

© 2018 by Chun Yin Ricky Chue. All rights reserved.

THE STRUCTURE AND SPATIAL DISTRIBUTION OF DARK MATTER HALOS

BY

CHUN YIN RICKY CHUE

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Astronomy
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Professor Brian Fields, Chair
Associate Professor Neal Dalal, Director of Research
Professor Athol Kembell
Professor Paul Ricker

Abstract

This is a study of how the distribution and properties of dark matter halos can be utilized as probes of fundamental questions in modern cosmology. Spatial clustering of dark matter halo carry a wealth of information regarding its evolution history and environment. In particular, I study halo assembly bias, which refers to the assembly history dependence of spatial clustering for dark matter halos at fixed mass, using observational and cosmological simulation data. Understanding and modeling assembly bias provides insight into the context of hierarchical structure formation theory. Apart from spatial clustering I also study shapes of dark matter halos and how to measure them using three-point galaxy statistics, which itself serves as an astrophysical constraint on properties of dark matter.

To my family and friends... For supporting and believing in me, always.

Acknowledgments

At the outset I would like to acknowledge those without whose support and company this journey would not have been nearly as enriching.

Firstly, I would like to give my sincere thanks to my advisor Neal Dalal, for his patience, constant support and confidence in me. I am glad to have met him during my formative years as a researcher. I highly value his knowledge, opinion and fervor for science. His guide is always clear with mathematical interpretation and physical insight. His advice not only guided me through the research, but also shaped my views and thinking. From my advisor, I learnt the way to convey difficult ideas and to work with collaborators. Also, I learnt to be humble. These things are going to be valuable in the rest of my life. I really appreciate very much for his guidance and support in these years.

Secondly, I thank the inspirational discussion with Susmita Adhikari, Arka Banerjee, Wenjuan Fang and Di Wen. I benefited a lot from their deep knowledge and friendly comments.

Next, I would also like to thank wholeheartedly the Department of Astronomy at UIUC and the support I have received from them throughout my stay in UIUC. Special thanks to my thesis committee members Prof. Brian Fields, Prof. Paul Ricker and Prof. Athol Kemball, for their valuable advice on my research work and career development.

I also take this opportunity to thank the Croucher Foundation in Hong Kong for providing support and funding to my scientific research. With the Croucher Scholarships for Doctoral Study, I had an opportunity of visiting the Institute for Advanced Study in Princeton University to share my research results with the top scientists and obtain valuable scientific advice.

I would also want to thank sincerely to Tak-Kei Lui, Ka-Ho Lo, Yat Sen Wong, Kwok Yan

Chan, Ka Wai Lo, Chi Hang Boyce Tsang, Kin Lam, Pak On Chan, Anthony Fan, Mervyn Cheng, Kwok Sun Tang, Lok Hei Ng, Erica Lai, Anatoly Pavlov, Alfred Tang, Sandra Braun, Julianna Cook and her adorable cat Luna, Fergus Chau, Andriey Chan, Tommy Chow, Michael Tang, Steve Wong, Jeffrey Jim, Hiu Yan Ha, Eddy Cheung, Jackie Ma, Sami Cheong, Wai-Ting Liao, Monica Huang, Yinghe Lü, Xilu Wang, Yixian Cao, Yiran Wang, Matias Carrasco-Kind and many other sincere friends in Hong Kong and Urbana-Champaign. We shared so many wonderful memories, and they have made my PhD life so much enjoyable.

Moreover, I would like to express my grateful thanks my younger sister Yin Ki Chue, my parents Pak Kui Chue and Chui Fun Lau, my grandparents Hon Chung Lau and Yin Tai Liu for their care and understanding during the tough period. I would not have got over the stress and hard times in PhD life without their support and love.

Lastly but importantly I would like to thank and acknowledge this country and her people for making this place a home away from home. Nothing creative can be accomplished in an environment that is inhospitable to diversity. I am glad to have met the beautiful people in this country that have made my stay here memorable and worthy of cherishing for a lifetime.

Table of Contents

List of Figures	viii
Chapter 1 Introduction	1
1.1 Gravitational instability and structure formation	3
1.2 Halo collapse model	7
1.3 Mass profile of dark matter halos	11
1.4 Shapes of dark matter halos	12
1.5 Spatial distribution of dark matter halos	13
1.6 Gravitational lensing - an astrophysical probe of dark matter halos	16
Chapter 2 Three-point galaxy-galaxy lensing as a probe of dark matter halo shapes	17
2.1 Introduction	17
2.2 Mass model	19
2.2.1 Three-Point Estimator	24
2.3 Results	27
2.3.1 Comparison with N-body simulations	29
2.3.2 Shape noise	34
2.3.3 Other systematics	36
2.4 Discussion	38
Chapter 3 Assembly bias in CFHTLenS	40
3.1 Introduction	40
3.2 Data	42
3.2.1 Lens and source selection	43
3.2.2 Satellite galaxy rejection	43
3.2.3 Selection of red and blue galaxy samples	45
3.2.4 Simulation data	47
3.3 Methodology	48
3.3.1 Galaxy-galaxy lensing	49
3.3.2 Finding consistent mass profiles	53
3.3.3 Correlation function measurements	55
3.4 Results	59
3.5 Discussion	61

Chapter 4	Some assembly required: assembly bias in massive dark matter halos	63
4.1	Motivation	63
4.2	Introduction	64
4.3	Simulation data	66
4.4	Analysis work and results	67
Chapter 5	Conclusions and Outlook	79
5.1	Probing dark matter halo shapes	79
5.2	Assembly bias of dark matter halos	81
5.2.1	Observational efforts of assembly bias in galaxy-sized halos	81
5.2.2	Revisiting assembly bias in cluster-sized halos	82
References		84
Appendix A	Chapter 3: Assembly bias in CFHTLenS	100
A.1	Effect of applying various satellite exclusion on the 2PCFs of the mocks	100
A.2	Effect of applying various satellite exclusion on CFHTLenS results	100

List of Figures

2.1	(a): Plot of the multipole moments of stacked halos. The solid curves show the isotropic component (monopole, κ_0) of the surface density profile and the dashed curves show the $\cos 2\theta$ component (quadrupole, κ_2). The convergence κ is proportional to surface density Σ . (b): Radial dependence of ellipticity, which we define as $\varepsilon(r) \equiv \kappa_2/(\eta \kappa_0)$, for three different mass bins. The blue, red and green colors correspond to three different mass bins. Note that, although the multipole moments vary by orders of magnitude, the ellipticity remains nearly constant across much of the range of interest.	20
2.2	Lens-shear-shear three point correlation function. We correlate the shear at sky positions $\vec{r}_1$ and $\vec{r}_2$ relative to foreground lens galaxies, and construct an estimator summing all such triangular configurations in the sky.	25
2.3	The top panels show the comparison between the average ellipticities of halos in various mass bins (solid curve), compared to the ellipticity determined from the 3-point estimator (dashed line). The bottom panel shows the local slope of the isotropic component (monopole) of the halos. The red curve corresponds to the smoothed slope of the 3D profile, and the green curve corresponds to the local slope of the projected 2D profile. The dashed curves show the slope of the NFW profile in 2D and 3D for reference. These slopes were measured from the stacked profiles after they were smoothed using a 6th order Savitzky-Golay filter over 17 nearest bins [1]. In the low mass bins, we observe significant departures from NFW slopes at large radii, possibly indicating the effects of nearby halos.	28
2.4	Example of line of sight projection effects. The halo at the origin has massive neighbors projecting nearby, which generate a large quadrupole moment that is unrelated to the halo's own ellipticity. Colors correspond to convergence κ , for a lens redshift of $z_l = 0.3$ and source redshift of $z_s = 0.5$	30
2.5	Effect of including satellites (subhalos) in our sample. The black solid curve shows the average ellipticity of halos estimated from N-body simulations in the mass bin $10^{12.4} - 10^{12.6} M_\odot h^{-1}$ using central galaxies only. The dashed lines show the ellipticity determined by the three-point estimator, applied to samples with various degrees of contamination by satellites. If satellites are not excluded (red line), the estimated ellipticity is significantly biased. Removing the satellites in the most dense environments eliminates much of the bias.	32

3.1	The stacked tangential shear profiles $\langle \Delta\Sigma(r) \rangle$ for mock galaxies, after the M_* stellar mass ranges have been adjusted to make the shear profiles consistent within $200h^{-1}\text{kpc}$. Left: $\langle \Delta\Sigma(r) \rangle$ for samples including both centrals and satellites. Right: $\langle \Delta\Sigma(r) \rangle$ for only central galaxies, i.e. with all satellites removed. The red (blue) data points represent the red (blue) sample.	51
3.2	These line plots present the fraction of mock halos $P(M)$ in mass bin $(M, M + dM)$ for the galaxy samples with consistent lensing profiles shown in Fig. 3.1. Red and blue lines represent the red and blue galaxies respectively. Solid and dashed lines represent the mass distributions for the subhalos and the host halos respectively. Left: $P(M)$ for all mock galaxies, the significant proportion at $M \geq 10^{12.5}h^{-1}M_\odot$ indicates there are satellites in the samples which are located in massive host halos. Right: $P(M)$ for only central galaxies. By construction, the halos in these samples are all host halos. After satellite removal, the red and blue samples have consistent halo mass distributions.	52
3.3	The 2PCFs of the mocks after pairing the lensing profiles of the red and blue samples. The red (blue) dots represent the red (blue) sample of the “All galaxies” sample, while the magenta (cyan) triangles represent the red (blue) sample of the “Central only” sample. The stellar mass bins $(\log_{10}(M_*/M_\odot))$ chosen in the mock have similar widths as those chosen in CFHTLenS data. Left: The 2PCFs with no satellite exclusion. The stellar mass ranges of the blue “All galaxies” and “Centrals only” samples are $[9.925, 10.925]$ and $[9.943, 10.943]$ respectively, whereas both of the red samples are kept at $[9.8, 10.8]$. The severe satellite contamination $f_{\text{sat,init.}} = 25.4\%$ can mimic the assembly bias signal, as demonstrated by the larger difference in 2PCFs of the red and blue galaxies for the “All galaxies” sample than those for the “Centrals only” sample. Right: The 2PCFs with the most stringent satellite exclusion adopted in this paper. The stellar mass ranges of the blue “All galaxies” and “Centrals only” samples are $[9.935, 10.935]$ and $[9.928, 10.928]$ respectively, whereas both of the red samples are kept at $[9.8, 10.8]$. Under this exclusion criterion, we have excluded $f_{\text{sat,excl.}} = 80.8\%$ of the satellites, and left with satellite contamination $f_{\text{sat}} = 9.1\%$. Nonetheless, the consistency between the 2PCFs of the “All galaxies” and “Centrals only” samples suggests that the residual contamination present in the sample is insufficient to solely account for the difference in the bias signals in the mock.	54
3.4	The stacked shear profiles of CFHTLenS data with the most stringent satellite exclusion adopted, i.e. $A = 1.5$, and if there is any neighboring galaxy whose stellar mass is at least 3 times more massive than the lens galaxy within $500h_{70}^{-1}\text{kpc}$ projected distance. The red (blue) dots represent the red (blue) sample. The stellar mass bins have been adjusted to allow the lensing profiles of the red and blue samples to be consistent by minimizing $\Delta\chi^2_S$. There are 4376 and 8206 galaxies in the red and blue samples respectively, and the stellar mass ranges $(\log_{10}(M_*/M_\odot))$ for the red and blue samples are $[9.05, 10.05]$ and $[9.5, 10.55]$ respectively. The results of $\Delta\chi^2_S$ minimization are shown in Table A.1.	57

3.5	Left: The 2PCFs of CFHTLenS with the following satellite exclusion criterion applied: $A = 1.5$, and if there is any neighboring galaxy whose stellar mass is at least 3 times more massive than the lens galaxy within $500h_{70}^{-1}\text{kpc}$ projected distance. The red (blue) dots represent the red (blue) sample. Right: The difference of the 2PCFs $\Delta w(\theta) \equiv w^{\text{red}}(\theta) - w^{\text{blue}}(\theta)$ corresponding to the left panel at $3 - 11h_{70}^{-1}\text{Mpc}$. The dashed lines show the $\Delta w(\theta) = 0$ level. The significance of difference of the 2PCFs is 1.67σ . The physical distance at $z = 0.3$ is shown at the top abscissae of the panels.	58
4.1	The different curves show stacked mass accretion histories for subsets of BigMDPL halos with $M_{\text{vir}} = 0.7 - 1 \times 10^{15}h^{-1}M_{\odot}$. The curves corresponds to the halos with the highest 25% of $a_{0.5}$ (brown), the lowest 25% of $a_{0.5}$ (green), the top 25% of α_g (red), and bottom 25% of α_g (blue). The width of each curve corresponds to the $1 - \sigma$ jackknife uncertainty on the mean MAH. As discussed in the text, α_g is a better indicator of large-scale bias than $a_{0.5}$, and it tends to split the halos more strongly on their early assembly histories ($f < 0.5$) for this mass range.	70
4.2	Plotted is the cross-correlation coefficient between the linear overdensity, smoothed on a scale of $30 h^{-1} \text{ Mpc}$, and the time when the halo acquires fraction f of its present-day mass, denoted a_f . The blue curve corresponds to high-mass halos in the mass range $M_{\text{vir}} = 1 - 2 \times 10^{15}h^{-1}M_{\odot}$, while the red curve is for low-mass halos with $M_{\text{vir}} = 1 - 2 \times 10^{14}h^{-1}M_{\odot}$. The width of each band corresponds to the $1 - \sigma$ uncertainty, determined by jackknife.	72
4.3	BigMDPL halos with $M_{\text{vir}} = 1 - 1.1 \times 10^{14}h^{-1}M_{\odot}$. Plotted are the stacked (average) MAHs for isolated halos in the top (red) and bottom (blue) quartiles of α_{opt} as defined in Eq. (4.4). The width of each curve corresponds to the $1 - \sigma$ jackknife uncertainty on the mean MAH. We have used a narrow mass bin in order to enforce that both subsets have the same average mass at $z = 0$	74
4.4	Phase space diagrams around BigMDPL halos with $M_{\text{vir}} = 1 - 1.2 \times 10^{14}h^{-1}M_{\odot}$. The color corresponds to the mass in neighbors (in units of $h^{-1}M_{\odot}$) at each pixel in the space of r and v_r . We have used a somewhat wider mass bin than in Fig. 4.3 in order to improve the statistical uncertainties. The top panel is for the low α_{opt} , corresponding to the blue curve in Fig. 4.3, while the bottom panel is for high α_{opt} , corresponding to the red curve in Fig. 4.3. For this mass range, r_{vir} is depicted by the vertical dotted white line at slightly less than $1 h^{-1} \text{ Mpc}$. Comparing the two panels, we can see that the sample in the lower panel has a significant amount of bound, virialized mass near splashback located just outside r_{vir}	76

- A.1 The 2PCFs of the mock galaxies with different satellite removal criteria. The red (blue) dots represent the red (blue) “All galaxies” sample respectively. The magenta (cyan) triangles represent the red (blue) “Centrals only” sample respectively. The stellar mass bins ($\log_{10}(M_*/M_\odot)$) chosen for the mock data have similar widths as those chosen in CFHTLenS data. In all cases, the stellar mass bins of both “All galaxies” and “Centrals only” red samples are kept at $[9.8, 10.8]$. (Left panels: 1 & 2): Keeping $A = 1.5$ and $M_{*,\text{nbr}}/M_{*,\text{lens}} \geq 3$ fixed, with $r_{\text{excl.}}$ varied between $300 - 400h^{-1}\text{kpc}$. In (1), the stellar mass ranges of the blue “All galaxies” and “Centrals only” samples are $[9.935, 10.935]$ and $[9.945, 10.945]$ respectively. In (2), the corresponding ranges are $[9.94, 10.94]$ and $[9.928, 10.928]$ respectively. (Middle panels: 3 & 4): Keeping $r_{\text{excl.}} = 500h^{-1}\text{kpc}$ and $M_{*,\text{nbr}}/M_{*,\text{lens}} \geq 3$ fixed, with A varied between 1.0 and 1.2. In (3), the stellar mass ranges of the blue “All galaxies” and “Centrals only” samples are $[9.928, 10.928]$ and $[9.928, 10.928]$ respectively. In (4), the corresponding ranges are $[9.928, 10.928]$ and $[9.925, 10.925]$ respectively. (Right panel: 5): Keeping $A = 1.5$ and $r_{\text{excl.}} = 500h^{-1}\text{kpc}$ fixed, with $M_{*,\text{nbr}}/M_{*,\text{lens}} \geq 5$, the stellar mass ranges of the blue “All galaxies” and “Centrals only” samples are $[9.945, 10.945]$ and $[9.938, 10.938]$ respectively. . . . 102
- A.2 Pairing up the tangential shear profiles of the CFHTLenS galaxies with different satellite removal criteria by minimizing $\Delta\chi^2_S$. (Left panels: 1 & 2): Keeping $A = 1.5$ and $M_{*,\text{nbr}}/M_{*,\text{lens}} \geq 3$ fixed, with $r_{\text{excl.}}$ varied between $300 - 400h_{70}^{-1}\text{kpc}$. (Middle panels: 3 & 4): Keeping $r_{\text{excl.}} = 500h_{70}^{-1}\text{kpc}$ and $M_{*,\text{nbr}}/M_{*,\text{lens}} \geq 3$ fixed, with A varied between 1.0 and 1.2. (Right panel: 5): Keeping $A = 1.5$ and $r_{\text{excl.}} = 500h_{70}^{-1}\text{kpc}$ fixed, with $M_{*,\text{nbr}}/M_{*,\text{lens}} \geq 5$. The red and blue data points represent the red and blue samples respectively. About 80% – 90% of the galaxies are excluded in different removal schemes. The results of pairing up lensing profiles are shown in Table A.1. 103
- A.3 The 2PCFs of the CFHTLenS galaxies with different satellite removal criteria after having the lensing profiles paired up. (Left panels: 1 & 2): Keeping $A = 1.5$ and $M_{*,\text{nbr}}/M_{*,\text{lens}} \geq 3$ fixed, with $r_{\text{excl.}}$ varied between $300 - 400h_{70}^{-1}\text{kpc}$. (Middle panels: 3 & 4): Keeping $r_{\text{excl.}} = 500h_{70}^{-1}\text{kpc}$ and $M_{*,\text{nbr}}/M_{*,\text{lens}} \geq 3$ fixed, with A varied between 1.0 and 1.2. (Right panel: 5): Keeping $A = 1.5$ and $r_{\text{excl.}} = 500h_{70}^{-1}\text{kpc}$ fixed, with $M_{*,\text{nbr}}/M_{*,\text{lens}} \geq 5$. The red and blue data points represent the red and blue samples respectively. The physical distance at $z = 0.3$ is shown at the top abscissae of the panels. The statistics of comparing the 2PCFs is summarized in Table A.2. 105
- A.4 The differences of the 2PCFs of the CFHTLenS galaxies at $3 - 11h_{70}^{-1}\text{Mpc}$ with different satellite removal criteria after having the lensing profiles paired up. The panel numbers correspond to the panels in Fig. A.3. The physical distance at $z = 0.3$ is shown at the top abscissae of the panels. The dashed lines show the $\Delta w(\theta) = 0$ level. The statistics of comparing the 2PCFs difference is summarized in Table A.2. 106

Chapter 1

Introduction

The “standard model” of cosmology has emerged over the past few decades, named the Λ CDM model. This model has gained spectacular successes in describing numerous observed large-scale properties of the Universe [2–17]. By combining the Λ CDM model prediction with observational data from modern cosmological surveys, we now understand the Universe mainly consists of components far beyond our current knowledge. Surprisingly, the luminous baryonic matter that we are familiar with, only represents a tiny amount of the energy content of the Universe [2, 18]. To explain the phenomena in our Universe, we need two key physical ingredients, dark matter and dark energy [2, 18], whose very existence represent physics beyond the Standard Model of particle physics.

While little is known about dark energy at present, dark matter is believed to be composed of a new particle undiscovered in any terrestrial accelerators [19–22]. For instance, a popular candidate for dark matter is the lightest supersymmetric partner to Standard Model particles [19–23]. Elucidating the nature of dark matter requires measuring properties and behaviors of the dark matter particle, such as its mass, its self-interactions, or its spatial distribution in the Universe. While terrestrial experiments have placed stringent constraints on any interactions between dark matter and ordinary matter, astrophysical observations can additionally constrain physics of dark matter particles. For instance, self-interactions in the dark matter sector can produce significant effects in the formation and evolution of cosmic structure that are potentially observable [24–27]. On the other hand, studying the spatial distribution of dark matter allows us to infer the evolution of large scale structures [28–30], and hence unveil the mystery of the Universe.

Dark matter is ubiquitous in the Universe. Not only does it reveal evidence of its existence

through rotation curves of spiral galaxies [31–33], but its necessity has also been detected by investigating galaxy clusters and large scale structures [7, 11, 34]. In fact, dark matter plays a crucial role in structure formation of the early Universe. Measurements of the CMB power spectrum back to the time of recombination ($z \sim 1100$) tell us that the matter distribution in the early Universe was highly homogeneous, with only a tiny inhomogeneities of the order $\delta T/T \sim 10^{-5}$ [2, 3, 35]. Nevertheless, it is exactly these initial minuscule inhomogeneities that planted the seeds evolving to the cosmic structure, galaxies and clusters we observe today [28, 29, 36, 37]. However, theories and numerical simulations have showed that the amount of baryonic matter present today is not sufficient to explain the currently observable inhomogeneity of matter structure. It appears to be possible only with the existence of large amount of dark matter [38, 39]. One of the end-point products created from the collapse of dark matter fields is called dark matter halos, which are gravitationally bound and virialized structures that have decoupled from the Hubble expansion [30, 40]. These highly non-linear bound objects evolve independently from the background expansion of the Universe. The average overdensity of the matter field within these halos is almost 200 times larger than the mean background density of the Universe [41], making them the strong gravitational potential wells within which all known structure in the Universe, such as galaxies, stars, solar systems, planets form and evolve.

As dark matter accounts for a significant amount of components in our Universe, understanding the structure and distribution of dark matter halos can provide us insights into some of the outstanding fundamental problems in modern cosmology. In the rest of this chapter, we firstly outline in section 1.1 how dark matter plays a role in the initial density perturbation which then evolves to the non-linear structures we observe today. Next, we outline the model explaining the collapse and formation of dark matter halos in section 1.2. We then give an introduction on mass profile, shapes and spatial distribution of halos in sections 1.3, 1.4 and 1.5 respectively. Last but not least, we describe an observable astrophysical probe of dark matter halos - gravitational lensing in section 1.6.

1.1 Gravitational instability and structure formation

Structure formation can be treated as an initial value problem, i.e. given the initial conditions as seeds for the primordial density perturbations of different species, we then follow the time evolution of the densities and compare with observed measures of structure. In the following, we outline the essence of the calculation. See also [28, 42–44] for a more sophisticated treatment of the problem.

We consider only gravitational interaction and pressure force. The gravity comes from the overdensities which have extra gravitational attraction over the background universe. On the other hand, the pressure force comes from the thermal pressure of baryons, when photons exert radiation pressure on baryons before decoupling. The fluid dynamics is used for calculation when the underlying fields of different matter species are investigated. For small scales and in the weak gravity regime, using linear theory in the Newtonian approach would be sufficient. In the following discussion, we consider an expanding Universe.

We firstly introduce the Lagrangian and Eulerian coordinate choices. The Lagrangian coordinate $\vec{r}$ is time-independent but it expands in physical space. Next, the Eulerian coordinate is a time-independent grid $\vec{x}$ fixed in physical space, mathematically,

$$\vec{x}(t) = a(t)\vec{r} \tag{1.1}$$

$$\nabla_{\vec{x}} = \frac{1}{a}\nabla_{\vec{r}} \tag{1.2}$$

where ∇ stands for the gradient of the coordinate. Consider firstly the unperturbed density field $\rho(\vec{x}, t) = \rho_0(t)$, velocity field $\vec{v}(\vec{x}, t) = \vec{v}_0 = (\dot{a}/a)\vec{x} = \dot{a}\vec{r}$ and gravitational potential field

$\Phi(\vec{x}, t) = \Phi_0$. From mass conservation equation,

$$\partial_t \rho_0 + \nabla \cdot (\rho_0 \vec{v}) = \dot{\rho}_0 + \rho_0 \frac{\dot{a}}{a} \nabla_{\vec{x}} \cdot \vec{x} = 0 \quad (1.3)$$

$$\dot{\rho}_0 + 3 \frac{\dot{a}}{a} \rho_0 = 0 \quad (1.4)$$

$$\Rightarrow \rho \propto a^{-3} \quad (1.5)$$

Then, from the Poisson equation,

$$\nabla^2 \Phi_0 = \frac{1}{x^2} \partial_x (x \partial_x \Phi_0) = 4\pi G \rho_0 \quad (1.6)$$

$$\Rightarrow \Phi_0 = \frac{2\pi G \rho_0}{3} x^2 = \frac{2\pi G \rho_0}{3} a^2 r^2 \quad (1.7)$$

$$\nabla_{\vec{x}} \Phi_0 = \frac{4\pi G \rho_0}{3} \vec{x} \quad (1.8)$$

$$\nabla_{\vec{r}} \Phi_0 = \frac{4\pi G \rho_0}{3} a \vec{r} \quad (1.9)$$

Lastly, from the Euler equation,

$$\frac{d(\dot{a}\vec{r})}{dt} = \ddot{a}\vec{r} = \frac{\ddot{a}}{a} \vec{x} = -\frac{4\pi G \rho_0}{3} \vec{x} \quad (1.10)$$

$$\Rightarrow \frac{\ddot{a}}{a} = -\frac{4\pi G \rho_0}{3} \quad (1.11)$$

Hence, the usual Friedmann equations of an expanding Universe are derived in the above equations with non-relativistic approximation.

Next, we add linear perturbations to the physical fields, denoting the perturbed values with the subscript 1, i.e.

$$\rho = \rho_0 + \rho_1 = \rho_0 [1 + \delta(\vec{x})] \quad (1.12)$$

$$\vec{v} = \vec{v}_0 + \vec{v}_1 \quad (1.13)$$

$$\Phi = \Phi_0 + \Phi_1 \quad (1.14)$$

Plugging these in the mass conservation, Poisson and Euler equations, and keeping only linear terms in perturbations in Lagrangian coordinates, i.e. $\vec{x}(t) = a(t)\vec{r}(t)$, $\vec{v}_1(t) = a(t)\vec{u}(t)$ and $\nabla = \nabla_{\vec{r}}$,

$$\dot{\vec{u}} + 2\frac{\dot{a}}{a}\vec{u} = -\frac{1}{a^2}\nabla\Phi_1 - \frac{1}{a}\frac{\nabla\delta p}{\rho_0} \quad (1.15)$$

$$\dot{\delta} = -\nabla \cdot \vec{u} \quad (1.16)$$

In particular, the term $2(\dot{a}/a)\vec{u} = -2H\vec{u}$ is called the Hubble drag, which removes kinetic energy from collapsing objects, and allows the total energy to decrease with time.

From Eqns. 1.15, 1.16 and the Poisson equation, the linearized density evolution in Fourier space, i.e. $\delta(\vec{r}) \sim e^{-i\vec{k}\cdot\vec{r}}$ is given by,

$$\ddot{\delta}_k + 2\frac{\dot{a}}{a}\dot{\delta}_k = \left(4\pi G\rho_0 - \frac{c_s^2 k^2}{a^2}\right)\delta_k \quad (1.17)$$

where the adiabatic sound speed, $c_s^2 = \partial p / \partial \rho$. With the expansion of the universe, the Hubble drag $-2H\dot{\delta}_k$ is present which opposes density growth. The Jeans' scale is defined as $k_J = 2\pi/\lambda_J = \sqrt{4\pi G\rho a^2/c_s^2}$. Regions with length scales larger than Jeans' scale, i.e. ($\lambda > \lambda_J$, $k < k_J$), would collapse under gravity. On the other hand, regions smaller than Jeans' scale, i.e. ($\lambda < \lambda_J$, $k > k_J$) would be expected to oscillate due to balancing effect between pressure force and gravity.

Consider large scales $\lambda \gg \lambda_J$, Eqn. 1.17 would be reduced to:

$$\ddot{\delta}_k + 2\frac{\dot{a}}{a}\dot{\delta}_k \approx 4\pi G\rho_0\delta_k \quad (1.18)$$

For a matter-dominated Universe, the two roots of Eqn. 1.18 are found to follow power law solutions, with power indices $s = 2/3$ and -1 , which correspond to the growing and decaying modes, i.e.

$$\delta_+(t) = \delta_+(t_i) \left(\frac{t}{t_i}\right)^{2/3} ; \quad \delta_-(t) = \delta_-(t_i) \left(\frac{t}{t_i}\right)^{-1} \quad (1.19)$$

The growing mode dominates. Hence, the scales larger than the Jeans' scales are referred to the unstable modes, each unstable Fourier mode grows with time as $\delta_k(t) \propto D(t) \sim t^{2/3} \sim a$ in a matter-dominated Universe. Note that each wavenumber k grows independently by the same factor. In real space, the linear growth factor $D(t)$ relates the overdensities at initial time t_i and time t , i.e. $\delta(t) = D(t)\delta(t_i)$. Simply put, the entire overdensity pattern grows with the same amplification.

An application of the linear theory can be applied to the structure formation problem. If the Universe were made of baryonic matter only, we should expect oscillation modes for the overdensities with little growth before recombination due to the coupling effect between baryons and photons. However, after decoupling, photons were no longer intertwined with baryons and therefore the density fluctuation of the underlying matter field should grow.

Given that matter density $\rho \propto a^{-3} \propto T^3$, we have $\delta\rho/\rho = 3\delta T/T$. As the CMB observation shows that the temperature fluctuation at last scattering was $\delta T/T \sim 10^{-5}$ at $z \sim 1100$ [2, 3, 35], one can deduce the matter density fluctuation at last scattering, which was $\delta(z = 1100) \sim 3 \times 10^{-5}$. Hence, one would expect the size of fluctuations today to be:

$$\delta_0 = \frac{D_0}{D_{\text{ls}}} \delta_{\text{ls}} = \frac{a_0}{a_{\text{ls}}} \delta_{\text{ls}} = (1 + z_{\text{ls}}) \delta_{\text{ls}} \sim 0.03 \ll 1 \quad (1.20)$$

The perturbation would then be still well in the linear regime, implying no non-linear structures would have formed. On the contrary, we do observe non-linear structures today such as galaxies and clusters. Hence, a baryonic matter-only Universe model is apparently contradicting the observational evidence in the real Universe.

Hence, we further consider density perturbation due to dark matter. We assume cold dark matter is pressureless, i.e. $c_s = 0$. During the radiation dominated epoch, i.e. $\rho_r \gg \rho_m$, we write the Jeans' equation for dark matter in Eqn. 1.17 as:

$$\ddot{\delta}_m + 2\frac{\dot{a}}{a}\dot{\delta}_m \approx 0 \quad (1.21)$$

The solution comprises of a logarithmic growing mode and a t^{-1} dependence decaying mode, i.e.

$$\delta_m(t) = D(t)\delta_m(t_i) = \left(D_1 \ln t + \frac{D_2}{t}\right) \delta_m(t_i) \quad (1.22)$$

As the growing mode scales with $\ln t$, dark matter perturbation hardly grows in the radiation dominated era until the end of the radiation era, namely the matter-radiation equality ($z_{\text{eq}} \sim 3 \times 10^4$). After the transition, the Universe enters the matter-dominated era and dark matter grows as $t^{2/3}$. Since dark matter does not couple with photons, they could have grown earlier, and hence more than baryon does. With the presence of dark matter perturbation, the size of fluctuation today would be:

$$\delta_{m,0} = \frac{D_{\text{ls}}}{D_{\text{eq}}} \delta_{b,0} \sim \frac{1 + z_{\text{eq}}}{1 + z_{\text{ls}}} \delta_b \sim 30 \times 0.03 \sim 1 \quad (1.23)$$

Therefore, dark matter can grow to nonlinearity today. Hence, the existence of collapsed cosmic structures requires collisionless dark matter.

1.2 Halo collapse model

The spherical collapse model provides a standard way to understand halo formation. This model follows the evolution of a shell enclosing a top hat density perturbation in a spatially flat, matter-dominated Einstein de-Sitter universe (i.e. $\Omega_m = 1$). See [28, 42–44] for a more sophisticated treatment of the spherical collapse model.

In the spherical collapse model, the spherical density perturbation of radius r_0 evolves independently as a closed universe, with the mass enclosed by the perturbation shell given by $M = 4\pi r_0^3 \Omega_m \rho_c / 3$, where ρ_c is the critical density of the universe, r_0 is the initial radius of the shell and Ω_m is the fractional matter overdensity of the closed universe. Also, let θ be the development angle that goes from 0 to 2π . Consider the equation of motion of the shell, from Gauss' law, the force on the shell enclosing a spherically symmetric density depends only on the

mass enclosed within it, therefore,

$$\frac{d^2 r(t)}{dt^2} = -\frac{GM(< r(t))}{r(t)^2} \quad (1.24)$$

Solving the equation, the parametric solution is given by,

$$r(\theta) = A(1 - \cos \theta) \quad (1.25)$$

$$t(\theta) = B(\theta - \sin \theta) \quad (1.26)$$

Define the scaled conformal time η by $\theta = H_0 \eta (\Omega_m - 1)^{1/2}$, hence

$$A = \frac{r_0 \Omega_m}{2(\Omega_m - 1)} \quad (1.27)$$

$$B = \frac{\Omega_m}{2H_0(\Omega_m - 1)^{3/2}} \quad (1.28)$$

To interpret the above equations physically, the initial perturbation expands with the Hubble flow, after some time it detaches from the Hubble flow and begins to deviate significantly. We firstly consider the linear perturbation, we hence take the two leading order terms of the solution, which are given by:

$$r(\theta) \approx A \left(\frac{\theta^2}{2} - \frac{\theta^4}{4} \right) \quad (1.29)$$

$$t(\theta) \approx B \left(\frac{\theta^3}{6} - \frac{\theta^5}{120} \right) \quad (1.30)$$

Relating as $A^3 = GMB^2$. Consider at early times, i.e. $\theta \rightarrow 0$, we consider only the first order leading terms of $r(\theta)$ and $t(\theta)$, and we get:

$$r(t) = \frac{A}{2} \left(\frac{6t}{B} \right)^{2/3} \propto t^{2/3} \quad (1.31)$$

so the evolution of the perturbation shell evolves similarly to an unperturbed matter dominated

universe with $\Omega_m = 1$.

Consider at later times, we consider also the next leading order terms. Reiterating solutions for $r(\theta)$ and $t(\theta)$, we get,

$$r(t) = (6t)^{2/3}(GM)^{1/3} \left[1 - \frac{1}{20} \left(\frac{6t}{B} \right)^{2/3} \right] \quad (1.32)$$

Spherical collapse conserves the mass inside the perturbation shell, which is given by $M = 4\pi r_0^3 \bar{\rho}_m / 3$. Therefore, if the linear overdensity enhances by an amount δ , the radius must have shrunk by an amount of dr , yielding $M = 4\pi r_0^3 \bar{\rho}_m (1 + \delta) (1 + dr)^3 / 3$. On the other hand, mass conservation yields $(1 + \delta) (1 + dr)^3 = 1$. By Taylor expanding to express δ in terms of dr , and substituting the expression of dr from Eqn. 1.32, we find that the matter density and linear overdensity perturbation at a given time t are:

$$\rho_m = \frac{3M}{4\pi r^3} = \frac{1}{6\pi t^2 G} \left[1 + \frac{3}{20} \left(\frac{6t}{B} \right)^{2/3} \right] \quad (1.33)$$

$$\delta \equiv \frac{\rho_m - \bar{\rho}_m}{\bar{\rho}_m} = \frac{3}{20} \left(\frac{6t}{B} \right)^{2/3} \quad (1.34)$$

From the above equation, we see the mean mass density at a given time t is $\bar{\rho}_m = 1/(6\pi t^2 G)$, which also holds true for non-linear perturbation calculation. We will describe some of the key epochs of evolution predicted by the linear perturbation theory. For instance, turnaround occurs when $\theta = \pi$, when $t = \pi B$, and halo collapse occurs when $\theta = 2\pi$, when $t = 2\pi B$. The linear overdensities of these the turnaround and collapse epochs correspond to $\delta_{\text{ta}} = 1.06$ and $\delta_{\text{col}} = 1.69$ respectively.

The simple spherical collapse model assumes no shell crossing, i.e. all shells collapsing at the same time creating an infinite overdensity at collapse. Nevertheless, collapse does not proceed to a point but reach a state called virial equilibrium. The Virial theorem governs the final equilibrium state of a system with energies $U = -2K$, where U and K are the gravitational potential energy and kinetic energy respectively. At turnaround, the radius of the shell is maximum, reaching r_{max} .

By applying the Virial theorem and conservation of energy at turnaround and virialized state, we obtain $U(r_{\text{max}}) = U(r_{\text{vir}})/2$. Since $U \propto 1/r$, we solve that $r_{\text{vir}} = r_{\text{max}}/2$. Consider the collapse time equals twice that of the turnaround time, and the fact that the turnaround time equals the free fall time, i.e.

$$t_{\text{ta}} = t_{\text{ff}} = \frac{1}{2}t_{\text{col}} = \sqrt{\frac{3\pi}{32G\rho_{\text{ta}}}} \quad (1.35)$$

which implies the mean density of perturbation at turnaround is

$$\rho_{\text{ta}} = \frac{3\pi}{32Gt_{\text{ta}}^2} \quad (1.36)$$

On the other hand, as the radius of the density contrast shrinks by a factor of 2 from the turnaround state to the collapsed (virialized) state, the density at collapsed state becomes 8 times that of the turnaround state, i.e.

$$\rho_{\text{col}} = 8\rho_{\text{ta}} = \frac{3\pi}{Gt_{\text{col}}^2} \quad (1.37)$$

If we compare the mean density of perturbation at collapse with the background mean matter density, i.e. $\bar{\rho}_m(t_{\text{col}}) = 1/(6\pi t_{\text{col}}^2 G)$, we obtain the non-linear overdensity at virialization as:

$$\Delta_{\text{vir}} = \frac{\rho_{\text{col}}}{\bar{\rho}_m} = 18\pi^2 \approx 178 \quad (1.38)$$

This is true for the Einstein-de Sitter cosmology ($\Omega_m = 1$). For general cosmologies, the non-linear overdensity Δ_{vir} can be calculated in a similar fashion. Physically, for lower Ω_m models, fluctuation of the same mass M and δ has a larger initial radius and smaller physical density, and hence takes longer to collapse. As that implies the mean density of matter becomes smaller when collapse happens, the overdensity Δ_{vir} therefore are larger for lower Ω_m models.

Bryan and Norman [41] proposed a fitting formula that accurately approximates Δ_c , the over-

density defined with respect to the critical density of the Universe, in open (i.e. $\Omega_\Lambda = 0$) and flat Λ CDM cosmologies, with error $\lesssim 1\%$ for $0.1 \leq \Omega_m \leq 1$, i.e.

$$\Delta_c(z) = 18\pi^2 + 82[\Omega_m(z) - 1] - 39[\Omega_m(z) - 1]^2 \quad (1.39)$$

where $\Omega_m(z)$ is the cosmological parameter of the mass density at redshift z . As an example, for the concordance Λ CDM cosmology with $(\Omega_m, \Omega_\Lambda) = (0.27, 0.73)$, $\Delta_c(z=0) \approx 97$. Using the relation $\rho_{\text{halo}} = \Delta_c \rho_c = \Delta_{\text{vir}} \rho_m$ and $\Omega_m = \rho_m / \rho_c$, we obtain $\Delta_{\text{vir}}(z=0) \approx 358$.

As one of the key insights from this model is that Δ_{vir} is independent of the initial size and the amplitude of the density perturbation. Therefore, it is handy, and hence many literatures utilize this overdensity to define the boundary of a halo. In other words, a halo is defined as the interior of the radius enclosing a matter overdensity of $\Delta_{\text{vir}} = 178$. Conventionally, literatures use this virialized overdensity Δ_{vir} to define the radius of the halo which bounds an average density of this threshold value. In most of the analyses presented in this thesis, we use virial mass as the mass definition of halos, unless otherwise specified. In particular, we shall see how the choice of this definition affects the assembly bias results of cluster-sized halos ($\sim 10^{14} M_\odot$) in chapter 4.

1.3 Mass profile of dark matter halos

Multiple cosmological simulations have confirmed a universal density profile of dark matter halo internal mass distribution, named the Navarro-Frenk-White profile (hereby NFW profile) [45, 46]. It has been demonstrated repeatedly that the NFW profile can successfully describe the internal structure of dark matter halo over a wide range of halo masses (from sub-galaxy mass to cluster-sized halo mass). Many attempts have been made to study and analyze this apparent universality of the dark matter halo density profile [47–52].

The NFW profile takes the functional form

$$\rho(r) = \frac{\delta_0 \rho_c}{(r/r_s)(1 + r/r_s)^2} \quad (1.40)$$

where ρ_c is the critical density of the universe, δ_0 is a characteristic (dimensionless) density, and r_s is the scale radius, at which the logarithmic slope of density profile $d \ln \rho / d \ln r = -2$. In Eqn. 1.40, one can deduce that at the inner region where $r \ll r_s$, the density scales with r^{-1} , whereas at the outer region where $r \gg r_s$, the density drops more rapidly and scales with r^{-3} . Another important parameter for halos is the concentration parameter c , which can be defined as the ratio of the virial radius r_{vir} and the scale radius r_s of the halo, i.e. $c = r_{\text{vir}}/r_s$. The concentration parameter can be likewise defined with respect to different density thresholds such as 200 times the critical density or mass density of the universe. The concentration parameter characterizes how compact the halo is, in the sense that a larger c corresponds a more compact halo. With the known concentration parameter, we can figure out the dimensionless characteristic density δ_0 as,

$$\delta_0 = \frac{\Delta_{\text{vir}}}{3} \frac{c^3}{\ln(c+1) - c/(c+1)} \quad (1.41)$$

where Δ_{vir} is the overdensity at virialization as defined in Eqn. 1.38. The NFW profile can be leveraged to determine the mass profile and hence total mass of dark matter halo, which we will discuss in depth in the following chapters.

1.4 Shapes of dark matter halos

As mentioned in section 1.2, dark matter halos arise from the initial density perturbation. Nevertheless, a real density perturbation is neither spherical nor homogenous. In fact, dark matter halos are found to be triaxial in cosmological N -body simulations. Depending on the dark matter self-interaction cross-section, halos with different ellipticities can be produced. The cold dark matter model predicts elliptical halos with axis ratio approaching 0.5:1 [53–56], since cold dark matter is

dissipationless, it retains signatures of anisotropy during the formation of halos. Alternative model such as the self-interacting dark matter model, in which particles can dissipate momentum through non-gravitational interactions, washes out the anisotropic mass distribution of dark matter halos since their formation, producing nearly spherical halos [57–59]. Therefore, measuring halo shapes can help probe the fundamental nature of dark matter itself, which is one of the most outstanding problems in modern cosmology. We elaborate on this topic in chapter 2.

1.5 Spatial distribution of dark matter halos

Apart from the mass structure of dark matter halos, we also study their spatial distribution. Details on this topic can be found in [60–64].

In the standard cosmological model, the spatial clustering of galaxies reflects the clustering of the dark matter halos hosting those galaxies. Since halos are made out of dark matter, naively we would expect the halo sample the underlying dark matter mass distribution, i.e. the number density of halos δ_h is simply proportional to the matter density δ_m ,

$$\delta_h(\vec{x}) = \frac{n_h(\vec{x}) - \bar{n}_h}{\bar{n}_h} = \frac{\rho_m(\vec{x}) - \bar{\rho}_m}{\bar{\rho}_m} = \delta_m(\vec{x}) \quad (1.42)$$

If that were the case, galaxies form and reside in dark matter halos, then the number density of galaxies would be an unbiased estimator of the local mass density, i.e. light traces matter. However, in reality, halo formation is not a random process. Rather, they only form where the (smoothed) density field has exceeded a threshold value, i.e. the critical overdensity predicted by the spherical collapse model, $\delta_{\text{collapse}} = 1.69$. It is this threshold which causes halos to be **biased** tracers of the underlying matter distribution, i.e.

$$\delta_h = b\delta_m \quad (1.43)$$

In cosmological simulations, the bias b is typically measured by comparing the cross-correlation

function between dark matter halos of mass M and dark matter particles $\xi_{hm}(r)$ with the auto-correlation function of dark matter particles $\xi_{mm}(r)$, in other words:

$$\xi_{hm}(r) = \langle \delta_h(\vec{x}) \delta_m(\vec{x} + \vec{r}) \rangle = b \langle \delta_m(\vec{x}) \delta_m(\vec{x} + \vec{r}) \rangle = b \xi_{mm}(r) \quad (1.44)$$

we see that $b = \langle \xi_{hm} / \xi_{mm} \rangle$, with $\langle \cdot \rangle$ denotes an ensemble averaging over large radii. Alternatively, we can express the density of a component $\delta(\vec{x})$ as a sum over wave modes,

$$\delta(\vec{x}) = \sum \delta(\vec{k}) e^{-i\vec{k} \cdot \vec{x}} \quad (1.45)$$

we then have,

$$\xi(\vec{r}) = \langle \sum_{\vec{k}} \sum_{\vec{k}'} \delta(\vec{k}) \delta^*(\vec{k}') e^{i(\vec{k}' - \vec{k}) \cdot \vec{r}} \rangle \quad (1.46)$$

Given the periodic boundary condition, all the cross terms with $\vec{k} \neq \vec{k}'$ cancel out, we then obtain,

$$\xi(\vec{r}) = \frac{V}{8\pi^3} \int |\delta(\vec{k})|^2 e^{-i\vec{k} \cdot \vec{r}} d^3k \quad (1.47)$$

where V is the concerned volume, and $|\delta(\vec{k})|^2$ is the power spectrum. As we investigate a sufficient large volume where the isotropic principle applies, we then have $P(k) = \langle |\delta(k)|^2 \rangle = |\delta(\vec{k})|^2$, the correlation function becomes

$$\xi(r) = \frac{V}{2\pi^2} \int k^3 P(k) \frac{\sin(kr)}{kr} \frac{dk}{k} \quad (1.48)$$

Therefore, the two-point correlation function of a density fluctuation field δ is the Fourier transform of the power spectrum $P(k)$. We can then also measure the bias b in the Fourier space, i.e.

$$P_{hm}(k) = b P_{mm}(k) \quad (1.49)$$

One can determine the two-point correlation functions by either directly carry out pair-counting of the objects, or use Fourier transform of the fluctuation power spectrum. At large scales where the linear theory is applicable, the bias b approaches a constant value.

The peak-background split theory established in [64] explains why halo bias should exist. Simply put, the initial density field contained a mixture of small and large wavelength modes. The large-scale density field peaks evolved as ‘background’ overdensity environments, which modulated and enhanced the probability of small-scale density field to exceed δ_{collapse} , and hence forming dark matter halos. This also explains why those halos display enhanced clustering. Following the derivation in [64], it is immediately clear that massive halos with $M > M_*$ are positively biased $b > 1$ (i.e. more strongly clustered than the underlying dark matter distribution), whereas the opposite is true for low mass halos ($M < M_*$). Here, M_* is the characteristic mass scale which depends on redshift. Today, at $z = 0$, $M_* \approx 10^{13} M_\odot$.

Analysis on cosmological simulations have verified the claim above, i.e. the halo bias depends on halo mass. However, multiple simulations have also reported that halo bias can depend significantly on halo properties besides virial mass. Subsequent work showed that halo bias can depend on a variety of halo properties like concentration [65, 66]. The first detections of this effect noted a dependence on assembly history [67], leading to the term ‘**halo assembly bias**’ [67]. The origin of the age and concentration dependence of halo bias in simulations is now fairly well understood [68]. Motivated by the rigorous theoretical framework as well as analysis results of previous numerical simulations, we investigate on the assembly bias for galaxy-sized halos ($\sim 10^{12} M_\odot$) using CFHTLenS observational data and cluster-sized halos ($\sim 10^{14} M_\odot$) using the BigMDPL cosmological simulation, described in chapters 3 and 4 respectively.

1.6 Gravitational lensing - an astrophysical probe of dark matter halos

Unlike ordinary matter, dark matter does not emit or absorb electromagnetic radiation at any significant level [69], and hence is not directly observed in telescopes. Instead, its existence can be inferred by its gravitational effect on ordinary matter or photons. One astrophysical probe of dark matter is by gravitational lensing, which refers to the deflection of light rays that pass by concentrations of matter. This deflection of light rays can magnify and distort the appearance of light sources which are seen behind massive objects such as galaxies. For instance, gravitational lensing by a foreground halo can cause a background circular light source to appear sheared and non-circular. This shearing effect can be utilized to detect the amount of lensing that has occurred, and thereby to infer the existence and measure the mass distribution of dark objects.

In practice, the universe contains few circular light sources on the sky. Instead, most galaxies have intrinsically non-circular shapes, which acts as a source of noise when we attempt to measure gravitational lensing. In most cases, the amplitude of this shape noise is orders of magnitude larger than the amount of shear that is produced by gravitational lensing (the weak lensing regime). As a result, it is nearly impossible to detect lensing using individual objects. Nevertheless, we can measure lensing **statistically**, by correlating the shapes of many galaxies on the sky. This statistical measurement of lensing is called galaxy-galaxy lensing, and has now been detected observationally with signal-to-noise (S/N) ratios of order several hundred. See [70] for detailed introduction of weak gravitational lensing.

We mainly utilize weak gravitational lensing to constrain shapes and spatial distribution of dark matter halos, which are the focused topics in this thesis. The rest of the thesis discusses the physics of dark matter halos, and the use of gravitational lensing as an observational probe to study these systems.

Chapter 2

Three-point galaxy-galaxy lensing as a probe of dark matter halo shapes

We propose a method to measure the ellipticities of dark matter halos using the lens-shear-shear 3-point correlation function. This method is immune to effects of galaxy-halo misalignments that can potentially limit 2-point galaxy-galaxy lensing measurements of halo anisotropy. Using a simple model for the projected mass distributions of dark matter halos, we construct an ellipticity estimator that sums over all possible triangular configurations of the 3-point function. By applying our estimator to halos from N-body simulations, we find that systematic errors in the recovered ellipticity will be at the $\lesssim 5\%$ fractional level. We estimate that future imaging surveys like LSST will have sufficient statistics to detect halo ellipticities using 3-point lensing ¹.

2.1 Introduction

In the cold dark matter model of cosmological structure formation, galaxies are believed to form inside of virialized objects called dark matter halos. The properties of these halos, like their internal structure or abundance, are related to the background cosmology and to the physics of dark matter particles. One example of this is the ellipticity of dark matter halos. In CDM cosmologies, halos are found to be triaxial, with axis ratios of the order of 0.5:1, with a significant scatter from object to object [53–56]. Alternative models, like self-interacting dark matter (SIDM) can produce significantly different shapes. Pure SIDM simulations generally produce halos with rounder shapes than CDM simulations [57–59], although the effects of baryons can modify these results

¹This chapter was previously published in the Journal of Cosmology and Astroparticle Physics as S. Adhikari, C. Y. R. Chue and N. Dalal, Three-point galaxy-galaxy lensing as a probe of dark matter halo shapes, JCAP 1501 (2015) 009, [1407.1128]. This chapter matches the published version.

[71].

Therefore, measurements of the shapes of dark matter halos may be used to probe the nature of dark matter. Accordingly, multiple groups have attempted to measure halo shapes using a variety of probes. In our own Galaxy, several groups have attempted to model the dynamics of the Sagittarius tidal stream in order to infer the underlying shape of the Milky Way’s halo [72–76]. In other galaxies, halo shapes have been probed using strong lensing and stellar dynamics [77, 78] on small scales, and satellite dynamics on larger scales [79].

Another probe of dark matter halo properties is weak gravitational lensing. The average radial profiles of dark matter halos have been inferred with high precision through measurement of the two-point cross-correlation between galaxies and tangential shear, called galaxy-galaxy lensing [80–82]. Circularly averaged statistics are insensitive to halo ellipticity, but in principle, anisotropy could be constrained by measuring shear not only as a function of radius r , but position angle θ as well. Unfortunately, because dark matter halos are dark, we cannot determine the orientations of halos, making it impossible to measure shear profiles as a function of position angle relative to the halo principal axes. We can, however, measure shear as a function of the position angle relative to the lens galaxies’ principal axes. If halos are perfectly aligned with their central galaxies, then such measurements may be used to determine the average halo ellipticity. This is the approach that has been used by most previous work [83–86]. This previous work, however, has yielded inconclusive results. For example, [86] report an average projected ellipticity of $e = 0.38 \pm 0.26$, which is consistent both with CDM predictions and with completely isotropic halos. Currently, statistical errors are a principal limitation of this measurement, but with the vastly increased sample sizes provided by future imaging surveys like LSST, the statistical errors may be reduced sufficiently to detect the expected signal. More worryingly, this method is likely limited by potentially severe systematic effects. First, the assumption that galaxies and their halos are perfectly aligned may be unrealistic [87]. [88] has argued that significant misalignments between galaxies and halos may be quite typical; the median misalignment angle in their simulations was $\sim 38^\circ$. Random misalignments act to wash out the halo anisotropy signal from galaxy-galaxy lensing. Even worse, they

complicate the interpretation of any measured anisotropy signal. Without knowledge of the misalignment distribution, we will not know how to translate stacked lensing signals into constraints on halo axis ratios. This effect is also not the only possible systematic. For example, if lens galaxies and background source galaxies are both lensed by foreground structures, this common lensing will tend to align their observed shapes, thereby contaminating the halo anisotropy signal [89]. Because of these systematic limitations, an alternative approach for measuring halo shapes with galaxy-galaxy lensing may be required – ideally, a method that does not require galaxies to align with their host halos. Such an approach is suggested by the recent work of [90], who find that halo ellipticities affect galaxy-galaxy lensing 3-point correlation functions. Although most previous work on galaxy-galaxy lensing has focused on 2-point statistics, higher order correlation functions are now becoming measurable in modern imaging surveys [91, 92]. In this section, we explore how halo ellipticities may be determined from measurements of the galaxy-shear-shear 3-point function.

2.2 Mass model

The 3D density profiles of halos in dissipationless CDM simulations have axis ratios of order $q \approx 0.5$, slowly varying with radius [54, 55]. Similarly, the 2D projected surface density Σ is anisotropic, with axis ratios closer to $q \sim 0.7$, again slowly increasing with radius. Because q is nearly constant with radius, we can write $\Sigma \propto R^{-\eta}$, where $R = (x^2 + y^2/q^2)^{1/2}$ is an ellipsoidal radial coordinate, and η is the logarithmic slope of the projected surface density. We will find it convenient below to work with the multipole moments of the density profile. In the limit of small ellipticity, we can write the multipole expansion of Σ in terms of q ,

$$\begin{aligned} \Sigma(r, \theta) &\propto r^{-\eta} [1 + \varepsilon \eta \cos 2\theta + \mathcal{O}(\varepsilon^2)] \\ &\equiv \Sigma_0(r) + \Sigma_2(r) \cos 2\theta + \dots \end{aligned} \tag{2.1}$$

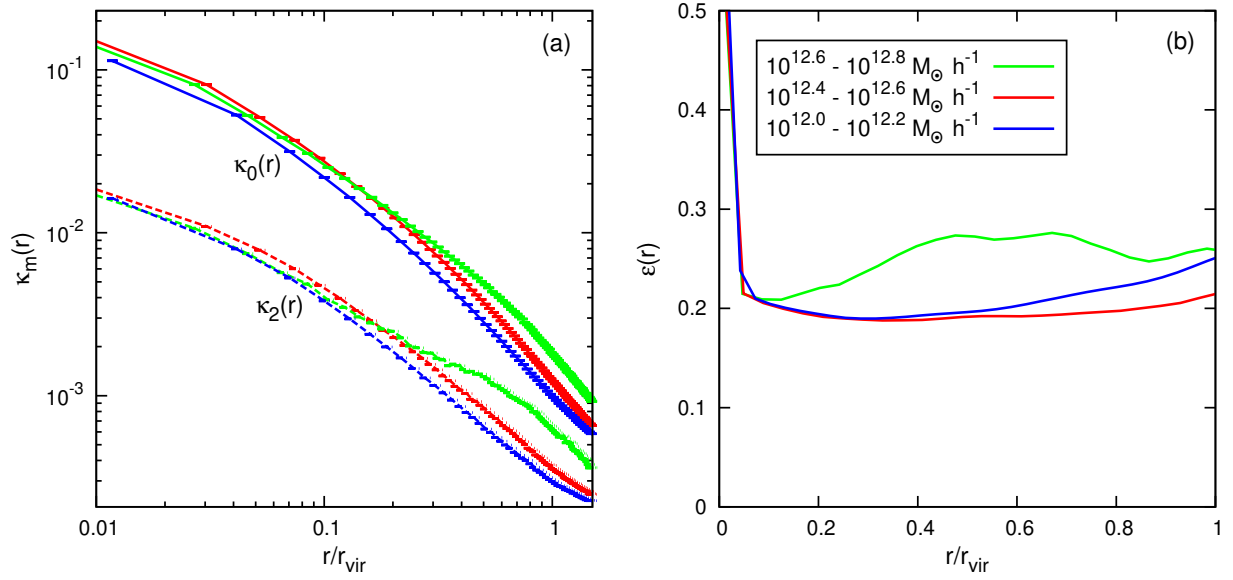


Figure 2.1: (a): Plot of the multipole moments of stacked halos. The solid curves show the isotropic component (monopole, κ_0) of the surface density profile and the dashed curves show the $\cos 2\theta$ component (quadrupole, κ_2). The convergence κ is proportional to surface density Σ . (b): Radial dependence of ellipticity, which we define as $\varepsilon(r) \equiv \kappa_2/(\eta \kappa_0)$, for three different mass bins. The blue, red and green colors correspond to three different mass bins. Note that, although the multipole moments vary by orders of magnitude, the ellipticity remains nearly constant across much of the range of interest.

where the multipole $\Sigma_m(r)$ is the coefficient of the $e^{im\theta}$ component of the azimuthal behavior, and we use $\varepsilon = (1 - q^2)/[2(1 + q^2)]$ to parameterize the ellipticity. For the typical axis ratios found in simulated halos, $\varepsilon \sim 0.2$, so we neglect higher order terms in the expansion.

We therefore model the mass distributions of halos as the sum of a monopole and quadrupole, and further assume that

$$\varepsilon \approx \frac{\Sigma_2(r)}{\eta(r)\Sigma_0(r)} \quad (2.2)$$

where $\eta = d \log \Sigma_0 / d \log r$. In Figure 2.1, we plot ε as defined in Eqn. (2.2), measured from stacked profiles of projected halos taken from the Bolshoi simulation [93]. We measure multipole moments from the particle positions, using

$$\Sigma_m(r) = \sum_i \frac{m_p \delta(r - r_i) e^{im\theta_i}}{2\pi r_i}, \quad (2.3)$$

where m_p is the particle mass, and r_i and θ_i are the radius and azimuthal angle for particle i . After computing $\Sigma_0(r)$ and $\Sigma_2(r)$ for each halo, we then stack the halos to compute $\langle \Sigma_0 \rangle(r)$ and $\langle |\Sigma_2| \rangle(r)$, and then ε . As expected, the ellipticity is fairly constant with radius, except very near the halo center where $\eta = d \log \Sigma_0 / d \log r \rightarrow 0$. Because ε is nearly constant with radius, then the radial dependence of the quadrupole may be predicted from the monopole, whose mean $\langle \Sigma_0(r) \rangle$ may be determined from real galaxy halos using galaxy-galaxy lensing 2-point statistics. Specifically, the mean tangential shear $\langle \gamma_+ \rangle$ profile around halos is related to the mean monopole via [70]

$$\langle \gamma_+ \rangle = \frac{\Delta \Sigma(r)}{\Sigma_{\text{crit}}} = \frac{\bar{\Sigma}_0(< r) - \Sigma_0(r)}{\Sigma_{\text{crit}}}, \quad (2.4)$$

where the lensing critical density Σ_{crit} is defined as

$$\Sigma_{\text{crit}} = \frac{c^2}{4\pi G} \frac{D_s}{D_{ds}D_d}, \quad (2.5)$$

and D_d and D_s are the angular diameter distances to the lens and the source respectively, while D_{ds} is the angular diameter distance from the lens to the source. Eqn. (2.4) can be inverted (up

to a mass-sheet degeneracy) to obtain the mean monopole profile $\Sigma_0(r)$ from the observed mean tangential shear profile $\langle\gamma_+\rangle(r)$ via

$$\frac{\Sigma_0(r) - \Sigma_0(r_{\max})}{\Sigma_{\text{crit}}} = 2 \int_r^{r_{\max}} \langle\gamma_+\rangle(R) \frac{dR}{R} - [\langle\gamma_+\rangle(r) - \langle\gamma_+\rangle(r_{\max})]. \quad (2.6)$$

Here, $r_{\max}$ is the largest radius over which the stacked tangential shear profile $\langle\gamma_+\rangle$ has been measured.

For circularly symmetric lenses, the tangential shear is the only nonzero component of the shear. When the surface density is anisotropic, however, the other component ($\gamma_\times$) becomes nonzero. In the same way that we can decompose the surface density into angular multipoles $\Sigma_m(r)$, we can similarly decompose the shear into multipoles $\gamma^{(m)}(r)$. The relation between the density and shear multipoles is straightforward. For convenience, we follow conventional notation and define the convergence as $\kappa = \Sigma/\Sigma_{\text{crit}}$, and define a 2D lensing potential ψ via

$$\nabla^2 \psi = 2\kappa \quad (2.7)$$

where the gradient is with respect to sky coordinates. In polar coordinates, this equation becomes

$$\kappa(r, \theta) = \frac{1}{2} \left[\frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} \right] \psi. \quad (2.8)$$

The two components of the shear are given by

$$\gamma_+ = \left[-\frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} \right] \psi \quad (2.9)$$

$$\gamma_\times = \left[-\frac{2}{r} \frac{\partial^2}{\partial r \partial \theta} + \frac{2}{r^2} \frac{\partial}{\partial \theta} \right] \psi \quad (2.10)$$

Next, let us decompose these fields into angular multipoles

$$\begin{aligned}\psi(r, \theta) &= \sum_{m=-\infty}^{\infty} \psi_m(r) e^{im\theta} \\ \kappa(r, \theta) &= \sum_{m=-\infty}^{\infty} \kappa_m(r) e^{im\theta}.\end{aligned}\tag{2.11}$$

Explicitly,

$$\begin{aligned}m = 0 : \quad \kappa_0(r) &= \frac{1}{2\pi} \int_0^{2\pi} \kappa(r, \theta) d\theta \\ m \geq 1 : \quad \kappa_m(r) &= \frac{1}{\pi} \int_0^{2\pi} \kappa(r, \theta) \cos m\theta d\theta.\end{aligned}\tag{2.12}$$

Solving the 2-d Poisson Eqn. (2.8), we obtain the multipole moments of ψ ,

$$\begin{aligned}\psi_0(r) &= \ln r \int_0^r r' \kappa_0(r') dr' + \int_r^\infty r' \ln r' \kappa_0(r') dr' \\ \psi_m(r) &= -\frac{1}{2m} \left[r^{-m} \int_0^r r'^{(m+1)} \kappa_m(r') dr' \right. \\ &\quad \left. + r^m \int_r^\infty r'^{(1-m)} \kappa_m(r') dr' \right].\end{aligned}\tag{2.13}$$

Then, using Eqns. (2.9) and (2.10), we may obtain the multipole moments of the two shear components. Because we keep only $m = 0$ and $m = 2$, and because we assume that $\kappa_2(r) = \varepsilon \eta(r) \kappa_0(r)$, we have

$$\gamma_+^{(0)}(r) = \frac{2}{r^2} \int_0^r r' \kappa_0(r') dr' - \kappa_0(r)\tag{2.14}$$

$$\begin{aligned}g_+(r) &= \left[-\kappa_0(r) \eta(r) + \frac{3}{r^4} \int_0^r r'^3 \kappa_0(r') \eta(r') dr' \right. \\ &\quad \left. + \int_r^\infty \frac{\kappa_0(r') \eta(r')}{r'} dr' \right]\end{aligned}\tag{2.15}$$

$$\begin{aligned}g_\times(r) &= \left[\frac{3}{r^4} \int_0^r r'^3 \kappa_0(r') \eta(r') dr' \right. \\ &\quad \left. - \int_r^\infty \frac{\kappa_0(r') \eta(r')}{r'} dr' \right],\end{aligned}\tag{2.16}$$

where we have defined, for the purpose of convenience, the functions g_+ and $g_\times$ such that the quadrupole components of the shear are $\gamma_+^{(2)} = \varepsilon g_+(r) \cos 2\theta$ and $\gamma_\times^{(2)} = \varepsilon g_\times(r) \sin 2\theta$. Note that, by definition, $\gamma_\times^{(0)} = 0$.

Given this model for the mass distributions of lenses, we can predict the shear at all locations around the lenses. The one unknown parameter is the ellipticity ε , which defines the amplitude of the quadrupole moment κ_2 in terms of the monopole moment κ_0 , which we assume may be determined using Eqn. (2.6). Because we have an expression for the shear at all locations, we can construct an estimator for the quantity ε .

2.2.1 Three-Point Estimator

As discussed in §2.1, [90] have shown that lensing 3-point functions are sensitive to halo ellipticities. However, they also show that lensing 3-point functions are also sensitive to many other terms, making it difficult to disentangle the signal in the bispectrum generated by halo ellipticity. Fortunately, given our model for halo mass distributions, it is straightforward for us to construct an estimator to measure halo ellipticity from lensing correlation functions. Following [90], we focus on the lens-shear-shear 3-point function. Measurement of this correlation function involves stacking the shear measured from pairs of source galaxies behind foreground lens galaxies. Because the number density of source pairs is low, especially at the small radii of interest for measuring internal halo properties ($r < r_{\text{vir}}$), we assume that shape noise in the source galaxies dominates measurement uncertainties. That is, we neglect the signal covariance compared to Poisson fluctuations in source counts. Because Poisson noise is white noise, the optimal estimator is then simply proportional to the expected signal from our model.

We therefore estimate the average lens ellipticity by summing over all lens-source-source triangles, weighting each triangle with a filter $\mathbf{F}$ that is given by the *predicted* model shear for each configuration of galaxies. Figure 2.2 illustrates the geometry on the sky. Suppose that we have measurements of the shear at positions $\vec{r}_1$ and $\vec{r}_2$ relative to the center of the lens halo. When we sum over all possible $\vec{r}_1$ and $\vec{r}_2$, the filter which weights each triangle is $\mathbf{F}(\vec{r}_1, \vec{r}_2) \propto \gamma^{(2)}(\vec{r}_1) \otimes \gamma^{(2)}(\vec{r}_2)$.

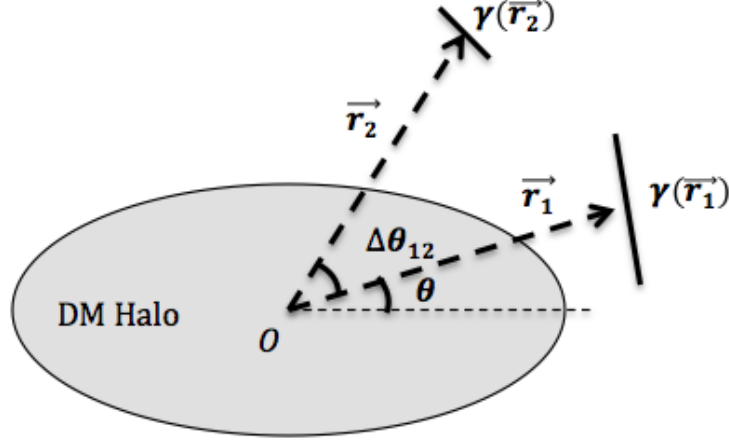


Figure 2.2: Lens-shear-shear three point correlation function. We correlate the shear at sky positions $\vec{r}_1$ and $\vec{r}_2$ relative to foreground lens galaxies, and construct an estimator summing all such triangular configurations in the sky.

Because the orientation angle of the lens halo is unknown, the filter must depend on the relative position angle of the sources, not their absolute position angles: $\mathbf{F}(\vec{r}_1, \vec{r}_2) = \mathbf{F}(|\vec{r}_1|, |\vec{r}_2|, \Delta\theta_{12})$. We compute $\mathbf{F}$ by averaging all possible triangles with one vertex at the center of the lens, and a constant opening angle, $\Delta\theta_{12}$, between position vectors to the lensed galaxies, of magnitude $|\vec{r}_1|$ and $|\vec{r}_2|$ (see Fig. 2.2). Since each shear has 2 components, the filter $\mathbf{F}$ is a 2×2 matrix, with components

$$\begin{aligned}
 F_{ij}(r_1, r_2, \Delta\theta_{12}) &= \langle g_i(\vec{r}_1) g_j(\vec{r}_2) \rangle \\
 &= \frac{\int r'_1 r'_2 dr'_1 dr'_2 d\theta'_1 d\theta'_2 g_i(\vec{r}_1) g_j(\vec{r}_2) \delta_{r'_1} \delta_{r'_2} \delta_{\Delta\theta'_{12}}}{\int r'_1 r'_2 dr'_1 dr'_2 d\theta'_1 d\theta'_2 \delta_{r'_1} \delta_{r'_2} \delta_{\Delta\theta'_{12}}},
 \end{aligned} \tag{2.17}$$

where indices i, j run over $+, \times$, and we define $\Delta\theta_{12} = \theta_2 - \theta_1 = \cos^{-1}(\vec{r}_1 \cdot \vec{r}_2 / r_1 r_2)$, along with $\delta_{r'_\alpha} \equiv \delta(r'_\alpha - r_\alpha)$ and $\delta_{\Delta\theta'_{12}} \equiv \delta(\Delta\theta'_{12} - \Delta\theta_{12})$. The integral in (2.17) covers the projected area in the sky, where r ranges from some $r_{\min}$ to $r_{\max}$, and θ ranges from 0 to 2π . Simplifying using the

δ functions, we obtain

$$\begin{aligned}
F_{++} &= \frac{1}{2}g_+(r_1)g_+(r_2)\cos 2\Delta\theta_{12} \\
F_{+\times} &= \frac{1}{2}g_+(r_1)g_\times(r_2)\sin 2\Delta\theta_{12} \\
F_{\times+} &= -\frac{1}{2}g_\times(r_1)g_+(r_2)\sin 2\Delta\theta_{12} \\
F_{\times\times} &= \frac{1}{2}g_\times(r_1)g_\times(r_2)\cos 2\Delta\theta_{12}
\end{aligned} \tag{2.18}$$

Eqns. (2.18) specify the elements of the filter weighting each possible triangle in the 3-point correlation function. We then evaluate our estimator by summing over all triangles, weighting the shear by $\mathbf{F}$. Explicitly, we evaluate

$$f_{\text{obs}} = \langle \boldsymbol{\gamma}(\vec{r}_1) \cdot \mathbf{F}(r_1, r_2, \Delta\theta_{12}) \cdot \boldsymbol{\gamma}(\vec{r}_2) \rangle, \tag{2.19}$$

where the expectation value implies averaging over all possible $\vec{r}_1$ and $\vec{r}_2$. Note that, because of the angular dependence of $\mathbf{F}$, Eqn. (2.19) is only sensitive to the quadrupolar component of the shear.

In order to translate f_{obs} into an estimate for the ellipticity ε , we need to know what result Eqn. (2.19) will give as a function of ε . We can compute this by inserting the predicted model shear into

the equation. Specifically, let us define

$$\begin{aligned}
f_{\text{model}} &= \langle \mathbf{g}(\vec{r}_1) \cdot \mathbf{F} \cdot \mathbf{g}(\vec{r}_2) \rangle \\
&= \frac{1}{2} \left\langle g_+^2(r_1) g_+^2(r_2) \cos 2\theta_1 \cos 2\theta_2 \cos 2\Delta\theta_{12} \right. \\
&\quad + g_+^2(r_1) g_\times^2(r_2) \cos 2\theta_1 \sin 2\theta_2 \sin 2\Delta\theta_{12} \\
&\quad - g_\times^2(r_1) g_+^2(r_2) \sin 2\theta_1 \cos 2\theta_2 \sin 2\Delta\theta_{12} \\
&\quad \left. + g_\times^2(r_1) g_\times^2(r_2) \sin 2\theta_1 \sin 2\theta_2 \cos 2\Delta\theta_{12} \right\rangle \\
&= \frac{1}{2} \left\{ \frac{\pi}{A} \int_{r_{\min}}^{r_{\max}} [g_+^2(r) + g_\times^2(r)] r dr \right\}^2 \\
&= \frac{1}{2} \left\{ \frac{\int_{r_{\min}}^{r_{\max}} [g_+^2(r) + g_\times^2(r)] r dr}{r_{\max}^2 - r_{\min}^2} \right\}^2.
\end{aligned} \tag{2.20}$$

Then

$$\varepsilon = \sqrt{\frac{f_{\text{obs}}}{f_{\text{model}}}} \tag{2.21}$$

This defines our estimator for the ellipticity ε . To reiterate, the ingredient in our expression is the radial profile of the average monopole density profile $\langle \kappa_0 \rangle(r)$, which may be reconstructed from the stacked tangential shear profile $\langle \gamma_+ \rangle(r)$. Given $\kappa_0(r)$, we may then determine the functions g_+ and $g_\times$ which enter the estimator. In the next section, we apply this estimator to samples of halos from N-body simulations, to gauge how well we can measure halo ellipticities for realistic objects.

2.3 Results

In §2.2, we proposed a 3-point estimator for halo anisotropy. In this section, we assess how well this estimator measures average halo ellipticities. First, we use simulated halos from cosmological N-body simulations to quantify systematic errors caused by the fact that the structure of realistic halos will not be as simple as our monopole+quadrupole mass model. Secondly, we quantify the statistical error associated with the finite number of lens-source-source triples. Because each

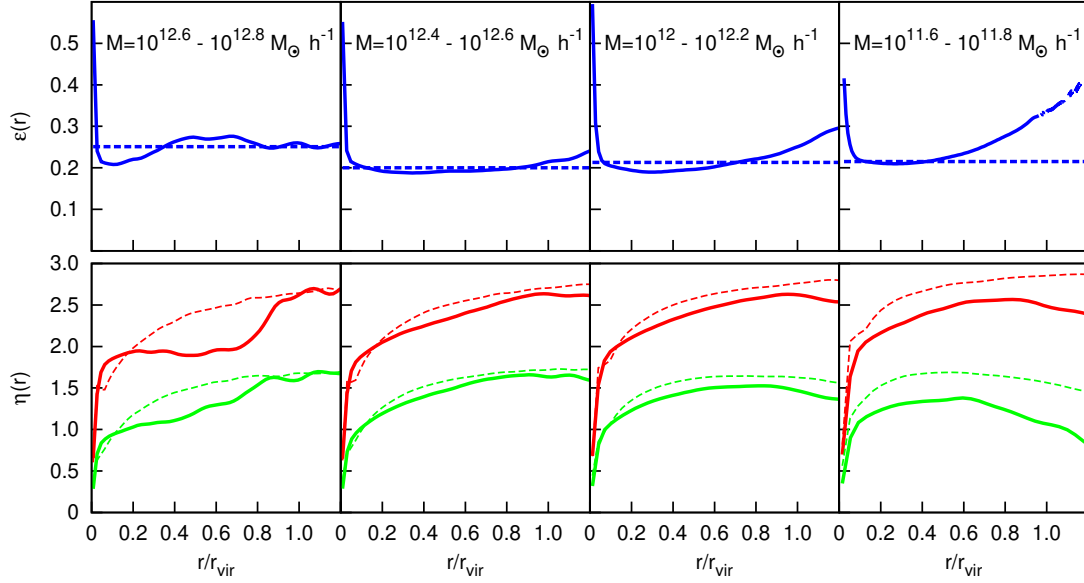


Figure 2.3: The top panels show the comparison between the average ellipticities of halos in various mass bins (solid curve), compared to the ellipticity determined from the 3-point estimator (dashed line). The bottom panel shows the local slope of the isotropic component (monopole) of the halos. The red curve corresponds to the smoothed slope of the 3D profile, and the green curve corresponds to the local slope of the projected 2D profile. The dashed curves show the slope of the NFW profile in 2D and 3D for reference. These slopes were measured from the stacked profiles after they were smoothed using a 6th order Savitzky-Golay filter over 17 nearest bins [1]. In the low mass bins, we observe significant departures from NFW slopes at large radii, possibly indicating the effects of nearby halos.

source galaxy provides an extremely noisy estimate of the shear, large numbers of triples are required to suppress this statistical shape noise.

2.3.1 Comparison with N-body simulations

We have applied our estimator (Eqn. 2.19) to halos from the publicly available Bolshoi simulation [93]. Using the BDMW catalog provided by the MultiDark database ², we selected halos with virial masses in the range $10^{11.7} - 10^{12.7} M_{\odot} h^{-1}$. We downloaded particles within $5r_{\text{vir}}$ of the halo center, to account for the mass within the halo as well as the nearby neighborhood. For each halo, we construct three projections, along the simulation box axes, to construct convergence and shear maps. From these shear maps, we then apply our estimator to measure the halo ellipticity ε . Figure 2.3 shows the results of our measurement across several mass bins. For comparison, the figure also plots the ellipticity directly measured from the projected mass profiles. In all cases, we find good agreement, despite several potential systematics discussed below.

First, our mass model assumes that ellipticity ε is constant with radius, meaning that the shape of the quadrupole $\kappa_2(r)$ of the mass distribution may be determined from the shape of the monopole profile $\kappa_0(r)$. For individual galaxies, the mass distributions are unknown. Galaxy-galaxy lensing can be used to reconstruct the mean monopole profile $\langle \kappa_0 \rangle(r)$, but individual halos will have radial profiles that vary from the mean. Because our estimator is not linear in the shear, this scatter in radial profiles can bias our measurement. To estimate the size of this potential bias, we generated artificial halos with radial profiles consistent with the Bolshoi halos (i.e. same M_{vir} and c_{vir}) but with specified values of ε . For the range of concentrations found in the mass range we have considered, we find a *fractional* bias in the reconstructed ε of $\sim 3 - 6\%$.

A second potential source of systematic errors arises from projections of other halos. Our mass model assumes that all shear is generated by the halos hosting the stacked galaxies. In reality, however, not all halos are isolated: other objects can project near the objects we are stacking and contaminate our measurement. Figure 2.4 shows one such example. Such projections can produce

²<http://hipacc.ucsc.edu/Bolshoi/MergerTrees.html>

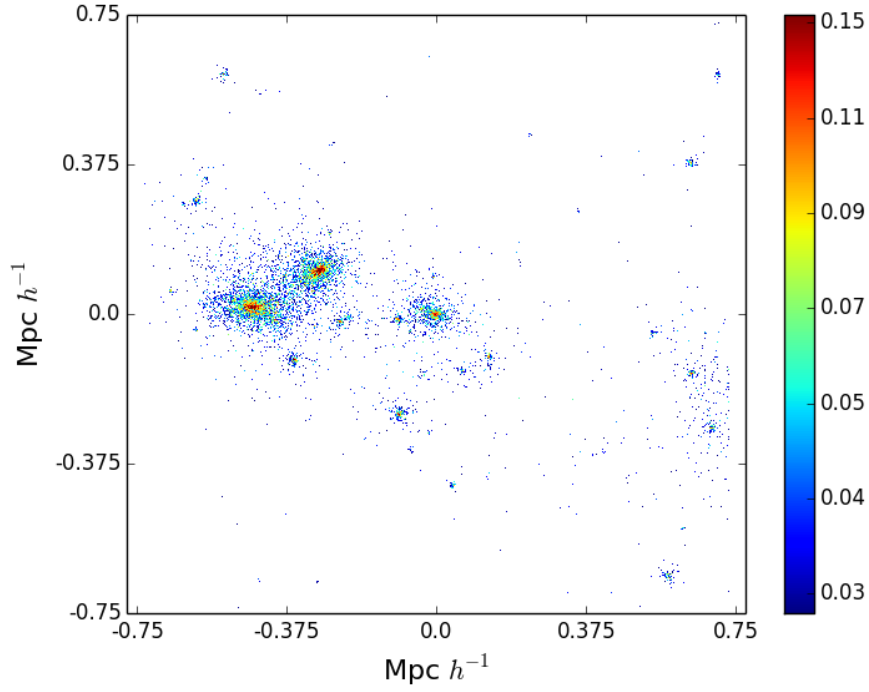


Figure 2.4: Example of line of sight projection effects. The halo at the origin has massive neighbors projecting nearby, which generate a large quadrupole moment that is unrelated to the halo’s own ellipticity. Colors correspond to convergence κ , for a lens redshift of $z_l = 0.3$ and source redshift of $z_s = 0.5$.

very large quadrupoles (and other multipoles) near certain halos, and because our estimator is not linear in the shear, this contamination can bias our results.

In general, there are two types of projections relevant to our measurement: galaxies that are correlated with the foreground lenses, and uncorrelated galaxies that randomly project into the line of sight. It is straightforward to correct for the uncorrelated projections. We could, for example, simply stack on random sky points instead of lens galaxies, and subtract this from our estimator. Mitigating the effects of correlated structures is not as easy. Perhaps the simplest approach would be to stack only galaxies that are relatively isolated, i.e. galaxies that are clearly central galaxies (not satellites), and that have no comparably bright galaxies nearby the line of sight. Such an approach should remove much of the contamination from nearby, correlated structures, but may not remove the contamination completely. Therefore, we need to estimate the effects of such contamination.

The largest source of contamination from correlated structures comes from satellite galaxies. Our estimator assumes that the lens galaxies are central galaxies within their halos, however a large fraction of galaxies ($\sim 20\%$) will be satellites living in massive hosts like clusters or groups. The quadrupole moments around satellite galaxies are much larger than those near central galaxies. To estimate this effect, we use subhalo abundance matching to find subhalos which could host galaxies that are similar to the central galaxies in our sample. We use the Rockstar catalog from MDR1 and find all halos and subhalos with similar V_{acc} , the circular velocity at time of accretion. For objects with $V_{\text{acc}} \approx 245\text{km/s}$, about 20% were subhalos rather than isolated halos. When we apply our estimator to the full sample of halos and subhalos, we find a large bias in the recovered ellipticity, $\varepsilon \approx 0.6$ instead of 0.2, as shown in Figure 2.5. Therefore if uncorrected, satellite contamination would significantly compromise our ability to measure halo ellipticity.

However, most of the contamination arises from satellites in the most massive hosts, and those satellites are the easiest to identify as satellites, since their local neighborhoods have a large galaxy overdensity. Removing those objects should therefore be relatively straightforward. Figure 2.5 shows how the contamination is reduced if we are able to remove various fractions of the satellite

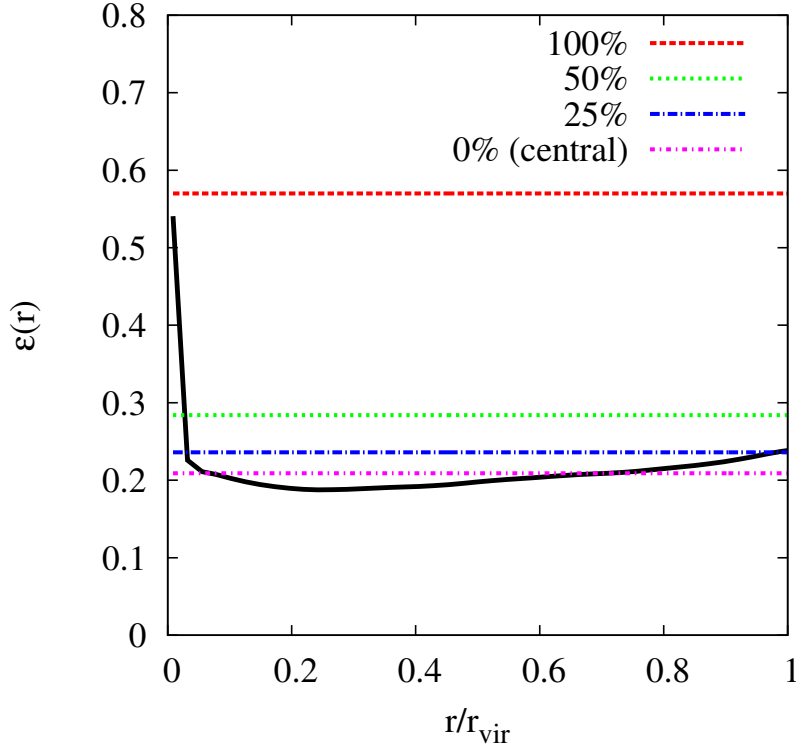


Figure 2.5: Effect of including satellites (subhalos) in our sample. The black solid curve shows the average ellipticity of halos estimated from N-body simulations in the mass bin $10^{12.4} - 10^{12.6} M_{\odot} h^{-1}$ using central galaxies only. The dashed lines show the ellipticity determined by the three-point estimator, applied to samples with various degrees of contamination by satellites. If satellites are not excluded (red line), the estimated ellipticity is significantly biased. Removing the satellites in the most dense environments eliminates much of the bias.

population. Here, we have ranked the satellites based on local overdensity. Removing 50% of the satellites removes the vast majority of the contamination, while removing 75% of the satellites gives a recovered ellipticity very close to the ellipticity for no satellites at all. Therefore, even a crude identification of satellite galaxies should suffice to eliminate most of the potential contamination.

The other type of contamination to consider is that from nearby halos. Fortunately, it appears that any bias due to projections of nearby correlated halos may not be large. In our calculations, we have not corrected for projections of nearby halos in any way. Our measurement is therefore contaminated by projections of other halos within $5 r_{\text{vir}}$ (as in Figure 2.4). Arguably, this should account for most of the correlated objects. The galaxy auto-correlation function behaves close to $\xi(r) \propto r^{-2}$ in 3D [29], so the number of galaxies with 3D radius $r > 5r_{\text{vir}}$ that project onto small radius should be about $\sim 1/5$ of the number of galaxies with 3D radius $r > r_{\text{vir}}$ projecting onto small radius. Because we have extracted particles out to $5r_{\text{vir}}$, we should account for about $\sim 80\%$ of the correlated projections.

Our calculations should therefore include most of the effect of projections of correlated structure, and as Figure 2.3 illustrates, the effect of those projections on the stacked profiles of central galaxies is likely to be small. Only in the lowest mass bin ($M \approx 10^{11.7} M_{\odot} h^{-1}$) do we observe any effects of the 2-halo term, and even there the recovered ε from the 3-point estimator is consistent with the halo ellipticity measured over the radial range where the 1-halo term is dominant. Nevertheless, when measuring halo ellipticity for real lenses, it will be important to restrict the analysis to the regime where the 1-halo term dominates, which may be determined by modeling the stacked tangential shear profile $\langle \gamma_+ \rangle$.

Another potential source of systematic error can arise due to the ‘twisting’ of halos. It is known that the principal axes of the isodensity surfaces in N-body halos are not constant with radius, but instead twist in orientation between small radii and large radii. Our simplistic mass model does not account for twisting of the principal axes, so a significant twisting could bias our results. To quantify how much twist we can tolerate, we created artificial halos in which we rotate

the direction of the principal axes following a logarithmic spiral in radius. We found that twist biases the estimated ellipticity by 10% when the halo axes rotate by more than π radians within one virial radius of the halo. In N-body simulations, halos do not show such high degrees of twist within their virial radius. Typically the rotation of the major axis is $\lesssim \pi/6$ within one virial radius [56], and for such low twist angles, the bias generated by twist is negligible. Therefore, it is safe to conclude that the twisting of halos does not significantly affect the measurement of ellipticity with the 3-point correlator based on our simple mass model, as we might have guessed based on the good agreement between the estimated ellipticity and true ellipticity of our N-body halos. Overall, our analysis of N-body halos suggests that systematic errors due to our simplistic mass model will not significantly bias our measurement of halo anisotropy.

2.3.2 Shape noise

In most regimes of weak lensing, the shear signal due to weak gravitational lensing is orders of magnitude weaker than the noise introduced by the intrinsic distribution of galaxy shapes and orientations. To estimate the magnitude of the errors induced by shape noise, let us first define the shape noise per galaxy $\vec{\mathcal{N}} = \{\mathcal{N}_+, \mathcal{N}_\times\}$. Each component of $\vec{\mathcal{N}}$ is assumed to be a Gaussian random variable with covariance

$$\langle \mathcal{N}_i \mathcal{N}_j \rangle = \sigma_\epsilon^2 \delta_{ij} \quad (2.22)$$

where the indices i and j correspond to the tangential and cross components of the noise, and $\sigma_\epsilon = 0.25$ [94]. The number density of source galaxies is $n(\vec{x}) = \sum_i \delta(\vec{x} - \vec{x}_i)$, with mean number density $\bar{n}$. Then the 2-point correlation of the shape noise is

$$\langle \mathcal{N}_i(\vec{x}_1) \mathcal{N}_j(\vec{x}_2) \rangle = \sigma_\epsilon^2 \delta_{ij} \frac{\delta(\vec{x}_1 - \vec{x}_2)}{\bar{n}} \quad (2.23)$$

Within an area A , where the number of galaxies is approximately $N = \bar{n}A$, the signal expectation value derived in Eqn. (2.20) is

$$S = \langle \gamma \cdot \mathbf{F} \cdot \gamma \rangle = \frac{\pi^2 \varepsilon^2}{2A^2} \left\{ \int [g_+^2(r) + g_\times^2(r)] r dr \right\}^2 \quad (2.24)$$

In comparison, the noise variance is

$$\begin{aligned} \sigma_N^2 &= \langle (\mathcal{N} \cdot \mathbf{F} \cdot \mathcal{N})^2 \rangle - (\langle \mathcal{N} \cdot \mathbf{F} \cdot \mathcal{N} \rangle)^2 \\ &= \frac{\sigma_\varepsilon^4 \pi^2}{\bar{n}^2 A^4} \left\{ \int [g_+^2(r) + g_\times^2(r)] r dr \right\}^2 \end{aligned} \quad (2.25)$$

where the expectation value is computed by both summing over all possible triangles in the sky and also by taking the ensemble average of the Gaussian noise field. In the second equality, we have used Eqn. (2.23) and Wick's theorem. Therefore the expected signal to noise per lens galaxy for the constructed estimator is

$$\frac{S}{\sigma_N} = \frac{\pi \varepsilon^2 \bar{n}}{2\sigma_\varepsilon^2} \int [g_+^2(r) + g_\times^2(r)] r dr. \quad (2.26)$$

As we might expect, the signal to noise ratio (SNR) per lens scales quadratically in the shear. Therefore the signal should be easiest to detect for more massive galaxies that produce stronger shear, as long as the abundance of galaxies does not fall steeply with mass. To get a sense of the expected SNR, we can perform a rough estimate by approximating the halo profile as isothermal ($\Sigma_0 \propto r^{-1}$), which Figure 2.3 shows is not a terrible approximation over the radial range of interest. To be concrete, suppose that the monopole profile is $\kappa_0(r) = b/(2r)$. Plugging this into Eqn. (2.26), we find that $S/\sigma_N \approx [\pi b^2 \bar{n} \varepsilon^2 / (8\sigma_\varepsilon^2)] \log(r_{\max}/r_{\min})$ per lens. Taking $\varepsilon = 0.2$, $\sigma_\varepsilon = 0.25$, $\bar{n} = 12 \text{ arcmin}^{-2}$ as appropriate for DES, $b = 1 \text{ arcsecond}$, and $r_{\max}/r_{\min} = 20$ gives $S/\sigma_N \approx 0.0025$, meaning that with 10^6 such lenses, we could detect the expected ellipticity at $\sim 2.5\sigma$. LSST will have more than twice the effective number density of sources [94], more than doubling the signal to noise of the 3-point estimator. At this point, it is perhaps worth comparing this

estimate with the corresponding signal/noise ratio for a 2-point estimator. Repeating the argument of §2.2 for the analogous 2-point estimator, we find that per lens, $(S/\sigma_N)_{2\text{pt}} = \sqrt{2(S/\sigma_N)_{3\text{pt}}}$, for halos that are perfectly aligned with their galaxies on the sky. Since the SNR per lens is much less than 1, this illustrates that 2-point estimates of halo anisotropy will have much greater statistical sensitivity than 3-point estimators. As noted above, however, this superior statistical power may be irrelevant if systematic effects due to halo misalignments remain uncertain.

2.3.3 Other systematics

Above, we discussed potential systematic errors which could arise if our simple mass model failed to describe actual halos adequately, due to effects such as twisting or satellite contamination. Besides these systematics in the mass model, our proposed measurement will also be liable to possible observational systematics associated with the lensing measurement. One obvious observational source of systematic error is point spread function (PSF) anisotropy. The PSF determines how the actual shape of a galaxy on the sky is related to the observed shape of a galaxy, measured by a camera on a telescope possibly beneath the distorting effects of the Earth’s atmosphere. Our ability to measure the PSF is frequently a limiting factor in our ability to measure the true shapes of weakly lensed galaxies, which degrades our ability to measure shear. In principle, this could be disastrous for the halo ellipticity measurement we have proposed. For example, if the PSF were uniformly anisotropic across the virial radius of a lens halo, leading to a spurious, uniform shear, this would exactly mimic the ellipticity signal we are seeking to detect. In practice, however, mitigating such effects in galaxy-galaxy lensing measurements should be straightforward, as long as the shape of the PSF is not strongly correlated with the number of foreground lens galaxies. For example, we can assess the extent of such PSF anisotropies by stacking on random sky points instead of lens galaxies. Even if PSF anisotropies are present, the ellipticity signal should appear as an excess correlation with lens galaxies, above what is seen around random sky points. As discussed in §2.3, the same test would also help remove the effects of masses uncorrelated with the lens galaxies.

Another potential astrophysical contamination of the signal arises from intrinsic alignment be-

tween galaxies [95–101]. Galaxies that form and evolve in the same local environment may be systematically aligned with each other due to long range tidal effects [102], consequently replicating the correlation that is produced by gravitational lensing. This effect can manifest itself in two ways, (i) nearby source galaxies can be preferentially aligned with each other, and (ii) lens-source pairs can be physically associated with each other, if (for example) a fraction of source galaxies are satellites of the lensing, foreground galaxy. Both these problems can be mitigated using redshift information, for example by excluding galaxy pairs with similar redshifts. Observationally, the contamination of galaxy-galaxy lensing due to alignments from lens-source correlations produced by photometric redshift errors has been shown to be exceedingly small in SDSS [103]. Stacking on random points, as discussed above, would also help to quantify and remove the effect of alignments of pairs of source galaxies.

Magnification of lenses could also produce a systematic effect on the signal. Our estimator correlates the number density of foreground galaxies, n_g , to the shear at two positions in the sky, γ_1 and γ_2 . The foreground galaxies (lenses) are lensed by matter distribution between the observer and the lens along the line of sight. This causes a modification of the clustering of lenses due to cosmic magnification along the line of sight. The variation from the unlensed number density is, to lowest order, linear in the lensing convergence, $\kappa_{<}$ [91]. In addition the shear itself has a contribution from the matter density, integrated along the line of sight to the redshift of the source. The combined effect therefore contributes to the 3-point correlator, $\langle n'_g \gamma'_1 \gamma'_2 \rangle$, terms like $\langle \kappa_{<} \gamma_{1<} \gamma_{2<} \rangle$. These third order shear correlations have been measured to be less than 10^{-7} [104, 105] for aperture scales of $\theta \sim 1'$ and source redshift $z_s \sim 1$, while [91] predict an upper bound to the effect of magnification of lenses on three point statistics of 10^{-8} for sources at $z_s \sim 0.4$. Therefore, it appears that this effect will not significantly contaminate the measurement of halo ellipticity.

Photometric redshifts can also lead to systematic errors for our estimator. Redshift errors produce errors in Σ_{crit} , which become density errors when converting from shear γ to $\Delta\Sigma$. Because our ellipticity estimator is nonlinear, this can bias the inferred ellipticity. Assuming that photometric redshift errors of source galaxies do not correlate with projected separation to the lens galaxies,

then a fractional systematic error in Σ_{crit} of size $\delta = \delta\Sigma_{\text{crit}}/\Sigma_{\text{crit}}$ produces a fractional error in the ellipticity of order δ . For modern surveys, systematic errors in photometric redshifts are expected to be at the level of $\lesssim 1\%$ [106–108], indicating that this source of bias will likely be subdominant compared to other systematics.

Another potential systematic is the effect of baryons, which can act to modify the halo axis ratios on small scales $\lesssim 0.25r_{\text{vir}}$ [109, 110]. Judging from Figure 2.3, our estimator is most sensitive to the ellipticity at somewhat larger radii, suggesting that baryonic effects will be limited. In principle, we can suppress any baryonic effects on our estimator by restricting the range of integration (r_{min} and r_{max} in Eqns. (2.17)–(2.19)) to exclude small-scale regions that may be contaminated [111]. Alternatively, given sufficient signal to noise, one could try to measure the ellipticity as a function of radius by subdividing the sample, for example by comparing triangles at large vs. small separation. Besides constraining any radial variation in ellipticity, 3-point lensing could also probe any twist, i.e. misalignments between the principal axes at small radii vs. large radii. We defer such possibilities to future work.

2.4 Discussion

We have shown that the lens-shear-shear three-point correlation function can be used to extract the ellipticity of dark matter halos, without the need to align the light profiles of galaxies that are being stacked. Using a simple model of the projected surface density profiles of dark matter halos, we constructed an estimator for halo ellipticity that sums over all triangular configurations of the 3-point function. We validated our estimator using simulated halos from the Bolshoi cosmological simulation, showing that the shear-derived estimator yields results consistent with the ellipticity measured directly from the particle data. We investigated potential sources of systematic error, and argued that they should be small, at the $\sim 5\%$ level, well below theoretical uncertainties. We also estimated the signal to noise ratios expected for imaging surveys, and found that deep imaging surveys should be able to detect halo ellipticities. The total signal to noise scales with the number

of lens-source-source triplets as $N_t^{1/2} \propto n_l^{1/2} n_s$, meaning that deep imaging surveys with large effective number densities of sources will be most sensitive. Ongoing surveys like PanSTARRS, DES, and HSC may be able to detect halo ellipticities at the $\sim 2\sigma$ level, while future surveys like Euclid or LSST should have sufficient sensitivity for a significant ($> 3\sigma$) detection. The same surveys will, of course, be able to measure 2-point galaxy-galaxy lensing with far greater signal to noise than 3-point lensing. However, if 2-point estimators are limited by systematic uncertainties, as suggested by theoretical work on galaxy-halo misalignments [88], then 3-point lensing could prove to be a useful probe of halo anisotropy. Indeed, a comparison of the ellipticity determined by the 2-point estimator vs. the 3-point estimator could be used to determine the typical misalignments between galaxies and their halos. If a galaxy is misaligned with its halo by angle θ_{mis} , then the ellipticity determined from the 2-point estimator is lowered by a factor $\cos(2\theta_{\text{mis}})$. Therefore, the average misalignment angle may be inferred as $\varepsilon_{2\text{pt}}/\varepsilon_{3\text{pt}} = \langle \cos(2\theta_{\text{mis}}) \rangle$.

In this section, we have investigated one particular application of the measurement of 3-point correlation functions. As theoretical work has shown [91], high-order correlation functions contain significant amounts of information above and beyond that encoded in better studied 2-point functions. The advent of deep, wide-area imaging surveys is now making the measurement of these high-order correlations practical across a range of spatial scales, suggesting that this will be a fruitful area of research for years to come.

Chapter 3

Assembly bias in CFHTLenS

Halo assembly bias is a robust prediction of nonlinear structure formation, however the detection of this effect in real galaxies has remained elusive. In this paper we analyze lensing and clustering data from the CFHTLenS survey in an attempt to detect assembly bias in galaxy clustering. Using galaxy-galaxy lensing, we construct samples of red and blue galaxies at $0.2 < z < 0.4$ that should have similar halo masses, after excluding potential satellites of candidate groups and clusters. We compare the 2-point clustering of these samples, and find $\sim 1\sigma$ difference in clustering amplitude of blue and red galaxies with similar mass, on ~ 10 Mpc scales. This result appears not to be in line with the sense of assembly bias expected in halos of this mass. We caution that residual satellite contamination or systematic errors in photometric redshifts could possibly affect the results. To mitigate these effects, we reject $\sim 80 - 90\%$ of our galaxy sample using several different satellite exclusion criteria. The burial of the assembly bias signal could be attributed to the limited sample size, or galaxy color does not correlate well with the halo formation history. Larger surveys like DES or HSC should easily be able to confirm or rule out the detection of assembly bias in galaxies.

3.1 Introduction

In the standard cosmological model, the spatial clustering of galaxies reflects the clustering of the dark matter halos hosting those galaxies. Because galaxies and halos arise from perturbations in the matter density, the large-scale clustering of galaxies and halos traces the large-scale clustering of matter, a phenomenon known as biasing [112]. Modeling the bias of observed tracers like galaxies or quasars is a powerful method for inferring the halo occupation of those tracers, which

itself provides valuable insight into the formation of galaxies (see [113–115] for recent examples) or the duty cycle of quasars [61].

The first step in modeling galaxy bias is to compute the bias of dark matter halos, and numerical simulations now provide a reliable method for the accurate prediction of halo bias [116, 117]. Simulations have shown that halo bias depends significantly on halo mass, in a manner that is qualitatively consistent with theoretical expectations from simple models of halo formation [60, 64, 118]. However, simulations have also found that halo bias can depend significantly on halo properties besides virial mass. The first detections of this effect noted a dependence on assembly history [67], leading to the term ‘assembly bias’ to be used to describe this effect. Subsequent work showed that halo bias can depend on a variety of halo properties like concentration [65, 66]. The origin of the age and concentration dependence of halo bias in simulations is now fairly well understood [68].

The dependence of halo clustering on assembly history should, naively, translate into a dependence of galaxy clustering on assembly history, since it would be surprising for galaxy assembly to be unrelated to the assembly of the underlying halos. Indeed, [119] have argued that halo assembly bias is crucial for understanding certain aspects of observed galaxy clustering. However, directly detecting assembly bias is observationally challenging. Assembly bias refers to the dependence of the bias of isolated halos on halo properties at fixed halo mass, which means that a detection of assembly bias requires constructing a sample of central galaxies that all reside in halos of similar mass. Measuring halo masses of individual galaxies is challenging, as is separating central galaxies from satellite galaxies.

Nevertheless, in spite of these difficulties, multiple groups have attempted to detect assembly bias. [120] constructed a group catalog from 2dFGRS data, and found that at fixed group mass, the clustering bias of groups diminishes with increasing star formation rate (SFR) of the central galaxy. However, based on a stacked lensing analysis, [121] found that the groups in the Yang catalog with equal reported masses actually had significantly different lensing masses, and that the mass errors systematically correlated with SFR. This underscores the necessity of obtaining

lensing masses or dynamical masses when attempting to observe assembly bias in galaxies.

Because of these challenges, [121] proposed an alternative method to observe galaxy assembly bias. They constructed samples of galaxies with differing SFR, and used galaxy-galaxy lensing to ensure that the average lensing masses of the two samples were consistent. They then compared the clustering amplitude of galaxy samples with consistent lensing masses but different SFR, and found no significant evidence for any difference in the bias at fixed lensing mass. Their result either implies that the SFR indicator they used does not correlate significantly with halo age, or (more interestingly) that perhaps halo assembly bias does not occur in nature for galaxy-sized halos.

More recently, [122] applied a similar method to redMaPPer galaxy clusters and reported a significant detection of assembly bias. They used stacked lensing to construct cluster samples with consistent lensing masses, and found that the bias of the clusters varies with galaxy concentration at fixed lensing mass. Their results appear to confirm the prediction of assembly bias in cluster-sized halos.

The confirmation of assembly bias in clusters makes the non-detection in galaxies all the more puzzling. To address this apparent contradiction between theory and observations, we revisit the subject of galaxy assembly bias. We apply the methods of [121] to a different galaxy sample, from the publicly available CFHTLenS survey. In §3.2, we describe the CFHTLenS dataset. In §3.3, we discuss our analysis method, and in §3.4 we present our results. Throughout this paper, we assume *WMAP7* [123] cosmological parameters: $(\Omega_M, \Omega_\Lambda, h, \sigma_8, w) = (0.27, 0.73, 0.70, 0.81, -1)$.

Wherever appropriate, the h -scaling of each quantity is shown explicitly using

$$h_{70} = H_0 / (70 \text{ km s}^{-1} \text{ Mpc}^{-1}).$$

3.2 Data

The galaxy-galaxy lensing and galaxy clustering analyses presented in this work are based on galaxy and shear catalogs produced by the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS) [107, 124–126], which are taken from the Wide components of the Canada-France-

Hawaii Telescope Legacy Survey (CFHTLS-Wide). The Wide fields cover an effective area of 154 deg², with four independent fields, W1–4, with individual field areas in the range 23–64 deg².

3.2.1 Lens and source selection

We used the public galaxy catalog from the CFHTLenS catalog query page ¹ with the following parameters: $\text{MASK} \leq 1$, $\text{Flag} < 3$ and $\text{CLASS_STAR} < 0.9$. Details of these parameters are in Appendix C of [125].

From this galaxy catalog, we pre-select lens galaxies at low photometric redshifts, $0.2 \leq z_{\text{lens}} \leq 0.4$, where z is the peak of the photometric redshift probability density function (photo- z PDF), unless explicitly specified otherwise. Following [127], we ensure that the apparent magnitudes of our lens samples are brighter than $m_{i'} \leq 23$. We discard lens galaxies for which any of the following catalog parameters equal -99.0 : stellar mass M_* , or the magnitudes in any of the five optical Sloan Digital Sky Survey (SDSS)-like bands m_{u^*} , $m_{g'}$, $m_{r'}$, $m_{i'}$ and $m_{z'}$.

To reduce the errors in the lensing signal due to the uncertainties in the photometric redshifts, we follow the approach of [128] and require that the difference between the redshift of the sources and the lenses $z_{\text{src}} - z_{\text{lens}} \geq 0.1$. We further ensure the lower bound of the 95% confidence interval of z_{src} to be greater than z_{lens} . We go to the CFHTLenS analysis depth threshold of $m_{i'} \sim 24.7$ for the sources [107, 124, 126].

3.2.2 Satellite galaxy rejection

Because we are interested in halo assembly bias, we need to remove satellite galaxies that are present in the galaxy catalog. Satellites reside in halos of larger mass than similar central galaxies, and will therefore have a larger bias. Since this mass difference and bias difference could mimic assembly bias, it is crucial to minimize contamination from satellites.

CFHTLenS provide a galaxy cluster catalog based on the 3D-Matched-Filter (3D-MF) opti-

¹www.cadc-ccda.hia-ihp.nrc-cnrc.gc.ca/en/community/CFHTLenS/query.html

mized galaxy cluster finder. This finder determines a likelihood map of clusters on the sky, and searches for significantly detected peaks. Details of the 3D-MF algorithm can be found in [129]. The cluster catalog provides positions, redshifts and the detection significance for each cluster candidate. As noted by [130, 131], many of the peaks at significance $< 5\sigma$ will be false detections. Nevertheless, in order to be conservative in excluding satellites, we use all candidates with significance $\geq 3.5\sigma$.

Given a list of clusters, we identify galaxies as satellites if they fall within the virialized region of each cluster. This requires estimating the size of each cluster. We do so using the following power-law relation between cluster mass and peak significance for the 3D-MF algorithm [131]:

$$\log \left[\frac{M_{200c}}{M_{\odot}} \right] = 0.161\sigma + 12.39. \quad (3.1)$$

The lowest-significance clusters, therefore, are expected to have masses $M_{200c} \sim 10^{13} M_{\odot}$. Given M_{200c} , we then calculate the radius of the cluster by

$$r_{200c} = \left[\frac{3M_{200c}}{4\pi(200)\rho_c(z)} \right]^{1/3}, \quad (3.2)$$

where $\rho_c(z)$ is the critical energy density of the universe evaluated at z_{cluster} . We label a galaxy as a potential satellite of an identified cluster if its projected separation from the cluster centroid on the sky satisfies $d \leq A \times r_{200c}$, where A is an arbitrary parameter. Since the boundary of a cluster is not clearly defined, we vary the value of this parameter in the range $1 \leq A \leq 1.5$, to study the effect of removing different percentages of galaxies as satellites. Additionally, along the line-of-sight direction, we also require potential satellite galaxies to fall within $\Delta z \leq 0.2(1 + z)$ of the cluster redshift, which corresponds to $\sim 5\sigma$ of the photo- z scatter for CFHTLenS galaxies [107, 124, 125]. Only those galaxies which do not meet the above two selection criteria are kept as centrals in our analysis. Usually 20%–40% of the galaxies are identified as satellites and removed using these selection criteria.

The 3D-MF catalog allows us to remove satellites of clusters, but satellites of poor groups could still be present in our sample. To suppress their contribution, we impose a further cut on our galaxy sample: we remove all galaxies for which a neighboring galaxy with stellar mass 3 times more massive, i.e. $M_{\star,\text{nbr}}/M_{\star,\text{lens}} \geq 3$ (or 5 times), is found within projected separation $r_{\text{excl.}}$, where $r_{\text{excl.}}$ is varied between 300 to $500h_{70}^{-1}\text{kpc}$. This exclusion further removes $\sim 50\%$ of the galaxies, leaving us with significantly less than half of the original sample. These cuts may be overly conservative, but they should leave us with a reasonably pure selection of central galaxies.

3.2.3 Selection of red and blue galaxy samples

With our sample of central lens galaxies, we can now attempt to detect assembly bias. To do so, we require proxies for halo mass, and halo assembly history, since the dark matter halo properties themselves are not directly observable. In this work, we use galaxy stellar mass as a proxy for halo mass, and galaxy color as a proxy for assembly history, under the assumption that redder galaxies have older stellar populations and therefore might reside in older dark matter halos.

Accordingly, we split our lenses into red and blue samples. We follow the definition of red and blue galaxies from [127], with $T_{\text{BPZ}} \leq 1.5$ and $2.0 \leq T_{\text{BPZ}} \leq 4.0$ respectively. Here, T_{BPZ} is a parameter which represents the spectral type of a CFHTLenS galaxy corresponding to its redshift, and is obtained from the best-fitting SED of the galaxy determined by the Bayesian photometric redshift (BPZ) code [125, 132].

Given our red and blue samples, we next construct stellar mass bins for the two samples that give similar halo masses. We select galaxies based on their stellar masses $M_{\star}$, and adjust the widths and ranges of the $M_{\star}$ cuts for the red and blue galaxy samples separately, to ensure that the lensing profiles for the red and blue samples are consistent, as determined by galaxy-galaxy lensing (see §3.3.1 below). We attempt to construct galaxy samples with masses $M \lesssim 10^{12}M_{\odot}$, the mass regime which the effect of assembly bias is expected to be significant. We perform simple Navarro-Frenk-White (NFW) profile fits [45, 46] at $z = 0.3$, whose projected surface mass density profiles are given in [133], together with baryonic lensing signals [134] from $\sim 20 - 200h_{70}^{-1}\text{kpc}$ by

χ^2 analyses [127] to ensure our galaxy samples are in this low mass regime. At low halo masses, however, the lensing signal becomes weak, so we are forced to use relatively wide stellar mass bins to achieve a reasonable signal-to-noise in the lensing measurements. Typically, as discussed below, the widths of the stellar mass bins that we use are $M_{\star,\text{max}}/M_{\star,\text{min}} \lesssim 10$, where $M_{\star,\text{max}}$ and $M_{\star,\text{min}}$ represent the central galaxies with the most and the least massive stellar masses in the sample respectively. Central galaxies in halos of this mass range have been found to follow a power-law relationship between halo masses and stellar masses, i.e. $M \propto M_{\star}^{0.6}$ [128, 135–137], which implies that a ten-fold increase in stellar mass would roughly correspond to about a range of $\Delta M/M \sim 4$ in the halo masses.

For our CFHTLenS galaxy samples, we find that blue centrals tend to have higher stellar masses than red centrals with similar host halo masses, in agreement with the results of [135]. This turns out to present a difficulty in our analysis, for the following reason. To ensure that the galaxies in our red and blue samples have consistent redshift distributions, we measure their cross-correlation coefficients r_c^2 by comparing the angular cross-correlations with the angular auto-correlations, described in more detail in §3.3.3. Specifically, we define

$$r_c^2(\theta) = \frac{w_{c,\text{all}}^2(\theta)}{w_c(\theta) w_{\text{all}}(\theta)}, \quad (3.3)$$

where the index c refers either to red or blue, w_c refers to the auto-correlation of sample c , w_{all} refers to the auto-correlation of all galaxies in the prescribed photo- z range, and $w_{c,\text{all}}$ refers to the angular cross-correlation between sample c and all galaxies. If the redshift distribution of sample c is consistent with the redshift distribution of all galaxies, then we would expect a strong cross-correlation, $r_c^2 \approx 1$. For most of the luminosity and color bins, this is indeed the case. However, for the brightest blue galaxies with r' -band absolute magnitude $M_{r'} < -21.0$, we find a much weaker cross-correlation coefficient, with $r_c^2 \sim 0.3$. At present, we do not understand the origin of this result, but one possible explanation could be systematic errors in the photo- z 's for this luminosity bin [138]. To avoid this possibility, we exclude blue galaxies with $M_{r'} < -21.0$ from our lens

samples. However, as galaxy stellar mass is strongly correlated with luminosity [139], excluding the brightest blue galaxies implies that the blue galaxies with the most massive stellar masses would be excluded, and since blue galaxies have much smaller halo masses than red galaxies with similar stellar masses, this makes it difficult to find red and blue samples with matching lensing profiles. We will return to this issue in §3.4.

3.2.4 Simulation data

To validate our analysis, and to provide theoretical comparisons for our measurements described below, we have applied much of our analysis to publicly available mock galaxy catalogs provided by [140], which were constructed on the ROCKSTAR [141, 142] halo catalogs ² of the Bolshoi N -body simulation [93].

The Bolshoi simulation was run in a periodic cube with volume $250h^{-1}\text{Mpc}$ on a side with 2048^3 dark matter particles, in a ΛCDM cosmology with $(\Omega_M, \Omega_\Lambda, \Omega_b, h, \sigma_8, n_s) = (0.27, 0.73, 0.042, 0.70, 0.82, 0.95)$. The simulation has mass resolution of $M_{\text{particle}} = 1.35 \times 10^8 h^{-1} M_\odot$ and force resolution of $1.0h^{-1}\text{kpc}$. The dark matter particles were traced from $z = 80$ to 0. For further details of the Bolshoi simulation, we refer the reader to Riebe et al. [143].

From this simulation, [140] have constructed mock galaxy catalogs built on the ROCKSTAR catalogs publicly available at cosmosim.org. The mock galaxy catalog uses the age-matching formalism to relate galaxy properties such as stellar mass and $(g - r)$ color to the properties of dark matter halos and subhalos. [119, 140] find that these catalogs reproduce the galaxy color probability distribution function, galaxy clustering, and lensing signals measured in SDSS, although [114] note a discrepancy in this model’s lensing predictions for locally brightest galaxies compared to observations. For our purposes, the precise validity of these catalogs is unimportant; we merely use them to check that our lensing analysis should not introduce systematic biases in mass determination and how the clustering behavior of the samples changes by imposing different satellite removal criteria.

²<http://cosmosim.org>

We calculate the mass profiles around the galaxies by downloading the Bolshoi particle data from the Particles416 table available in cosmosim.org. We project particles within $1.0h^{-1}\text{Mpc}$ around each halo center and calculate the projected differential mass profile $\Delta\Sigma(r)$. We then stack the mass profiles for the halo samples to provide mock simulations of the galaxy-galaxy lensing analysis described below.

We implement similar satellite exclusion schemes in the mock data as in CFHTLenS. Firstly, we identify the massive clusters with $M > 10^{13}h^{-1}M_{\odot}$ using the Bolshoi halo catalog at $z = 0$. Nevertheless, [129] report different detection rates for 3D-MF algorithm on galaxy clusters with different mass ranges. In order to mimic the completeness of the 3D-MF method to identify the satellites residing in massive clusters, we randomly select mock halos at different mass ranges according to the detection rates (with photometric redshift errors) indicated in Table 3 of [129]. Next, we exclude a mock galaxy as a potential satellite of a mock cluster if its projected separation from the cluster centroid satisfies $d \leq A \times r_{200c}$, where $1 \leq A \leq 1.5$. We further impose the additional cut identical to what we do in CFHTLenS to remove galaxies in poor groups. We then calculate the 2-point auto-correlations (details in §3.3.3) for the mock galaxies and see why satellite exclusion is necessary in our analysis.

3.3 Methodology

To detect assembly bias, we require a method to determine the (average) lensing profiles of our red and blue galaxy samples. Given red and blue samples with consistent lensing profiles (and hence halo masses), we can then compare their clustering to search for assembly bias. We use galaxy-galaxy lensing to measure lensing profiles, and use projected angular 2-point auto-correlations (2PCF) to infer the relative biases of the samples.

3.3.1 Galaxy-galaxy lensing

Galaxy-galaxy lensing measures the cross-correlation between foreground lens galaxies and the apparent distortions to the shapes of background source galaxies, induced by weak lensing shear. Specifically, the stacked azimuthally averaged tangential shear $\langle\gamma_+\rangle$ profile around lenses is related to the projected surface mass density of the lenses by [70, 144]:

$$\langle\gamma_+\rangle = \frac{\Delta\Sigma(r)}{\Sigma_{\text{crit}}} = \frac{\bar{\Sigma}(< r) - \Sigma(r)}{\Sigma_{\text{crit}}}, \quad (3.4)$$

where $\bar{\Sigma}(< r)$ is the averaged projected surface mass density enclosed by radius r , and $\Sigma(r)$ is the mean projected surface mass density at radius r . The lensing critical density Σ_{crit} is defined as

$$\Sigma_{\text{crit}} = \frac{c^2}{4\pi G} \frac{D_s}{D_{ds}D_d}, \quad (3.5)$$

where c and G are speed of light and gravitational constant, D_d and D_s are the angular diameter distances to the lens and the source respectively, while D_{ds} is the angular diameter distance from the lens to the source. For lens galaxy j and source galaxy k , we compute the following quantities of the lens-source pair:

$$g_{jk}(r) = \frac{c^2}{4\pi G} \frac{w_k e_{jk}^+(r) \langle\eta_1\rangle_{jk}(r)}{1 + m_k}$$

$$W_{jk} = w_k \langle\eta_2\rangle_{jk}(r), \quad (3.6)$$

where w_k is the LENSFIT weight for the shear profile of source k , $e_{jk}^+(r)$ is the tangential ellipticity of the source k at projected physical distance r about lens j .

The CFHTLenS galaxy catalog provides the raw ellipticity components, e_1 and e_2 , along the celestial coordinate frame. These quantities are known to be affected by a small additive calibration c_2 and a multiplicative calibration m [124, 126, 145]. For the additive bias, the one affecting e_1 is consistent with zero, but not for the one affecting e_2 , i.e. $\langle c_2 \rangle \neq 0$. c_2 must be subtracted from

e_2 for individual sources. e_2 is defined relative to a decreasing RA in CFHTLenS, e_2 has to be multiplied by -1 so the conventional definition of RA/DEC reference frame is followed [125]. From e_1 , e_2 and the angle of the great circle passing through the lens-source pair with declination, we can calculate $e_{jk}^+(r)$. Following [127], we check that the stacked lensing profiles of the galaxy samples do not change significantly if we do not apply the additive calibration.

For the multiplicative calibration correction m , instead of applying the calibration corrections statistically on all the galaxy pairs as suggested in [126], we apply the corrections on individual sources, as is implied in $g_{jk}(r)$ in Eqn. 3.6. We keep only the source galaxies with $-0.5 \leq m \leq 0.2$. This range of m is chosen to maintain the source count and avoid the spurious shear signal by sources with extreme values of m , especially those with $1 + m \rightarrow 0$. We check that the stacked lensing profiles are consistent in the cases when we apply the calibration corrections individually and statistically on all the galaxy pairs.

We make use of the normalized photo- z PDFs $p(z)$, i.e. $\int_0^\infty p(z)dz = 1$, for the lens j and source k to evaluate $\langle \eta_1 \rangle_{jk}(r)$ and $\langle \eta_2 \rangle_{jk}(r)$. The residual photo- z calibration bias in the lensing measurement should be improved by using the full $p(z)$ [146]. We firstly define $E_1(z_j)$ and $E_2(z_j)$ as:

$$\begin{aligned} E_1(z_j) &= \int_{z_j}^\infty p(z_j)p(z_k) \frac{D_{ds}(z_j, z_k)D_d(z_j)}{D_s(z_k)} dz_k \\ E_2(z_j) &= \int_{z_j}^\infty p(z_j)p(z_k) \left[\frac{D_{ds}(z_j, z_k)D_d(z_j)}{D_s(z_k)} \right]^2 dz_k. \end{aligned} \quad (3.7)$$

Here we get the lensing efficiencies of the lens-source pair as functions of the redshift z_j of lens j . We also compute the projected angular separation θ_{jk} of lens j and source k . We then make use of the photo- z PDF of lens j and θ_{jk} to calculate the lensing efficiencies of the lens-source pair as functions of the projected physical distance r , i.e. $E_1(z_j)$ and $E_2(z_j)$ can be mapped into $\langle \eta_1 \rangle_{jk}(r)$ and $\langle \eta_2 \rangle_{jk}(r)$ respectively.

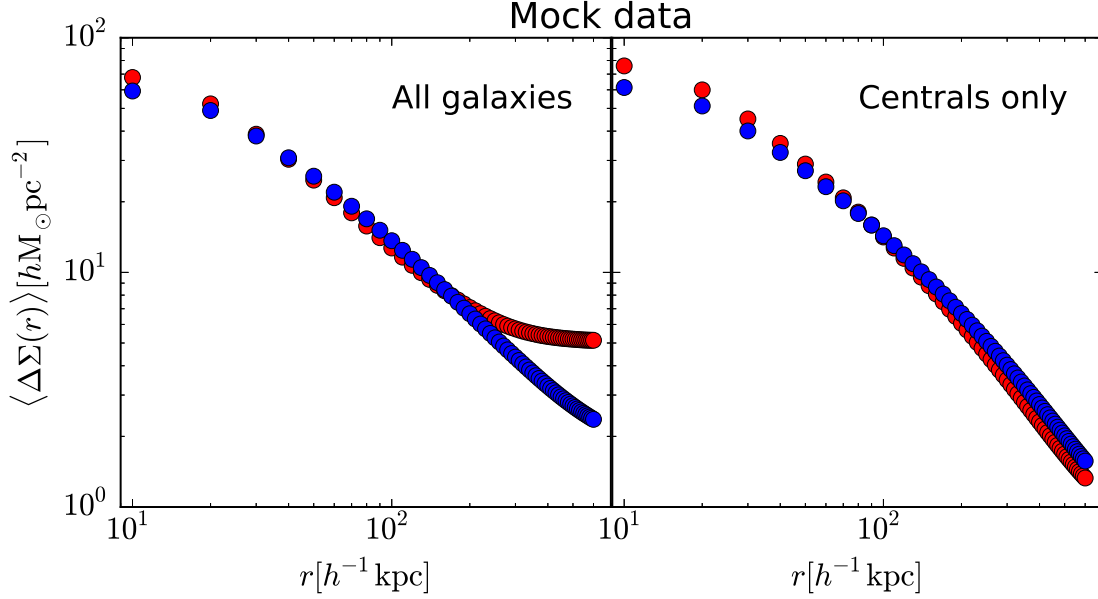


Figure 3.1: The stacked tangential shear profiles $\langle \Delta \Sigma(r) \rangle$ for mock galaxies, after the $M_{\star}$ stellar mass ranges have been adjusted to make the shear profiles consistent within $200h^{-1}\text{kpc}$. Left: $\langle \Delta \Sigma(r) \rangle$ for samples including both centrals and satellites. Right: $\langle \Delta \Sigma(r) \rangle$ for only central galaxies, i.e. with all satellites removed. The red (blue) data points represent the red (blue) sample.

Next, for lens j , we sum up the shear signal for the source galaxies around it by

$$\Delta \Sigma(r)_j = \frac{\sum_k g_{jk}(r)}{\sum_k W_{jk}}. \quad (3.8)$$

We estimate the error of the shear signal at distance r for each lens, $\delta \Delta \Sigma(r)_j$, by bootstrapping the source galaxies around the lens 5×10^2 times. We use this error estimate $\delta \Delta \Sigma(r)_j$ to inverse-variance weight the shear signal from individual lenses contributing to the stacked shear profile.

We then stack the shear signals of the lens galaxies by calculating the weighted mean, i.e.

$$S(r) \equiv \langle \Delta \Sigma(r) \rangle = \frac{\sum_j \Delta \Sigma(r)_j [\delta \Delta \Sigma(r)_j]^{-2}}{\sum_j [\delta \Delta \Sigma(r)_j]^{-2}}, \quad (3.9)$$

the error of the stacked mass profile $\langle \Delta \Sigma(r) \rangle$ is obtained by bootstrapping the lens samples by

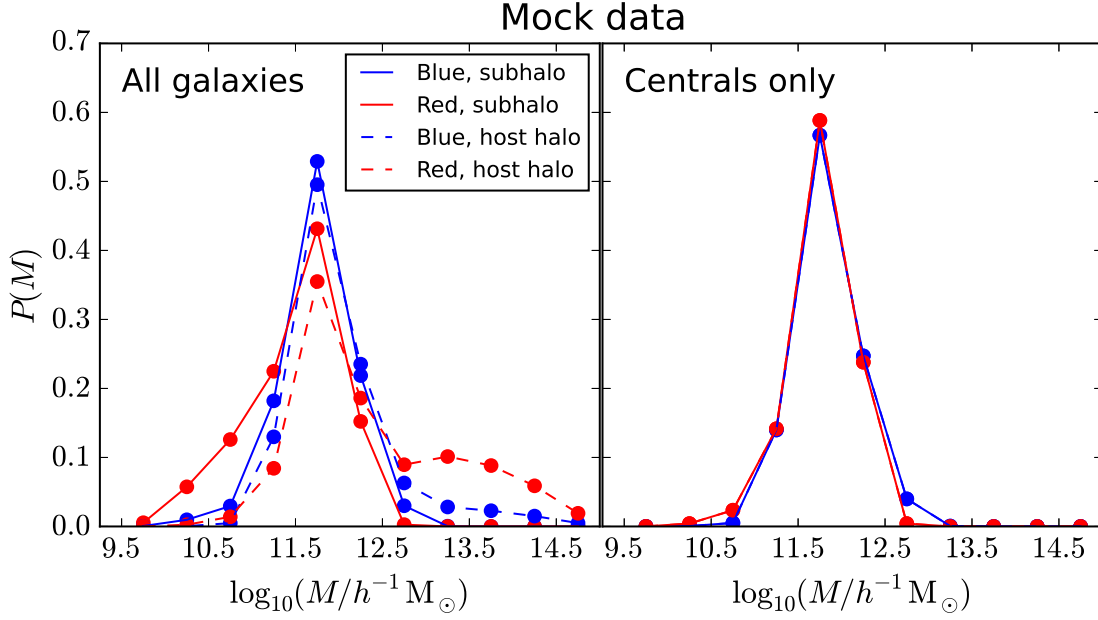


Figure 3.2: These line plots present the fraction of mock halos $P(M)$ in mass bin $(M, M+dM)$ for the galaxy samples with consistent lensing profiles shown in Fig. 3.1. Red and blue lines represent the red and blue galaxies respectively. Solid and dashed lines represent the mass distributions for the subhalos and the host halos respectively. Left: $P(M)$ for all mock galaxies, the significant proportion at $M \geq 10^{12.5} h^{-1} M_{\odot}$ indicates there are satellites in the samples which are located in massive host halos. Right: $P(M)$ for only central galaxies. By construction, the halos in these samples are all host halos. After satellite removal, the red and blue samples have consistent halo mass distributions.

$N_S = 10^3$ times. We keep track of the full covariance matrix [147], i.e.

$$C_{m,n} = \frac{1}{N_S - 1} \sum_{p=1}^{N_S} [S_p(r_m) - S(r_m)] [S_p(r_n) - S(r_n)], \quad (3.10)$$

where $S(r)$ is the stacked mass profile of the galaxy sample, $S_p(r)$ is the p th bootstrap stacked sample, and r_m and r_n are the m th and n th radial bins respectively. The bootstrap errors, σ_m can be obtained from the diagonal elements of the covariance matrix, i.e. $(\sigma_m)^2 = C_{m,m}$.

As a systematic check, we measure the cross shear, which is 45° to that of the tangential shear, and find that it is consistent with zero.

3.3.2 Finding consistent mass profiles

We measure stacked tangential shear profiles $\langle \Delta \Sigma(r) \rangle$ from $\sim 20 - 200 h_{70}^{-1} \text{kpc}$, spaced logarithmically in 7 radial bins. We minimize $\Delta \chi_S^2$ [148], defined as

$$\Delta \chi_S^2 = \sum_{m,n} [S^{\text{red}}(r_m) - S^{\text{blue}}(r_m)] (C_{m,n}^{\text{red}} + C_{m,n}^{\text{blue}})^{-1} [S^{\text{red}}(r_n) - S^{\text{blue}}(r_n)], \quad (3.11)$$

where $S^{\text{red}}(r) - S^{\text{blue}}(r)$ is the difference of the stacked lensing profiles for the red and blue samples, and $(C_{m,n}^{\text{red}} + C_{m,n}^{\text{blue}})^{-1}$ is the inverse of the sum of the covariance matrices for the red and blue samples respectively. We iterate this step by adjusting the stellar mass cuts of the red and blue samples until $\Delta \chi_S^2$ is minimized. We then check to make sure the lensing profiles of the two samples are consistent with each other, which implies the two samples should have similar (average) halo masses.

Testing the procedure

To validate our method for producing consistent shear profiles for red and blue galaxies, we apply the same technique to mock galaxies from the [119] catalog described in §3.2.4. We follow [119] and [149], and apply the following stellar mass-dependent color cut to separate the red and blue mock galaxies, i.e.

$$(g - r) = 0.76 + 0.15 [\log_{10}(M_{\star}) - 10.0]. \quad (3.12)$$

After distinguishing the galaxies into red and blue categories, we repeat the same exercise on the mock galaxy catalog as we performed on the CFHTLenS observational data, i.e. we adjust the $M_{\star}$ stellar mass ranges and widths for the red and blue galaxies to allow their mass profiles agree with each other, and therefore the halo masses. We ensure the widths of $M_{\star}$ are comparable to those of the galaxy samples we select in CFHTLenS data, as then we can estimate the mass distribution of the halos included in the samples.

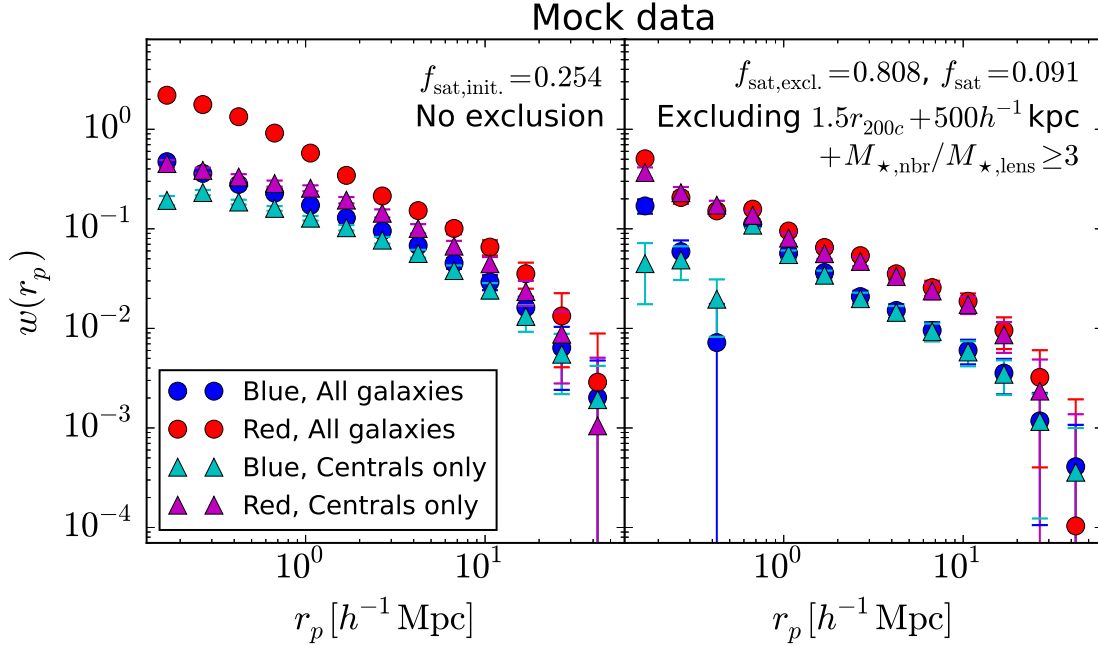


Figure 3.3: The 2PCFs of the mocks after pairing the lensing profiles of the red and blue samples. The red (blue) dots represent the red (blue) sample of the “All galaxies” sample, while the magenta (cyan) triangles represent the red (blue) sample of the “Central only” sample. The stellar mass bins ($\log_{10}(M_{\star}/M_{\odot})$) chosen in the mock have similar widths as those chosen in CFHTLenS data. Left: The 2PCFs with no satellite exclusion. The stellar mass ranges of the blue “All galaxies” and “Centrals only” samples are $[9.925, 10.925]$ and $[9.943, 10.943]$ respectively, whereas both of the red samples are kept at $[9.8, 10.8]$. The severe satellite contamination $f_{\text{sat,init.}} = 25.4\%$ can mimic the assembly bias signal, as demonstrated by the larger difference in 2PCFs of the red and blue galaxies for the “All galaxies” sample than those for the “Centrals only” sample. Right: The 2PCFs with the most stringent satellite exclusion adopted in this paper. The stellar mass ranges of the blue “All galaxies” and “Centrals only” samples are $[9.935, 10.935]$ and $[9.928, 10.928]$ respectively, whereas both of the red samples are kept at $[9.8, 10.8]$. Under this exclusion criterion, we have excluded $f_{\text{sat,excl.}} = 80.8\%$ of the satellites, and left with satellite contamination $f_{\text{sat}} = 9.1\%$. Nonetheless, the consistency between the 2PCFs of the “All galaxies” and “Centrals only” samples suggests that the residual contamination present in the sample is insufficient to solely account for the difference in the bias signals in the mock.

Fig. 3.1 shows the stacked tangential shear profiles $\langle \Delta\Sigma(r) \rangle$ for the mock catalog. Both the cases with all satellites removed (Centrals only) and without any satellite exclusion (All galaxies) are shown. The shear profiles for the red and blue samples are aligned by minimizing $\Delta\chi_S^2$. The stacked shear profiles are in good agreement with each other within $r \sim 200h^{-1}\text{kpc}$, which coincides with the virial radius of the halo samples at $M \leq 10^{12}h^{-1}M_\odot$. This implies the two samples should have consistent halo masses. At scales beyond $200h^{-1}\text{kpc}$, the plateauing behavior of the stacked shear profiles for the “All galaxies” sample suggests there is significant satellite contamination.

Fig. 3.2 shows the normalized halo mass distributions for the red and blue samples with matching lensing profiles both for the “All galaxies” and “Centrals only” cases. There is a significant proportion of host halos reside at $M \geq 10^{12.5}h^{-1}M_\odot$ for the “All galaxies” sample, suggesting there is satellite contamination. Particularly, the red samples suffer from more severe satellite contamination than the blue samples. For the “Central galaxies” case, the curves are narrower than those in “All galaxies” case, the red and blue curves have similar shapes. Moreover, a majority of halos in this case reside in $10^{11} \leq (M/h^{-1}M_\odot) \leq 10^{12}$, which is the target mass regime for the selected sample. This suggests that our procedure for matching the stacked lensing profiles by $\Delta\chi_S^2$ minimization does indeed produce samples with consistent halo mass distributions.

3.3.3 Correlation function measurements

We calculate the 2PCF using the standard Landy-Szalay estimator [150].

$$w(\theta) = \frac{N_{DD} - 2N_{DR} + N_{RR}}{N_{RR}}, \quad (3.13)$$

where N_{DD} , N_{DR} and N_{RR} represent the normalized numbers of data-data, data-random and random-random pairs in a given angular separation bin respectively. We calculate $w(\theta)$ using 20 angular bins, spaced logarithmically between 0.003° and 3° .

To calculate N_{DR} and N_{RR} , we generate random galaxy catalogs for the four CFHTLenS Wide

patches, using the same areas and masked regions of the Wide patches, discarding any randoms falling within masked areas, $\text{MASK} > 1$. We use 10 times more random galaxies than data galaxies in order to suppress Poisson noise from our random realizations.

To estimate the errors of the 2PCF, we compute the covariance matrix by the jackknife resampling method. We follow [125] and use $N_w = 54$ jackknife samples across all four Wide patches. The covariance matrix of the 2PCF is determined by [147, 151, 152]

$$C_{m,n}^w = \frac{N_w - 1}{N_w} \sum_{p=1}^{N_w} [w_p(\theta_m) - w(\theta_m)] [w_p(\theta_n) - w(\theta_n)], \quad (3.14)$$

where $w(\theta)$ is the weighted average (by number of data-data pairs in each Wide patch) of the correlation measurement from the entire galaxy sample, $w_p(\theta)$ is the weighted average obtained by omitting the p th subsample of the data, and θ_m and θ_n are the m th and n th angular bins respectively. Similar to bootstrap errors, the jackknife bin errors, σ_m^w can be obtained from the diagonal elements of the covariance matrix, $C_{m,m}^w$. We check our 2PCF calculations reproduce the results of [125] when applied to the same galaxy sample (see their Fig. 15).

Fig. 3.3 shows the 2PCFs measured from the mock galaxy samples. $f_{\text{sat,init.}}$ represents the satellite contamination in the initial sample, i.e. without any satellite removal. $f_{\text{sat,excl.}}$ is defined as the ratio between the number of satellites being excluded to the number of satellites in the original sample, whereas f_{sat} is defined as the ratio between number of satellites to the total number of galaxies in the final sample, hence it indicates the level of residual satellite contamination after the satellite removal procedures. The stellar mass bins ($\log_{10}(M_*/M_\odot)$) chosen for the mock data have similar widths as those chosen in CFHTLenS data. In the left plot, there is no satellite exclusion applied to the “All galaxies” (dots) and “Centrals only” (triangles) samples (See §3.3.2). The stellar mass ranges of the blue “All galaxies” and “Centrals only” samples are $[9.925, 10.925]$ and $[9.943, 10.943]$ respectively, whereas both of the red samples are kept at $[9.8, 10.8]$. The “Centrals only” generally have lower amplitudes than the “All galaxies” at the same radial bins, suggesting that removing satellites would decrease the clustering signals of the galaxy samples. $f_{\text{sat,init.}} =$

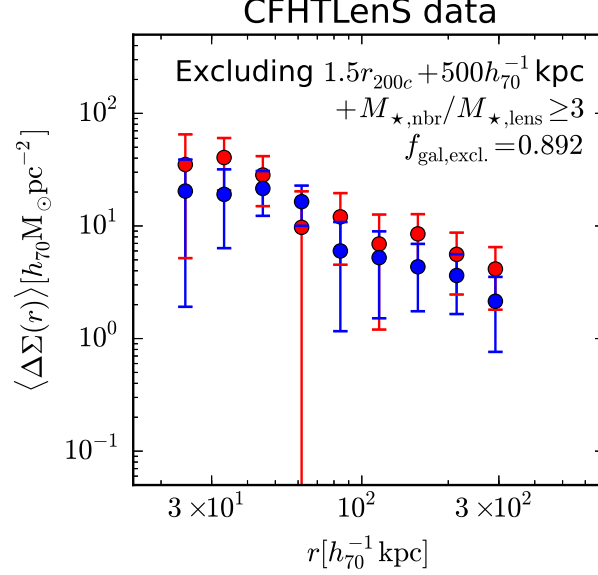


Figure 3.4: The stacked shear profiles of CFHTLenS data with the most stringent satellite exclusion adopted, i.e. $A = 1.5$, and if there is any neighboring galaxy whose stellar mass is at least 3 times more massive than the lens galaxy within $500h_{70}^{-1}\text{kpc}$ projected distance. The red (blue) dots represent the red (blue) sample. The stellar mass bins have been adjusted to allow the lensing profiles of the red and blue samples to be consistent by minimizing $\Delta\chi_S^2$. There are 4376 and 8206 galaxies in the red and blue samples respectively, and the stellar mass ranges ($\log_{10}(M_{\star}/M_{\odot})$) for the red and blue samples are $[9.05, 10.05]$ and $[9.5, 10.55]$ respectively. The results of $\Delta\chi_S^2$ minimization are shown in Table A.1.

25.4% of the galaxies are satellites in the mock sample before any exclusion is implemented. Severe satellite contamination would mimic the assembly bias signal, as demonstrated by the larger difference between the 2PCFs of the red and blue galaxies in the “All galaxies” sample than those of the “Centrals only” sample. Therefore satellite removal is indispensable in our analysis.

Our procedure for excluding satellites reduces the assembly bias measured from the simulations, but does not eliminate it entirely. We verify this statement by applying the same exclusion criteria on both “All galaxies” and “Centrals only” samples. The right plot of Fig. 3.3 shows the 2PCFs of the red and blue galaxies with the most stringent satellite removal scheme adopted in this paper applied, i.e. we firstly exclude the potential satellites if they are located within $1.5r_{200c}$ ($A = 1.5$) of the mock clusters in the projected plane, and we further reject the satellites in poor groups by searching if there is any neighboring galaxy with at least more than 3 times stellar

CFHTLenS data

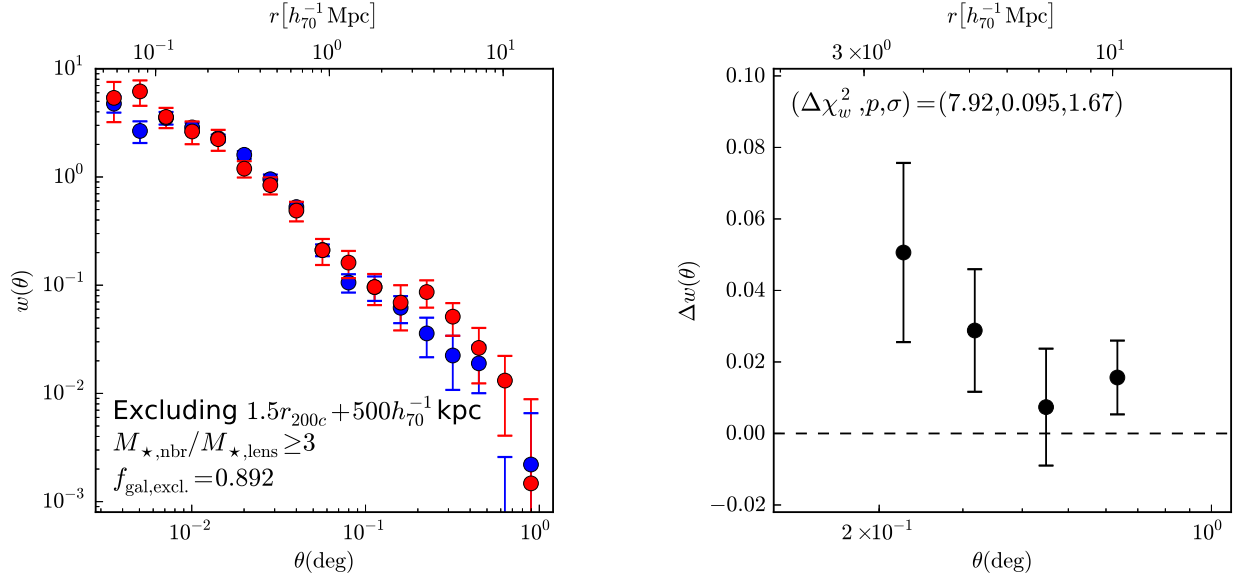


Figure 3.5: Left: The 2PCFs of CFHTLenS with the following satellite exclusion criterion applied: $A = 1.5$, and if there is any neighboring galaxy whose stellar mass is at least 3 times more massive than the lens galaxy within $500h_{70}^{-1}\text{kpc}$ projected distance. The red (blue) dots represent the red (blue) sample. Right: The difference of the 2PCFs $\Delta w(\theta) \equiv w^{\text{red}}(\theta) - w^{\text{blue}}(\theta)$ corresponding to the left panel at $3 - 11h_{70}^{-1}\text{Mpc}$. The dashed lines show the $\Delta w(\theta) = 0$ level. The significance of difference of the 2PCFs is 1.67σ . The physical distance at $z = 0.3$ is shown at the top abscissae of the panels.

mass of the lens galaxy within $r_{\text{excl.}} = 500h^{-1}\text{kpc}$ projected distance. The stellar mass ranges $\log_{10}(M_{\star}/M_{\odot})$ of the blue “All galaxies” and “Centrals only” samples are $[9.935, 10.935]$ and $[9.928, 10.928]$ respectively, whereas both of the red samples are kept at $[9.8, 10.8]$. Through these procedures, $f_{\text{sat,excl.}} = 80.8\%$ of the satellites are excluded and the final satellite contamination is $f_{\text{sat}} = 9.1\%$ in this case. However, the residual satellite contamination after the satellite removal is insufficient to solely account for the assembly bias signal found in the mocks. For instance, when the same removal scheme is applied on both the “All galaxies” and the “Centrals only” (i.e. *without any satellites*) samples, the 2PCFs between the dots and the triangles of the same galaxy sample are consistent, and yet the assembly bias signals are still clearly present, albeit weakened. We further investigate how the clustering behaviors of red and blue mock samples change when different satellite removal criteria are imposed (see Appendix A.1).

3.4 Results

Using the methods described in §3.3, we now present our approach to search for assembly bias. As described in §3.2.2, we use the cluster catalog of [130, 131] to remove potential satellites. The parameter A describes how aggressively we mask regions near identified groups and clusters; below we show the result for $A = 1.5$, which correspond to excluding $\sim 40\%$ of the galaxies in the sample. Additionally, we exclude all galaxies with neighbors whose stellar mass is at least 3 times more massive (i.e. $M_{\star,\text{nbr}}/M_{\star,\text{lens}} \geq 3$) within $r_{\text{excl.}} = 500h_{70}^{-1}\text{kpc}$ projected separation, which ultimately exclude $\sim 90\%$ of the initial sample, as indicated by $f_{\text{gal,excl.}}$ in Figures 3.4 and 3.5. We present the CFHTLenS results with other satellite exclusion schemes in Appendix A.2.

For this sample of “centrals”, we adjust the $M_{\star}$ stellar mass widths and ranges for the red and blue galaxies separately, to obtain samples with consistent galaxy-galaxy lensing profiles. We judge consistency of the profiles by minimizing $\Delta\chi_S^2$ of the shear profiles (see Eqn. 3.11). Our shear measurements are at $r < 200h_{70}^{-1}\text{kpc}$, which is safely in the 1-halo regime for the masses of interest, i.e. $M \lesssim 10^{12}h_{70}^{-1}M_{\odot}$.

As noted above in §3.2.3, we exclude the brightest blue galaxies with $M_{r'} < -21.0$ as lenses. Because red galaxies have significantly larger halo masses than blue galaxies of similar stellar masses, also galaxy luminosity and stellar masses are strongly correlated, this exclusion of the brightest blue galaxies makes it difficult to find samples with matching lensing profiles. Table A.1 lists the stellar mass cuts of the galaxy samples stacked in Figures 3.4 and 3.5, along with the fraction of galaxies being excluded $f_{\text{gal,excl.}}$ and the $\Delta\chi_S^2$ of the sample pairs. The samples denoted in the Table and depicted in Fig. 3.4 do appear to have matching lensing profiles, however we are unable to construct samples at other stellar mass bins. If we use galaxies with significantly more massive stellar masses, most of the blue sample will overlap with the problematic magnitude range found above because of the strong correlation between galaxy luminosity and stellar masses, and if we use significantly least massive bins, we run out of red galaxies in the catalog.

Nevertheless, the lensing profiles do appear consistent for the samples shown in Fig. 3.4. After

ensuring that the stacked shear profiles of the two samples are statistically consistent, we proceed to measure the 2PCFs $w(\theta)$ for the samples. The left panel of Fig. 3.5 compare the 2PCFs of the red and blue samples with the same satellite exclusion criterion applied in Fig. 3.5. In order to quantify the significance of difference of the 2PCFs between the red and blue samples, i.e. $\Delta w(\theta) \equiv w^{\text{red}}(\theta) - w^{\text{blue}}(\theta)$, we follow the same calculation in Eqn. 3.14, but on $\Delta w(\theta)$ instead of the 2PCFs of the individual samples. We then calculate $\Delta\chi_w^2$, which is defined as,

$$\Delta\chi_w^2 = \sum_{m,n} \Delta w(\theta_m) \cdot (C_{m,n}^{\Delta w})^{-1} \cdot \Delta w(\theta_n), \quad (3.15)$$

where $(C_{m,n}^{\Delta w})^{-1}$ is the inverse of the covariance matrix of $\Delta w(\theta)$. For this calculation, we consider the radial bins from $3 - 11h_{70}^{-1}\text{Mpc}$, which is in the 2-halo regime as we are interested in the large-scale clustering behaviors of the galaxy samples. Given 4 degrees of freedom in the radial range and $\Delta\chi_w^2$, we compute the two-tailed p -value, the probability for the 2PCFs of the red and blue samples are drawn from the same distribution. The $\Delta\chi_w^2$, p -values and the detection significance for different exclusion criteria are listed in Table A.2. In the right panel of Fig. 3.5, $\Delta w(\theta)$ and the jackknife errors for different exclusion criteria for the 4 radial bins from $3 - 11h_{70}^{-1}\text{Mpc}$ are illustrated. The physical projected separation at $z = 0.3$ is shown at the top abscissae of both panels for comparison.

In Fig. 3.5, where 89.2% of the galaxies are being identified as potential satellites, we obtain $(\Delta\chi_w^2, p) = (8.05, 0.090)$, which translates to a 1.67σ difference between the 2PCFs of the red and blue samples between $3 - 11h_{70}^{-1}\text{Mpc}$. The significance of difference can also be demonstrated in the right panel, in which the jackknife errors are comparable to the deviation of the data points from the $\Delta w(\theta) = 0$ level. Particularly at the third radial bin at $r \approx 7h_{70}^{-1}\text{Mpc}$, $\Delta w(\theta)$ is consistent with zero. Therefore, our measurements in CFHTLenS galaxy samples do not favor a significant evidence of assembly bias detection. To test whether this is an overly restrictive satellite exclusion criterion, we loosen the rejection criteria by varying A , $r_{\text{excl.}}$ or $M_{\star, \text{nbr}}/M_{\star, \text{lens}}$ and repeat the same exercise. Nonetheless, the result does not change appreciably as indicated in Appendix A.2.

3.5 Discussion

Nominally, the detection of assembly bias remains elusive with our measurements with CFHTLenS galaxies. To ensure the rigorousness of our analysis, we have to exclude systematic effects that could mimic the signal we are attempting to observe. In previous work, there have been two principal bugaboos that have plagued studies of assembly bias: uncertainty in halo mass, and satellite contamination. We have attempted to mitigate both of these potential systematics. First, our lensing measurements give us some confidence that the halo mass distributions of the red and blue galaxy samples are consistent. The halo masses of our galaxy samples are of order $M \lesssim 10^{12} h_{70}^{-1} M_{\odot}$, for redshifts in the range $0.2 \leq z \leq 0.4$, and halo bias is a relatively weak function of halo mass in this regime. From Fig. 9 of Bhattacharya et al. [153], the large scale halo bias $b(M)$ changes only $\sim 20\%$ when the central halo mass is changed by a factor of $\Delta M/M \sim 4$, for halos in this mass regime. This justifies our choice of selecting galaxy samples with wide stellar mass ranges as we ensure the scatter of halo bias in our galaxy samples is small.

In Fig. 3.4, the blue sample has a slightly lower amplitude of the lensing profile than the red sample even though they are statistically consistent (i.e. $\Delta\chi_S^2 = 3.19$ over 7 radial bins from $20 - 200 h_{70}^{-1} \text{kpc}$), and we find a 1.67σ difference in their 2PCFs. To understand whether this subtle mass difference would contribute to the difference in 2PCFs, we select the new red (blue) samples which have slightly higher (lower) lensing profile amplitudes compared to the initial red (blue) samples used in Figures 3.4 and 3.5 with the $\Delta\chi_S^2$ between the new and initial samples ~ 3 , and we find the difference of 2PCFs between the two samples is about 1σ .

To address the systematic of residual contamination from satellites, we have used quite aggressive cuts to remove potential satellites, excluding $\sim 90\%$ of the initial lens sample. The resulting shear profiles shown in Fig. 3.4 do not exhibit signs of satellite contamination on $\sim \text{Mpc}$ scales, and the 2PCFs shown in Fig. 3.5 resemble the results from our mocks where satellites have been excluded. In case if any cluster satellites that are not removed by the 3D-MF algorithm due to incompleteness at mass ranges $10^{13.5} \leq M/h_{70}^{-1} M_{\odot} \leq 10^{14}$, many of which should be removed

by our $r_{\text{excl.}} = 500h_{70}^{-1}\text{kpc}$ exclusion of galaxies with more massive neighbors. Such an exclusion radius is chosen as it coincides with the virial radius of clusters at this halo mass range. Therefore, we ensure our analysis does address the systematics caused by the two major obstacles in the previous attempts on assembly bias detection: scatter in halo bias due to uncertainty in halo mass, and contamination from residual satellites.

Our results appears to be consistent with the recent results of [121]. We have adopted nearly the same methodology used in that work. There are yet some differences between our analysis and theirs. While [121] analyze a spectroscopic sample from SDSS, we have analyzed a photometric sample from CFHTLenS. We are relying on the validity of photometric redshifts, which could potentially have systematic errors compared to spectroscopic redshifts. In light of this, we check the cross-correlation coefficients of the galaxies in our red and blue samples with all galaxies in the prescribed photo- z range and abandon the brightest blue galaxies with possible systematic errors in their photo- z 's. We caution that the redshift errors of the remaining samples could still bias our results.

Similarly, the removal of satellites in our work may not be as complete as that in [121], since groups and clusters (and their members) may be identified far more reliably using spectroscopic data. To account for the limitation of identifying satellites using only photo- z information, we adopt quite conservative satellite cuts to ensure a high purity of central galaxy samples, and therefore the galaxy sample size is compromised. The non-detection of assembly bias in our measurements could be attributed to the small sample size. Another possibility is that the galaxy color might not correlate well with halo age, potentially diluting the assembly bias signal in our sample. With numerous endeavors on attempting to detect assembly bias for galaxy-sized halos, it is proven to be incredibly difficult to confirm observationally despite the robust theoretical prediction of its existence. Further work will be required in order to resolve this potential discrepancy between theory and observation.

Chapter 4

Some assembly required: assembly bias in massive dark matter halos

4.1 Motivation

We study halo assembly bias for cluster-sized halos. Previous work has found little evidence for correlations between large-scale bias and halo mass assembly history for simulated cluster-sized halos, in contrast to the significant correlation found between bias and concentration for halos of this mass. This difference in behavior is surprising, given that both concentration and assembly history are closely related to the same properties of the linear-density peaks that collapse to form halos. Using publicly available simulations, we show that significant assembly bias is indeed found in the most massive halos with $M \sim 10^{15} M_{\odot}$, using essentially any definition of halo age. For lower halo masses $M \sim 10^{14} M_{\odot}$, no correlation is found between bias and the commonly used age indicator $a_{0.5}$, the half-mass time. We show that this is a mere accident, and that significant assembly bias exists for other definitions of halo age, including those based on the time when the halo progenitor acquires some fraction f of the ultimate mass at $z = 0$. For halos with $M_{\text{vir}} \sim 10^{14} M_{\odot}$, the sense of assembly bias changes sign at $f = 0.5$. We explore the origin of this behavior, and argue that it arises because standard definitions of halo mass in halo finders do not correspond to the collapsed, virialized mass that appears in the spherical collapse model used to predict large-scale clustering. Because bias depends strongly on halo mass, these errors in mass definition can masquerade as or even obscure the assembly bias that is physically present. More physically motivated halo definitions using splashback should be free of this particular defect of

standard halo finders ¹.

4.2 Introduction

The clustering of tracers of cosmological large-scale structure, such as galaxies, quasars, clusters, or voids, may be used to probe the clustering of the underlying matter field. The clustering strength of any particular tracer does not exactly match the clustering of total matter, but instead is generally biased relative to matter clustering [154]. On large scales, in the linear regime of structure formation for standard cosmologies with cold dark matter and gravity described by Einstein’s general relativity, the bias for any tracer tends towards a constant value that becomes independent of scale [e.g. 34, 60]. For dark matter halos, the linear bias is a strong function of halo mass, with the most massive halos clustering far more strongly than typical dark matter particles, while the smallest halos cluster less strongly than typical particles [60, 61, 154]. Qualitatively, one may think of highly biased halos ($b \gg 1$) as preferentially forming in regions of high density, while halos with low bias (e.g., $b < 1$) tend to avoid high-density regions.

In addition to its mass dependence, halo bias can also depend on other halo properties such as mass assembly history [67] or properties like concentration, spin, etc. [66]. Although not as strong as the mass dependence, these secondary dependencies of halo bias can be quite significant, in some cases leading to variations in linear bias of more than a factor of 2 for halos of fixed mass. Because secondary bias can be quite significant, a number of studies have explored the impact of such biases on the galaxy-halo connection; see Ref. [30] for a recent review of this topic and for a more comprehensive review of work on secondary biases. The most well-studied of these secondary biases have been assembly bias, the dependence of bias on mass assembly history (MAH), and the concentration bias, referring to the dependence on halo concentration. In general, secondary biases exhibit significant mass dependence. For example, the concentration

¹This chapter was previously submitted on arXiv.org as C. Y. R. Chue, N. Dalal and M. White, Some assembly required: assembly bias in massive dark matter halos, arXiv:[1804.04055v1]. This chapter matches the submitted version.

bias actually reverses in sign as halo mass is varied, with high concentration associated with high bias for small halos but with low bias for the largest halos [65].

Much of this behavior in halo bias is not difficult to understand in the context of hierarchical structure formation [68]. Because halos tend to arise from peaks of the linear density field [60, 155], the properties of halos are related to the properties of the corresponding initial peaks. For example, peaks with steep slopes tend to produce halos with high concentration, while peaks with shallow slopes tend to lead to halos with low concentration [e.g. 47]. Additionally, because the linear density field is continuous, the slopes of initial peaks are also correlated with their local environments. At fixed peak height, peaks with steep slopes tend to be found in relatively lower density environments than peaks with shallow slopes. This accounts for the concentration bias seen at high halo masses [68], however this does not explain the opposite behavior seen at low halo mass. At lower masses, another process starts to dominate over the effect of peak slopes in producing concentration bias (and assembly bias). Among low-mass halos below the nonlinear mass scale ($M \ll M_*$), a significant fraction of order 20% ceases to grow in mass, due to environmental effects. Because halo concentration is related to assembly history [156], the halos that stop growing exhibit the highest concentrations. At the same time, the environmental effects that shut down halo growth (e.g., strong tides or high velocity dispersion) are also associated with high density regions. For this reason, at low masses high concentrations become correlated with high local density, i.e. high bias. This effect is unimportant at the very highest masses because the biggest halos dominate their environments.

A corollary of the argument explaining concentration bias is that very similar behavior should be found in assembly bias. At high masses, the same peak properties that determine halo concentration also determine halo assembly histories, and at low masses, the environmental effects that lead to high concentration also arrest the growth of halo mass. The expected assembly bias is indeed found in low-mass halos [67], but at higher masses, the evidence is far less clear. [66] found no significant assembly bias at high mass in their simulations, and [157] argued that cluster-sized halos exhibit no detectable assembly bias in Λ CDM simulations. If correct, this result would be

remarkable and would require a dramatic rethinking of halo formation in general. The prediction of assembly bias follows from the continuity of the linear density field, given the known result that the formation of the most massive halos closely follows the prediction of the spherical collapse model [158] that formation occurs when the smoothed linear density reaches a critical value, $\bar{\delta} = \delta_c \approx 1.686$ [68, 159]. Since the linear density field is indeed continuous, the prediction of nonzero assembly bias at high mass would seem to be inescapable. Motivated by this surprising claim, we investigate halo assembly bias for massive cluster-sized halos in Λ CDM simulations.

4.3 Simulation data

Since we focus on only the most massive halos which tend to be rare, we utilize simulations with large volume. Most of the results we present below are derived from the BigMDPL simulation [160], publicly available at <https://www.cosmosim.org> [143]. This simulation is part of the MultiDark simulation suite, and contains 3840^3 particles in a box of comoving side length of $2.5 \ h^{-1} \text{ Gpc}$ for a flat Λ CDM cosmology with $\Omega_m \approx 0.307$, $h = 0.6777$, $\sigma_8 = 0.8228$ and $n_s = 0.96$, corresponding to particle mass $m_p = 2.36 \times 10^{10} h^{-1} M_\odot$. We use the Rockstar [141] halo catalogs and merger trees publicly provided at <https://www.cosmosim.org>. To derive mass accretion histories, we follow the main branch of the Rockstar merger tree, using the mmp (most massive progenitor) flag. As a sanity check, we have also examined other simulations, including the MDPL2 simulation from the same MultiDark suite, as well as a series of $L = 640 \ h^{-1} \text{ Mpc}$ simulations run for this investigation. As a check on the Rockstar results, we have computed halo catalogs and merger trees using a different method for the 640 Mpc boxes, as described in Ref. [161]. In all cases, we find results consistent with the BigMDPL simulation results, so the discussion below will focus on that simulation since it provides the best statistics due to its large volume.

4.4 Analysis work and results

For the BigMDPL simulation, we measure the linear bias for halo samples by first computing the halo-matter cross spectrum $P_c(k)$ and the matter auto-spectrum $P_m(k)$, and then defining the bias b by a least-squares fit for $P_c(k) = b P_m(k)$ for $k < 0.1 h \text{ Mpc}^{-1}$. Because the matter field is not made publicly available for this simulation, as a proxy for the matter field we use the set of all halos and subhalos with $M_{\text{peak}} \geq 5 \times 10^{11} h^{-1} M_{\odot}$ in the $z = 0$ Rockstar catalog. These halos should be nearly unbiased on large scales, but it is worth noting that formally all of our quoted bias values really correspond to the ratio b/b_{tracer} where b_{tracer} is the mean bias of our tracer population of subhalos.

To start, we first examine halos with $M_{\text{vir}} = 0.7 - 1 \times 10^{15} h^{-1} M_{\odot}$. Previous work has shown significant concentration bias for halos in this mass range, and the BigMDPL simulation gives consistent results. Rank ordering the halos based on the concentration values reported in the Rockstar catalogs, we measure mean linear biases for the subsets with the highest 25% and lowest 25% of c_{vir} . The quartile with highest c_{vir} gives $b_{c\text{-high}} = 4.4 \pm 0.08$, while the quartile with lowest concentration gives $b_{c\text{-low}} = 5.2 \pm 0.08$, as expected for the concentration bias at these high halo masses.

Next, we turn to assembly bias. Similar to the concentration split, we can split halos into the oldest and youngest quartiles, using some definition of halo age. In previous literature [66, 157], the half-mass time $a_{0.5}$ has been the most common definition of age. This is defined as the scale factor when a given halo's most massive progenitor first acquires a fraction $f = 0.5$ of the final mass at $z = 0$. From the Rockstar merger trees, we can readily determine a_f for any fraction including $f = 0.5$. Halos with a small a_f assembled fraction f of their mass relatively earlier, and therefore may be considered to be older, while conversely halos with larger a_f may be considered to be younger. If we split halos in this mass range ($M_{\text{vir}} = 0.7 - 1 \times 10^{15} h^{-1} M_{\odot}$) then the top and bottom quartiles give $b_{a\text{-high}} = 4.9 \pm 0.07$ and $b_{a\text{-low}} = 4.6 \pm 0.07$. Therefore we do find significant assembly bias in high mass halos, with the expected sign, but the amplitude is about

half as strong as the concentration bias for the same halos. We find similar results for even higher masses or from other simulations, albeit with larger uncertainties. It is reassuring that this basic prediction of Gaussian statistics is confirmed, but the weaker amplitude relative to concentration bias is somewhat surprising. One possibility is that $a_{0.5}$ may simply be noisier than concentration. This quantity is derived by tracking M_{vir} along the merger tree, but M_{vir} itself is a noisy estimate of the true virialized mass in a halo for a variety of reasons, including the presence of substructure, or the fact that the nominal virial radius r_{vir} can be either larger or smaller than the actual virialized region around a halo, the splashback radius [1, 162].

If the assembly bias seen using $a_{0.5}$ is weak simply due to noise in the MAH, then we could improve the significance by using the entire MAH to classify halos into ‘young’ or ‘old’. As is well known, halo mass accretion histories exhibit a variety of behaviors [e.g. 156], so there is little reason to expect an arbitrarily chosen number like $a_{0.5}$ to capture the aspects of halo assembly that relate to large-scale environment. However, since the entire MAH has many degrees of freedom, it may not be immediately obvious what definition of age that we should use instead of $a_{0.5}$.

The approach that we use is to perform a linear operation on the MAH to assign a single number to each halo, and then rank order based on that number. To choose what linear operation to perform on the MAH, note that we can predict how the MAH should change when we raise or lower the large-scale linear density, using Gaussian statistics and the spherical collapse model. The starting point is again the spherical collapse result that collapse occurs when the linear density smoothed over radius R exceeds the collapse threshold, $\bar{\delta}(R) \geq \delta_c$. The model predicts that the set of halos of mass M therefore should have $\bar{\delta}(R_L) = \delta_c$, where $R_L = (3M/4\pi\bar{\rho}_m)^{1/3}$ is the Lagrangian radius corresponding to mass M in the notation of [47]. The linear density profile interior to R_L determines the assembly history of that halo [68]. Therefore, to predict how the assembly history changes when we vary the large-scale environment, we simply need to know the expected value of $\bar{\delta}(R)$ at $R < R_L$ as a function of the large-scale environmental density δ_{long} . This is readily determined from the Gaussian statistics of the linear density field. In general, for Gaussian distributed quantities $\mathbf{X}$ and $\mathbf{Y}$ with zero mean, the expected value of $\mathbf{X}$ conditioning

on the value of $\mathbf{Y}$ is given by

$$\langle \mathbf{X} | \mathbf{Y} \rangle = \langle \mathbf{X} \mathbf{Y} \rangle \langle \mathbf{Y} \mathbf{Y} \rangle^{-1} \mathbf{Y}. \quad (4.1)$$

In our case, $\mathbf{X}$ consists of the interior profile $\bar{\delta}(R)$ for $R < R_L$, and $\mathbf{Y}$ consists of the pair of quantities $\bar{\delta}(R_L) = \delta_c$ and δ_{long} on some large scale. For concreteness, we define δ_{long} as the linear density smoothed with a top hat filter of radius $30 h^{-1} \text{ Mpc}$.

Eq. (4.1) gives us the expected profile for a peak of size R_L in a background overdensity δ_{long} , and if we know the linear growth factor $D(a)$ as a function of a , we can translate that peak profile into a mass assembly history by setting the collapse radius at each time a such that $\bar{\delta}(R)D(a) = \delta_c$. Since Eq. (4.1) is linear in δ_{long} , then for small δ_{long} the response of the halo MAH is also linear in δ_{long} . If we think of the MAH as a vector $\mathbf{h}$, then its expected linear response to δ_{long} may be written as $\mathbf{h} = \text{const} + \mathbf{g} \delta_{\text{long}}$, where the vector $\mathbf{g}$ encodes the linear response computed above. This immediately suggests a sensible choice for the linear operation to perform on the actual MAH to assign an age to each halo: the inner product between $\mathbf{h}$ and the expected response vector $\mathbf{g}$. To define an inner product on the space of possible assembly histories, however, we need some notion of a metric on that space, i.e. a matrix to allow us to compute distances and dot products between vectors. One obvious choice for this metric is the inverse covariance matrix of all MAH's for halos in the mass bin being considered, $\mathbf{C}_h^{-1} = (\langle \mathbf{h} \mathbf{h} \rangle - \langle \mathbf{h} \rangle \langle \mathbf{h} \rangle)^{-1}$.

Our procedure, therefore, is to define the ‘age’ of each halo from its MAH $\mathbf{h}$ as

$$\alpha_g = \mathbf{g}^T \cdot \mathbf{C}_h^{-1} \cdot \mathbf{h}, \quad (4.2)$$

where $\mathbf{g}$ is computed from Gaussian statistics as described above, and $\mathbf{C}_h^{-1}$ is computed from the ensemble of MAH's of the halo mass bin under consideration. Defined in this way, halos with high α_g are expected to be more highly biased than halos with low α_g , as long as halos are forming according to spherical collapse. When we apply this age definition to halos in the same

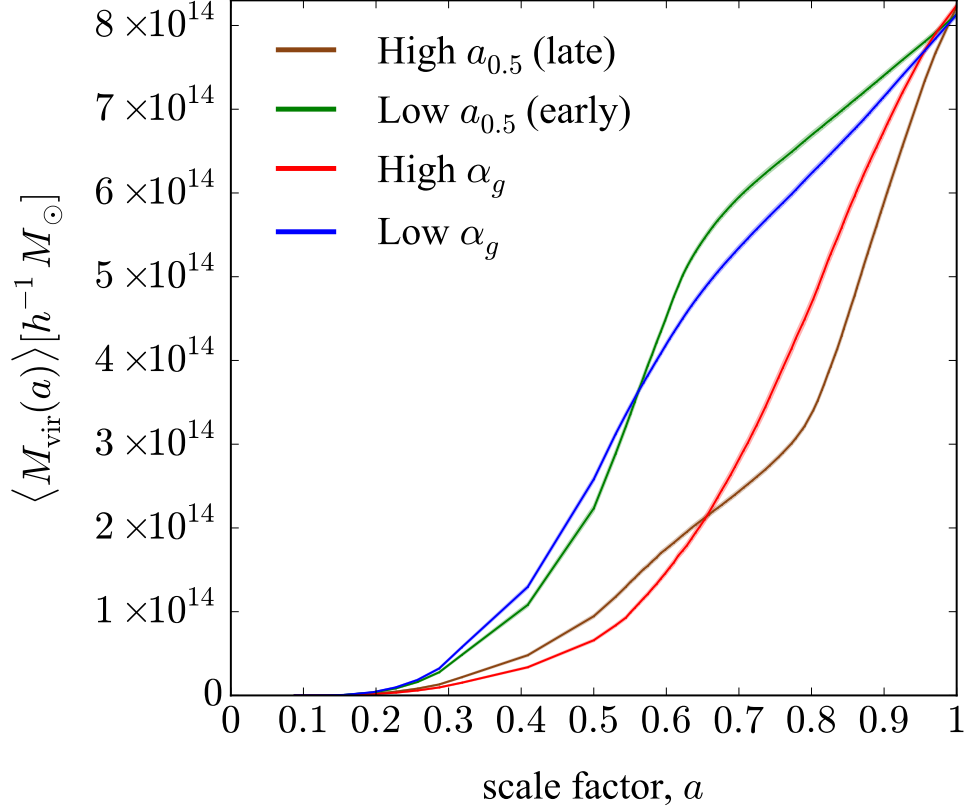


Figure 4.1: The different curves show stacked mass accretion histories for subsets of BigMDPL halos with $M_{\text{vir}} = 0.7 - 1 \times 10^{15} h^{-1} M_{\odot}$. The curves corresponds to the halos with the highest 25% of $a_{0.5}$ (brown), the lowest 25% of $a_{0.5}$ (green), the top 25% of α_g (red), and bottom 25% of α_g (blue). The width of each curve corresponds to the $1 - \sigma$ jackknife uncertainty on the mean MAH. As discussed in the text, α_g is a better indicator of large-scale bias than $a_{0.5}$, and it tends to split the halos more strongly on their early assembly histories ($f < 0.5$) for this mass range.

mass range ($M_{\text{vir}} = 0.7 - 1 \times 10^{15} h^{-1} M_{\odot}$) considered above, the bias of the high α_g quartile is $b_{\alpha\text{-high}} = 5.0 \pm 0.07$, while the low α_g quartile gives $b_{\alpha\text{-low}} = 4.5 \pm 0.07$. Evidently, using the entire MAH does enhance the amplitude of assembly bias, though the overall signal is still slightly smaller than the amplitude of the concentration bias. In Fig. 4.1 we plot the stacked MAH's for the top and bottom quartiles of α_g , along with stacked MAH's for the top and bottom quartiles of $a_{0.5}$. The MAH's selected by α_g differ more at early times than the MAH's selected by $a_{0.5}$.

Therefore, the highest mass halos do exhibit clear assembly bias, as required theoretically. This may seem to contradict previous results [66, 157], but note that so far we have focused on

halos with $M \sim 10^{15} M_\odot$, whereas previous works studied smaller clusters with $M \sim 10^{14} M_\odot$. Therefore, we next consider halos with $M_{\text{vir}} = 1 - 2 \times 10^{14} h^{-1} M_\odot$. When we split these halos using $a_{0.5}$, we do not find significant differences in the biases of the oldest or youngest halos. This agrees with previous work, but is contrary to the results for the higher mass sample.

To understand this change in behavior, in Fig. 4.2 we plot the cross-correlation coefficient between the large-scale density and the mass accretion history. As above, the large-scale density is defined as $\delta_{\text{long}} = \delta_{30}$, the overdensity smoothed over a scale of $30 h^{-1}$ Mpc. We characterize the MAH using a_f , the scale factor when a halo reaches fraction f of its $z = 0$ mass. The cross-correlation coefficient is defined as

$$\text{corr}(a_f, \delta_{\text{long}}) = \frac{\langle (a_f - \bar{a}_f)(\delta_{\text{long}} - \bar{\delta}_{\text{long}}) \rangle}{[\langle (a_f - \bar{a}_f)^2 \rangle \langle (\delta_{\text{long}} - \bar{\delta}_{\text{long}})^2 \rangle]^{1/2}}, \quad (4.3)$$

where $\bar{a}_f = \langle a_f \rangle$, $\bar{\delta}_{\text{long}} = \langle \delta_{\text{long}} \rangle$, and averages are computed over the sample of halos being considered. A positive cross-correlation means that increasing the large-scale density increases a_f , i.e. delays the time when the halo acquires mass fraction f .

As Fig. 4.2 shows, there is no significant correlation between large-scale density and $a_{0.5}$ for $M \sim 10^{14} M_\odot$ halos, but this appears to be an accident. If we use some other fraction, like $a_{0.2}$ or $a_{0.8}$, then we do find significant correlations. The early part of the MAH behaves similarly to the behavior for the $M \sim 10^{15} M_\odot$ halos: younger halos are associated with higher density. But the later part of the MAH, for $f \lesssim 1$, has the opposite correlation. The cross-over happens to occur near $f = 0.5$, by accident. Note that this is dependent on the mass of the sample. For even lower masses, the cross-over occurs at even smaller f , and for higher masses it occurs at higher f (or does not occur at all in the most massive halos, as shown in the blue curve in Fig. 4.2). Note that the significant correlations between large-scale density and MAH that we find do not necessarily contradict the results of [157], who found that the stacked MAH's for halos with $M \sim 10^{14} M_\odot$ found in large-scale over-densities were very similar to those found in large-scale under-densities. When we perform the same exercise, we also find similar MAH's with percent-level differences.

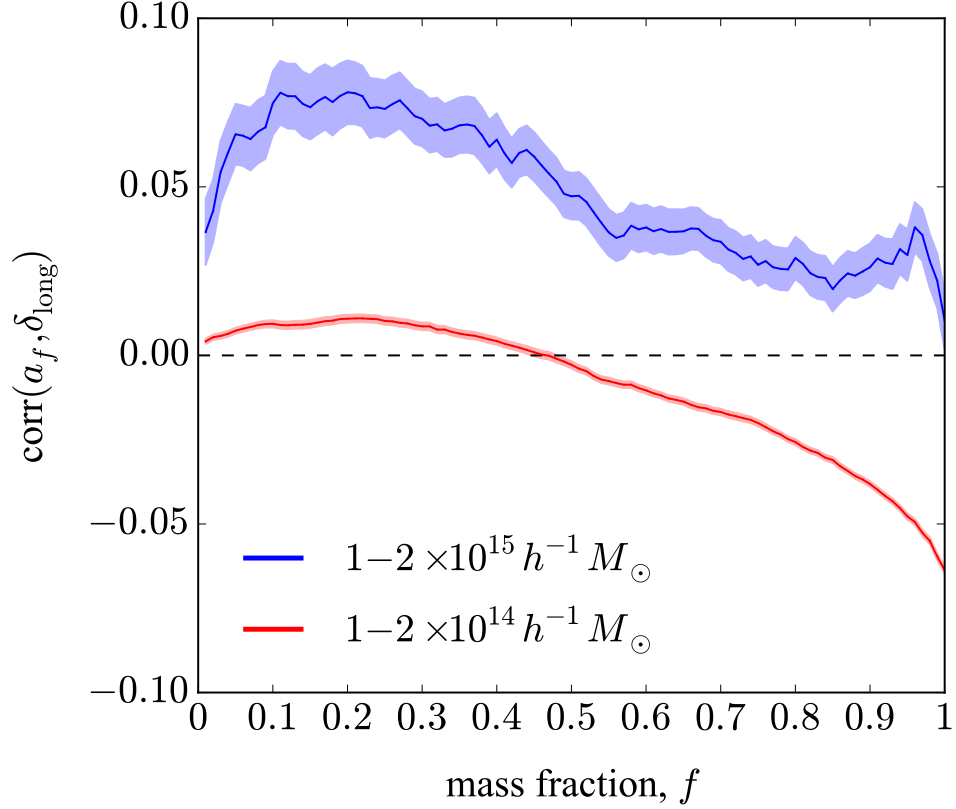


Figure 4.2: Plotted is the cross-correlation coefficient between the linear overdensity, smoothed on a scale of $30 h^{-1}$ Mpc, and the time when the halo acquires fraction f of its present-day mass, denoted a_f . The blue curve corresponds to high-mass halos in the mass range $M_{\text{vir}} = 1 - 2 \times 10^{15} h^{-1} M_{\odot}$, while the red curve is for low-mass halos with $M_{\text{vir}} = 1 - 2 \times 10^{14} h^{-1} M_{\odot}$. The width of each band corresponds to the $1 - \sigma$ uncertainty, determined by jackknife.

However, that is exactly the amplitude of difference that is expected. The amplitude of density fluctuations on large scales in the linear regime is small by definition, percent-level for the scales of interest here. Because the expected level of assembly bias is of order unity, not order 100, these percent-level overdensities on large scales should correspond to percent level variations in the MAH's, as observed.

The significant correlations at $f \neq 0.5$ imply that assembly bias is present in halos of this mass range, i.e. there are correlations between large-scale density and assembly history. Accidentally, $a_{0.5}$ is insensitive to this assembly bias, however we can use other metrics for halo age to find significant assembly bias. For example, we can once again use the ‘theoretical’ template α_g to select old or young halos, which does indeed give nonzero assembly bias. Alternatively, we can derive the optimal definition of halo age to maximize the difference in bias between old and young subsets. We do so by cross-correlating halo MAH's with their large-scale density. As before, we quantify the large-scale density as $\delta_{\text{long}} = \delta_{30}$, the overdensity centered on a halo smoothed over a $30 h^{-1}$ Mpc radius. Similarly, again let us write $\mathbf{h}$ as the MAH for a halo. If δ and $\mathbf{h}$ are Gaussian distributed then the optimal definition of age for a halo with a MAH $\mathbf{h}$ is given by $\alpha_{\text{opt}} = \mathbf{d}^T \cdot (\mathbf{h} - \bar{\mathbf{h}})$, where

$$\mathbf{d} = \mathbf{C}_h^{-1} \langle (\mathbf{h} - \bar{\mathbf{h}})(\delta_{\text{long}} - \bar{\delta}_{\text{long}}) \rangle. \quad (4.4)$$

In other words, we take the inner product of each MAH with the part of the MAH correlated with large-scale density, where the inner product over the space of MAH's is defined using the inverse covariance of MAH's as the metric. Note that to avoid over-fitting, when evaluating Eq. (4.4) for each cluster, we exclude all halos in the spatial octant centered on that cluster in computing the ensemble averages. In labeling this definition optimal, what we mean is that this definition should maximize the difference in large-scale bias of the two samples, using only the mass accretion histories, as long as the underlying assumption of Gaussianity is approximately satisfied. Non-gaussianity will make this definition sub-optimal for the purpose of splitting halos into high-bias

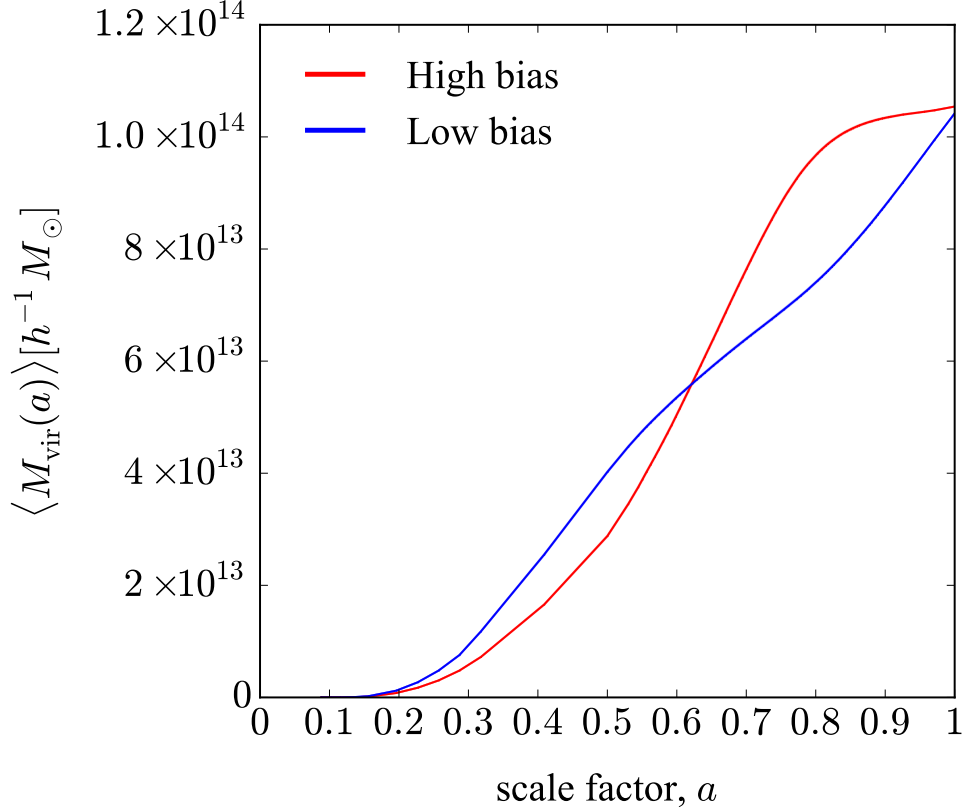


Figure 4.3: BigMDPL halos with $M_{\text{vir}} = 1 - 1.1 \times 10^{14} h^{-1} M_{\odot}$. Plotted are the stacked (average) MAHs for isolated halos in the top (red) and bottom (blue) quartiles of α_{opt} as defined in Eq. (4.4). The width of each curve corresponds to the $1 - \sigma$ jackknife uncertainty on the mean MAH. We have used a narrow mass bin in order to enforce that both subsets have the same average mass at $z = 0$.

and low-bias subsets, but as long as we do detect assembly bias any suboptimality does not impact our conclusions significantly.

Fig. 4.3 shows the average MAH’s for the halos in this mass range, split into top and bottom quartiles using α_{opt} . The quartile with high α_{opt} (red curve) has a mean linear bias $b = 2.2 \pm 0.02$, while the quartile with low α_{opt} (blue curve) has a mean linear bias $b = 2.0 \pm 0.02$. As expected, there is significant assembly bias among halos in this mass range, in that we can split halos into samples with higher or lower bias using only their MAH’s. It is difficult to say which subset is older or younger: at low mass fractions, the blue subset is significantly older, while at high mass fractions, the red subset is significantly older.

One striking property of the red curve in Fig. 4.3 is that the mean MAH nearly plateaus at late times, $a > 0.85$. This lack of growth in halo mass is quite surprising for cluster-sized halos. Even if the physical mass distribution around the cluster remains static in time, the nominal virial mass will grow simply due to the decrease in the mean matter density as the universe expands, an effect called pseudo-evolution [163]. For a static mass profile $M(r)$ around a halo, pseudo-evolution gives a minimum growth rate of $d \log M_{\text{vir}} / d \log a = (d \log \rho_{\text{vir}} / d \log a) \times [1 + 3 / (d \log \bar{\rho} / d \log r|_{r=r_{\text{vir}}})]$, where $\bar{\rho}(r) = 3M(r)/(4\pi r^3)$, and $\rho_{\text{vir}} = \Delta_{\text{vir}} \rho_m$ for virial overdensity Δ_{vir} [41] and mean matter density $\rho_m = \Omega_m \rho_{\text{crit}}$. Since these clusters tend to have somewhat low concentrations, e.g. $c_{\text{vir}} \sim 6$, then for NFW outer profiles we would expect $d \log M_{\text{vir}} / d \log a > 0.5$ even if the density profiles around the halos remain static in time. Of course, the outer profiles of these halos can be steeper than NFW, due to the splashback feature [1], but that steepening would only affect the pseudo-evolution rate of M_{vir} when the splashback radius is $r_{\text{sp}} \lesssim r_{\text{vir}}$, which only occurs for high accretion rates [162]. For the low growth rates shown in the red curve ($d \log M_{\text{vir}} / d \log a \approx 0.18$), the splashback radius should be outside r_{vir} , implying that the NFW profile should be a reasonable approximation. We will return to this topic later, but for now, the point is that the observed growth rate in this subset of clusters is even less than the minimal pseudo-evolution rate for static mass distributions. In order for the average M_{vir} to grow so slowly with time, mass must be physically removed from within r_{vir} for at least some fraction of the clusters in the red subset.

One possible explanation for this could be that many of the halos in the red subset are in extreme environments capable of stripping mass from these cluster-sized halos. To check for this, we search for more massive neighbors ($M_{\text{vir}} \geq 2 \times 10^{14} h^{-1} M_{\odot}$) within a few Mpc of these clusters. We find that only a tiny, percent-level fraction of the halos (excluding subhalos) have massive nearby neighbors capable of tidally stripping the clusters.

If tides are unimportant, then some other explanation is required to account for the slow growth in M_{vir} . To clarify the origin of this behavior, we plot in Fig. 4.4 the average (stacked) phase space density for the two subsets of high and low α_{opt} . Using the catalog of all Rockstar halos and subhalos with $M_{\text{peak}} > 5 \times 10^{11} h^{-1} M_{\odot}$, we compute the mass in neighboring objects as a

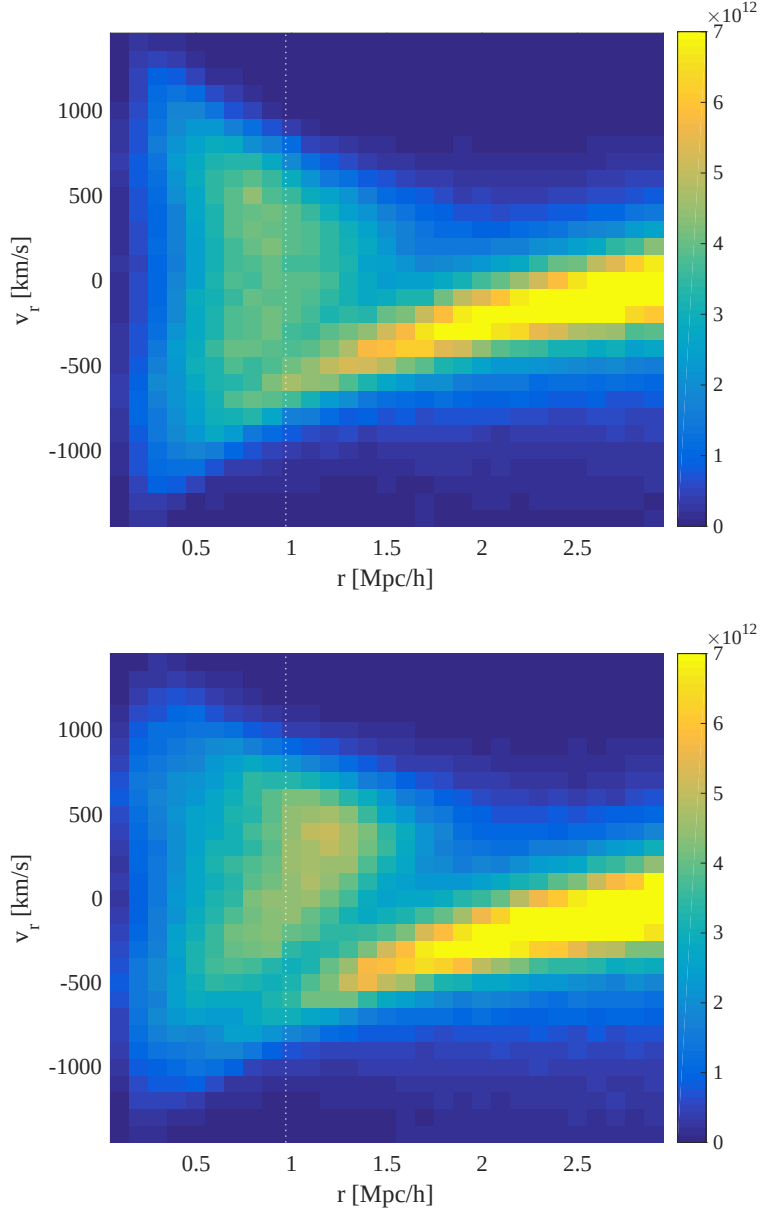


Figure 4.4: Phase space diagrams around BigMDPL halos with $M_{\text{vir}} = 1 - 1.2 \times 10^{14} h^{-1} M_{\odot}$. The color corresponds to the mass in neighbors (in units of $h^{-1} M_{\odot}$) at each pixel in the space of r and v_r . We have used a somewhat wider mass bin than in Fig. 4.3 in order to improve the statistical uncertainties. The top panel is for the low α_{opt} , corresponding to the blue curve in Fig. 4.3, while the bottom panel is for high α_{opt} , corresponding to the red curve in Fig. 4.3. For this mass range, r_{vir} is depicted by the vertical dotted white line at slightly less than $1 h^{-1} \text{ Mpc}$. Comparing the two panels, we can see that the sample in the lower panel has a significant amount of bound, virialized mass near splashback located just outside r_{vir} .

function of distance and radial velocity relative to each cluster. For the clusters being considered, with $M_{\text{vir}} = 1 - 1.2 \times 10^{14} h^{-1} M_{\odot}$ at $z = 0$, the virial radius is approximately $r_{\text{vir}} \approx 0.97 h^{-1}$ Mpc. Fig. 4.4 immediately explains why the high α_{opt} subset has stopped growing in M_{vir} since $a \sim 0.85$: that subset of clusters has a large portion of splashback mass beyond the nominal r_{vir} . Much of that mass just outside r_{vir} was previously inside the virial radius one crossing time in the past, which corresponds to $a \sim 0.85$. Therefore, mass has indeed been removed from within r_{vir} for these clusters, but not because of tidal stripping, but instead merely because this recently accreted mass is on wide orbits that extend beyond r_{vir} . Although we do not have access to the particle data for this simulation, we can estimate the amount of this extra mass using the population of neighboring halos and subhalos as a proxy for dark matter mass. Very roughly it appears that the splashback mass for the high α_{opt} sample is larger than M_{vir} by about 60%.

Therefore, the physical explanation for the slow mass growth in the red curve of Fig. 4.3 may actually be quite mundane. Simply put, these clusters have been assigned the wrong masses. Their actual physical masses are larger than the quoted virial masses, and therefore it is no surprise that they are more highly biased. The problem is that the virial mass definition used in most halo finders does not actually measure the bound, virialized mass around a halo (i.e., the mass within the splashback radius), but instead measures the mass within an arbitrarily chosen density threshold. Relatedly, the quoted masses in the catalog account only for material within a spherical surface, whereas the actual splashback surfaces around simulated halos can deviate significantly from spherical shapes [164, 165]. The problem we described above may not be specific to the Δ_{vir} definition of halo mass, but instead could arise for other similarly arbitrary definitions like $200c$ or $200m$. Indeed, if we repeat the same calculation for halos selected in bins of M_{200m} , we again find that the high- α_{opt} sample with high bias has a significant amount of mass located just outside r_{200m} . Adopting even lower density thresholds to produce even larger halo radii could suffer the opposite problem of overestimating halo masses, due to uncollapsed mass prematurely being included in the halo, leading to halos with Lagrangian densities well below the spherical collapse threshold. This would similarly generate spurious assembly bias. A more physically correct halo

mass definition using the splashback feature should avoid such problems and thereby mitigate this spurious behavior in assembly bias. Fortunately, implementations of splashback halo masses for simulations now exist [164, 165], so it should be possible to avoid this problem in future analyses.

This issue with mass definitions may also explain why the assembly bias signal found using mass accretion histories was somewhat weaker than the signal found using halo concentrations, even though mass assembly history and density profile are both related to the same properties of the initial peaks that collapse to form halos. The splashback radius can be larger or smaller than the arbitrarily chosen overdensity radii like r_{vir} or r_{200} used in halo finders, depending on the physical accretion of mass onto halos, meaning that at all times there are errors in the derived halo boundary and halo mass. In principle, these errors could possibly generate enough noise in the derived MAH's to erase some of the assembly bias signal that is physically present.

One question that may arise is why the effect of halo mass definitions does not also corrupt the assembly bias signal for higher masses (e.g. $M \sim 10^{15} M_{\odot}$) the way that it does for lower mass clusters. It is certainly possible to find clusters in this higher mass range whose apparent MAH's exhibit the plateau shown in Fig. 4.3, but their proportion appears to be far smaller among $10^{15} M_{\odot}$ clusters than it is among $10^{14} M_{\odot}$ clusters. We have not explored this question in detail, but a plausible explanation may simply be that clusters with such high mass are much more rare, corresponding to $\sim 3\sigma$ fluctuations of the linear density, rather than $\sim 2\sigma$ fluctuations. Any cluster with $M_{\text{vir}} \approx 10^{15} M_{\odot}$ that has a significant amount of mass outside r_{vir} would therefore be an even more massive cluster and would correspond to an even rarer fluctuation. The fraction of such objects therefore should be smaller at $M \sim 10^{15} M_{\odot}$ than at $10^{14} M_{\odot}$, simply because the mass function is so much steeper at the higher mass.

Chapter 5

Conclusions and Outlook

The spectacular successes of the Λ CDM model in describing numerous observed large-scale properties of the universe rely on two major key components, namely dark matter and dark energy. In particular, dark matter is believed to be a dominant form of mass in the Universe [2], and yet the nature of dark matter remains mysterious. Therefore, understanding dark matter remains one of the most important outstanding problems in modern fundamental physics.

Astrophysical observations can put constraints on multiple physical properties of dark matter. Nevertheless, Unlike ordinary matter, dark matter does not emit or absorb electromagnetic radiation at any significant level [69]. Instead, its existence can be inferred by its gravitational effect on ordinary matter or electromagnetic radiation. In this thesis, we first discuss the development of a new spatial 3-point lens-shear-shear correlation technique to probe dark matter halo shapes in chapter 2, which can be applied to observational data to put constraints on dark matter self-interaction cross-section. Next, we discuss the attempt to detect assembly bias, the age dependence of halo clustering, for galaxy-sized halos using the CFHTLenS observational data in chapter 3. Lastly, inspired by the claims from several papers [66, 157] that there is no significant halo assembly bias evidence at high mass ($\sim 10^{14} M_{\odot}$) in cosmological simulations, we investigate this topic using the BigMDPL simulation data [160] in chapter 4.

5.1 Probing dark matter halo shapes

Different cosmological models predict significantly different shapes of dark matter halos. CDM cosmologies predict triaxial halos with axis ratios of the order of 0.5:1 [53–56], while pure SIDM

simulations predict halos with much rounder shapes [57–59]. Therefore, one probable astrophysical probes of the dark matter particle properties is to probe the shape of the dark matter halos.

One method to probe the structure of dark matter halos is by galaxy-galaxy lensing [80–82]. Efforts have been made to measure halo shapes by lens-shear 2-point correlation functions [83–86] to extract lensing shear as a function of position angle with respect to the lens galaxies’ principal axes, with the assumption that the lens galaxies and hosting halos being perfectly aligned. This previous work, however, has yielded inconclusive results.

In light of this, we propose a new methodology to probe the shapes of dark matter halos, namely the lens-shear-shear 3-point correlation function. Using this methodology, the anisotropic signal of lensing shear is immune of to galaxy-halo misalignments which can potentially limit 2-point lens-shear for halo anisotropy measurement. With a simple mathematical model for the projected mass distributions of dark matter halos, we construct an ellipticity estimator that sums over all possible triangular configurations of the 3-point function.

As a test for rigorousness of our methodology and model, we apply our estimator to halos from the Boishoi N-body cosmological simulations. Fig. 2.3 shows the result of applying the estimator on ensemble of halos ranging from $10^{11.6}M_{\odot}h^{-1}$ to $10^{12.8}M_{\odot}h^{-1}$, with systematic errors in the recovered ellipticity in the $\lesssim 5\%$ fractional level. In particular, the significant error at large radii for the low mass bins might be an indication of effects of nearby halos. Fig. 2.5 demonstrates the prediction accuracy of our estimator as a function of satellite contamination. This shows that satellite contamination and nearby environments could potentially hinder the ability of measuring halo anisotropy by the 3-point function. This indicates that satellite removal is necessary should this method be applied to measure halo shapes in real observations.

Previously, such high-order correlations have not been measured in lensing surveys, because a high number density of galaxies are required. With the advent of deep cosmological surveys such as the Dark Energy Survey (DES), Hyper Suprime-Cam (HSC) and LSST, it is now becoming possible to measure such three-point functions in lensing.

5.2 Assembly bias of dark matter halos

Another robust prediction by the standard cosmological model is halo assembly bias, the age dependence of halo clustering. Theoretical framework [68] and plenty of numerical simulations [65–67] have demonstrated that halo bias can depend significantly on halo properties besides virial mass, such as concentration and age. Interest in studies of halo assembly bias has risen recently because of the great advent of cosmological surveys and the availability of large cosmological simulations. In this section, we conclude our efforts of attempts to detect assembly bias for low mass halos using CFHTLenS observations and investigation of high mass halo assembly bias in cosmological simulations.

5.2.1 Observational efforts of assembly bias in galaxy-sized halos

Halo assembly bias is predicted to be significant for the low-mass halos ($\sim 10^{12} M_{\odot}$). However, recent studies [121] reported no significant evidence for any bias difference in two galaxy samples at fixed mass with different star-formation rate (SFR). Their result either implies that the SFR indicator they used does not correlate significantly with halo age, or (more interestingly) that perhaps halo assembly bias does not occur in nature for galaxy-sized halos. To address this apparent contradiction between theory and observations, we revisit this subject by using the publicly available CFHTLenS survey data.

Satellites reside in halos of larger mass than similar central galaxies, and will therefore have a larger bias. Since this mass difference and bias difference could mimic assembly bias, it is crucial to minimize contamination from satellites. We excluded $\sim 90\%$ of the galaxies. These cuts may be overly conservative, but they should leave us with a reasonably pure selection of central galaxies.

We then separate the lens galaxies into red and blue samples, with the parameter T_{BPZ} available in CFHTLenS. We further construct galaxy samples with similar masses $\lesssim 10^{12} M_{\odot}$ by performing simple NFW profile fits with weak gravitational lensing. Fig. 3.4 illustrates the stacked projected surface density of the lens galaxies for the red and blue samples, with consistent masses. We

then measure the 2-pt auto-correlation of the galaxy samples as a measurement of clustering of those galaxies, as demonstrated in Fig. 3.5. We do not find significant difference in clustering for $3h_{70}^{-1}\text{Mpc} \leq r \leq 10h_{70}^{-1}\text{Mpc}$.

Our results appear to be consistent with [121]. We caution that the null detection of assembly bias might be attributed to the small sample size due to removal of $\sim 90\%$ galaxies. Another possibility is that galaxy color might not correlate well with halo age, potentially diluting the assembly bias signal in our sample. Further work will be required in order to resolve this potential discrepancy between theory and observation. For example, the Dark Energy Survey (DES) covers 5000 deg^2 to 24^{th} magnitude in SDSS bands [18], while the Subaru Hyper-Suprime Camera survey (HSC) covers 1400 deg^2 to $r \approx 26$. Analysis of either of these data sets should easily be able to verify whether galaxies exhibit assembly bias at levels found in simulations of dark matter halos.

5.2.2 Revisiting assembly bias in cluster-sized halos

The standard cosmological model predicts the existence of assembly bias for high mass halos ($\geq 10^{14}M_{\odot}$). Nevertheless, multiple studies reported no significant assembly bias signal for cluster-sized halos when analyzing simulation data [66, 157]. If correct, that would require a dramatic rethinking of halo formation in general. Motivated by this surprising claim, we revisit the halo assembly bias for massive cluster-sized halos in ΛCDM simulations, namely the BigMDPL simulation [160].

There are multiple ways to define halo age. In previous literature [66, 157], the half-mass time, $a_{0.5}$ has been the most common definition of age, which is defined as the scale factor when a given halo’s most massive progenitor first acquires a mass fraction $f = 0.5$ of the final mass at $z = 0$. However, there is little physical reason that an arbitrarily chosen number such as $a_{0.5}$ would characterize the aspects of halo assembly bias that relate to large-scale environment. Even worse, for the mass range that we are considering (i.e. $\sim 10^{14}M_{\odot}$), we find that $a_{0.5}$ does not correlate to the large-scale environment (see Fig. 4.2).

In light of that, we leverage the entire mass accretion history of halos to construct an optimal

definition of halo age using Gaussian statistics, which would maximize the difference in large-scale bias of the two samples at fixed mass. Under that definition, we are capable to separate two halo samples with same average mass that have different large-scale biases (see Fig. 4.3). In particular, for the halo sample which exhibit high bias (red), we argue that mass must be physically removed from within the virial radius at late time (i.e. $a > 0.85$) given its slow growth rate.

Motivated by this observation, in Fig. 4.4 we investigate the phase space diagrams around the two halo samples, which characterize the relative motion and position of neighboring halos to the halo samples that we are concerned. We find that the high bias sample contains significantly more amount of matter just outside the virial radius of the halos, which was previously inside the r_{vir} one crossing time at $a \sim 0.85$. We argue that mass has been removed from r_{vir} for the high bias clusters because this recently accreted mass is on wide orbits that extend beyond r_{vir} . In other words, those high bias halos have been assigned the wrong masses, with the actual physical masses being larger than the quoted virial masses, and therefore they are more strongly biased due to the strong relation between halo mass with bias.

Finally, we argue that the more physically motivated splashback radius should provide a more correct halo mass definition, which could avoid such behavior. With the cosmological simulations with splashback halo masses coming into play, future analyses should be able to provide a more clear picture on the assembly bias of cluster-sized halos.

References

- [1] B. Diemer and A. V. Kravtsov. Dependence of the Outer Density Profiles of Halos on Their Mass Accretion Rate. *The Astrophysical Journal*, 789:1, July 2014. doi: 10.1088/0004-637X/789/1/1.
- [2] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, and et al. Planck 2015 results. XIII. Cosmological parameters. *Astronomy & Astrophysics*, 594:A13, September 2016. doi: 10.1051/0004-6361/201525830.
- [3] G. Hinshaw, J. L. Weiland, R. S. Hill, N. Odegard, D. Larson, C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, N. Jarosik, E. Komatsu, M. R. Nolte, L. Page, D. N. Spergel, E. Wollack, M. Halpern, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright. Five-Year Wilkinson Microwave Anisotropy Probe Observations: Data Processing, Sky Maps, and Basic Results. *The Astrophysical Journal Supplement*, 180:225–245, February 2009. doi: 10.1088/0067-0049/180/2/225.
- [4] C. L. Reichardt, P. A. R. Ade, J. J. Bock, J. R. Bond, J. A. Brevik, C. R. Contaldi, M. D. Daub, J. T. Dempsey, J. H. Goldstein, W. L. Holzapfel, C. L. Kuo, A. E. Lange, M. Lueker, M. Newcomb, J. B. Peterson, J. Ruhl, M. C. Runyan, and Z. Staniszewski. High-Resolution CMB Power Spectrum from the Complete ACBAR Data Set. *The Astrophysical Journal*, 694:1200–1219, April 2009. doi: 10.1088/0004-637X/694/2/1200.
- [5] M. L. Brown, P. Ade, J. Bock, M. Bowden, G. Cahill, P. G. Castro, S. Church, T. Culverhouse, R. B. Friedman, K. Ganga, W. K. Gear, S. Gupta, J. Hinderks, J. Kovac, A. E. Lange, E. Leitch, S. J. Melhuish, Y. Memari, J. A. Murphy, A. Orlando, C. O’Sullivan, L. Piccirillo, C. Pryke, N. Rajguru, B. Rusholme, R. Schwarz, A. N. Taylor, K. L. Thompson, A. H. Turner, E. Y. S. Wu, M. Zemcov, and QUaD Collaboration. Improved Measurements of the Temperature and Polarization of the Cosmic Microwave Background from QUaD. *The Astrophysical Journal*, 705:978–999, November 2009. doi: 10.1088/0004-637X/705/1/978.
- [6] B. A. Reid, W. J. Percival, D. J. Eisenstein, L. Verde, D. N. Spergel, R. A. Skibba, N. A. Bahcall, T. Budavari, J. A. Frieman, M. Fukugita, J. R. Gott, J. E. Gunn, Ž. Ivezić, G. R. Knapp, R. G. Kron, R. H. Lupton, T. A. McKay, A. Meiksin, R. C. Nichol, A. C. Pope, D. J. Schlegel, D. P. Schneider, C. Stoughton, M. A. Strauss, A. S. Szalay, M. Tegmark, M. S. Vogeley, D. H. Weinberg, D. G. York, and I. Zehavi. Cosmological constraints from the clustering of the Sloan Digital Sky Survey DR7 luminous red galaxies. *Monthly Notices*

of the *Royal Astronomical Society*, 404:60–85, May 2010. doi: 10.1111/j.1365-2966.2010.16276.x.

- [7] W. J. Percival, B. A. Reid, D. J. Eisenstein, N. A. Bahcall, T. Budavari, J. A. Frieman, M. Fukugita, J. E. Gunn, Ž. Ivezić, G. R. Knapp, R. G. Kron, J. Loveday, R. H. Lupton, T. A. McKay, A. Meiksin, R. C. Nichol, A. C. Pope, D. J. Schlegel, D. P. Schneider, D. N. Spergel, C. Stoughton, M. A. Strauss, A. S. Szalay, M. Tegmark, M. S. Vogeley, D. H. Weinberg, D. G. York, and I. Zehavi. Baryon acoustic oscillations in the Sloan Digital Sky Survey Data Release 7 galaxy sample. *Monthly Notices of the Royal Astronomical Society*, 401:2148–2168, February 2010. doi: 10.1111/j.1365-2966.2009.15812.x.
- [8] M. Hicken, W. M. Wood-Vasey, S. Blondin, P. Challis, S. Jha, P. L. Kelly, A. Rest, and R. P. Kirshner. Improved Dark Energy Constraints from ~100 New CfA Supernova Type Ia Light Curves. *The Astrophysical Journal*, 700:1097–1140, August 2009. doi: 10.1088/0004-637X/700/2/1097.
- [9] R. Kessler, A. C. Becker, D. Cinabro, J. Vanderplas, J. A. Frieman, J. Marriner, T. M. Davis, B. Dilday, J. Holtzman, S. W. Jha, H. Lampeitl, M. Sako, M. Smith, C. Zheng, R. C. Nichol, B. Bassett, R. Bender, D. L. Depoy, M. Doi, E. Elson, A. V. Filippenko, R. J. Foley, P. M. Garnavich, U. Hopp, Y. Ihara, W. Ketzeback, W. Kollatschny, K. Konishi, J. L. Marshall, R. J. McMillan, G. Miknaitis, T. Morokuma, E. Mörtzell, K. Pan, J. L. Prieto, M. W. Richmond, A. G. Riess, R. Romani, D. P. Schneider, J. Sollerman, N. Takanashi, K. Tokita, K. van der Heyden, J. C. Wheeler, N. Yasuda, and D. York. First-Year Sloan Digital Sky Survey-II Supernova Results: Hubble Diagram and Cosmological Parameters. *The Astrophysical Journal Supplement*, 185:32–84, November 2009. doi: 10.1088/0067-0049/185/1/32.
- [10] A. Vikhlinin, A. V. Kravtsov, R. A. Burenin, H. Ebeling, W. R. Forman, A. Hornstrup, C. Jones, S. S. Murray, D. Nagai, H. Quintana, and A. Voevodkin. Chandra Cluster Cosmology Project III: Cosmological Parameter Constraints. *The Astrophysical Journal*, 692:1060–1074, February 2009. doi: 10.1088/0004-637X/692/2/1060.
- [11] A. Mantz, S. W. Allen, D. Rapetti, and H. Ebeling. The observed growth of massive galaxy clusters - I. Statistical methods and cosmological constraints. *Monthly Notices of the Royal Astronomical Society*, 406:1759–1772, August 2010. doi: 10.1111/j.1365-2966.2010.16992.x.
- [12] A. G. Riess, L. Macri, S. Casertano, M. Sosey, H. Lampeitl, H. C. Ferguson, A. V. Filippenko, S. W. Jha, W. Li, R. Chornock, and D. Sarkar. A Redetermination of the Hubble Constant with the Hubble Space Telescope from a Differential Distance Ladder. *The Astrophysical Journal*, 699:539–563, July 2009. doi: 10.1088/0004-637X/699/1/539.
- [13] S. H. Suyu, P. J. Marshall, M. W. Auger, S. Hilbert, R. D. Blandford, L. V. E. Koopmans, C. D. Fassnacht, and T. Treu. Dissecting the Gravitational lens B1608+656. II. Precision Measurements of the Hubble Constant, Spatial Curvature, and the Dark Energy Equation of State. *The Astrophysical Journal*, 711:201–221, March 2010. doi: 10.1088/0004-637X/711/1/201.

- [14] R. Fadely, C. R. Keeton, R. Nakajima, and G. M. Bernstein. Improved Constraints on the Gravitational Lens Q0957+561. II. Strong Lensing. *The Astrophysical Journal*, 711:246–267, March 2010. doi: 10.1088/0004-637X/711/1/246.
- [15] R. Massey, J. Rhodes, A. Leauthaud, P. Capak, R. Ellis, A. Koekemoer, A. Réfrégier, N. Scoville, J. E. Taylor, J. Albert, J. Bergé, C. Heymans, D. Johnston, J.-P. Kneib, Y. Mellier, B. Mobasher, E. Semboloni, P. Shopbell, L. Tasca, and L. Van Waerbeke. COSMOS: Three-dimensional Weak Lensing and the Growth of Structure. *The Astrophysical Journal Supplement*, 172:239–253, September 2007. doi: 10.1086/516599.
- [16] L. Fu, E. Semboloni, H. Hoekstra, M. Kilbinger, L. van Waerbeke, I. Tereno, Y. Mellier, C. Heymans, J. Coupon, K. Benabed, J. Benjamin, E. Bertin, O. Doré, M. J. Hudson, O. Ilbert, R. Maoli, C. Marmo, H. J. McCracken, and B. Ménard. Very weak lensing in the CFHTLS wide: cosmology from cosmic shear in the linear regime. *Astronomy & Astrophysics*, 479:9–25, February 2008. doi: 10.1051/0004-6361:20078522.
- [17] T. Schrabback, J. Hartlap, B. Joachimi, M. Kilbinger, P. Simon, K. Benabed, M. Bradač, T. Eifler, T. Erben, C. D. Fassnacht, F. W. High, S. Hilbert, H. Hildebrandt, H. Hoekstra, K. Kuijken, P. J. Marshall, Y. Mellier, E. Morganson, P. Schneider, E. Semboloni, L. van Waerbeke, and M. Velander. Evidence of the accelerated expansion of the Universe from weak lensing tomography with COSMOS. *Astronomy & Astrophysics*, 516:A63, June 2010. doi: 10.1051/0004-6361/200913577.
- [18] The Dark Energy Survey Collaboration. The Dark Energy Survey. *ArXiv Astrophysics e-prints*, October 2005.
- [19] L. Bergström. Dark matter evidence, particle physics candidates and detection methods. *Annalen der Physik*, 524:479–496, October 2012. doi: 10.1002/andp.201200116.
- [20] N. F. Bell, A. J. Galea, J. B. Dent, T. D. Jacques, L. M. Krauss, and T. J. Weiler. Searching for dark matter at the LHC with a mono-Z. *Physical Review D*, 86(9):096011, November 2012. doi: 10.1103/PhysRevD.86.096011.
- [21] A. Askew, S. Chauhan, B. Penning, W. Shepherd, and M. Tripathi. Searching for dark matter at hadron colliders. *International Journal of Modern Physics A*, 29:1430041, September 2014. doi: 10.1142/S0217751X14300415.
- [22] L. Baudis. Dark matter detection. *Journal of Physics G Nuclear Physics*, 43(4):044001, August 2016. doi: 10.1088/0954-3899/43/4/044001.
- [23] A. Ibarra, D. Tran, and C. Weniger. Indirect Searches for Decaying Dark Matter. *International Journal of Modern Physics A*, 28:1330040, October 2013. doi: 10.1142/S0217751X13300408.
- [24] D. N. Spergel and P. J. Steinhardt. Observational Evidence for Self-Interacting Cold Dark Matter. *Physical Review Letters*, 84:3760–3763, April 2000. doi: 10.1103/PhysRevLett.84.3760.

- [25] J. Miralda-Escudé. A Test of the Collisional Dark Matter Hypothesis from Cluster Lensing. *The Astrophysical Journal*, 564:60–64, January 2002. doi: 10.1086/324138.
- [26] D. Clowe, M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky. A Direct Empirical Proof of the Existence of Dark Matter. *The Astrophysical Journal Lett.*, 648:L109–L113, September 2006. doi: 10.1086/508162.
- [27] M. Bradač, D. Clowe, A. H. Gonzalez, P. Marshall, W. Forman, C. Jones, M. Markevitch, S. Randall, T. Schrabback, and D. Zaritsky. Strong and Weak Lensing United. III. Measuring the Mass Distribution of the Merging Galaxy Cluster 1ES 0657-558. *The Astrophysical Journal*, 652:937–947, December 2006. doi: 10.1086/508601.
- [28] P. J. E. Peebles. *The large-scale structure of the universe*. 1980.
- [29] P. J. E. Peebles. *Principles of Physical Cosmology*. 1993.
- [30] R. H. Wechsler and J. L. Tinker. The Connection between Galaxies and their Dark Matter Halos. *ArXiv e-prints*, April 2018.
- [31] M. Persic, P. Salucci, and F. Stel. The universal rotation curve of spiral galaxies - I. The dark matter connection. *Monthly Notices of the Royal Astronomical Society*, 281:27–47, July 1996. doi: 10.1093/mnras/281.1.27.
- [32] J. F. Navarro. The Cosmological Significance of Disk Galaxy Rotation Curves. *ArXiv Astrophysics e-prints*, July 1998.
- [33] E. Corbelli and P. Salucci. The extended rotation curve and the dark matter halo of M33. *Monthly Notices of the Royal Astronomical Society*, 311:441–447, January 2000. doi: 10.1046/j.1365-8711.2000.03075.x.
- [34] R. J. Scherrer and D. H. Weinberg. Constraints on the Effects of Locally Biased Galaxy Formation. *The Astrophysical Journal*, 504:607–611, September 1998. doi: 10.1086/306113.
- [35] G. F. Smoot, C. L. Bennett, A. Kogut, E. L. Wright, J. Aymon, N. W. Boggess, E. S. Cheng, G. de Amici, S. Gulkis, M. G. Hauser, G. Hinshaw, P. D. Jackson, M. Janssen, E. Kaita, T. Kelsall, P. Keegstra, C. Lineweaver, K. Loewenstein, P. Lubin, J. Mather, S. S. Meyer, S. H. Moseley, T. Murdock, L. Rokke, R. F. Silverberg, L. Tenorio, R. Weiss, and D. T. Wilkinson. Structure in the COBE differential microwave radiometer first-year maps. *The Astrophysical Journal Lett.*, 396:L1–L5, September 1992. doi: 10.1086/186504.
- [36] E. V. Linder and A. Jenkins. Cosmic structure growth and dark energy. *Monthly Notices of the Royal Astronomical Society*, 346:573–583, December 2003. doi: 10.1046/j.1365-2966.2003.07112.x.
- [37] D. Huterer, D. Kirkby, R. Bean, A. Connolly, K. Dawson, S. Dodelson, A. Evrard, B. Jain, M. Jarvis, E. Linder, R. Mandelbaum, M. May, A. Raccanelli, B. Reid, E. Rozo, F. Schmidt, N. Sehgal, A. Slosar, A. van Engelen, H.-Y. Wu, and G. Zhao. Growth of cosmic structure: Probing dark energy beyond expansion. *Astroparticle Physics*, 63:23–41, March 2015. doi: 10.1016/j.astropartphys.2014.07.004.

- [38] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White. The evolution of large-scale structure in a universe dominated by cold dark matter. *The Astrophysical Journal*, 292: 371–394, May 1985. doi: 10.1086/163168.
- [39] V. R. Eke, S. Cole, and C. S. Frenk. Cluster evolution as a diagnostic for Ω . *Monthly Notices of the Royal Astronomical Society*, 282, September 1996. doi: 10.1093/mnras/282.1.263.
- [40] W. H. Press and P. Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *The Astrophysical Journal*, 187:425–438, February 1974. doi: 10.1086/152650.
- [41] G. L. Bryan and M. L. Norman. Statistical Properties of X-Ray Clusters: Analytic and Numerical Comparisons. *The Astrophysical Journal*, 495:80–99, March 1998. doi: 10.1086/305262.
- [42] T. Padmanabhan. *Structure Formation in the Universe*. May 1993.
- [43] J. A. Peacock. *Cosmological Physics*. January 1999.
- [44] A. V. Kravtsov and S. Borgani. Formation of Galaxy Clusters. *Annual review of astronomy and astrophysics*, 50:353–409, September 2012. doi: 10.1146/annurev-astro-081811-125502.
- [45] J. F. Navarro, C. S. Frenk, and S. D. M. White. The Structure of Cold Dark Matter Halos. *The Astrophysical Journal*, 462:563, May 1996. doi: 10.1086/177173.
- [46] J. F. Navarro, C. S. Frenk, and S. D. M. White. A Universal Density Profile from Hierarchical Clustering. *The Astrophysical Journal*, 490:493–508, December 1997.
- [47] N. Dalal, Y. Lithwick, and M. Kuhlen. The Origin of Dark Matter Halo Profiles. *ArXiv e-prints*, October 2010.
- [48] A. Dekel, I. Arad, J. Devor, and Y. Birnboim. Dark Halo Cusp: Asymptotic Convergence. *The Astrophysical Journal*, 588:680–695, May 2003. doi: 10.1086/374328.
- [49] E. Salvador-Solé, J. Viñas, A. Manrique, and S. Serra. Theoretical dark matter halo density profile. *Monthly Notices of the Royal Astronomical Society*, 423:2190–2202, July 2012. doi: 10.1111/j.1365-2966.2012.21066.x.
- [50] A. A. El-Zant, Y. Hoffman, J. Primack, F. Combes, and I. Shlosman. Flat-cored Dark Matter in Cuspy Clusters of Galaxies. *The Astrophysical Journal Lett.*, 607:L75–L78, June 2004. doi: 10.1086/421938.
- [51] A. Manrique, A. Raig, E. Salvador-Solé, T. Sanchis, and J. M. Solanes. On the Origin of the Inner Structure of Halos. *The Astrophysical Journal*, 593:26–37, August 2003. doi: 10.1086/376403.

- [52] D. Syer and S. D. M. White. Dark halo mergers and the formation of a universal profile. *Monthly Notices of the Royal Astronomical Society*, 293:337, February 1998. doi: 10.1046/j.1365-8711.1998.01285.x.
- [53] J. Dubinski and R. G. Carlberg. The structure of cold dark matter halos. *The Astrophysical Journal*, 378:496–503, September 1991. doi: 10.1086/170451.
- [54] Y. P. Jing and Y. Suto. Triaxial Modeling of Halo Density Profiles with High-Resolution N-Body Simulations. *The Astrophysical Journal*, 574:538–553, August 2002. doi: 10.1086/341065.
- [55] B. Allgood, R. A. Flores, J. R. Primack, A. V. Kravtsov, R. H. Wechsler, A. Faltenbacher, and J. S. Bullock. The shape of dark matter haloes: dependence on mass, redshift, radius and formation. *Monthly Notices of the Royal Astronomical Society*, 367:1781–1796, April 2006. doi: 10.1111/j.1365-2966.2006.10094.x.
- [56] M. D. Schneider, C. S. Frenk, and S. Cole. The shapes and alignments of dark matter halos. *Journal of Cosmology and Astroparticle Physics*, 5:030, May 2012. doi: 10.1088/1475-7516/2012/05/030.
- [57] R. Davé, D. N. Spergel, P. J. Steinhardt, and B. D. Wandelt. Halo Properties in Cosmological Simulations of Self-interacting Cold Dark Matter . *The Astrophysical Journal*, 547:574–589, 2001.
- [58] A. H. G. Peter, M. Rocha, J. S. Bullock, and M. Kaplinghat. Cosmological simulations with self-interacting dark matter - II. Halo shapes versus observations. *Monthly Notices of the Royal Astronomical Society*, 430:105–120, March 2013. doi: 10.1093/mnras/sts535.
- [59] M. Rocha, A. H. G. Peter, J. S. Bullock, M. Kaplinghat, S. Garrison-Kimmel, J. Oñorbe, and L. A. Moustakas. Cosmological simulations with self-interacting dark matter - I. Constant-density cores and substructure. *Monthly Notices of the Royal Astronomical Society*, 430: 81–104, March 2013. doi: 10.1093/mnras/sts514.
- [60] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay. The statistics of peaks of Gaussian random fields. *The Astrophysical Journal*, 304:15–61, May 1986. doi: 10.1086/164143.
- [61] S. Cole and N. Kaiser. Biased clustering in the cold dark matter cosmogony. *Monthly Notices of the Royal Astronomical Society*, 237:1127–1146, April 1989. doi: 10.1093/mnras/237.4.1127.
- [62] A. Cooray and R. Sheth. Halo models of large scale structure. *Physics Reports*, 372:1–129, December 2002. doi: 10.1016/S0370-1573(02)00276-4.
- [63] P. Valageas. Large-scale bias of dark matter halos. *Astronomy & Astrophysics*, 525:A98, January 2011. doi: 10.1051/0004-6361/201015699.
- [64] H. J. Mo and S. D. M. White. An analytic model for the spatial clustering of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 282:347–361, September 1996.

- [65] R. H. Wechsler, A. R. Zentner, J. S. Bullock, A. V. Kravtsov, and B. Allgood. The Dependence of Halo Clustering on Halo Formation History, Concentration, and Occupation. *The Astrophysical Journal*, 652:71–84, November 2006. doi: 10.1086/507120.
- [66] L. Gao and S. D. M. White. Assembly bias in the clustering of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 377:L5–L9, April 2007. doi: 10.1111/j.1745-3933.2007.00292.x.
- [67] L. Gao, V. Springel, and S. D. M. White. The age dependence of halo clustering. *Monthly Notices of the Royal Astronomical Society*, 363:L66–L70, October 2005. doi: 10.1111/j.1745-3933.2005.00084.x.
- [68] N. Dalal, M. White, J. R. Bond, and A. Shirokov. Halo Assembly Bias in Hierarchical Structure Formation. *The Astrophysical Journal*, 687:12–21, November 2008. doi: 10.1086/591512.
- [69] V. Trimble. Existence and nature of dark matter in the universe. *Annual review of astronomy and astrophysics*, 25:425–472, 1987. doi: 10.1146/annurev.aa.25.090187.002233.
- [70] M. Bartelmann and P. Schneider. Weak gravitational lensing. *Physical Reports*, 340:291–472, January 2001. doi: 10.1016/S0370-1573(00)00082-X.
- [71] M. Kaplinghat, R. E. Keeley, T. Linden, and H.-B. Yu. Tying Dark Matter to Baryons with Self-interactions. *ArXiv e-prints*, November 2013.
- [72] A. Helmi. Velocity Trends in the Debris of Sagittarius and the Shape of the Dark Matter Halo of Our Galaxy. *The Astrophysical Journal*, 610:L97–L100, 2004.
- [73] D. R. Law and S. R. Majewski. The Sagittarius Dwarf Galaxy: A Model for Evolution in a Triaxial Milky Way Halo. *The Astrophysical Journal*, 714:229–254, 2010.
- [74] N. Deg and L. Widrow. The Sagittarius stream and halo triaxiality. *Monthly Notices of the Royal Astronomical Society*, 428:912–922, 2013.
- [75] R. Ibata, G. F. Lewis, N. F. Martin, M. Bellazzini, and M. Correnti. Does the Sagittarius Stream Constrain the Milky Way Halo to be Triaxial? *The Astrophysical Journal Lett.*, 765: L15, 2013.
- [76] C. Vera-Ciro and A. Helmi. Constraints on the Shape of the Milky Way Dark Matter Halo from the Sagittarius Stream. *The Astrophysical Journal Lett.*, 773:L4, 2013.
- [77] G. van de Ven, J. Falcón-Barroso, R. M. McDermid, M. Cappellari, B. W. Miller, and P. T. de Zeeuw. The Einstein Cross: Constraint on Dark Matter from Stellar Dynamics and Gravitational Lensing. *The Astrophysical Journal*, 719:1481–1496, August 2010. doi: 10.1088/0004-637X/719/2/1481.

- [78] S. H. Suyu, S. W. Hensel, J. P. McKean, C. D. Fassnacht, T. Treu, A. Halkola, M. Norbury, N. Jackson, P. Schneider, D. Thompson, M. W. Auger, L. V. E. Koopmans, and K. Matthews. Disentangling Baryons and Dark Matter in the Spiral Gravitational Lens B1933+503. *The Astrophysical Journal*, 750:10, May 2012. doi: 10.1088/0004-637X/750/1/10.
- [79] J. Bailin, C. Power, P. Norberg, D. Zaritsky, and B. K. Gibson. The anisotropic distribution of satellite galaxies. *Monthly Notices of the Royal Astronomical Society*, 390:1133–1156, November 2008. doi: 10.1111/j.1365-2966.2008.13828.x.
- [80] E. S. Sheldon, D. E. Johnston, J. A. Frieman, R. Scranton, T. A. McKay, A. J. Connolly, T. Budavári, I. Zehavi, N. A. Bahcall, J. Brinkmann, and M. Fukugita. The Galaxy-Mass Correlation Function Measured from Weak Lensing in the Sloan Digital Sky Survey. *The Astronomical Journal*, 127:2544–2564, May 2004. doi: 10.1086/383293.
- [81] R. Mandelbaum, U. Seljak, R. J. Cool, M. Blanton, C. M. Hirata, and J. Brinkmann. Density profiles of galaxy groups and clusters from SDSS galaxy-galaxy weak lensing. *Monthly Notices of the Royal Astronomical Society*, 372:758–776, October 2006. doi: 10.1111/j.1365-2966.2006.10906.x.
- [82] D. E. Johnston, E. S. Sheldon, R. H. Wechsler, E. Rozo, B. P. Koester, J. A. Frieman, T. A. McKay, A. E. Evrard, M. R. Becker, and J. Annis. Cross-correlation Weak Lensing of SDSS galaxy Clusters II: Cluster Density Profiles and the Mass–Richness Relation. *ArXiv e-prints*, September 2007.
- [83] H. Hoekstra, H. K. C. Yee, and M. D. Gladders. Properties of Galaxy Dark Matter Halos from Weak Lensing . *The Astrophysical Journal*, 606:67–77, 2004.
- [84] R. Mandelbaum, C. M. Hirata, T. Broderick, U. Seljak, and J. Brinkmann. Ellipticity of dark matter haloes with galaxy-galaxy weak lensing. *Monthly Notices of the Royal Astronomical Society*, 370:1008–1024, 2006.
- [85] L. C. Parker, H. Hoekstra, M. J. Hudson, L. van Waerbeke, and Y. Mellier. The Masses and Shapes of Dark Matter Halos from Galaxy-Galaxy Lensing in the CFHT Legacy Survey . *The Astrophysical Journal*, 669:21–31, 2007.
- [86] E. van Uitert, H. Hoekstra, T. Schrabback, D. G. Gilbank, M. D. Gladders, and H. K. C. Yee. Constraints on the shapes of galaxy dark matter haloes from weak gravitational lensing. *Astronomy & Astrophysics*, 545:A71, September 2012. doi: 10.1051/0004-6361/201219295.
- [87] A. Tenneti, R. Mandelbaum, T. Di Matteo, Y. Feng, and N. Khandai. Galaxy shapes and intrinsic alignments in the MassiveBlack-II simulation. *Monthly Notices of the Royal Astronomical Society*, 441:470–485, June 2014. doi: 10.1093/mnras/stu586.
- [88] P. Bett. Halo shapes from weak lensing: the impact of galaxy-halo misalignment. *Monthly Notices of the Royal Astronomical Society*, 420:3303–3323, March 2012. doi: 10.1111/j.1365-2966.2011.20258.x.

- [89] P. J. Howell and T. G. Brainerd. Galaxy-galaxy lensing by non-spherical haloes - I. Theoretical considerations. *Monthly Notices of the Royal Astronomical Society*, 407:891–909, September 2010. doi: 10.1111/j.1365-2966.2010.16979.x.
- [90] P. Simon, P. Schneider, and D. Kübler. Towards an understanding of third-order galaxy-galaxy lensing. *Astronomy & Astrophysics*, 548:A102, December 2012. doi: 10.1051/0004-6361/201218993.
- [91] P. Simon, T. Erben, P. Schneider, C. Heymans, H. Hildebrandt, H. Hoekstra, T. D. Kitching, Y. Mellier, L. Miller, L. Van Waerbeke, C. Bonnett, J. Coupon, L. Fu, M. J. Hudson, K. Kuijken, B. T. P. Rowe, T. Schrabback, E. Semboloni, and M. Velander. CFHTLenS: higher order galaxy-mass correlations probed by galaxy-galaxy-galaxy lensing. *Monthly Notices of the Royal Astronomical Society*, 430:2476–2498, April 2013. doi: 10.1093/mnras/stt069.
- [92] L. Fu, M. Kilbinger, T. Erben, C. Heymans, H. Hildebrandt, H. Hoekstra, T. D. Kitching, Y. Mellier, L. Miller, E. Semboloni, P. Simon, L. Van Waerbeke, J. Coupon, J. Harnois-Déraps, M. J. Hudson, K. Kuijken, B. Rowe, T. Schrabback, S. Vafaei, and M. Velander. CFHTLenS: cosmological constraints from a combination of cosmic shear two-point and three-point correlations. *Monthly Notices of the Royal Astronomical Society*, 441:2725–2743, July 2014. doi: 10.1093/mnras/stu754.
- [93] A. A. Klypin, S. Trujillo-Gomez, and J. Primack. Dark Matter Halos in the Standard Cosmological Model: Results from the Bolshoi Simulation. *The Astrophysical Journal*, 740: 102, October 2011. doi: 10.1088/0004-637X/740/2/102.
- [94] C. Chang, M. Jarvis, B. Jain, S. M. Kahn, D. Kirkby, A. Connolly, S. Krughoff, E.-H. Peng, and J. R. Peterson. The effective number density of galaxies for weak lensing measurements in the LSST project. *Monthly Notices of the Royal Astronomical Society*, 434:2121–2135, September 2013. doi: 10.1093/mnras/stt1156.
- [95] R. A. C. Croft and C. A. Metzler. Weak-Lensing Surveys and the Intrinsic Correlation of Galaxy Ellipticities. *The Astrophysical Journal*, 545:561–571, December 2000. doi: 10.1086/317856.
- [96] A. Heavens, A. Refregier, and C. Heymans. Intrinsic correlation of galaxy shapes: implications for weak lensing measurements. *Monthly Notices of the Royal Astronomical Society*, 319:649–656, December 2000. doi: 10.1046/j.1365-8711.2000.03907.x.
- [97] R. G. Crittenden, P. Natarajan, U.-L. Pen, and T. Theuns. Spin-induced Galaxy Alignments and Their Implications for Weak-Lensing Measurements. *The Astrophysical Journal*, 559: 552–571, October 2001. doi: 10.1086/322370.
- [98] C. M. Hirata and U. Seljak. Intrinsic alignment-lensing interference as a contaminant of cosmic shear. *Physical Review D*, 70(6):063526, September 2004. doi: 10.1103/PhysRevD.70.063526.

- [99] C. M. Hirata, R. Mandelbaum, U. Seljak, J. Guzik, N. Padmanabhan, C. Blake, J. Brinkmann, T. Budávári, A. Connolly, I. Csabai, R. Scranton, and A. S. Szalay. Galaxy-galaxy weak lensing in the Sloan Digital Sky Survey: intrinsic alignments and shear calibration errors. *Monthly Notices of the Royal Astronomical Society*, 353:529–549, September 2004. doi: 10.1111/j.1365-2966.2004.08090.x.
- [100] R. Mandelbaum, C. M. Hirata, M. Ishak, U. Seljak, and J. Brinkmann. Detection of large-scale intrinsic ellipticity-density correlation from the Sloan Digital Sky Survey and implications for weak lensing surveys. *Monthly Notices of the Royal Astronomical Society*, 367: 611–626, April 2006. doi: 10.1111/j.1365-2966.2005.09946.x.
- [101] C. M. Hirata, R. Mandelbaum, M. Ishak, U. Seljak, R. Nichol, K. A. Pimbblet, N. P. Ross, and D. Wake. Intrinsic galaxy alignments from the 2SLAQ and SDSS surveys: luminosity and redshift scalings and implications for weak lensing surveys. *Monthly Notices of the Royal Astronomical Society*, 381:1197–1218, November 2007. doi: 10.1111/j.1365-2966.2007.12312.x.
- [102] J. Blazek, M. McQuinn, and U. Seljak. Testing the tidal alignment model of galaxy intrinsic alignment. *Journal of Cosmology and Astroparticle Physics*, 5:010, May 2011. doi: 10.1088/1475-7516/2011/05/010.
- [103] J. Blazek, R. Mandelbaum, U. Seljak, and R. Nakajima. Separating intrinsic alignment and galaxy-galaxy lensing. *Journal of Cosmology and Astroparticle Physics*, 5:041, May 2012. doi: 10.1088/1475-7516/2012/05/041.
- [104] M. Jarvis, M. Takada, B. Jain, and G. Bernstein. Weak Lensing Cosmology with LSST: Three-Point Shear Correlations. In *American Astronomical Society Meeting Abstracts*, volume 36 of *Bulletin of the American Astronomical Society*, page 108.17, December 2004.
- [105] E. Semboloni, T. Schrabback, L. van Waerbeke, S. Vafaei, J. Hartlap, and S. Hilbert. Weak lensing from space: first cosmological constraints from three-point shear statistics. *Monthly Notices of the Royal Astronomical Society*, 410:143–160, January 2011. doi: 10.1111/j.1365-2966.2010.17430.x.
- [106] C Sánchez, M Carrasco Kind, H Lin, R Miquel, FB Abdalla, A Amara, M Banerji, C Bonnett, R Brunner, D Capozzi, et al. Photometric redshift analysis in the dark energy survey science verification data. *Monthly Notices of the Royal Astronomical Society*, 445(2):1482–1506, 2014.
- [107] H. Hildebrandt, T. Erben, K. Kuijken, L. van Waerbeke, C. Heymans, J. Coupon, J. Benjamin, C. Bonnett, L. Fu, H. Hoekstra, T. D. Kitching, Y. Mellier, L. Miller, M. Velander, M. J. Hudson, B. T. P. Rowe, T. Schrabback, E. Semboloni, and N. Benítez. CFHTLenS: improving the quality of photometric redshifts with precision photometry. *Monthly Notices of the Royal Astronomical Society*, 421:2355–2367, April 2012. doi: 10.1111/j.1365-2966.2012.20468.x.

- [108] J. Benjamin, L. van Waerbeke, B. Ménard, and M. Kilbinger. Photometric redshifts: estimating their contamination and distribution using clustering information. *Monthly Notices of the Royal Astronomical Society*, 408:1168–1180, October 2010. doi: 10.1111/j.1365-2966.2010.17191.x.
- [109] S. Kazantzidis, A. V. Kravtsov, A. R. Zentner, B. Allgood, D. Nagai, and B. Moore. The Effect of Gas Cooling on the Shapes of Dark Matter Halos. *The Astrophysical Journal Lett.*, 611:L73–L76, August 2004. doi: 10.1086/423992.
- [110] S. E. Bryan, S. T. Kay, A. R. Duffy, J. Schaye, C. Dalla Vecchia, and C. M. Booth. The impact of baryons on the spins and shapes of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 429:3316–3329, March 2013. doi: 10.1093/mnras/sts587.
- [111] T. Baldauf, R. E. Smith, U. Seljak, and R. Mandelbaum. Algorithm for the direct reconstruction of the dark matter correlation function from weak lensing and galaxy clustering. *Physical Review D*, 81(6):063531, March 2010. doi: 10.1103/PhysRevD.81.063531.
- [112] N. Kaiser. Evolution and clustering of rich clusters. *Monthly Notices of the Royal Astronomical Society*, 222:323–345, September 1986.
- [113] H. Guo, Z. Zheng, P. S. Behroozi, I. Zehavi, C.-H. Chuang, J. Comparat, G. Favole, S. Gottloeber, A. Klypin, F. Prada, S. A. Rodriguez-Torres, D. H. Weinberg, and G. Yepes. Modelling Galaxy Clustering: Halo Occupation Distribution versus Subhalo Matching. *ArXiv e-prints*, August 2015.
- [114] R. Mandelbaum, W. Wang, Y. Zu, S. White, B. Henriques, and S. More. Strong bimodality in the host halo mass of central galaxies from galaxy-galaxy lensing. *ArXiv e-prints*, September 2015.
- [115] Y. Zu and R. Mandelbaum. Mapping stellar content to dark matter halos. II. Halo mass is the main driver of galaxy quenching. *ArXiv e-prints*, September 2015.
- [116] U. Seljak and M. S. Warren. Large-scale bias and stochasticity of haloes and dark matter. *Monthly Notices of the Royal Astronomical Society*, 355:129–136, November 2004. doi: 10.1111/j.1365-2966.2004.08297.x.
- [117] J. L. Tinker, B. E. Robertson, A. V. Kravtsov, A. Klypin, M. S. Warren, G. Yepes, and S. Gottlöber. The Large-scale Bias of Dark Matter Halos: Numerical Calibration and Model Tests. *The Astrophysical Journal*, 724:878–886, December 2010. doi: 10.1088/0004-637X/724/2/878.
- [118] J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser. Excursion set mass functions for hierarchical Gaussian fluctuations. *The Astrophysical Journal*, 379:440–460, October 1991. doi: 10.1086/170520.
- [119] A. P. Hearin, D. F. Watson, M. R. Becker, R. Reyes, A. A. Berlind, and A. R. Zentner. The dark side of galaxy colour: evidence from new SDSS measurements of galaxy clustering and lensing. *Monthly Notices of the Royal Astronomical Society*, 444:729–743, October 2014. doi: 10.1093/mnras/stu1443.

- [120] X. Yang, H. J. Mo, and F. C. van den Bosch. Observational Evidence for an Age Dependence of Halo Bias. *The Astrophysical Journal Lett.*, 638:L55–L58, February 2006. doi: 10.1086/501069.
- [121] Y.-T. Lin, R. Mandelbaum, Y.-H. Huang, H.-J. Huang, N. Dalal, B. Diemer, H.-Y. Jian, and A. Kravtsov. On Detecting Halo Assembly Bias with Galaxy Populations. *ArXiv e-prints*, April 2015.
- [122] H. Miyatake, S. More, M. Takada, D. N. Spergel, R. Mandelbaum, E. S. Rykoff, and E. Rozo. Evidence of Halo Assembly Bias in Massive Clusters. *ArXiv e-prints*, June 2015.
- [123] E. Komatsu, K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. R. Nolta, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *The Astrophysical Journal Supplement*, 192:18, February 2011. doi: 10.1088/0067-0049/192/2/18.
- [124] C. Heymans, L. Van Waerbeke, L. Miller, T. Erben, H. Hildebrandt, H. Hoekstra, T. D. Kitching, Y. Mellier, P. Simon, C. Bonnett, J. Coupon, L. Fu, J. Harnois Déraps, M. J. Hudson, M. Kilbinger, K. Kuijken, B. Rowe, T. Schrabback, E. Semboloni, E. van Uitert, S. Vafaei, and M. Velander. CFHTLenS: the Canada-France-Hawaii Telescope Lensing Survey. *Monthly Notices of the Royal Astronomical Society*, 427:146–166, November 2012. doi: 10.1111/j.1365-2966.2012.21952.x.
- [125] T. Erben, H. Hildebrandt, L. Miller, L. van Waerbeke, C. Heymans, H. Hoekstra, T. D. Kitching, Y. Mellier, J. Benjamin, C. Blake, C. Bonnett, O. Cordes, J. Coupon, L. Fu, R. Gavazzi, B. Gillis, E. Grocutt, S. D. J. Gwyn, K. Holhjem, M. J. Hudson, M. Kilbinger, K. Kuijken, M. Milkeraitis, B. T. P. Rowe, T. Schrabback, E. Semboloni, P. Simon, M. Smit, O. Toader, S. Vafaei, E. van Uitert, and M. Velander. CFHTLenS: the Canada-France-Hawaii Telescope Lensing Survey - imaging data and catalogue products. *Monthly Notices of the Royal Astronomical Society*, 433:2545–2563, August 2013. doi: 10.1093/mnras/stt928.
- [126] L. Miller, C. Heymans, T. D. Kitching, L. van Waerbeke, T. Erben, H. Hildebrandt, H. Hoekstra, Y. Mellier, B. T. P. Rowe, J. Coupon, J. P. Dietrich, L. Fu, J. Harnois-Déraps, M. J. Hudson, M. Kilbinger, K. Kuijken, T. Schrabback, E. Semboloni, S. Vafaei, and M. Velander. Bayesian galaxy shape measurement for weak lensing surveys - III. Application to the Canada-France-Hawaii Telescope Lensing Survey. *Monthly Notices of the Royal Astronomical Society*, 429:2858–2880, March 2013. doi: 10.1093/mnras/sts454.
- [127] M. Velander, E. van Uitert, H. Hoekstra, J. Coupon, T. Erben, C. Heymans, H. Hildebrandt, T. D. Kitching, Y. Mellier, L. Miller, L. Van Waerbeke, C. Bonnett, L. Fu, S. Giodini, M. J. Hudson, K. Kuijken, B. Rowe, T. Schrabback, and E. Semboloni. CFHTLenS: the relation between galaxy dark matter haloes and baryons from weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 437:2111–2136, January 2014. doi: 10.1093/mnras/stt2013.

- [128] A. Leauthaud, J. Tinker, K. Bundy, P. S. Behroozi, R. Massey, J. Rhodes, M. R. George, J.-P. Kneib, A. Benson, R. H. Wechsler, M. T. Busha, P. Capak, M. Cort  s, O. Ilbert, A. M. Koekemoer, O. Le F  vre, S. Lilly, H. J. McCracken, M. Salvato, T. Schrabback, N. Scoville, T. Smith, and J. E. Taylor. New Constraints on the Evolution of the Stellar-to-dark Matter Connection: A Combined Analysis of Galaxy-Galaxy Lensing, Clustering, and Stellar Mass Functions from $z = 0.2$ to $z = 1$. *The Astrophysical Journal*, 744:159, January 2012. doi: 10.1088/0004-637X/744/2/159.
- [129] M. Milkeraitis, L. van Waerbeke, C. Heymans, H. Hildebrandt, J. P. Dietrich, and T. Erben. 3D-Matched-Filter galaxy cluster finder - I. Selection functions and CFHTLS Deep clusters. *Monthly Notices of the Royal Astronomical Society*, 406:673–688, July 2010. doi: 10.1111/j.1365-2966.2010.16720.x.
- [130] J. Ford, H. Hildebrandt, L. Van Waerbeke, T. Erben, C. Laigle, M. Milkeraitis, and C. B. Morrison. Cluster magnification and the mass-richness relation in CFHTLenS. *Monthly Notices of the Royal Astronomical Society*, 439:3755–3764, April 2014. doi: 10.1093/mnras/stu225.
- [131] J. Ford, L. Van Waerbeke, M. Milkeraitis, C. Laigle, H. Hildebrandt, T. Erben, C. Heymans, H. Hoekstra, T. Kitching, Y. Mellier, L. Miller, A. Choi, J. Coupon, L. Fu, M. J. Hudson, K. Kuijken, N. Robertson, B. Rowe, T. Schrabback, and M. Velander. CFHTLenS: a weak lensing shear analysis of the 3D-Matched-Filter galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 447:1304–1318, February 2015. doi: 10.1093/mnras/stu2545.
- [132] N. Ben  tez. Bayesian Photometric Redshift Estimation. *The Astrophysical Journal*, 536: 571–583, June 2000. doi: 10.1086/308947.
- [133] C. O. Wright and T. G. Brainerd. Gravitational Lensing by NFW Halos. *The Astrophysical Journal*, 534:34–40, May 2000. doi: 10.1086/308744.
- [134] A. Leauthaud, J. Tinker, P. S. Behroozi, M. T. Busha, and R. H. Wechsler. A Theoretical Framework for Combining Techniques that Probe the Link Between Galaxies and Dark Matter. *The Astrophysical Journal*, 738:45, September 2011. doi: 10.1088/0004-637X/738/1/45.
- [135] S. More, F. C. van den Bosch, M. Cacciato, R. Skibba, H. J. Mo, and X. Yang. Satellite kinematics - III. Halo masses of central galaxies in SDSS. *Monthly Notices of the Royal Astronomical Society*, 410:210–226, January 2011. doi: 10.1111/j.1365-2966.2010.17436.x.
- [136] J. L. Tinker, A. Leauthaud, K. Bundy, M. R. George, P. Behroozi, R. Massey, J. Rhodes, and R. H. Wechsler. Evolution of the Stellar-to-dark Matter Relation: Separating Star-forming and Passive Galaxies from $z = 1$ to 0. *The Astrophysical Journal*, 778:93, December 2013. doi: 10.1088/0004-637X/778/2/93.
- [137] J. Coupon, S. Arnouts, L. van Waerbeke, T. Moutard, O. Ilbert, E. van Uitert, T. Erben, B. Garilli, L. Guzzo, C. Heymans, H. Hildebrandt, H. Hoekstra, M. Kilbinger, T. Kitching,

- Y. Mellier, L. Miller, M. Scodeggio, C. Bonnett, E. Branchini, I. Davidzon, G. De Lucia, A. Fritz, L. Fu, P. Hudelot, M. J. Hudson, K. Kuijken, A. Leauthaud, O. Le Fèvre, H. J. McCracken, L. Moscardini, B. T. P. Rowe, T. Schrabback, E. Semboloni, and M. Velander. The galaxy-halo connection from a joint lensing, clustering and abundance analysis in the CFHTLenS/VIPERS field. *Monthly Notices of the Royal Astronomical Society*, 449:1352–1379, May 2015. doi: 10.1093/mnras/stv276.
- [138] A. Choi, C. Heymans, C. Blake, H. Hildebrandt, C. A. J. Duncan, T. Erben, R. Nakajima, L. Van Waerbeke, and M. Viola. CFHTLenS and RCSLenS: Testing Photometric Redshift Distributions Using Angular Cross-Correlations with Spectroscopic Galaxy Surveys. *ArXiv e-prints*, December 2015.
- [139] M. Bernardi, F. Shankar, J. B. Hyde, S. Mei, F. Marulli, and R. K. Sheth. Galaxy luminosities, stellar masses, sizes, velocity dispersions as a function of morphological type. *Monthly Notices of the Royal Astronomical Society*, 404:2087–2122, June 2010. doi: 10.1111/j.1365-2966.2010.16425.x.
- [140] A. P. Hearin and D. F. Watson. The dark side of galaxy colour. *Monthly Notices of the Royal Astronomical Society*, 435:1313–1324, October 2013. doi: 10.1093/mnras/stt1374.
- [141] P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu. The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. *The Astrophysical Journal*, 762:109, January 2013. doi: 10.1088/0004-637X/762/2/109.
- [142] P. S. Behroozi, R. H. Wechsler, H.-Y. Wu, M. T. Busha, A. A. Klypin, and J. R. Primack. Gravitationally Consistent Halo Catalogs and Merger Trees for Precision Cosmology. *The Astrophysical Journal*, 763:18, January 2013. doi: 10.1088/0004-637X/763/1/18.
- [143] K. Riebe, A. M. Partl, H. Enke, J. Forero-Romero, S. Gottlöber, A. Klypin, G. Lemson, F. Prada, J. R. Primack, M. Steinmetz, and V. Turchaninov. The MultiDark Database: Release of the Bolshoi and MultiDark cosmological simulations. *Astronomische Nachrichten*, 334:691–708, August 2013. doi: 10.1002/asna.201211900.
- [144] P. Schneider. Weak Gravitational Lensing. *ArXiv Astrophysics e-prints*, September 2005.
- [145] C. Heymans, L. Van Waerbeke, D. Bacon, J. Berge, G. Bernstein, E. Bertin, S. Bridle, M. L. Brown, D. Clowe, H. Dahle, T. Erben, M. Gray, M. Hetterscheidt, H. Hoekstra, P. Hudelot, M. Jarvis, K. Kuijken, V. Margoniner, R. Massey, Y. Mellier, R. Nakajima, A. Refregier, J. Rhodes, T. Schrabback, and D. Wittman. The Shear Testing Programme - I. Weak lensing analysis of simulated ground-based observations. *Monthly Notices of the Royal Astronomical Society*, 368:1323–1339, May 2006. doi: 10.1111/j.1365-2966.2006.10198.x.
- [146] R. Mandelbaum, U. Seljak, C. M. Hirata, S. Bardelli, M. Bolzonella, A. Bongiorno, M. Carollo, T. Contini, C. E. Cunha, B. Garilli, A. Iovino, P. Kampczyk, J.-P. Kneib, C. Knobel, D. C. Koo, F. Lamareille, O. Le Fèvre, J.-F. Le Borgne, S. J. Lilly, C. Maier, V. Mainieri, M. Mignoli, J. A. Newman, P. A. Oesch, E. Perez-Montero, E. Ricciardelli, M. Scodeggio,

- J. Silverman, and L. Tasca. Precision photometric redshift calibration for galaxy-galaxy weak lensing. *Monthly Notices of the Royal Astronomical Society*, 386:781–806, May 2008. doi: 10.1111/j.1365-2966.2008.12947.x.
- [147] P. Norberg, C. M. Baugh, E. Gaztañaga, and D. J. Croton. Statistical analysis of galaxy surveys - I. Robust error estimation for two-point clustering statistics. *Monthly Notices of the Royal Astronomical Society*, 396:19–38, June 2009. doi: 10.1111/j.1365-2966.2009.14389.x.
- [148] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C++ : the art of scientific computing*. 2002.
- [149] F. C. van den Bosch, D. Aquino, X. Yang, H. J. Mo, A. Pasquali, D. H. McIntosh, S. M. Weinmann, and X. Kang. The importance of satellite quenching for the build-up of the red sequence of present-day galaxies. *Monthly Notices of the Royal Astronomical Society*, 387: 79–91, June 2008. doi: 10.1111/j.1365-2966.2008.13230.x.
- [150] S. D. Landy and A. S. Szalay. Bias and variance of angular correlation functions. *The Astrophysical Journal*, 412:64–71, July 1993. doi: 10.1086/172900.
- [151] R. Scranton, D. Johnston, S. Dodelson, J. A. Frieman, A. Connolly, D. J. Eisenstein, J. E. Gunn, L. Hui, B. Jain, S. Kent, J. Loveday, V. Narayanan, R. C. Nichol, L. O’Connell, R. Scoccimarro, R. K. Sheth, A. Stebbins, M. A. Strauss, A. S. Szalay, I. Szapudi, M. Tegmark, M. Vogele, I. Zehavi, J. Annis, N. A. Bahcall, J. Brinkman, I. Csabai, R. Hindsley, Z. Ivezic, R. S. J. Kim, G. R. Knapp, D. Q. Lamb, B. C. Lee, R. H. Lupton, T. McKay, J. Munn, J. Peoples, J. Pier, G. T. Richards, C. Rockosi, D. Schlegel, D. P. Schneider, C. Stoughton, D. L. Tucker, B. Yanny, and D. G. York. Analysis of Systematic Effects and Statistical Uncertainties in Angular Clustering of Galaxies from Early Sloan Digital Sky Survey Data. *The Astrophysical Journal*, 579:48–75, November 2002. doi: 10.1086/342786.
- [152] I. Zehavi, M. R. Blanton, J. A. Frieman, D. H. Weinberg, H. J. Mo, M. A. Strauss, S. F. Anderson, J. Annis, N. A. Bahcall, M. Bernardi, J. W. Briggs, J. Brinkmann, S. Burles, L. Carey, F. J. Castander, A. J. Connolly, I. Csabai, J. J. Dalcanton, S. Dodelson, M. Doi, D. Eisenstein, M. L. Evans, D. P. Finkbeiner, S. Friedman, M. Fukugita, J. E. Gunn, G. S. Hennessy, R. B. Hindsley, Ž. Ivezić, S. Kent, G. R. Knapp, R. Kron, P. Kunszt, D. Q. Lamb, R. F. Leger, D. C. Long, J. Loveday, R. H. Lupton, T. McKay, A. Meiksin, A. Merrelli, J. A. Munn, V. Narayanan, M. Newcomb, R. C. Nichol, R. Owen, J. Peoples, A. Pope, C. M. Rockosi, D. Schlegel, D. P. Schneider, R. Scoccimarro, R. K. Sheth, W. Siegmund, S. Smee, Y. Snir, A. Stebbins, C. Stoughton, M. SubbaRao, A. S. Szalay, I. Szapudi, M. Tegmark, D. L. Tucker, A. Uomoto, D. Vanden Berk, M. S. Vogele, P. Waddell, B. Yanny, and D. G. York. Galaxy Clustering in Early Sloan Digital Sky Survey Redshift Data. *The Astrophysical Journal*, 571:172–190, May 2002. doi: 10.1086/339893.
- [153] S. Bhattacharya, K. Heitmann, M. White, Z. Lukić, C. Wagner, and S. Habib. Mass Function Predictions Beyond Λ CDM. *The Astrophysical Journal*, 732:122, May 2011. doi: 10.1088/0004-637X/732/2/122.

- [154] N. Kaiser. On the spatial correlations of Abell clusters. *The Astrophysical Journal Lett.*, 284:L9–L12, September 1984. doi: 10.1086/184341.
- [155] A. D. Ludlow and C. Porciani. The peaks formalism and the formation of cold dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 413:1961–1972, May 2011. doi: 10.1111/j.1365-2966.2011.18282.x.
- [156] R. H. Wechsler, J. S. Bullock, J. R. Primack, A. V. Kravtsov, and A. Dekel. Concentrations of Dark Halos from Their Assembly Histories. *The Astrophysical Journal*, 568:52–70, March 2002. doi: 10.1086/338765.
- [157] Y.-Y. Mao, A. R. Zentner, and R. H. Wechsler. Beyond assembly bias: exploring secondary halo biases for cluster-size haloes. *Monthly Notices of the Royal Astronomical Society*, 474: 5143–5157, March 2018. doi: 10.1093/mnras/stx3111.
- [158] J. E. Gunn and J. R. Gott, III. On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution. *The Astrophysical Journal*, 176:1, August 1972. doi: 10.1086/151605.
- [159] B. E. Robertson, A. V. Kravtsov, J. Tinker, and A. R. Zentner. Collapse Barriers and Halo Abundance: Testing the Excursion Set Ansatz. *The Astrophysical Journal*, 696:636–652, May 2009. doi: 10.1088/0004-637X/696/1/636.
- [160] A. Klypin, G. Yepes, S. Gottlöber, F. Prada, and S. Heß. MultiDark simulations: the story of dark matter halo concentrations and density profiles. *Monthly Notices of the Royal Astronomical Society*, 457:4340–4359, April 2016. doi: 10.1093/mnras/stw248.
- [161] J. D. Cohn and M. White. Dark matter halo abundances, clustering and assembly histories at high redshift. *Monthly Notices of the Royal Astronomical Society*, 385:2025–2033, April 2008. doi: 10.1111/j.1365-2966.2008.12972.x.
- [162] S. Adhikari, N. Dalal, and R. T. Chamberlain. Splashback in accreting dark matter halos. *Journal of Cosmology and Astroparticle Physics*, 11:019, November 2014. doi: 10.1088/1475-7516/2014/11/019.
- [163] B. Diemer, S. More, and A. V. Kravtsov. The Pseudo-evolution of Halo Mass. *The Astrophysical Journal*, 766:25, March 2013. doi: 10.1088/0004-637X/766/1/25.
- [164] P. Mansfield, A. V. Kravtsov, and B. Diemer. Splashback Shells of Cold Dark Matter Halos. *The Astrophysical Journal*, 841:34, May 2017. doi: 10.3847/1538-4357/aa7047.
- [165] B. Diemer, P. Mansfield, A. V. Kravtsov, and S. More. The Splashback Radius of Halos from Particle Dynamics. II. Dependence on Mass, Accretion Rate, Redshift, and Cosmology. *The Astrophysical Journal*, 843:140, July 2017. doi: 10.3847/1538-4357/aa79ab.

Appendix A

Chapter 3: Assembly bias in CFHTLenS

A.1 Effect of applying various satellite exclusion on the 2PCFs of the mocks

Here we present the 2PCFs with various levels of satellite removal for the mock galaxy catalog. Particularly we present the cases by varying $r_{\text{excl.}}$ (left panels), A (middle panels) or $M_{\star, \text{nbr}}/M_{\star, \text{lens}}$ (right panels) in Fig. A.1.

With different removal schemes, the proportion of satellites removed $f_{\text{sat, excl.}}$ varies between 66.8% to 79.6%, and the residual satellite contamination f_{sat} ranges from 9.3% to 12.8%. Nonetheless, in all cases the 2PCFs of the “All galaxies” (dots) and the “Centrals only” (triangles) of the same galaxy sample are consistent. It demonstrates that with different ways of removing satellites, the difference in the clustering between the red and blue samples are still present even all satellites have been removed (“Centrals only”). It indicates that the assembly bias signals found in simulation cannot be solely attributed to the presence of the residual satellite contamination using our satellite removal criteria.

A.2 Effect of applying various satellite exclusion on CFHTLenS results

Here we present the results of pairing up lensing profiles (Fig. A.2) of the red and blue samples, the corresponding 2PCFs (Fig. A.3) and the difference of 2PCFs (Fig. A.4) for the CFHTLenS

galaxy catalog. We present the cases by varying A , $r_{\text{excl.}}$ or $M_{\star,\text{nbr}}/M_{\star,\text{lens}}$. The reader is advised to compare with Figures 3.4 and 3.5 for the case with the most stringent exclusion scheme adopted in this paper (i.e. $A = 1.5$, $r_{\text{excl.}} = 500h_{70}^{-1}\text{kpc}$ and $M_{\star,\text{nbr}}/M_{\star,\text{lens}} \geq 3$).

Under different satellite removal schemes, the percentage of galaxies being excluded $f_{\text{gal,excl.}}$ varies between 78.9% to 87.8%. Nevertheless, the significance of difference of the 2PCFs in all cases does not change appreciably at $\lesssim 1.8\sigma$ level (see Fig. A.4). The statistics of comparing the 2PCFs is summarized in Table A.2.

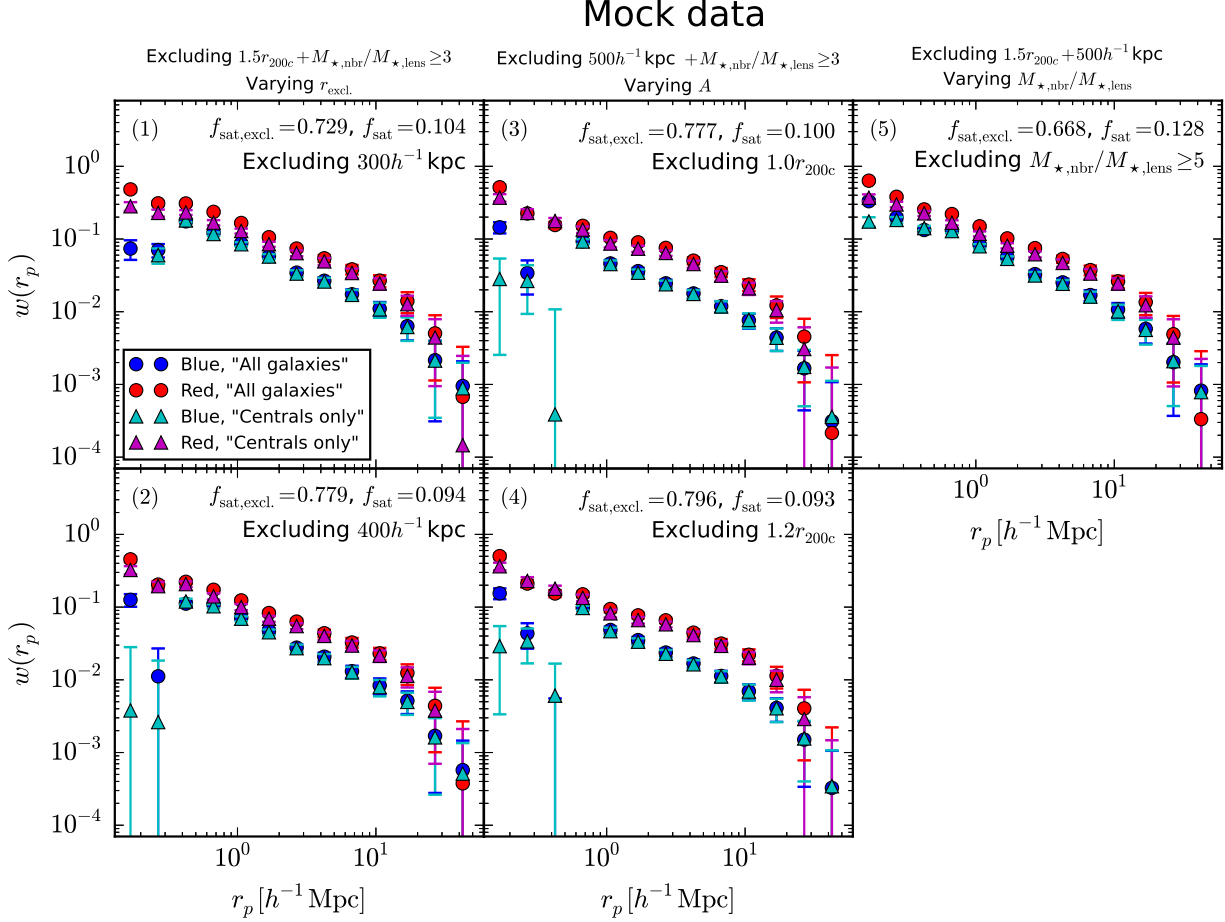


Figure A.1: The 2PCFs of the mock galaxies with different satellite removal criteria. The red (blue) dots represent the red (blue) “All galaxies” sample respectively. The magenta (cyan) triangles represent the red (blue) “Centrals only” sample respectively. The stellar mass bins ($\log_{10}(M_{\star}/M_{\odot})$) chosen for the mock data have similar widths as those chosen in CFHTLenS data. In all cases, the stellar mass bins of both “All galaxies” and “Centrals only” red samples are kept at $[9.8, 10.8]$. (Left panels: 1 & 2): Keeping $A = 1.5$ and $M_{\star,\text{nbr}}/M_{\star,\text{lens}} \geq 3$ fixed, with $r_{\text{excl.}}$ varied between $300 - 400h^{-1}$ kpc. In (1), the stellar mass ranges of the blue “All galaxies” and “Centrals only” samples are $[9.935, 10.935]$ and $[9.945, 10.945]$ respectively. In (2), the corresponding ranges are $[9.94, 10.94]$ and $[9.928, 10.928]$ respectively. (Middle panels: 3 & 4): Keeping $r_{\text{excl.}} = 500h^{-1}$ kpc and $M_{\star,\text{nbr}}/M_{\star,\text{lens}} \geq 3$ fixed, with A varied between 1.0 and 1.2 . In (3), the stellar mass ranges of the blue “All galaxies” and “Centrals only” samples are $[9.928, 10.928]$ and $[9.928, 10.928]$ respectively. In (4), the corresponding ranges are $[9.928, 10.928]$ and $[9.925, 10.925]$ respectively. (Right panel: 5): Keeping $A = 1.5$ and $r_{\text{excl.}} = 500h^{-1}$ kpc fixed, with $M_{\star,\text{nbr}}/M_{\star,\text{lens}} \geq 5$, the stellar mass ranges of the blue “All galaxies” and “Centrals only” samples are $[9.945, 10.945]$ and $[9.938, 10.938]$ respectively.

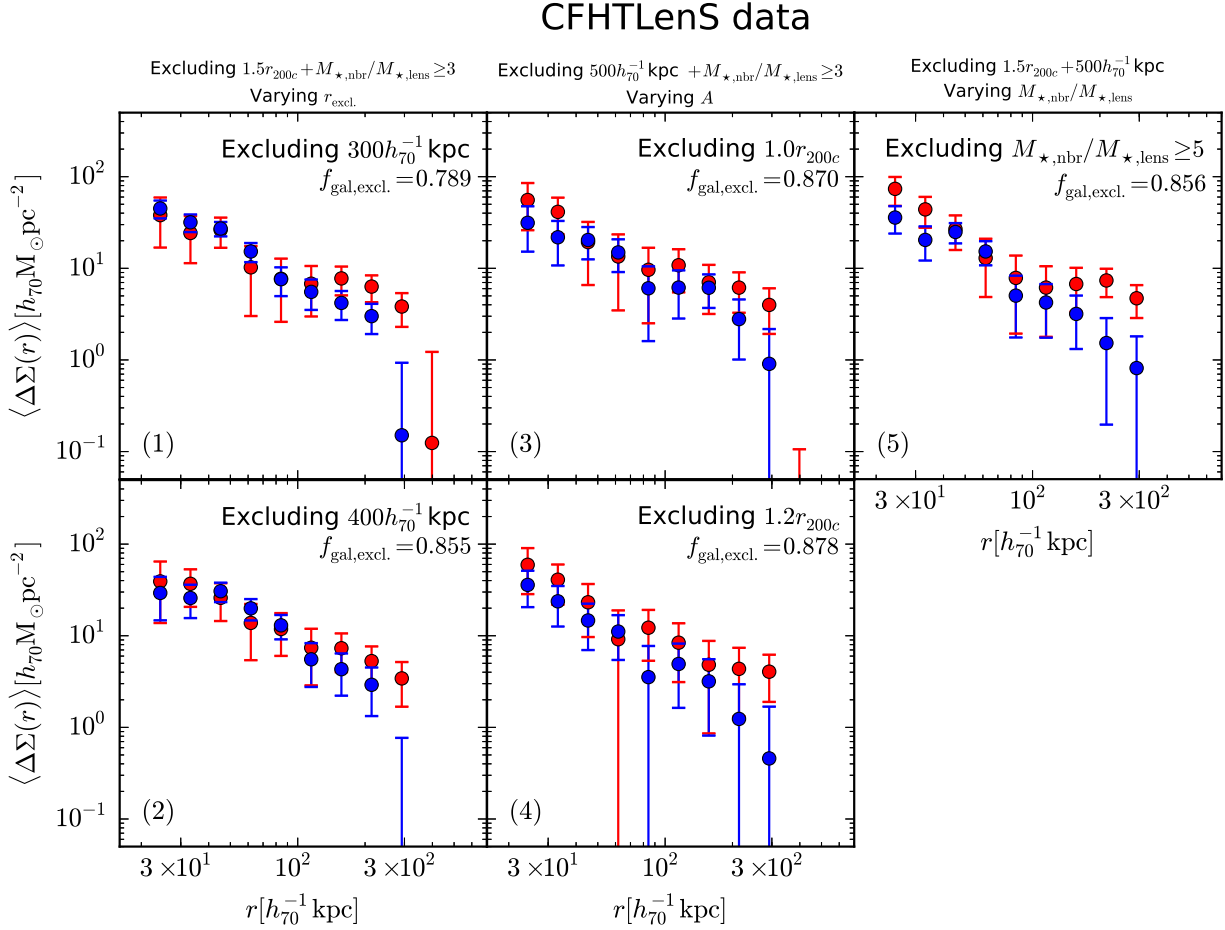


Figure A.2: Pairing up the tangential shear profiles of the CFHTLenS galaxies with different satellite removal criteria by minimizing $\Delta\chi^2_S$. (Left panels: 1 & 2): Keeping $A = 1.5$ and $M_{\star, \text{nbr}}/M_{\star, \text{lens}} \geq 3$ fixed, with $r_{\text{excl.}}$ varied between $300 - 400h_{70}^{-1} \text{kpc}$. (Middle panels: 3 & 4): Keeping $r_{\text{excl.}} = 500h_{70}^{-1} \text{kpc}$ and $M_{\star, \text{nbr}}/M_{\star, \text{lens}} \geq 3$ fixed, with A varied between 1.0 and 1.2 . (Right panel: 5): Keeping $A = 1.5$ and $r_{\text{excl.}} = 500h_{70}^{-1} \text{kpc}$ fixed, with $M_{\star, \text{nbr}}/M_{\star, \text{lens}} \geq 5$. The red and blue data points represent the red and blue samples respectively. About $80\% - 90\%$ of the galaxies are excluded in different removal schemes. The results of pairing up lensing profiles are shown in Table A.1.

$A^{(1)}$	$r_{\text{excl.}}^{(2)}$	$M_{\star, \text{back}}/M_{\star, \text{lens}}^{(3)}$	$f_{\text{gal, excl.}}^{(4)}$	$M_{\star}^{\text{red}(5)}$	$N_{\text{lens}}^{\text{red}(6)}$	$M_{\star}^{\text{blue}(7)}$	$N_{\text{lens}}^{\text{blue}(8)}$	$\Delta\chi_S^{2,(9)}$
1.0	500	≥ 3	0.870	[9.05, 10.05]	5073	[9.45, 10.5]	10325	1.78
1.2	500	≥ 3	0.878	[9.05, 10.05]	4819	[9.42, 10.47]	10572	2.62
1.5	500	≥ 5	0.856	[9.05, 10.05]	6763	[9.35, 10.35]	17372	3.51
1.5	300	≥ 3	0.789	[9.05, 10.05]	9627	[9.34, 10.34]	27906	2.25
1.5	400	≥ 3	0.855	[9.05, 10.05]	6570	[9.5, 10.55]	12636	1.55
1.5	500	≥ 3	0.892	[9.05, 10.05]	4376	[9.5, 10.55]	8206	3.19

Table A.1: Results from pairing up red and blue samples with $\Delta\chi_S^2$ minimization. (1): Values of A in which galaxies within $d = A \times r_{200c}$ from the cluster center are identified as satellites. (2 & 3): Values of projected separation $r_{\text{excl.}}$ (in $h_{70}^{-1}\text{kpc}$) at which a lens galaxy is rejected if there is any neighbor whose stellar mass is at least 3 times (or 5 times) more massive than its own. (4): Fraction of galaxies being rejected. (5): $M_{\star}$ bin that the red lens samples are selected (in $\log_{10}(M_{\star}/M_{\odot})$). (6): Number of red lenses. (7): $M_{\star}$ bin that the blue lens samples are selected (in $\log_{10}(M_{\star}/M_{\odot})$). (8): Number of blue lenses. (9): $\Delta\chi_S^2$ of the sample pairs, see Eqn. 3.11. The last row presents the results in the main text (See Fig. 3.4).

$A^{(1)}$	$r_{\text{excl.}}^{(2)}$	$M_{\star, \text{back}}/M_{\star, \text{lens}}^{(3)}$	$f_{\text{gal, excl.}}^{(4)}$	$\Delta\chi_w^{2,(5)}$	$p^{(6)}$	$\sigma^{(7)}$
1.0	500	≥ 3	0.870	7.31	0.121	1.55
1.2	500	≥ 3	0.878	8.56	0.073	1.79
1.5	500	≥ 5	0.856	8.01	0.091	1.69
1.5	300	≥ 3	0.789	8.05	0.090	1.70
1.5	400	≥ 3	0.855	5.76	0.218	1.23
1.5	500	≥ 3	0.892	7.92	0.095	1.67

Table A.2: Results of the statistics of difference in 2PCFs between the red and blue samples after pairing up their lensing profiles. (1): Values of A in which galaxies within $d = A \times r_{200c}$ from the cluster center are identified as satellites. (2 & 3): Values of projected separation $r_{\text{excl.}}$ (in $h_{70}^{-1}\text{kpc}$) at which a lens galaxy is rejected if there is any neighbor whose stellar mass is at least 3 times (or 5 times) more massive than its own. (4): Fraction of galaxies being rejected. (5): $\Delta\chi_w^2$ of the 2PCFs of the red and blue samples, see Eqn. 3.15. (6): p -value corresponding to $\Delta\chi_w^2$. (7): Significance of difference of the 2PCFs (in sigma). The last row presents the results in the main text (See Fig. 3.5).

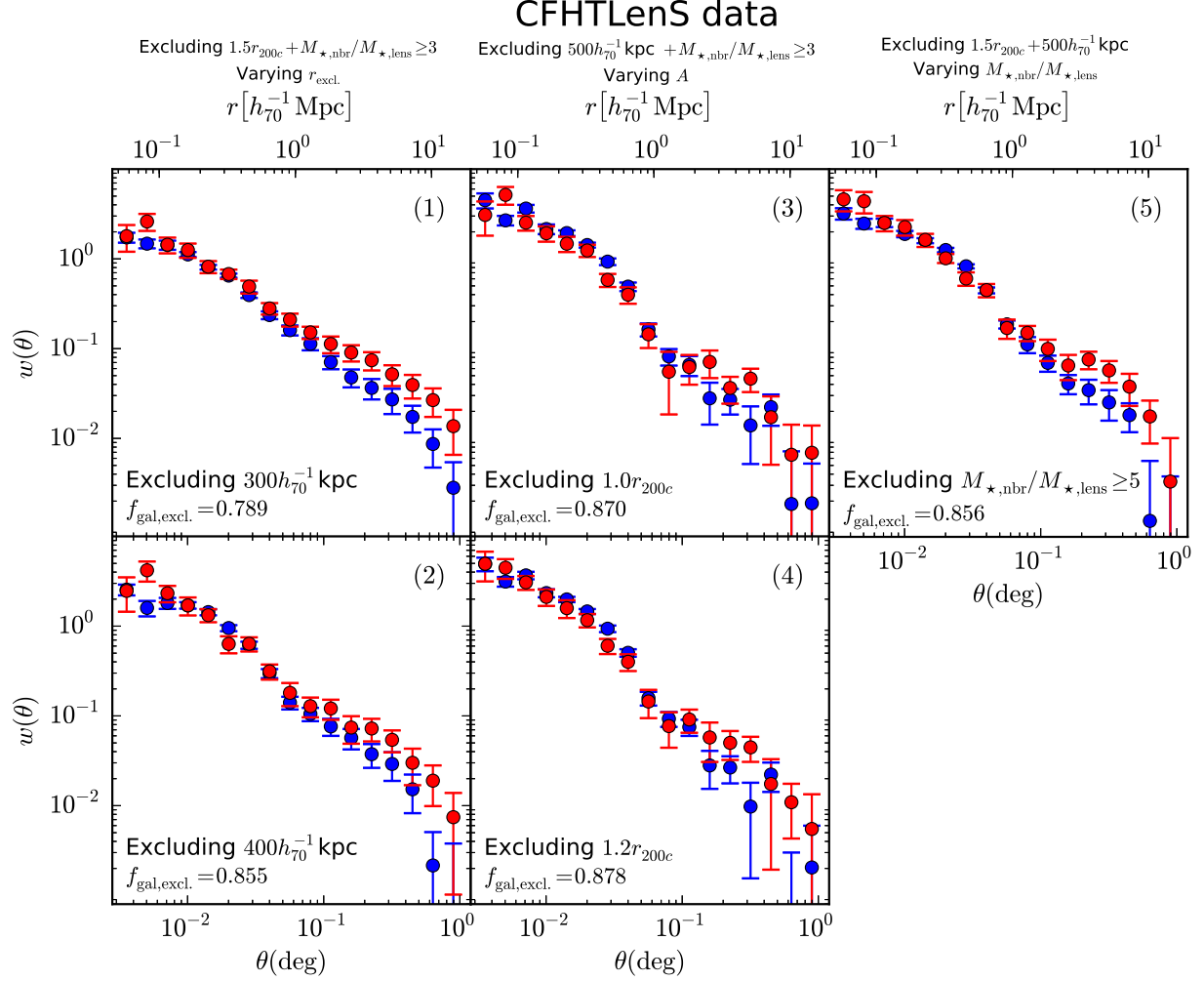


Figure A.3: The 2PCFs of the CFHTLenS galaxies with different satellite removal criteria after having the lensing profiles paired up. (Left panels: 1 & 2): Keeping $A = 1.5$ and $M_{\star,\text{nbr}}/M_{\star,\text{lens}} \geq 3$ fixed, with $r_{\text{excl.}}$ varied between $300 - 400h_{70}^{-1}\text{kpc}$. (Middle panels: 3 & 4): Keeping $r_{\text{excl.}} = 500h_{70}^{-1}\text{kpc}$ and $M_{\star,\text{nbr}}/M_{\star,\text{lens}} \geq 3$ fixed, with A varied between 1.0 and 1.2. (Right panel: 5): Keeping $A = 1.5$ and $r_{\text{excl.}} = 500h_{70}^{-1}\text{kpc}$ fixed, with $M_{\star,\text{nbr}}/M_{\star,\text{lens}} \geq 5$. The red and blue data points represent the red and blue samples respectively. The physical distance at $z = 0.3$ is shown at the top abscissae of the panels. The statistics of comparing the 2PCFs is summarized in Table A.2.

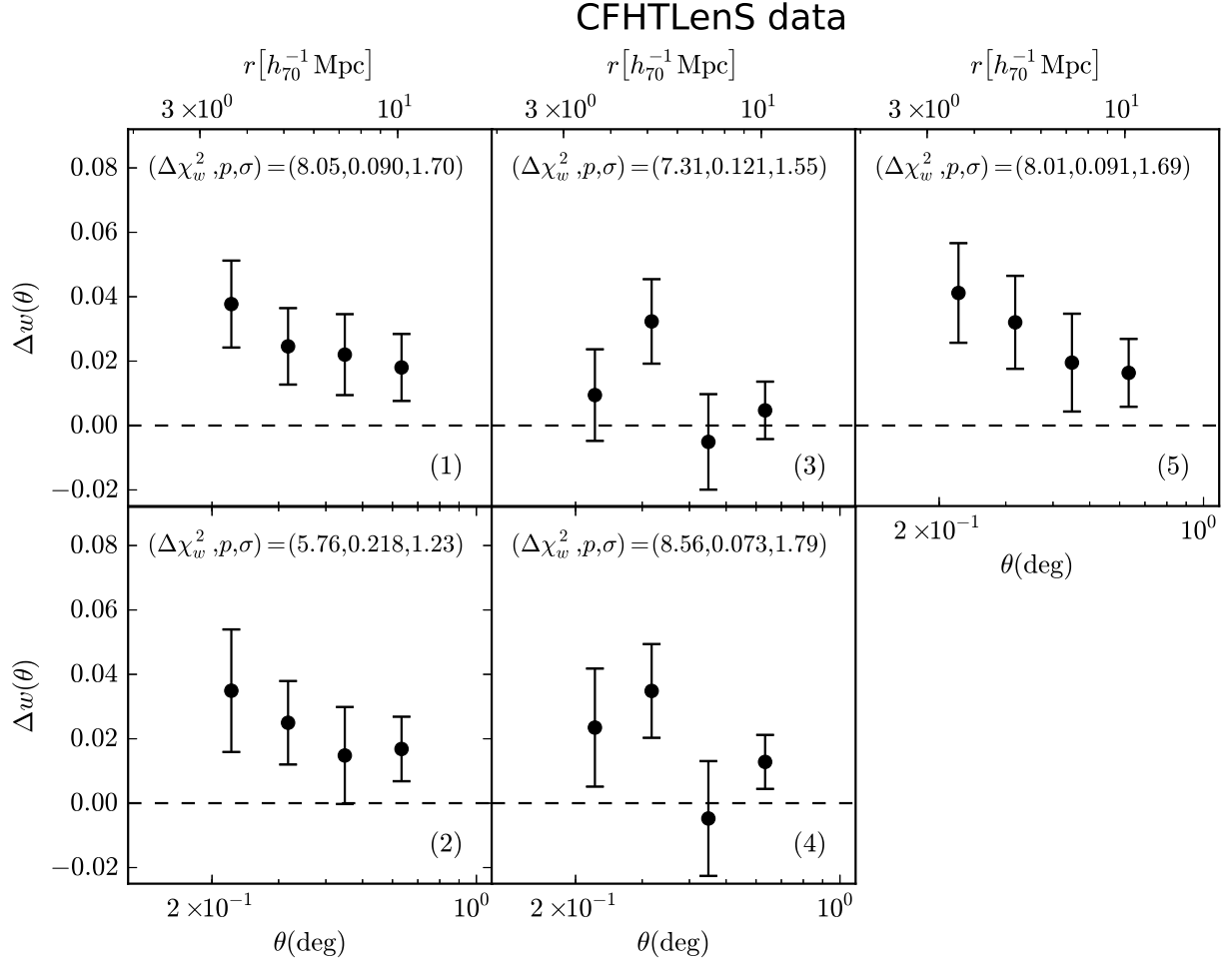


Figure A.4: The differences of the 2PCFs of the CFHTLenS galaxies at $3 - 11 h_{70}^{-1} \text{ Mpc}$ with different satellite removal criteria after having the lensing profiles paired up. The panel numbers correspond to the panels in Fig. A.3. The physical distance at $z = 0.3$ is shown at the top abscissae of the panels. The dashed lines show the $\Delta w(\theta) = 0$ level. The statistics of comparing the 2PCFs difference is summarized in Table A.2.