3072)^3$ particles in a periodic box of $\sim (500\,h^{-1}\mathrm{Mpc})^3$, and assuming a cosmological model given by WMAP5 (Dunkley et al. 2009). Dark matter halos are identified with the FoF algorithm.

For simulated halos in a given mass bin, we estimate the total momentum of dark matter particles within R_{200}, $\mathcal{K}_{200,\mathrm{sph}}$, and compare it with that in a cylinder defined by $r_\perp \leq R_{200}$ and $r_\parallel \leq l_{\mathrm{cyl}}$, $\mathcal{K}_{200,\mathrm{cyl}}$, where $r_\perp$ and $r_\parallel$ are the distances from the halo center that are perpendicular and parallel to the LOS, respectively, and l_{cyl} is half the length of the cylinder along the LOS. When estimating $\mathcal{K}_{200,\mathrm{cyl}}$, we exclude contribution from other halos with masses above $M_{200} \sim 10^{12.4}\mathrm{M}_\odot$, since our method takes into account the projections of such halos. Fig. B.1 shows $(\mathcal{K}_{200,\mathrm{cyl}} - \mathcal{K}_{200,\mathrm{sph}})/\mathcal{K}_{200,\mathrm{sph}}$ as a function of l_{cyl} for halos of different masses. We only select the halos with $|v_r| > 80\,\mathrm{km\ s^{-1}}$ because the results from halos with $|v_r| < 80\,\mathrm{km\ s^{-1}}$ are noisy. The solid lines are the predictions from the simulation. Fig. B.1 also shows the mass fraction and the mass-weighted velocity within the cylinder with respect to that within R_{200}, in the middle and right panels, respectively. Note that the quantities are plotted beyond the box size of the simulation. We take the average matter density inside the simulation box to extrapolate the matter distribution outside the box. We use the peculiar velocity correlation predicted by the linear theory following Wang et al. (2018), to estimate the mass-weighted velocity outside the simulation box. The momentum from the

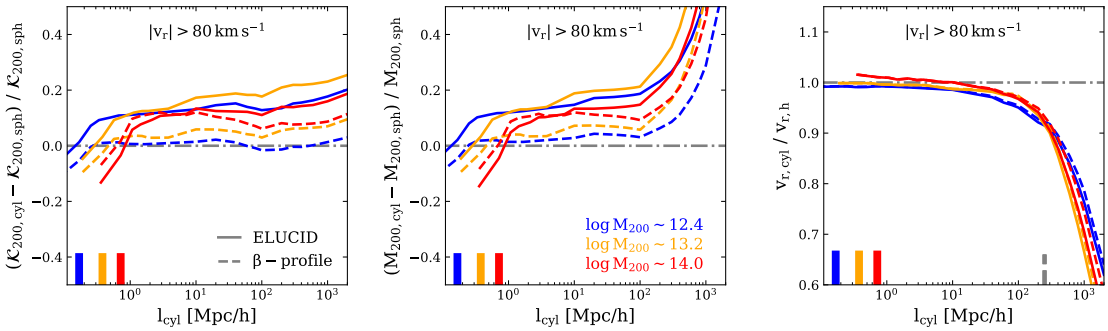

Figure B.1. The total momentum (left panel), total mass (middle), and average velocity (right) of the dark matter particles within the cylinder of varying length, $l_{\rm cyl}$, with respect to that within a sphere of R_{200} at the halo centers (see the text for details on the construction of cylinders), predicted directly from the ELUCID simulation (solid lines) and from the β-profile fit to the simulation (dashed). The averages for halos with $|v_r| > 80\,\mathrm{km\ s^{-1}}$ are plotted, separately for halos of different masses, as indicated. The thick solid ticks represent the values of R_{200} for the given halo masses. The gray vertical tick in the right panel indicates half of the simulation box size. The mass in cylinders outside the simulation box are estimated assuming the average density inside the simulation box. The velocities and momenta outside the simulation box are computed from the velocities predicted by the linear theory.

cylinder outside the simulation box is obtained by multiplying the extrapolation of the matter distribution with the mass-weighted velocity from the linear theory. The test result shows that the LOS contribution from particles outside halos is about 20%. Because the *Planck* beam size is up to three times as large as R_{200} of the low-mass halos in our sample depending on redshift, gas distributed at $r_\perp > R_{200}$ may also contaminate the signal to be detected from halos. To see this, we have carried out similar tests for the cylinders of $r_\perp \leq 2R_{200}$ and $3R_{200}$, and found that the momentum contribution from the particles outside halos is about 50% and 70%, respectively.

However, this does not mean that the results presented in §6.4 are over-estimated by these amounts. In fact, our simultaneous matching method uses the projected β-profile, which includes mass distributions outside halos on the one hand but suppress the contributions from large distances on the other. To see this, we fit the projected profiles from the simulation with the projected β-profile for each $l_{\rm cyl}$, and compute the

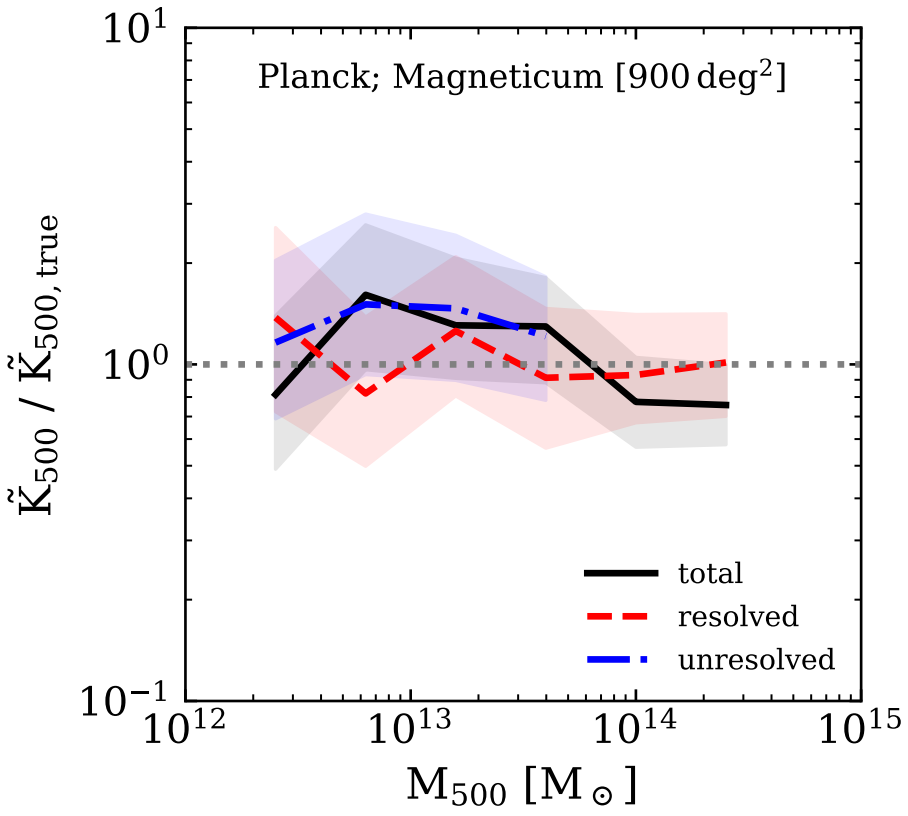

Figure B.2. The $\tilde{K}_{500}$ obtained by applying our method to the realistic kSZE mock maps for the *Planck* constructed based on the Magneticum simulation (Dolag et al. 2016), normalized by the true $\tilde{K}_{500}$ directly measured from the simulation. The black solid line shows the result for all halos. The red and blue lines show the results obtained for halos with R_{500} greater ('resolved') or smaller ('unresolved') than the *Planck* beam, respectively. The bands show 68% ranges of the posterior distribution.

corresponding LOS contamination. The results, shown by the dashed lines, indicate that the momentum recovered over-estimates the true values by $\sim 10\%$ or less. We find similar results from the same test based on another simulation, the IllustrisTNG (Pillepich et al. 2018).

To quantify the net impact of contamination more directly, we apply our matched filter method to the kSZE light-cone map, constructed from the Magneticum simulation (Dolag et al. 2016), and examine how our method recovers the true $\tilde{K}_{500}$

directly measured from the simulation. The Magneticum is a set of cosmological gas simulations of various volumes and resolutions, run with an improved version of GADGET-3 (Springel 2005). They adopted a WMAP7 cosmology (Komatsu et al. 2011). The specific run chosen for our analysis samples a box of $L = 352\,h^{-1}\mathrm{Mpc}$ with $2 \times (1584)^3$ particles, which results in a mass resolution of $m_{\mathrm{DM}} \sim 10^9 \mathrm{M}_\odot$ for the dark matter particles. Based on the simulation, we construct the kSZE light-cone maps using the SMAC code (Dolag et al. 2005), which cover the redshift range of $0.0173 < z < 0.194$ and $900\,\mathrm{deg}^2$ of the sky. We generate the kSZE based on the dark matter particles only, assuming that gas particles follow the dark matter distribution with a universal ratio between the gas and total mass densities. We then generate the CMB temperature maps at the same three frequencies as used in our analysis, and add the primordial CMB anisotropy using the CMB power spectrum. Finally, we degrade the map with the *Planck* beam function, and add the instrument noise. To the constructed temperature maps, we apply our method to constrain the model parameters and estimate the corresponding $\tilde{K}_{500}$ for halos in the same mass bins as in our analysis. Fig. B.2 shows the $\tilde{K}_{500}$ recovered by our method with respect to the true value. The black solid line shows the result for all halos. The red and blue lines show the results for halos with R_{500} greater ('resolved') or smaller ('unresolved') than the *Planck* beam size, respectively, where two different sets of model parameters are assumed for the 'resolved' and 'unresolved' halos to model the simulated maps. The shaded bands show the 68% range of the posterior distribution for each case. Within the statistical uncertainty, the $\tilde{K}_{500}$ recovered by our method is unbiased relative to the true value. The results for the 'unresolved' halos tend to be higher than both those for the 'resolved' halos and the true $\tilde{K}_{500}$, but the increase is insignificant given the statistical uncertainty.

Based on the tests presented above, we conclude that our results shown in the main text are not affected significantly by the contamination of projection effects.

BIBLIOGRAPHY

[1] Abazajian K. N. et al., 2009, ApJS, 182, 543

[2] Abell G. O., 1958, ApJS, 3, 211

[3] Abell G. O., Corwin H. G. Jr., Olowin R. P., 1989, ApJS, 70, 1

[4] Aguirre A., Hernquist L., Schaye J., Katz N., Weinberg D. H., Gardner J., 2001, ApJ, 561, 521

[5] Agustsson I., Brainerd T. G., 2006, ApJ, 644, 25

[6] Agustsson I., Brainerd T. G., 2006, ApJ, 650, 550

[7] Albareti F. D. et al., 2017, ApJS, 233, 25

[8] Anglés-Alcázar D., Faucher-Giguère C.-A., Kereš D., Hopkins P. F., Quataert E., Murray N., 2017, MNRAS, 470, 4698

[9] Arnaud M., Pratt G. W., Piffaretti R., Böhringer H., Croston J. H., Pointecouteau E., 2010, A&A, 517, 92

[10] Azzaro M., Patiri S. G., Prada F., Zentner A. R., 2007, MNRAS, 376, L43

[11] Baldry I. K. et al., 2004, ApJ, 600, 681

[12] Baldry I. K., Glazebrook K., Driver S. P., 2008, MNRAS, 388, 945

[13] Baldry I. K. et al., 2012, MNRAS, 421, 621

[14] Battaglia N., Bond J. R., Pfrommer C., Sievers J. L., 2012, ApJ, 758, 74

[15] Battaglia N., Bond J. R., Pfrommer C., Sievers J. L., 2013, ApJ, 777, 123

[16] Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379

[17] Behroozi P. S. et al., 2013, ApJ, 763, 18

[18] Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57

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

[20] Bennett C. L. et al., 2003, ApJS, 148, 1

[21] Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587

[22] Berlind A. A. et al., 2006, ApJS, 167, 1

[23] Bilicki M., Jarrett T. H., Peacock J. A., Cluver M. E., Steward L., 2014, ApJS, 210, 9

[24] Blanton M. R. et al., 2003, ApJ, 592, 819

[25] Blanton M. R. et al., 2005, AJ, 129, 2562

[26] Blazek J., McQuinn M., Seljak U., 2011, JCAP, JCAP05(2011) 010

[27] Böhringer H. et al., 2007, A&A, 469, 363

[28] Booth C. M., Schaye J., 2009, MNRAS, 398, 53

[29] Borthakur J. K. et al., 2016, ApJ, 833, 259

[30] Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645

[31] Brainerd T. G., 2005, ApJ, 628, L101

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

[33] Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

[34] Bullock J. S.,Wechsler R. H., Somerville R. S., 2002, MNRAS, 329, 246

[35] Butcher H., Oemler A. Jr., 1978, ApJ, 226, 559

[36] Butcher H., Oemler A. Jr., 1984, ApJ, 285, 426

[37] Cen R., Ostriker J. P., 1999, ApJ, 514, 1

[38] Chabrier G., 2003, PASP, 115, 763

[39] Charlot S., Fall S. M., 2000, ApJ, 539, 718

[40] Chisholm J., Tremonti C. A., Leitherer C., Chen Y., Wofford A., 2016, MNRAS, 457, 3133

[41] Choi Y. Y., Han D. H., Kim S. S., 2010, Journal of Korean Astronomical Society, 43, 191

[42] Cohen M., Wheaton Wm. A., Megeath S. T., 2003, AJ, 126, 1090

[43] Cole S. et al., 2001, MNRAS, 326, 255

[44] Colless M. et al., 2001, MNRAS, 328, 1039

[45] Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201

[46] Conroy C., Wechsler R. H., 2009, ApJ, 696, 620

[47] Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486

[48] Crain R. A. et al., 2015, MNRAS, 450, 1937

[49] Crook A. C. et al., 2007, ApJ, 655, 790

[50] Croton D. J. et al., 2006, MNRAS, 365, 11

[51] Danforth C. W., Shull J. M., 2008, ApJ, 679, 194

[52] Davé R., Hernquist L., Katz N., Weinberg D. H., 1999, ApJ, 511, 521

[53] Davé R. et al., 2001, ApJ, 552, 473

[54] David L. P., Jones C., Forman W., Vargas I. M., Nulsen P., 2006, ApJ, 653, 207

[55] Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371

[56] Dawson K. S. et al., 2013, AJ, 145, 10

[57] de Graaff A., Cai Y.-C., Heymans C., Peacock J. A., 2019, A&A, 624, 48

[58] Dekel A., Silk J., 1986, ApJ, 303, 39

[59] Dolag K., Jubelgas M., Springel V., Borgani S., Rasia E., 2004, ApJ, 606, 97

[60] Dolag K., Hansen F. K., Roncarelli M., Moscardini L., 2005, MNRAS, 363, 29

[61] Dolag K., Komatsu E., Sunyaev R., 2016, MNRAS, 463, 1797

[62] Donoso E., O'Mill A., Lambas D. G., 2006, MNRAS, 369, 479

[63] Dressler A., 1980, ApJ, 236, 351

[64] Duarte M., Mamon G. A., 2015, MNRAS, 453, 3848

[65] Dubois Y. et al., 2014, MNRAS, 444, 1453

[66] Eke V. R. et al., 2004, MNRAS, 348, 866

[67] Ettori S., Baldi A., Balestra I., Gastaldello F., Molendi S., Tozzi P., 2015, A&A, 578, A46

[68] Fabjan D., Borgani S., Tornatore L., Saro A., Murante G., Dolag K., 2010, MNRAS, 401, 1670

[69] Fall S. M., Efstathiou G., 1980, MNRAS, 193, 189

[70] Faltenbacher A. et al., 2007, ApJL, 662, L71

[71] Faltenbacher A., Li C., White S. D. M. et al., 2009, Research in Astronomy and Astrophysics, 9, 41

[72] Faucher-Giguère C.-A., Kereš D., Dijkstra M., Hernquist L., Zaldarriaga M., 2010, ApJ, 725, 633

[73] Faucher-Giguère C.-A., Kereš D., Ma C. P., 2011a, MNRAS, 417, 2982

[74] Faucher-Giguère C.-A., Kereš D., 2011b, MNRAS, 412, 118

[75] Faucher-Giguère C.-A., Hopkins P. F., Kereš D., Muratov A. L., Quataert E., Murray N., 2015, MNRAS, 449, 987

[76] Faucher-Giguère C.-A., Feldmann R., Quataert E., Kereš D., Hopkins P. F., Murray N., 2016, MNRAS, 461, L32

[77] Feldmann R., Hopkins P. F., Quataert E., Faucher-Giguère C. A., Kereš D., 2016, MNRAS, 458, L14

[78] Feroz F., Hobson M. P., Bridges, M., 2009, MNRAS, 398, 1601

[79] Ferrarese L., Merritt D., 2000, ApJ, 539, 9

[80] Finkbeiner D. P. et al., 2016, ApJ, 822, 66

[81] Fukugita M., Hogan C. J., Peebles P. J. E., 1998, ApJ, 503, 518

[82] Fukugita M., Peebles P. J. E., 2004, ApJ, 616, 643

[83] Furlanetto S. R., Loeb A., 2004, ApJ, 611, 642

[84] Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C., 2005, MNRAS, 362, 41

[85] Gastaldello F., Buote D. A., Humphrey P. J., Zappacosta L., Bullock J. S., Brighenti F., Mathews W. G., 2007, ApJ, 669, 158

[86] Genel S. et al., 2014, MNRAS, 445, 175

[87] Goto T., 2005, MNRAS, 359, 1415

[88] Gott III J. R. et al., 2005, ApJ, 624, 463

[89] Governato F. et al., 2004, ApJ, 607, 688

[90] Governato F. et al., 2010, Nature, 463, 203

[91] Greco J. P., Hill J. C., Spergel D. N., Battaglia N., 2015, ApJ, 808, 151

[92] Guedes J., Callegari S., Madau P., Mayer L., 2011, ApJ, 742, 76

[93] Guo Q., White S. D. M., Li C., Boylan-Kolchin M., 2010, MNRAS, 404, 1111

[94] Guo Q. et al., 2011, MNRAS, 413, 101

[95] Guo Q. et al., 2013, MNRAS, 428, 1351

[96] Guo H. et al., 2015, MNRAS, 453, 4368

[97] Haehnelt M. G., Tegmark M., 1996, MNRAS, 279, 545

[98] Hafen Z. et al., 2018, arXiv:1811.11753

[99] Hahn O., Carollo C. M., Porciani C., Dekel A., 2007, MNRAS, 381, 41

[100] Hand N. et al., 2012, Physical Review Letters, 109, 1101

[101] Hawley D. L., Peebles P. J. E., 1975, AJ, 80,477

[102] Hearin A. P., Watson D. F., 2013, MNRAS, 435, 1313

[103] Hearin A. P., Watson D. F., Becker M. R., Reyes R., Berlind A. A., Zentner A. R., 2014, MNRAS, 444, 729

[104] Heckman T. M., Alexandroff R. M., Borthakur S., Overzier R., Leitherer C., 2015, ApJ, 809, 147

[105] Hernández-Monteagudo C., Ma Y.-Z., Kitaura F. S., Wang W., Génova-Santos R., Macías-Pérez J., Herranz D., 2015, Physical Review Letters, 115, 191301

[106] Herranz D. et al., 2002, MNRAS, 336, 1057

[107] Hill J. C., Spergel D. N., 2014, JCAP, 2, 030

[108] Hill J. C., Ferraro S., Battaglia N., Liu J., Spergel D. N., 2016, Physical Review Letters, 117, 051301

[109] Hill J. C., Baxter E. J., Lidz A., Greco J. P., Jain B., 2018, Physical Review D, 97, 083501

[110] Hinshaw G. et al., 2013, ApJS, 208, 19

[111] Hirata C. M., Mandelbaum R., Ishak M. et al., 2007, MNRAS, 381, 1197

[112] Hirschmann M., Dolag K., Saro A., Bachmann L., Borgani S., Burkert A., 2014, MNRAS, 442, 2304

[113] Hojjati A., McCarthy I. G., Harnois-Deraps J., Ma Y.-Z., Van Waerbeke L., Hinshaw G., Le Brun A. M. C., 2015, JCAP, 10, 047

[114] Holmberg E., 1969, Ark. Astron., 5, 305

[115] Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C. A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581

[116] Hopkins P. F. et al., 2018, MNRAS, 480, 800

[117] Huchra J. P., Geller M. J., 1982, ApJ, 257, 423

[118] Huchra J. P., Davis M., Latham D., Tonry J., 1983, ApJS, 52, 89

[119] Huchra J. P. et al., 2012, ApJS, 199, 26

[120] Hummels C. B., Bryan G. L., Smith B. D., Turk M. J., 2013, MNRAS, 430, 1548

[121] Hurier G., Macías-Pérez J. F., Hildebrandt S., 2013, A&A, 558, 118

[122] Jiang F., van den Bosch F. C., 2014, MNRAS, 440, 193

[123] Jing Y. P., Mo H. J., Börner G., 1998, ApJ, 494, 1

[124] Joachimi B., Mandelbaum R., Abdalla F. B., Bridle S. L., 2011, 527, A26

[125] Jones D. H. et al., 2004, MNRAS, 355, 747

[126] Jones D. H., Saunders W., Read M., Colless M., 2005, PASA, 22, 277

[127] Jones D. H. et al., 2009, MNRAS, 399, 683

[128] Jones T., Stark D. P., Ellis R. S., 2012, ApJ, 751, 51

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

[130] Kashlinsky A., Atrio-Barandela F., Ebeling H., Edge A., Kocevski D., 2010, ApJ, 712, L81

[131] Katz N., Weinberg D. H., Hernquist L., 1996, ApJS, 105, 19.

[132] Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999, MNRAS, 303, 188

[133] Kauffmann G. et al., 2003, MNRAS, 341, 33

[134] Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2

[135] Kereš D., Katz N., Fardal M., Davé R., Weinberg D. H., 2009, MNRAS, 395, 160

[136] Khandai N., Di Matteo T., Croft R., Wilkins S. M., Feng Y., Tucker E., DeGraf C., Liu M. S., 2015, MNRAS, 450, 1349

[137] Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102

[138] Komatsu E. et al., 2011, ApJS, 192, 18

[139] Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35

[140] Kravtsov A. V., Gnedin O. Y., Klypin A. A., 2004, ApJ, 609, 482

[141] Kravtsov A. V., 2013, ApJL, 764, L31

[142] Kroupa P., 2001, MNRAS, 322, 231

[143] Kroupa P., Theis C., Boily C. M., 2005, A&A, 431, 517

[144] Lan T.-W., Ménard B., Mo H. J., 2016, MNRAS, 459, 3998

[145] Lavaux G., Hudson M. J., 2011, MNRAS, 416, 2840

[146] Leauthaud A. et al., 2012, ApJ, 744, 159

[147] Le Brun A. M. C., McCarthy I. G., Schaye J., Ponman T. J., 2014, MNRAS, 441, 1270

[148] Le Brun A. M. C., McCarthy I. G., Melin J.-B., 2015, MNRAS, 451, 3868

[149] Lee K.-S. et al., 2012, ApJ, 752, 66

[150] Li Y., Mo H. J., van den Bosch F.C., Lin W. P., 2007, MNRAS, 379, 689

[151] Li Y., Mo H. J., Gao L., 2008, MNRAS, 389, 1419

[152] Li C., White S. D. M., 2009, MNRAS, 398, 2177

[153] Li C., Jing Y. P., Faltenbacher A., Wang J., 2013, ApJ, 770, L12

[154] Lim S. H., Mo H. J., Wang H., Yang X., 2016, MNRAS, 455, 499

[155] Lim S. H., Mo H. J., Lan T.-W., Ménard B., 2017, MNRAS, 464, 3256

[156] Lim S. H., Mo H. J., Lu Y., Wang H., Yang X., 2017, MNRAS, 470, 2982

[157] Lim S. H., Mo H. J., Wang H., Yang X., 2017, arXiv:1712.08619

[158] Lim S. H., Mo H. J., Li R., Liu Y., Ma Y.-Z., Wang H., Yang X., 2018, ApJ, 854, 181

[159] Lim S. H., Mo H. J., Wang H., Yang X., 2018, ApJ, 480, 4017

[160] Li M., Angulo R. E., White S. D. M., Jasche J., 2014, MNRAS, 443, 2311

[161] Lu Y., Mo H. J., 2007, MNRAS, 377, 617

[162] Lu Y., Kereš D., Katz N., Mo H. J., Fardal M., Weinberg, M. D., 2011, MNRAS, 416, 660

[163] Lu Y., Mo H. J., Katz N., Weinberg M. D., 2012, MNRAS, 421, 1779

[164] Lu Y., Mo H. J., Lu Z., Katz N., Weinberg M. D., 2014, MNRAS, 443, 1252

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

[166] Lu Z., Mo H. J., Lu Y., 2015, MNRAS, 450, 606

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

[168] Lu Y., Yang X., Shi F., Mo H. J., Tweed D., Wang H., Zhang Y., Li S., Lim S. H., 2016, ApJ, 832, 39

[169] Luo A.-L. et al., 2015, RAA, 15, 1095

[170] Ma Y.-Z., Van Waerbeke L., Hinshaw G., Hojjati A., Scott D., Zuntz J., 2015, JCAP, 9, 046

[171] MacGillivray H. T., Dodd R. J., McNally B. V., Corwin H. G. Jr, 1982, MNRAS, 198, 605

[172] Majewski S. R., 1994, ApJ, 431, L17

[173] Mandelbaum R. et al., 2006, MNRAS, 372, 758

[174] Mantz A. B., Allen S. W., Morris R. G., Simionescu A., Urban O., Werner N., Zhuravleva I., 2017, MNRAS, 472, 2877

[175] Marchesini D., van Dokkum P. G., Förster Schreiber N. M., Franx M., Labbé I., Wuyts S., 2009, ApJ, 701, 1765

[176] Marinacci F. et al., 2018, MNRAS, 480, 5113

[177] Martin D. C. et al., 2005, ApJL, 619, L1

[178] Martin C. L. et al., 2012, ApJ, 760, 127

[179] McAlpine S. et al., 2016, Astronomy and Computing, 15, 72

[180] McCarthy I. G. et al., 2010, MNRAS, 406, 822

[181] McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, MNRAS, 465, 2936

[182] McConnell N. J., Ma C. P., Gebhardt K., Wright S. A., Murphy J. D., Lauer T. R., Graham J. R., Richstone D. O., 2011, Nature, 480, 215

[183] McDonald M. et al., 2016, ApJ, 826, 124

[184] McIntosh D. H., Bell E. F., Weinberg M. D., Katz N., 2006, MNRAS, 373, 1321

[185] Melin J.-B., Bartlett J. G., Delabrouille J., 2005, A&A, 429, 417

[186] Melin J.-B., Bartlett J. G., Delabrouille J., 2006, A&A, 459, 341

[187] Mo H. J., White S. D. M., 1996, MNRAS, 282, 347

[188] Mo H. J., Mao S., White S. D. M., 1998, MNRAS, 295, 319

[189] Mo H. J., Mao S., White S. D. M., 1999, MNRAS, 304, 175

[190] Mo H. J., Mao S., 2002, MNRAS, 333, 768

[191] Mo H. J., Mao S., 2004, MNRAS, 353, 829

[192] Mo H. J., Yang X., van den Bosch F. C., Katz N., 2005, MNRAS, 363, 1155

[193] Mo H. J., van den Bosch F. C., White S. D. M., 2010, Galaxy Formation and Evolution. Cambridge University Press, New York, NY

[194] Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, ApJ, 710, 903

[195] Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121

[196] Moustakas J. et al., 2013, ApJ, 767, 50

[197] Muratov A. L., Kereš D., Faucher-Giguère C.-A., Hopkins P. F., Quataert E., Murray N., 2015, MNRAS, 454, 2691

[198] Muratov A. L. et al., 2017, MNRAS, 468, 4170

[199] Naiman J. P. et al., 2018, MNRAS, 477, 1206

[200] Nath B. B., Silk J., 2001, MNRAS, 327, 5

[201] Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

[202] Neistein E., Weinmann S. M., Li C., Boylan-Kolchin M., 2011, MNRAS, 414, 1405

[203] Nelson D., Genel S., Vogelsberger M., Springel V., Sijacki D., Torrey P., Hernquist L., 2015, MNRAS 448, 59

[204] Nelson D., et al., 2018, MNRAS, 477, 450

[205] Neto A. F. et al., 2007, MNRAS, 381, 1450

[206] Newman S. F. et al., 2012, ApJ, 761, 43

[207] Okamoto T., Eke V. R., Frenk C. S., Jenkins A., 2005, MNRAS, 363, 1299

[208] Okumura T., Jing Y. P., Li C., 2009, ApJ, 694, 214

[209] Olive K. A., Steigman G., Walker T. P., 2000, Physics Reports, 333, 389

[210] Oppenheimer B. D., 2018, MNRAS, 480, 2963

[211] Pakmor R., Hachinger S., Röopke F. K., Hillebrandt W., 2011, A&A, 528, 117

[212] Panter B., Jimenez R., Heavens A. F., Charlot S., 2007, MNRAS, 378, 1550

[213] Parkinson H., Cole S., Helly J., 2008, MNRAS, 383, 557

[214] Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144

[215] Pérez-González P. G. et al., 2008, ApJ, 675, 234

[216] Pillepich A., et al., 2018, MNRAS 473, 4077

[217] Plagge T. et al., 2010, ApJ, 716, 1118

[218] Planck Collaboration I., 2011, A&A, 536, 1

[219] Planck Collaboration V., 2013, A&A, 550, 131

[220] Planck Collaboration XI., 2013, A&A, 557, 52

[221] Planck Collaboration XVI, 2014, A&A, 571, 16

[222] Planck Collaboration XIII., 2016, A&A, 594, 13

[223] Planck Collaboration XXII., 2016, A&A, 594, 11

[224] Planck Collaboration XXXVII., 2016, A&A, 586, 140

[225] Planck Collaboration LIII., 2018, A&A, 617, 48

[226] Poggianti B. M., 1997, A&AS, 122, 399

[227] Popesso P., Biviano A., Bhringer H., Romaniello M., 2006, A&A, 445, 29

[228] Postman M., Geller M. J., 1984, ApJ, 281, 95

[229] Pratt G. W., Croston J. H., Arnaud M., Böhringer H., 2009, A&A, 498, 361

[230] Prochaska J. X. et al., 2017, ApJ, 837, 169

[231] Rasera Y., Teyssier R., 2006, A&A, 445, 1

[232] Remazeilles M. et al., 2011, MNRAS, 418, 467

[233] Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, ApJ, 771, 30

[234] Rodríguez-Puebla A., Avila-Reese V., Yang X., Foucaud S., Drory N., Jing Y. P., 2015, ApJ, 799, 130

[235] Rong Y., Yi S., Zhang S., Tu H., 2015, MNRAS, 451, 2536

[236] Rosas-Guevara Y. M. et al., 2016, MNRAS, 462, 190

[237] Rubin K. H. R. et al., 2014, ApJ, 794, 156

[238] Rudie G. C., Steidel C. C., Pettini M., Trainor R. F., Strom A. L., Hummels C. B., Reddy N. A., Shapley A. E., 2019, arXiv:1903.00004

[239] Sales L., Lambas D. G., 2004, MNRAS, 348, 1236

[240] Salpeter E. E., 1955, ApJ, 121, 161

[241] Schaan E. et al., 2016, Physical Review D, 93, 082002

[242] Schaye J., Dalla Vecchia C., 2008, MNRAS, 383, 1210

[243] Schaye J. et al., 2010, MNRAS, 402, 1536

[244] Schaye J. et al., 2015, MNRAS 446, 521

[245] Schlegel D. J., Finkbeiner D. P., Davis, M., 1998, ApJ, 500, 525

[246] Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, ApJ, 546, 20

[247] Seljak U., 2000, MNRAS, 318, 203

[248] Sharp N. A., Lin D. N. C., White S. D. M., 1979, MNRAS, 187, 287

[249] Shen J., Abel T., Mo H. J., Sheth R. K., 2006, ApJ, 645, 783

[250] Shen S. Y. et al., 2016, RAA, 16, 43

[251] Sheth R. K., Mo H. J., Tormen, G., 2001, MNRAS, 323, 1

[252] Shi J., Wang H., Mo H. J., 2015, ApJ, 807, 37

[253] Sijacki D., Springel V., Di Matteo T., Hernquist L., 2007, MNRAS, 380, 877

[254] Sijacki D., Vogelsberger M., Genel S., Springel V., Torrey P., Snyder G. F., Nelson D., Hernquist L., 2015, MNRAS, 452, 575

[255] Simard L., Mendel J. T., Patton D. R., Ellison S. L., McConnachie A. W., 2011, ApJS, 196, 11

[256] Skilling J., 2006, Bayesian Analysis, 1, 833

[257] Skrutskie M. F. et al., 2006, AJ, 131, 1163

[258] Smargon A., Mandelbaum R., Bahcall N., Niederste-Ostholt M., 2012, MNRAS, 423, 856

[259] Smit R., Bouwens R. J., Franx M., Illingworth G. D., Labbe I., Oesch P. A., van Dokkum P. G., 2012, ApJ, 756, 14S

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

[261] Somerville R. S., Davé R., 2015, ARAA, 53, 51

[262] Song M. et al., 2016, ApJ, 825, 5

[263] Springel V., Hernquist L., 2003, MNRAS, 339, 289

[264] Springel V., 2005, MNRAS, 364, 1105

[265] Springel V. et al., 2005, Nature, 435, 629

[266] Springel V., 2010, MNRAS, 401, 791

[267] Springel V. et al., 2018, MNRAS, 475, 676

[268] Steidel C. C., Erb D. K., Shapley A. E., Pettini M., Reddy N., Bogosavljević M., Rudie G. C., Rakic O., 2010, ApJ, 717, 289

[269] Stocke, J. T. et al., 2013, ApJ, 763, 148

[270] Sunyaev R. A., Zeldovich Y. B., 1972, Comments on Astrophysics and Space Physics, 4, 173

[271] Sun M., Voit G. M., Donahue M., Jones C., Forman W., Vikhlinin A., 2009, ApJ, 693, 1142

[272] Tanimura H., Hinshaw G., McCarthy I. G., Van Waerbeke L., Ma Y.-Z., Mead A., Hojjati A., Tröster T., 2018, MNRAS, 483, 223

[273] Tauber J. A. et al., 2010, A&A, 520, 1

[274] Tinker J. L., Kravtsov A. V., Klypin A., Abazajian K., Warren M. S., Yepes G., Gottlober S., Holz D. E., 2008, ApJ, 688, 709

[275] Torrey P., Vogelsberger M., Genel S., Sijacki D., Springel V., Hernquist L., 2014, MNRAS, 438, 1985

[276] Tremonti C. et al., 2004, ApJ, 613, 898

[277] Tumlinson J., Peeples M. S., Werk J. K., 2017, ARAA, 55, 389

[278] Tully R. B., Rizzi L., Shaya E. J., Courtois H. M., Makarov D. I., Jacobs B. A., 2009, AJ, 138, 323

[279] Tully R. B., 2015, AJ, 149, 171

[280] Vale A., Ostriker J. P., 2004, MNRAS, 353, 189

[281] Vale A., Ostriker J. P., 2006, MNRAS, 371, 1173

[282] van de Voort F. et al., 2016, MNRAS, 463, 4533

[283] van de Voort F., Springel V., Mandelker N., van den Bosch F. C., Pakmor R., 2019, MNRAS, 482, 85

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

[285] van den Bosch F. C., Yang X., Mo H. J., Weinmann S. M., Macciò A. V., More S., Cacciato M., Skibba R., Kang X., 2007, MNRAS, 376, 841

[286] van den Bosch F. C., Jiang F., Hearin A., Campbell D., Watson D., Padmanabhan N., 2014, MNRAS, 445, 1713

[287] Van Waerbeke L., Hinshaw G., Murray N., 2014, Physical Review D, 89, 023508

[288] Vikram V., Lidz A., Jain B., 2017, MNRAS, 467, 2315

[289] Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, MNRAS, 436, 3031

[290] Vogelsberger M. et al., 2014, MNRAS, 444, 1518

[291] Vogelsberger M. et al., 2014, Nature, 509, 177

[292] Wang L., Li C., Kauffmann G., De Lucia G., 2006, MNRAS, 371, 537

[293] Wang L., Jing Y. P., 2010, MNRAS, 402, 1796

[294] Wang H., Mo H. J., Guo Y., van den Bosch F. C., Yang X., 2009, MNRAS, 394, 398

[295] Wang H., Mo H. J., Jing Y. P., Yang X., Wang Y., 2011, MNRAS, 413, 1973

[296] Wang H., Mo H. J., Yang X., van den Bosch F. C., 2012, MNRAS, 420, 1809

[297] Wang H., Mo H. J., Yang X., van den Bosch F. C., 2013, ApJ, 772, 63

[298] Wang H., Mo H. J., Yang X., Jing Y. P., Lin W. P., 2014, ApJ, 794, 94

[299] Wang H. et al., 2016, ApJ, 831, 164

[300] Wang Y., Rooney C., Feldman H. A., Watkins R., 2018, MNRAS, 480, 5332

[301] Watson D. F., Berlind A. A., McBride C. K., Hogg D. W., Jiang T., 2012, ApJ, 749, 83

[302] Watson D. F., Hearin A. P., Berlind A. A., Becker M. R., Behroozi P. S., Skibba R. A., Reyes R., Zentner A. R., van den Bosch F. C., 2015, MNRAS, 446, 651

[303] Weinberger R. et al., 2017, MNRAS, 465, 3291

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

[305] Weinmann S. M., Kauffmann, G., van den Bosch F. C., Pasquali A., McIntosh D. H., Mo H. J., Yang X., Guo Y., 2009, MNRAS, 394, 1213

[306] Weisz D. R. et al., 2011, ApJ, 739, 5

[307] Werk J. K. et al., 2014, ApJ, 792, 8

[308] Werk J. K. et al., 2016, ApJ, 833, 54

[309] Werner M. W. et al., 2004, ApJS, 154, 1

[310] Werner N., Urban O., Simionescu A., Allen S. W., 2013, Nature, 502, 656

[311] White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

[312] White S. D. M., Frenk C. S., 1991, ApJ, 379, 52

[313] White M., 2001, MNRAS, 321, 1

[314] Wiersma R. P. C., Schaye J., Smith B. D., 2009a, MNRAS, 393, 99

[315] Willett K. W. et al., 2013, MNRAS, 435, 2835

[316] Wright E. L. et al., 2010, AJ, 140, 1868

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

[318] Yang X., Mo H. J., Jing Y. P., van den Bosch F. C., Chu Y. Q., 2004, MNRAS, 350, 1153

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

[320] Yang X., van den Bosch F. C., Mo H. J., Mao S., Kang X., Weinmann S. M., Guo Y., Jing Y. P., 2006, MNRAS, 369, 1293

[321] Yang X., Mo H. J., van den Bosch F. C., Pasquali A., Li C., Barden M., 2007, ApJ, 671, 153

[322] Yang X., Mo H. J., van den Bosch F. C., 2008, MNRAS, 676, 248

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

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

[325] Zaritsky D., Smith R., Frenk C. S., White S. D. M., 1997, ApJ, 478, 53

[326] Zehavi I. et al., 2004, ApJ, 608, 16

[327] Zehavi I. et al., 2011, ApJ, 736, 59

[328] Zhang Y., Yang X., Wang H., Wang L., Mo H. J., van den Bosch F. C., 2013, ApJ, 779, 160

[329] Zhao D. H., Mo H. J., Jing Y. P., Börner G., 2003, MNRAS, 339, 12

[330] Zhao D. H., Jing Y. P., Mo H. J., Börner G., 2009, ApJ, 707, 354

[331] Zheng Z., Coil A. L., Zehavi I., 2007, ApJ, 667, 760

[332] Zheng X. Z., Bell E. F., Papovich C., Wolf C., Meisenheimer K., Rix H.-W., Rieke G. H., Somerville R., 2007, ApJ, 661, L41

[333] Zhao D. H., Jing Y. P., Mo H. J., Börner G., 2009, ApJ, 707, 354

[334] Zwicky F., Herzog E., Wild P., Karpowicz M., Kowal C. T., 1961-1968, Catalogue of Galaxies and Clusters of Galaxies (Pasadena: California Institute of Technology)