EXPLORING THE LIGHT AND DARKNESS: HIERARCHICAL BAYESIAN ANALYSES ON EARLY-TYPE GALAXIES AND DARK MATTER

by

Yiping Shu

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Physics

Department of Physics and Astronomy

The University of Utah

August 2014

THE UNIVERSITY OF UTAH GRADUATE SCHOOL

STATEMENT OF DISSERTATION APPROVAL

This dissertation of
Yiping Shu

has been approved by the following supervisory committee members:

Adam S. Bolton	, Chair	April 29, 2014
		Date Approved
Benjamin C. Bromley	, Member	April 29, 2014
		Date Approved
Stephane LeBohec	, Member	April 29, 2014
		Date Approved
Jordan Gerton	, Member	April 29, 2014
		Date Approved
Lajos Horvath	, Member	April 29, 2014
		Date Approved

and by Carleton DeTar, Chair of

the Department of
Physics and Astronomy

and by David B. Kieda of the Graduate School.

ABSTRACT

Being the end-product of the hierarchical merging scenario, *early-type* (elliptical and lenticular) galaxies (ETGs) are the "live" fossil records that permit compelling tests of galaxy formation theories within a cosmological context. Also, ETGs can be extremely luminous and serve as the ideal cosmological tracers in the Universe. Additionally, the mysterious *dark matter* (DM), which is believed to constitute almost 85% of the observed mass portion, acts as the host of galaxies and hence plays a pivotal role in shaping the observed Universe. A thorough understanding of DM, including its nature, properties, and structures provides crucial insights into the fundamental laws of physics and cosmology.

On the luminous side, a hierarchical Bayesian determination of the velocity-dispersion function of approximately 430000 massive luminous red galaxies observed by the Baryon Oscillation Spectroscopic Survey (BOSS) is performed. We use the full velocity-dispersion likelihood function for each galaxy to make a self-consistent determination of the velocity-dispersion distribution parameters as a function of absolute magnitude and redshift. Parameterizing the distribution at each point in the luminosity-redshift plane with a log-normal form, we detect significant evolution in the width of the distribution toward higher intrinsic scatter at higher redshifts, which indicates a more diverse heterogeneity in ETGs at earlier cosmic time.

On the dark side, I report the discovery of 40 strong gravitational lenses in the SLACS for the Masses (S4TM) Survey and 33 additional systems with single-lensed images in S4TM and SLACS Surveys, for which upper limits of the Einstein radii are determined. A hierarchical Bayesian analysis reveals strong evidence (4σ) of variations of the total mass-density structure toward shallower profiles at larger velocity dispersion when upper limits are incorporated. Estimating the stellar masses based on the *HST* I-band photometry, we find a significant trend of higher dark-matter fraction at higher velocity dispersion. A Salpeter initial mass function is substantially disfavored for all but the most massive lens galaxies by predicting stellar masses in excess of the total lensing-measured mass. An approach of constraining mass structure via a joint analysis of lensing and stellar kinematics is also outlined, the application of which on a sample of strong lenses shows a 4σ evolution trend in the sense of steeper mass profiles at later cosmic times.

To my beloved fiancée – Minne Wang, and my parents.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Pursuing a Ph.D. degree is a long and hard process, which not only requires hard work from myself, but also needs all sorts of support and help from other people. At the end of my Ph.D. period, it brings me great joy to say "thank you" to everyone without whom I would not have been here. First of all, I want to thank my advisor Professor **Adam Bolton** and the remaining committee members Professor **Benjamin Bromley**, Professor **Stephane LeBohec**, Professor **Jordan Gerton**, and Professor **Lajos Horvath** for guiding me through this whole process. I would also like to thank all the people that came to my thesis defense to support me.

Adam, if I were restricted to use only one sentence to summarize my feelings about you, I would go for this old Chinese proverb — *"Who teaches me for a day is my father for a lifetime."* I am truly fortunate and grateful to have you be my advisor for the past four years and I really appreciate all the support and guidance you have given to me. I still remember the first time we met in your office on the second floor of INSCC in 2010. You showed me the beautiful mosaic of the SLACS lenses, something that I didn't foresee would compose such a big part of my research. My astronomy career then started from there. Four years' working with you has been a pleasant and memorable experience in my life. You have impacted me in every aspect of my life. I have learned how to do research, how to write academic papers, how to communicate with people, and most importantly, how to stay strong when failure happens. You have also set up the standards of being a perfect advisor for me, something that I will try to follow if I am lucky enough to be someone else's advisor.

Haojie, I am glad to be your roommate. Thank you for keeping the apartment nice and clean. I appreciate all your suggestions whenever I encountered a problem. I wish you all the best in the future. **Hong**, you are very caring and always ready to give advice. I am glad that you came to Utah as a postdoc so that we had a chance to know each other. **Antonio, Mark, and Raphael**, do not ever forget all the workouts, basketball games, dancing parties, and BBQs. They were great fun. Dr. **Joel Brownstein**, I enjoyed all the discussions with you, whether on gravitational lensing or plants. And I hope you were not too bothered by my snores in Sunspot, New Mexico, when we were sharing a room for

a SDSS meeting at the APO. **Matt**, thanks for the comments on my defense, the help on my thesis, and telling me all the secrets that I need to know about the graduation staff in general. You are just like a big brother. Professor **Kyle Dawson**, you have continuously helped me to become a better researcher ever since we first met in 2009. Professor **Zheng Zheng**, I enjoyed the Thanksgiving dinner with your family. And also thank you very much for arranging the Beijing visit for me. Professor **Anil Seth**, thank you for coming to my celebration party. Professor **Inese Ivans**, the chocolate you gave to me was delicious.

There is also a group of staff members in the Department of Physics and Astronomy who have made remarkable efforts to ensure my stay in the department and the U. was enjoyable and trouble-free. In particular, I want to thank Grad Secretary **Jackie Hadley** for patiently answering all my dumb questions, Payroll Accountant **Vicki Nielsen** for managing my paychecks month by month, **Sareah Gardner** for taking care of all my travel plans and reimbursements, and Editorial Assistant **Mary Ann Woolf** for assigning TA/RA jobs every semester.

Other than the people in the U.S., I have also received endless support from my parents **Xinhua Shu** and **Chunxian Pan** in China. They perfectly elaborate the idea that "father's love is like a mountain, while mother's love is like a river." They always care about my life and research, cheer for every single piece of achievement I have accomplished, and encourage me whenever I feel down. I am sorry that I have not been able to celebrate the Lunar New Year, the most important date for a Chinese family, with them for 6 consecutive years. I love you, my parents. I also want to thank my fiancée's parents **Zhiyuan Wang** and **Meiying Xu** for handing me the best gift in my life — their lovely daughter.

Now it comes to my sweetheart, my fiancée **Minne**. I love you so much. You are simply my other half. You have made my life colorful and full of love ever since we met. Keeping a long-distance relationship is definitely not easy. I see everything that you have done and I understand how much love and courage that requires. I am grateful to your understanding and consolation when I had a really hard time to find a postdoctoral position, and to you sacrificing your career future in China for a chance of being closer to me. I love you!

Yiping Shu
May 14, 2014

CHAPTER 1

INTRODUCTION

1.1 Brief History of the Universe

The *Universe* is the totality of time, space, and any matter/energy within it. The currently observable universe is about 46 billion light years in radius (Gott et al., 2005), and the age of the universe is 13.813 ± 0.058 billion years according to the latest Planck results (Planck Collaboration et al., 2013).

The currently well-accepted model for explaining the early development of the universe — the Big Bang theory (Einstein, 1916; Friedmann, 1924; Hubble, 1929; Lemaître, 1931, also see Figure 1.1) believes that our universe originated from a quantum singularity with infinite density and temperature and underwent a short period ($10^{-33} - 10^{-32}$ seconds) of exponential inflation due to quantum fluctuations. As the universe expanded, photons, together with the baryons they were coupled to, started to cool down. At some point when the universe was cool enough, protons and electrons combined into neutral atoms (mostly hydrogens) and photons decoupled from baryons and traveled freely ever since. This epoch happened at about 379,000 years ($z \approx 1100$) after the Big Bang and is referred to as "Recombination" by cosmologists. The "relic radiation" was studied by a number of scientists (McKellar, 1941; Dicke et al., 1946; Gamow, 1948; Alpher and Herman, 1948) and eventually detected "accidentally" by Penzias and Wilson (1965), which earned them the 1978 Nobel Prize in Physics. Since the universe had been growing so dramatically during the inflationary epoch, this relic radiation known as the Cosmic Microwave Background (CMB) can be well-characterized by a perfect black-body spectrum at a temperature of 2.72548 ± 0.00057 K (Fixsen, 2009).

Although the universe could be considered as homogeneous and isotropic after inflation, small quantum fluctuations at the beginning of the Big Bang persisted and led to small density fluctuations after recombination. Slightly denser regions attracted nearby matter and thus became even denser due to gravitational instability. After a period of "dark Ages" in which only 21-cm hydrogen line was emitted, first stars and quasars started to form due to gravitational collapse of the accumulated molecular clouds that exceeded the Jeans mass

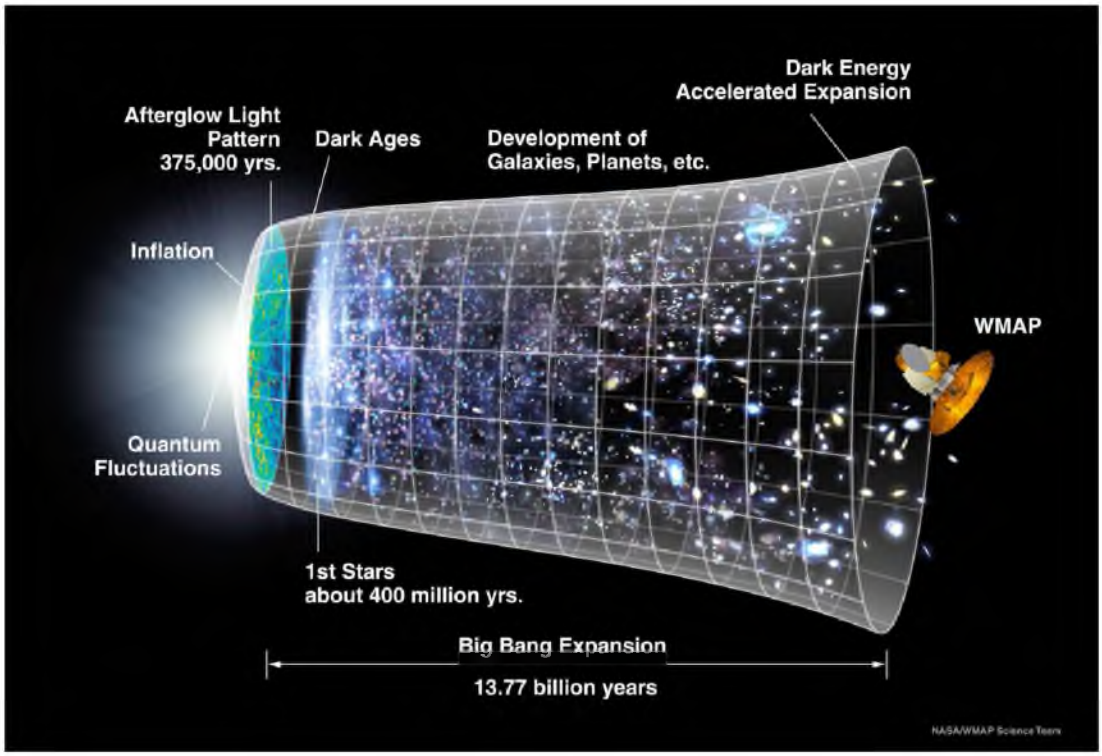

Figure 1.1: The chronology of our universe as seen in the Big Bang theory. Please refer to text for details. Credit: NASA/WMAP Science Team, (via http://wmap.gsfc.nasa.gov/media/060915/060915_CMB_Timeline150.jpg)

(Jeans, 1902) at about 400 million years after the Big Bang. The intense radiation from the first stars and quasars reionized the neutral universe to be a plasma in the "Reionization" epoch. Later on, a large volume of matter collapsed to hierarchically form stars, galaxies, and all the other objects we now observe.

The ΛCDM model is so far the most successful model in the Big Bang Theory, in which the universe is composed by a cosmological constant term denoted as Λ (also known as the "dark energy") and Cold Dark Matter (CDM). According to the latest study of the CMB by the Planck collaboration, the fraction of dark energy density to the critical density $\rho_{\rm crit}$ for which the spatial geometry is flat, Ω_Λ, is 0.685 ± 0.017 (Planck Collaboration et al., 2013). The rest 0.315 ± 0.017 resides in the "matter sector" that includes both ordinary matter (0.049) and CDM (0.266), which is the focus of this dissertation work.

1.1.1 Luminous Component

Although only constituting 15.5% of the total mass sector, the ordinary matter builds up literally all the fascinating objects in the universe that we are able to observe in various wavelength ranges, including planets, stars, star clusters, interstellar medium, quasars, galaxies, intergalactic medium, galaxy clusters, etc. Figure 1.2 shows the very impressive Hubble-Ultra-Deep-Field (HUDF) observation of a small sky region in the constellation Fornax that had been observed repeatedly for several months by the Hubble Space Telescope (*HST*), which contains about 10,000 of the most distant/oldest galaxies that have ever been imaged in the optical band with various brightnesses, colors, shapes, and ages (a few hundred million years after the Big Bang).

All the observed structures, especially galaxies, are built up in a so-called "bottom-up" fashion. At first, a giant molecular cloud starts to collapse after exceeding the Jeans mass for which the internal gas pressure can no longer balance the gravitational attraction. As it becomes dense and hot enough, nuclear fusion is ignited, and the radiation pressure prevents the gas from further collapsing. A star is then formed. The process happens everywhere in the universe and because of gravitational instability, stars accumulate to form star clusters and small galaxies. They interact with each other to form more massive galaxies by merging. Galaxies become further bounded to galaxy groups, clusters, and superclusters.

Galaxies can be classified according to their morphologies. Figure 1.3 shows the classification scheme invented by Hubble (1936a), also known as the "Hubble tuning fork" diagram. The "handle" is comprised of smooth, featureless galaxies with generally elliptical shapes, known as "elliptical" galaxies (denoted by E). The integer following E represents the degree of the observed ellipticity. On the right are "spiral" galaxies with disk-like structures

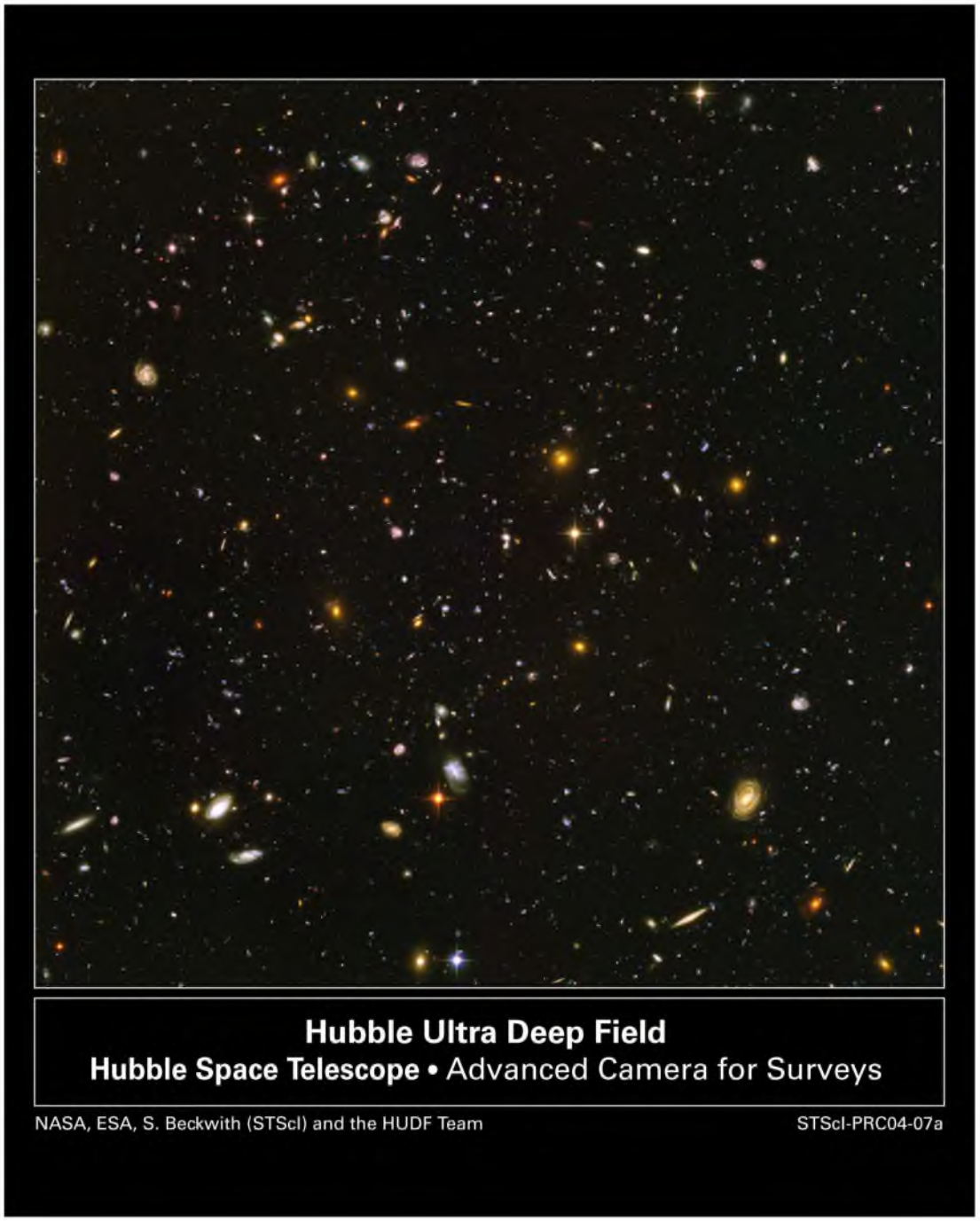

Figure 1.2: Hubble-Ultra-Deep-Field of about 10000 most distant galaxies observed by the HST from September 2003 to January 2004. Credit: NASA, the European Space Agency, S. Beckwith (STScI), and the HUDF Team, (via http://imgsrc.hubblesite.org/hu/db/images/hs-2004-07-a-print.jpg)

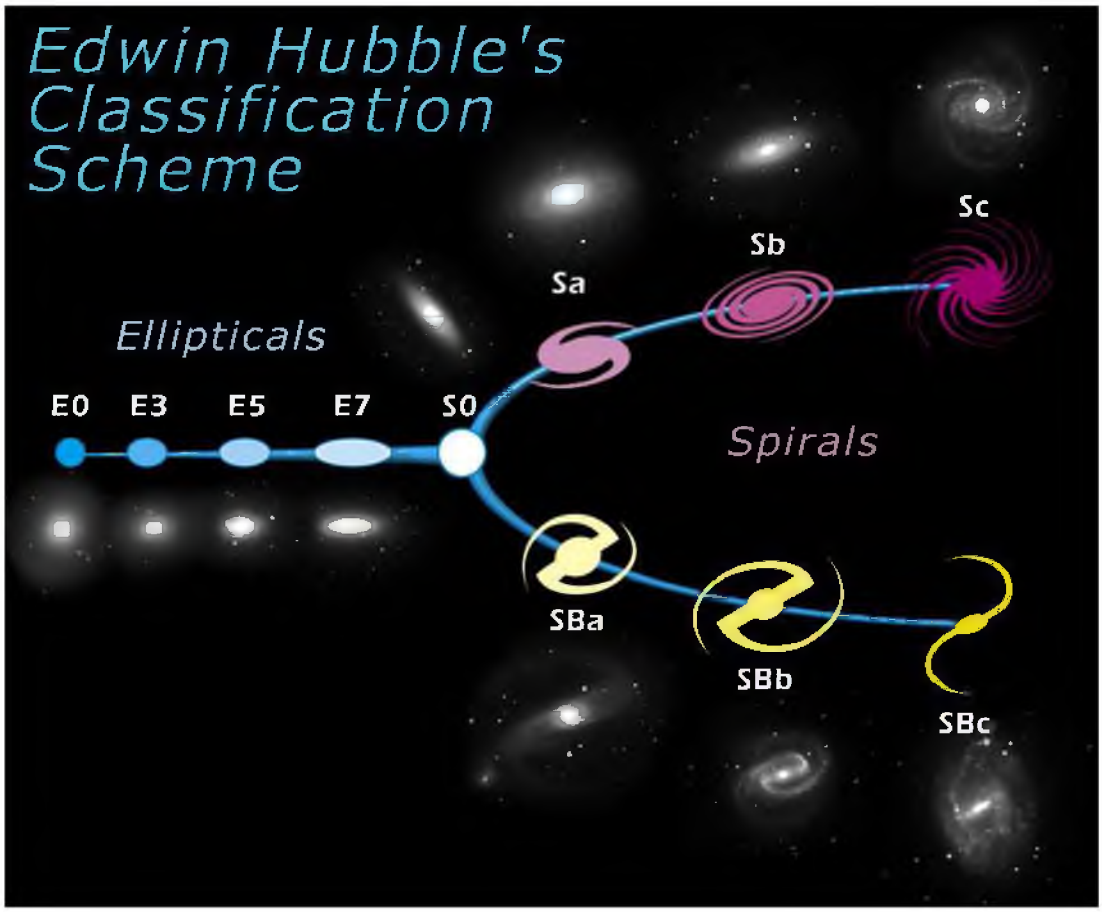

Figure 1.3: Hubble tuning fork diagram of the Hubble sequence. Credit: Public Domain, (via Wikimedia Commons, http://commons.wikimedia.org/wiki/File:HubbleTuningFork.jpg, User: Cosmo0)

and evident spiral arms. The two "tines" are two subcategories of spiral galaxies, one with bar-like structures and one without. Sitting in between are the "lenticular" galaxies (designated S0) that have bright prominent central bulges and visible disk components but no visible spiral arms. Although not shown in the Hubble tuning fork diagram, there are also irregular galaxies which have no obvious regular shapes. Elliptical and lenticular galaxies can be furthered grouped into "early-type" galaxies (ETGs). Further studies have shown that ETGs are passively evolving (little ongoing star formation, mergers, and interactions) and share many common properties, among which are spectral features and scaling relations. Spiral and irregular galaxies are referred to together as "late-type" galaxies.

Named as "early" type, massive ETGs are believed to be the end-product of the hierarchical merging scenario (Toomre and Toomre, 1972; White and Frenk, 1991; Kauffmann et al., 1993a; Baugh et al., 1996; Cole et al., 2000). Although featureless in shape, ETGs are of particular interest because they are the "live" fossil records that allow us to test theories of galaxy formation within a cosmological context by studying their structures, properties, and formation histories. Also, massive ETGs are the most luminous and highly clustered objects that serve as the ideal cosmological tracers of clustering and the large-scale structure (Eisenstein et al., 2005a; Percival et al., 2007; Anderson et al., 2012).

There are several well-established empirical scaling relations among the kinematic and photometric properties observed in ETGs. The *Faber-Jackson* relation (FJR) was first derived by Faber and Jackson (1976) using 25 E and S0 galaxies, the central line-of-sight velocity dispersions of which were determined by a subjective visual comparison method and a Fourier transform method independently. They adopted the means of the visual method as the measured velocity dispersions mostly because the visual method was found to be less sensitive to noise. A strong correlation between the so-measured velocity dispersion and the B-band absolute magnitude of the "normal" elliptical galaxies (as compared to the smaller and fainter "dwarf" ellipticals) was detected as displayed in Figure 1.4. They obtained a simple relation of

$$L_B \propto \sigma^4 \tag{1.1}$$

The same analysis was done with larger samples and led to similar relations in both elliptical and lenticular galaxies (e.g., Tonry and Davis, 1981; de Vaucouleurs and Olson, 1982; Jorgensen et al., 1996).

In his thesis work in 1977, John Kormendy saw an anticorrelation between the V-band surface brightness (in magnitude) B_e at the effective radius and the effective radius r_e derived from de Vaucouleurs' fits of a small sample of "normal" ellipticals (Kormendy, 1977),

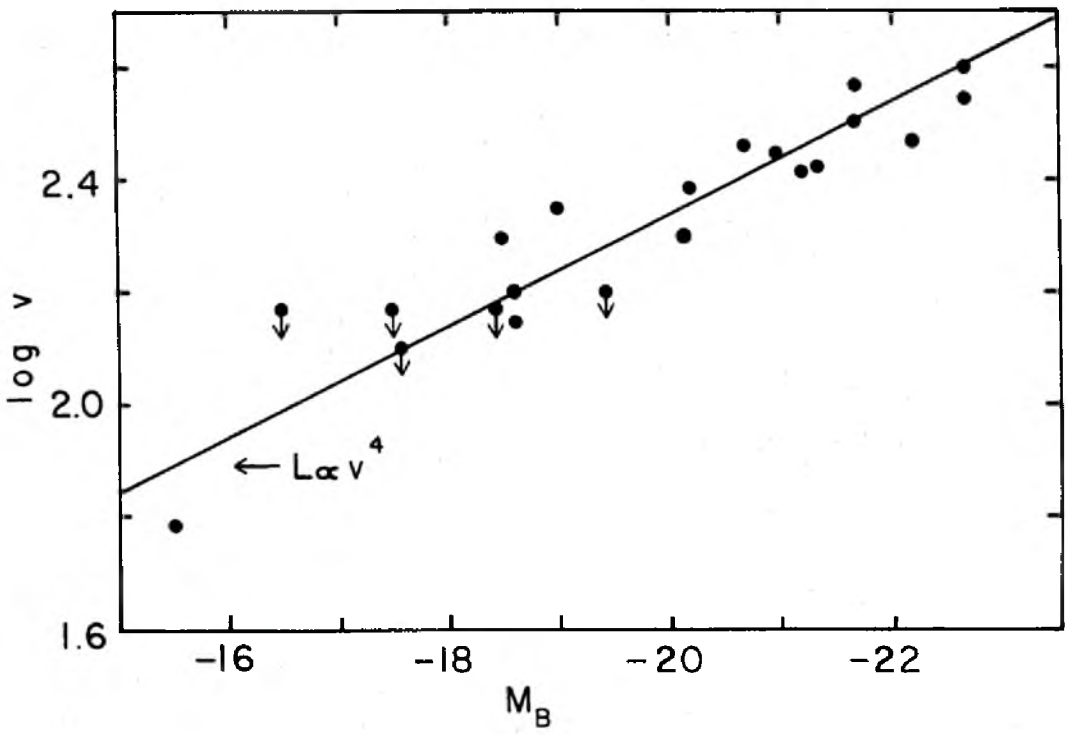

Figure 1.4: Correlation between the line-of-sight velocity dispersion and absolute magnitude for elliptical galaxies detected by Faber and Jackson (1976). The lowest velocity dispersion value (60 km s^{-1}, corresponding to M32) was taken from Richstone and Sargent (1972). Credit: Figure 16 in the referred publication: "Velocity dispersions and mass-to-light ratios for elliptical galaxies," Faber, S. M. and Jackson, R. E., 1976, ApJ, 204, 668. Reproduced by permission of the AAS.

later known as the *Kormendy Relation* (KR). A de Vaucouleurs' fit is a two-dimensional fit to the surface brightness of elliptical galaxies first introduced by de Vaucouleurs (1948). A de Vaucouleurs' profile is parameterized as

$$I(R) = I_e \; e^{-7.669[(R/R_e)^{1/4}-1]} \tag{1.2}$$

where R_e is called the *effective* radius or equivalently the *half-light* radius within which the enclosed light is equal to one half of the total light, and I_e is the surface brightness at $R = R_e$. Figure 1.5 is a snapshot of the original version of the KR. A linear least squares fit yielded

$$B_e = 3.02 \log_{10} r_e + 19.74 \tag{1.3}$$

Considering the fact that the surface brightness in physical units I_e is related to B_e and luminosity L as

$$\begin{aligned} B_e &\equiv -2.5 \log_{10} I_e \\ L &\propto I_e \times R_e^2 \end{aligned} \tag{1.4}$$

more luminous elliptical galaxies are more diffuse (with larger R_e).

Dressler et al. (1987) defined a new photometric size parameter D_n as the diameter within which the mean surface brightness drops to some fiducial value (chosen to be $\Sigma = 20.75$ mag/arcsec^2 in that work) and found a tight correlation between D_n and the central velocity dispersion σ for a sample of elliptical galaxies in 6 rich clusters, Coma, Virgo, Fornax, Perseus, A2199, and DC2345-28, in the direction that galaxies with larger velocity dispersions are more extended in terms of the light distribution (Figure 1.6). Shortly after that, Dressler (1987) showed that this correlation extended to lenticular galaxies (S0) as well (Figure 1.7).

Furthermore, if distributed in the three-dimensional space spanned by $\log_{10} \sigma$, $\log_{10} R_e$, and $\log_{10}\langle I \rangle_e$, elliptical galaxies populate a well-defined thin plane (Figure 1.8) known as the *Fundamental Plane* (FP, Dressler et al., 1987; Djorgovski and Davis, 1987). The parameterized form of the plane can be written as

$$\log_{10} R_e = a \log_{10} \sigma + b \log_{10}\langle I \rangle_e + c \tag{1.5}$$

in which $\log_{10}\langle I \rangle_e$ is defined as the mean surface brightness within R_e. Using the fact that the integrated light within R_e is one half of the total light, we have

$$\langle I \rangle_e = \frac{L/2}{\pi R_e^2} \tag{1.6}$$

Converting luminosity L to mass M by assuming a mass-to-light ratio Υ,

$$\log_{10} R_e = 0.5 \log_{10} M - 0.5 \log_{10} \Upsilon - 0.5 \log_{10}\langle I \rangle_e + \text{constant} \tag{1.7}$$

If the galaxy is virialized (which is in general valid), the dynamical mass is related to the velocity dispersion as

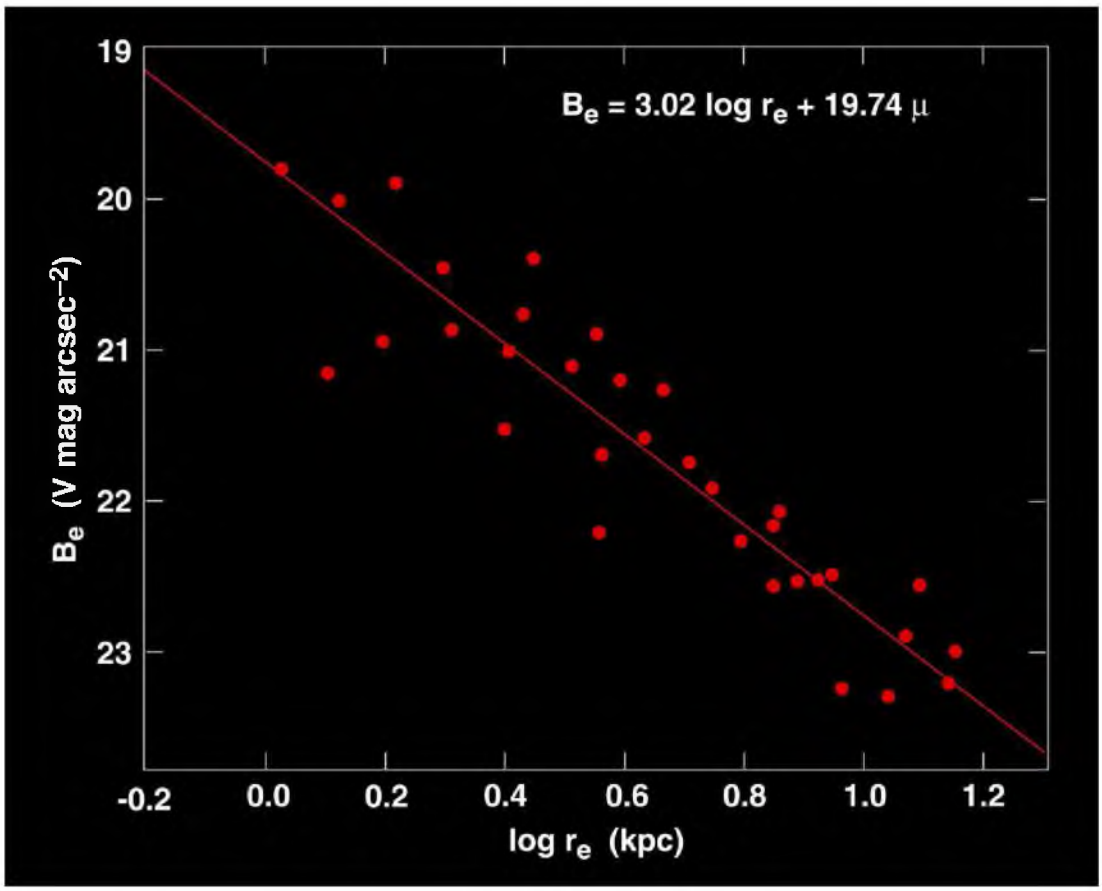

Figure 1.5: Correlation between the surface brightness and size of "normal" ellipticals detected by Kormendy (1977). Credit: J. Kormendy, (via http://chandra.as.utexas.edu/~kormendy/Be-logre-big.gif)

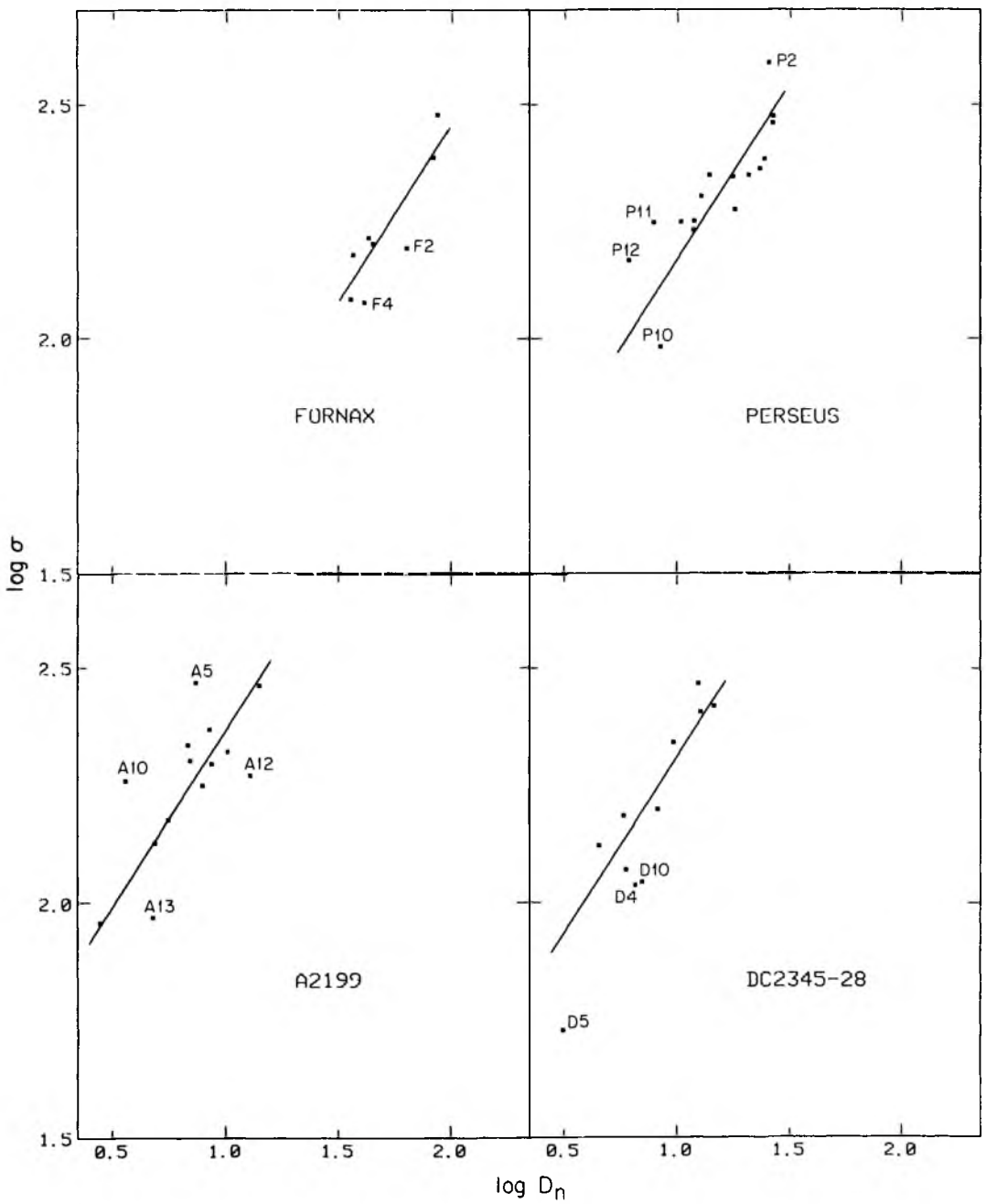

Figure 1.6: Correlations between D_n and central velocity dispersion of ellipticals in clusters FORNAX, PERSEUS, A2199, and DC2345-28. Credit: Figure 2 in the referred publication: "Spectroscopy and photometry of elliptical galaxies. I - A new distance estimator," Dressler, Alan, et al., 1987, ApJ, 313, 42. Reproduced by permission of the AAS.

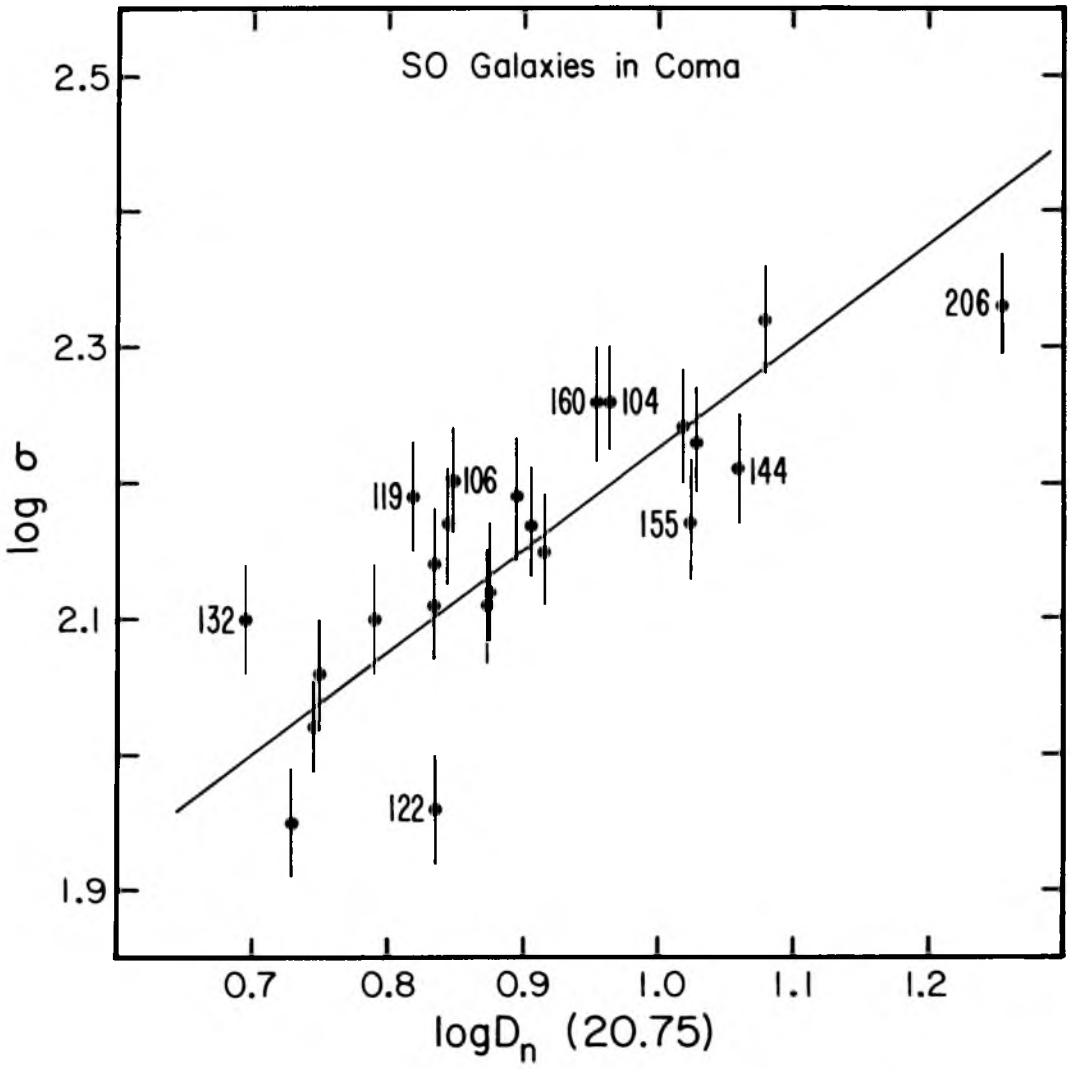

Figure 1.7: Correlation between D_n and central velocity dispersion of S0 galaxies in the Coma cluster. Credit: Figure 5 in the referred publication: "The Dn-sigma relation for bulges of disk galaxies - A new, independent measure of the Hubble constant," Dressler, Alan, 1987, ApJ, 317, 1. Reproduced by permission of the AAS.

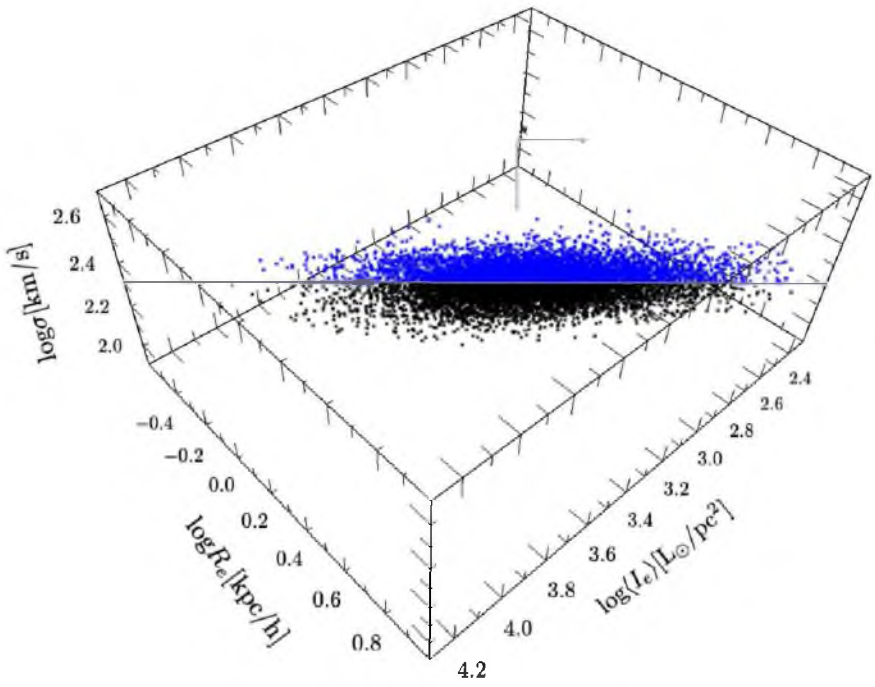

Figure 1.8: Near-infrared Fundamental Plane of ETGs in the $\log_{10}\sigma$, $\log_{10}R_e$, and $\log_{10}\langle I\rangle_e$ space observed by the 6dF Galaxy Survey. Blue and black symbols represent ETGs above and under the best-fitting plane. Credit: Figure 9 in the referred publication: "The 6dF Galaxy Survey: the near-infrared Fundamental Plane of early-type galaxies," Magoulas, Christina, et al., 2012, MNRAS, 427, 245.

$$M_{\rm dyn} \propto \frac{\sigma^2 R_e}{G} \tag{1.8}$$

Combining the above two equations, we have

$$\log_{10} R_e = 2\log_{10}\sigma - \log_{10}\langle I\rangle_e - \log_{10}\Upsilon + \text{constant} \tag{1.9}$$

If the mass-to-light ratio Υ is constant, we would expect $a = 2$ and $b = -1$. However, the observed values of coefficients a and b vary from sample to sample and from one photometric band to another, which is known as the "tilt" of the FP. For instance, Jorgensen et al. (1996) found $a = 1.24\pm0.07$ and $b = -0.82\pm0.02$ using Gunn r-band (Thuan and Gunn, 1976) data for 226 E and S0 galaxies. By simply restricting to $\log_{10}\sigma > 2.0$, they noticed a significant change in coefficient a to $a = 1.35 \pm 0.05$. Pahre et al. (1998) analyzed the near-infrared data of 251 ETGs and obtained $a = 1.53\pm0.08$ and $b = -0.79\pm0.03$. A much larger sample of nearly 9000 ETGs selected from the Sloan Digital Sky Survey (SDSS, York et al., 2000) yielded $a = 1.49 \pm 0.05$ and $b = -0.75 \pm 0.01$ in the Sloan r band (Bernardi et al., 2003c).

It is clear that the empirical scaling relations introduced above are the two-dimensional projections of the FP, which is quite useful for various purposes. The FJR can be used as a good distance-estimator as it estimates the true luminosity from a distance-independent quantity σ with little scatter, which can be compared with the observed apparent magnitude to estimate the distance to the galaxy (e.g., de Vaucouleurs and Olson, 1982; Dressler et al., 1987; Paturel and Garnier, 1992). The Kormendy relation has been employed to test the expansion of the universe and cosmology (e.g., Pahre et al., 1996; Moles et al., 1998). The D_n-σ relation has been shown to be capable of constraining the expansion rate or equivalently the Hubble constant (e.g., Dressler, 1987; Kelson et al., 1999).

Although it has been well-accepted that the origin and tilt of the FP is related to the galaxy formation/evolution processes in ETGs and dark matter that will be introduced in the next section, how exactly the FP is affected is still under investigation. For example, the deviations in the coefficients from the canonical Virial Theorem's predictions can be caused by a mass/luminosity-dependent mass-to-light ratio. In particular, a larger mass-to-light ratio for more massive and luminous ETGs can shift the coefficients in the right direction (Trujillo et al., 2004; Cappellari et al., 2006a; Bolton et al., 2007; D'Onofrio et al., 2008; Tortora et al., 2009; Cappellari et al., 2013). Also, the wavelength-dependence of the coefficients indicates some entanglement between galaxy mass, structure, and stellar population (e.g., Scodeggio et al., 1998; Oñorbe et al., 2005; La Barbera et al., 2010b; D'Onofrio et al., 2013; Dutton et al., 2013a). Furthermore, studies of the time/redshift evolution of the FP provide fruitful insights in understanding the history of the Universe (Gebhardt et al., 2003; Fritz et al., 2005; Fernández Lorenzo et al., 2011; Shu et al., 2012).

Other factors such as metallicity and environment also have nonnegligible effects on the tilt of the FP (Kochanek et al., 2000a; Boylan-Kolchin et al., 2005; La Barbera et al., 2010a).

1.1.2 Dark Component

Despite its large portion in the total mass budget, the existence of the mysterious dark matter (DM) was not known until the 1930s, when measurements of the orbital velocities of stars within galaxies led to a suggestion of "missing mass" (Oort, 1932; Zwicky, 1933). The most compelling evidence of the existence of dark matter then came in 1970 when Rubin and Ford (1970) made measurements of the rotational velocities spectroscopically in the Andromeda galaxy (a.k.a. M31) as a function of the distance from the center. The original data points are shown in Figure 1.9. Measurements in the central 16 arcminutes region (about 3.2 kpc) as represented by the open circles were done using the [N II] emission line and measurements in the outer region were done using [H II], [O III], and other emission lines. It is clear that the rotational velocity profile is unexpectedly flat in the outer region of M31, which, according to Newton's laws of motion (Newton, 1760), indicates that the total enclosed mass is approximately linear in radius. While the luminous mass drops significantly at large radii, there must be some extended *dark* component that provides necessary gravitational attraction to maintain this *flat-rotation* curve. Later on, other techniques such as gravitational lensing and CMB anisotropy studies have further supported the existence of dark matter (Bolton et al., 2008a; Coe et al., 2010; Dietrich et al., 2012; Krawczyk et al., 2013; Hinshaw et al., 2013; Planck Collaboration et al., 2013). Note that various modified gravity theories, as an alternative of DM theory, also provide reasonable explanation of the flat-rotation curve without the inclusion of DM. However, unless the modified gravity theories can do a better job in explaining observational facts such as the structure formation and CMB anisotropies (Slosar et al., 2005), they would not be able to challenge the currently widely-accepted DM theory.

So far, the consensus is that DM is made up of one or more species of exotic, elementary particles that, not like the usual protons, electrons, and neutrons, only interact gravitationally in order to match its dark feature. The most commonly proposed candidates are Weakly Interacting Massive Particles (WIMPs), axions, and sterile neutrinos. There are tons of detectors searching for evidence of DM particles either directly or indirectly, but no conclusive result has been found. The latest study of the CMB power spectrum by the Planck collaboration suggests that DM density $\rho_c = 0.266\ \rho_{\rm crit}$ assuming a flat ΛCDM cosmology (Planck Collaboration et al., 2013). Additionally, based on various observations, DM particles are thought to be "cold" and move at a speed much less than the speed of

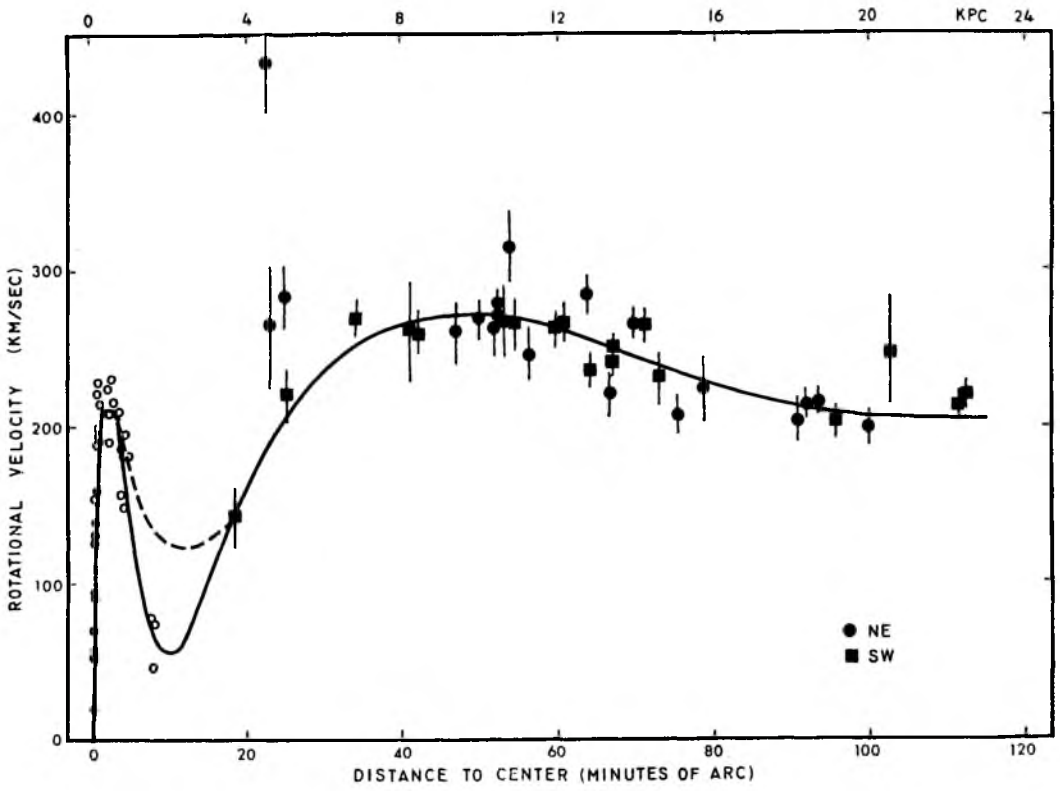

Figure 1.9: Original rotation curve for the Andromeda galaxy (a.k.a. M31) as measured by Rubin and Ford (1970). Velocities in the central 16 arcminutes region (about 3.2 kpc, open circles) were measured from the [N II] emission line and outer measurements were done using [H II], [O III] and other emission lines. Solid circles are data points from the Northeast arm and solid squares are from the Southwest arm. Credit: Figure 9 in the referred publication: "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions," Rubin, Vera, C. and Ford, W. Kent, Jr., 1970, ApJ, 159, 379. Reproduced by permission of the AAS.

light (e.g., Peebles, 1982; Blumenthal et al., 1984).

Several lines of evidence have suggested that DM plays a pivotal role in galaxy formation. Particularly, the flat-rotation curve indicates that DM distributes much more extendedly than the galaxy itself. Therefore, White and Rees (1978) proposed a "two-stage" theory to describe galaxy formation and clustering, according to which DM forms structures with various scales in the "bottom-up" fashion; then gas starts to cool and condense within the DM gravitational potential wells and ignites the star formation that eventually leads to luminous galaxies residing in the centers of DM halos. During these processes, the interaction between dark and luminous matter determines the properties of both DM and galaxies. The theory has been further improved since the 1980s and become the preferred solution for galaxy formation.

Nevertheless, the nature of DM is still an open question. N-body DM-only numerical simulations have suggested it has a somewhat "universal" density profile with a r^{-1} inner profile and a r^{-3} drop-off at large radii, independent of the halo mass (Navarro et al., 1996, 1997), known as the Navarro-Frenk-White (NFW) profile

$$\rho(r) = \frac{\rho_0}{(r/r_s)(1 + r/r_s)^2} \tag{1.10}$$

where r_s is a scale radius that differs from halo to halo. However, various observations of DM-dominated galaxies yield inconsistent (shallower) inner density slopes against high-resolution numerical simulations (Moore et al., 1999b; Diemand et al., 2005; Graham et al., 2006; Navarro et al., 2010), which is known as the "cuspy core problem." The solution of the cuspy core problem relies mostly on the baryonic physics that has not been captured in DM-only simulations such as gas cooling, gravitational heating, and energetic feedbacks (e.g, Governato et al., 2012; Weinberg et al., 2013; Velliscig et al., 2014). Figure 1.10 shows the effect of baryon outflows on the DM density profiles from a high-resolution N-Body+Smoothed-particle hydrodynamics (SPH) simulations ran by Governato et al. (2012). The dash-dot black line shows the DM density profile prediction by a DM-only N-body simulation. The three colored lines represent the redshift evolution of the DM density profile by including gas outflows, star formation, and other baryonic physics. Finally the solid black line is the prediction at $z = 0$. It is clear that the inclusion of baryonic physics is able to flatten the inner profile of a galaxy. However, how exactly all these effects compete and shape the galaxy mass structure is still unknown.

Another big failure of the ΛCDM model is the "missing satellite problem," which is the significant excess (by a order of magnitude) in the number of DM subhalos predicted by numerical cosmological simulations compared to that of the observed satellite galaxies in

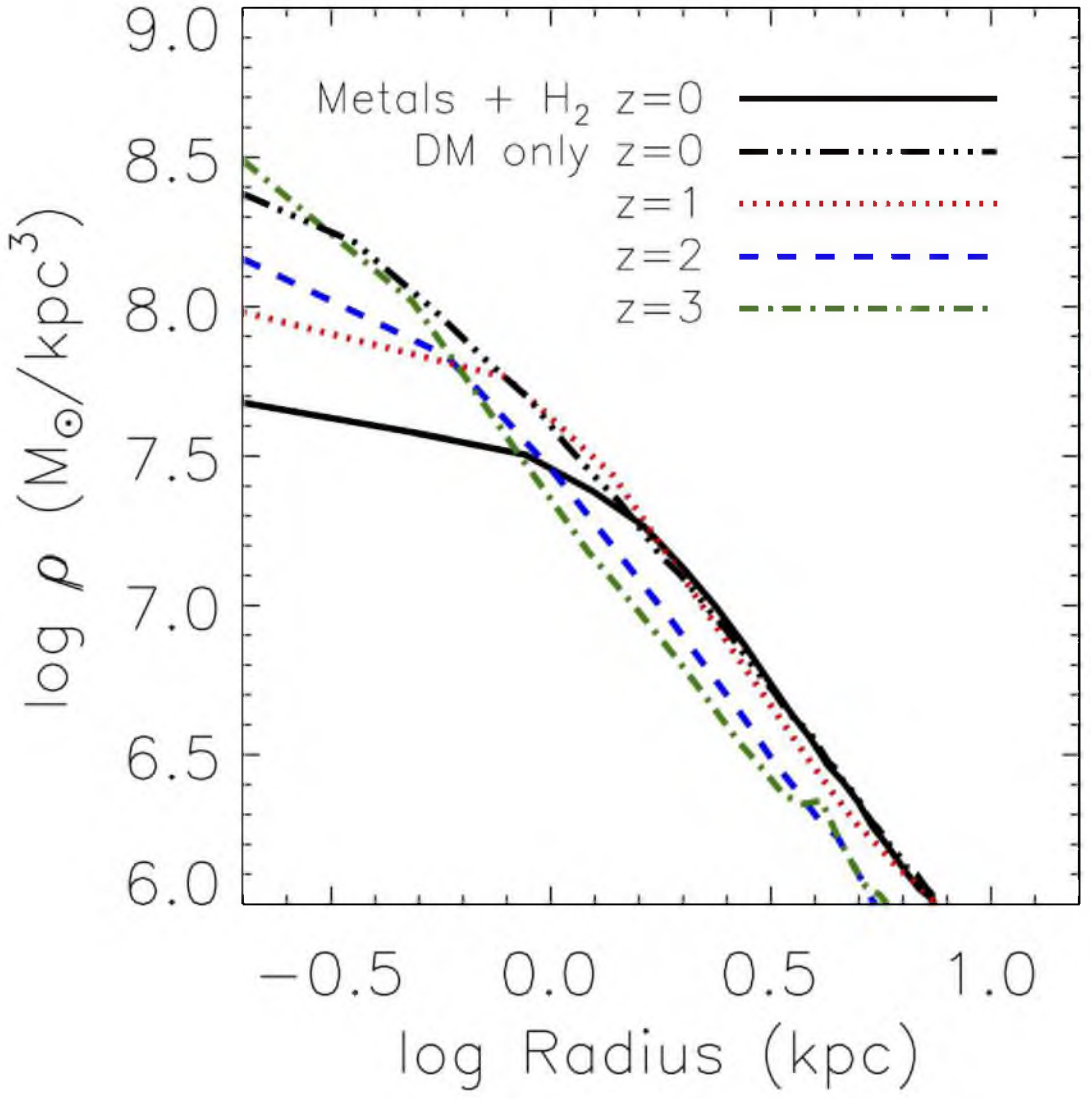

Figure 1.10: DM density profiles predicted by various simulations. Dash-dot black curve is the result of a DM-only simulation, while different colored lines show the evolution of the profile when baryonic physics is included. The solid black line is the prediction at $z = 0$ from a DM+SPH simulation. Credit: Figure 3 in the referred publication "Cuspy no more: how outflows affect the central dark matter and baryon distribution in cold dark matter galaxies," Governato, F., et al., 2012, MNRAS, 422, 1231.

the Local Group (Klypin et al., 1999; Moore et al., 1999a). Figure 1.11 shows a comparison of the mass functions between the observed results in the Milky Way (filled circles) and Virgo Cluster (open circles with error bars) and numerical predictions. Although the Virgo Cluster data agree well with numerical simulations, observations from our own Milky Way show a definite deficit in the abundance of substructures. The lack of observed satellites in galaxy scale indicates either a suppression in the galaxy formation process in subhalos, or more importantly, an insufficient understanding of the nature of DM particles.

1.2 Gravitational Lensing in a Nutshell

Gravity, as one of the four fundamental forces in the Universe, has been known for centuries for governing motions from falling apples to celestial objects. It pulls objects toward the center of the gravitational field as stated in one of Sir Isaac Newton's most famous works - *Philosophiae Naturalis Principia Mathematica* (Newton, 1760, the original book was published in 1687 in Latin). However, the motion of massless photons had not been well-understood until Albert Einstein's *General Theory of Relativity* (GR, Einstein, 1916), a novel idea of interpreting gravity as a pure geometric effect of space and time. Then Chwolson (1924) first discussed the gravitational lensing (GL) effect which, in analogy to an optical lens, is the bending of the photon trajectories under the influence of gravity of a massive object which is called the "lens." Einstein (1936) quantitatively studied the GL effect in the framework of GR, which I will briefly recap here.

1.2.1 Deflection of Photons in Gravitational Field

In GR, a freely moving photon follows the so-called "geodesic," a generalized "straight line" in curved spacetime, which is determined by the geodesic equation

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\lambda^2} + \Gamma^\mu_{\nu\rho}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\rho}{\mathrm{d}\lambda} = 0 \tag{1.11}$$

in which x^μ, x^ν, x^ρ are the spacetime coordinates with all Greek indices taking the value [0, 1, 2, 3]; λ is the "affine parameter" which uniquely corresponds to the position along the geodesic; and $\Gamma^\mu_{\nu\rho}$ is the "Christoffel symbol" purely determined by the spacetime metric $g_{\mu\nu}$. For photons, an extra constraint is that the speed is always the speed of light. In GR's language, that is

$$g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda} = 0 \tag{1.12}$$

Considering a locally Minkowski spacetime perturbed by the gravitational potential of a massive object, under the weak-field approximation, metric $g_{\mu\nu}$ can be written as

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{1.13}$$

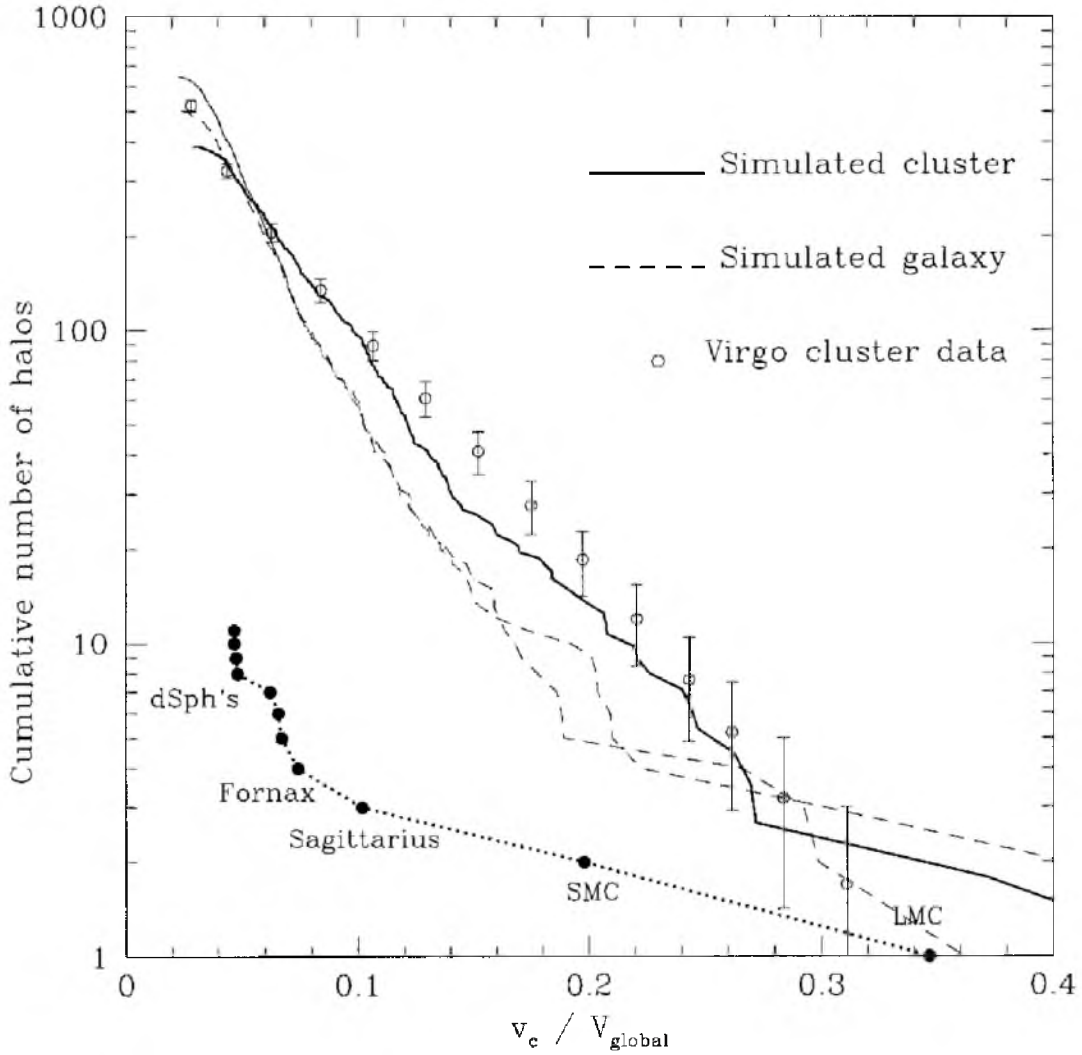

Figure 1.11: Mass functions of substructures within the Milky Way (filled circles) and Virgo Cluster (open circles) as a function of their circular velocity. The solid line is the numerical prediction for a simulated cluster, and the two dashed lines are the predictions for simulated galaxies at different ages. Credit: Figure 2 in the referred publication: "Dark Matter Substructure within Galactic Halos," Moore, Ben, et al., 1999, ApJ, 524, 19. Reproduced by permission of the AAS.

where $\eta_{\mu\nu}$ is the Minkowski metric as

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

and $h_{\mu\nu}$ is the metric components due to perturbation, which, in an appropriate coordinate system, is

$$\begin{pmatrix} \frac{2\Phi}{c^2} & 0 & 0 & 0 \\ 0 & \frac{2\Phi}{c^2} & 0 & 0 \\ 0 & 0 & \frac{2\Phi}{c^2} & 0 \\ 0 & 0 & 0 & \frac{2\Phi}{c^2} \end{pmatrix}$$

We can then work out the speed of photons in the gravitational field as

$$v = ||\dot{x}^i|| = (\frac{1+h_{00}}{1-h_{ii}})^{1/2} c \tag{1.14}$$

in which c is the speed of light in vacuum. The effect of the gravitational potential Φ on photons can therefore be expressed as the light path in a medium with effective refractive index n defined as

$$n \equiv \frac{c}{v} = (\frac{1-h_{ii}}{1+h_{00}})^{1/2} \approx 1 - \frac{2\Phi}{c^2} \tag{1.15}$$

The last step in the above equation is valid under the weak-field approximation in which $\Phi/c^2 << 1$. From Snell's law, it can be shown that the deflection angle is related to the gradient of the refractive index as

$$\vec{\hat{\alpha}} = -\int \nabla_\perp n \, \mathrm{d}l = \frac{2}{c^2} \int \nabla_\perp \Phi \, \mathrm{d}l \tag{1.16}$$

In all cases of interest, the deflection angle is small and the integration can be approximated by integrating along the unperturbed light ray instead of the actual light path.

For a point mass M acting as the lens, the gravitational potential is

$$\Phi(b, x) = -\frac{GM}{(b^2 + x^2)^{1/2}} \tag{1.17}$$

with G the gravitational constant, b the impact parameter, and x the position along the unperturbed light ray. Therefore, the deflection angle is

$$\vec{\hat{\alpha}} = \frac{2}{c^2} \int_{-\infty}^{\infty} \nabla_\perp \Phi(b, x) \, \mathrm{d}x = \frac{2}{c^2} \int_{-\infty}^{\infty} \frac{\partial \Phi(b, x)}{\partial b} \, \mathrm{d}x = \frac{4GM}{c^2 b^2} \vec{b} \tag{1.18}$$

To study the deflection by a realistic lens that usually has a 3D mass-density distribution, we first introduce the thin lens approximation. Note that most of the deflection occurs when $|x| \sim b$ which is almost negligible compared with distances between the lens and the source and between the observer and the lens. So the lens can be considered thin and replaced by a sheet of mass perpendicular to the line-of-sight. The plane of the mass sheet is referred to as the "lens plane" with a surface mass-density

$$\Sigma(\vec{\xi}) = \int \rho(\vec{\xi}, x) \, \mathrm{d}x \tag{1.19}$$

As a result, the deflection angle at position $\vec{\xi}$ is the summation of deflections by individual

mass-element

$$\vec{\hat{\alpha}} = \frac{4G}{c^2} \int \frac{\Sigma(\vec{\xi'})(\vec{\xi} - \vec{\xi'})}{|\vec{\xi} - \vec{\xi'}|^2} \, \mathrm{d}^2\vec{\xi'} \tag{1.20}$$

In a special case of circularly symmetric mass-density distribution, we can simplify the integral using the residual theorem to

$$\vec{\hat{\alpha}} = \frac{4G}{c^2\xi} \int_0^{\xi} \Sigma(\xi') 2\pi\xi' \, \mathrm{d}\xi' = \frac{4GM(\xi)}{c^2\xi} \tag{1.21}$$

1.2.2 Lens Geometry and Lens Equation

Figure 1.12 displays a gravitational lens system that consists of a background source at a distance D_s from the observer, and a massive object as the lens located at a distance D_d along the line-of-sight between the observer and the source. The distance between the lens and the source is D_{ds}. Note that all the distances defined above are angular diameter distances and hence $D_{ds} \neq D_s - D_d$ in general. The impact parameter is ξ and the deflection angle at this position $\vec{\hat{\alpha}}$ is given by Equation 1.21. The source has an angular position of β as seen by the observer and an image is generated at an angle θ. From geometry, it is clear that $D_s\beta + D_{ds}\hat{\alpha} = D_s\theta$ under the small-angle approximation. Therefore, the positions of the image and the source are related by the following equation known as the "lens equation"

$$\vec{\beta} = \vec{\theta} - \vec{\alpha}(\vec{\theta}) \tag{1.22}$$

where $\vec{\alpha} = \vec{\hat{\alpha}}\frac{D_{ds}}{D_s}$.

Consider a circularly symmetric lens and a very special case when the source lies exactly on the optical axis ($\vec{\beta} = 0$). The lens equation becomes

$$1 - \frac{M(R)}{\pi R^2 \Sigma_{crit}} = 0 \tag{1.23}$$

where $R = D_d\theta$ and the critical surface density Σ_{crit} is defined as

$$\Sigma_{crit} = \frac{c^2}{4\pi G} \frac{D_s}{D_{ds} D_d} \tag{1.24}$$

It is clear that if the average surface density of the lens within an arbitrary radius is always less than the critical density Σ_{crit}, then Equation 1.23 can never be satisfied and hence no extra images will be seen by the observer at distance D_d. To ensure multiple images, there has to be a radius in the lens plane within which the average surface density is equal to Σ_{crit}. Due to the rotational symmetry in this case, the source is imaged to a complete ring which is known as the "Einstein ring" (Figure 1.13). The corresponding radius in the lens plane is called the "Einstein radius" θ_{Ein}, which, as will be shown later, is a very important and useful quantity because it is a direct indicator of the lens mass.

$$\theta_{\mathrm{Ein}} = \left[\frac{4GM(\theta_{\mathrm{Ein}})}{c^2} \frac{D_{ds}}{D_d D_s}\right]^{1/2} \tag{1.25}$$

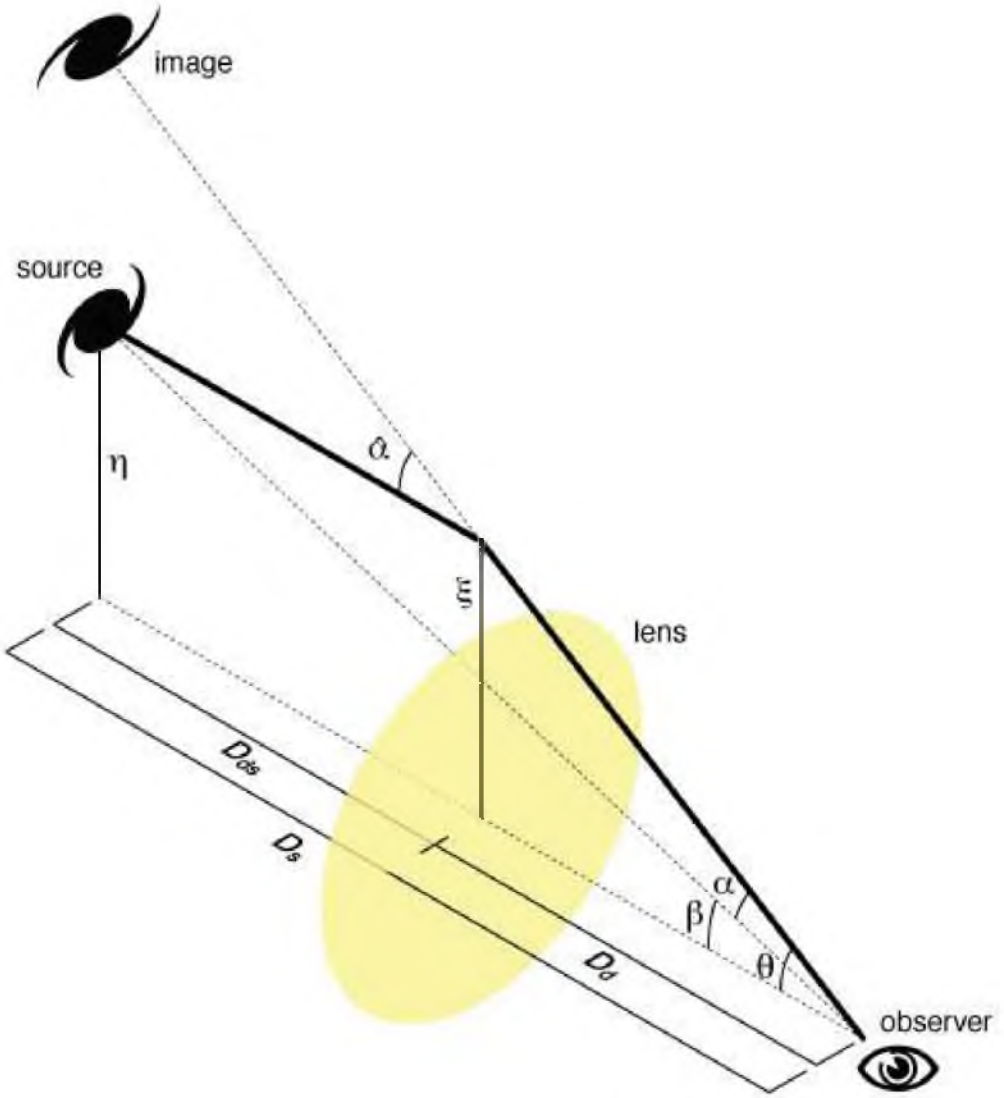

Figure 1.12: Illustration of the gravitational lensing effect. A massive object (yellow blob) along the line-of-sight direction between the observer and the background source acts as a gravitational lens. Credit: Public Domain, (via Wikimedia Commons, http://commons.wikimedia.org/wiki/File:Gravitational-lensing-angles.png, Author: Michael Sachs)

Figure 1.13: A gravitational lens system with an Einstein ring discovered in the SLACS for the Masses Survey (Shu et al. 2014, to be submitted). The bright blob in the center is an early-type galaxy selected spectroscopically from the Sloan Digital Sky Survey (SDSS) acting as the lens. The bluish ring (Einstein ring) around the lens is the lensed image of a star-forming galaxy behind the lens. False color scales are employed for better illustration.

1.2.3 Time Delay and Magnification

Due to the change in the refractive index and the light path, the arrival time of a signal from the source will be delayed as compared to that in the absence of the lens. This delay is known as the "time delay." Obviously, the time-delay contains contributions from two parts, gravitational delay t_{grav} due to the change in the speed of photons and geometric delay t_{geom} due to the change in the light path.

$$t_{grav} = \int \frac{\mathrm{d}l}{v} - \frac{\mathrm{d}l}{c} = \frac{1}{c}\int n - 1\,\mathrm{d}l = -\frac{2}{c^3}\int \Phi\,\mathrm{d}l = -\frac{2}{c^3}\int \Phi(D_d\vec{\theta}, x)\,\mathrm{d}x \tag{1.26}$$

where the integration is along the unperturbed light path and $D_d\vec{\theta}$ is the two-dimensional coordinates in the lens plane. We can define an effective lensing potential $\psi(\vec{\theta})$ as

$$\psi(\vec{\theta}) = \frac{2}{c^2}\frac{D_{ds}}{D_d D_s}\int \Phi(D_d\vec{\theta}, x)\,\mathrm{d}x \tag{1.27}$$

Then

$$t_{grav} = -\frac{1}{c}\frac{D_d D_s}{D_{ds}}\psi(\vec{\theta}) \tag{1.28}$$

The change in the light path can be easily derived from the laws of sines and cosines to be

$$\Delta l = \frac{1}{2}\frac{D_d D_s}{D_{ds}}(\vec{\theta} - \vec{\beta})^2 \tag{1.29}$$

and t_{geom} is

$$t_{geom} = \frac{\Delta l}{c} = \frac{1}{2c}\frac{D_d D_s}{D_{ds}}(\vec{\theta} - \vec{\beta})^2 \tag{1.30}$$

Note that both t_{grav} and t_{geom} are derived in the lens plane. To convert them to the observed quantities in the observer plane, an extra $(1 + z_d)$ factor is needed. So eventually the time-delay function is

$$t(\vec{\theta}) = \frac{1 + z_d}{c}\frac{D_d D_s}{D_{ds}}[\frac{1}{2}(\vec{\theta} - \vec{\beta})^2 - \psi(\vec{\theta})] \tag{1.31}$$

Note that the time-delay function explicitly contains the angular diameter terms which are cosmology-related quantities. So time-delays can be used to constrain cosmological parameters either independently or jointly with other methods (e.g., Blandford and Narayan, 1992; Jackson, 2007; Treu, 2010; Suyu et al., 2010, 2013).

Another property of GL is the magnification μ. According to the *Liouville's theorem*, GL preserves the surface brightness, but changes the apparent size (solid angle) of the source. The net effect is a flux magnification in lensed images. The solid-angle elements of the image $\delta\theta^2$ and the source $\delta\beta^2$ are related by the Jacobi matrix

$$\delta\theta^2 = \begin{vmatrix} \frac{\partial\theta_1}{\partial\beta_1} & \frac{\partial\theta_1}{\partial\beta_2} \\ \frac{\partial\theta_2}{\partial\beta_1} & \frac{\partial\theta_2}{\partial\beta_2} \end{vmatrix} \delta\beta^2 \tag{1.32}$$

We can define a matrix $\mathcal{A}$ which is

$$\mathcal{A} \equiv \frac{\partial\vec{\beta}}{\partial\vec{\theta}} = (\delta_{ij} - \frac{\partial\alpha_i}{\partial\theta_j}) = (\delta_{ij} - \frac{\partial^2\psi(\vec{\theta})}{\partial\theta_i\partial\theta_j}) \tag{1.33}$$

It is easy to see that the magnification is the inverse of the determinant of matrix $\mathcal{A}$

$$\mu = \frac{\delta\theta^2}{\delta\beta^2} = \begin{vmatrix} \frac{\partial\theta_1}{\partial\beta_1} & \frac{\partial\theta_1}{\partial\beta_2} \\ \frac{\partial\theta_2}{\partial\beta_1} & \frac{\partial\theta_2}{\partial\beta_2} \end{vmatrix} = \frac{1}{\det\mathcal{A}} \tag{1.34}$$

Define the following quantities

$$\kappa = \frac{1}{2}(\psi_{11} + \psi_{22}) \tag{1.35}$$

$$\gamma_1 = \frac{1}{2}(\psi_{11} - \psi_{22}) \tag{1.36}$$

$$\gamma_2 = \psi_{12} = \psi_{21} = \frac{1}{2}(\psi_{12} + \psi_{21}) \tag{1.37}$$

With these definitions, we have

$$\mathcal{A} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix} \tag{1.38}$$

and

$$\mu = \frac{1}{(1-\kappa)^2 - \gamma_1^2 - \gamma_2^2} \tag{1.39}$$

μ is very sensitive to the gravitational potential as it depends on the second derivative of the potential. Therefore, it has been used as an effective probe of the lens mass distribution. The feasibility of using flux-ratio anomaly to detect substructures in the lens was first discussed by Mao and Schneider (1998) and then supported by various observations and simulations (e.g., Metcalf and Zhao, 2002; Moustakas and Metcalf, 2003; Keeton et al., 2003; Kochanek and Dalal, 2004; McKean et al., 2007; Xu et al., 2009; MacLeod et al., 2013).

1.2.4 Singular Isothermal Ellipsoid Model

According to the definition of the effective potential $\psi(\vec{\theta})$ (Equation 1.27), we can easily see that the deflection angle $\vec{\alpha}$ is the gradient of $\psi(\vec{\theta})$

$$\vec{\alpha} = \nabla_\theta \psi(\vec{\theta}) \tag{1.40}$$

and $\psi(\vec{\theta})$ itself satisfies the following two-dimensional Poisson's equation

$$\nabla_\theta^2 \psi(\vec{\theta}) = 2\kappa = 2\frac{\Sigma(\vec{\theta})}{\Sigma_{crit}} \tag{1.41}$$

In principle, we can solve for $\psi(\vec{\theta})$ for an arbitrary surface-density distribution using Green's function as

$$\psi(\vec{\theta}) = \frac{1}{\pi}\int \kappa(\vec{\theta'}) \ln|\vec{\theta} - \vec{\theta'}| \, \mathrm{d}^2\theta' \tag{1.42}$$

from which the deflection angle can be further derived. However, there are only a small fraction of density profiles that permit analytical solutions to the direct integration, among which the Singular Isothermal Ellipsoid (SIE) model has the largest degree of generality and at the same time well matches the observations.

In an SIE lens model, the two-dimensional density profile can be written as

$$\Sigma(\theta_x, \theta_y) = \frac{\sigma_{\mathrm{SIE}}^2}{2G} \frac{\sqrt{q}}{\sqrt{\theta_x^2 + q^2\theta_y^2}} \tag{1.43}$$

where σ_{SIE} has the dimensionality of velocity and represents the characteristic velocity

dispersion of the lens, and q is the minor-to-major axis ratio. The total mass within an intermediate axis R is

$$M(R) = \frac{\pi\sigma_{\rm SIE}^2}{G}R \tag{1.44}$$

The radius within which the average surface density is equal to the critical density, namely the Einstein radius, can be found as

$$\theta_{\rm Ein} = 4\pi\frac{\sigma_{\rm SIE}^2}{c^2}\frac{D_{ds}}{D_s} \tag{1.45}$$

This model is called "isothermal" because $\sigma_{\rm SIE}$ is assumed to be independent of the radius. In thermal equilibrium, the temperature T is proportional to the velocity dispersion $\sigma_{\rm SIE}$ as a result of the virial theorem. Once $\sigma_{\rm SIE}$ does not depend on radius, the temperature is constant across the entire galaxy and therefore, the galaxy is isothermal.

As proven by Kormann et al. (1994), the potential $\psi(\vec{\theta})$ can be solved analytically to be

$$\psi(r,\phi) = \frac{\sqrt{q}}{\sqrt{1-q^2}}r\theta_{\rm Ein}[\sin\phi\ \arcsin(\sqrt{1-q^2}\sin\phi) + \cos\phi\,\mathrm{arcsinh}(\frac{\sqrt{1-q^2}}{q}\cos\phi)] \tag{1.46}$$

The deflection angle is simply the gradient of $\psi(\vec{\theta})$

$$\alpha_x = \theta_{\rm Ein}\frac{\sqrt{q}}{\sqrt{1-q^2}}\mathrm{arcsinh}(\frac{\sqrt{1-q^2}}{q}\cos\phi) \tag{1.47}$$

$$\alpha_y = \theta_{\rm Ein}\frac{\sqrt{q}}{\sqrt{1-q^2}}\arcsin(\sqrt{1-q^2}\sin\phi) \tag{1.48}$$

1.3 Hierarchical Bayesian Inference

Astronomy has been significantly revolutionized after entering the new century by the developments of mirror production, adaptive optics, robotic optical fiber, space techniques and etc. Dedicated large all-sky surveys in various wavelength ranges such as the Sloan Digital Sky Survey (SDSS-I,II,III, York et al., 2000; Frieman et al., 2008; Eisenstein et al., 2011a), Dark Energy Survey (DES), Wide-field Infrared Survey Explorer (WISE, Wright et al., 2010) and the Galaxy And Mass Assembly survey (GAMA, Driver et al., 2011) have shed light on the Universe unprecedentedly with enormous data (GBs to TBs) observed nightly. To "survive" from this data-explosion situation, effective tools which enable unambiguous studies of large populations based on typically low signal-to-noise ratio (SNR) datasets are highly demanded. One such tool is the hierarchical Bayesian inference, which can deconvolve the observational uncertainties and hence probe the intrinsic properties of the entire population.

Bayesian inference is build upon the Bayes' theorem which states that the conditional probability of event A given event B multiplied by the prior probability of event B is equal to the conditional probability of event B given A multiplied by the prior probability of event A, and is also equal to the joint probability of event A and B. In a mathematician's

language, that is

$$\Pr(A|B)\Pr(B) = \Pr(B|A)\Pr(A) = \Pr(A, B) \tag{1.49}$$

Bayesian inference is used to determine how likely a hypothesis $\mathscr{H}$ is true given the measured data D. Typically, a hypothesis $\mathscr{H}$ can be represented by a particular model described by a set of physical parameters $\boldsymbol{\lambda}$. So we have

$$\Pr(\boldsymbol{\lambda}|D, \mathscr{H}) = \frac{\Pr(D|\boldsymbol{\lambda}, \mathscr{H})\Pr(\boldsymbol{\lambda}|\mathscr{H})}{\Pr(D|\mathscr{H})} \tag{1.50}$$

in which $\Pr(\boldsymbol{\lambda}|D, \mathscr{H})$ is the *posterior* probability density function (PDF) of parameter $\boldsymbol{\lambda}$ given the data D and assuming hypothesis $\mathscr{H}$, $\Pr(D|\boldsymbol{\lambda}, \mathscr{H})$ is the *likelihood* function of $\boldsymbol{\lambda}$, $\Pr(\boldsymbol{\lambda}|\mathscr{H})$ is the *prior* PDF of $\boldsymbol{\lambda}$ and $\Pr(D|\mathscr{H})$ is the *evidence* Z. Notice that

$$Z = \int \Pr(D|\boldsymbol{\lambda}, \mathscr{H})\Pr(\boldsymbol{\lambda}|\mathscr{H})\,\mathrm{d}\boldsymbol{\lambda} \tag{1.51}$$

One relies on the evidence to determine which hypothesis/model better describes the observation. For instance, to select between $\mathscr{H}_1$ and $\mathscr{H}_2$, we simply write down the ratio of the two evidences, known as the *Bayes factor* K, as

$$K = \frac{Z_1}{Z_2} = \frac{\Pr(D|\mathscr{H}_1)}{\Pr(D|\mathscr{H}_2)} = \frac{\int \Pr(D|\boldsymbol{\lambda}_1, \mathscr{H}_1)\Pr(\boldsymbol{\lambda}_1|\mathscr{H}_1)\,\mathrm{d}\boldsymbol{\lambda}_1}{\int \Pr(D|\boldsymbol{\lambda}_2, \mathscr{H}_2)\Pr(\boldsymbol{\lambda}_2|\mathscr{H}_2)\,\mathrm{d}\boldsymbol{\lambda}_2} \tag{1.52}$$

In statistics, a value of $K > 1$ indicates that $\mathscr{H}_1$ is more strongly supported by the observation than $\mathscr{H}_2$. More specifically, according to the interpretation by Jeffreys (1961), $1 < K < 3$ means "barely worth mentioning," $10 < K < 30$ corresponds to a "strong" confidence of favoring $\mathscr{H}_1$, and $K > 100$ is "decisive."

Under a particular hypothesis, Bayesian inference can be used to constrain physical parameters $\boldsymbol{\lambda}$ based on the observational data D. Recall that the posterior PDF of $\boldsymbol{\lambda}$ is purely determined by the likelihood function and the prior PDF of $\boldsymbol{\lambda}$ as the evidence itself does not depend on $\boldsymbol{\lambda}$. Once the likelihood function of $\boldsymbol{\lambda}$ is constructed, under a reasonable assumption of the prior (typically uniform or Gaussian distribution), one can map out the posterior PDF of $\boldsymbol{\lambda}$.

A *hierarchical* Bayesian model is introduced when the parameters $\boldsymbol{\lambda}$ themselves can be parameterized by another set of parameters $\boldsymbol{\theta}$, in which case, $\boldsymbol{\theta}$ is identified as the *hyperparameters*. The likelihood function of the hyperparameters $\boldsymbol{\theta}$ can be constructed as

$$\mathscr{L}(\boldsymbol{\theta}|D, \mathscr{H}) \equiv \Pr(D|\boldsymbol{\theta}, \mathscr{H}) = \int \Pr(D|\boldsymbol{\lambda}, \mathscr{H})\Pr(\boldsymbol{\lambda}|\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\lambda} \tag{1.53}$$

CHAPTER 2

EVOLUTION OF VELOCITY-DISPERSION FUNCTION OF LUMINOUS RED GALAXIES: A HIERARCHICAL BAYESIAN MEASUREMENT[†]

We present a hierarchical Bayesian determination of the velocity-dispersion function of approximately 430,000 massive luminous red galaxies(LRGs) observed at relatively low spectroscopic signal-to-noise ratio (SNR ∼3–5 per $69\,\mathrm{km\,s^{-1}}$) by the Baryon Oscillation Spectroscopic Survey (BOSS) of the Sloan Digital Sky Survey III (SDSS-III). We marginalize over spectroscopic redshift errors, and use the full velocity-dispersion likelihood function for each galaxy to make a self-consistent determination of the velocity-dispersion distribution parameters as a function of absolute magnitude and redshift, correcting as well for the effects of broadband magnitude errors on our binning. Parameterizing the distribution at each point in the luminosity–redshift plane with a log-normal form, we detect significant evolution in the width of the distribution toward higher intrinsic scatter at higher redshifts. Using a subset of deep reobservations of BOSS galaxies, we demonstrate that our distribution-parameter estimates are unbiased regardless of spectroscopic SNR. We also show through simulation that our method introduces no systematic parameter bias with redshift. We highlight the advantage of the hierarchical Bayesian method over frequentist "stacking" of spectra, and illustrate how our measured distribution parameters can be adopted as informative priors for velocity-dispersion measurements from individual noisy spectra.

2.1 Introduction

Massive elliptical galaxies (EGs: Hubble, 1936a) are one of the most important classes of astrophysical objects for galaxy evolution and cosmology. They represent the end stage of

[†]Published as **Shu, Yiping**, Bolton, A. S., Schlegel, D. J., Dawson, K. S., Wake, D. A., Brownstein, J. R., Brinkmann, J., Weaver, B. A., "*Evolution of Velocity-Dispersion Function of Luminous Red Galaxies: A Hierarchical Bayesian Measurement*," 2012, AJ, 143, 90"

hierarchical galaxy-formation processes (e.g., Kauffmann et al., 1993b; Baugh et al., 1996), and therefore their properties and scaling relations represent a key test for theories of galaxy formation within a cosmological context. In addition, since they are the most luminous and highly clustered galaxies, they serve as ideal cosmological tracers of clusters and large-scale structure (e.g., Eisenstein et al., 2005b).

To a first approximation, EGs are "pressure-supported" rather than rotationally supported (e.g., Bertola and Capaccioli, 1975; Illingworth, 1977; Binney, 1978), with their stellar motions characterized by a velocity dispersion σ. Among the many observational parameters of massive elliptical galaxies, σ is unique in its direct sensitivity to the depth of the galaxy's gravitational potential (and therefore to its mass), and in its relatively weak dependence on observational aperture. In combination with galaxy sizes (i.e., half-light radii), velocity dispersions can be used to determine "dynamical masses" that are independent of stellar-population assumptions (e.g., Padmanabhan et al., 2004; Bolton et al., 2008d). Dynamical masses can then in turn be used to trace the evolution of EGs at fixed mass (e.g., van der Marel and van Dokkum, 2007; van der Wel et al., 2008; Cappellari et al., 2009), indicating a nuanced dynamical history despite generally passive star-formation histories at $z < 1$ (e.g., Thomas et al., 2005; Cool et al., 2008). Stellar velocity dispersion is also the most important single predictor of strong gravitational lensing cross-sections (e.g., Turner et al., 1984; Bolton et al., 2008c), and can be used in combination with strong lensing observations to constrain the central mass-density structure of elliptical galaxies at cosmological distances (e.g., Koopmans and Treu, 2002; Treu and Koopmans, 2004; Koopmans et al., 2006a). Stellar velocity dispersions are tied to nearly all other properties of EGs through multiple empirical scaling relations. Faber and Jackson (1976) found a correlation between luminosities of early-type galaxies and their velocity dispersions σ known as the Faber-Jackson Relation (FJR). The relation of Kormendy (1977) ties the surface brightness $\langle \mathrm{I} \rangle_{\mathrm{e}}$ with the effective radius $\mathrm{R_e}$. Both the FJR and Kormendy relations can be viewed as projections of the "fundamental plane" (FP, e.g., Djorgovski and Davis, 1987; Dressler et al., 1987; Bernardi et al., 2003c) within the space spanned by $\log_{10} \mathrm{R_e}$, $\langle \mathrm{I} \rangle_{\mathrm{e}}$ and $\log_{10} \sigma$. Furthermore, central black hole mass has been found to be correlated with the velocity dispersion of the bulge via the $\mathrm{M_{BH}} - \sigma$ relation(e.g., Ferrarese and Merritt, 2000; Gebhardt et al., 2000; Kormendy and Bender, 2009). Together, these relations provide multiple constraints on the structure, formation, and evolution of EGs .

Although velocity dispersion plays a starring role in the study of EGs, it is an "expensive" observable that must be measured spectroscopically. Hence, large samples of galaxies with

well-measured velocity dispersions across cosmic time are largely unavailable. Measurements of σ are made by quantifying the line-of-sight Doppler broadening of absorption lines relative to a set of template stellar spectra, either via the Fourier method (e.g., Sargent et al., 1977; Tonry and Davis, 1979) or the direct-fitting method (e.g., Burbidge et al., 1961; Rix and White, 1992). Both methods rely on the quality of galaxy spectra: for spectra of low signal-to-noise ratio (SNR), uncertainties in the measured stellar velocity dispersion can be very large and significantly non-Gaussian. This aspect is of particular concern for galaxies at cosmological distance (faint even if luminous), which can only be measured at high SNR through substantial investment of spectroscopic observing time and aperture.

In this paper, we introduce a hierarchical Bayesian statistical method to measure the parameters of the distribution of stellar velocity dispersions within a population of galaxies that has been observed with relatively low spectroscopic signal-to-noise ratio. We apply the method to approximately 430,000 luminous red galaxy (LRG) targets from the Baryon Oscillation Spectroscopic Survey (BOSS: Schlegel et al., 2009), one of four survey projects within the Sloan Digital Sky Survey III (SDSS-III: Eisenstein et al., 2011b). We quantify the evolution of the velocity-dispersion function of BOSS galaxies, and detect significant evolution in the intrinsic population RMS of $\log_{10}\sigma$ at fixed absolute magnitude since $z \approx 0.8$.

This chapter is organized as follows. In Section 2.2, we describe the sample selection and the method for velocity dispersion measurement. Section 2.3 presents our statistical method for the measurement of the distribution of stellar velocity dispersions within a population of galaxies, including a verification using high-SNR reobservations of a subsample of galaxies and a test for redshift-dependent systematic biases. Section 2.4 presents the results of our application of this method to the BOSS sample, showing the evolution of the velocity-dispersion function at fixed magnitude. Discussion and conclusions are presented in Section 2.5. Throughout the paper, we assume a standard general-relativistic cosmology with $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$ and $\rm H_0 = 70\ km\ s^{-1}\ Mpc^{-1}$.

2.2 Spectroscopic Data

2.2.1 Sample Selection

We use spectroscopic data obtained by the BOSS project via the 2.5-m SDSS telescope located at Apache Point Observatory in Sunspot, New Mexico (Gunn et al., 2006). The primary science goal of BOSS is the detection of the baryon acoustic feature in the two-point correlation function of galaxies (and quasar absorption systems), from which to constrain the

distance–redshift relation and the nature of dark energy. BOSS also offers a unique resource for the study of the properties and evolution of massive galaxies. The BOSS footprint covers approximately 10,000 deg^2 in five imaging filters (*ugriz*, Fukugita et al., 1996), and will by 2014 obtain spectra of about 1.5 million LRGs out to redshift z $\simeq$ 0.8. Note that the majority of the BOSS LRG targets are massive EGs, although there is a small fraction of late-type galaxies as well as unresolved multiples, particularly at the higher redshift end (Masters et al., 2011).

The BOSS spectra are broadly comparable to SDSS-I (York et al., 2000) spectra in resolution ($R \approx 2000$), and cover a wavelength range from 3,600Å to 10,000Å. The primary design goal of BOSS is to measure as many redshifts as efficiently as possible, in order to map the largest possible volume of the universe. Consequently, the SNR of the galaxy spectra is significantly lower than in SDSS-I, with typical SNR values of 3 to 5 per 69 km s^{-1} (rebinned) pixel, as compared with $\gtrsim$ 10 per pixel in SDSS-I. Thus, although the BOSS spectroscopic database is by far the largest available for the study of massive galaxies, the individual spectra are well below the SNR threshold of about 10 per Å generally regarded as a minimum for acceptable velocity-dispersion measurement on a galaxy-by-galaxy basis. Motivated by this context, we develop the Bayesian analysis method presented below.

Spectroscopic calibration, extraction, classification, and redshift measurement of all BOSS galaxy spectra are carried out using the `idlspec2d` software (see, e.g., Aihara et al., 2011), written originally for SDSS-I and recently updated to handle the data format and noise regime of BOSS. In selecting our analysis sample, we make the following cuts based upon the redshift pipeline output:

- We use only the best spectroscopic observation of any given galaxy target as some objects are observed more than once (`SPECPRIMARY`=1 according to SDSS terminology).
- We use only objects that were both targeted as galaxies and spectroscopically confirmed as galaxies
- We require a confident redshift measurement with no warning flags (`ZWARNING` = 0 according to SDSS terminology).[1]

These cuts return approximately 430,000 galaxies from the first 1.5 years of BOSS spectroscopic observations, with redshifts ranging from zero to 1, but concentrated primarily over the interval $0.2 < z < 0.8$.

For all selected galaxies, we use the measured spectroscopic redshifts and SDSS broadband imaging colors to compute absolute rest-frame V-band magnitudes and associated

[1]For BOSS galaxies, the specifically relevant flag is `ZWARNING_NOQSO` = 0 (Bolton et al., in prep.)

uncertainties via the `sdss2bessell` routine implemented in the `kcorrect` software of Blanton and Roweis (2007).

The details of the BOSS galaxy target selection, and the corresponding incompletenesses, are the subject of a separate paper (Padmanabhan et al., in preparation). Here we summarize the target selection cuts for the two main galaxy target classes that we focus upon in our current study. The first is the CMASS sample (for "constant mass"), which is selected photometrically to deliver LRGs of approximately constant stellar mass over the redshift interval $0.3 < z < 0.8$, and which constitutes approximately 76% of the galaxies selected above. The second sample, LOWZ, is selected to target LRGs at lower redshifts, and represents the remaining 24% of the selected galaxies. Defining the following quantities (Eisenstein et al., 2001a; Cannon et al., 2006):

$$c_{\parallel} = 0.7(g-r) + 1.2[(r-i) - 0.18] \tag{2.1}$$

$$c_{\perp} = (r-i) - (g-r)/4.0 - 0.18 \tag{2.2}$$

$$d_{\perp} = (r-i) - (g-r)/8.0 \tag{2.3}$$

$$\text{ifiber2} = i\text{-band fiber magnitude for } 2'' \text{ fibers}, \tag{2.4}$$

the CMASS sample is defined by the photometric cuts:

$$17.5 < i < 19.9 \tag{2.5}$$

$$r - i < 2 \tag{2.6}$$

$$d_{\perp} > 0.55 \tag{2.7}$$

$$\text{ifiber2} < 21.7 \tag{2.8}$$

$$i < 19.86 + 1.60(d_{\perp} - 0.80) \tag{2.9}$$

as well as a cut to exclude galaxies with major-axis half-light radii greater than $8''$. Equations (2.5) and (2.7) aim to select galaxies between redshifts $z \sim 0.4$–0.8, while Equation (2.9) attempts to impose a cut at constant stellar mass across this redshift range. The LOWZ sample is defined by the cuts:

$$r < 13.5 + c_{\parallel}/0.3 \tag{2.10}$$

$$|c_{\perp}| < 0.2 \tag{2.11}$$

$$16 < r < 19.6\ . \tag{2.12}$$

Equation (2.10) sets up a magnitude threshold as a function of redshift and Equation (2.11) picks out low-redshift galaxies specifically.

The redshift–absolute magnitude distributions of these two BOSS galaxy samples, with associated 1D projections, are plotted in Figure 2.1. In the following analysis, we will treat the two populations separately, since the combined sample does not define a simple locus

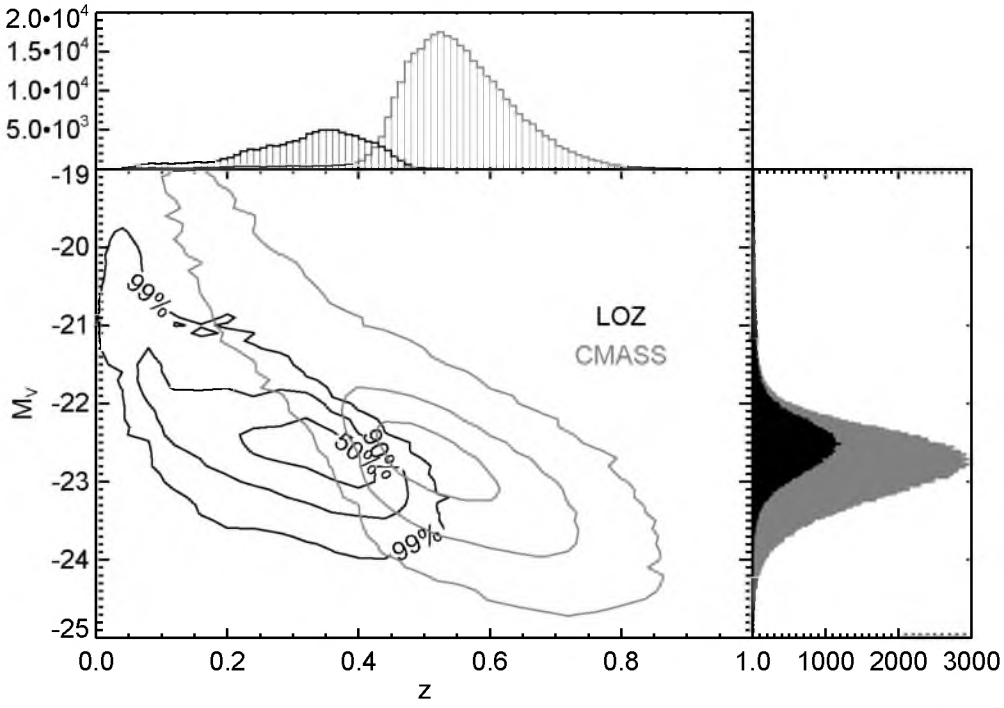

Figure 2.1: Distribution of galaxies for our sample, along with histograms of redshift z and V-band absolute magnitudes M_V. LOWZ galaxies *(black)* and CMASS galaxies *(gray)* are plotted separately. For both the samples, contours are drawn at constant number density in the z–M_V plane, enclosing 50%, 90%, and 99% of the sample. Credit: Figure 1 in the referred publication "Evolution of the Velocity-dispersion Function of Luminous Red Galaxies: A Hierarchical Bayesian Measurement," Shu, Yiping, et al., 2012, AJ, 143, 90. Reproduced by permission of the AAS.

in luminosity–redshift space, with LOWZ galaxies being of generally higher luminosity over the redshift range where the two samples overlap.

2.2.2 Velocity Dispersion Extraction

Our strategy for extracting velocity-dispersion information is to make use of the full velocity-dispersion likelihood function for each galaxy spectrum. To do this, we make use of the IDL routine `vdispfit` within the `idlspec2d` product of spectroscopic analysis software. This software has been extensively tested in the SDSS-I, and has been upgraded for the analysis of BOSS data. Velocity dispersions measured with this software have been the basis for multiple studies of the dynamics of EGs (e.g., Bernardi et al., 2003a,b,c; Sheth et al., 2003; Padmanabhan et al., 2004; Koopmans et al., 2006a). To summarize briefly: `vdispfit` uses a set of stellar eigenspectra derived from a principal-component analysis (PCA) decomposition of the ELODIE stellar spectrum library (Prugniel and Soubiran, 2001). The eigenspectra are convolved and binned to the resolution and sampling of the BOSS spectra, then broadened by Gaussian kernels of different trial velocity dispersions. The broadened templates are then shifted to the redshift of the galaxy under consideration. After masking out regions containing common emission lines, a linear least-squares fit is performed to obtain a best-fit model spectrum at each trial velocity dispersion. The resulting curve of χ^2 as a function of trial velocity dispersion encodes the likelihood function of velocity dispersion given the data. For measurements from high signal-to-noise spectra, the position of the minimum χ^2 is adopted as the maximum-likelihood estimate of the galaxy's velocity dispersion. Below, rather than adopt these estimates, we will work with the full likelihood function.

In this procedure, we must choose the number of stellar eigenspectra to use in forming the template basis. The pipeline analysis of SDSS-I data used the first 24 PCA modes. For the much lower signal-to-noise BOSS data, an acceptable χ^2 can be obtained using only the first 5 PCA modes, and hence we restrict our basis to this smaller number of eigenspectra so as to avoid fitting noise fluctuations.

As described above, before being fit to the galaxy spectra, the stellar eigenspectra are shifted by the appropriate galaxy redshifts. If the redshifts have nonnegligible errors, the corresponding offsets can introduce a bias into the measured velocity dispersion. Although the BOSS spectra provide redshifts with a precision well in excess of what is required for large-scale structure studies and absolute-magnitude determinations, their errors can be nonnegligible on the scale of internal galaxy velocity dispersions. Therefore, we implement a marginalization over redshift errors in our analysis. Specifically, we modify the

`vdispfit` routine to take a radial velocity-marginalization range Δz (expressed in constant-velocity pixels) and the redshift error δz (the $\pm 68\%$ confidence interval as estimated by the `idlspec2d` pipeline) as arguments. Then we calculate $\chi^2(\sigma, z)$ for a set of trial redshifts in the range $z \pm \Delta z$ and define a new effective $\chi^2(\sigma)$ by integrating over z as

$$\chi^2(\sigma) = -2 \ln \left(\int_{z-\Delta z}^{z+\Delta z} dz \, \exp \left[-\frac{\chi^2(\sigma, z)}{2} \right] \, p(z) \right) \tag{2.13}$$

We assume a Gaussian probability distribution for z given by

$$p(z) \propto \exp \left[-\frac{(z - z_{\rm best})^2}{2\delta z^2} \right] \tag{2.14}$$

where $z_{\rm best}$ is the best-estimate redshift from the BOSS spectroscopic pipeline. The choice of a Gaussian prior is made because the galaxy redshifts have been determined using absorption and emission-line information over the full optical range of the BOSS spectrograph, whereas the velocity-dispersion fitting is done only over the 4100–6800 Å rest-frame range covered by the ELODIE spectra, while also masking the wavelength positions of common emission lines. We also explored the use of a flat prior to completely marginalize over redshift in the velocity dispersion analysis, and found only a negligible change (at most a few percent) in the derived relations. For most galaxies, the effect of this marginalization on the χ^2 curve is insignificant, but since we wish to avoid introducing any spurious broadening into our population analysis, we apply the procedure to all spectra.

In this work, we do not make any aperture correction for velocity dispersions, although the angular BOSS fiber radius of $1''$ subtends a different physical length scale as a function of redshift. Since aperture velocity dispersions are seen in the local universe to depend on aperture radius only to a weak power of approximately 0.04 to 0.06 (e.g., Jørgensen et al., 1995; Mehlert et al., 2003a; Cappellari et al., 2006b), this effect should be relatively insignificant. For example, taking a redshift range spanning the majority of our CMASS sample, the angular size of a fixed physical length at $z = 0.8$ is about 72% of its angular size at $z = 0.4$. Assuming the velocity dispersion within an aperture decreases as the aperture to the power -0.05 (a representative compromise value between the previous three references), this would correspond to a systematic change in measured velocity dispersion of about 1.7%, which is well below the level of other uncertainties in our analysis. In addition, the typical atmospheric seeing of approximately $1.8''$ delivered to the BOSS spectroscopic focal plane will dilute the significance of the varying projected fiber scale. Essentially, BOSS velocity dispersions will represent a fair luminosity-weighted average value over the half-light radius of most target galaxies, which have half-light radii on the order of $1''$.

2.3 Statistical Population Analysis Formalism

The results of Bernardi et al. (2003b) suggest that the distribution of velocity dispersions for early-type galaxies at fixed luminosity can be well-approximated by a log-normal function. Motivated by this, we will assume a Gaussian distribution in $\log_{10}\sigma$ with mean m and intrinsic scatter s:

$$p(\log_{10}\sigma|m,s) = \frac{1}{\sqrt{2\pi}s}\exp[\frac{-(\log_{10}\sigma - m)^2}{2s^2}] \tag{2.15}$$

We will treat m and s as functions of redshift and absolute magnitude, although we will suppress this dependence in our notation for convenience. Compared to the SDSS-I studies by Bernardi et al. (2003b) and Sheth et al. (2003), we have a much larger sample with greater redshift coverage, so we may investigate the evolution of both the mean and intrinsic scatter of $\log_{10}\sigma$ with redshift and luminosity as encoded by these two distribution parameters (see also Bezanson et al. 2011 for a complementary analysis in terms of photometric velocity-dispersion proxies). Our strategy will be to analyze samples binned by an interval of 0.04 in redshift z, and by 0.1 in absolute magnitude M_V.

2.3.1 Frequentist Approach

As mentioned above, the SNR of BOSS galaxy spectroscopy is typically rather low, especially at the high-redshift end of the survey. Therefore, point estimation of the velocity dispersion of individual galaxies is of questionable reliability. Hence, we resort to analyzing the data by binning galaxies in the z–M_V plane, requiring at least 100 galaxies in every single bin. The most obvious first approach to determining the mean velocity dispersion in these bins is to remove the small relative redshift differences within the bin, stack all the spectra directly, and analyze the resulting high-SNR combination (see Figure 2.2). Although we do not adopt this method for our ultimate determinations of m and s, it is instructive to consider how such an approach relates to these parameters.

While a velocity dispersion can be measured at high SNR from the stacked spectrum, the measured value bears a nontrivial relation to the parameters m and s, which we now derive. Assuming equal luminosities within the bin (which basically holds by construction due to binning in absolute magnitude), what we measure from the stack $\sigma^2_{\rm stack}$ is the population-weighted expectation value of σ^2, i.e.

$$\sigma^2_{\rm stack} = \langle\sigma^2\rangle = \int \sigma^2\, p(\log_{10}\sigma|m,s)\,{\rm d}\log_{10}\sigma \tag{2.16}$$

The variance of $\sigma^2_{\rm stack}$ is given by

$${\rm Var}(\sigma^2_{\rm stack}) = \frac{1}{N}{\rm Var}(\sigma^2) = \frac{1}{N}(\langle\sigma^4\rangle - \langle\sigma^2\rangle^2) \tag{2.17}$$

with N being the number of galaxies in the bin.

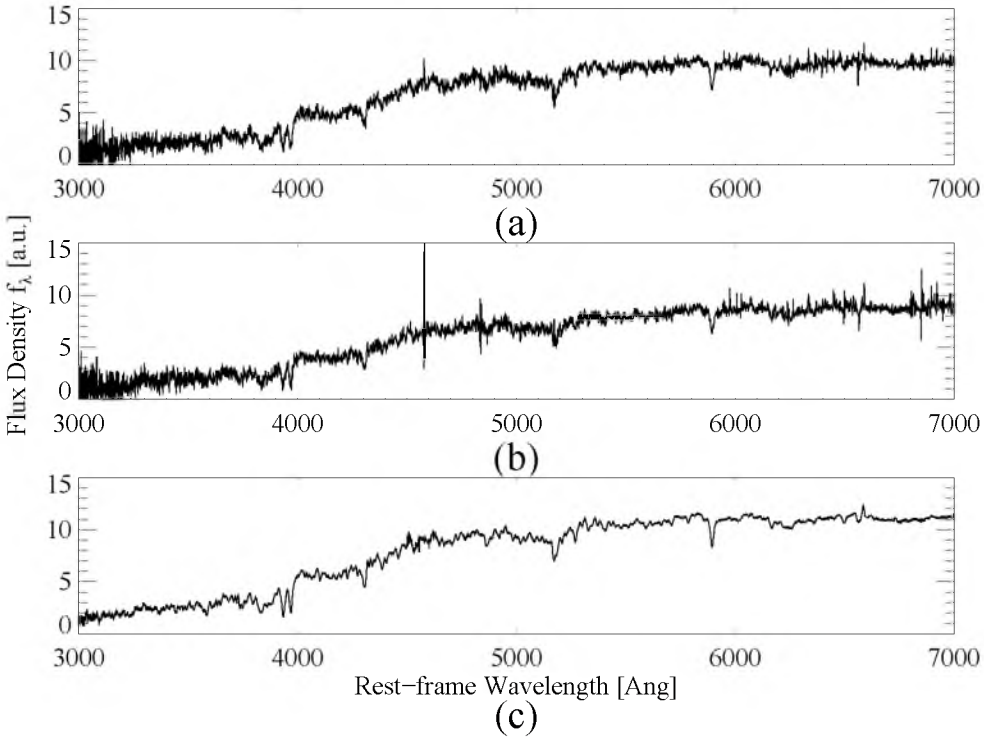

Figure 2.2: Results for stacking of spectra within a single redshift-luminosity bin. Panels (a) and (b) show typical individual spectra, while panel (c) shows the high-SNR stacked spectrum for that bin, resulting from averaging the spectra of ~ 200 galaxies. Credit: Figure 2 in the referred publication "Evolution of the Velocity-dispersion Function of Luminous Red Galaxies: A Hierarchical Bayesian Measurement," Shu, Yiping, et al., 2012, AJ, 143, 90. Reproduced by permission of the AAS.

Making use of the following relation, which can be derived for our log-normal form of Equation (2.15):

$$\begin{aligned} \langle \sigma^n \rangle &= \int \sigma^n \, p(\log_{10}\sigma | m, s) \, \mathrm{d}\log_{10}\sigma \\ &= 10^{[n\,m + n^2\,\ln(10)\,s^2/2]} \end{aligned} \tag{2.18}$$

we find that

$$\sigma^2_{\mathrm{stack}} = 10^{[2\,m + 2\,\ln(10)\,s^2]} \tag{2.19}$$

$$Var(\sigma^2_{\mathrm{stack}}) = \frac{(\sigma^2_{\mathrm{stack}})^2}{N}[10^{4\,\ln(10)\,s^2} - 1] \tag{2.20}$$

Thus we see that the velocity dispersion measured from the stacked spectrum is not given by the mean log-σ value alone, but rather includes a contribution from the width of the population distribution as well. In principle, if a good estimator of $\mathrm{Var}(\sigma^2_{\mathrm{stack}})$ can be obtained, the system can be closed and solved for m and s independently. Although we attempted to estimate $\mathrm{Var}(\sigma^2_{\mathrm{stack}})$ via bootstrap resampling within each bin, we found the treatment of observational errors and varying signal-to-noise ratio among the spectra to be intractable within this framework. Rather than working further from measurements of stacked spectra, we proceed to the hierarchical Bayesian method described in the following section.

2.3.2 Hierarchical Bayesian Approach

To constrain the distribution parameters m and s within each redshift–magnitude bin, we consider the following expansion of the likelihood function $\mathscr{L}(m, s|\{\vec{d}\})$ in the bin:

$$\begin{aligned} \mathscr{L}(m, s|\{\vec{d}\}) &= \Pr(\{\vec{d}\}|m, s) = \prod_i \Pr(\vec{d_i}|m, s) \\ &= \prod_i \int \Pr(\vec{d_i}|\log_{10}\sigma) \Pr(\log_{10}\sigma|m, s) \, \mathrm{d}\log_{10}\sigma \end{aligned} \tag{2.21}$$

Here $\{\vec{d}\}$ is the set of all spectra in the bin, with each element $\vec{d_i}$ representing the spectrum of the i^{th} galaxy. The expression $\Pr(\vec{d_i}|\log_{10}\sigma)$ is related to the $\chi^2(\log_{10}\sigma)$ function by

$$\Pr(\vec{d_i}|\log_{10}\sigma) \propto \exp\left[-\frac{\chi^2_i(\log_{10}\sigma)}{2}\right] \tag{2.22}$$

and $\Pr(\log_{10}\sigma|m, s)$ is given by Equation (2.15). Translating into Bayesian terms, we have a posterior probability for m and s given by

$$\Pr(m, s|\{\vec{d}\}) \propto \Pr(\{\vec{d}\}|m, s) \Pr(m, s) \tag{2.23}$$

with $\Pr(m, s)$ being the prior probability distribution for m and s. For simplicity, we assume a uniform prior on m and s over a reasonable range. In actuality, we find that the likelihood is quite strongly peaked in each bin, so the exact nature and range of the prior are insignificant.

2.3.3 Verification

To verify the correct functioning of our Bayesian approach, we make use of data from BOSS plate 3851. Due to a CCD failure on one of the two BOSS spectrographs that temporarily suspended normal survey operations, 500 of the 1000 targets on this plate were plugged and observed for a total integration time of 7 hours (28 exposures of 15 minutes each) over the course of several nights ending on 2010 April 12, denoted within the SDSS-III database by the modified Julian date (MJD) of 55298. Subsequent to the replacement of the failed CCD, the entire plate was replugged and observed for a more typical BOSS integration time of 1.75 hours total on MJD 55302. The set of reobserved targets allows us to compare BOSS galaxy spectra of standard survey depth with spectra of the same objects at essentially double the nominal survey SNR. We use these repeat observations to verify that our method (1) does not have a signal-to-noise ratio-dependent bias in the estimation of velocity-dispersion distribution parameters, and (2) reproduces the known distribution of velocity dispersions within a controlled sample, as measured from the high-SNR set of spectra.

Between the deep and shallow reobservations, there are 308 galaxies which have equal redshifts (within $\Delta z = \pm 0.005$) and classifications for both observation dates. Since the sample is heterogeneous in magnitude and redshift, we select a subsample with a controlled distribution in velocity dispersion. We restrict our attention to galaxies that have their individual velocity dispersions measured at SNR of 10 or more from the 7-hour observations, and that have redshifts between 0.4 and 0.6. We then select a random subsample of 125 galaxies from this set so as to have a Gaussian histogram in $\log_{10}\sigma$ with a mean of $m = 2.33$ and an intrinsic RMS scatter of $s = 0.07$. The histogram of this subsample, along with the histogram of the same sample as constructed from galaxy-by-galaxy measurements using the 1.75-hour observations, is shown in Figure 2.3.

The frequentist formulas given by Equations (2.19) and (2.20) do not account for observational error, and hence we do not use them to solve for m and s estimates for our relatively low signal-to-noise BOSS survey data. However, our subsample of high signal-to-noise 7-hour observations allows us to test them, which we do before proceeding to the verification of our Bayesian analysis framework. First, we use Equation (2.19) with a mean of 2.33 and an intrinsic scatter of 0.07 to predict a value of $\sigma_{\rm stack} = 219\,{\rm km\,s^{-1}}$, which is in very good agreement with the result of $(222 \pm 12)\,{\rm km\,s^{-1}}$ that we obtain by fitting the stacked spectrum of this set of 125 galaxies directly. Similarly, we predict $[{\rm Var}(\sigma^2_{\rm stack})]^{1/4} = 40\,{\rm km\,s^{-1}}$ from Equation (2.20), which is in reasonable agreement with

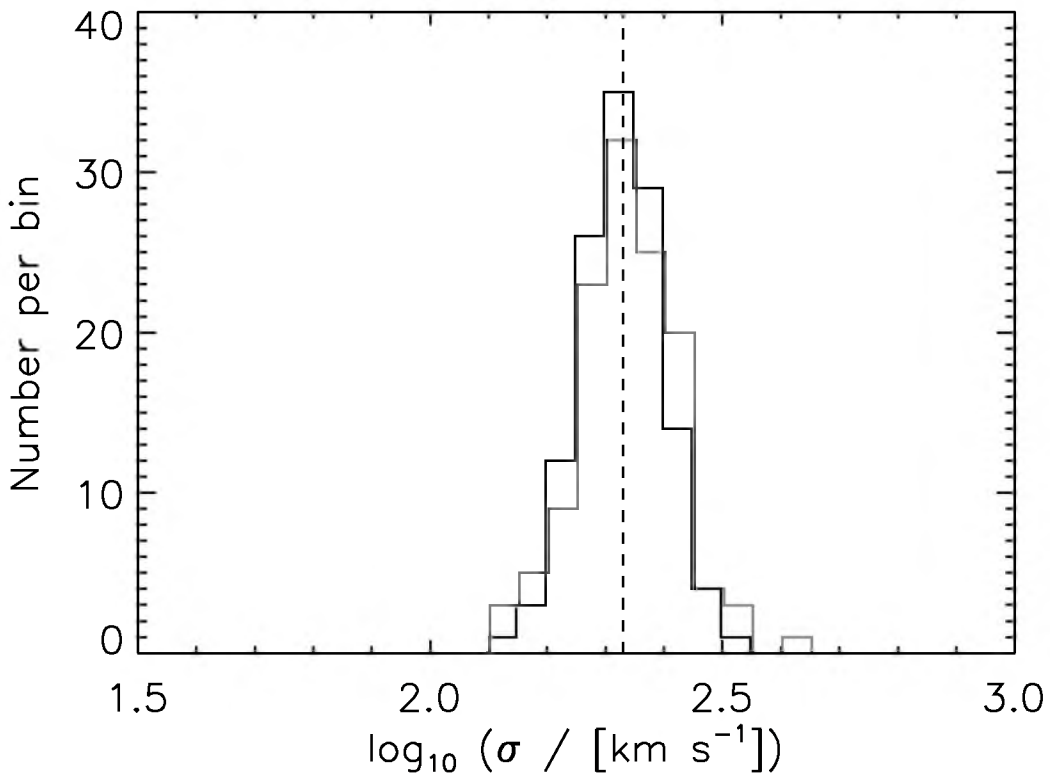

Figure 2.3: Engineered subsample Gaussian histogram in $\log_{10}\sigma$ constructed using measurements from 7-hour BOSS observations *(black)*, with a histogram of the same subsample using velocity dispersions measured from 1.75-hour observations *(gray)*. The two histograms have been given a slight relative horizontal offset, for display purposes. The vertical dashed line indicates the mean $\log_{10}\sigma$ value of 2.33 for the subsample. Note the relative broadening of the 1.75-hour histogram due to the effects of observational error. Credit: Figure 3 in the referred publication "Evolution of the Velocity-dispersion Function of Luminous Red Galaxies: A Hierarchical Bayesian Measurement," Shu, Yiping, et al., 2012, AJ, 143, 90. Reproduced by permission of the AAS.

the value of $46\,\mathrm{km\,s^{-1}}$ obtained through a bootstrap resampling process. In both cases, the agreement is not exact because there is still some observational error even in the 7-hour data, but as mentioned above, we will pass to the Bayesian framework to quantify these effects.

We next carry out the estimation of the m and s parameters of the selected subsample of objects, using the Bayesian approach described above, for both the 7-hour and the 1.75-hour data sets. Figure 2.4 shows the resulting posterior probability density for these parameters as estimated from both data sets. As expected, we see that the posterior PDF is tighter for the 7-hour data. More importantly, we see no significant bias in the posterior PDF between the low-SNR and high-SNR data sets. This is especially significant for the estimation of the s parameter: if we were not handling our observational uncertainties correctly, we might expect to infer a broader intrinsic distribution (higher s value) from the noisier data, but this not the case. We also see that the parameters used to engineer the subsample are recovered with no significant bias in m. We see a slight offset of the 7-hour maximum posterior s value from the input value used to engineer the sample. This is in the direction and of the size to expected given the observational error of the 7-hour individual-spectrum velocity dispersion measurements, which have an RMS signal-to-noise of about 17. This corresponds to an observational broadening of about 0.025 dex in the engineered histogram of Figure 2.3, which is deconvolved by the Bayesian parameter estimation procedure to give the lower recovered s value seen in Figure 2.4.

Another concern is that there might be a systematic bias with redshift, since the spectral regions used by `vdispfit` in fitting for velocity dispersions (rest frame wavelength range from 4,100Å to 6,800Å) move to the redder and noisier parts of the spectrum as the redshift gets higher. In order to test this, we construct another controlled subsample with 152 galaxies of redshift $z < 0.2$ and very high SNR. Then we take the best-fit template combination models of those 152 galaxies returned by `vdispfit` and redshift them to progressively higher redshift bins, giving them a uniform random distribution over a bin width of $\Delta z = 0.04$ in each case (to match our actual binning). At each new redshift, the model spectra are added to sky-subtracted BOSS sky fibers to simulate realistic survey noise, and scaled individually in flux to give a typical median SNR at that redshift bin. We then analyse the simulated redshifted samples with our Bayesian method to estimate the posterior PDF of m and s. The results are shown in Figure 2.5 and Figure 2.6, for 5 separate redshift bins. We see that the recovered parameters are consistent within observational error across all redshifts, with no apparent redshift-dependent bias.

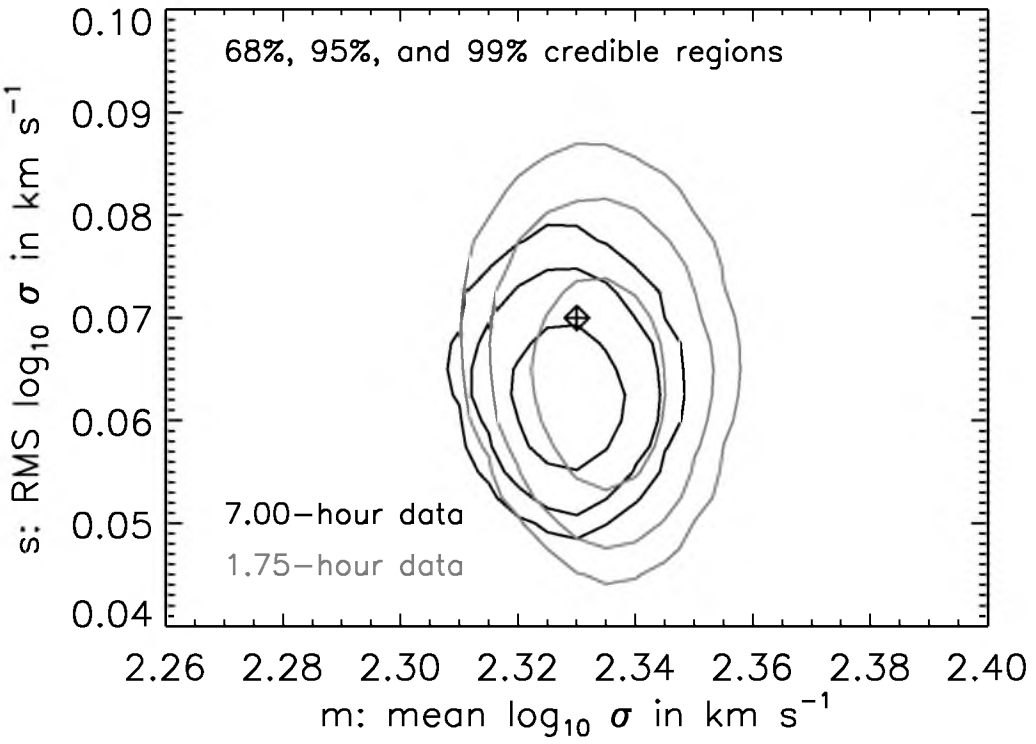

Figure 2.4: Credible-region contours of constant posterior probability density for m and s parameters measured from the engineered test subsample of galaxies observed with both 7-hour integrations *(black)* and 1.75-hour integrations *(gray)*. The symbol is the location of the parameters chosen for the construction of the test subsample. The offset in s between the contours and the symbol is a result of the proper deconvolution of observational uncertainty that is implemented by the Bayesian method. Credit: Figure 4 in the referred publication "Evolution of the Velocity-dispersion Function of Luminous Red Galaxies: A Hierarchical Bayesian Measurement," Shu, Yiping, et al., 2012, AJ, 143, 90. Reproduced by permission of the AAS.

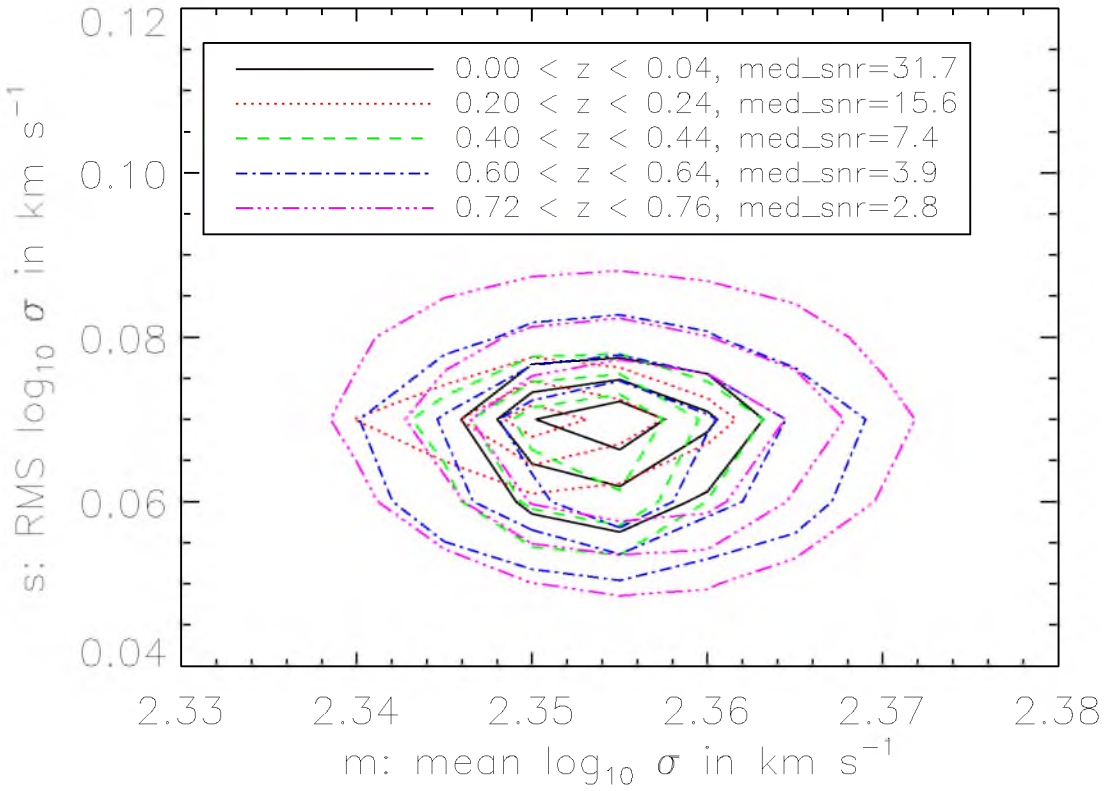

Figure 2.5: Contours of constant posterior probability density (68%, 95%, and 99%) for m and s parameters obtained from a controlled subsample of 152 galaxies in 5 different redshift bins with gradually reduced SNRs. Credit: Figure 5 in the referred publication "Evolution of the Velocity-dispersion Function of Luminous Red Galaxies: A Hierarchical Bayesian Measurement," Shu, Yiping, et al., 2012, AJ, 143, 90. Reproduced by permission of the AAS.

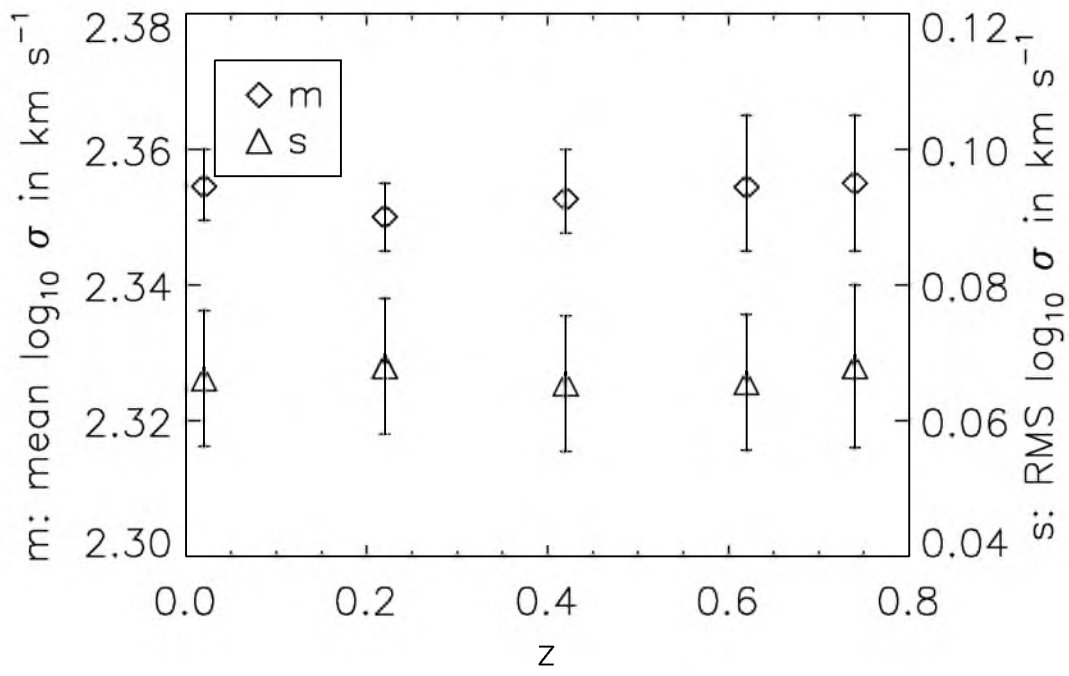

Figure 2.6: The best estimated m (Diamond) and s (Triangle) values for the controlled subsample of 152 galaxies at 5 different redshift bins. Credit: Figure 6 in the referred publication "Evolution of the Velocity-dispersion Function of Luminous Red Galaxies: A Hierarchical Bayesian Measurement," Shu, Yiping, et al., 2012, AJ, 143, 90. Reproduced by permission of the AAS.

Finally, to rule out any significant dependence of our measurement on airmass and fiber position within the BOSS spectrographs, we make use of data from plates 3615, 3647, 4238, and 4239. These four plates cover roughly the same area of sky, but with different plate drillings that place the same objects in very different fibers within the spectrograph system. They were also observed over a range of different airmasses on multiple nights. From these plates, we construct several subsamples of spectra, all of which include the same galaxies, but are drawn from different plates and/or observations. As with the previous tests, we recover consistent estimates of m and s from the analysis of all these samples.

Based on the above three tests, we conclude that our method recovers accurate estimates of the population velocity-dispersion distribution parameters.

2.3.4 Magnitude Error Correction

Our method of determining $\Pr(\vec{d}_i|\log_{10}\sigma)$ incorporates an explicit marginalization over redshift error, and propagates all observational uncertainty in the velocity dispersion of a given galaxy. Our binning in redshift and absolute magnitude introduces additional error possibilities that we must account for. In the case of redshift, the errors are negligible relative to the bin width of $\Delta z = 0.04$, and are unlikely to contribute any artificial broadening to our determination of the redshift dependence of m and s. The absolute magnitude errors are, however, nonnegligible in comparison to the magnitude bin width of $\Delta \mathrm{M_V} = 0.1$, and thus we use the following technique to estimate and compensate for the broadening effect of the observational scattering of galaxies between absolute-magnitude bins (see Figure 2.1).

Suppose (m, s) are the true values within a bin, and (m_1, s_1) are the values that we determine in the presence of absolute-magnitude errors. We assume that

$$m_1 = m + \delta m \tag{2.24}$$

$$s_1^2 = s^2 + \delta s^2 \tag{2.25}$$

where δm and δs are the biases introduced by magnitude errors. To estimate and remove these biases, we add additional random errors to all our galaxy absolute magnitudes M_V to give

$$M_V' = M_V + \epsilon \delta M_V \tag{2.26}$$

where ϵ is a normally distributed random number with mean 0 and standard deviation 1, and δM_V are the galaxy-by-galaxy absolute-magnitude errors estimated by `sdss2bessell` (propagated from SDSS *ugriz* apparent magnitude errors). We repeat our analysis, binning instead in M_V', and denoting the new distribution parameter results by m_2 and s_2. We assume these new determinations are related to m_1 and s_1 in the same way as m_1 and s_1

are related to m and s, which implies that

$$m_2 = m + 2\,\delta m \tag{2.27}$$

$$s_2^2 = s^2 + 2\,\delta s^2 \tag{2.28}$$

Thus the biases due to absolute magnitude errors δm and δs can be removed to yield

$$m = 2\,m_1 - m_2 \tag{2.29}$$

$$s = \sqrt{2\,s_1^2 - s_2^2} \tag{2.30}$$

In practice, we find typical values for δm of 0.01, and for δs of 0.04.

2.4 Results: Evolution of the Velocity-Dispersion Function

In this section, we present the results of the application of our hierarchical Bayesian velocity-dispersion distribution measurement technique to the approximately 103,000 galaxies in our LOWZ sample and 330,000 galaxies in our CMASS sample.

2.4.1 LOWZ Sample

The LOWZ sample extends to $z \approx 0.5$. The two-dimensional contour plots of m and s (Figure 2.7) and scatter plots in different redshift bins (Figure 2.8) show that the mean m is strongly correlated with absolute magnitude, while the intrinsic scatter s shows no significant variation. Tracks of constant stellar mass assuming the LRG stellar population model of Maraston et al. (2009) have also been overplotted in Figure 2.7, and used to convert from an absolute-magnitude to a stellar-mass baseline in Figure 2.8. Galaxies in the LOWZ sample have estimated stellar masses between approximately $10^{11} M_{\odot}$ and $10^{12} M_{\odot}$.

To quantify the variation of the m and s parameters with redshift and absolute magnitude, we consider a simple model specified by:

$$m^0 = \mathrm{A_m^0 M_V + B_m^0 \log_{10}(1+z)} \quad + C_m^0 \tag{2.31}$$

$$s^0 = \mathrm{A_s^0 M_V + B_s^0 \log_{10}(1+z)} \quad + C_s^0 \tag{2.32}$$

with the "0" superscript denoting the LOWZ sample specifically. Performing a linear least squares fit to the individual bin data points, we obtain

$$A_m^0 = -0.0880 \pm 0.0012 \qquad A_s^0 = 0.006 \pm 0.002$$

$$B_m^0 = -0.087 \pm 0.018 \qquad B_s^0 = -0.08 \pm 0.03$$

$$C_m^0 = 0.37 \pm 0.02 \qquad C_s^0 = 0.20 \pm 0.04 \tag{2.33}$$

We can translate the resulting scaling into the standard form for the FJR, with luminosity $L \propto \sigma^x$ by recognizing that $x = -0.4/A_m^0$. The resulting value of $x = 4.55 \pm 0.06$ is in reasonable agreement with the canonical local-universe value of $x = 4$. Thus, BOSS

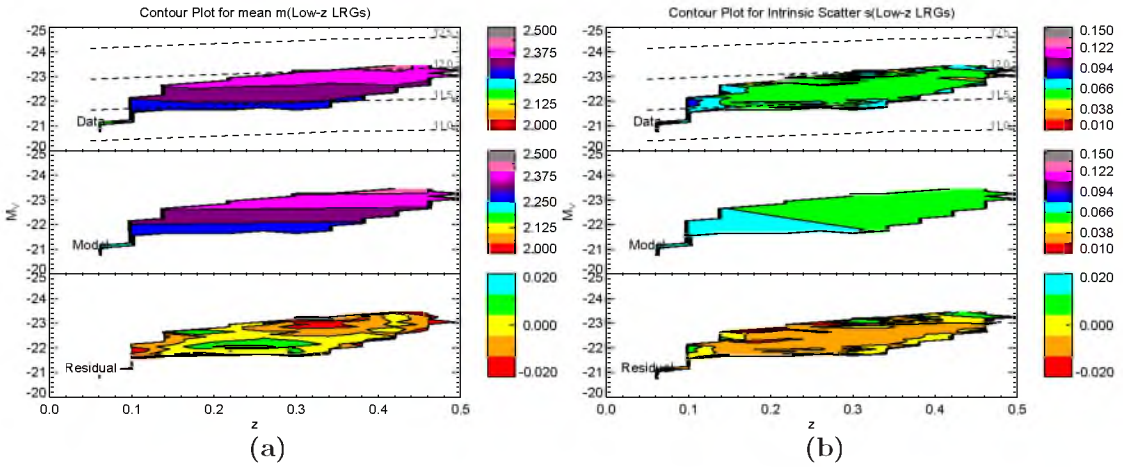

Figure 2.7: Contour plots of m & s for LOWZ sample galaxies. Top panels in (a) and (b) show the map of maximum posterior probability density functions of m and s, respectively, over the range of the plane with bins containing at least 100 galaxies. Middle panels show low-order bivariate model fits to these maps constructed as described in the text, and residuals (top minus middle) are shown in the bottom panels. Dashed lines in the top panels show tracks of constant stellar mass from the LRG population model of Maraston et al. (2009). Credit: Figure 7 in the referred publication "Evolution of the Velocity-dispersion Function of Luminous Red Galaxies: A Hierarchical Bayesian Measurement," Shu, Yiping, et al., 2012, AJ, 143, 90. Reproduced by permission of the AAS.

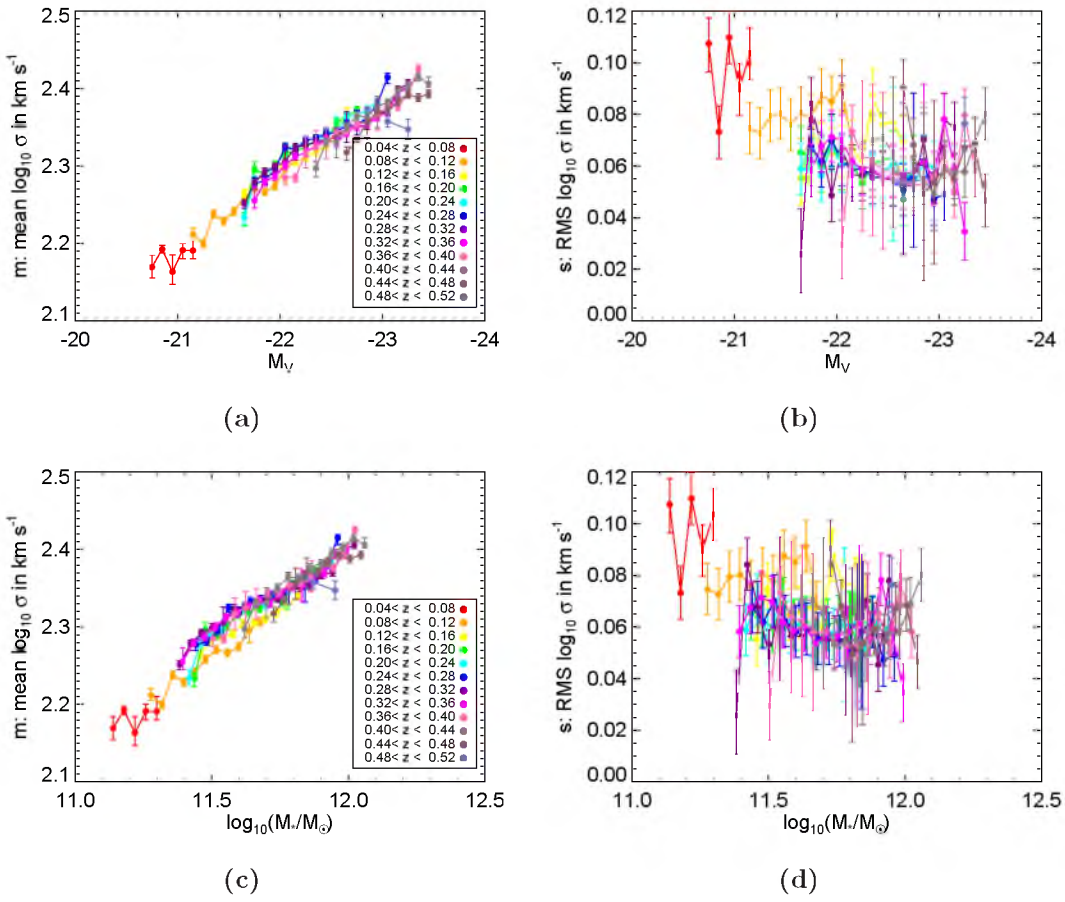

Figure 2.8: Scatter plots of m & s versus M_V (panels (a) and (b)) and $\log_{10}(M_*/M_\odot)$ (panels (c) and (d)) for LOWZ sample galaxies in different redshift ranges. Credit: Figure 8 in the referred publication "Evolution of the Velocity-dispersion Function of Luminous Red Galaxies: A Hierarchical Bayesian Measurement," Shu, Yiping, et al., 2012, AJ, 143, 90. Reproduced by permission of the AAS.

LOW-Z LRGs define an FJR whose slope and scatter has little dependence on redshift and luminosity; there is correspondingly little evidence for dynamical evolution in this sample since roughly z = 0.5.

2.4.2 CMASS Sample

The CMASS galaxy sample extends from $z \approx 0.3$ to $z \approx 0.8$. The results of our σ distribution parameter measurements are shown in Figures 2.9 and 2.10, once again using tracks of constant stellar mass based on the Maraston et al. (2009) population model. Using the same model form as used for the LOWZ sample above,

$$m^1 = \mathrm{A_m^1 M_V + B_m^1 \log_{10}(1+z)} \quad + C_m^1 \tag{2.34}$$

$$s^1 = \mathrm{A_s^1 M_V + B_s^1 \log_{10}(1+z)} \quad + C_s^1 \tag{2.35}$$

(with the "1" superscript denoting the CMASS sample specifically), and again doing a linear least-squares fit, we find that

$$A_m^1 = -0.1128 \pm 0.0010 \qquad A_s^1 = 0.0263 \pm 0.0016$$

$$B_m^1 = -0.77 \pm 0.02 \qquad B_s^1 = 0.82 \pm 0.04$$

$$C_m^1 = -0.089 \pm 0.019 \qquad C_s^1 = 0.52 \pm 0.03$$

In the case of the CMASS sample, the FJR is still apparent, but the scaling exponent in $L \propto \sigma^x$ is now $x = 3.55 \pm 0.03$. This observation that the FJR becomes "shallower" at higher redshift can be interpreted in terms of mass-dependent star-formation history (e.g., Cowie et al., 1996; di Serego Alighieri et al., 2005), with less massive (lower σ) galaxies having undergone more recent star formation and thus fading more rapidly with cosmic time relative to more massive galaxies.

There is a clear evolution in the zero-point of the m versus M_V relation (upper left panel in Figure 2.10) with redshift. This evolution is essentially eliminated in the lower left panel of Figure 2.10, which translates to a baseline of constant stellar mass. Hence, the evolution of the m versus M_V relation in the CMASS sample is consistent with passive stellar evolution.

It can easily be seen from Figures 2.9 and 2.10 that s is no longer constant with redshift at fixed luminosity or stellar mass. The significance of this result is encapsulated in the nonzero value of $B_s^1 = 0.82 \pm 0.04$ given above. To quantify this result in more detail, we fit the s versus M_V relation with a linear model at each redshift bin, and plot the zero-point of this relation as a function of redshift in Figure 2.11. We see that within the CMASS sample, the intrinsic width s of the velocity-dispersion function at fixed magnitude or stellar mass *decreases* with cosmic time (i.e., broader distribution width at higher redshift), especially

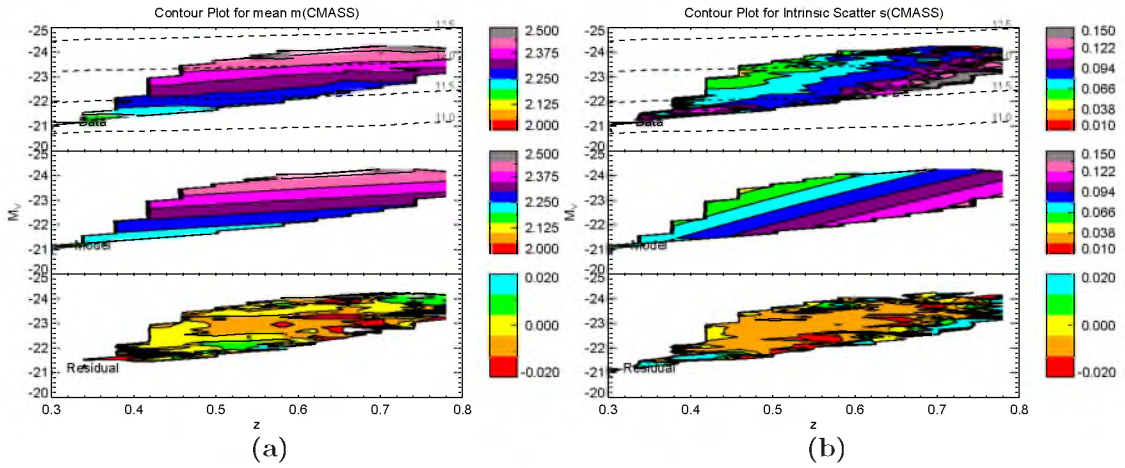

Figure 2.9: The same as Figure 2.7 but for CMASS galaxies. Credit: Figure 9 in the referred publication "Evolution of the Velocity-dispersion Function of Luminous Red Galaxies: A Hierarchical Bayesian Measurement," Shu, Yiping, et al., 2012, AJ, 143, 90. Reproduced by permission of the AAS.

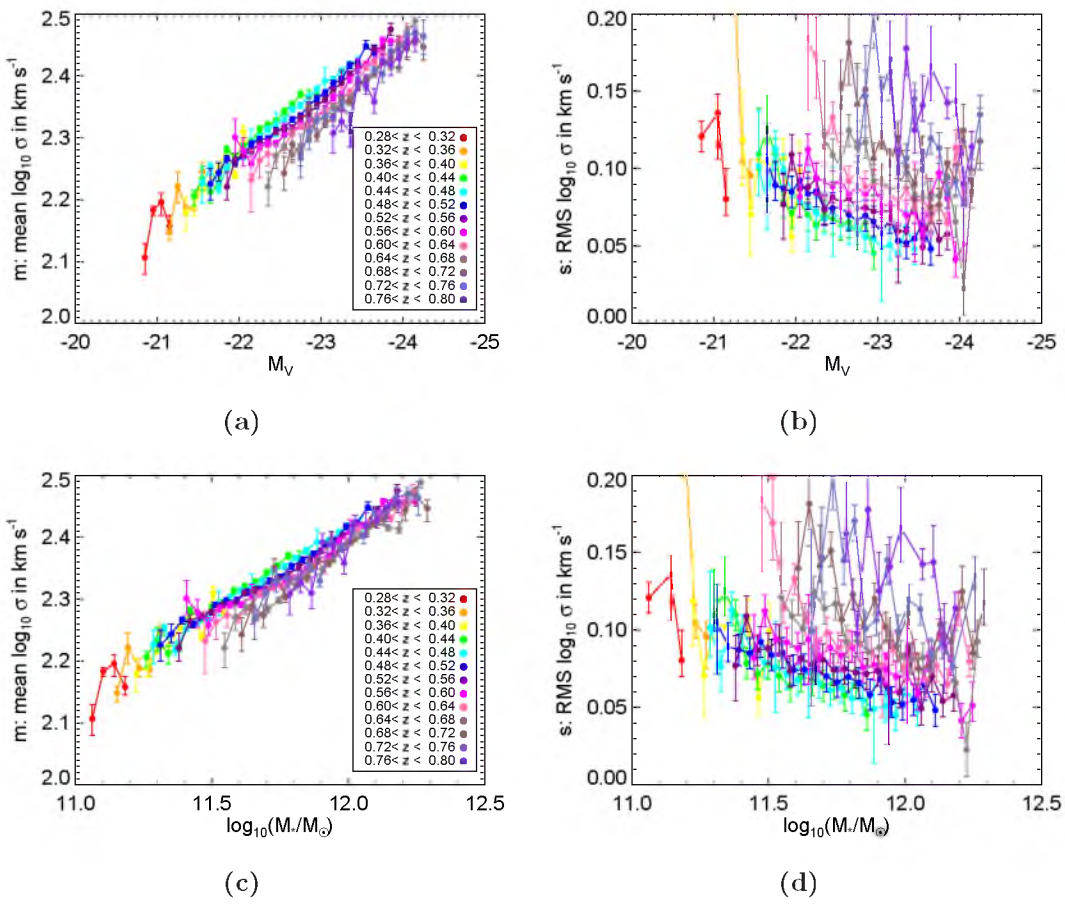

Figure 2.10: The same as Figure 2.8 but for CMASS galaxies. (Note that the scales of these panels are expanded relative to Figure 2.8.) The increase of intrinsic scatter with redshift can be seen in the right-hand figure. Credit: Figure 10 in the referred publication "Evolution of the Velocity-dispersion Function of Luminous Red Galaxies: A Hierarchical Bayesian Measurement," Shu, Yiping, et al., 2012, AJ, 143, 90. Reproduced by permission of the AAS.

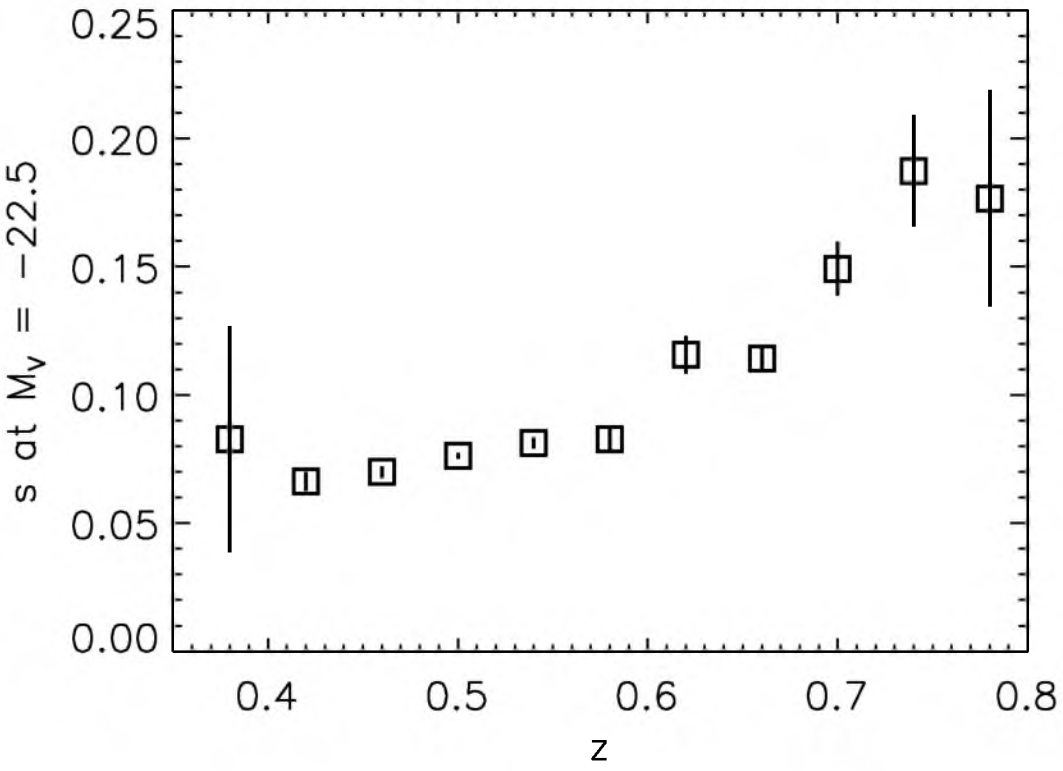

Figure 2.11: Variation of the intrinsic width s of the CMASS population distribution in $\log_{10}\sigma$ as a function of redshift. Credit: Figure 11 in the referred publication "Evolution of the Velocity-dispersion Function of Luminous Red Galaxies: A Hierarchical Bayesian Measurement," Shu, Yiping, et al., 2012, AJ, 143, 90. Reproduced by permission of the AAS.

at redshifts $z > 0.6$. This is consistent with our tentative detection of evolution in the FJR slope between the LOWZ and CMASS samples, in the sense that a given range in luminosity encompasses a larger range of velocity dispersions at higher redshift, but the signal is too large to be explained by this effect alone (since the FJR slope is not seen to evolve significantly *within* the CMASS sample alone). We are therefore seeing increased dynamical heterogeneity at fixed luminosity in the CMASS sample at higher redshifts.

We note that the apparent increase in the intrinsic σ distributions at high redshift cannot be explained in terms of surface-brightness selection effects. Through the FP or Kormendy relations, velocity dispersion at fixed luminosity is correlated with surface brightness. At the high-redshift end of the CMASS sample, we can expect a degree of incompleteness at both ends of the surface-brightness distribution. On the one hand, relatively low surface-brightness galaxies will have fainter magnitudes within the BOSS spectroscopic fiber, and will thus be less likely to be targeted, and less likely to have confident and correct spectroscopic redshift measurements even if targeted. On the other hand, relatively high surface-brightness galaxies (again, at fixed luminosity) run the risk of being unresolved in star–galaxy separation. Consequently, we might expect the distribution of velocity dispersion at fixed magnitude to be made more narrow at high redshift by these considerations, which goes in the opposite sense to the trend we detect.

2.4.3 Application to Individual Spectra

Our results characterize the dynamical properties of the population of LRGs targeted by BOSS. The parameters of the population can in turn be used to inform our estimates of the velocity dispersion values of individual noisy BOSS spectra. For this application, we want to use distribution parameters *uncorrected* for broadband magnitude errors, since these same errors will be present in the photometric data for the individual galaxies whose spectra we wish to analyze.

For LOW-Z LRGs, without magnitude error correction, we have

$$m^0 = -0.0829 M_V - 0.042 \log_{10}(1+z) + 0.48 \tag{2.36}$$

$$s^0 = 0.006 M_V - 0.09 \log_{10}(1+z) + 0.22 \tag{2.37}$$

and for CMASS galaxies, without magnitude error correction, we have

$$m^1 = -0.0973 M_V - 0.60 \log_{10}(1+z) + 0.23 \tag{2.38}$$

$$s^1 = 0.0240 M_V + 0.76 \log_{10}(1+z) + 0.49 \tag{2.39}$$

We can then take the posterior probability $\Pr(m, s|\{\vec{d}\})$ from the entire sample as a prior probability for the analysis of an individual galaxy spectrum. The posterior probability for

$\log_{10}\sigma$ of the spectrum is then

$$\Pr(\log_{10}\sigma|\vec{d_i}) \propto \Pr(\vec{d_i}|\log_{10}\sigma)\Pr(m,s|\{\vec{d}\}) \tag{2.40}$$

Loosely speaking, if the observational error in the velocity dispersion measured from a single spectrum is comparable to the intrinsic width s of the particular population from which it is drawn, then the data and the prior will contribute equally to the determination of the posterior PDF of $\log_{10}\sigma$. If the observational error is small, the effect of the prior will be correspondingly minor, while if the observational error is large, the posterior PDF will be determined primarily by the prior.

The application of this method can thus permit a more precise σ estimate for individual galaxies, by making use of the collective information about the population from which it was drawn. It is important, however, to note that if the spectra under consideration are somehow selected to be biased towards either higher or lower velocity dispersions, then the prior will pull them systematically towards the population mean, giving posterior PDFs that are biased relative to the true σ values. We must also be sure only to apply this method to subsamples of spectra that are much smaller than the population samples used to determine the distribution parameters.

2.5 Discussion and Conclusion

In this paper, we have presented a new technique for estimating the velocity-dispersion function of LRGs from large numbers of low SNR spectra. This method incorporates the effects of observational uncertainties in spectroscopic redshift, velocity dispersion, and broadband magnitude. We have compared our method favorably to the more traditional approach of "stacking" multiple spectra; our new approach can perhaps be termed "Bayesian stacking." We have also indicated how the results of our method can be used as informative priors to provide more precise estimates of the velocity dispersions of individual galaxies, provided that those galaxies are an unbiased selection from the parent distribution at their particular redshift and luminosity.

We have applied our technique to a sample of 430,000 galaxy spectra from the BOSS project of the SDSS-III, covering the redshift range from zero to unity, concentrated between approximately $z = 0.2$ and $z = 0.8$. For the higher-redshift CMASS target sample (approximately 76% of our galaxies), we detect a highly significant increase in the intrinsic width of the velocity-dispersion distribution at higher redshifts, indicative of greater galaxy diversity at fixed luminosity at earlier cosmic times. For the lower-redshift LOWZ galaxy sample, we find little evolution in the velocity-dispersion distribution below $z \approx 0.5$. Although

the CMASS and LOWZ samples do not form a single uniform sample (LOWZ galaxies being generally more luminous than CMASS targets over the range of redshift where the two overlap), our results suggest that dynamical evolution of massive LRGs is much more significant over the interval $0.5 < z < 1.0$ as compared to $0 < z < 0.5$.

Future applications of this method to the BOSS galaxy samples will focus on the effects of observational selection on the deduced population evolution. We also plan to divide our analysis further by rest-frame color, so as to differentiate between galaxies of different stellar population at a given redshift and magnitude. By making a more accurate division of the sample in terms of stellar mass and star-formation history, we hope to separate the signatures of dynamical and stellar-population evolutionary channels, and to thereby obtain a more detailed picture of LRG population evolution and a more powerful discriminant between theoretical scenarios. This approach can also determine whether the effect of increased population scatter in $\log_{10}\sigma$ at high redshift is due to greater dynamical diversity, greater stellar-population diversity, or to some combination of the two effects.

Our measurements can also have important implications for the statistics of gravitational lensing, by constraining the total lensing cross-section in massive elliptical galaxies between redshift 0 and 1. Although a precise application to gravitational-lensing statistics must await a proper treatment of completeness, our current results can be combined with published luminosity functions (e.g., Cimatti et al., 2006; Cool et al., 2008) to place a lower limit on the integrated lensing cross-section.

The application of hierarchical Bayesian methods such as the one presented here may hold the key to reconciling the tension between redshift surveys designed for constraining cosmological parameters and those designed for the study of galaxy evolution. The former goal generally dictates an SNR just sufficient to measure redshift for as many galaxies over as large a volume of the universe as possible, while the latter goal traditionally requires observations at high enough SNR to precisely constrain multiple physical parameters for each galaxy. However, if the ultimate goal of galaxy-evolution studies is to measure the distribution of physical parameters within a statistically significant sample of galaxies, then Bayesian methods can remove the need to measure those parameters precisely on a galaxy-by-galaxy basis. In fact, there may indeed be an objective galaxy-evolution case for trading fewer high-SNR spectra for more low-SNR spectra, so as to reduce the effects of sample variance. If cosmological experimental designs can also accommodate the more permissive (e.g., magnitude-limited) targeting desired for galaxy population studies, then both goals may be well-served by the same redshift survey.

CHAPTER 3

THE SLOAN LENS ACS SURVEY. XIII. EXTENDING STRONG LENSING TO LOWER MASSES[†]

We present observational results from a new Hubble Space Telescope (HST) Snapshot program to extend the methods of the Sloan Lens ACS (SLACS) Survey to lower lens-galaxy masses. We confirm and model 40 new galaxy-scale strong lenses from this program, which we supplement with 58 lenses previously discovered by SLACS. In addition, we determine upper limits to the masses of an additional 33 galaxies (18 new and 15 from legacy SLACS data) based on single images of background galaxies without detectable lensed counter-images. Incorporating these lensing measurements and upper limits together in a single statistical analysis, we find a significantly less-than-unity slope of 0.78 ± 0.05 for the $\log_{10}\sigma_*$-$\log_{10}\sigma_{\rm SIE}$ relation, which corresponds to an evidence at 4σ that the total mass-density profile of early-type galaxies (parameterized by the ratio of stellar to lensing velocity dispersions) varies systematically within the population in the sense of having shallower profiles at larger lens-galaxy velocity dispersions. This trend provides evidence of variation in the efficiency of dissipative baryonic processes as a function of galaxy mass. The trend is only evident when upper limits are incorporated into the analysis, highlighting the importance of including both "lenses" and "nonlenses" for an unbiased treatment of the lens population when extending to mass ranges with lower lensing cross-sections. By scaling simple stellar population models to the *HST* I-band data under a variety of assumptions, we identify a strong trend of increasing dark-matter fraction at higher velocity dispersions, which can alternatively be interpreted as a trend in the stellar initial mass function (IMF). Consistent with previous finding and the suggestion of a non-universal IMF, we find that a Salpeter IMF is ruled out for galaxies below $10^{10.8}M_\odot$ within one half of the half-light radii, while it is acceptable above this mass. Considered together, our mass-profile and dark-matter-fraction trends with increasing galaxy mass could both be explained by an

[†]Submitted to the Astrophysical Journal

increasing relative contribution on kiloparsec scales from a dark-matter halo with a spatial profile more extended than that of the stellar component.

3.1 Introduction

Early-type galaxies (ETGs), classified by their morphology, compose one of the two main categories of galaxies (Hubble, 1926, 1936b). Although considered to be relatively *"dead"* and *"featureless"* as a consequence of their little star formation activities and smooth light distributions, ETGs play a crucial role in studying the evolution of galaxies, the nature of dark matter, and cosmology. Being the endproducts of hierarchical merging scenario (Toomre and Toomre, 1972; White and Frenk, 1991; Kauffmann et al., 1993a; Cole et al., 2000), their structures, properties, and formation histories can be used as a compelling test of the Λ Cold Dark Matter (ΛCDM) paradigm. Additionally, ETGs can be extremely luminous and therefore can be used as powerful cosmological tracers of the large-scale structure (Eisenstein et al., 2005a; Percival et al., 2007; Anderson et al., 2012).

However, the formation and evolution of ETGs are still puzzling and further investigations are highly demanded. Concerning the mass-density profile of ETGs, N-body DM-only numerical simulations have revealed a somewhat "universal" density profile with a r^{-1} inner profile and a r^{-3} drop-off at large radii, independent of the halo mass (Navarro et al., 1996, 1997). Later on, various observations of DM-dominated galaxies yield inconsistent inner density slopes with numerical simulations (Moore et al., 1999b; Graham et al., 2006; Navarro et al., 2010), the tension of which can be loosened by taking baryonic physics into account. Gas cooling permits baryons to condense in the central regions of galaxies, and therefore makes the mass distribution more centrally concentrated (e.g. Gnedin et al., 2004; Gustafsson et al., 2006; Abadi et al., 2010; Velliscig et al., 2014). Heating due to dynamical friction and supernovae (SN)/Active Galactic Nucleus (AGN) feedback, on the other hand, can soften the central density concentration (e.g. Nipoti et al., 2004; Romano-Díaz et al., 2008; Governato et al., 2010; Duffy et al., 2010; Martizzi et al., 2012; Velliscig et al., 2014). The strength of these competing effects differs from galaxy to galaxy and hence studying the dependences of the shape of the mass-density profile in the central region on galaxy mass, redshift, and other structural quantities unravels the formation and evolution of ETGs.

The stellar initial mass function (IMF) is an empirical relation quantifying the relative fraction of stars as a function of the stellar mass at the time when the whole population formed. Salpeter (1955) first quantified the IMF as a simple power-law function using main-sequence stars in the solar neighborhood. Later on, various modifications have been

considered at the low-mass end and the most commonly used forms are the Kroupa IMF (Kroupa, 2001) and the Chabrier IMF (Chabrier, 2003). Apparently, the IMF of a galaxy should depend on the environmental properties of the molecular cloud it originated from such as metallicity, temperature, and density, and is therefore non-universal. Having knowledge of the form and the variation of the IMF provides deep insights in understanding the role of the environment during star formation and galaxy evolution processes. Recently, several pieces of evidence convince that the IMF indeed varies (e.g. van Dokkum and Conroy, 2010; Strader et al., 2011; Cappellari et al., 2012; Sonnenfeld et al., 2012; Spiniello et al., 2012; Ferreras et al., 2013; La Barbera et al., 2013; Conroy et al., 2013; Tortora et al., 2013; Brewer et al., 2014; Spiniello et al., 2014).

Strong gravitational lensing (GL) has its unique power among the many techniques for the study of ETGs. As a pure gravity-dependent effect, GL provides highly accurate measurements of total mass that are robust against different models and assumptions about galaxy properties. Therefore, it provides the best estimation of the total projected mass within the so-called *Einstein radius* enclosed by the lensed images of the background object. Various lensing surveys have been conducted in the past decade and led to numerous important results. The Lenses Structure and Dynamics (LSD) Survey aimed to measure the stellar kinematics of a small sample of E/S0 galaxy lenses and combine it with GL to constrain the central mass distribution (Koopmans and Treu, 2002, 2003a; Treu and Koopmans, 2002, 2004). The SLACS survey (Bolton et al., 2006; Treu et al., 2006; Koopmans et al., 2006b; Gavazzi et al., 2007; Bolton et al., 2008a; Gavazzi et al., 2008; Bolton et al., 2008b; Treu et al., 2009; Auger et al., 2009, 2010; Newton et al., 2011) is by far the most productive survey for galaxy-scale strong lenses with known lens and source redshifts, with a discovery of over 90 spectroscopically-selected lenses confirmed by high-resolution Hubble Space Telescope (*HST*) follow-up. SLACS observes relatively low-redshift ($z_L \lesssim 0.4$) ETG lens candidates selected from the Luminous Red Galaxy (LRG, Eisenstein et al., 2001b) and MAIN (Strauss et al., 2002) galaxy samples of the Sloan Digital Sky Survey (SDSS, York et al., 2000). The SLACS survey has yielded multiple novel results on the structure and dynamics of ETGs, which are detailed in the previous papers of this series. Recently, the technique of spectroscopic lens selection has been extended to earlier cosmic time (higher redshift) by the BOSS Emission-Line Lens Survey (BELLS Brownstein et al., 2012), which has confirmed 25 strong lenses ($0.4 \lesssim z_L \lesssim 0.7$) using data from the Baryon Oscillation Spectroscopic Survey (Dawson et al., 2013) of the SDSS-III (Eisenstein et al., 2011a).

The SLACS sample is a unique resource for studies of the structure of ETGs, but

it has been significantly biased toward the high-mass end due to several related factors. First, strong lensing cross section (an approximation of the lensing possibility in general) increases with the lens galaxy mass, so high-mass ETGs are more likely to act as strong lenses. Second, even if a low-mass galaxy acts as a strong lens, the characteristic angular separation of the lensed images will be small and hard to resolve even at space-based imaging resolution. Third, low-mass galaxies can be intrinsically too faint to be selected for SDSS spectroscopy. Fourth, for the preceding reasons, high-mass SLACS candidates have previously been prioritized for *HST* follow-up, in order to maximize the survey success rate.

In order to extend the power of strong lensing to low-mass galaxies, an extension of the SLACS survey known as "SLACS for the Masses" (hereafter S4TM, *HST* Snapshot Program 12210) was initiated in April 2012 with a focus on lens candidates with lower masses and smaller predicted Einstein radii as compared to SLACS lenses. While the lensing confirmation rate of S4TM is lower than that of previous SLACS *HST* programs, it importantly achieves a wider lens-mass baseline in combination with previous SLACS lenses. We refer the readers to the catalog paper by Brownstein et al. 2014 (in preparation) for a full description of the S4TM program details. In this paper, we present the first scientific results on the total mass-density profile and dark-matter content of an extended ETG sample combining the S4TM lenses and previous SLACS lenses. We use a hierarchical Bayesian method to infer the mass-profile scaling relation of the combined lens sample, and estimate stellar masses through single stellar-population (SSP) model scalings to the observed *HST* photometry.

This chapter is organized as follows. In Section 3.2, we briefly describe our lens identification technique using the SDSS spectroscopy and high-resolution imaging data observed by the *HST*. Section 3.3 describes our parametric lens-modeling technique. We then derive the main findings in Sections 3.4 and 3.5 with regards to the study of the mass-density profile and dark-matter fraction of the ETGs. Discussion and conclusions are presented in Section 3.6. Throughout the paper, we assume a standard cosmology with $\Omega_{\rm m} = 0.274$, $\Omega_\Lambda = 0.726$ and $\rm H_0 = 70\,km\,s^{-1}\,Mpc^{-1}$ (WMAP7, Komatsu et al., 2011).

3.2 Lens Candidate Identification

The S4TM survey is a snapshot program designed to extend strong gravitational lensing observations toward lower masses and relatively smaller Einstein radii as compared to previous SLACS programs. Using the same lens searching technique as SLACS, 137 lens candidates were identified from the seventh data release (DR7) of the SDSS (Abazajian

et al., 2009) and awarded as *HST* snapshot targets in Observing Cycle 18. The details about the lens selection technique can be found in Bolton et al. (2006, 2008a) and Brownstein et al. 2014 (in preparation). The basic approach is to search for high-redshift emission lines such as [OII] doublets, Hβ, and [OIII] superimposed on the spectra of SDSS target galaxies at lower redshifts. Such emission lines, associated with star-forming galaxies more distant along the same line of sight, indicate the presence of a candidate lensing system, and also allow us to simultaneously determine the redshifts of the background objects.

Between 2010 September and 2012 June, 118 out of 137 candidates were successfully observed with an exposure time of 420 s each with a single exposure through the F814W filter of the Wide Field Channel (WFC) of the Advanced Camera for Surveys (ACS). The images were visually inspected using the *ACSPROC* software, a GUI tool implemented by Brownstein et al. (2012). By searching for lensed features in the b-spline-subtracted residual images (Bolton et al., 2006, 2008a), we have confirmed 40 strong gravitational lenses with clear and definite multiple lensed images or even complete Einstein rings (classified as "grade-A"), 8 systems with suspect multiple images due to various limitations, including low signal-to-noise ratio (SNR), spiral galaxies as the lenses, and lens galaxies being in a group environment (classified as "grade-B"), as well as 18 systems showing clear images of the background objects but no clear counter-images (classified as "grade-C"). Note that we only consider lens candidates that are isolated ETGs for consistency. Grade-B lenses are therefore excluded from this work for now. We also exclude one grade-C lens (SDSSJ1310 + 0220) from now on as it turned out to be a face-on late-type galaxy with strong emission lines after an examination of its SDSS spectrum. Tables in Brownstein et al. 2014 (in preparation) give a summary of all the lens-galaxy properties; images of these systems can be found in the same paper.

3.3 Lens Modeling and Sample Definition

For the foreground lens galaxies, we consider a singular isothermal ellipsoid (SIE) lens model (Kassiola and Kovner, 1993; Kormann et al., 1994; Keeton and Kochanek, 1998; Bolton et al., 2008a) that is generalized from a singular isothermal sphere (SIS) model in which the 2D surface mass density falls off as R^{-1}, but consists of elliptical iso-density contours specified by position angle $P.A.$ and minor-to-major axis ratio $q_{\rm SIE}$. We do not include external shear in our lens model as it has been shown by Koopmans et al. (2006b), Treu et al. (2009), and confirmed again using our grade-A lens sample to be a minor effect (a few percent). The SIE lens model is characterized by the lensing strength $b_{\rm SIE}$

(a.k.a. Einstein radius, specified according to an "intermediate axis" convention for elliptical models) defined as

$$b_{\rm SIE} = 4\pi \frac{\sigma^2_{\rm SIE}}{c^2} \frac{d_{LS}}{d_S} \tag{3.1}$$

where d_{LS} and d_S are the angular diameter distances from the lens and the observer to the source, respectively. $\sigma_{\rm SIE}$ is a characteristic velocity-dispersion parameter of the lens galaxy, which is related to the total enclosed mass within the Einstein radius as

$$M_{\rm Ein} = \frac{4\pi^2}{Gc^2} \frac{d_L d_{LS}}{d_S} \sigma^4_{\rm SIE}. \tag{3.2}$$

Note that since the strong-lensing cross-section scales as $b^2_{\rm SIE}$, and $b_{\rm SIE}$ in turn scales as $\sigma^2_{\rm SIE}$, the cross-section for strong lensing scales linearly with mass (although this picture is somewhat nuanced due to the spatially extended nature of the mass profile). The surface brightness distribution of sources are represented by either one or multiple Sérsic components with the form

$$I(x,y) \propto \exp[-\frac{1}{2}(\frac{qx^2 + y^2/q}{\sigma^2})^{n/2}] \tag{3.3}$$

with the axis ratio q, width σ, and exponent n as free parameters.

For a particular SIE lens model and specific composition of the source, one can generate the predicted lensed images via the ray-tracing technique according to the analytical expressions of the lens equation (Kormann et al., 1994). Then a Levenberg-Marquardt nonlinear least-squares fit (MPFIT, Moré, 1978; Markwardt, 2009) to the observed lensed images is performed to obtain the best-fit parameters for both the lens and the source.

3.3.1 Verification Test

Before applying our lens modeling code to the newly discovered lens sample, we did a verification test using the SLACS grade-A lens sample (Bolton et al., 2008a) which has been thoroughly studied. Restricting on the 58 modeled grade-A ETG lenses in the SLACS survey with well-measured velocity-dispersion values from SDSS, we re-did the SIE fittings and found highly consistent results. Figure 3.1 shows the ratios of the derived Einstein radii in this work to the published SLACS results. No significant bias is observed: the average ratio is 0.999 with an RMS of 0.005. Hence, our code is able to measure Einstein radii successfully within uncertainties and without bias relative to previously published SLACS Einstein radii.

3.3.2 Lens Modeling: Grade-A Lenses

For each of the 40 newly discovered grade-A lenses with distinct multiple lensed images in the S4TM survey, the fitting strategy is relatively straightforward. We use the position

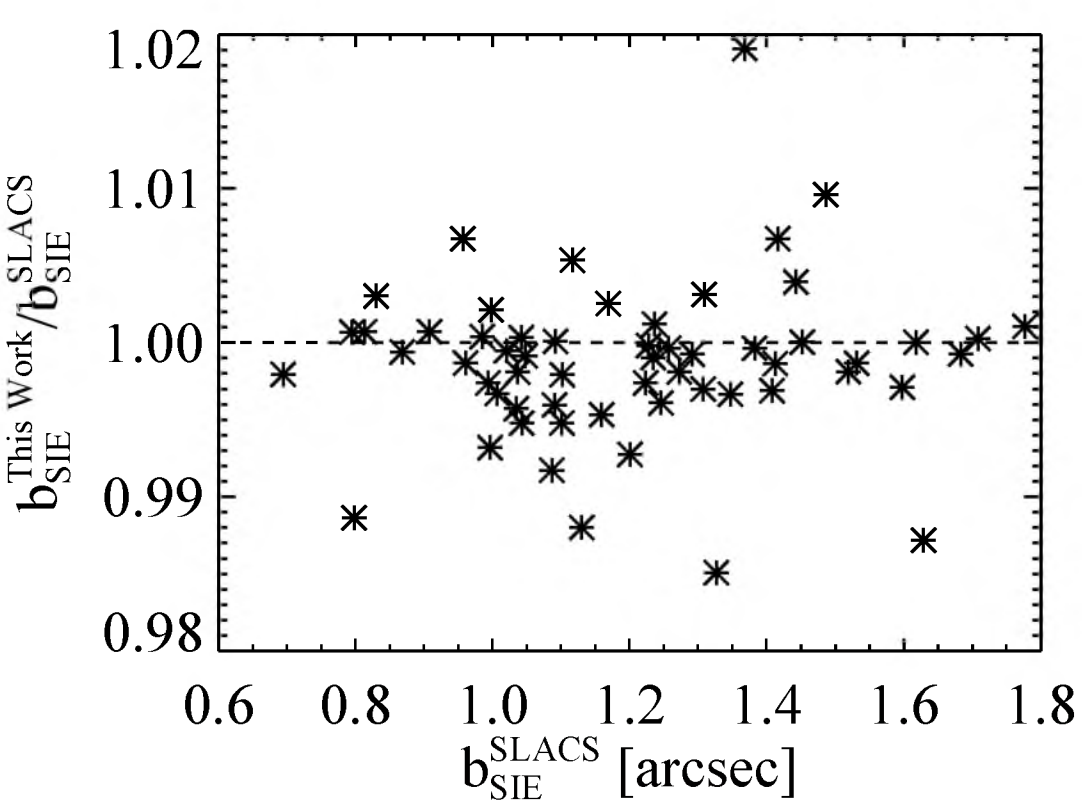

Figure 3.1: A comparison of the derived Einstein radii between this work and the published SLACS values for 58 grade-A lenses confirmed by the SLACS Survey. The ratios are consistent with 1 with a rms = 0.005 which suggests that our code works well in reproducing robust lens models.

angle $P.A._*$ and axis ratio q_* extracted from the b-spline fit to the light distribution as initial guesses. The starting value for $b_{\rm SIE}$ is determined from the separation between a lensed image and its counter-image. Depending on the configuration of the lensing features, one or multiple source components are considered to ensure a reasonably good fit.

Figure 3.2 compares the axis ratios and position angles of the light distribution to those of the mass distribution for the grade-A subsample. The top panel displays the ratios of the minor-to-major axis ratio as a function of the lensing velocity dispersion $\sigma_{\rm SIE}$ which is consistent with 1.0 with a rms scatter of 0.2. No correlation with the velocity dispersion (an approximation of the total mass) is observed. The bottom panel visualizes the difference in the position angle $\Delta P.A. = P.A. - P.A._*$ of lenses with respect to the axis ratios. $\langle \Delta P.A. \rangle = -6°$ with a rms spread of 34°. Clearly, as $q_{\rm SIE} \rightarrow 1$, the position angles become ill-determined and the scatter increases significantly. However, in general, the hypothesis of light tracing mass is valid in terms of both the match of isophotal and isodensity contours and the position-angle alignment and indicates little external perturbing potential. In general, our SIE lens models with multiple parameterized sources can successfully recover the overall lensing features as well as small details.

3.3.3 Lens Modeling: Grade-C Lenses

In comparison with "grade-A" lenses with multiple images or even complete Einstein rings, the systems we refer to as "grade-C lenses" (i.e., without counter-images) are less informative about the mass structure of the foreground galaxy (in addition to being less visually striking). However, while these grade-C lenses do not provide accurate lens mass measurements, the do provide accurate lens-mass *upper limits*. Furthermore, as will be seen later, grade-C systems are relatively less massive, and hence, the upper limits that they provide are an essential element of our program to extend strong-lensing science to lower lens masses. Indeed, the inclusion of such single-image systems to a lens ensemble analysis makes the selection function less sharply dependent on lens galaxy mass as compared to the grade-A lens sample alone.

Nevertheless, in most previous gravitational lensing studies, grade-C lens systems (and the upper mass limits that they provide) have been ignored because of the difficulty or impossibility of obtaining definite lens mass models. (A significant counterexample is the use of lensing "flexion" to constrain mass models in systems that allow this technique: Goldberg and Leonard (2007).) This difficulty is particularly pronounced within pure imaging surveys for lenses, because background galaxies with no clear multiple imaging can easily be confused with satellite galaxies of the foreground lens. For the case of the S4TM program, however,

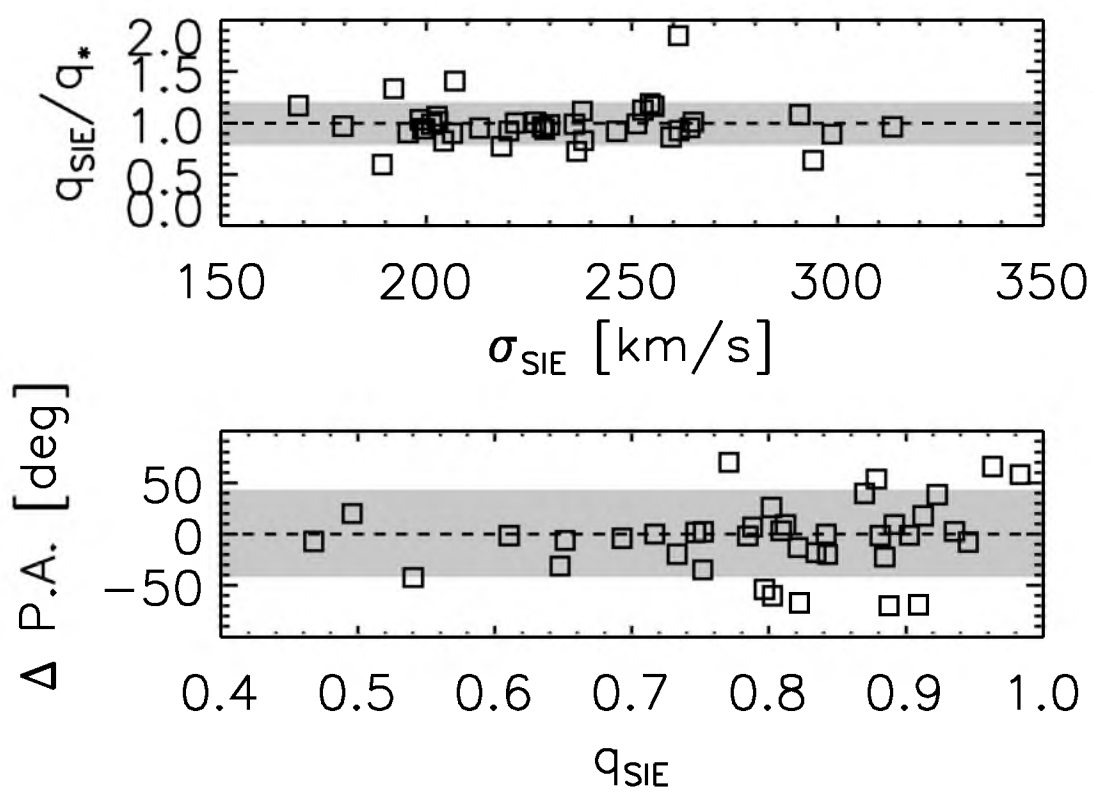

Figure 3.2: Axis ratio and position angle comparisons between the light and mass distributions. Panel (a): Ratio between the minor-to-major axis ratio determined from SIE model fitting $q_{\rm SIE}$ and that measured from light distribution q_*. Panel (b): Discrepancy in the position angle determined from SIE model fitting $P.A.$ and that measured from light distribution $P.A._*$. In both panels, the shaded gray regions indicate their rms spreads (see text for details).

we have a strong prior for the identification of singly imaged background galaxies due to the original spectroscopic detection of a second redshift along the line of sight.

In this work, we explicitly incorporate grade-C lenses into our analysis, by first determining their associated Einstein-radius upper limits, and subsequently incorporating these upper limits into our ensemble analysis of the scaling relations of the lens population. Our approach to the first step is to explore a sequence of lens models of incrementally increasing Einstein radius until we find the value at which an unobserved counter-image is predicted by the model. In detail, for each grade-C lens, at each trial Einstein radius, we fix the P.A. and axis ratio $q_{\rm SIE}$ to values derived from the light distribution, fit for the source as we do for the grade-A subsample, and record the best-fit χ^2 value. The special requirement of fixing P.A. and $q_{\rm SIE}$ stems from the fact that the fitting code can converge to unphysical lens models with very small axis ratios and small lensing strength for grade-C systems. As shown previously by Koopmans et al. (2006b), Bolton et al. (2008b), Barnabè et al. (2011) and confirmed again by the grade-A subsample in the S4TM survey (Figure 3.2), the hypothesis of a mass quadrupole following the light quadrupole is generally valid, and the idea of fixing P.A. and $q_{\rm SIE}$ is well-motivated. Figures in the Appendix show the χ^2 curves for 18 S4TM and 15 SLACS grade-C lenses. To decide the upper limit, we use the χ^2 curve with respect to the trial Einstein radius as a guide and search for a point after which the model starts to predict counter-images that should have been observed and the slope of the χ^2 curve changes significantly. We also tested an objective method of determining the upper limits by choosing the point at which $\Delta\chi^2 = 144$ with respect to the minimum value. The results reported below do not change significantly with this alternative procedure.

3.3.4 Combined Sample

The Einstein-radius upper limits derived for grade-C lenses can be combined with Einstein-radius measurements for grade-A lenses to perform unbiased analyses of the mass structure of ETGs across a wider range of galaxy masses. In the following sections, we combine these measurements with stellar velocity dispersions and broadband photometry to constrain the mass-density profile and dark-matter fraction of ETGs as a function of galaxy mass.

Combining the S4TM survey and the SLACS survey generates a data set including 98 (40+58) grade-A and 33 (18+15) grade-C ETG lens systems. The mean redshift for the foreground lenses is $\langle z_L \rangle = 0.18$ and $\langle z_S \rangle = 0.58$ for the background sources. The average Einstein radius in arcsecond $\langle b_{\rm SIE} \rangle$, the average Einstein radius in physical units $\langle R_{\rm Ein} \rangle$ and the total 2-D projected mass $\langle M_{\rm Ein} \rangle$ for grade-A and grade-C subsamples are summarized

in Table 3.1. Figure 3.3 shows the distribution of the Einstein radius $b_{\rm SIE}$ and total 2-D enclosed mass $M_{\rm Ein}$ for all the 131 lenses. The distributions and median values of the stellar masses (derived from the *HST* F814W photometry assuming a Chabrier IMF and the fiducial stellar-population model of Section 5 below) for various subsamples are also plotted in Figure 3.4. Note that the S4TM lenses are generally less massive compared to the SLACS lenses and grade-C lenses are generally less massive than grade-A lenses.

3.4 Mass Structure Analysis

In this section, we combine our Einstein-radius measurements and upper limits with lens-galaxy stellar velocity dispersions measured from SDSS spectroscopy in order to constrain the mass-density profile of ETG lenses as a function of lens-galaxy mass. The inclusion of new S4TM systems and grade-C lenses allows us to explore a broader range of galaxy masses than previous studies.

3.4.1 Velocity-Dispersion Proxy

To investigate the degree of central concentration of ETG mass profiles, we employ an observational proxy defined by the ratio of stellar velocity dispersion to "lensing velocity dispersion" $\sigma_{\rm SIE}$, defined in relation to the observable parameter $b_{\rm SIE}$ through Equation (3.1). It was suggested by Kochanek (1994) that this ratio, later denoted as $f_{\rm SIE} = \sigma_*/\sigma_{\rm SIE}$, should be approximately unity for isothermal mass models, greater than unity for models more centrally concentrated than isothermal, and less than unity for models less centrally concentrated. Successive studies have confirmed this idea and showed that $f_{\rm SIE}$ can be used as an empirical indicator of the mass-density slope (Kochanek et al., 2000b; Treu and Koopmans, 2002, 2004; Treu et al., 2006; Koopmans et al., 2006b; Bolton et al., 2008b; Auger et al., 2010). We take this approach in the current work in order to investigate physical trends and their significance in the simplest possible manner, and defer a more detailed and self-consistent joint analysis of gravitational lensing and stellar kinematics to future papers.

As mentioned above, our measurements or upper limits for $\sigma_{\rm SIE}$ come directly from our lens-modeling constraints on the Einstein radius for grade-A and grade-C lenses. The other ingredient for our present analysis is the stellar velocity dispersion σ_*, which is the standard deviation of velocities of stars within a galaxy. This quantity is determined spectroscopically by measuring the broadening of the galaxy spectrum due to the luminosity-weighted superposition of Doppler-shifted absorption lines from individual stars. Instead of adopting the preexisting SDSS stellar velocity-dispersion values calculated using a set

Table 3.1: Average values of the Einstein radius in arcsecond, physical units, and the total projected mass for grade-A and grade-C lens samples.

Lens Sample	$\langle b_{\rm SIE}\rangle$	$\langle R_{\rm Ein}\rangle$	$\langle \log_{10}(M_{\rm Ein}/M_\odot)\rangle$
Grade-A	$1.11''$	3.36 kpc	11.14
Grade-C†	$0.93''$	2.42 kpc	10.87

† Note that all the numbers are upper limits for grade-C lenses.

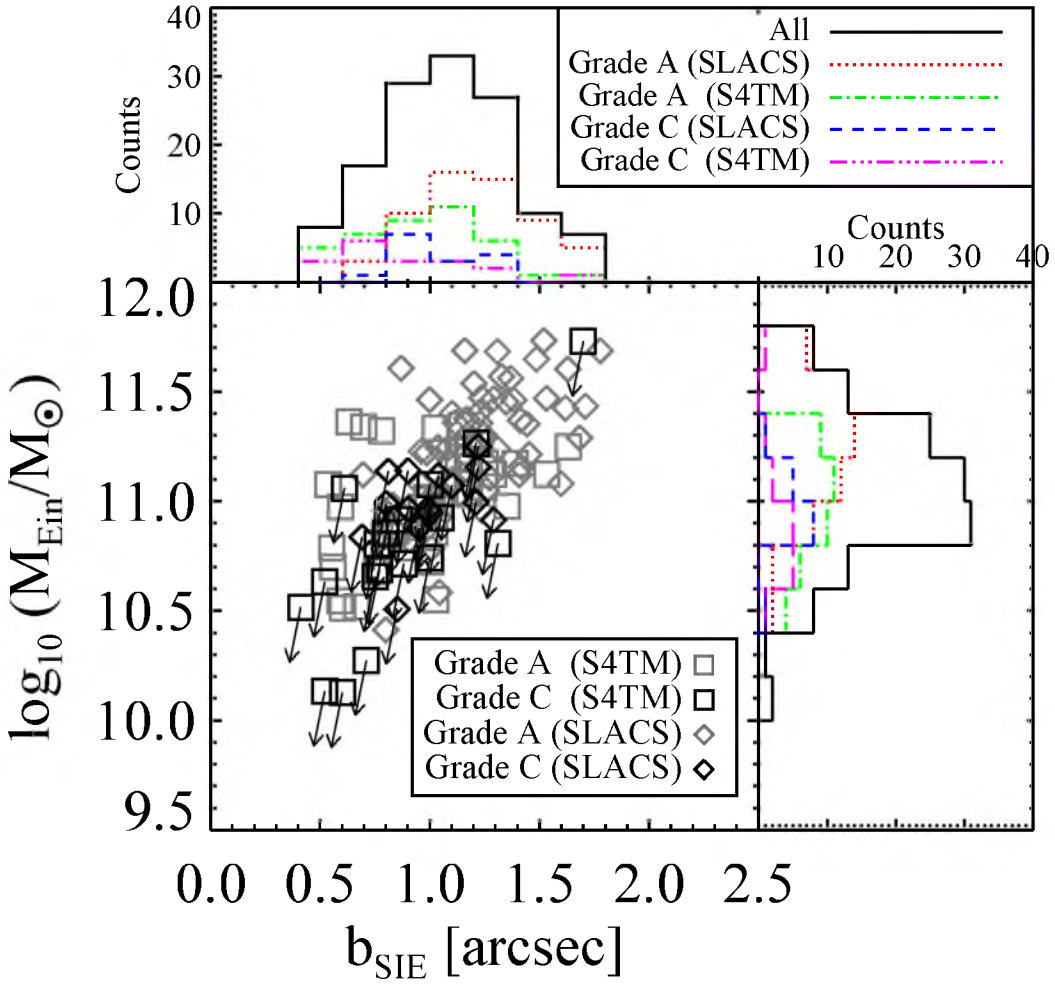

Figure 3.3: Distribution of the predicted Einstein radii $b_{\rm SIE}$ and total two-dimensional enclosed masses $M_{\rm Ein}$ for all the modeled lenses. Plotted in gray are grade-A lenses, with 40 from the S4TM Survey (squares) and 58 from the SLACS Survey (diamonds). Grade-C lenses are in black with left arrows indicating these are upper limits. Overplotted are the projected histogram distributions of $b_{\rm SIE}$ and $M_{\rm Ein}$, respectively, for various subsamples.

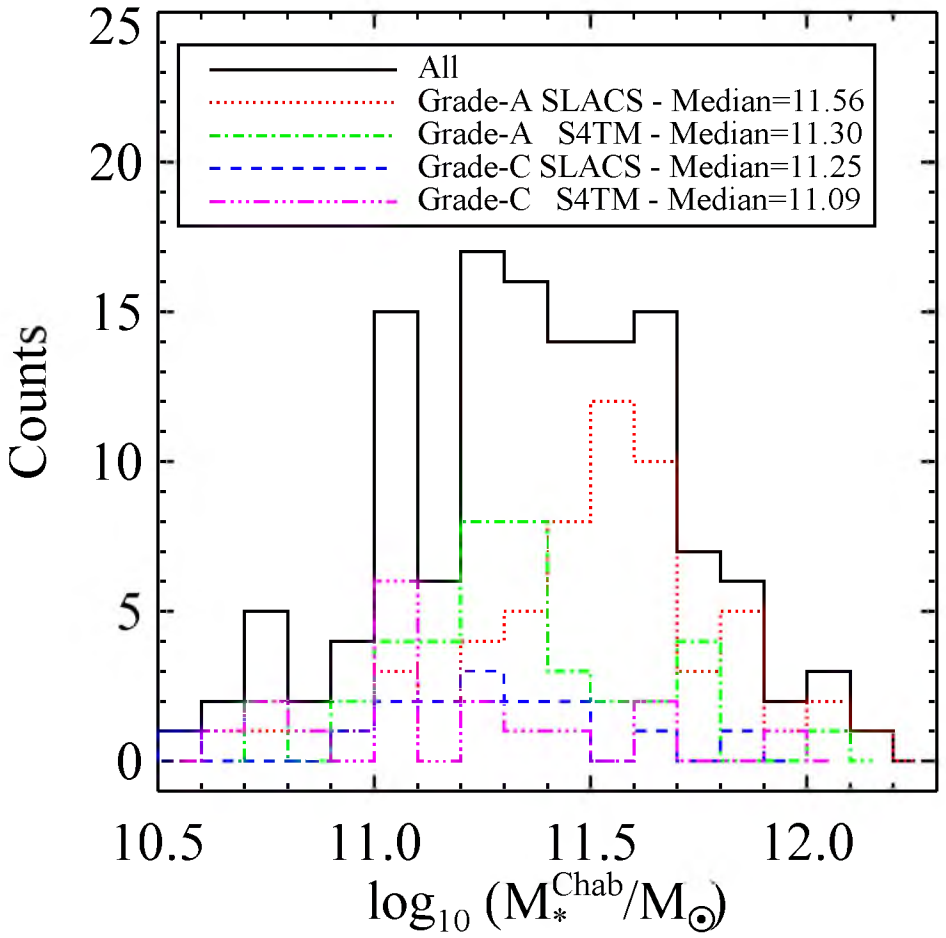

Figure 3.4: Distributions of the stellar mass $M_*^{\rm Chab}$ for the whole grade-A+C sample (black), SLACS grade-A subsample (red), S4TM grade-A subsample (green), SLACS grade-C subsample (blue), and S4TM grade-C subsample (magenta) and the median values for each subsample. The stellar mass is derived from the *HST* I-band photometry by the SPS analysis assuming a Chabrier IMF and other parameters (please refer to Section 5 for details).

of stellar spectra templates derived from ELODIE stellar spectrum library (Prugniel and Soubiran, 2001) with wavelength coverage 4100Å-6800Å, we generate a new set of stellar templates from the Indo-US library (Valdes et al., 2004) patched and extended by selected synthetic spectra in POLLUX database (Palacios et al., 2010), covering the full wavelength range of observed spectra. More details about this procedure can be found in Bolton et al. (2012a,b). The resulting velocity dispersions $\sigma_{\rm ap}$ are then corrected to values within one half of the half-light radius $\mathrm{R_{half}}/2$ (a quantity comparable to the Einstein radius) following a compromise prescription between Jorgensen et al. (1995), Mehlert et al. (2003b), and Cappellari et al. (2006a) as

$$\sigma_* = \sigma_{\rm ap} \times (\frac{1.5''}{\mathrm{R_{half}}/2})^{0.05} \tag{3.4}$$

where $1.5''$ is the angular radius of the SDSS fiber. The half-light radius $R_{\rm half}$ is derived from a core-Sérsic fit (Graham et al., 2003) to the surface brightness distribution of each lens galaxy as explained in Brownstein et al. 2014 (in preparation). Figure 3.5 shows the distribution of all the lenses in terms of their $\log_{10}\sigma_*$ and $\log_{10}\sigma_{\rm SIE}$.

3.4.2 Hierarchical Bayesian Analysis

Our primary interest here is in the physical scaling relation between σ_* and $\sigma_{\rm SIE}$ that encodes the variation of the mass-density profile within the ETG lens population. In the formalism of statistics, we treat this as a *conditional* probability density function (PDF) of σ_* given $\sigma_{\rm SIE}$. In order to deduce the parameters of this conditional PDF correctly, we must also account for the *marginal* PDF of $\sigma_{\rm SIE}$ alone within the sample. The product of the conditional and marginal PDF's gives the *joint* PDF of σ_* and $\sigma_{\rm SIE}$ together, which we infer from the sample data according to the procedure described in this section.

3.4.2.1 Methodology

We parameterize the mean scaling relation between $\log_{10}\sigma_*$ and $\log_{10}\sigma_{\rm SIE}$ as

$$\log_{10}\sigma_* = a \times \log_{10}\sigma_{\rm SIE} + b \tag{3.5}$$

with a slope a and intercept b. To constrain the parameters of this relation, we use the hierarchical Bayesian method (see e.g., Shu et al., 2012; Bolton et al., 2012b) with all the ingredients summarized in Table 3.2. This method makes full use of all the observed information, deconvolves the observational uncertainties, and offers unbiased estimations of population "hyperparameters" that we are interested in. The hierarchical Bayesian approach also allows for straightforward inclusion of Einstein-radius upper limits in the analysis.

Here we describe all the ingredients in the hierarchical Bayesian analysis. x, y represent

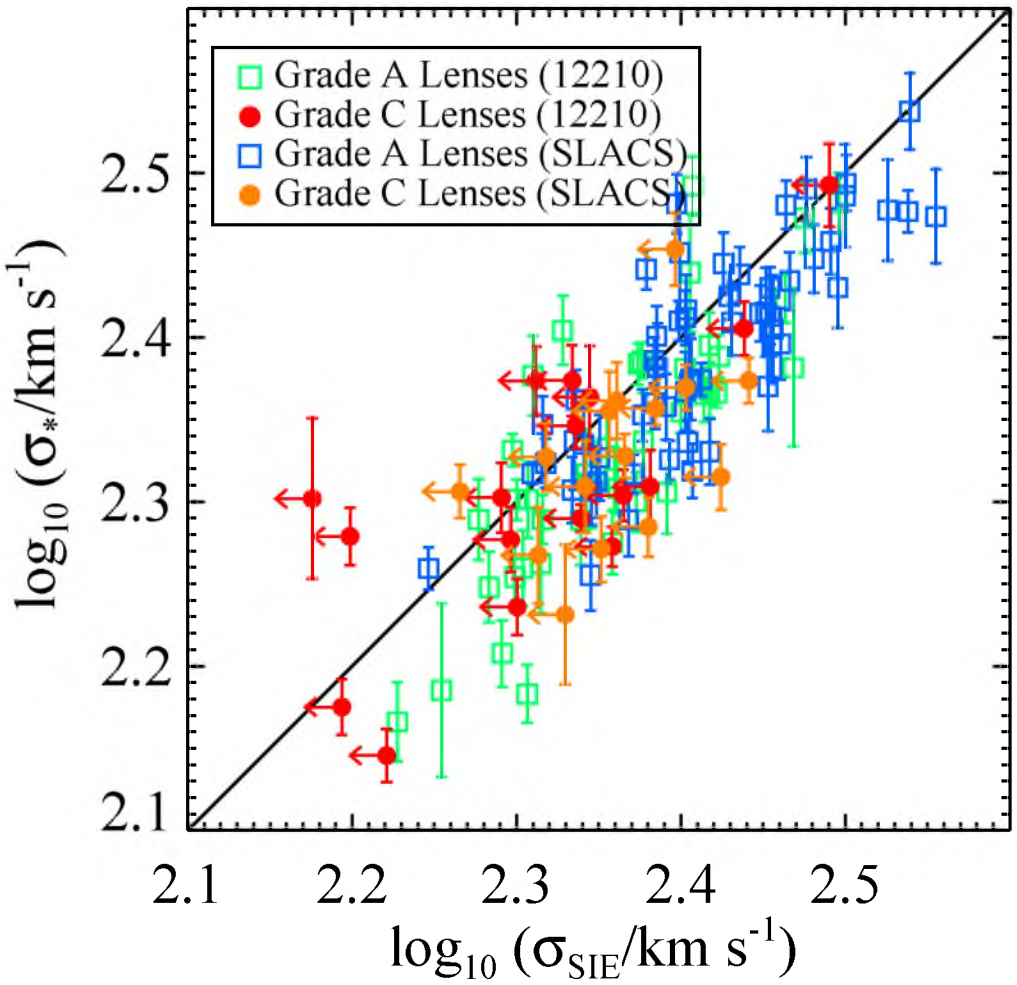

Figure 3.5: Distribution of $\log_{10}\sigma_*$ and $\log_{10}\sigma_{\rm SIE}$ for all the modeled lenses. Filled circles with left arrows indicating the upper limits represent grade-C lenses, with 18 from the S4TM survey (red) and 15 from the SLACS survey (chocolate). 40 S4TM and 58 SLACS grade-A lenses are shown by green and blue squares, respectively.

Table 3.2: Ingredients for the hierarchical Bayesian analysis.

Symbol	Definition
x	Parameter 1: $\log_{10}\sigma_{\rm SIE}$
y	Parameter 2: $\log_{10}\sigma_{*}$
x_i	Data 1: derived value of $\log_{10}\sigma_{\rm SIE}$ for lens i
y_i	Data 2: measured value of $\log_{10}\sigma_{*}$ for lens i
a	Hyperpar. 1: slope of $\log_{10}\sigma_{*}$-$\log_{10}\sigma_{\rm SIE}$
b	Hyperpar. 2: intercept of $\log_{10}\sigma_{*}$-$\log_{10}\sigma_{\rm SIE}$
δ	Hyperpar. 3: intrinsic scatter in $\log_{10}\sigma_{*}$
m	Hyperpar. 4: mean in $\log_{10}\sigma_{\rm SIE}$
s	Hyperpar. 5: intrinsic width in $\log_{10}\sigma_{\rm SIE}$
$\vec{\theta}$	Vector of hyperparameters
$\Pr(x_i\|x)$	Likelihood function of x given x_i
$\Pr(y_i\|y)$	Likelihood function of y given y_i
$\Pr(y,x\|\vec{\theta},\mathscr{H})$	Joint PDF of parameters given $\vec{\theta}$
$\Pr(y\|x,a,b,\delta,\mathscr{H})$	PDF of parameter y given x, a, b, and δ
$\Pr(x\|m,s,\mathscr{H})$	Marginal PDF of parameter x

the two physical parameters $\log_{10}\sigma_{\rm SIE}$, $\log_{10}\sigma_*$, and x_i, y_i are the corresponding observed values for the i^{th} lens galaxy. The five hyperparameters (compacted as $\vec{\theta}$) we want to determine are i) a: the slope of $\log_{10}\sigma_*$-$\log_{10}\sigma_{SIE}$ relation; ii) b: the intercept; iii) δ: the intrinsic scatter in $\log_{10}\sigma_*$; iv) m: the mean in $\log_{10}\sigma_{\rm SIE}$ distribution; and v) s: the intrinsic width in $\log_{10}\sigma_{\rm SIE}$ distribution. $\Pr(y|y_i)$ is the likelihood function of y given y_i, which can be written as

$$\Pr(y_i|y) = \frac{1}{\sqrt{2\pi}\delta_i}\exp[-\frac{(y-y_i)^2}{2\delta_i^2}] \tag{3.6}$$

$\Pr(x_i|x)$ is the likelihood function of x given x_i. Under the assumption that the Einstein radii for grade-A lenses are accurately determined through the lens modeling, the resulted likelihood functions of x can be well-described by various *Dirac delta functions*

$$\Pr_{_A}(x_i|x) = \delta(x-x_i) \tag{3.7}$$

while for grade-C lenses, since only upper limits are estimated, the probability of having x greater than the measured value x_i is assumed to be 0. Therefore, we introduce the *Heaviside step functions* as the likelihood functions of parameter x in the following manner

$$\Pr_{_C}(x_i|x) = [1-H(x-x_i)] \tag{3.8}$$

Note that in principle, $\Pr_{_C}(x_i|x)$ can not be normalized. However, as will be shown later, it is always combined with other Gaussian-like PDFs that drop rapidly at large values so that the normalization is no longer an issue. The joint PDF of parameters y and x given $\vec{\theta}$ and a hypothesis $\mathscr{H}$, $\Pr(y,x|\vec{\theta},\mathscr{H})$, is equal to

$$\Pr(y,x|\vec{\theta},\mathscr{H}) = \Pr(y|x,a,b,\delta,\mathscr{H})\Pr(x|m,s,\mathscr{H}) \tag{3.9}$$

where $\Pr(y|x,a,b,\delta,\mathscr{H})$ is the conditional PDF of y given x and three hyperparameters and $\Pr(x|m,s,\mathscr{H})$ is the marginal PDF of x. Followed by the parameterized model stated by Equation (3.5), $\Pr(y|x,a,b,\delta,\mathscr{H})$ can be expressed as

$$\Pr(y|x,a,b,\delta,\mathscr{H}) = \frac{1}{\sqrt{2\pi}\delta}\exp[-\frac{(ax+b-y)^2}{2\delta^2}] \tag{3.10}$$

with an assumed intrinsic scatter δ in y. $\Pr(y|x,a,b,\delta)$ is of physical interest as it quantifies the relation between y and x and can be adopted in related studies. The last piece is the marginal PDF of x, which is characterized by a Gaussian with mean m and intrinsic scatter s:

$$\Pr(x|m,s,\mathscr{H}) = \frac{1}{\sqrt{2\pi}s}\exp[-\frac{(x-m)^2}{2s^2}] \tag{3.11}$$

Following the same strategy used by Shu et al. (2012), the likelihood function of hyperparameters $\vec{\theta}$ given the observed data $\{\vec{y}\}$, $\{\vec{x}\}$ and a hypothesis $\mathscr{H}$ is defined as

$$\mathscr{L}(\vec{\theta}|\{\vec{y}\},\{\vec{x}\},\mathscr{H}) = \prod_{i=0}^{N}\Pr(y_i,x_i|\vec{\theta},) = \prod_{i=0}^{N}\iint\Pr(y_i,x_i|y,x)\Pr(y,x|\vec{\theta},\mathscr{H})\,{\rm d}x{\rm d}y$$

In general, y and x are independent variables and $\Pr(y_i,x_i|y,x)$ can be split as

$$\mathscr{L}(\vec{\theta}|\{\vec{y}\},\{\vec{x}\}) = \prod_{i=0}^{N} \iint \Pr(y_i|y)\Pr(x_i|x)\Pr(y,x|\vec{\theta},\mathscr{H})\,\mathrm{d}x\mathrm{d}y$$
$$= \prod_{i=0}^{N} \iint \Pr(y_i|y)\Pr(x_i|x)\Pr(y|x,a,b,\delta,\mathscr{H})\Pr(x|m,s,\mathscr{H})\,\mathrm{d}x\mathrm{d}y \quad (3.12)$$

The posterior PDF of the hyperparameters that we are interested in is simply related to the likelihood via the *Bayes' rule* as

$$\Pr(\vec{\theta}|\vec{y},\vec{x},\mathscr{H}) = \frac{\mathscr{L}(\vec{\theta}|\vec{y},\vec{x},\mathscr{H})\Pr(\vec{\theta}|\mathscr{H})}{\Pr(\vec{y},\vec{x}|\mathscr{H})}, \quad (3.13)$$

from which we can infer the relation between $\log_{10}\sigma_*$ and $\log_{10}\sigma_{\rm SIE}$.

3.4.2.2 Marginal PDF of x

To determine the mean m and intrinsic scatter s for the marginal PDF of x, we write out the likelihood function of m and s as

$$\mathscr{L}(m,s|\vec{x},\mathscr{H}) = \Pr(\vec{x}|m,s,\mathscr{H}) = \prod_{i=0}^{N}\Pr(x_i|m,s,\mathscr{H}) = \prod_{i=0}^{N}\int_{-\infty}^{\infty}\Pr(x_i|x)\Pr(x|m,s,\mathscr{H})\,\mathrm{d}x$$

which is a product of two parts corresponding to grade-A and grade-C lenses, respectively.

$$\mathscr{L}(m,s|\vec{x},\mathscr{H}) = \prod_{i=0}^{N_A}\int_{-\infty}^{\infty}\delta(x-x_i)\Pr(x|m,s,\mathscr{H})\,\mathrm{d}x \times \prod_{j=0}^{N_C}\int_{-\infty}^{\infty}[1-H(x-x_j)]\Pr(x|m,s,\mathscr{H})\,\mathrm{d}x$$
$$= \prod_{i=0}^{N_A}\frac{1}{\sqrt{2\pi}s}\exp[-\frac{(x_i-m)^2}{2s^2}] \times \prod_{j=0}^{N_C}\Phi(\frac{x_j-m}{s})$$

where $\Phi(x)$ is the cumulative distribution function (CDF) of the standard normal distribution. In a format of the more convenient log-likelihood, we have

$$\ln\mathscr{L}(m,s|\vec{x},\mathscr{H}) = [-N_A\ln(\sqrt{2\pi}s) - \sum_{i=0}^{N_A}\frac{(x_i-m)^2}{2s^2}] + \sum_{j=0}^{N_C}\ln[\Phi(\frac{x_j-m}{s})]$$

in which the first two terms come from pure grade-A lens subsample and the last term is contributed by grade-C lenses.

Exploration of the 2D m-s space yields a log-likelihood distribution which can be converted to the posterior PDF of m and s through the *Bayes' rule* as

$$\Pr(m,s|\vec{x},\mathscr{H}) = \frac{\mathscr{L}(m,s|\vec{x},\mathscr{H})\Pr(m,s,\mathscr{H})}{\Pr(\vec{x}|\mathscr{H})}. \quad (3.14)$$

A flat prior $\Pr(m,s,\mathscr{H})$ in m and s is assumed for simplicity and the evidence $\Pr(\vec{x}|\mathscr{H})$ is a normalization constant. Plotted in Fig. 3.6 are the resulted posterior probability distribution contours for m and s (solid black lines). For comparison, the posterior PDF for the pure grade-A lens subsample is over-plotted in dashed gray lines. The contours represent 68%, 95%, and 99.7% confidence levels, respectively. It is suggested from the maximum-likelihood estimation (MLE) that, for the grade-A lens subsample, the best-estimated values are $m = 2.391 \pm 0.010$ and $s = 0.068 \pm 0.007$ (all the quoted error bars are the averages of the upper and lower bounds). By including grade-C lenses, it pulls down the mean by 0.029

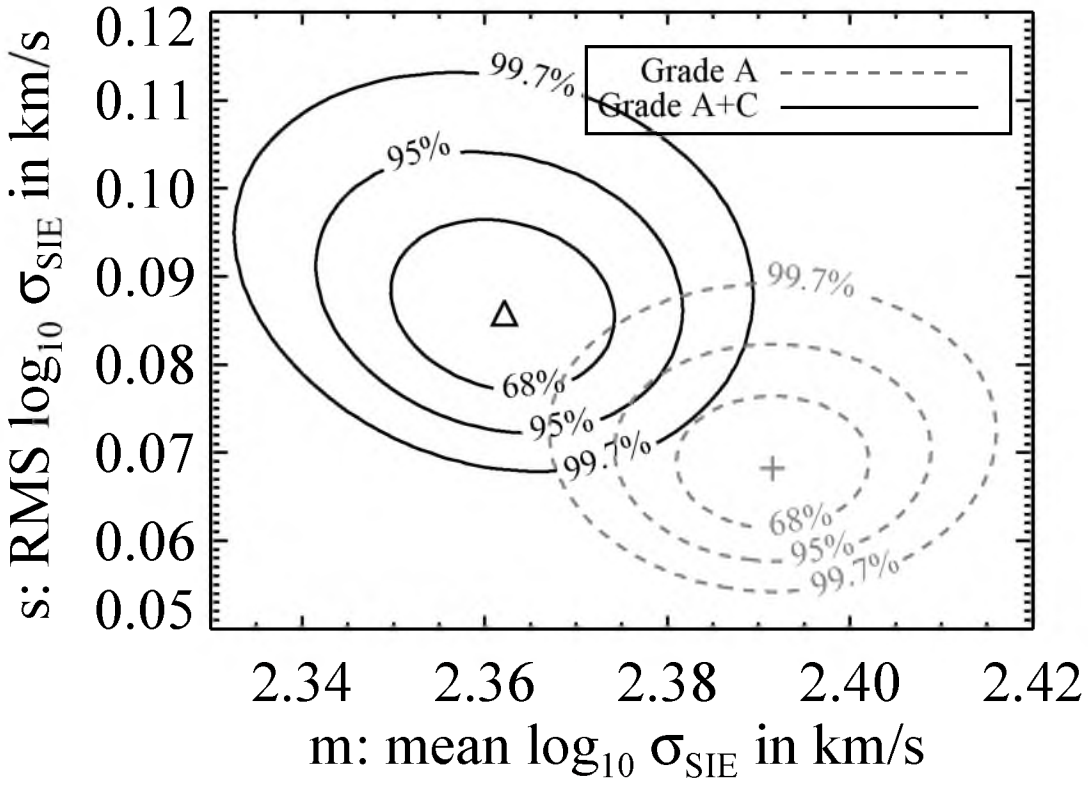

Figure 3.6: The two-dimensional distribution of the posterior PDF of m and s for the combined sample (grade A+B lenses; solid black lines) and pure grade-A sample (dashed gray lines). Contours represent the 68%, 95%, and 99.7% confidence levels, respectively. The triangle and the plus symbols each indicate the peak points in the posterior PDFs for these two samples.

dex to $m_{\rm best} = 2.362 \pm 0.012$ as expected since grade-C lenses are relatively less massive and the derived $\log_{10}\sigma_{\rm SIE}$ values are slightly smaller. Also, the distribution becomes broader with an intrinsic scatter $s_{\rm best} = 0.086 \pm 0.009$.

3.4.2.3 Hyperparameter Determination

Finally, by substituting $m_{\rm best}$ and $s_{\rm best}$ back to Equation (3.12), we can write the likelihood function as

$$\mathscr{L} \propto \prod_{i=0}^{N} \iint \Pr(y_i|y)\Pr(x_i|x)\Pr(y|x,a,b,\delta,\mathscr{H})\Pr(x|m_{\rm best},s_{\rm best})\,dxdy \propto \mathscr{L}_A \times \mathscr{L}_C$$

The full computation is outlined in the Appendix. Eventually, we have

$$\ln\mathscr{L} = \ln\mathscr{L}_A + \ln\mathscr{L}_C = -N\ln s_{\rm best} - \sum_{i=0}^{N_A}[\frac{1}{2}\ln(\delta_i^2+\delta^2) + \frac{(ax_i+b-y_i)^2}{2(\delta_i^2+\delta^2)} + \frac{(x_i-m_{\rm best})^2}{2s_{\rm best}^2}]+$$
$$\sum_{j=0}^{N_C}\{\ln[\Phi(\frac{s_{\rm best}^2 a(ax_j+b-y_j)+(x_j-m_{\rm best})(\delta_j^2+\delta^2)}{s_{\rm best}\sqrt{s_{\rm best}^2a^2+\delta^2+\delta_j^2}\sqrt{\delta^2+\delta_j^2}})]-$$
$$\frac{1}{2}\ln(s_{\rm best}^2a^2+\delta_j^2+\delta^2) - \frac{(am_{\rm best}+b-y_j)^2}{2(s_{\rm best}^2a^2+\delta^2+\delta_j^2)}\} \tag{3.15}$$

One issue we found was that the slope a and intercept b correlate with each other extremely strongly and resulted in extremely narrow confidence regions. An easy solution to that is to re-phrase Equation (3.5) as

$$\log_{10}\sigma_* = a' \times (\log_{10}\sigma_{\rm SIE} - m_{\rm best}) + (b' + m_{\rm best}) \tag{3.16}$$

The likelihood function has exact the same form as in Equation (3.15) but with a and b replaced by a' and $b' - a'm_{\rm best} + m_{\rm best}$ appropriately. Similar to what we have done previously, the posterior PDF of the hyperparameters is related to this new likelihood function as

$$\Pr(\vec{\theta'}|\vec{y},\vec{x}) \propto \mathscr{L}(\vec{\theta'}|\vec{y},\vec{x}) \tag{3.17}$$

Note that here we change $\vec{\theta}$ to $\vec{\theta'}$.

Known as $f_{\rm SIE}$, the ratio of σ_* to $\sigma_{\rm SIE}$ has been used as an empirical estimator of the logarithmic mass-density slope γ' (Treu et al., 2009; Auger et al., 2010). Using the σ_*-$\sigma_{\rm SIE}$ relation found above, by definition, we have

$$f_{\rm SIE} = \frac{\sigma_*}{\sigma_{\rm SIE}} = 10^{y-x} = 10^{[(a'-1)(y-m_{\rm best})+b']/a'} = 10^{(a'-1)(x-m_{\rm best})+b'} \tag{3.18}$$

The physical interpretation of hyperparameters a' and b' then become straightforward. If a' is exactly unity, $f_{\rm SIE}$ (or equivalently the logarithmic mass-density slope γ') is independent of σ_*, an indicator of the mass of lens galaxies. A less-than-unity a' indicates an anti-correlation between γ' and σ_*, namely galaxies with smaller/larger σ_* are more/less centrally concentrated. b' is related to the $f_{\rm SIE}$ value at the mean $\sigma_{\rm SIE}$ value ($\sigma_{\rm SIE} \approx 230$ km/s).

The mass-density profile is (approximately speaking) isothermal if $b' = 0$, sub-isothermal if $b' < 0$, and super-isothermal if $b' > 0$.

Marginalizing over the δ' dimension, the 2D posterior PDF of a' and b' is shown in Figure 3.7. Once again, there are two sets of contours for grade-A subsample (dashed gray lines) and the combined sample (solid black lines) with the best-estimated values indicated by a plus and a triangle symbol separately. The marginal PDFs of a' and b' are obtained by marginalizing over either b' or a' and shown in Fig. 3.8. The best-estimated values for a' and b' and the uncertainties are extracted from simple Gaussian fits to the corresponding marginal PDFs, and summarized in Table 3.3.

We see that for the grade-A lens subsample, the overall mass-density profile is essentially consistent with a mass-independent model at about 1.04σ ($P(a' < 1.0) = 85.086\%$), as found in previous SLACS studies (e.g., Bolton et al., 2008b; Auger et al., 2010). However, the inclusion of grade-C systems in the analysis significantly shifts this result and provides evidence for a mass-dependent density profile at about 4σ ($P(a' < 1.0) = 99.996\%$). The sense of this trend is for lower mass (i.e., lower velocity-dispersion) ETGs to have steeper mass-density profiles and higher mass ETGs to have shallower profiles.

Consulting the empirical relation between the mass-density profile slope γ' and the observable $f_{\rm SIE}$ from Auger et al. (2010), we find that a value of $b' = -0.0338$ corresponds on average to an isothermal profile ($\gamma' = 2$). Considering the best-fit value for b' for the grade-A subsample, we find that it is somewhat inconsistent with isothermal at about 2.6σ. Including grade-C systems as well, the best b' value is more strongly inconsistent with isothermal at 4.8σ. In both cases, the offset is in the sense of having a slightly super-isothermal profile at the central lensing velocity-dispersion of the samples, consistent with previous SLACS findings.

It is worth noting that we only assign upper limits of Einstein radii to grade-C lenses. This procedure will by nature tend to tilt the $\log_{10}\sigma_* - \log_{10}\sigma_{\rm SIE}$ relation and bias the mass-density estimation to a steeper profile. However, the observed change in the slope is too huge to be caused purely by the upper-limit treatment and therefore, is connected to lens-galaxy mass. Additionally, the connection between $f_{\rm SIE}$ and the density slope is rather empirical and may not be valid throughout the whole sample. So a more robust way of estimating the total mass-density profile is demanded.

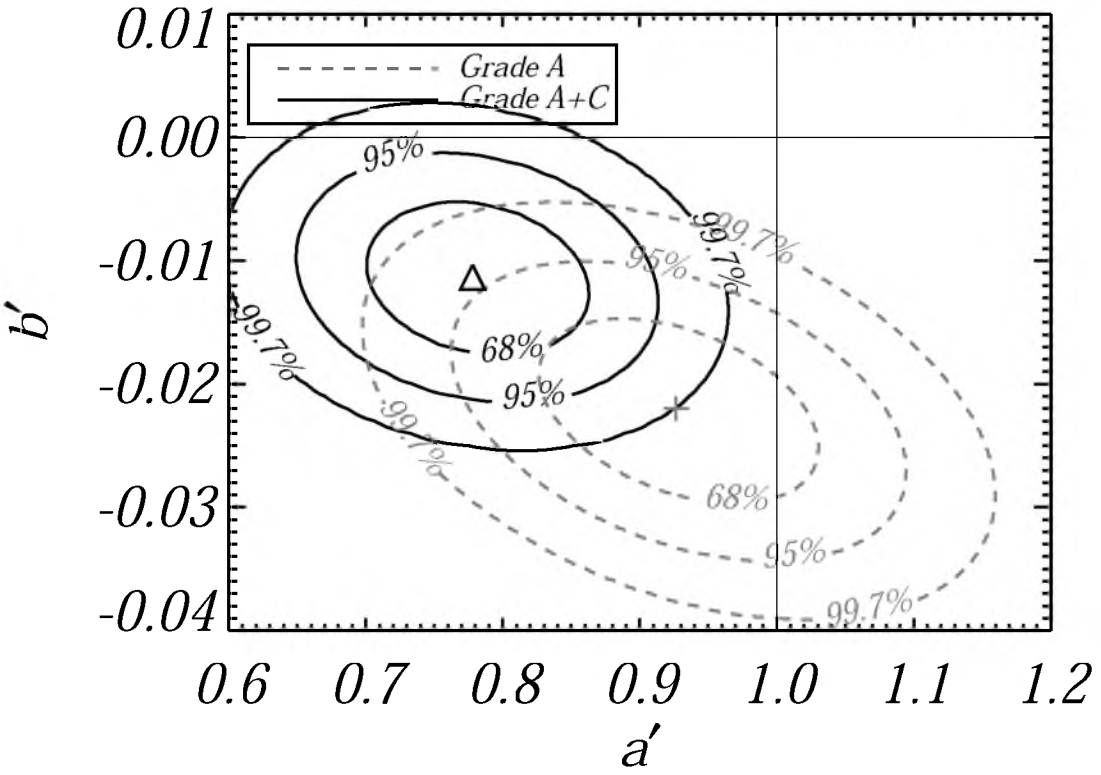

Figure 3.7: The posterior probability density distributions of (a', b') for the combined sample (grade A+C lenses, solid black contours) and pure grade-A subsample (dashed gray contours). 68%, 95%, and 99.7% confidence levels are plotted accordingly. The triangle and the plus symbols each represent the best-estimated values for a' and b' for these two samples. The horizontal line ($b' = 0$) corresponds to an isothermal mass-density profile at $\sigma_{\rm SIE} \approx 230$ km/s, and the vertical line ($a' = 1$) corresponds to a velocity-dispersion-independent mass-density profile.

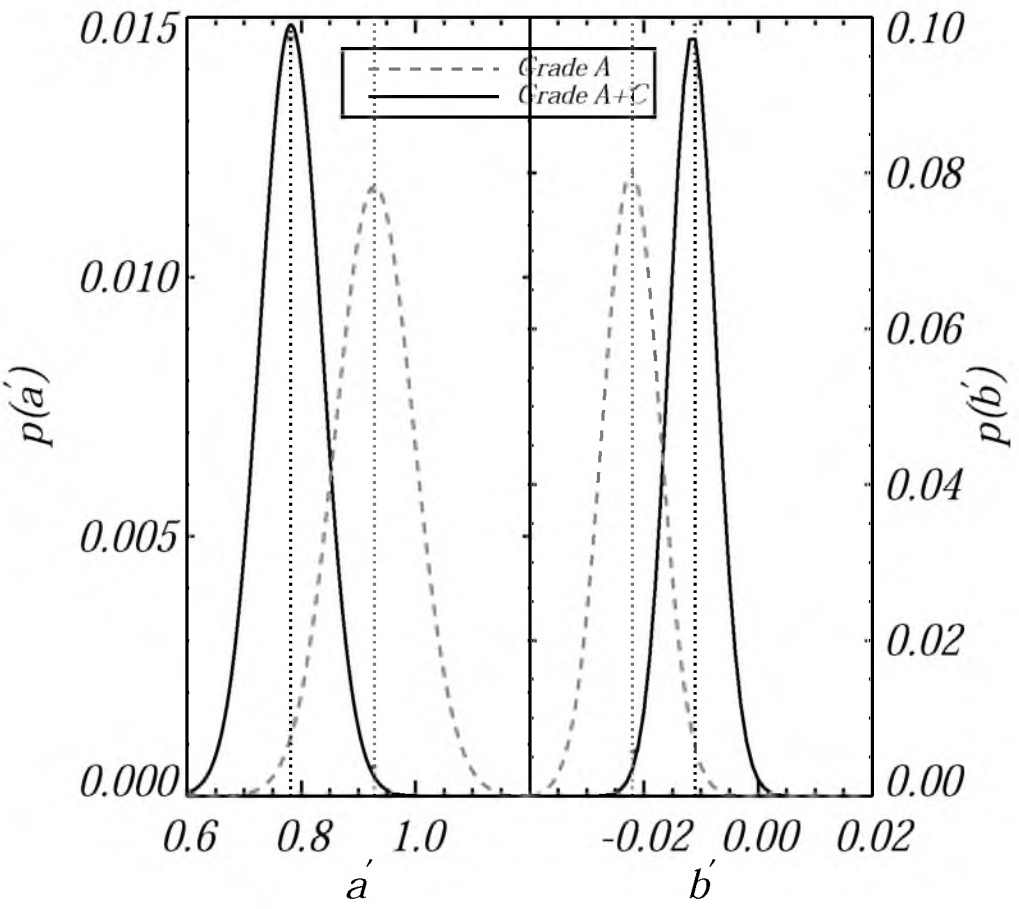

Figure 3.8: The marginal PDFs of a' and b' for the combined sample (grade-A+C lenses, solid black curves) and pure grade-A lens sample (dashed gray curves). Dotted lines corresponds to the peak positions in the marginal PDFs.

Table 3.3: The best-estimated values of the two hyperparameters a' and b' derived from the corresponding marginal PDFs for grade-A sample (2nd row) and grade A+C sample (3rd row). By including grade-C lenses, the slopes become significant shallower.

Lens Sample	$a'_{\rm best}$	$b'_{\rm best}$
Grade A	0.93 ± 0.07	-0.022 ± 0.005
Grade A+C	0.78 ± 0.05	-0.011 ± 0.004

3.5 Stellar Masses and Initial Mass Function

In addition to allowing measurements of the shape of the lens mass-density profile, strong lensing data can be combined with photometry and stellar-population diagnostics to constrain the dark-matter fraction and/or stellar IMF in the lens (e.g. Treu et al., 2010; Ferreras et al., 2010; Spiniello et al., 2011; Grillo and Christensen, 2011; Spiniello et al., 2012; Sonnenfeld et al., 2012; Brewer et al., 2014). For this purpose, we estimate stellar masses of lens galaxies based on scaling of SSP models to *HST* I-band photometry under a range of stellar-population assumptions, and adopting either a Chabrier or Salpeter IMF. This simplicity is motivated by uniformity, since although multiband *HST* photometry is available for many of the SLACS lenses, all the new S4TM systems currently have I-band data alone. High-resolution *HST* imaging is essential to masking the contribution from lensed background sources when performing lens-galaxy photometric modeling (Brownstein et al. 2014 in preparation). For this reason, we disregard multiband SDSS photometric magnitudes in our analysis.

To translate photometry into stellar masses, we make use of SSP models obtained with the Flexible Stellar Population Synthesis (FSPS: Conroy et al., 2009; Conroy and Gunn, 2010). Without colors or narrow-band indices, we must necessarily make assumptions about population parameters such as the formation time after the Big Bang $t_{\rm form}$, metallicity, and dust in the FSPS code. Since all the lens galaxies are by selection ETGs at relatively low redshifts, we adopt a reference model with typical values of $t_{\rm form} = 4$ Gyrs, solar metallicity, and no dust (Gallazzi et al., 2006; Carson and Nichol, 2010). We cross-check our stellar mass estimations with the values obtained by Auger et al. (2009) from multiband *HST* photometric data and a Bayesian stellar population analysis approach for 52 confirmed SLACS lenses in common and find good agreement with no bias observed. To quantify the systematic uncertainty of the simple treatment, we also consider lower- and upper-bound models. Our lower-bound model is dust-free and metal-poor ($\log_{10} Z/Z_{\odot} = -0.30$), while our upper-bound model is dusty and metal-rich (the optical depth for the dust attenuation $\tau = 0.95$, $\log_{10} Z/Z_{\odot} = 0.20$)[1]. The level of the resulting systematic variation in estimated stellar mass is around 0.5 dex. Figure 3.9 shows the stellar masses of the 130 lenses from both the S4TM survey (filled symbols) and the SLACS survey (open symbols) for the reference model as a function of the stellar velocity dispersion. In accordance with the well-known Faber-Jackson relation (FJR, Faber and Jackson, 1976), galaxies with higher

[1]The definitiona and physical meaning of the dust parameter can be found in Charlot and Fall (2000); Conroy et al. (2009).

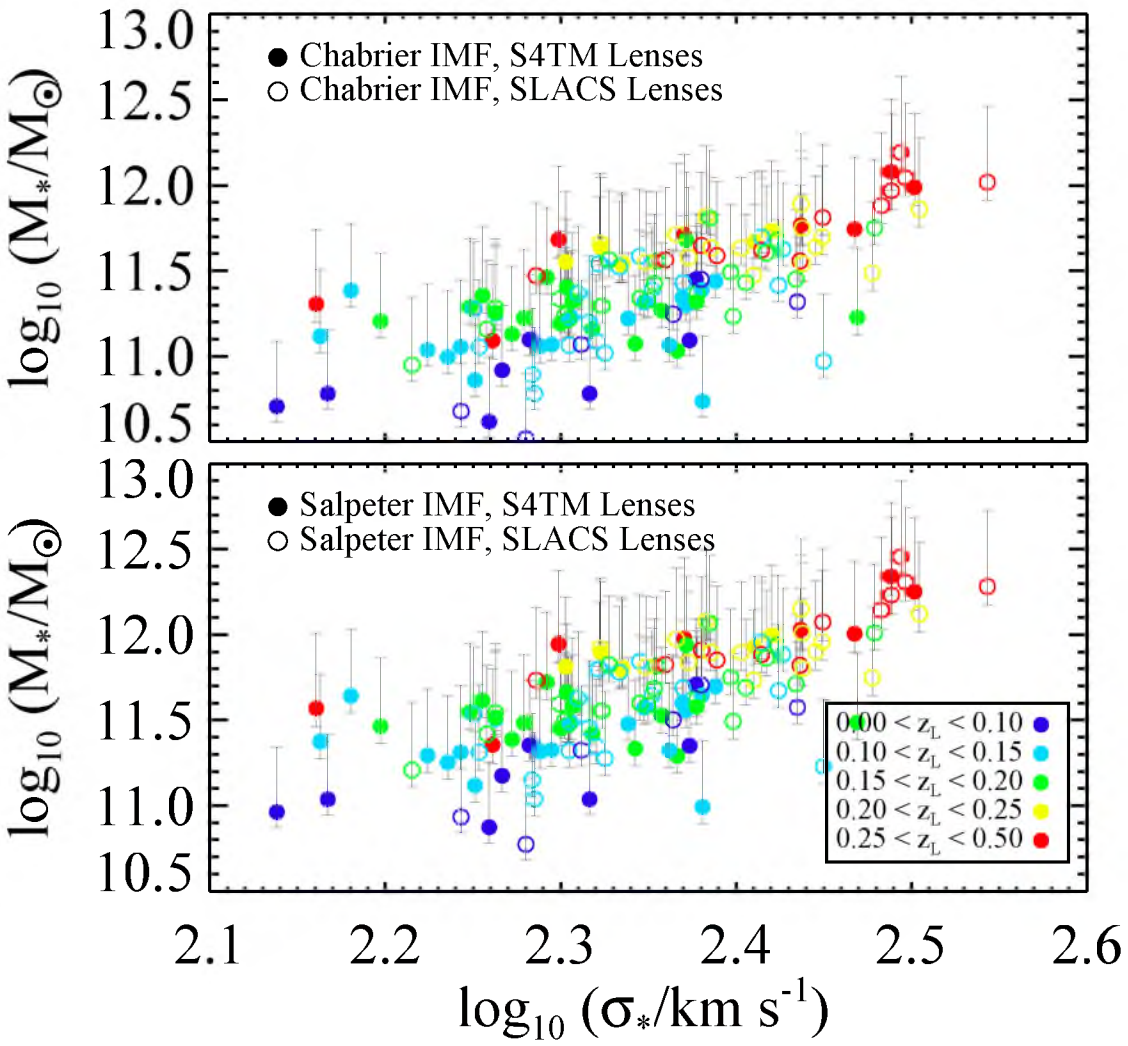

Figure 3.9: Stellar masses M_* of the 130 lenses inferred from the SPS analysis for the two IMFs as a function of the stellar velocity dispersion. Filled symbols are for the S4TM lenses and open symbols for the SLACS lenses. Colors represent the redshifts of the lens galaxies. Gray error bars represent the systematic variations (see text for details on how to determine the error bars). The correlation of scatter with redshift is primarily driven by Malmquist bias in the parent samples.

velocity dispersions also have higher stellar masses on average.

We next examine the relationships between the projected dark-matter fraction within one half of the half-light radius $f_{\rm dm}$ defined as $f_{\rm dm} \equiv 1 - M_*(< {\rm R_{half}}/2)/{\rm M_{R_{half}/2}}$ and the total mass inferred from lensing, the stellar mass, the lensing velocity dispersion, the stellar velocity dispersion, the half-light radius, and the ratio of the Einstein radius to the half-light radius. The half-light radius $\rm R_{half}$ is determined by the core-Sérsic fit as the radius within which the enclosed light is one half of the total profile light. The stellar mass within one half of the half-light radius $M_*(< {\rm R_{half}}/2)$ is interpolated according to the underlying core-Sérsic profile under the assumption of a constant stellar-mass-to-light ratio for each galaxy. And the total mass within one half of the half-light radius $M_{\rm R_{half}/2}$ is interpolated according to an isothermal density profile in which $M(< R) \propto R$. In Fig. 3.10, circiles correspond to predictions by the model of $t_{\rm form} = 4$ Gyrs, solar metallicity and dust-free for the two IMFs, respectively. The colors encode the redshifts of lens galaxies and downward arrows indicate the upper-limits for grade-C lenses. Gray error bars show the systematic variations in $f_{\rm dm}$ throughout the parameter space as explained above. From Fig. 3.10, we see general trends toward higher dark-matter fractions in galaxies with higher masses, larger velocity dispersions, and bigger sizes, consistent with detections by Auger et al. (2010) using confirmed SLACS lenses. In all cases, the intrinsic scatter in $f_{\rm dm}$ is appreciable. Of particular note, by implying a *negative* dark-matter fraction, the data strongly disfavor a Salpeter IMF for large fraction of the lenses, especially for those with total masses within one half of the half-light radius $M_{\rm R_{half}/2}$ less than approximately $10^{10.8} M_\odot$, or equivalently with stellar veolocity dispersions σ_* smaller than approximately 180 km/s. This confirms a similar finding from Brewer et al. (2012) based on a much smaller number of spiral lens galaxies.

We have several avenues to improve the stellar-mass estimation. First of all, the significant scatter and systematic variation in $f_{\rm dm}$ and the IMF that we have not taken into account here will weaken any observed trends. A proper way to handle them is heavily required. On the lensing side, a simple SIE model for the total mass distribution as considered in this work, although a good approximation, is not able to distinguish the contributions from dark and baryonic matter. Furthermore, the known mass-density-profile-IMF degeneracy (Treu et al., 2010; Oguri et al., 2013) can not been taken care of based on the current data. On the SPS side, age, metallicity, dust, and other parameters in the SPS models need to be better constrained. Also, it has been studied that the SPS technique highly relies on several IMF-sensitive spectral features such as NaI, CaII, FeH, TiO, and

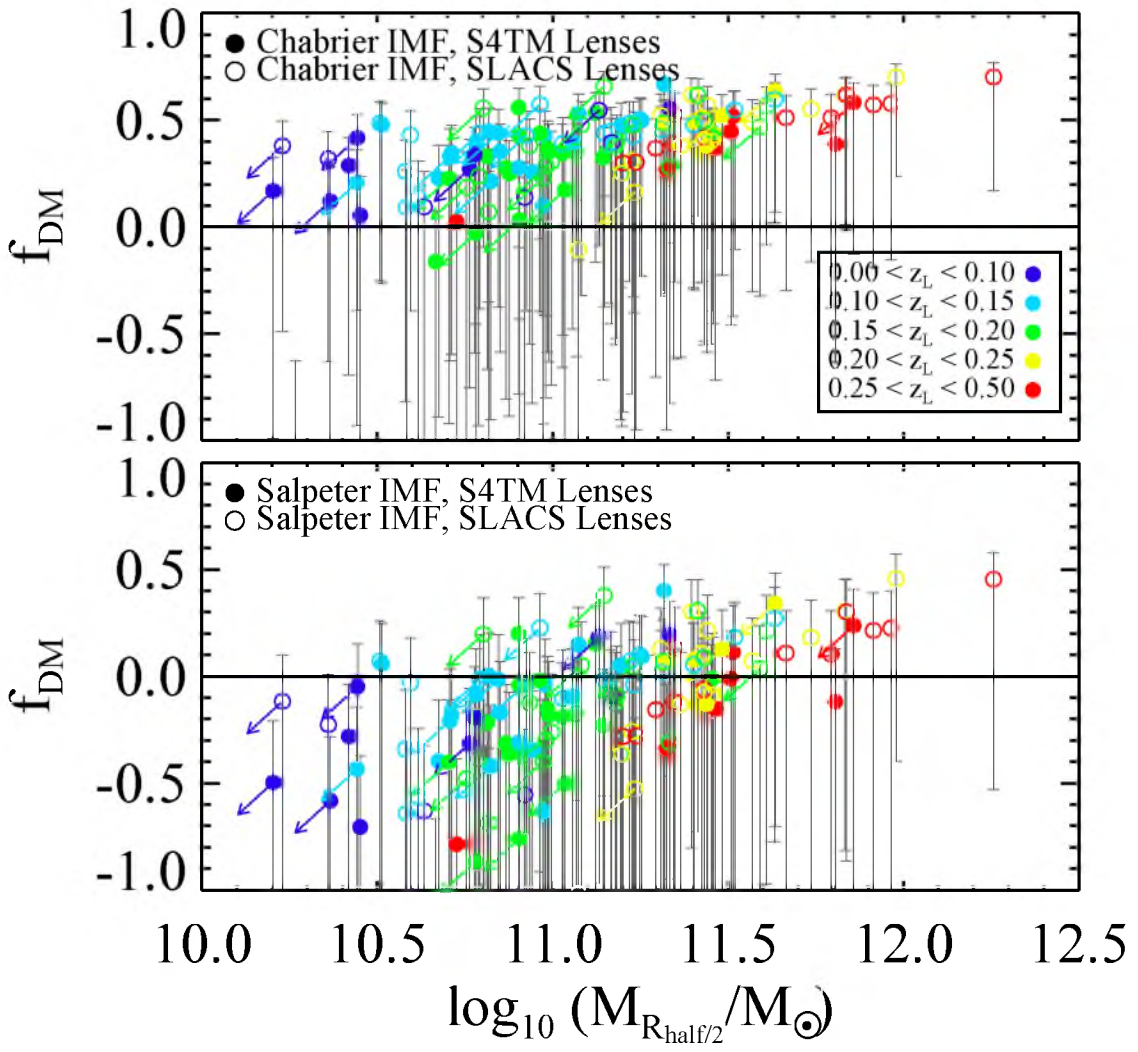

Figure 3.10: Relations between the dark-matter fraction within the Einstein radius $f_{\rm dm}$ and $\log_{10} M_{\rm R_{half}/2}$, $\log_{10} M_*^{\rm Chabrier}$, $\log_{10} \sigma_{\rm SIE}$, $\log_{10} \sigma_*$, $\log_{10} {\rm R_{half}}$, and ${\rm R_{SIE}/R_{half}}$. The arrows indicate the upper-limits of $f_{\rm dm}$, $\log_{10} M_{\rm R_{half}/2}$, $\log_{10} \sigma_{\rm SIE}$, and ${\rm R_{SIE}/R_{half}}$, for grade-C lenses and colors represent the redshifts of the lens galaxies. Gray error bars represent the systematic variations. Either a Chabrier or Salpeter IMF is considered.

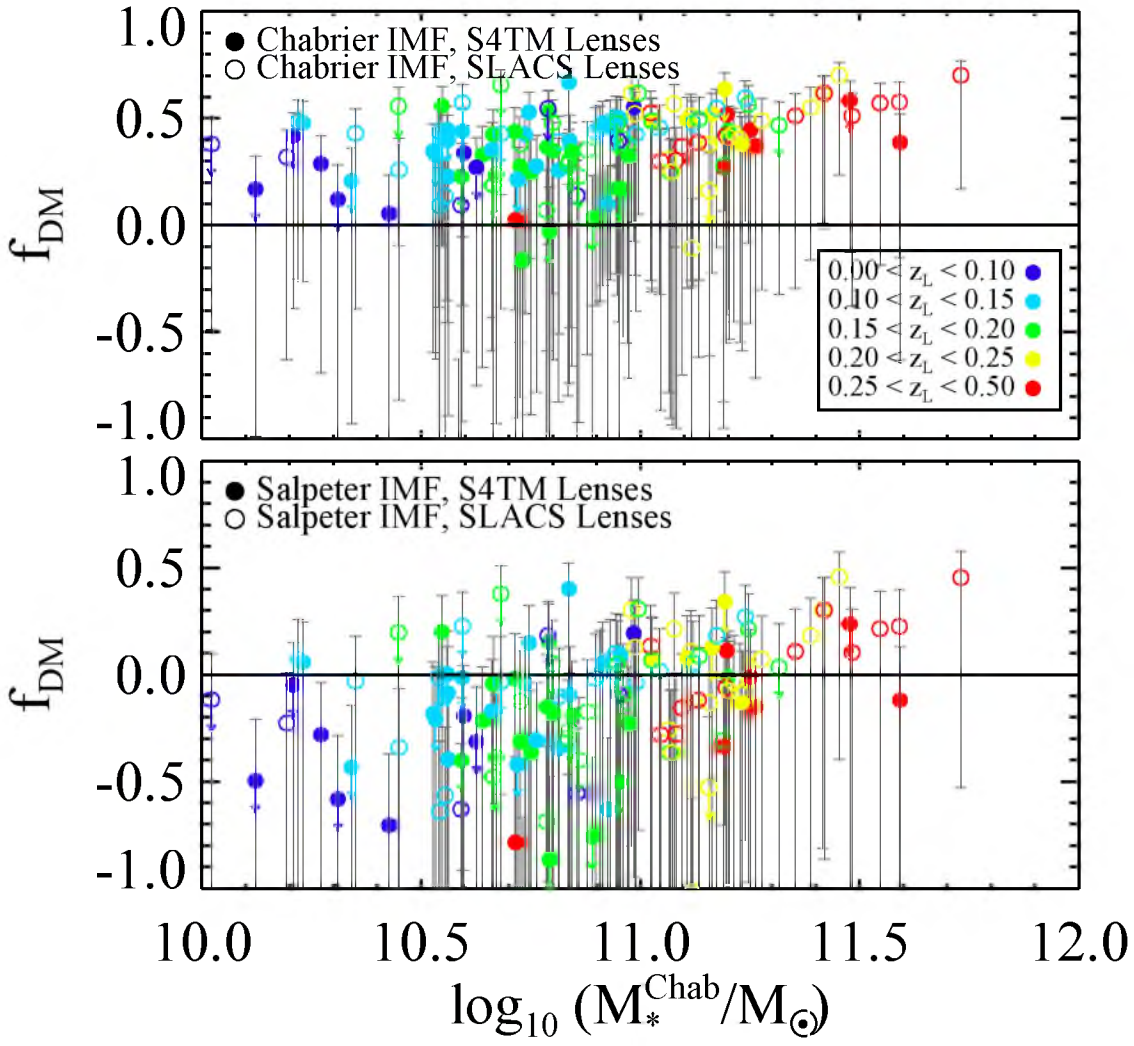

Figure 3.10: *Continued*

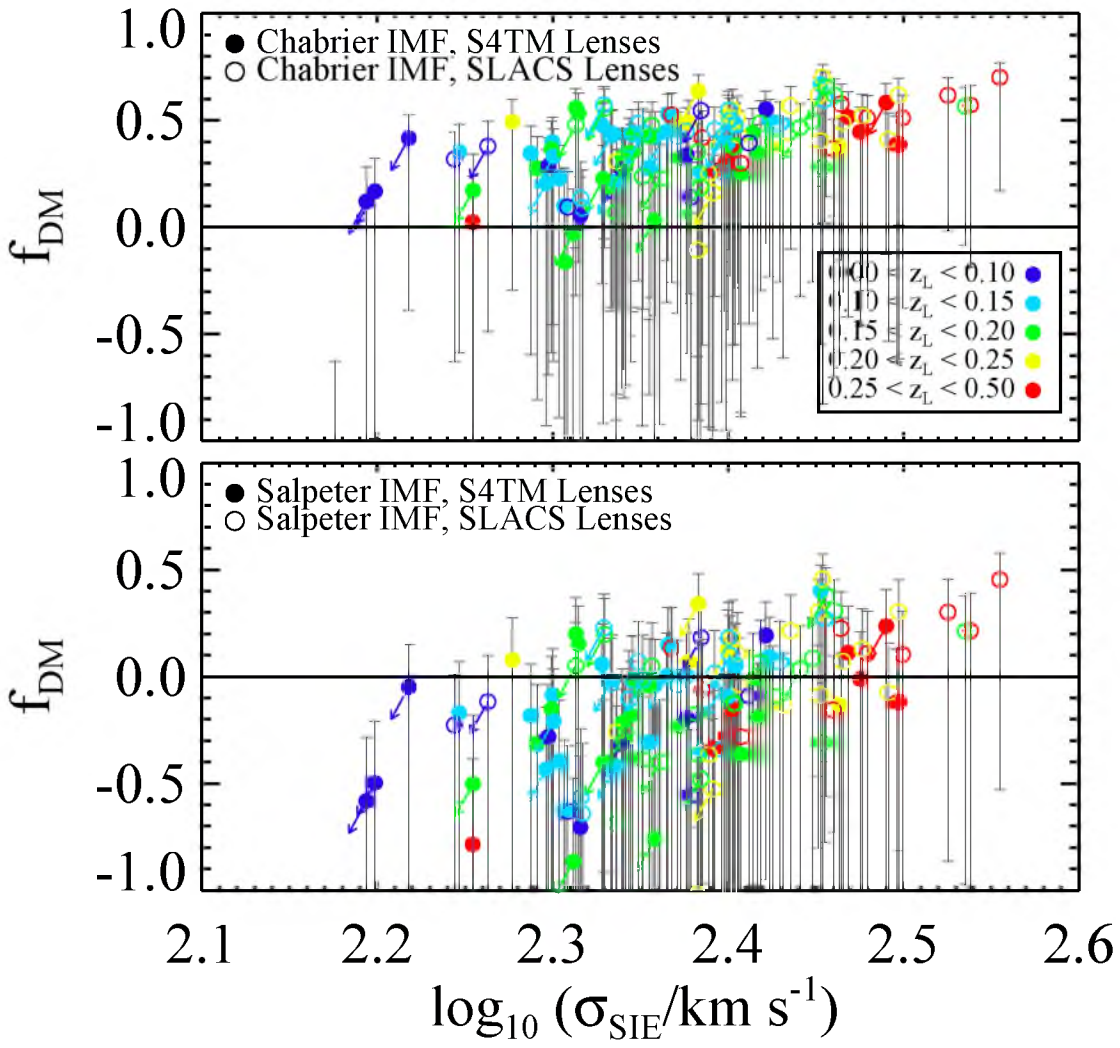

Figure 3.10: *Continued*

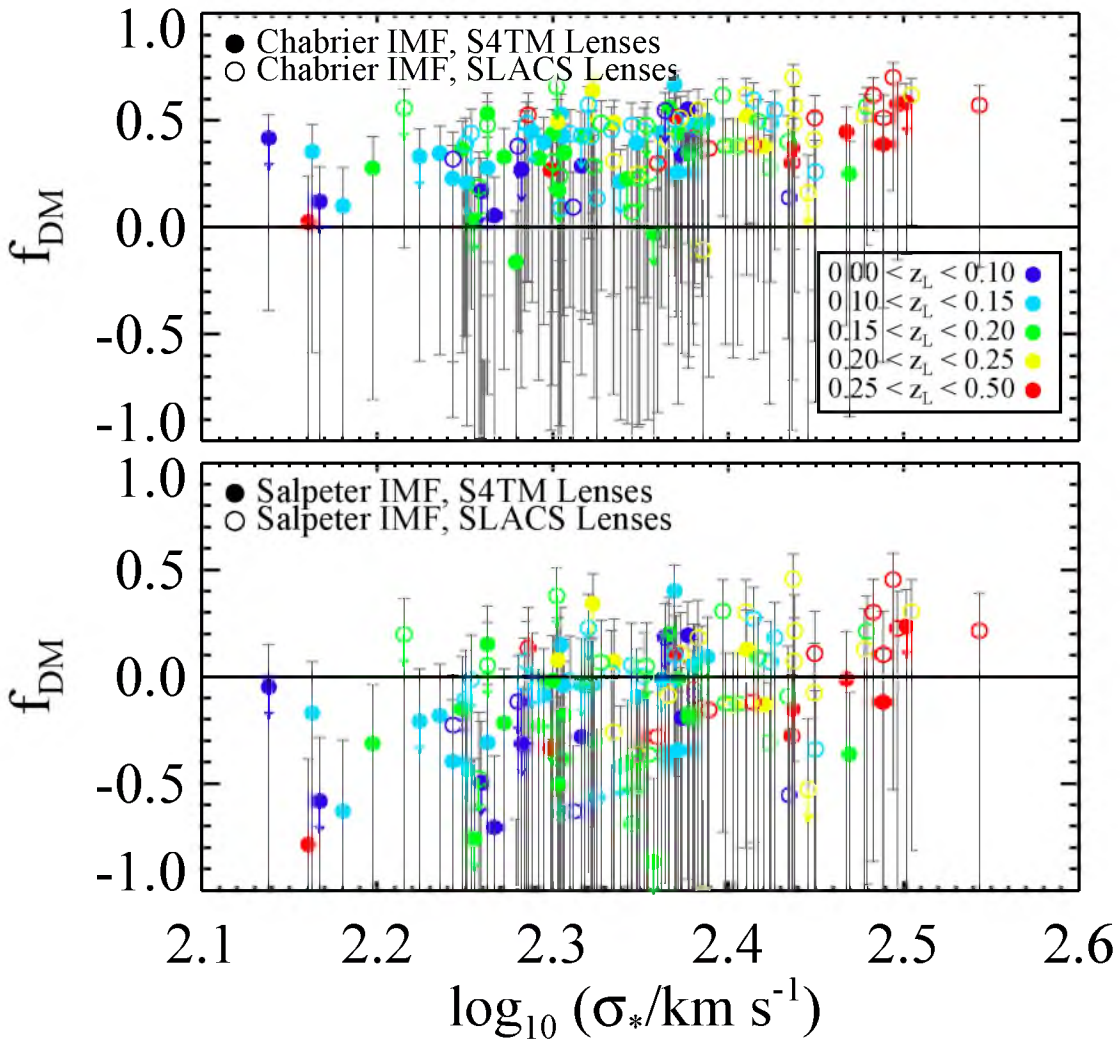

Figure 3.10: *Continued*

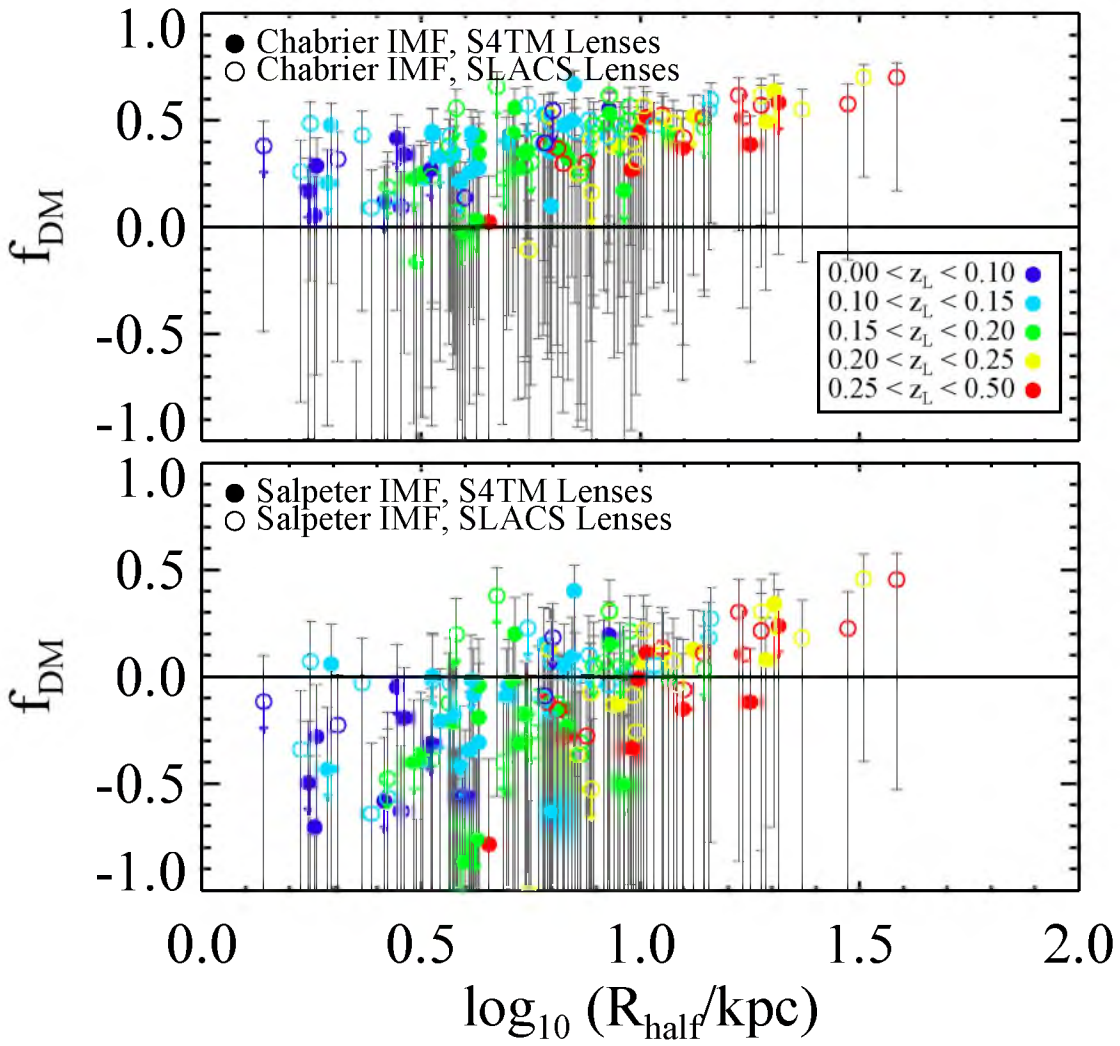

Figure 3.10: *Continued*

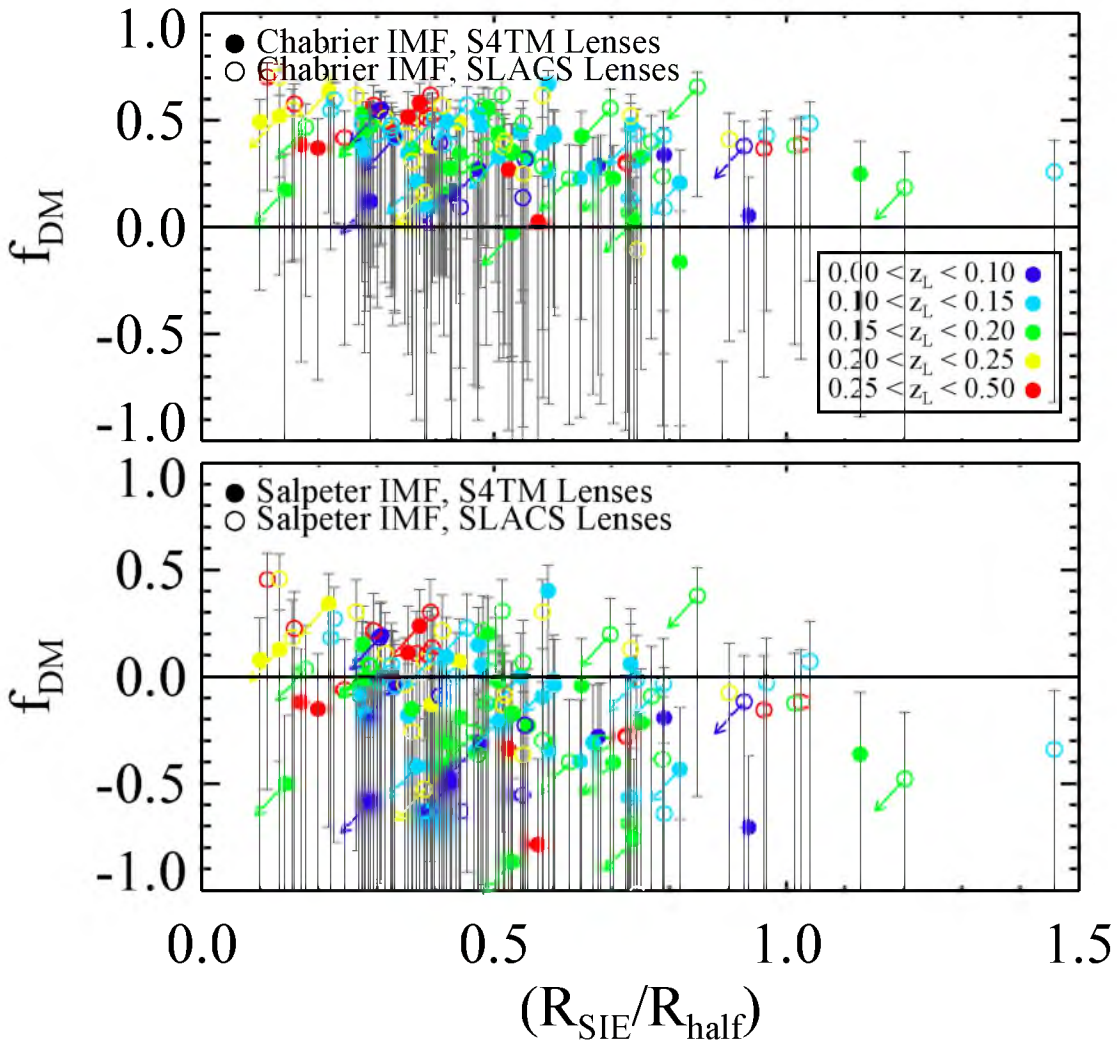

Figure 3.10: *Continued*

CaH1 (Conroy and van Dokkum, 2012; Spiniello et al., 2014). Improper interpretations in the SPS model, or lack of coverage of these features in the observed spectra, can lead to significant systematics in constraining the IMF (Conroy and van Dokkum, 2012; Spiniello et al., 2014).

3.6 Discussion and Conclusion

In this paper, we report the discovery of 40 strong gravitational lenses with clear and definite multiple images (classified as "grade-A") and another 18 single-image lenses (classified as "grade-C") from the S4TM survey (*HST* Program ID 12210 , Brownstein et al. 2014 (in preparation)), which by design selects lens galaxies with lower masses and smaller Einstein radii compared to the previous SLACS survey. Along with findings in the SLACS survey, we construct a statistically significant and more complete ensemble of over 100 gravitational lenses, including 98 grade-As and 33 grade-Cs. This combined sample probes ETGs with a mean lens redshift of $\langle z_L \rangle = 0.18$ and total enclosed mass within the Einstein radius $M_{\rm Ein}$ as low as $1.34 \times 10^{10} M_{\odot}$. All lenses have been modeled individually and measurements/upper limits of the Einstein radii have been obtained for grade-A/C lens galaxies appropriately.

3.6.1 Discussion

We have discovered clear evidence for the dependence of the total mass-density profile on galaxy velocity dispersion, in the sense that less massive (lower velocity-dispersion) lens galaxies have more centrally concentrated (super-isothermal) profiles. We have obtained this result by performing a hierarchical Bayesian analysis of the relation between $\log_{10} \sigma_*$ and $\log_{10} \sigma_{\rm SIE}$ for the combined lens ensemble. The inclusion of grade-C lenses is essential to this discovery: the significance of the trend is about 4-σ when including grade-A and grade-C lenses together in the analysis, but only about 1-σ (i.e., consistent with no trend) when analyzing grade-A lenses alone. This can be attributed both to the fact that the grade-C lenses extend the mass baseline of the measurement to lower masses, and to the fact that excluding grade-C lenses will bias the sample towards higher values of $\sigma_{\rm SIE}$ at fixed σ_*.

A trend of mass-density profile γ' upon the surface stellar mass density $\Sigma_* = M_*/(2\pi R_{\rm eff}^2)$ has been found in Auger et al. (2010), Dutton and Treu (2013), and Sonnenfeld et al. (2013) using either SLACS lenses or lenses from the Strong Lensing Legacy Survey (SL2S, Gavazzi et al., 2012). In particular, they found that galaxies with denser stellar mass densities have steeper profiles. Since the stellar velocity dispersion can be approximated as

$\sigma_*^2 \propto M_*/R_{\rm eff} \propto \Sigma_* R_{\rm eff}$, the γ'-σ_* relation and γ'-Σ_* relation suggest that as our lens galaxies grow their mass, they become spatially less concentrated (smaller Σ_*) and dynamically "hotter" (larger σ_*). This can be understood in terms of the expectations from baryonic physics within DM halos. Dissipative gas processes lead to higher baryon densities in the center of DM halos, and also steepen the DM halo profile through the effect of adiabatic contraction (AC: e.g., Blumenthal et al. 1986). Energetic feedback processes from SN and AGN tend to heat gas and counteract central condensation. These processes compete against each other in their effect on the mass distribution in DM halos. For less massive galaxies, the impact of feedback is less significant as compared to AC and hence leads to a more centrally concentrated halo. The importance of feedback increases as galaxies become more massive, resulting in shallower density profiles. Metallicity, environment, and other processes are also responsible for this competition. In order to determine whether the effect is regulated primarily by velocity dispersion or by stellar density, multiband data sufficient for detailed stellar-population analysis will be required for the full S4TM sample.

We have applied a simplified SPS analysis to the *HST* I-band photometric data to estimate the stellar masses of the lens galaxies assuming either a Chabrier or a Salpeter IMF. Age, metallicity, and dust have been chosen to match the typical values for passively evolving ETGs at low redshifts. A clear correlation between the projected dark matter fraction and the total mass is observed for both IMFs, consistent with previous findings (e.g., Tortora et al., 2009; Auger et al., 2010; Cappellari et al., 2012; Conroy et al., 2013; Brewer et al., 2014). There are two possible interpretations of this result: as a true increase in dark-matter fraction with velocity dispersion, or as a trend in the stellar IMF with velocity dispersion. The first interpretation aligns with the overall expectation of decreased star-formation efficiency with increasing halo mass for halos above $\sim 10^8\,M_\odot$(e.g., Behroozi et al., 2010). This interpretation could also explain our observed trend in dark-matter profile slope: for a fixed stellar profile shape, an increased fractional mass contribution from a more spatially extended DM halo will result in a shallower total-mass density profile. Alternatively, a trend towards a more bottom-heavy IMF in more massive galaxies can cause the apparent effect of an increased DM fraction when a single IMF is assumed across all masses (e.g., Treu et al., 2010; van Dokkum and Conroy, 2010; Strader et al., 2011; Sonnenfeld et al., 2012; Spiniello et al., 2012; Ferreras et al., 2013; La Barbera et al., 2013; Geha et al., 2013; Conroy et al., 2013; Tortora et al., 2013; Spiniello et al., 2014). The Salpeter IMF is in any event disfavored at the low-mass end ($M_{\rm R_{half}/2} \lesssim 10^{10.8} M_\odot$), since it results in unphysical negative DM fractions.

3.6.2 Conclusion and Future Work

To conclude, in this paper,

1. We report the discovery of 40 new grade-A and 33 new grade-C ETG lenses from the S4TM and SLACS surveys. Besides the measurements for grade-A lenses, upper limits of the Einstein radii for grade-C lenses are determined for the first time. Combining with 58 grade-A ETG lenses from the SLACS survey, we construct an unbiased ETG lens ensemble with wider mass coverage than previous strong-lens samples;

2. Applying a hierarchical Bayesian method which utilizes both the measurements and upper limits of the Einstein radii, we study the correlation between $\log_{10}\sigma_*$ and $\log_{10}\sigma_{\rm SIE}$, and find a less-than-unity slope of 0.78 ± 0.05 which corresponds to a significant ($\approx 4\sigma$) dependence of total mass-density profile on the lens mass in the sense that more massive ETGs possess shallower profiles (as quantified by the ratio of σ_* to $\sigma_{\rm SIE}$ which serves as a proxy for the logarithmic mass-density profile slope γ'). We have shown that this trend is only significant when grade-C lenses are included (the slope is 0.93 ± 0.07 for the grade-A only subsample) which, highlights the importance of grade-C lenses to enabling a wider and less biased coverage of lens masses;

3. Stellar masses of lens galaxies are estimated based on their *HST* I-band photometry and SPS models assuming either a Chabrier or Salpeter IMF. The resulting DM fractions within one half of the half-light radius $f_{\rm DM}$ for each IMF model are found to be strongly correlated with the lens mass/velocity dispersion in the sense that more massive ETGs have larger DM fractions, or alternatively mass-dependent IMFs (or a combination of both effects). A Salpeter IMF is ruled out for ETGs with mass less than $10^{10.8} M_\odot$ or velocity dispersion smaller than 180 km/s by implying negative $f_{\rm DM}$.

The analysis of our new S4TM lens sample in combination with other lens samples can be improved with spatially resolved long-slit or integral field spectroscopy in order to determine the two-dimensional stellar kinematics of the lenses, which can in turn enable detailed lensing-plus-dynamical modeling to better constrain the mass-density profile for individual galaxies and eventually break the mass-sheet degeneracy (Barnabè et al., 2009a; McKean et al., 2010; van de Ven et al., 2010; Newman et al., 2011; Barnabè et al., 2011; Dutton et al., 2011). Multiband photometry or spectroscopy covering a wide wavelength range from near ultra-violet (NUV) to near infrared (NIR) for the lens galaxies would similarly yield better constraints on the age, metallicity, dust, and other parameters in the

lens-galaxy SPS models. Finally, more sophisticated lens models with separate components for dark matter and stars would also improve upon our current single-component total-mass models (e.g., Gavazzi et al., 2007; Dutton et al., 2011; Barnabè et al., 2012; Vegetti et al., 2010, 2012). With uniform data on a comprehensive lens sample from the SLACS, S4TM, and BELLS surveys covering a wide range in lens redshift ($0.1 < z < 0.7$) and total enclosed mass ($10^{10}\mathrm{M}_{\odot} < \mathrm{M}_{\mathrm{Ein}} < 10^{12}\mathrm{M}_{\odot}$), we can fully explore the variation of ETG mass structure across galaxy mass and cosmic time through a joint analysis of strong lensing, stellar dynamics, and stellar populations.

CHAPTER 4

JOINT ANALYSIS OF STRONG LENSING AND STELLAR KINEMATICS[†]

In this chapter, I outline in detail the idea of a joint analysis of strong GL and two-dimensional stellar kinematics as a powerful technique to constrain the mass-density profiles of lens galaxies and separately study the luminous and dark components. Applying this joint analysis by Bolton et al. (2012b) to a combined lens sample discovered by the SLACS and BELLS Surveys over the redshift interval from 0.1 to 0.6, we detected a significant redshift dependence of the logarithmic total mass-density slope γ' with magnitude $\mathrm{d}\langle\gamma'\rangle/\mathrm{d}z = -0.60 \pm 0.15$. This 4σ evolution trend toward shallower mass profiles at higher redshifts unravels the significant role of off-axis major dry mergers during the process of galaxy assembly over the past 6 Gyrs.

4.1 Introduction

The total mass-density profile of ETGs is a highly instructive fossil record of galaxy formation and evolution. DM-only simulations find that the mass distribution can be well-fitted by a universal "NFW" profile (Navarro et al., 1996, 1997). However, tension arises when comparing to various observational facts (Moore et al., 1999b; Graham et al., 2006; Navarro et al., 2010), the solution of which lies in the physical processes brought by the baryons reside in real galaxies, such as adiabatic contraction, heating mechanisms, and other energetic feedbacks (Gnedin et al., 2004; Nipoti et al., 2004; Gustafsson et al., 2006; Romano-Díaz et al., 2008; Abadi et al., 2010; Governato et al., 2010; Duffy et al., 2010; Martizzi et al., 2012; Velliscig et al., 2014). A detailed study of the mass structure of ETGs permits examination and quantification of the effects of baryonic physics.

Strong GL studies have already found that the total mass-density profile in the central regions of massive ETGs at low redshift ($z < 1$) can be well-described by a simple power-

[†]A method implemented by Yiping Shu, applied in the cited paper—Bolton A. S., Brownstein, J. R., Kochanek, C. S. , **Shu, Yiping**, Schlegel, D. J., Eisenstein D. J., Wake, D. A., Connolly, N., Maraston, C., Arneson, R. A., Weaver, B. A., "*The BOSS Emission-Line Lens Survey.II. Investigating Mass-density Profile Evolution in the SLACS+BELLS Strong Gravitational Lens Sample*," 2012, ApJ, 757, 82.

law model (Koopmans et al., 2006b; Gavazzi et al., 2007; Koopmans et al., 2009; Bolton et al., 2012b). However, for a general power-law density profile other than the SIE model, the essential deflection angle for gravitational lens modeling can only be approximated by rather expensive numerical approaches. Additionally, when considering GL signals alone, the known mass-profile, mass-sheet, and families of other degeneracies severely prevent an accurate determination of the true mass-density profile (e.g., Gorenstein et al., 1988; Saha, 2000; Wucknitz, 2002; Treu et al., 2010; Liesenborgs and De Rijcke, 2012; Schneider and Sluse, 2013). On the other hand, the two-dimensional stellar kinematics can in principle unravel the mass structure of ETGs because the motion of any individual star within the galaxy is purely governed by its gravitational potential and the bulk effect is characterized by the Jeans' equations (Jeans, 1915; Binney and Tremaine, 1987). However, in reality, this technique is limited by the projection effect and insufficient observational power.

The joint analysis of strong GL and two-dimensional stellar kinematics has been shown to be a promising technique to break some of the known degeneracies and lead to a tighter constraint on the total mass-density profile in the central regions of massive ETGs (Treu and Koopmans, 2002, 2004; Grillo et al., 2008; Koopmans et al., 2009; Ruff et al., 2011; Barnabè et al., 2011; Spiniello et al., 2011). However, the lack of a significant lens sample with a wide redshift coverage and an appropriate analyzing tool has strongly limited the thorough exploration of the mass structure. In this work, we took advantage of a combined lens sample obtained by the SLACS (Bolton et al., 2008a) and BELLS (Brownstein et al., 2012) Surveys using the same selection technique and implemented a Bayesian approach to investigate the mass structures of individual lens galaxies and any evolution trend across the entire lens-galaxy population.

4.2 Dynamical Modeling

Following the notations in Binney and Tremaine (1987), the distribution function (DF) $f(\vec{r}, \vec{v}, t)$ of stars within a galaxy satisfies the continuity equation

$$\frac{\partial f}{\partial t} + \sum_{\alpha=1}^{6} \frac{\partial}{\partial w_\alpha}(\dot{w}_\alpha f) = 0 \tag{4.1}$$

where $\vec{w} = (\vec{r}, \vec{v})$, $\dot{\vec{w}} = (\vec{v}, -\nabla\Phi)$ and Φ is the gravitational potential. This equation can be further simplified a s

$$\frac{\partial f}{\partial t} + \sum_{\alpha=1}^{6} \dot{w}_\alpha \frac{\partial f}{\partial w_\alpha} = 0 \tag{4.2}$$

because v_i and x_i are independent variables and Φ does not depend on v_i. Equation 4.2 is known as the "collisionless Boltzmann equation."

In a spherical coordinate system, the collisionless Boltzmann equation becomes

$$\frac{\partial f}{\partial t}+v_r\frac{\partial f}{\partial r}+\frac{v_\theta}{r}\frac{\partial f}{\partial \theta}+\frac{v_\phi}{r\sin\theta}\frac{\partial f}{\partial \phi}+(\frac{v_\theta^2+v_\phi^2}{r}-\frac{\partial\Phi}{\partial r})\frac{\partial f}{\partial v_r}+\frac{1}{r}(v_\phi^2\cot\theta-v_rv_\theta-\frac{\partial\Phi}{\partial\theta})\frac{\partial f}{\partial v_\theta}(4.3)$$
$$-\frac{1}{r}(v_rv_\phi+v_\theta v_\phi\cot\theta+\frac{1}{\sin\theta}\frac{\partial\Phi}{\partial\phi})\frac{\partial f}{\partial v_\phi}=0$$

For a spherical symmetric system, it is reduced to

$$\frac{\partial f}{\partial t}+v_r\frac{\partial f}{\partial r}+(\frac{v_\theta^2+v_\phi^2}{r}-\frac{\mathrm{d}\Phi}{\mathrm{d}r})\frac{\partial f}{\partial v_r}+\frac{1}{r}(v_\phi^2\cot\theta-v_rv_\vartheta)\frac{\partial f}{\partial v_\theta} \tag{4.4}$$
$$-\frac{1}{r}(v_rv_\phi+v_\theta v_\phi\cot\theta)\frac{\partial f}{\partial v_\phi}=0$$

Define the velocity moments as

$$\nu(r)\bar{v_r^{2i}v_\theta^{2j}v_\phi^{2k}}=\int v_r^{2i}v_\theta^{2j}v_\phi^{2k}f(\vec{r},\vec{v},t)\,\mathrm{d}^3\vec{v} \tag{4.5}$$

where $\nu(r)$ is the spatial density of stars and also the normalization factor

$$\nu(r)=\int f(\vec{r},\vec{v},t)\,\mathrm{d}^3\vec{v}. \tag{4.6}$$

Note that any odd velocity moment actually vanishes. Multiplying Equation 4.4 by v_r and integrating over $\mathrm{d}^3\vec{v}$, we get the *three-dimensional spherical Jeans equation* as

$$\frac{\mathrm{d}}{\mathrm{d}r}[\nu(r)\bar{v_r^2}]+\frac{2\bar{v_r^2}-\bar{v_\theta^2}-\bar{v_\phi^2}}{r}\nu(r)=-\nu\frac{\mathrm{d}\Phi}{\mathrm{d}r} \tag{4.7}$$

We assume an Osipkov-Merritt (Osipkov, 1979; Merritt, 1985b,a) parameterization of the anisotropy β

$$\beta(r)=1-\frac{\bar{v_\theta^2}}{\bar{v_r^2}}, \tag{4.8}$$

and the velocity moments are invariant under rotation about the galactic center, namely

$$\bar{v_\theta^2}=\bar{v_\phi^2} \tag{4.9}$$

Then the spherical Jeans equation becomes

$$\frac{\mathrm{d}}{\mathrm{d}r}[\nu(r)\bar{v_r^2}]+\frac{2\beta}{r}[\nu(r)\bar{v_r^2}]=-\nu\frac{\mathrm{d}\Phi}{\mathrm{d}r} \tag{4.10}$$

which is the building block of our dynamical modeling as it connects the gravitational potential to the observable radial stellar velocity dispersion. Note that there are 4 unknowns $\nu(r)$, $\beta(r)$, $M(r)$, and $\bar{v_r^2}$ in total in the spherical Jeans equation, but only projected two-dimensional luminosity and velocity-dispersion distributions can be measured. It implies that we can *not* obtain a unique mass model $M(r)$ without making further assumptions on the anisotropy $\beta(r)$, which is the well-known "mass-anisotropy" degeneracy.

One feasible solution is to assume a constant anisotropy parameter, namely β does not depend on the radius r. Then we can solve the spherical Jeans equation (Equation 4.10)

$$\nu(r)\sigma_r^2(r)r^{2\beta}=\int_r^\infty GM(<r')\nu(r')r'^{2\beta-2}\,\mathrm{d}r' \tag{4.11}$$

Note that here I replace the radial velocity moment $\bar{v_r^2}$ by the radial velocity dispersion $\sigma_r(r)$ defined as

$$\sigma_r^2(r)=\bar{v_r^2}-\bar{v_r}^2 \tag{4.12}$$

under the assumption that the mean radial velocity is zero for a spherical symmetric velocity distribution. The three-dimensional luminosity density $\nu(r)$ is related to the observed two-dimensional surface brightness density $I(R)$ by the Abel transform

$$\nu(r) = -\frac{1}{\pi}\int_r^\infty \frac{\mathrm{d}I(R)}{\mathrm{d}R}\frac{\mathrm{d}R}{\sqrt{R^2-r^2}} \tag{4.13}$$

So given a particular total mass-density profile, one can try to work out the integral on the right-hand side of Equation 4.11 to obtain $\sigma_r(r)$. Then the line-of-sight velocity dispersion profile $\sigma_\parallel^2(R)$ is given by

$$I(R)\sigma_\parallel^2(R) = 2\int_R^{+\infty}(1-\beta\frac{R^2}{r^2})\frac{\nu(r)\sigma_r^2(r)r\,\mathrm{d}r}{\sqrt{r^2-R^2}} \tag{4.14}$$

4.3 Application to Lens Galaxies

The application of the above method that I implemented was done by Bolton et al. (2012b). In what follows, I summarize the procedures and findings in that work. The lens sample includes 79 confirmed grade-A lenses discovered by the SLACS (57, Bolton et al., 2008a) and BELLS (22, Brownstein et al., 2012) Surveys, two high-resolution *HST* imaging follow-ups on spectroscopically selected lens candidates from the Sloan and BOSS surveys, respectively. It is a quite homogeneous sample in the sense that all the lens systems were selected in exactly the same way and lens galaxies have comparable stellar masses. In terms of redshift, SLACS lenses are relatively nearby, and BELLS lenses are the high-redshift counterpart with $0.3 < z_L < 0.7$.

For each lens, we obtain the line-of-sight stellar velocity dispersion σ_i from the observed galaxy spectrum based on the method explained in Chapter 2 and Bolton et al. (2012a). However, in this work, instead of using the ELODIE eigenspectra, we generate a new set of stellar templates from the Indo-US library (Valdes et al., 2004) patched and extended by selected synthetic spectra in POLLUX database (Palacios et al., 2010), covering the full wavelength range of observed spectra. We restrict to A-K stars when generating the eigenspectra and only use the first five principal components as the velocity-dispersion template basis. At each trial velocity-dispersion value, we perform a least-squares fitting to the observed lens galaxy spectrum and record the best-fit chi^2. Eventually, we obtain a $\chi_i^2(\sigma_i)$ function for each galaxy, which will be further used to constrain the logarithmic density slope γ'.

The surface-brightness distribution of each lens galaxy is modeled by the so-called "Nuker" profile described in (Lauer et al., 1995)

$$I(R) = 2^{(\beta-\gamma)/\alpha} I_b(\frac{R_b}{R})^\gamma[1+(\frac{R}{R_b})^\alpha]^{(\gamma-\beta)/\alpha} \tag{4.15}$$

which is a broken power law with break radius R_b and sharpness of transition characterized

by α. The best-fit parameters are obtained by optimizing the point-spread-function (PSF) convolved Nuker models to best fit the *HST* I-band reduced imaging data inside a circular region of $6''$ centered on the lens galaxy. The robustness of this Nuker model was verified by comparing the results obtained by assuming different surface-brightness profiles and weighting schemes.

The total mass-density profile is chosen to be a simple power-law model following Koopmans et al. (2006b, 2009)

$$\rho(r) = \rho_0 r^{-\gamma'} \tag{4.16}$$

The strong-lensing observations allow us to eliminate ρ_0 in the power-law model by requiring that the total mass within the Einstein radius for each lens should be

$$M(< D_d\theta_{\rm Ein}) = \frac{c^2}{4G}\frac{D_d D_s}{D_{ds}}\theta^2_{\rm Ein} \tag{4.17}$$

Adopting these ingredients into the spherical Jeans equation, we obtain the line-of-sight velocity dispersion profiles for each lens galaxy as

$$\sigma^2_{\parallel}(R,\gamma') = \frac{2}{I(R)}\int_R^{+\infty}\frac{\nu(r)\sigma^2_r(r)r\,{\rm d}r}{\sqrt{r^2-R^2}} \tag{4.18}$$

Note that for simplicity, we assume the anisotropy parameter $\beta = 0$. To compare with observations, we convert $\sigma^2_{\parallel}(R,\gamma')$ to the luminosity-weighted line-of-sight velocity dispersion within the observational apertures R_{ap} ($1.5''$ in radius for SLACS lenses and $1''$ in radius for BELLS lenses) as

$$\sigma^2_{\rm pred}(\gamma') = \frac{\int_0^{R_{ap}} W(R)I(R)\sigma^2_{\parallel}(R,\gamma')2\pi R\,{\rm d}R}{\int_0^{R_{ap}} W(R)I(R)2\pi R\,{\rm d}R} \tag{4.19}$$

where $W(R)$ is the window function which takes care of the blurring due to the $1.8''$ seeing. It is worth noting that $\sigma_{\rm pred}(\gamma')$ depends on the specific mass profile used and therefore depends on the logarithmic density slope γ'.

4.4 Mass Structure Evolution

To explore the mass structure evolution among the lens galaxies, we parameterize the conditional PDF of γ' at a given z as a Gaussian form

$$\Pr(\gamma'|z,\gamma_0,\gamma_z,s_\gamma) = \frac{1}{\sqrt{2\pi}s_\gamma}\exp\{-\frac{[\gamma'-(\gamma_0+\gamma_z(z-0.25))]^2}{2s^2_\gamma}\} \tag{4.20}$$

with γ_0 the mean value at $z = 0.25$, γ_z the evolution magnitude of the mean γ', and s_γ the intrinsic scatter in γ'. To quantify these parameters based on the joint analysis of GL and stellar kinematics as outlined in the previous sections, we again apply the hierarchical Bayesian method in which γ_0, γ_z, and s_γ are the hyperparameters that need to be determined. The likelihood function of the hyperparameters for the entire lens population is

$$\mathscr{L}(\gamma_0, \gamma_z, s_\gamma|\vec{d}) \equiv \Pr(\vec{d}|\gamma_0, \gamma_z, s_\gamma) = \prod_i \Pr(d_i|\gamma_0, \gamma_z, s_\gamma)$$
$$= \prod_i \int \Pr(d_i|\gamma')\Pr(\gamma'|\gamma_0, \gamma_z, s_\gamma)\,\mathrm{d}\gamma' \quad (4.21)$$

The first term in the integral $\Pr(d_i|\gamma')$ is the conditional PDF of data d_i given γ' which comes from comparing the observed line-of-sight stellar velocity dispersion σ_i to the predicted value for a particular γ' value $\sigma_{\rm pred}$

$$\Pr(d_i|\gamma') \propto \exp[-\frac{\chi^2(\sigma_{\rm pred}(\gamma'))}{2}] \quad (4.22)$$

The posterior PDF of the hyperparameters $\Pr(\gamma_0, \gamma_z, s_\gamma|\vec{d})$ is given by the Bayes' theorem

$$\Pr(\gamma_0, \gamma_z, s_\gamma|\vec{d}) = \mathscr{L}(\gamma_0, \gamma_z, s_\gamma|\vec{d})\Pr(\gamma_0, \gamma_z, s_\gamma)/\Pr(\vec{d}) \quad (4.23)$$

Again, the prior of data $\Pr(\vec{d})$ can be simply treated as a normalization factor and the prior of the hyperparameters is assumed to be uniform in γ_0, γ_z, and $\log_{10} s_\gamma$. So the posterior PDF of the hyper parameters is proportional to the likelihood function

$$\Pr(\gamma_0, \gamma_z, s_\gamma|\vec{d}) = \mathscr{L}(\gamma_0, \gamma_z, s_\gamma|\vec{d}) \quad (4.24)$$

Bolton et al. (2012b) first derived the best-estimated γ' values for individual galaxies based on the observed $\chi^2_i(\sigma_i)$ and predicted $\sigma_{\rm pred}$ and plotted them in Figure 4.1. The black diamonds represent SLACS lenses at lower redshifts and blue squares represent BELLS lenses at higher redshifts. It is clear that the logarithmic density slope of a lens galaxy is correlated with its redshift. The solid line represents the best-fit relation from a linear regression and the four gray lines show the amounts of variation by allowing $\pm 1\sigma$ change in the best-fit slope and the zero-point of this linear relation.

We discretized Equation 4.21 by using a gridded range of γ' values $1.1 < \gamma' < 2.9$ with a step-size of 0.02 and obtained the posterior PDF of the hyperparameters. Figure 4.2 visualizes the projected line-of-sight velocity dispersion $\sigma_\parallel$ of a particular lens as a function of the power-law index γ' and radius R. Figure 4.3 shows the confidence regions of the marginal posterior PDFs of γ_0 and γ_z for the entire lens-galaxy population based on the joint analysis of strong GL and stellar kinematics. Black and gray contours are the results assuming a Nuker profile or a de Vaucouleurs profile, respectively. First of all, we can see that the two surface-brightness profiles yield almost the same posterior PDFs. Secondly, there is a very significant deviation (4-5σ) from nonevolution mass structure for this lens sample, toward the trend that lens galaxies at lower redshifts have steeper mass-density profiles in their central regions. Taken at the face value, it seems like this work is in tension with previous finding of little redshift evolution of the mass-density profile from studies using only SLACS lenses (e.g., Koopmans et al., 2006b; Gavazzi et al., 2007). However, it can be easily seen from Figure 4.1 that this disagreement is due to the insufficient redshift coverage in SLACS lenses. By including high-redshift BELLS lenses, the anticorrelation

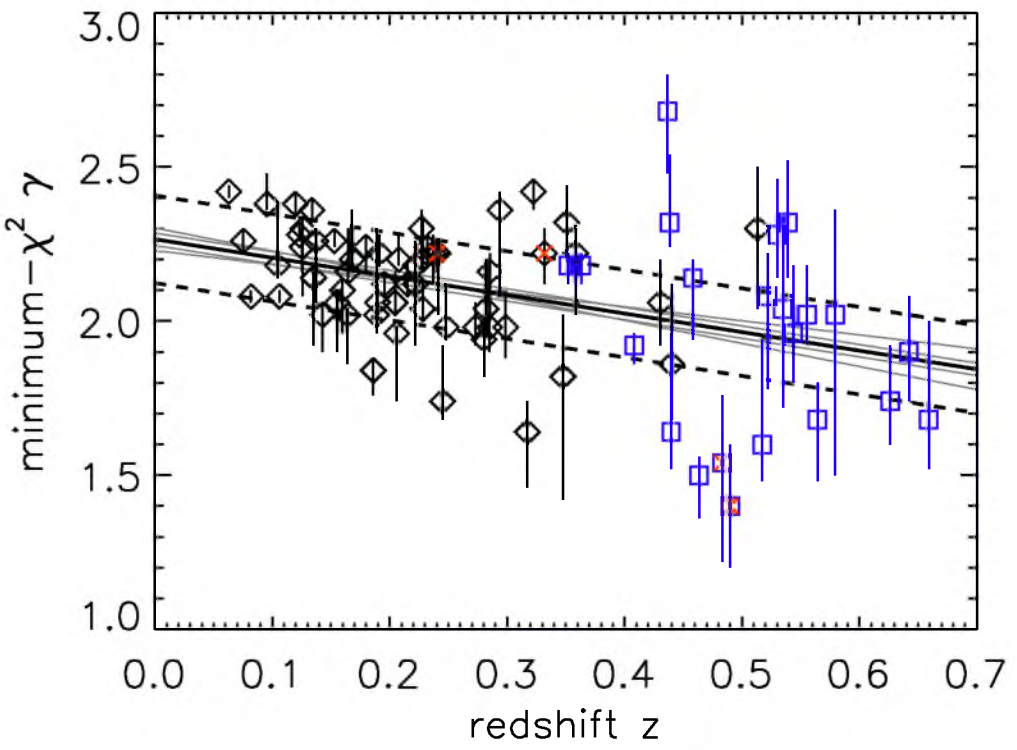

Figure 4.1: Best-estimated γ' values for individual galaxies as a function of lens redshift. Black diamonds are for SLACS lenses while blue squares are for BELLS lenses. The four red crosses indicate four lens galaxies with computed stellar velocity dispersions either greater than 316 km s^{-1} or less than 158 km s^{-1}. The meanings of all the lines are explained in the text. Credit: Figure 2 in the referred publication: "The BOSS Emission-Line Lens Survey. II. Investigating Mass-density Profile Evolution in the SLACS+BELLS Strong Gravitational Lens Sample," Bolton, A.S., et al., 2012, ApJ, 757, 82. Reproduced by permission of the AAS.

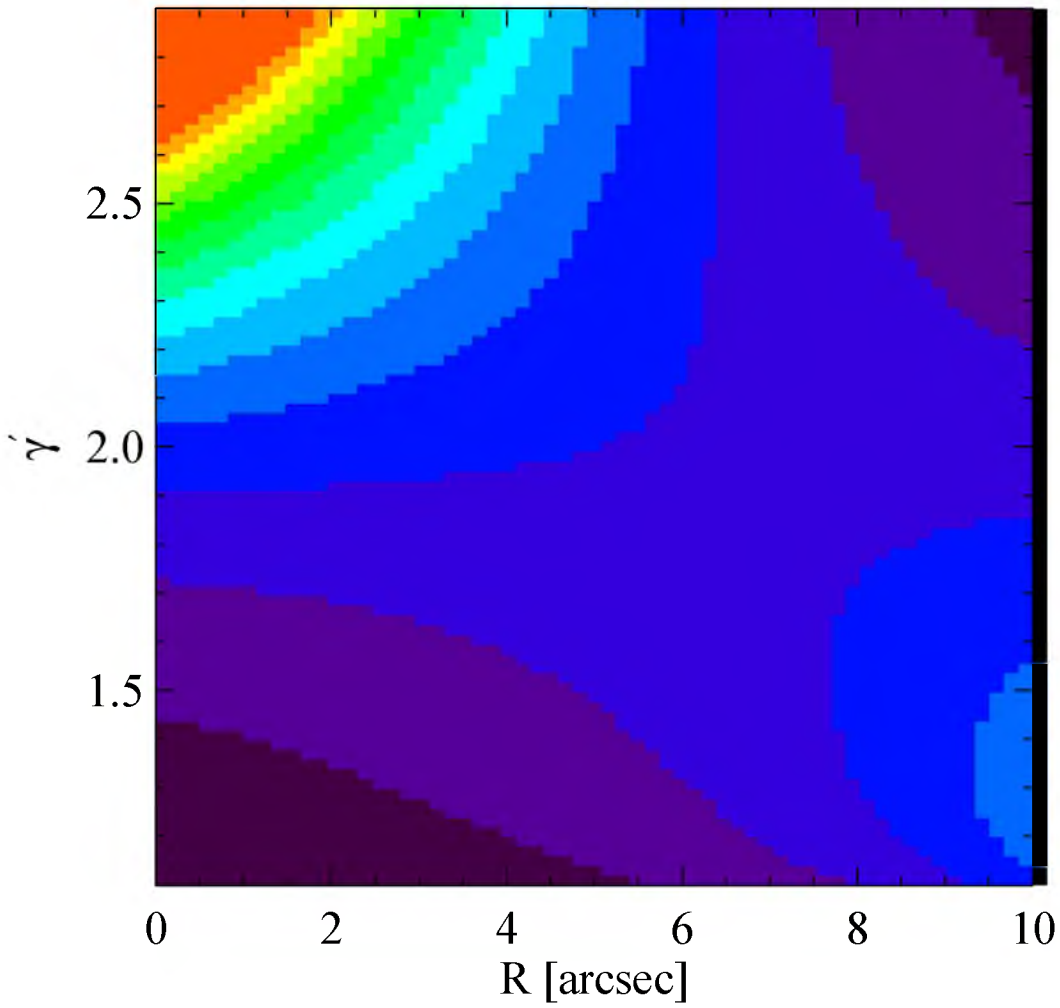

Figure 4.2: Projected line-of-sight velocity dispersion $\sigma_{\|}$ of a particular lens galaxy as a function of the power-law index γ' and radius R.

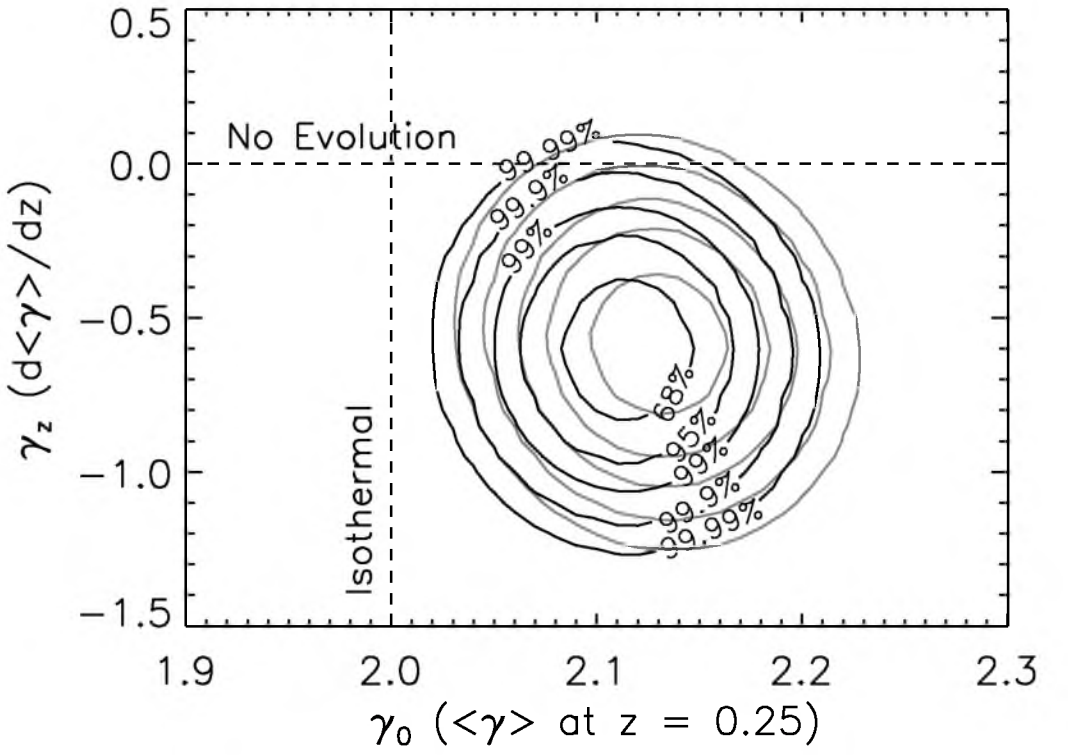

Figure 4.3: Marginal posterior PDFs of the mean and redshift evolution hyperparameters obtained by a hierarchical Bayesian analysis. Black contours are the results assuming a Nuker profile for the surface-brightness distribution while gray contours are for the de Vaucouleurs profile-based results. The horizontal dashed line indicates null evolution in the logarithmic density slope γ' and the vertical dashed line indicates an isothermal profile with $\gamma' = 2$. Credit: Figure 2 in the referred publication: "The BOSS Emission-Line Lens Survey. II. Investigating Mass-density Profile Evolution in the SLACS+BELLS Strong Gravitational Lens Sample," Bolton, A.S., et al., 2012, ApJ, 757, 82. Reproduced by permission of the AAS.

becomes much more significant. Finally, using strong GL as a mass constraint, solving the spherical Jeans equation, and comparing with the observed luminosity-weighted stellar velocity dispersion nail down the magnitude of the redshift-evolution of mean logarithmic slope to be $\gamma_z = \mathrm{d}\langle\gamma'\rangle/\mathrm{d}z = -0.60 \pm 0.15$.

4.5 Discussions

The detailed discussions of this work such as the selection effects and physical interpretations can be found in Bolton et al. (2012b). Here I just want to emphasize that this joint analysis of strong GL and stellar kinematics has been shown by various studies to be a very effective approach to better constrain the mass structure of ETGs in their central few kpcs' regions. However, it is clear that currently, this method is a highly simplified implementation of the dynamical model. In order to extend to a more general form, we need to treat the anisotropy β as a free parameter. Also if the two-dimensional stellar velocity dispersion profile of a lens galaxy can be measured by long-slit or integral field unit (IFU) instruments, we can directly compare it with the predicted two-dimensional line-of-sight velocity dispersion profile to get a much better understanding of the anisotropy and the mass structure, as compared to that obtained by using a single aperture-integrated velocity dispersion value. It has been suggested by Agnello et al. (2013) and Richardson and Fairbairn (2014) that the joint use of projected virial theorem can tightly constrain the anisotropy or even completely break the mass-anisotropy degeneracy. Furthermore, a more detailed lens model other than the simple SIE model is demanded for removing the mass-profile degeneracy and other associated biases. It is now obvious that the total mass-density profiles of most early-type lens galaxies are not isothermal. So assuming an SIE profile during the lens modeling in general introduces a bias in the estimation of the Einstein radius (especially for systems without rings) and hence in the estimation of the total mass. It is also worth decomposing the total mass distribution into a stellar component and a dark component, or even considering DM substructures for a more detailed characterization of the mass structure (e.g., Vegetti et al., 2012; Barnabè et al., 2012; Schneider and Sluse, 2013; Suyu et al., 2013). Multiband photometry or spectra of lens galaxies can be further used to yield a better constraint on the stellar mass based on stellar population synthesis (SPS) models (e.g., Spiniello et al., 2011; Dutton et al., 2013b; Conroy et al., 2013; Brewer et al., 2014; Spiniello et al., 2014).

CHAPTER 5

SUMMARY AND FUTURE WORK

5.1 Summary

In this dissertation, I investigated the properties and evolutions of the "mass sector" in our universe using a hierarchical Bayesian method. The theme of this dissertation is two-fold. Firstly, I studied the evolution of the luminous mass-component, i.e., massive early-type galaxies which are the most-luminous, end-product of hierarchical galaxy-formation processes. The other half of the dissertation focused on the characterization of the dark mass-component known as the mysterious dark matter.

I used a dynamical probe to unravel the evolution trend of ETGs. The luminosity-weighted line-of-sight stellar velocity dispersion σ_*, which quantifies the standard deviation of the velocities of stars within a galaxy, is a direct probe of the gravitational potential/dynamical mass and is strongly correlated with the size and surface brightness of the galaxy through the FP. The enormously large LRG sample observed by the BOSS Survey provided a perfect playground for galaxy evolution studies. I obtained the stellar velocity dispersions from the observed spectra using a velocity-dispersion extraction code specially customized for this low SNR data. Then instead of using the point-estimations of velocity dispersion for individual galaxies, I performed a hierarchical Bayesian analysis to explore the properties of the entire galaxy population. Binning the sample in the redshift-luminosity space and parameterizing the conditional PDF of $\log_{10}\sigma_*$ as a Gaussian distribution with mean m and intrinsic scatter s, I detected a strong evolution trend in the intrinsic scatter s for the CMASS sample ($0.4 < z < 0.8$) in the direction that LRGs at higher z have larger s values (Figure 2.10), which indicates that they are more dynamically diverse. Also, the mean m of the CMASS sample correlates linearly with galaxy magnitude with a varying slope, which is interpreted as a mass-dependent star-formation history. On the other hand, the LOWZ sample does not show much variation in either the intrinsic scatter or the mean (Figure 2.8). Further investigation is needed to quantify the selection effect.

To study the dark component, I relied on both the strong gravitational lensing technique and single stellar population models. I discovered 73 new gravitational lens systems in total

from the SLACS and S4TM Surveys with 40 S4TM grade-A lenses, 18 S4TM grade-C lenses, and 15 SLACS grade-C lenses. A customized lens modeling code was used to determine the upper limits of the Einstein radii for the 33 grade-C lenses. The relation between the observed stellar velocity dispersion σ_* and the lensing velocity dispersion $\sigma_{\rm SIE}$ was studied via a hierarchical Bayesian method and used to estimate the logarithmic total mass density slope γ' empirically. We detected a significant (4σ) anticorrelation between γ' and σ_* (Figure 3.8), which can be explained by baryonic physics such as adiabatic contraction, gravitational heating, and energetic feedbacks. I performed core-Sersic fits to the observed *HST* I-band surface brightnesses of the lens galaxies, based on which the stellar masses were estimated according to SSP models with two commonly used IMFs — the Salpeter and the Chabrier IMFs adopted. The dark-matter fraction f_{dm} within one half of the half-light radius was found to be correlated with the galaxy mass, stellar velocity dispersion, and half-light radius (Figure 3.10). That was interpreted as more massive ETGs possess larger fractions of DM in their central regions as a result of the competition among different cooling and heating mechanisms. Another interesting finding was that the Salpeter IMF is substantially disfavored by our lens sample especially at the low-mass end, as it predicted more stellar masses than the total masses constrained by gravitational lensing.

5.2 Future Work

According to the ΛCDM paradigm, the universe is composed by a cosmological constant term (dark energy), a *dark* component (DM), and a *luminous* component (stars and the gas and dust in the interstellar and intergalactic medium). Understanding the nature of these fundamental building blocks and the way they interact to form the "visible" universe is crucial to decode both the physics of galaxy formation and the nature of the dark matter that galaxies assemble on. Here I will discuss two future projects in detail that provide vital insight on the darkness and light in the universe using the powerful tool of strong gravitational lensing (GL). On the dark side, DM substructures will be uncovered by a sophisticated lens modeling technique and permit the exploration of the nature of DM particles. On the luminous side, the stellar initial mass function (IMF) will be thoroughly studied via a joint analysis of GL, stellar dynamics (SD), and stellar population synthesis (SPS) models. The probed universality/variation of the IMF reveals the controlling factors in galaxy formation and evolution processes.

5.2.1 For The Darkness — A Grand Lens Model

Despite the remarkable success of ΛCDM paradigm in describing the large-scale structure of the Universe and in explaining numerous observational results, the nature of the DM itself, which accounts for more than 80% of the mass in the universe, is still an unsolved fundamental question in astrophysics and physics. One notable issue that remains challenging for over a decade is the missing satellite problem (MSP), which is the significant excess in the number of DM subhalos predicted by numerical cosmological simulations compared to that of the observed satellite galaxies in the Local Group (Klypin et al., 1999; Moore et al., 1999a). The lack of observed satellite galaxies indicates either a suppression in the galaxy formation process in subhalos, or more importantly, an insufficient understanding of DM particles. For instance, the dynamic friction and tidal stripping owing to the central galaxy will enhance the subhalo disruption. Also if the dark matter particles were warm (WDM) instead of cold, the free-streaming due to thermal motion of particles would smear out small-scale structure while persisting the overall success on large scales (Bode et al., 2001). Consequently, properties like the abundance and mass function would be different from ΛCDM predictions.

However, the detection of DM substructure has been substantially limited by its dark feature as most astronomical detections heavily rely on the object's electromagnetic radiation. Luckily, there is the gravitational lensing effect, which is the convergence of the background light under the influence of the *gravity* of the foreground massive object (Einstein, 1936). As a pure gravitational phenomenon, GL is the perfect tool to study substructures within DM halos. The current schemes of detecting substructures through GL rely on independently modeling observational anomalies, including the flux-ratio anomaly, time-delay anomaly, and astrometric anomaly.

Figure 5.1 provides an illustration on how these schemes work using a mock lens system. The lens and the source are put at the average redshifts of lens systems discovered in the SLACS for The Masses (S4TM) Survey (She et al. 2014 to be submitted). In this illustration, we adopt a substructure with mass 1E−4 of the lens mass. In building a realistic simulation, we use the sky noise, point spread function (PSF), and image configuration of a real S4TM lens. Adding the substructure significantly affects the positions and brightnesses of the arcs (top) and results in up to 10% change in the flux ratio (bottom left) and a few days difference in the arrival time which can not be explained simultaneously otherwise.

Although many valuable results have been discovered by the above techniques (Mao and Schneider, 1998; Moustakas and Metcalf, 2003; Keeton et al., 2003; Kochanek and

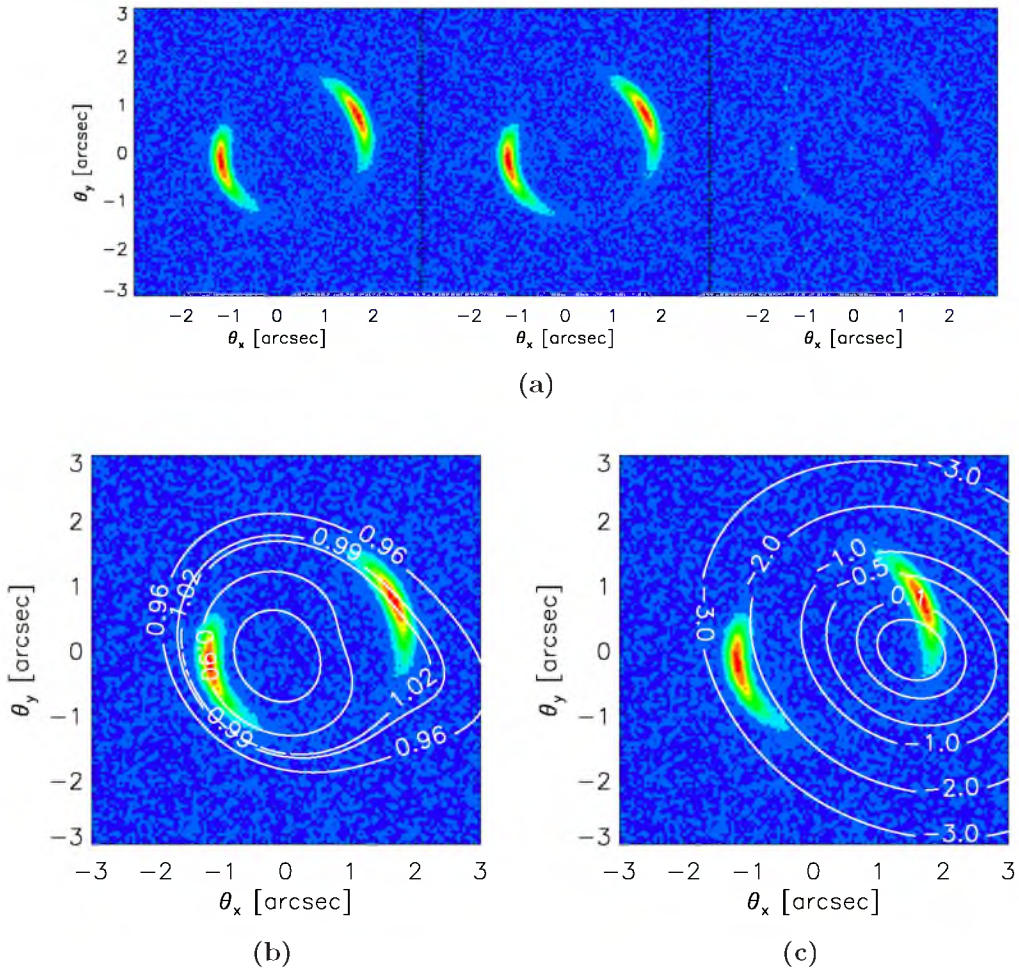

Figure 5.1: Illustration of the effects of a DM substructure on various observables. Panel (a): the lensed images without and with the consideration of the substructure and their difference; Panel (b): the unperturbed lensed images overlaid by the contours of the ratio of magnifications normalized by the value at the brightest peak; Panel (c): the same image overlaid by the contours of the differences (in days) in the arrival time with respect to the value at the brightest peak.

Dalal, 2004; Koopmans, 2005; Keeton and Moustakas, 2009; Vegetti et al., 2012), each has limitations if applied individually. The flux-ratio anomaly is only sensitive to perturbations from substructures on scales comparable to that of the source or larger. The time-delay measurements usually require instruments with supreme spatial resolution and time resolution, which can only be achieved by space-based satellites such as the proposed *OMEGA* mission (Moustakas et al., 2008). More importantly, as discussed by Keeton (2009), these different observables might have different dependence on the mass function and spatial distribution of substructures. As for the SK, although there are studies combining it with GL to put constraints on the total mass density profile and its evolution across cosmic time (Koopmans and Treu, 2003b; Barnabè et al., 2009b; Bolton et al., 2012c), substructures have not been included at all.

The Grand Lens Model that I propose will simultaneously fit all the available observables. However, it is not simply integrating all the existing methods into a whole. Instead, it is a specially designed, *fully pixelized* lens model. One extraordinary advantage of a fully pixelized model is that we are no longer restricted by the few analytically solvable profiles which might not be appropriate for describing the substructures. Also, it has the potential to even probe substructures in the sources. Furthermore, this model will work with the two-dimensional mass distribution rather than the potential so that the interpretation of the substructure distribution is extremely straightforward. The lensing potential can be reconstructed via the direct integration method (Saha and Williams, 1997) or fast Fourier transform (FFT) method (Wayth and Webster, 2006) as needed, from which the surface brightness, flux-ratio, and time-delay are derived. The SK information is predicted by the Jeans' equation to help breaking the mass-sheet degeneracy. The goodness of the fit is determined according to a penalty function and the significance of various models is characterized by the likelihood function in the framework of Bayesian statistics. A Markov-Chain Monte Carlo (MCMC) computation can be used to estimate relevant uncertainties.

To date, there are nearly 400 galaxy-galaxy and 100+ galaxy-QSO lensing events uncovered by various discovery programs. However, only a small fraction of galaxy-QSO systems have time-delay measurements and the resolution is typically insufficient. So a revisit of the known lens systems is quite necessary and will surely lead to a vast number of discoveries of substructure. The proposed *the Observatory for Multi-Epoch Gravitational-Lens Astrophysics (OMEGA)* mission is the ideal tool for this purpose as it can track the variability of gravitationally-lensed quasars in six broad-band filters covering the wavelength range from near-ultraviolet (NUV) to the near-infrared (NIR) in a very fast

manner (hours to days) and obtain detailed 2-dimensional spectroscopy in a relatively slower manner (days to weeks). This will provide unprecedentedly high-resolution measurements in the time-delay ($\sim 0.5\%$) and relative fluxes (subpercent level), which are crucial in probing the extremely weak signals caused by DM substructures as shown in Figure 5.1. The detailed SK information can also be extracted from the ultra-deep, 2-dimensional spectroscopy of the lensing galaxy and get compared with the dynamical prediction to further constrain the mass profile. Surveys enabled by the next-generation telescopes such as the James Webb Space Telescope and the Large Synoptic Survey Telescope will discover thousands of new gravitational lenses for future follow-ups and extend the study to a much wider range in galaxy redshift and mass.

Applying the Grand Lens Model to the hundreds and potentially thousands of lens systems, I expect to detect substructures in most of the lens systems with their positions, masses, and mass-density profiles well-determined. Besides the case-by-case study where the signals are typically noisy and uncertain, I will apply a hierarchical Bayesian method (Shu et al., 2012) to deconvolve observational uncertainties and explore the intrinsic behaviors of the abundance, spatial distribution, and mass function of substructure using the entire lens population. Further comparisons with cosmological simulation predictions will revolutionize our understanding of DM particles, including the collisional nature, thermal nature, and particle nature (e.g., Natarajan et al., 2002; Dalal and Kochanek, 2002; Kaplinghat, 2005).

5.2.2 Project II. For The Light — Starlight + Lensing

The other project that I propose is to further the study of the stellar initial mass function (IMF) using a combination of GL, stellar dynamics (SD) and stellar population synthesis (SPS) to test its universality/variation with the help of the comprehensive sample of gravitational lenses.

Apparently, the spectral energy distribution (SED), colors, and many other luminous observables of the stellar population and thus galaxies are strongly related to the stellar IMF, which is an empirical relation quantifying the relative fraction of stars as a function of the stellar mass at the time when the whole population formed. Having knowledge of the stellar IMF is also important for understanding galaxy evolution. Salpeter (1955) first quantified the IMF of main-sequence stars in the solar neighborhood as a power-law function with an exponent of -2.35. To date, the most commonly used forms of IMF are the Kroupa IMF (Kroupa, 2001) and the Chabrier IMF (Chabrier, 2003), two variations from the Salpeter IMF (Salpeter, 1955) at the low-mass end. However, a vigorous discussions has been going on for over 50 years as to whether or not the stellar IMF is universal and

what kinds of variables control the variation (Trager et al., 2000; Gilmore, 2001; Davé, 2008; Bastian et al., 2010).

The classical method of inferring the IMF is from the present-day star counts assuming certain stellar evolution history. Obviously, that can only be done in very nearby galaxies. For distant galaxies in which individual stars can not be resolved, two independent modeling techniques have been widely used recently. The first one relies on the SPS models by varying IMF, age, metallicity, and other parameters to explain the observed spectra or broad-band colors. The other technique is a dynamical probe by fitting the projected two-dimensional SD predictions to the observed SK measurements to derive the mass-to-light ratio for comparison. Although these two techniques have recently revealed some strong pieces of evidence that the stellar IMF is not universal, but rather depends on the properties and formation history of the galaxy itself (Cappellari et al., 2012; Conroy et al., 2013), potential systematic uncertainties still exist. The SPS technique highly relies on the surface-gravity-sensitive spectral features NaI, CaII, and FeH at 0.82μm, 0.86μm, and 0.99μm to break the degeneracy between low-mass stars and dark matter. Improper interpretations in the SPS model, or lack of coverage of these features in the observed spectra, can lead to up to 50% change in the mass-to-light ratio normalization (Conroy and van Dokkum, 2012). The dynamical probe requires assumptions on the DM halo distribution which is very model-dependent. Even by including GL constraint, a degeneracy between the IMF and DM density profile still prohibits unambiguous disentanglement of the luminous and dark components (Treu et al., 2010). Furthermore, those studies mainly focus on nearby, very massive early-type galaxies and substructures have not been taken into account at all.

In this project, I will extend the study to galaxies with a wider range of redshift, mass, and morphology. Additionally, I will perform two analyses using SD+GL and SPS to independently estimate the stellar masses, the combination of which will largely break existing degeneracies and yield tighter constraints. The Grand Lens Model will separate the mass distribution into the dark smooth component, dark substructure, and luminous component and utilize both the lensing signals and detailed two-dimensional SK measurements to break the mass-anisotropy, mass-sheet degeneracies, and determine the mass composition as described in the first project. On the other hand, multiband photometry or ideally spectroscopy covering a wide wavelength range from NIR to NUV for lens galaxies will permit the determinations of stellar mass for a range of IMFs through the SPS analysis (Conroy et al., 2009). In this process, carefully disentangling the light from the background sources is required to minimize the possible color bias which affects the stellar mass estimations.

Age, metallicity, dust, and other parameters in the SPS analysis should be marginalized over. Stellar masses obtained by the two methods can get compared and this joint analysis will determine the stellar IMF on a galaxy-to-galaxy base.

Figure 3.10 shows the discovery of a clear correlation between the DM fraction f_{dm} and the total enclosed mass M_{Ein} determined by GL within the Einstein radius for over 100 confirmed galaxy-galaxy lenses in the Sloan Lens ACS (SLACS) Survey and the S4TM Survey (Shu et al. 2013 to be submitted). Stellar masses are calculated by the SPS technique based on the background-corrected magnitudes assuming different IMFs. Given the single HST band photometry, we have to make assumptions on the age, metallicity, and dust. The symbols on the figure correspond to a typical model with a formation time of 4 Gyrs, solar metallicity, and no dust and error bars indicate the extreme values of f_{dm} by varying the age, metallicity, and dust. Although the systematic uncertainties are huge, the hard limits on the total mass provided by GL ($f_{dm} \geq 0.0$) clearly favors a Kroupa IMF for this set of galaxies. Deep, multiband photometry will allow a full exploration of the parameter space and permit more accurate estimations on the stellar mass.

On the observational side, as mentioned before, only a small fraction of the known lens systems have high signal-to-noise ratio (SNR) spectroscopy and detailed, robust two-dimensional SK measurements outer to large enough radii. The *OMEGA* mission will substantially bridge this gap in the next 3 years. Meanwhile, as a pioneer project, I propose to do spectroscopic follow-ups for a sample of lens systems (~ 100) selected from the Sloan Lens ACS Survey, S4TM Survey and BOSS Emission Line Lens Survey to cover a wide range in the lens redshift ($0.1 < z < 0.7$) and total enclosed mass ($10^{10} M_{\odot} < M_{\rm Ein} < 10^{12} M_{\odot}$). The spectroscopic and SK information can be collected by instruments with long-slit spectrograph or Integral Field Spectrograph (IFS), the Keck telescopes for example. The joint analysis on this well-chosen sample will test the detected correlation between the stellar IMF and galaxy mass and unravel any variation of the IMF across cosmic time. Finally convolving with the ultra-deep, comprehensive lens sample available in the near future, I expect to find unambiguous evidence for the correlation/variability of the stellar IMF using the hierarchical Bayesian method.

APPENDIX

BAYESIAN INFERENCE AND EXTRA FIGURES

Here we derive the procedure to calculate the likelihood function of the hyperparameters $\mathscr{L}(\vec{\theta}|\{\vec{y}\},\{\vec{x}\},\mathscr{H})$. Using the fact that the product of two Gaussian distributions is a scaled Gaussian distribution as

$$\mathcal{N}(\mu_1,\sigma_1^2)\mathcal{N}(\mu_2,\sigma_2^2)=\frac{\exp[-\frac{(\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}]}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\mathcal{N}(\frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_1^2+\sigma_2^2},\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}) \tag{A.1}$$

We obtain

$$\mathscr{L}_A\propto\prod_{i=0}^{N_A}\int\frac{1}{\delta_i\delta}\exp\{-\frac{\delta_i^2+\delta^2}{2\delta_i^2\delta^2}[y-\frac{\delta_i^2(ax+b)+\delta^2y_i}{\delta_i^2+\delta^2}]^2\}\,\mathrm{d}y$$
$$\int\delta(x-x_i)\exp[-\frac{(ax+b-y_i)^2}{2(\delta_i^2+\delta^2)}]\frac{1}{s_{\rm best}}\exp[-\frac{(x-m_{\rm best})^2}{2s_{\rm best}^2}]\,\mathrm{d}x \tag{A.2}$$

and

$$\mathscr{L}_C\propto\prod_{j=0}^{N_C}\int\frac{1}{\delta_j\delta}\exp\{-\frac{\delta_j^2+\delta^2}{2\delta_j^2\delta^2}[y-\frac{\delta_j^2(ax+b)+\delta^2y_j}{\delta_j^2+\delta^2}]^2\}\,\mathrm{d}y$$
$$\int[1-H(x-x_j)]\exp[-\frac{(ax+b-y_j)^2}{2(\delta_j^2+\delta^2)}]\frac{1}{s_{\rm best}}\exp[-\frac{(x-m_{\rm best})^2}{2s_{\rm best}^2}]\,\mathrm{d}x \tag{A.3}$$

Going through some math, we obtain

$$\mathscr{L}_A\propto\prod_{i=0}^{N_A}\frac{\exp[-\frac{(am_{\rm best}+b-y_i)^2}{2(s_{\rm best}^2a^2+\delta_i^2+\delta^2)}]}{s_{\rm best}\sqrt{\delta_i^2+\delta^2}}\times\exp\{-\frac{[s_{\rm best}^2a(ax_i+b-y_i)+(x_i-m_{\rm best})(\delta_i^2+\delta^2)]^2}{2s_{\rm best}^2(s_{\rm best}^2a^2+\delta^2+\delta_i^2)(\delta^2+\delta_i^2)}\} \tag{A.4}$$

$$\mathscr{L}_C\propto\prod_{j=0}^{N_C}\frac{\exp[-\frac{(am_{\rm best}+b-y_j)^2}{2(s_{\rm best}^2a^2+\delta^2+\delta_j^2)}]}{\sqrt{s_{\rm best}^2a^2+\delta^2+\delta_j^2}}\times\Phi(\frac{s_{\rm best}^2a(ax_j+b-y_j)+(x_j-m_{\rm best})(\delta_j^2+\delta^2)}{s_{\rm best}\sqrt{s_{\rm best}^2a^2+\delta^2+\delta_j^2}\sqrt{\delta^2+\delta_j^2}}) \tag{A.5}$$

Figures below are the reduced chi-square curves as a function of the Einstein radius and the lens models for all the 33 (18+15) grade-C lenses from S4TM and SLACS surveys.

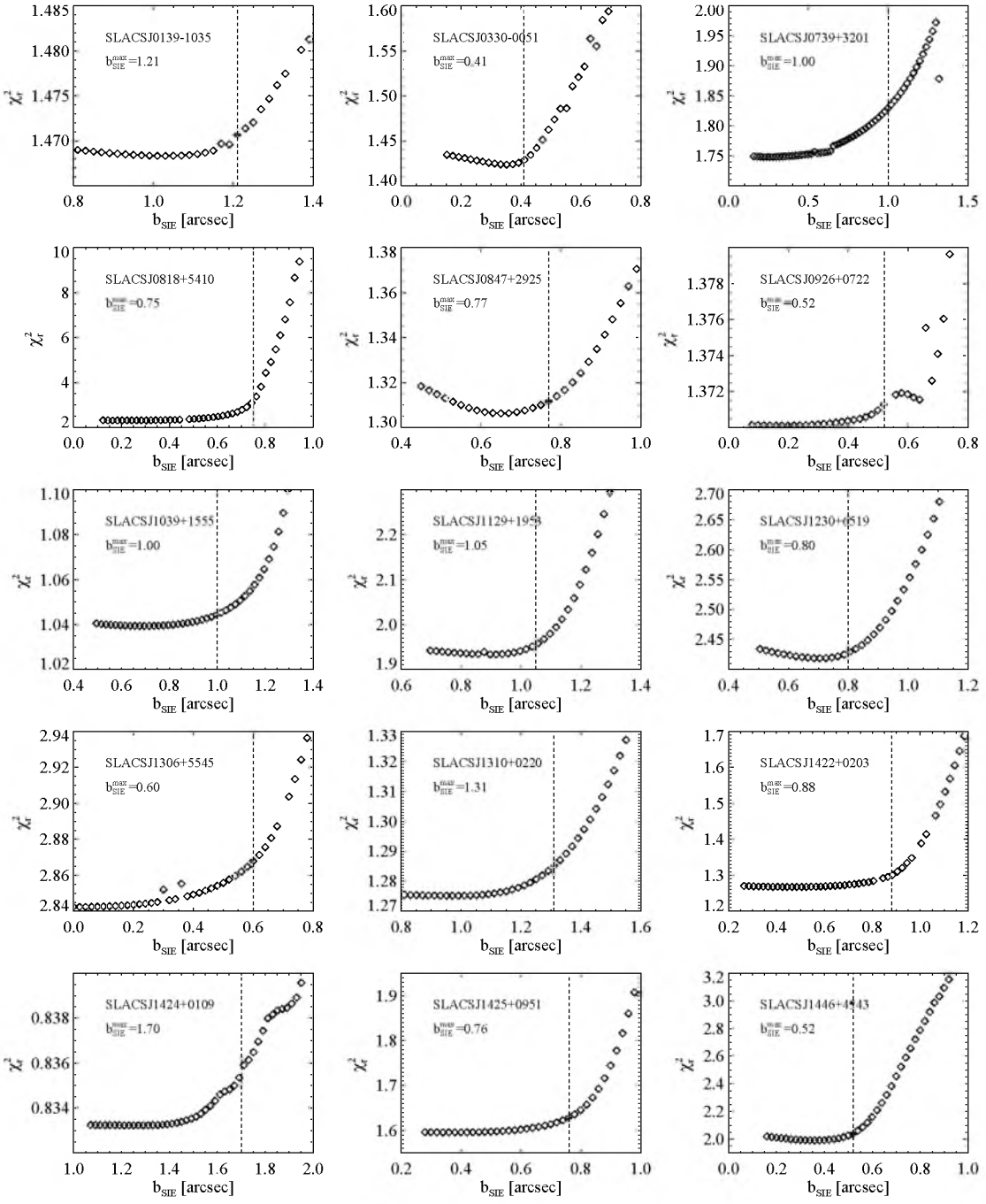

Figure A.1: The reduced chi-square χ_r^2 as functions of trial Einstein radius $b_{\rm SIE}$ for all the 18 grade-C lenses with no counter-images in the S4TM survey. The vertical dashed lines indicate the locations of the upper limits of Einstein radii. (Please refer to the context for details about how to determine the upper limits.)

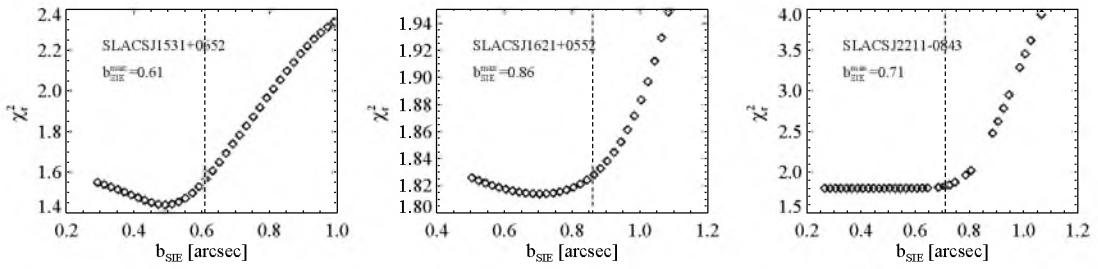

Figure A.1: *Continued*

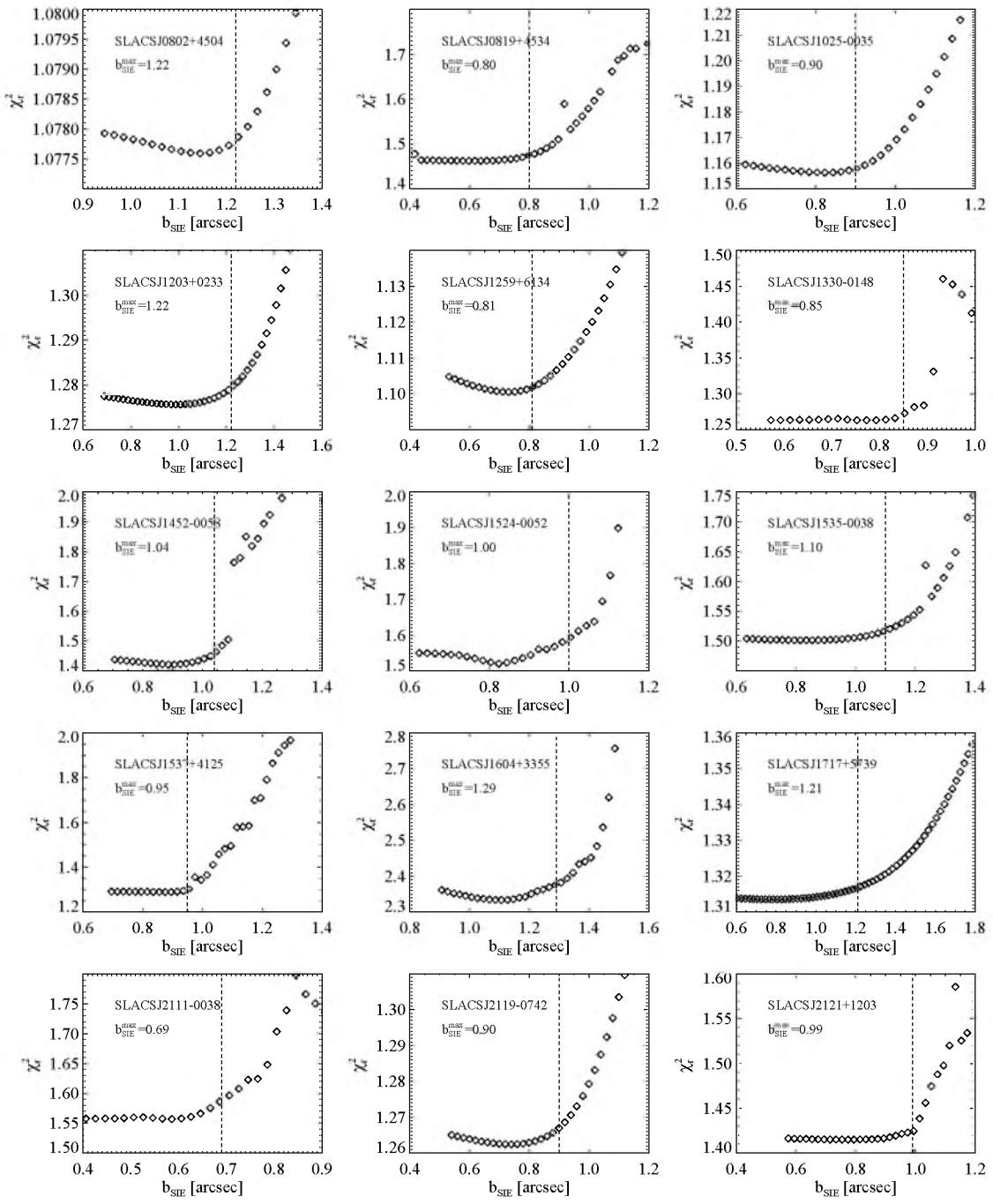

Figure A.2: The reduced chi-square χ_r^2 as functions of trial Einstein radius $b_{\rm SIE}$ for all the 15 grade-C lenses with no counter-images in the SLACS survey. The vertical dashed lines indicate the locations of the upper limits of Einstein radii. (Please refer to the context for details about how to determine the upper limits.)

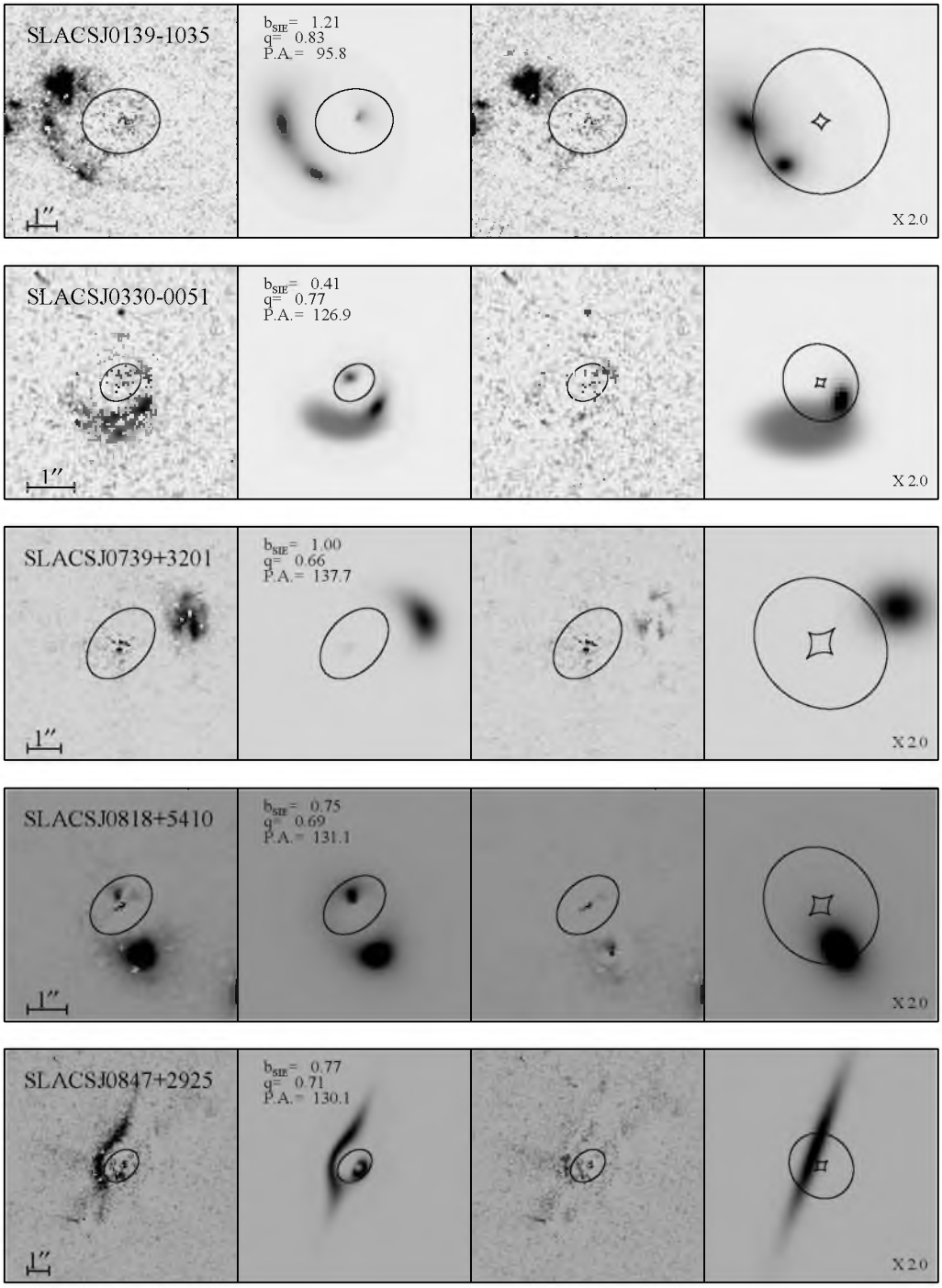

Figure A.3: Lens models for 18 newly discovered grade-C gravitational lenses from S4TM survey. On each row, from left to right, the b-spline subtracted data image, the predicted SIE lens model image, the residual, and the source distribution, respectively, overlaid by critical lines (lens plane) and caustics (source plane). Remember for these grade-C lenses, we adopt upper limits of Einstein radii to the lens models as well as fix P.A. and $q_{\rm SIE}$. Other useful information such as the target ID, fitting parameters, and χ^2 values are also present in different panels. All images are oriented such that north is up and east to the left.

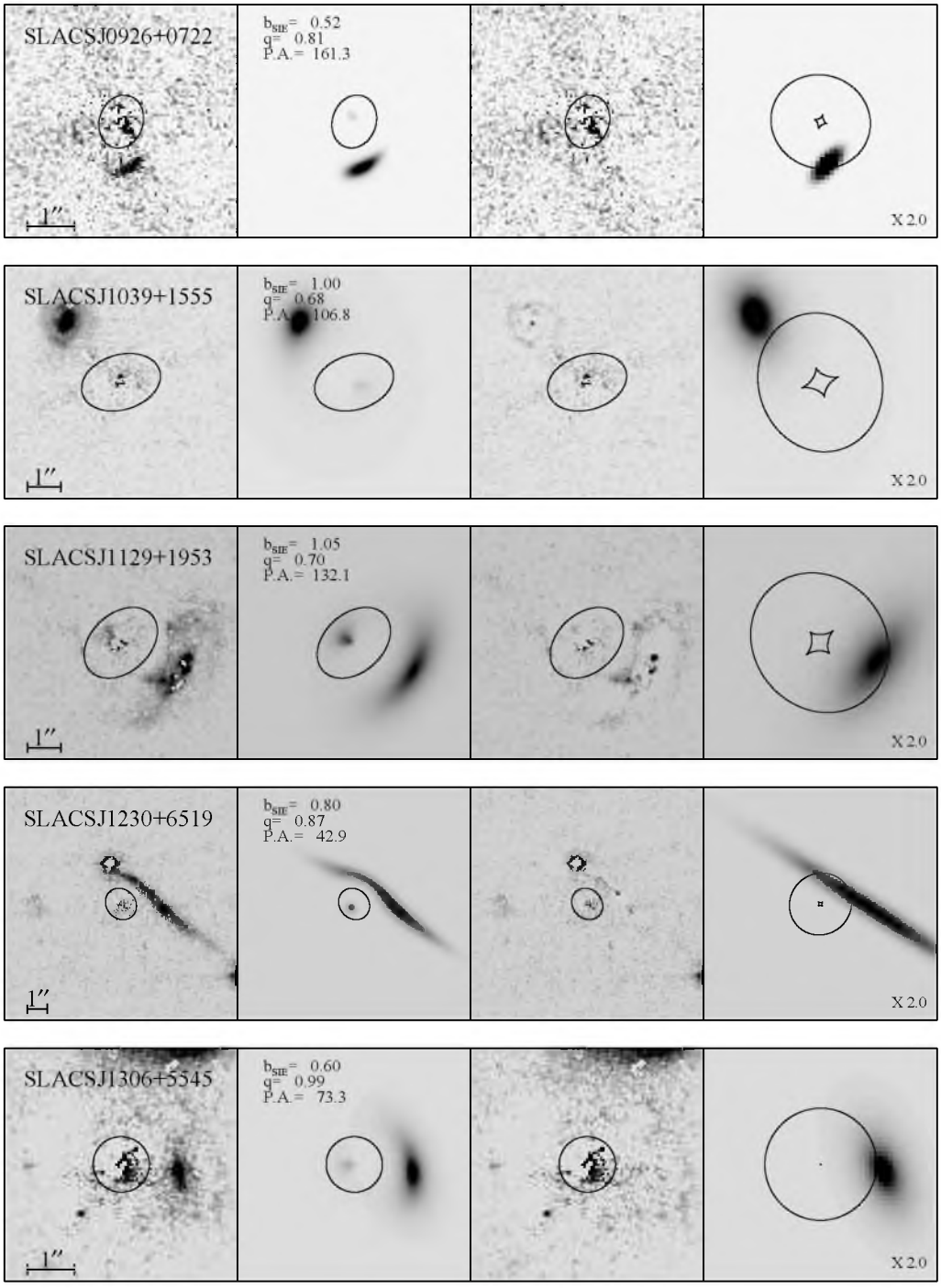

Figure A.3: *Continued*

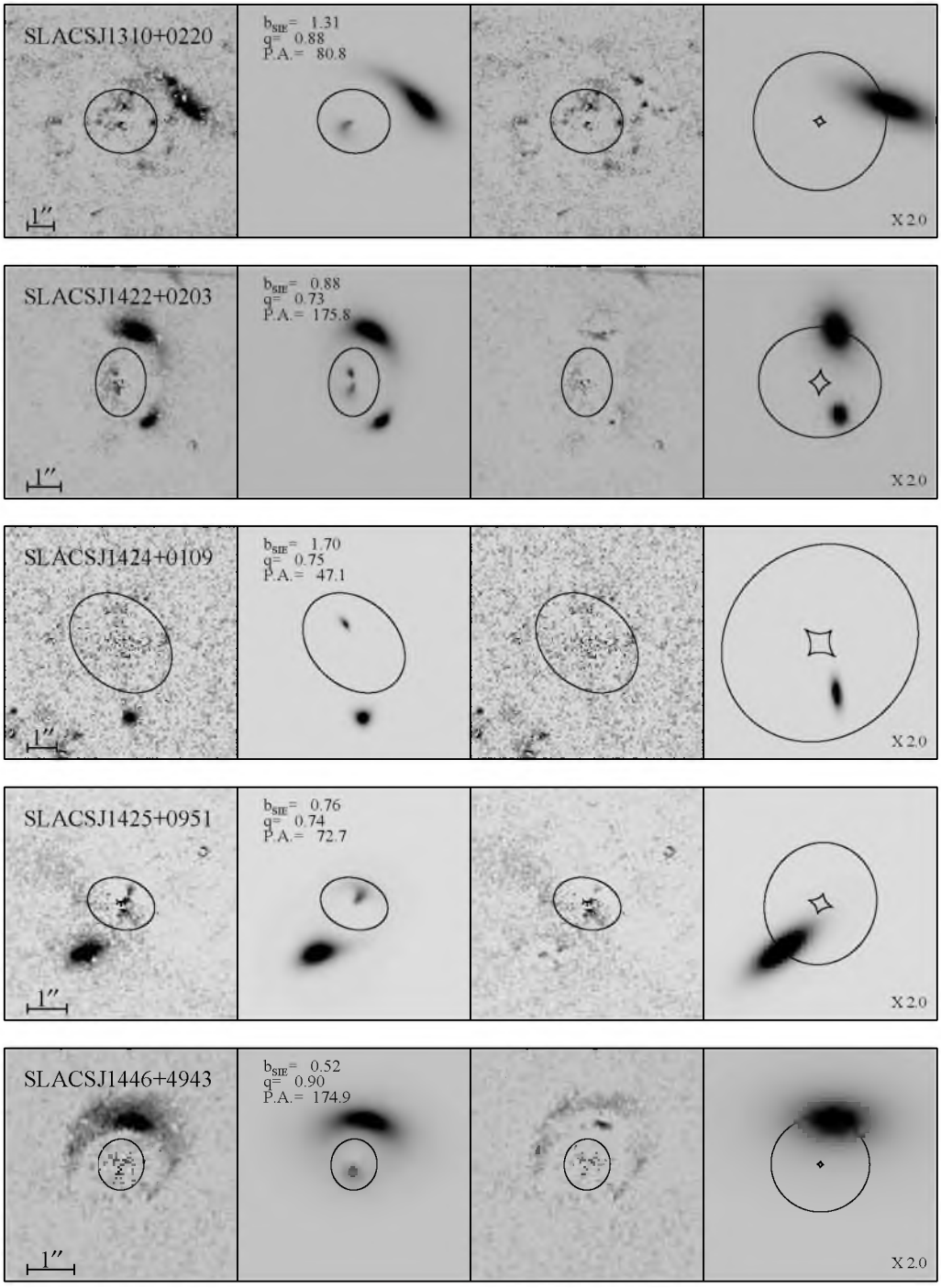

Figure A.3: *Continued*

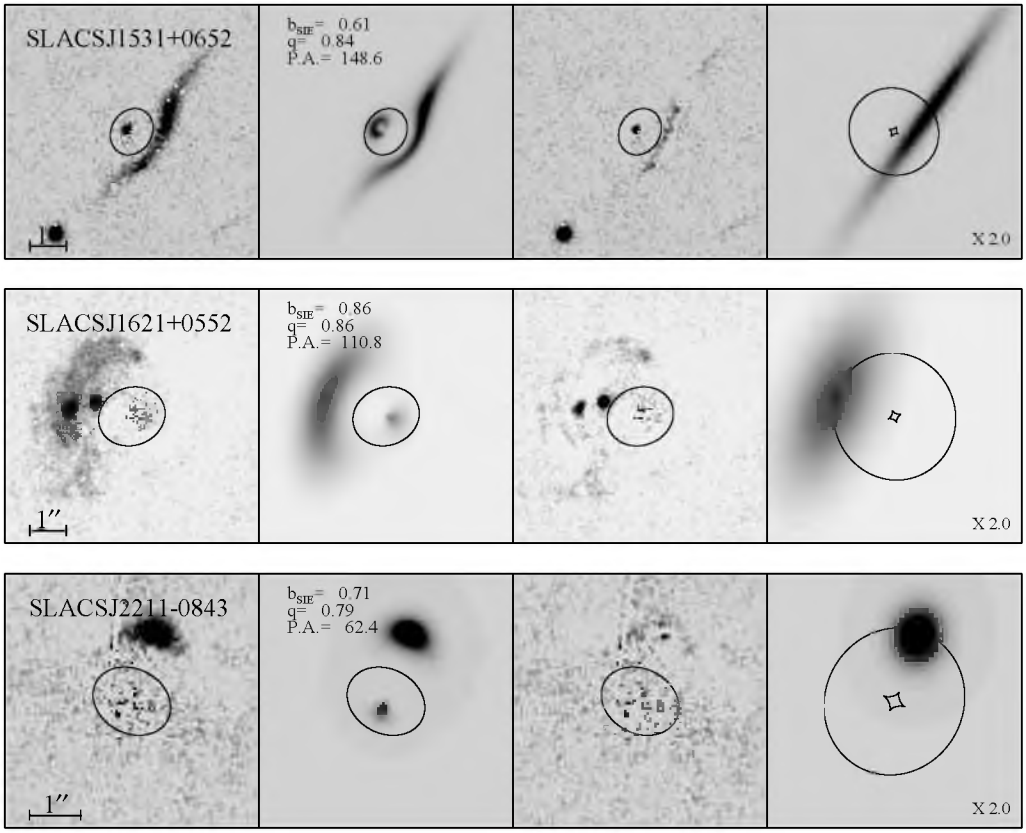

Figure A.3: *Continued*

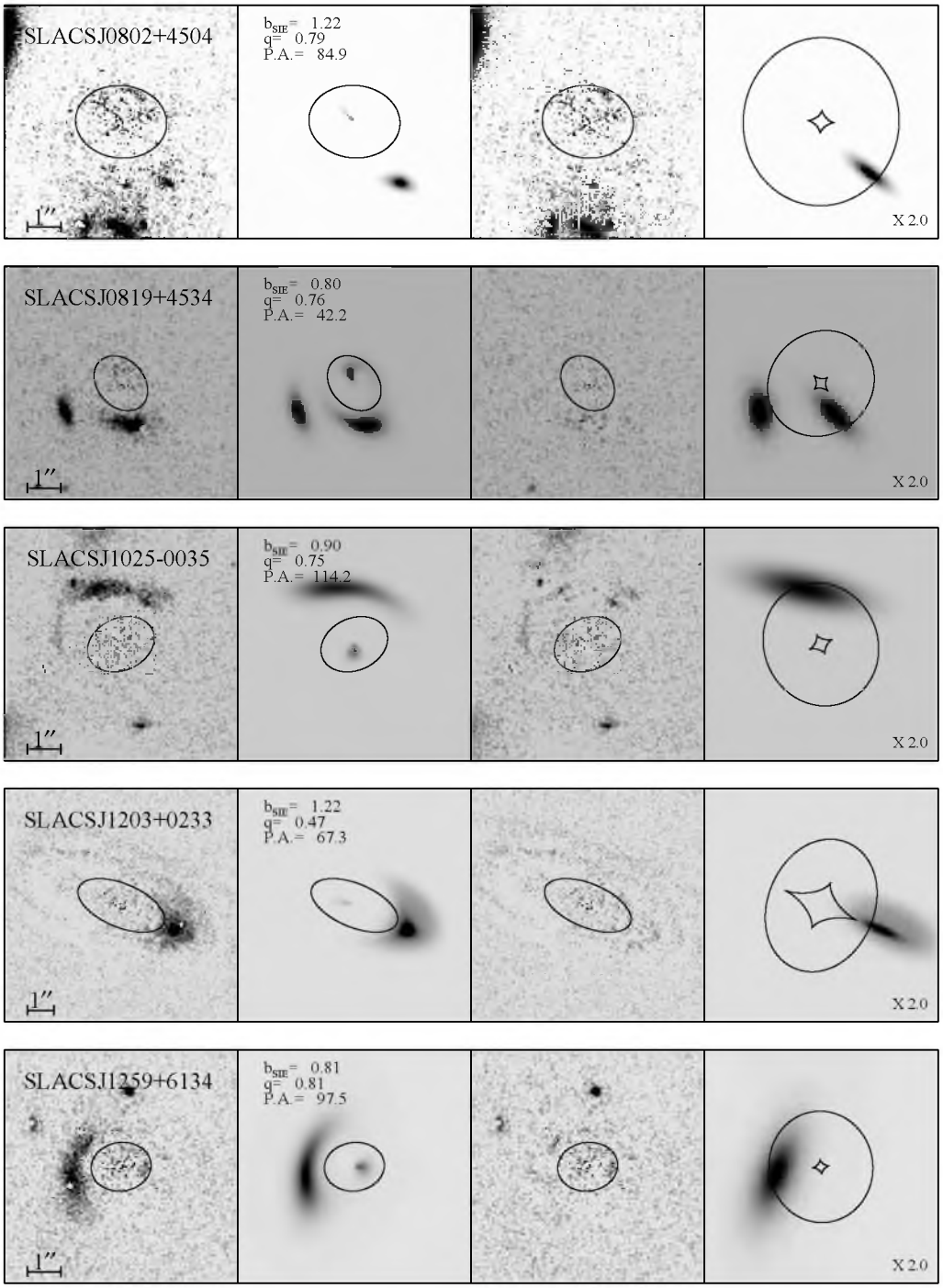

Figure A.4: Lens models for 15 newly discovered grade-C gravitational lenses from the SLACS survey. On each row, from left to right, it displays the b-spline subtracted data image, the predicted SIE lens model image, the residual, and the source distribution, respectively, overlaid by critical lines (lens plane) and caustics (source plane). Remember for these grade-C lenses, we adopt upper limits of Einstein radii to the lens models as well as fix P.A. and $q_{\rm SIE}$. Other useful information such as the target ID, fitting parameters, and χ^2 values are also present in different panels. All images are oriented such that north is up and east to the left.

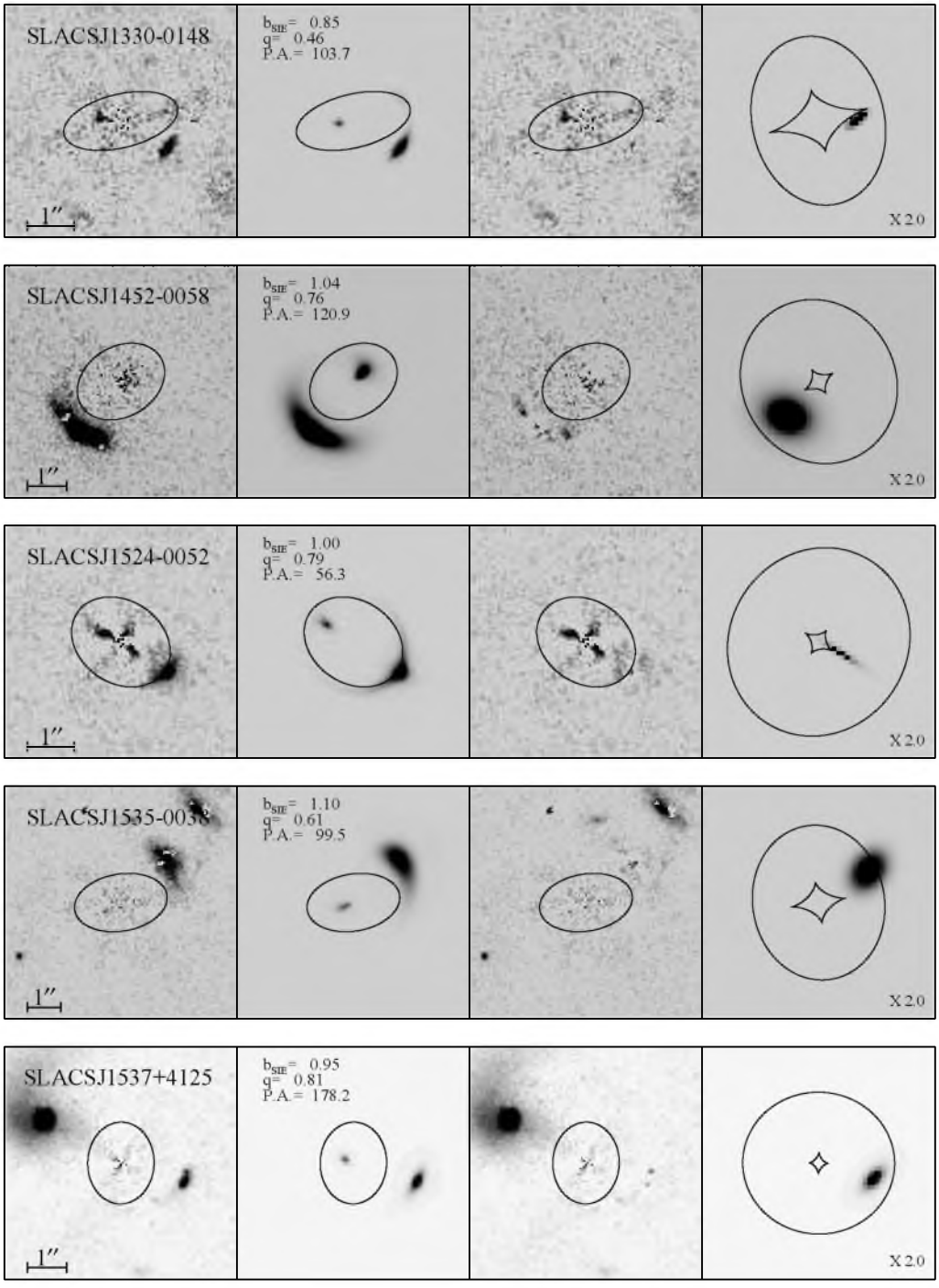

Figure A.4: *Continued*

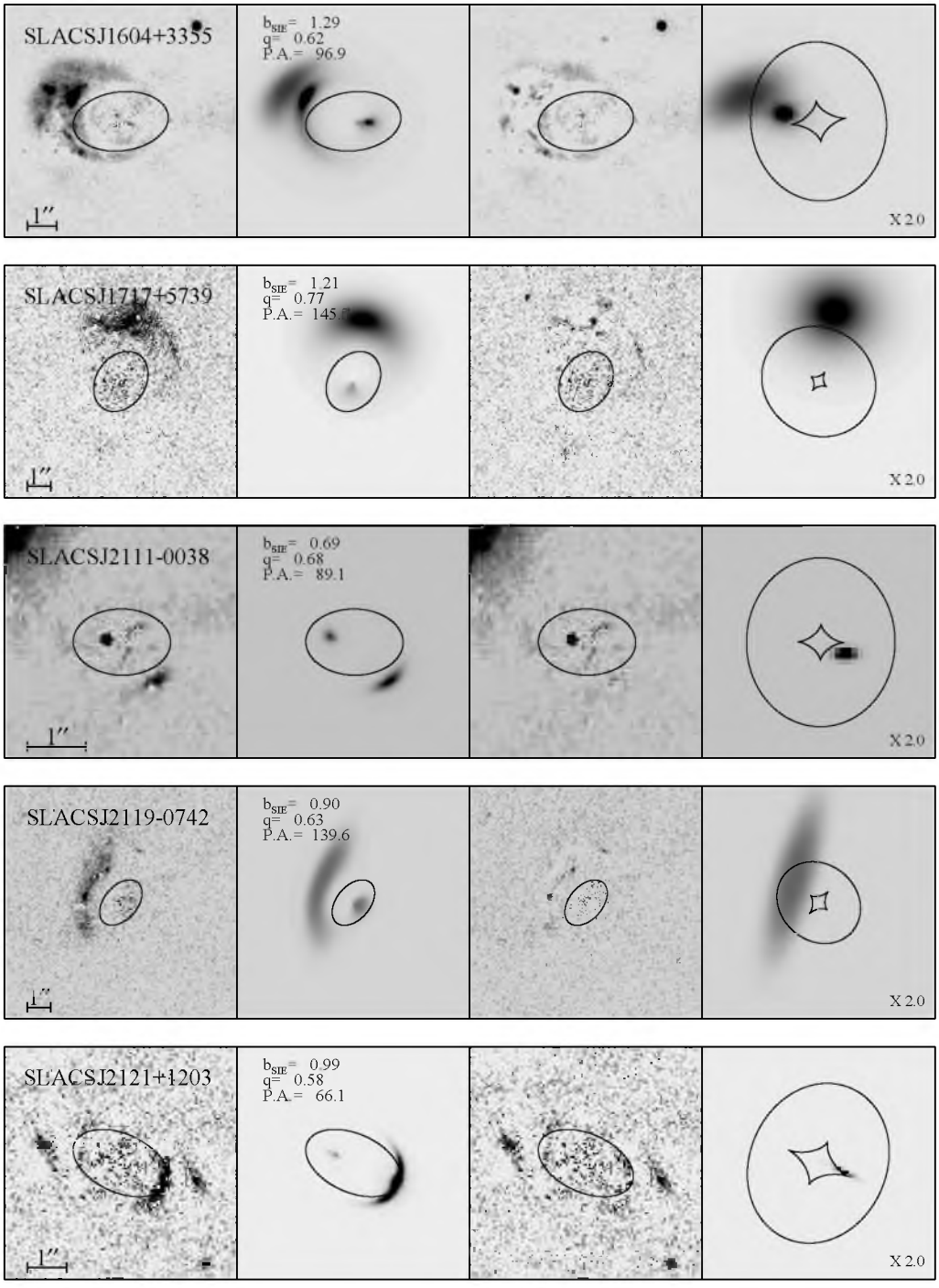

Figure A.4: *Continued*

REFERENCES

Abadi, M. G., Navarro, J. F., Fardal, M., Babul, A., and Steinmetz, M. 2010, MNRAS, 407, 435

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

Agnello, A., Auger, M. W., and Evans, N. W. 2013, MNRAS, 429, L35

Aihara, H., et al. 2011, ApJS, 193, 29

Alpher, R. A., and Herman, R. C. 1948, Physical Review, 74, 1737

Anderson, L., et al. 2012, MNRAS, 427, 3435

Auger, M. W., Treu, T., Bolton, A. S., Gavazzi, R., Koopmans, L. V. E., Marshall, P. J., Bundy, K., and Moustakas, L. A. 2009, ApJ, 705, 1099

Auger, M. W., Treu, T., Bolton, A. S., Gavazzi, R., Koopmans, L. V. E., Marshall, P. J., Moustakas, L. A., and Burles, S. 2010, ApJ, 724, 511

Barnabè, M., Czoske, O., Koopmans, L. V. E., Treu, T., and Bolton, A. S. 2011, MNRAS, 415, 2215

Barnabè, M., Czoske, O., Koopmans, L. V. E., Treu, T., Bolton, A. S., and Gavazzi, R. 2009a, MNRAS, 399, 21

—. 2009b, MNRAS, 399, 21

Barnabè, M., et al. 2012, MNRAS, 423, 1073

Bastian, N., Covey, K. R., and Meyer, M. R. 2010, ARA&A, 48, 339

Baugh, C. M., Cole, S., and Frenk, C. S. 1996, MNRAS, 283, 1361

Behroozi, P. S., Conroy, C., and Wechsler, R. H. 2010, ApJ, 717, 379

Bernardi, M., et al. 2003a, AJ, 125, 1817

—. 2003b, AJ, 125, 1849

—. 2003c, AJ, 125, 1866

Bertola, F., and Capaccioli, M. 1975, ApJ, 200, 439

Bezanson, R., et al. 2011, ApJL, 737, L31

Binney, J. 1978, MNRAS, 183, 501

Binney, J., and Tremaine, S. 1987, Galactic dynamics

Blandford, R. D., and Narayan, R. 1992, ARA&A, 30, 311

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

Blumenthal, G. R., Faber, S. M., Flores, R., and Primack, J. R. 1986, ApJ, 301, 27

Blumenthal, G. R., Faber, S. M., Primack, J. R., and Rees, M. J. 1984, Natur, 311, 517

Bode, P., Ostriker, J. P., and Turok, N. 2001, ApJ, 556, 93

Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T., Gavazzi, R., Moustakas, L. A., Wayth, R., and Schlegel, D. J. 2008a, ApJ, 682, 964

Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T., and Moustakas, L. A. 2006, ApJ, 638, 703

Bolton, A. S., Burles, S., Treu, T., Koopmans, L. V. E., and Moustakas, L. A. 2007, ApJL, 665, L105

Bolton, A. S., Treu, T., Koopmans, L. V. E., Gavazzi, R., Moustakas, L. A., Burles, S., Schlegel, D. J., and Wayth, R. 2008b, ApJ, 684, 248

Bolton, A. S., et al. 2008c, ApJ, 682, 964

—. 2008d, ApJ, 684, 248

—. 2012a, AJ, 144, 144

—. 2012b, ApJ, 757, 82

—. 2012c, ApJ, 757, 82

Boylan-Kolchin, M., Ma, C.-P., and Quataert, E. 2005, MNRAS, 362, 184

Brewer, B. J., Marshall, P. J., Auger, M. W., Treu, T., Dutton, A. A., and Barnabè, M. 2014, MNRAS, 437, 1950

Brewer, B. J., et al. 2012, MNRAS, 422, 3574

Brownstein, J. R., et al. 2012, ApJ, 744, 41

Burbidge, E. M., Burbidge, G. R., and Fish, R. A. 1961, ApJ, 133, 393

Cannon, R., et al. 2006, MNRAS, 372, 425

Cappellari, M., et al. 2006a, MNRAS, 366, 1126

—. 2006b, MNRAS, 366, 1126

—. 2009, ApJL, 704, L34

—. 2012, Natur, 484, 485

—. 2013, MNRAS, 432, 1709

Carson, D. P., and Nichol, R. C. 2010, MNRAS, 408, 213

Chabrier, G. 2003, PASP, 115, 763

Charlot, S., and Fall, S. M. 2000, ApJ, 539, 718

Chwolson, O. 1924, Astronomische Nachrichten, 221, 329

Cimatti, A., Daddi, E., and Renzini, A. 2006, A&A, 453, L29

Coe, D., Benítez, N., Broadhurst, T., and Moustakas, L. A. 2010, ApJ, 723, 1678

Cole, S., Lacey, C. G., Baugh, C. M., and Frenk, C. S. 2000, MNRAS, 319, 168

Conroy, C., Dutton, A. A., Graves, G. J., Mendel, J. T., and van Dokkum, P. G. 2013, ApJL, 776, L26

Conroy, C., and Gunn, J. E. 2010, ApJ, 712, 833

Conroy, C., Gunn, J. E., and White, M. 2009, ApJ, 699, 486

Conroy, C., and van Dokkum, P. G. 2012, ApJ, 760, 71

Cool, R. J., et al. 2008, ApJ, 682, 919

Cowie, L. L., Songaila, A., Hu, E. M., and Cohen, J. G. 1996, AJ, 112, 839

Dalal, N., and Kochanek, C. S. 2002, ApJ, 572, 25

Davé, R. 2008, MNRAS, 385, 147

Dawson, K. S., et al. 2013, AJ, 145, 10

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

de Vaucouleurs, G., and Olson, D. W. 1982, ApJ, 256, 346

di Serego Alighieri, S., et al. 2005, A&A, 442, 125

Dicke, R. H., Beringer, R., Kyhl, R. L., and Vane, A. B. 1946, Physical Review, 70, 340

Diemand, J., Zemp, M., Moore, B., Stadel, J., and Carollo, C. M. 2005, MNRAS, 364, 665

Dietrich, J. P., Werner, N., Clowe, D., Finoguenov, A., Kitching, T., Miller, L., and Simionescu, A. 2012, Natur, 487, 202

Djorgovski, S., and Davis, M. 1987, ApJ, 313, 59

D'Onofrio, M., et al. 2008, ApJ, 685, 875

—. 2013, MNRAS, 435, 45

Dressler, A. 1987, ApJ, 317, 1

Dressler, A., Lynden-Bell, D., Burstein, D., Davies, R. L., Faber, S. M., Terlevich, R., and Wegner, G. 1987, ApJ, 313, 42

Driver, S. P., et al. 2011, MNRAS, 413, 971

Duffy, A. R., Schaye, J., Kay, S. T., Dalla Vecchia, C., Battye, R. A., and Booth, C. M. 2010, MNRAS, 405, 2161

Dutton, A. A., Macciò, A. V., Mendel, J. T., and Simard, L. 2013a, MNRAS, 432, 2496

Dutton, A. A., and Treu, T. 2013, ArXiv e-prints

Dutton, A. A., et al. 2011, MNRAS, 417, 1621

—. 2013b, MNRAS, 428, 3183

Einstein, A. 1916, Annalen der Physik, 354, 769

—. 1936, Science, 84, 506

Eisenstein, D. J., et al. 2001a, AJ, 122, 2267

—. 2001b, AJ, 122, 2267

—. 2005a, ApJ, 633, 560

—. 2005b, ApJ, 633, 560

—. 2011a, AJ, 142, 72

—. 2011b, AJ, 142, 72

Faber, S. M., and Jackson, R. E. 1976, ApJ, 204, 668

Fernández Lorenzo, M., Cepa, J., Bongiovanni, A., Pérez García, A. M., Ederoclite, A., Lara-López, M. A., Pović, M., and Sánchez-Portal, M. 2011, A&A, 526, A72

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

Ferreras, I., La Barbera, F., de la Rosa, I. G., Vazdekis, A., de Carvalho, R. R., Falcón-Barroso, J., and Ricciardelli, E. 2013, MNRAS, 429, L15

Ferreras, I., Saha, P., Leier, D., Courbin, F., and Falco, E. E. 2010, MNRAS, 409, L30

Fixsen, D. J. 2009, ApJ, 707, 916

Friedmann, A. 1924, Zeitschrift fur Physik, 21, 326

Frieman, J. A., et al. 2008, AJ, 135, 338

Fritz, A., Ziegler, B. L., Bower, R. G., Smail, I., and Davies, R. L. 2005, MNRAS, 358, 233

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

Gallazzi, A., Charlot, S., Brinchmann, J., and White, S. D. M. 2006, MNRAS, 370, 1106

Gamow, G. 1948, Natur, 162, 680

Gavazzi, R., Treu, T., Koopmans, L. V. E., Bolton, A. S., Moustakas, L. A., Burles, S., and Marshall, P. J. 2008, ApJ, 677, 1046

Gavazzi, R., Treu, T., Marshall, P. J., Brault, F., and Ruff, A. 2012, ApJ, 761, 170

Gavazzi, R., Treu, T., Rhodes, J. D., Koopmans, L. V. E., Bolton, A. S., Burles, S., Massey, R. J., and Moustakas, L. A. 2007, ApJ, 667, 176

Gebhardt, K., et al. 2000, ApJL, 539, L13

—. 2003, ApJ, 597, 239

Geha, M., et al. 2013, ApJ, 771, 29

Gilmore, G. 2001, in Starburst Galaxies: Near and Far, ed. L. Tacconi and D. Lutz, 34

Gnedin, O. Y., Kravtsov, A. V., Klypin, A. A., and Nagai, D. 2004, ApJ, 616, 16

Goldberg, D. M., and Leonard, A. 2007, ApJ, 660, 1003

Gorenstein, M. V., Shapiro, I. I., and Falco, E. E. 1988, ApJ, 327, 693

Gott, III, J. R., Jurić, M., Schlegel, D., Hoyle, F., Vogeley, M., Tegmark, M., Bahcall, N., and Brinkmann, J. 2005, ApJ, 624, 463

Governato, F., et al. 2010, Natur, 463, 203

—. 2012, MNRAS, 422, 1231

Graham, A. W., Erwin, P., Trujillo, I., and Asensio Ramos, A. 2003, AJ, 125, 2951

Graham, A. W., Merritt, D., Moore, B., Diemand, J., and Terzić, B. 2006, AJ, 132, 2701

Grillo, C., and Christensen, L. 2011, MNRAS, 418, 929

Grillo, C., Lombardi, M., and Bertin, G. 2008, A&A, 477, 397

Gunn, J. E., et al. 2006, AJ, 131, 2332

Gustafsson, M., Fairbairn, M., and Sommer-Larsen, J. 2006, PhRvD, 74, 123522

Hinshaw, G., et al. 2013, ApJS, 208, 19

Hubble, E. 1926, Contributions from the Mount Wilson Observatory / Carnegie Institution of Washington, 324, 1

—. 1929, Proceedings of the National Academy of Science, 15, 168

Hubble, E. P. 1936a, Realm of the Nebulae (New Haven: Yale University Press)

—. 1936b, Realm of the Nebulae

Illingworth, G. 1977, ApJL, 218, L43

Jackson, N. 2007, Living Reviews in Relativity, 10, 4

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

—. 1915, MNRAS, 76, 70

Jeffreys, H. 1961, The Theory of Probability (Oxford: Oxford University Press)

Jørgensen, I., Franx, M., and Kjærgaard, P. 1995, MNRAS, 276, 1341

Jorgensen, I., Franx, M., and Kjaergaard, P. 1995, MNRAS, 276, 1341

—. 1996, MNRAS, 280, 167

Kaplinghat, M. 2005, PhRvD, 72, 063510

Kassiola, A., and Kovner, I. 1993, ApJ, 417, 450

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

—. 1993b, MNRAS, 264, 201

Keeton, C. R. 2009, ArXiv e-prints

Keeton, C. R., Gaudi, B. S., and Petters, A. O. 2003, ApJ, 598, 138

Keeton, C. R., and Kochanek, C. S. 1998, ApJ, 495, 157

Keeton, C. R., and Moustakas, L. A. 2009, ApJ, 699, 1720

Kelson, D. D., et al. 1999, ApJ, 514, 614

Klypin, A., Kravtsov, A. V., Valenzuela, O., and Prada, F. 1999, ApJ, 522, 82

Kochanek, C. S. 1994, ApJ, 436, 56

Kochanek, C. S., and Dalal, N. 2004, ApJ, 610, 69

Kochanek, C. S., et al. 2000a, ApJ, 543, 131

—. 2000b, ApJ, 543, 131

Komatsu, E., et al. 2011, ApJS, 192, 18

Koopmans, L. V. E. 2005, MNRAS, 363, 1136

Koopmans, L. V. E., and Treu, T. 2002, ApJL, 568, L5

—. 2003a, ApJ, 583, 606

—. 2003b, ApJ, 583, 606

Koopmans, L. V. E., Treu, T., Bolton, A. S., Burles, S., and Moustakas, L. A. 2006a, ApJ, 649, 599

—. 2006b, ApJ, 649, 599

Koopmans, L. V. E., et al. 2009, ApJL, 703, L51

Kormann, R., Schneider, P., and Bartelmann, M. 1994, A&A, 284, 285

Kormendy, J. 1977, ApJ, 218, 333

Kormendy, J., and Bender, R. 2009, ApJL, 691, L142

Krawczyk, M., Sokołowska, D., Swaczyna, P., and Świeżewska, B. 2013, Journal of High Energy Physics, 9, 55

Kroupa, P. 2001, MNRAS, 322, 231

La Barbera, F., de Carvalho, R. R., de La Rosa, I. G., and Lopes, P. A. A. 2010a, MNRAS, 408, 1335

La Barbera, F., Ferreras, I., Vazdekis, A., de la Rosa, I. G., de Carvalho, R. R., Trevisan, M., Falcón-Barroso, J., and Ricciardelli, E. 2013, MNRAS, 433, 3017

La Barbera, F., Lopes, P. A. A., de Carvalho, R. R., de La Rosa, I. G., and Berlind, A. A. 2010b, MNRAS, 408, 1361

Lauer, T. R., et al. 1995, AJ, 110, 2622

Lemaître, G. 1931, MNRAS, 91, 483

Liesenborgs, J., and De Rijcke, S. 2012, MNRAS, 425, 1772

MacLeod, C. L., Jones, R., Agol, E., and Kochanek, C. S. 2013, ApJ, 773, 35

Mao, S., and Schneider, P. 1998, MNRAS, 295, 587

Maraston, C., Strömbäck, G., Thomas, D., Wake, D. A., and Nichol, R. C. 2009, MNRAS, 394, L107

Markwardt, C. B. 2009, in Astronomical Society of the Pacific Conference Series, Vol. 411, Astronomical Data Analysis Software and Systems XVIII, ed. D. A. Bohlender, D. Durand, & P. Dowler, 251–+

Martizzi, D., Teyssier, R., Moore, B., and Wentz, T. 2012, MNRAS, 422, 3081

Masters, K. L., et al. 2011, MNRAS, 418, 1055

McKean, J. P., et al. 2007, MNRAS, 378, 109

—. 2010, MNRAS, 404, 749

McKellar, A. 1941, Publications of the Dominion Astrophysical Observatory Victoria, 7, 251

Mehlert, D., Thomas, D., Saglia, R. P., Bender, R., and Wegner, G. 2003a, A&A, 407, 423

—. 2003b, A&A, 407, 423

Merritt, D. 1985a, MNRAS, 214, 25P

—. 1985b, AJ, 90, 1027

Metcalf, R. B., and Zhao, H. 2002, ApJL, 567, L5

Moles, M., Campos, A., Kjaegaard, P., Fasano, G., and Bettoni, D. 1998, ApJL, 495, L31

Moore, B., Ghigna, S., Governato, F., Lake, G., Quinn, T., Stadel, J., and Tozzi, P. 1999a, ApJL, 524, L19

Moore, B., Quinn, T., Governato, F., Stadel, J., and Lake, G. 1999b, MNRAS, 310, 1147

Moré, J. 1978, in Lecture Notes in Mathematics, Vol. 630, Numerical Analysis, ed. G. Watson (Springer Berlin Heidelberg), 105–116

Moustakas, L. A., and Metcalf, R. B. 2003, MNRAS, 339, 607

Moustakas, L. A., et al. 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7010, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series

Natarajan, P., Loeb, A., Kneib, J.-P., and Smail, I. 2002, ApJL, 580, L17

Navarro, J. F., Frenk, C. S., and White, S. D. M. 1996, ApJ, 462, 563

—. 1997, ApJ, 490, 493

Navarro, J. F., et al. 2010, MNRAS, 402, 21

Newman, A. B., Treu, T., Ellis, R. S., and Sand, D. J. 2011, ApJL, 728, L39

Newton, E. R., Marshall, P. J., Treu, T., Auger, M. W., Gavazzi, R., Bolton, A. S., Koopmans, L. V. E., and Moustakas, L. A. 2011, ApJ, 734, 104

Newton, I. 1760, Philosophiae naturalis principia mathematica, vol. 1 - 4

Nipoti, C., Treu, T., Ciotti, L., and Stiavelli, M. 2004, MNRAS, 355, 1119

Oñorbe, J., Domínguez-Tenreiro, R., Sáiz, A., Serna, A., and Artal, H. 2005, ApJL, 632, L57

Oguri, M., Rusu, C. E., and Falco, E. E. 2013, ArXiv e-prints

Oort, J. H. 1932, Bulletin of the Astronomical Institutes of the Netherlands, 6, 249

Osipkov, L. P. 1979, Pisma v Astronomicheskii Zhurnal, 5, 77

Padmanabhan, N., et al. 2004, New Astron., 9, 329

Pahre, M. A., Djorgovski, S. G., and de Carvalho, R. R. 1996, ApJL, 456, L79

—. 1998, AJ, 116, 1591

Palacios, A., Gebran, M., Josselin, E., Martins, F., Plez, B., Belmas, M., and Lèbre, A. 2010, A&A, 516, A13+

Paturel, G., and Garnier, R. 1992, A&A, 254, 93

Peebles, P. J. E. 1982, ApJL, 263, L1

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

Percival, W. J., Cole, S., Eisenstein, D. J., Nichol, R. C., Peacock, J. A., Pope, A. C., and Szalay, A. S. 2007, MNRAS, 381, 1053

Planck Collaboration et al. 2013, ArXiv e-prints

Prugniel, P., and Soubiran, C. 2001, A&A, 369, 1048

Richardson, T., and Fairbairn, M. 2014, ArXiv e-prints

Richstone, D., and Sargent, W. L. W. 1972, ApJ, 176, 91

Rix, H.-W., and White, S. D. M. 1992, MNRAS, 254, 389

Romano-Díaz, E., Shlosman, I., Hoffman, Y., and Heller, C. 2008, ApJL, 685, L105

Rubin, V. C., and Ford, Jr., W. K. 1970, ApJ, 159, 379

Ruff, A. J., Gavazzi, R., Marshall, P. J., Treu, T., Auger, M. W., and Brault, F. 2011, ApJ, 727, 96

Saha, P. 2000, AJ, 120, 1654

Saha, P., and Williams, L. L. R. 1997, MNRAS, 292, 148

Salpeter, E. E. 1955, ApJ, 121, 161

Sargent, W. L. W., Schechter, P. L., Boksenberg, A., and Shortridge, K. 1977, ApJ, 212, 326

Schlegel, D., White, M., and Eisenstein, D. 2009, in Astronomy, Vol. 2010, astro2010: The Astronomy and Astrophysics Decadal Survey, 314–+

Schneider, P., and Sluse, D. 2013, A&A, 559, A37

Scodeggio, M., Gavazzi, G., Belsole, E., Pierini, D., and Boselli, A. 1998, MNRAS, 301, 1001

Sheth, R. K., et al. 2003, ApJ, 594, 225

Shu, Y., Bolton, A. S., Schlegel, D. J., Dawson, K. S., Wake, D. A., Brownstein, J. R., Brinkmann, J., and Weaver, B. A. 2012, AJ, 143, 90

Slosar, A., Melchiorri, A., and Silk, J. I. 2005, PhRvD, 72, 101301

Sonnenfeld, A., Treu, T., Gavazzi, R., Marshall, P. J., Auger, M. W., Suyu, S. H., Koopmans, L. V. E., and Bolton, A. S. 2012, ApJ, 752, 163

Sonnenfeld, A., Treu, T., Gavazzi, R., Suyu, S. H., Marshall, P. J., Auger, M. W., and Nipoti, C. 2013, ApJ, 777, 98

Spiniello, C., Koopmans, L. V. E., Trager, S. C., Czoske, O., and Treu, T. 2011, MNRAS, 417, 3000

Spiniello, C., Trager, S., Koopmans, L. V. E., and Conroy, C. 2014, MNRAS, 438, 1483

Spiniello, C., Trager, S. C., Koopmans, L. V. E., and Chen, Y. P. 2012, ApJL, 753, L32

Strader, J., Caldwell, N., and Seth, A. C. 2011, AJ, 142, 8

Strauss, M. A., et al. 2002, AJ, 124, 1810

Suyu, S. H., Marshall, P. J., Auger, M. W., Hilbert, S., Blandford, R. D., Koopmans, L. V. E., Fassnacht, C. D., and Treu, T. 2010, ApJ, 711, 201

Suyu, S. H., et al. 2013, ArXiv e-prints

Thomas, D., Maraston, C., Bender, R., and Mendes de Oliveira, C. 2005, ApJ, 621, 673

Thuan, T. X., and Gunn, J. E. 1976, PASP, 88, 543

Tonry, J., and Davis, M. 1979, AJ, 84, 1511

Tonry, J. L., and Davis, M. 1981, ApJ, 246, 680

Toomre, A., and Toomre, J. 1972, ApJ, 178, 623

Tortora, C., Napolitano, N. R., Romanowsky, A. J., Capaccioli, M., and Covone, G. 2009, MNRAS, 396, 1132

Tortora, C., Romanowsky, A. J., and Napolitano, N. R. 2013, ApJ, 765, 8

Trager, S. C., Faber, S. M., Worthey, G., and González, J. J. 2000, AJ, 120, 165

Treu, T. 2010, ARA&A, 48, 87

Treu, T., Auger, M. W., Koopmans, L. V. E., Gavazzi, R., Marshall, P. J., and Bolton, A. S. 2010, ApJ, 709, 1195

Treu, T., Gavazzi, R., Gorecki, A., Marshall, P. J., Koopmans, L. V. E., Bolton, A. S., Moustakas, L. A., and Burles, S. 2009, ApJ, 690, 670

Treu, T., Koopmans, L. V., Bolton, A. S., Burles, S., and Moustakas, L. A. 2006, ApJ, 640, 662

Treu, T., and Koopmans, L. V. E. 2002, ApJ, 575, 87

—. 2004, ApJ, 611, 739

Trujillo, I., Burkert, A., and Bell, E. F. 2004, ApJL, 600, L39

Turner, E. L., Ostriker, J. P., and Gott, III, J. R. 1984, ApJ, 284, 1

Valdes, F., Gupta, R., Rose, J. A., Singh, H. P., and Bell, D. J. 2004, ApJS, 152, 251

van de Ven, G., Falcón-Barroso, J., McDermid, R. M., Cappellari, M., Miller, B. W., and de Zeeuw, P. T. 2010, ApJ, 719, 1481

van der Marel, R. P., and van Dokkum, P. G. 2007, ApJ, 668, 756

van der Wel, A., Holden, B. P., Zirm, A. W., Franx, M., Rettura, A., Illingworth, G. D., and Ford, H. C. 2008, ApJ, 688, 48

van Dokkum, P. G., and Conroy, C. 2010, Natur, 468, 940

Vegetti, S., Koopmans, L. V. E., Bolton, A., Treu, T., and Gavazzi, R. 2010, MNRAS, 408, 1969

Vegetti, S., Lagattuta, D. J., McKean, J. P., Auger, M. W., Fassnacht, C. D., and Koopmans, L. V. E. 2012, Natur, 481, 341

Velliscig, M., van Daalen, M. P., Schaye, J., McCarthy, I. G., Cacciato, M., Le Brun, A. M. C., and Dalla Vecchia, C. 2014, ArXiv e-prints

Wayth, R. B., and Webster, R. L. 2006, MNRAS, 372, 1187

Weinberg, D. H., Bullock, J. S., Governato, F., Kuzio de Naray, R., and Peter, A. H. G. 2013, ArXiv e-prints

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

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

Wright, E. L., et al. 2010, AJ, 140, 1868

Wucknitz, O. 2002, MNRAS, 332, 951

Xu, D. D., et al. 2009, MNRAS, 398, 1235

York, D. G., et al. 2000, AJ, 120, 1579

Zwicky, F. 1933, Helvetica Physica Acta, 6, 110