

# **Systematic Errors for Weak Lensing Surveys**

by

Chien-Hao Lin

Submitted in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

at

Carnegie Mellon University  
Department of Physics  
Pittsburgh, Pennsylvania

Advised by Professor Rachel Mandelbaum

August 24, 2020



## Abstract

In the era of precision cosmology, a wide range of cosmological surveys, such as LSST of Rubin Observatory, DESI, Euclid and Roman Space Telescope will precisely probe the large-scale structure and expansion history of the universe, shedding light on the nature of the dark sector. The unprecedented high-quality data from upcoming weak lensing surveys is expected to reduce the statistical uncertainty in weak lensing measurements compared to current surveys by roughly an order of magnitude. This will be an opportunity to explore mysteries in contemporary cosmology, including the apparent tensions between the early and the late universe as well as the nature of dark energy and dark matter. In the next generation of weak lensing surveys, the analysis requires a better understanding and more careful control of systematic uncertainties to avoid their dominating over statistical uncertainties. Low-level sources of systematic uncertainty that are presently ignored in current weak lensing analyses will become significant as the precision of the measurements increases.

In the first part of the thesis, we studied the systematic bias of weak gravitational lensing due to the assumption of Gaussian likelihoods. Weak gravitational lensing has been proven to be a powerful probe to constrain cosmology. The standard approach that converts the weak lensing measurements into constraints on cosmological parameters is to perform likelihood analyses by assuming the underlying likelihood distribution of the shear two-point statistics is a multivariate Gaussian likelihood function. However, it has been known in the literature that the likelihood function of weak lensing shear two-point statistics is not strictly Gaussian. It is therefore important to quantify potential cosmological parameter biases that may arise due to the assumption of an incorrect likelihood function, so as to determine whether an alternative way of modeling the likelihood function is needed. We study the significance of non-Gaussianity in the likelihood of weak lensing shear two-point correlation functions using simulated weak lensing data. We find significant non-Gaussianity in the marginal distributions of the correlation functions, indicating that the multivariate likelihood distributions are non-Gaussian. In order to estimate the bias in cosmological parameters due to carrying out the analysis while assuming an incorrect likelihood distribution, we construct likelihood models with both parametric and non-parametric methods and then infer the posteriors of  $\Omega_m$  and  $\sigma_8$  using the MCMC stochastic sampling method. To properly assess likelihood models and avoid overfitting, we adopt cross-validation and perform non-parametric two-samples tests. Based on the results on simulated data and estimates of how the non-Gaussianity scales with survey area, we demonstrate that neglecting the non-Gaussianity of the likelihood is not a significant source of bias for wide-field imaging surveys such as LSST.

The rest part of the thesis is devoted to the systematic biases for the upcoming near infrared weak lensing survey. Recently, there have been more studies of the impact of

detector non-idealities on weak lensing. Since Roman Space Telescope will focus on infrared imaging, it will use CMOS based near-infrared detectors. Unlike optical CCDs, near-infrared detectors typically suffer from different detector non-idealities. Among the systematic effects in these detectors that will be relevant for Roman Space Telescope are the light polarization effect and the image persistence effect. Persistence is the phenomenon of retaining a small but non-negligible fraction of the flux in images from the previous exposures after a reset. This leads to images being contaminated by bright stars in the previous exposures or during the telescope slews. In our work, we implemented image simulations, combined with the dithering strategies, to investigate impact of persistence on weak lensing measurements. Light polarization effect, on the other hand, is the impact of light polarization of galaxies on weak lensing. Since the efficiency of detectors and the response of optical devices depend on the polarization of incoming light rays, if the light from galaxies is linearly polarized, the existing analysis methods that ignore the polarization information could lead to biases in the weak lensing results. Even though the linear polarization of light from galaxies has been observed, polarization-induced systematic errors have not generally been considered for optical weak lensing analyses. We investigated two polarization-induced systematic biases: the selection bias in favor of galaxies with specific orientations and the polarization-dependent PSF uncertainty. The selection bias arises from the dependence of the transmitted intensity of polarized light on the angle of polarization. The polarization-dependent PSF, on the other hand, is a result of the polarization-dependent optical aberrations. We build toy models to obtain for the first time an estimate for both of these polarization-induced biases and show that both biases are comparable to the Roman Space Telescope systematics tolerance level, indicating the need for more detailed studies.

## Acknowledgments

Pursuing a PhD is a long journey that can not be completed without all the help and support over the years. First and foremost, I wish to express my deepest gratitude to my advisor, Rachel Mandelbaum. Rachel has always been the strongest support during the days of my PhD. It is her encouragement and patience that helped me overcome the difficult times. She is a great teacher who led my exploration in a field of which I knew almost nothing when I first started. She is the primary resource for getting my science questions answered and she is always patient to my ignorant questions. She is also a great mentor who guided me to become a better researcher in cosmology. I am constantly inspired by her profound knowledge, her persistence in pursuing high quality research and her passion in science and in her work. My sincere thanks also go to Professor Shirley Ho for her valuable advice and guidance in my first year and to the members of my committee, Professors Scott Dodelson, Jefferey Newman and Matthew Walker for helpful academic suggestions and career advice.

I am grateful to my research collaborators, Joachim Harnois-Déraps, Tim Eifler, Taylor Pospisil, Ann Lee, Sukhdeep Singh, Christopher Hirata, Brent Tan, Michael Troxel and Mike Jarvis for their excellent work and making various research projects possible. I also thank the LSST DESC review committee, Seth Digel, Alan Heavens, Joe Zuntz and Rogério Rosenfeld for their useful comments during the long and repeated review process.

Many colleagues of mine at CMU deserve a lot of thanks. In particular, I would like to thank Hung-Jin Huang, my predecessor, for many useful discussions on weak lensing and constructive career advice. I also thank Yao-Yuan Mao, Markus Rau, Danielle Leonard, Francois Lanusse, Husni Almoubayyed, Yesukhei Jagvaral, Tianqing Zhang, Simon Samuroff, Duncan Campbell, Carolina Nunez, Shadab Alam, Sebastian Fromenteau and Layne Price for all the great academic discussions and assistance.

I would like to thank my friends Mei-Yu Wang, Hsiu-Hsien Lin, Bai-Cian Ke, Yen-Chi Chen, Kuan-Wei Huang, Zhonghao Luo, Abel Sun, Mandy Huang, Yueying Ni, Matthew Ho, Siddharth Satpathy, Elena Giusarma, I-Hsuan Kao, Yufeng Shen, Xiaochuan Wu, Vikesh Siddhu, Nianyi Chen, Andresa Rodrigues de Campos, Siyu He, Yingzhang Chen, Yuchi He, Yizhou He and He Liu for their company, conversations and dinners. I am also grateful for the support and the friendly environment that the CMU Physics department and the McWilliams center provide. I appreciate the support from the graduate program heads Prof. Manfred Paulini and Fred Gilman, and the members of staff in the physics department as a graduate student.

The last word of acknowledgment is dedicated to my family and the love of my life, Yi-Jhen, for their love, support and understanding. They selflessly encouraged me to explore my own life in my own way. Without them, I cannot have the opportunities and experiences that have made me who I am today.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Contemporary Cosmology . . . . .	1
1.1.1	Metrics and Distances . . . . .	2
1.1.2	Standard Model and Dark Components of the Universe . . . . .	3
1.2	Weak Lensing . . . . .	5
1.2.1	Gravitational Lensing . . . . .	5
1.2.2	Cosmic Shear . . . . .	8
1.2.3	Bayesian Inference . . . . .	9
1.2.4	Systematic Errors . . . . .	10
1.3	Structure of the Thesis . . . . .	11
<b>2</b>	<b>Non-Gaussianity in the Weak Lensing Correlation Function Likelihood</b>	<b>13</b>
2.1	Abstract . . . . .	13
2.2	Introduction . . . . .	14
2.3	Simulations . . . . .	17
2.4	Method . . . . .	18
2.4.1	Cosmic shear correlation function data vector . . . . .	18
2.4.2	Modeling of observables . . . . .	20
2.4.3	Likelihood functions and data covariance matrix . . . . .	22
2.4.4	PCA transformation . . . . .	23
2.4.5	Multivariate likelihood models . . . . .	23
2.4.6	Skewness and Kurtosis . . . . .	26
2.5	Likelihood model assessment . . . . .	27
2.6	Results . . . . .	29
2.6.1	Biases due to non-Gaussian distributions of $\xi_{\pm}$ . . . . .	29
2.6.2	Modeling the distributions of principal components . . . . .	33
2.6.3	Biases due to non-Gaussian distributions of principal components . . . . .	36
2.7	Conclusions and Discussions . . . . .	36
2.8	Third Central Moment . . . . .	44
2.8.1	Convergence . . . . .	44
2.8.2	Shear . . . . .	47
2.9	Difference between mean and mode . . . . .	47

<b>3</b>	<b>The Impact of Light Polarization Effects on Weak Lensing Systematics</b>	<b>49</b>
3.1	Abstract . . . . .	49
3.2	Introduction . . . . .	49
3.3	Optical/NIR Polarization of galaxies . . . . .	51
3.3.1	Toy Model for Galaxy Polarizations . . . . .	52
3.3.2	Corrections to the toy model . . . . .	53
3.4	Polarization dependent Optical response and PSF . . . . .	53
3.4.1	Polarization Repsonce Difference . . . . .	53
3.4.2	Polarization-Induced Astigmatism . . . . .	54
3.5	Selection Bias . . . . .	55
3.5.1	Defining a Galaxy Population . . . . .	55
3.5.2	Effect of Polarization on Magnitude . . . . .	56
3.5.3	Magnitude Cut . . . . .	57
3.5.4	Selection bias estimation process . . . . .	57
3.5.5	Parameters in the simplified toy model . . . . .	58
3.6	PSF Errors . . . . .	59
3.6.1	Image Simulation with Galsim . . . . .	59
3.6.2	PSF shape measurement . . . . .	59
3.7	Results . . . . .	60
3.7.1	Selection bias . . . . .	60
3.7.2	PSF size and shape errors . . . . .	61
3.8	Discussion . . . . .	64
<b>4</b>	<b>Image Persistence Effect on Weak Lensing</b>	<b>66</b>
4.1	Abstract . . . . .	66
4.2	Introduction . . . . .	66
4.3	Persistence Effect . . . . .	68
4.3.1	Theory of trapped charge carriers . . . . .	68
4.3.2	Persistence model . . . . .	68
4.4	Simulation . . . . .	71
4.4.1	Galsim . . . . .	71
4.4.2	Galaxy catalog . . . . .	71
4.4.3	Implementation of persistence effect . . . . .	71
4.4.4	simulation steps . . . . .	71
4.5	Results . . . . .	72
4.5.1	Persistence effect in simulations . . . . .	72
4.5.2	Correlation between shape errors of galaxies . . . . .	73
4.6	Conclusions . . . . .	73
<b>5</b>	<b>Conclusion and Outlook</b>	<b>79</b>

# Chapter 1

## Introduction

### 1.1 Contemporary Cosmology

The modern understanding of the universe is built upon a few great progresses both in theory and observation in the past century. In 1910s, Einstein completed his theory of General Relativity (GR) that generalizes Newtonian gravity and constructed the first model of the Universe based on the new theory of gravity and spacetime. He introduced the cosmological constant  $\Lambda$ , which provides repulsive gravity that counters the attractive gravity by matter to yield a static solution for the universe.

In 1929, Hubble discovered that galaxies farther away from us are receding at higher velocities. This discovery implied that the universe is expanding homogeneously, and the expansion of the universe has been well accepted since then. In 1960s, the discovery of Cosmic Microwave Background radiation (CMB) provided evidence for the Hot Big Bang Theory and proved that the universe was dominated by radiation at early times. And in 1970s, the significant discrepancy between the observed galaxy rotation curve and the predicted one by the gravity of observed matter led to the postulated concept of dark matter.

Since the 2000s, the field of cosmology entered the precision era with the launch of several cosmological experiments. One of the most important discovery of contemporary cosmology is that the expansion of the Universe is accelerating. This was first discovered by determining distance as a function of galaxy redshift using supernova (Riess et al., 1998; Perlmutter et al., 1999). The accelerated expansion was further confirmed by the studies of CMB, baryon acoustic oscillation (BAO), galaxy clustering and weak lensing. The extensive evidence for cosmic acceleration suggested the existence of dark energy that provides negative pressure to drive the acceleration and revived the concept of cosmological constant  $\Lambda$ . The flat  $\Lambda$ -Cold Dark Matter ( $\Lambda$ CDM) model has become a widely accepted model in modern cosmology.

### 1.1.1 Metrics and Distances

The Standard Model is built around GR, and the space-time separations in GR between events are described by a metric. In the 1930s, Robertson and Walker (independently) showed that there are only three possible spacetime metrics for a universe that is homogeneous and isotropic. They can be written

$$ds^2 = -c^2 dt^2 + a^2(t)[d\chi^2 + S_k^2(\chi)d\Omega^2], \quad (1.1)$$

where

$$S_k(\chi) = \begin{cases} \sin(\chi) & k = 1 \\ \chi & k = 0 \\ \sinh(\chi) & k = -1 \end{cases} \quad (1.2)$$

In this notation,  $k$  is a constant indicating the geometry of the space and  $a(t)$  is a dimensionless scale factor. It is defined so that  $a(t_0) = 1$  at the time  $t_0$  (usually taken to be the present) when the curvature radius is  $R_0$ .  $\chi$  is the comoving distance defined as

$$d\chi = \frac{dr}{\sqrt{1 - kr^2}} = c \frac{dt}{a(t)} \quad (1.3)$$

In observations, we often measure the angular size of objects, which can then be converted to physical size using the line of sight distance,  $D_A$  (called angular diameter distance)

$$D_A(a) = aS_k(\chi) \quad (1.4)$$

## 1.1.2 Standard Model and Dark Components of the Universe

The explanation of the history of the Universe is derived from the widely accepted Lambda Cold Dark Matter cosmological model of the big bang cosmology. This standard model, usually abbreviated as  $\Lambda$ CDM model, is a mathematical model that parametrizes the cosmology with only six parameters under the framework of General Relativity and the Friedman equation.  $\Lambda$ CDM model assumes that the universe is composed of photons, neutrinos, ordinary matter, cold dark matter and dark energy, which take the form of a constant energy density, referred to as the cosmological constant  $\Lambda$ . Though simple, the flat  $\Lambda$ CDM model has become a most accepted model in modern cosmology. Various cosmological probes from the larger sample size of supernova, higher resolution CMB data, baryon acoustic oscillation (BAO), galaxy clustering and weak lensing, and massive clusters in X-ray surveys, have shown consistency with the  $\Lambda$ CDM model. Despite the success of  $\Lambda$ CDM, the recent highest-resolution CMB measurements from the Planck experiment has revealed a tendency of tension with other low redshift cosmological probes (e.g., Bernal et al. (2016)).

### Dark Energy

The cosmological constant was first introduced by Einstein into his theory of General Relativity in 1917 for cosmological considerations (Einstein, 1917). The cosmological constant terms provided a repulsive force that countered the attraction by gravity to keep the universe stable and static. The cosmological constant term was later ignored by physicists for decades after Hubble discovered that the universe was expanding in 1929.

To determine how different components of the Universe drive the cosmological expansion, we can derive the Friedman equation from Einstein's field equations and relate the Hubble parameter with energy densities:

$$H^2(a) = \frac{8\pi G}{3} [\rho_r(a) + \rho_m(a) + \rho_k(a) + \rho_\Lambda(a)], \quad (1.5)$$

where the Hubble parameter is defined as  $H(a) \equiv \dot{a}/a$ , and  $\rho_k$  and  $\rho_\Lambda$  are the energy densities of the spatial curvature and cosmological constant with  $\rho_\Lambda = \frac{\Lambda c^2}{8\pi G}$ ,  $\rho_k = \frac{-3kc^2}{a^2 8\pi G}$ . The Friedman equation can be further simplified by defining the equation of states that relates the pressure  $p$  and the mass density  $\rho$  of a perfect fluid:

$$p = w\rho c^2 \quad (1.6)$$

The solution of the scale factor for a single perfect fluid is

$$\rho \propto a^{-3(1+w)}. \quad (1.7)$$

The case of  $w = 0$  corresponds to a matter-dominated universe, implying that  $\rho_m \propto a^{-3}$ . Similarly, a radiation dominated universe has  $w = 1/3$  and thus  $\rho_r \propto a^{-4}$ . And for cosmological parameter  $\Lambda$ ,  $w = -1$  and  $\rho_\Lambda = \text{const.}$

We can also rewrite the Friedman’s acceleration equation using the equation of states:

$$\frac{\ddot{a}}{a} = \frac{-4\pi G}{3}\rho(1 + 3w) \quad (1.8)$$

It was the supernova experiments that revived the cosmological constant in late 1990s when two groups of supernova observers independently reported the critical discovery that the expansion rate of the universe is expanding (Riess et al., 1998; Perlmutter et al., 1999). In Eq. 1.8, the cosmological constant with  $w = -1$  can drive the accelerated expansion under the framework of  $\Lambda$ CDM. The observational evidence for the cosmic acceleration has grown since then. Although a wide range of measurements, including larger supernova experiments (e.g., Wood-Vasey et al. (2007); Conley et al. (2011)), high-resolution CMB measurements (e.g., Larson et al. (2011); Planck Collaboration et al. (2014b, 2016, 2018)), BAO (e.g., Eisenstein et al. (2005); Percival et al. (2010)) and precise measurements of  $H_0$  (e.g., Riess et al. (2009); Freedman et al. (2012)) reinforced the cosmic acceleration since the 2000s, the origin of the acceleration remains mysterious. Inspired by the profound implications of cosmic acceleration, a number of ambitious surveys, including DESI (DESI Collaboration et al., 2016), Euclid (Laureijs et al., 2011), Rubin Observatory LSST (Ivezić et al., 2019), and Roman Space Telescope (Spergel et al., 2015) are planned to shed light on the origin of cosmic acceleration and determine the properties of dark energy.

## Dark Matter

The first evidence for the existence of matter that does not interact with light was presented in observations of the Coma cluster in the 1930s by Fritz Zwicky (Zwicky, 1937), who measured the velocities of the Coma member galaxies and determined that they were moving faster than their combined stellar mass could bind together gravitationally. Further evidence was announced in the 1970s with Vera Rubin’s measurements of galaxy rotation curves (Rubin & Ford, 1970), which showed that the total mass of galaxies increased significantly with radius even though very little additional stellar mass was present at these larger radii. More recent imaging of the merging Bullet cluster, in lensing and X-ray provide another instance of a cluster scale object with apparently non-identical mass and light distributions (e.g., Clowe et al. (2007)). The inference of large amount of mass from gravitational lensing, and the CMB experiment, has left the existence of dark matter as a key component of the Universe. Extensive observational evidence has now made the existence of the dark matter component very difficult to dispute.

In addition to the evidence, determining the nature of dark matter has become an outstanding problem of modern physics. Over the last several decades, an extensive experimental program has sought to determine the cosmological origin, fundamental constituents, and interaction mechanisms of dark matter. Various collider experiments attempt to directly detect dark matter particles or indirectly detect dark matter through the products of dark matter particles (for reviews, see Feng (2010); Boveia & Doglioni (2018)). Despite these extensive efforts, astrophysical and cosmological observations remain the only empirical measurement of dark matter. So far, the observational data is well described by a simple,

non-relativistic, collisionless, cold dark matter model. However, many alternative models of dark matter predict observable deviations from CDM, which are testable with current and future experiments. In addition, fundamental properties of dark matter, such as particle mass, time evolution and coupling to the Standard Model particles, can leave message in the macroscopic distribution of dark matter (Feng, 2010).

## 1.2 Weak Lensing

Weak gravitational lensing is a powerful probe of cosmology as it is sensitive to both geometry and growth of the structure in the universe (see Weinberg et al. (2013); Kilbinger (2015); Dodelson (2017) for reviews). Weak gravitational lensing is the deflection of light by the gravitational field of large-scale structure, which leads to minute distortions of the observed galaxy images compared to their original shapes in the galaxy source plane. In the weak regime, lensing introduces small but coherent distortions in the shapes (shear) and sizes (convergence) of background sources. These distortions contain the information of the large scale structure between the observer and the source. Measuring the correlation functions of the galaxy shapes is therefore a way to measure the growth of structure and the geometry of the Universe (e.g., Bartelmann & Schneider, 2001; Hoekstra & Jain, 2008; Kilbinger, 2015; Mandelbaum, 2018) and hence a promising avenue to constrain cosmology (e.g., Huff et al. (2014); Jee et al. (2016); Hildebrandt et al. (2018); Troxel et al. (2018a); Hikage et al. (2019b)). Despite the potential, weak lensing measurements are also difficult. Typical weak lensing shear is of order 0.01 compared to intrinsic galaxy ellipticities of order 0.2. In a lensing survey we need to accurately measure shapes of a large number of galaxies, while keeping any systematic errors to a level lower than statistical errors. The signals from weak lensing measurement are subject to a number of systematics originating from shape measurement errors, noisy photometric redshifts measurements, modeling uncertainties and astrophysical systematics. Therefore, the biggest challenge of weak lensing lies in estimating and studying the sources of systematics errors. This is also the main focus of this dissertation. With the next generation of weak lensing surveys, such as the Legacy Survey of Space and Time of Rubin Observatory (LSST Ivezić et al. 2019), the Roman Space Telescope and *Euclid*, we expect data sets that are both wider and deeper compared to current surveys in the near future. Though current results are dominated by statistical uncertainties, the high-quality data from upcoming lensing surveys is expected to reduce the statistical uncertainty in weak lensing measurements compared to current surveys and the systematics will become important in the future. It will therefore be crucial to improve our understanding of various systematic errors in the weak lensing.

### 1.2.1 Gravitational Lensing

In this section, we summarize the formalism of gravitational lensing under the framework of General Relativity. We will derive some simple and useful relations of gravitational lens-

ing that provides intuitive insights to the lensing effect. For more detailed formalism, we refer the reader to review paper of weak lensing (Bartelmann & Schneider, 2001; Weinberg et al., 2013; Kilbinger, 2015). The derivation below will follow the notations in Bartelmann & Schneider (2001).

In general relativity, a point mass with mass  $M$  deflects a light ray with impact factor  $\xi$  by an angle of

$$\hat{\alpha} = \frac{4GM}{c^2\xi}. \quad (1.9)$$

Since Eq. 1.9 is linear in  $M$ , this equation can be generalized to an array of lens easily in the cases where the gravitational field is weak and linearized. Now suppose we have  $N$  points on the lens plane with masses  $M_i$  and impact factors  $\xi_i$ . The deflection angle by the  $N$  lenser will be

$$\hat{\alpha} = \frac{4G}{c^2} \sum_i M_i \frac{\vec{\xi} - \vec{\xi}_i}{|\vec{\xi} - \vec{\xi}_i|^2}. \quad (1.10)$$

Under the thin lens approximation that the size of the lens is much smaller than the distances between the source, lens and the observer, we can project the 3-D mass of the lens onto the 2-D lens plane and define the surface density,

$$\Sigma(\vec{\xi}) = \int \rho(\vec{\xi}, z) dz, \quad (1.11)$$

where  $\rho$  is the 3-D mass density. The total deflection of Eq. 1.10 becomes

$$\hat{\alpha} = \frac{4G}{c^2} \int d^2\xi' \Sigma(\vec{\xi}') \frac{\vec{\xi} - \vec{\xi}'}{|\vec{\xi} - \vec{\xi}'|^2}. \quad (1.12)$$

Figure 1.1 shows the diagram of the lensing system. This figure contains the source and the lens plane, both perpendicular to the vertical optical axis, indicated by the solid dashed line. From the diagram, we can relate the angles  $\theta$ ,  $\alpha$  and  $\beta$  with the lens equation:

$$\vec{\theta} D_s = \vec{\beta} D_s + \hat{\alpha} D_{ds}, \quad (1.13)$$

Or

$$\vec{\beta} = \vec{\theta} - \vec{\alpha}(\vec{\theta}), \quad (1.14)$$

with  $\vec{\alpha} \equiv \frac{D_{ds}}{D_s} \hat{\alpha}$ .

If we define a dimension-less surface density (convergence) with  $\Sigma(\vec{\xi})$ :

$$\kappa(\vec{\theta}) \equiv \frac{\Sigma(D_d \vec{\theta})}{\Sigma_{cr}}, \quad \text{with} \quad \Sigma_{cr} = \frac{c^2}{4\pi G} \frac{D_s}{D_d D_{ds}}, \quad (1.15)$$

we can rewrite Eq. 1.12 as

$$\vec{\alpha}(\vec{\theta}) = \frac{1}{\pi} \int d^2\theta' \kappa(\vec{\theta}') \frac{\vec{\theta} - \vec{\theta}'}{|\vec{\theta} - \vec{\theta}'|^2}. \quad (1.16)$$

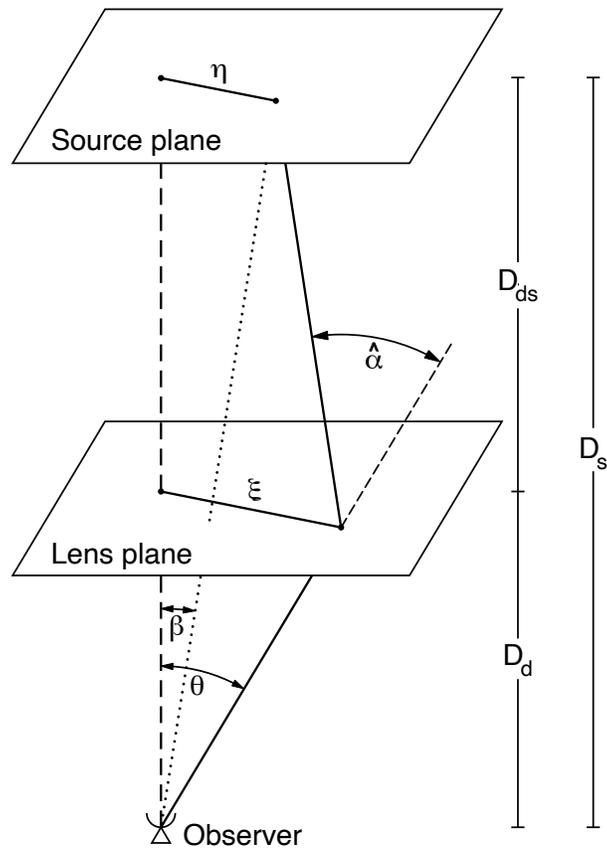


Figure 1.1: Sketch of a typical gravitational lens system (from Bartelmann & Schneider (2001)).

It is natural to define another scalar quantity which is commonly called the effective lensing potential

$$\psi(\vec{\theta}) = \frac{1}{\pi} \int d^2\theta' \kappa(\vec{\theta}') \ln(\vec{\theta} - \vec{\theta}'), \quad (1.17)$$

so that the deflection vector is the gradient of the lensing potential

$$\vec{\alpha} = \nabla_{\theta} \psi \quad (1.18)$$

and it satisfies the Poisson equation

$$2\kappa(\vec{\theta}) = \nabla_{\theta}^2 \psi. \quad (1.19)$$

Given Eq. 1.14 and Eq. 1.18, we can derive the Jacobian matrix of the image distortion as

$$\mathcal{A}(\vec{\theta}) = \frac{\partial \vec{\beta}}{\partial \vec{\theta}} = \delta_{ij} - \frac{\partial \alpha_i}{\partial \theta_j} = \delta_{ij} - \frac{\partial^2 \psi(\vec{\theta})}{\partial \theta_i \partial \theta_j} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}, \quad (1.20)$$

where the convergence  $\kappa$  and the shear components  $\gamma_1$  and  $\gamma_2$  are defined as

$$\kappa = \frac{1}{2} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \psi, \quad (1.21)$$

$$\gamma_1 = \frac{1}{2} \left( \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} \right) \psi, \quad (1.22)$$

$$\gamma_2 = \frac{\partial^2 \psi}{\partial x \partial y}. \quad (1.23)$$

## 1.2.2 Cosmic Shear

The formalism in the last section is derived for one single thin lens plane. But in practice, a source galaxy is sheared by an infinite amount of continuous lens planes before it is observed by the observer on the earth. This effect is known as cosmic shear. Cosmic shear is the distortion of images of distant galaxies due to weak gravitational lensing by the large-scale structure in the Universe. Such images are coherently deformed by the tidal field of matter inhomogeneities along the line of sight. The idea of making the cosmic shear measurement was proposed in 1967 (Kristian, 1967), and was finally detected by several groups in the year of 2000 (e.g., Wittman et al. (2000); Bacon et al. (2000); Van Waerbeke et al. (2000)).

In Sect. 1.2.1 we define the convergence of the lens plane as the normalized surface density. We can extend the formalism to the case of cosmic shear by defining the effective convergence summing over all the source planes behind a given lens plane

$$\kappa_{eff}(\vec{\theta}, z_L) = \sum_s \kappa_s = \frac{4\pi G D_d \Sigma(\vec{\theta}, z_L)}{c^2} \int_{z_L}^{\infty} \frac{D_{ds}(z_s)}{D_s(z_s)} p(z_s) dz_s, \quad (1.24)$$

where  $p(z_s)$  is the source density. Summing over  $z_L$ , we obtain the total effective convergence which corresponds to the signal of cosmic shear:

$$\kappa_{eff}(\vec{\theta}) = \int g(\chi)\delta(\chi)d\chi, \quad (1.25)$$

$$g(\chi) = \frac{3\Omega_m H_0^2}{2c^2 a(\chi)} \int p(\chi') \frac{S_k(\chi)S_k(\chi' - \chi)}{S_k(\chi')} d\chi', \quad (1.26)$$

where the kernel  $g(\chi)$  is the lensing efficiency. As with the analogue of a classical lens, the lens is most efficient when positioned approximately mid-way between the observer and the source. As shown in Eq. 1.25, the cosmic shear signal is determined by the lensing efficiency which represents the geometry of the lens-source system as well as the matter density distribution. This allows the cosmic shear to probe both the geometry and growth of structure of the universe.

### 1.2.3 Bayesian Inference

The modern cosmology is pretty much built upon observations of the universe to study its origin and evolution. Especially, cosmology has evolved into a data-rich field in the last few decades when new telescopes and surveys allowed the production of a large amount of data. The observed data are used to test the predictions of cosmological theories and lead to improvement of cosmological models. Such analyses are usually done through statistics, in particular, Bayesian statistics. Here we outline the key concepts of Bayesian inference. For reviews of Bayesian inference and its applications to cosmology, please see Trotta (2008); Padilla et al. (2019).

In the frequentist's point of view, the probability of an event is simply the number of times that the event occurs in the limit of an infinite series of measurements. This interpretation is intuitive in the context of coin flip experiments, where the results are random. One can simply perform the random tosses many times and records the series of head and tails to estimate the probability of getting each results. The frequentist interpretation of probability makes less sense in the context of measuring cosmological parameters since there is only one universe. This problem can be avoided by the Bayesian stance, which views the probability as a measure of the degree of belief. Under this definition, the probability of a cosmological parameter is the observer's degree of confidence in the inferred parameter. Repetitive measurements on the single universe just improve our degree of belief.

Now suppose we have a column vector of the observed data (e.g., measurement of shear two-point correlation function or CMB fluctuations) denoted by  $\mathbf{D}$ . The observation is based on a specific cosmology with cosmological parameters  $\theta$ . Given a model  $M$ , the data vector can be predicted from the input parameters  $\theta$ . In the case of parameter inference, we would like to obtain the probability distribution of  $\theta$  given the model and the observed data. Bayes' theorem states that,

$$P(\theta|\mathbf{D}, M) = \frac{\mathcal{L}(\mathbf{D}|\theta, M)p(\theta|M)}{p(\mathbf{D}|M)}. \quad (1.27)$$

In this relation,  $P(\theta|\mathbf{D}, M)$  is called the posterior probability which expresses the conditional probability of the desired parameters.  $p(\theta|M)$  is called the prior and encodes the prior knowledge about the model before getting the observed data.  $p(\mathbf{D}|M)$  is the evidence of the model, which acts as a normalization factor in the relation. The first term,  $\mathcal{L}(\mathbf{D}|\theta, M)$ , is called the likelihood and represents the probability of observed data  $\mathbf{D}$  for a specific set  $\theta$  values.

In practice, it is typically very difficult to compute the posterior distribution of the parameters analytically, and thus numerical tools for estimating the posteriors are essential in these cases. Techniques including the Markov Chain Monte Carlo (MCMC) (Hastings, 1970; Metropolis et al., 1953) and the Nested Sampling algorithm (Skilling, 2006) have been approved to be computationally efficient, with publicly available packages implementing the algorithms, such as EMCEE (Foreman-Mackey et al., 2013), MultiNest (Feroz et al., 2009) and COSMOMC (Lewis & Bridle, 2002).

## 1.2.4 Systematic Errors

The typical magnitude of the weak lensing shear signal is of order 0.01 compared to intrinsic galaxy ellipticities of order 0.2. Thus, one major challenge in a lensing survey is to accurately measure shapes of a large number of galaxies, while keeping any systematic errors to a level lower than statistical errors. Note that the term "systematic errors" could ambiguously mean both systematic biases, which are the biases that can be identified, estimated and removed, and systematic uncertainties, which are the residual uncertainties after removing biases. A great amount of progress have been accumulated in the past decade to overcome various challenges. Much effort has been committed to understand observational systematic effects, such as uncertainties in the measurement of galaxy shapes (e.g., Massey et al., 2013; Mandelbaum et al., 2015) and photometric redshifts. On the astrophysical side, major progress has been made in modeling systematic biases that affect the interpretation of the weak lensing signal, such as the intrinsic alignment of galaxies (e.g., Mandelbaum et al., 2011; Joachimi et al., 2015; Troxel & Ishak, 2015; Krause et al., 2016), the non-linear evolution of the dark matter density field (e.g., Takahashi et al., 2012; Heitmann et al., 2014), and baryonic effects that modify the latter (e.g., Eifler et al., 2015; Mead et al., 2015; Chisari et al., 2018). For a more detailed review on the technical developments of weak lensing, from raw images to scientific inferences, one can take Mandelbaum (2018) for reference.

Currently, we have three weak lensing survey programs running, including the Kilo-Degree Survey (KiDS), the Hyper Suprime-Cam survey (HSC) and the Dark Energy Survey (DES). The ongoing surveys are excellent test grounds for the weak lensing systematic errors in preparation for the upcoming surveys in 2020s. With the next generation of weak lensing surveys, such as the Legacy Survey of Space and Time of Rubin Observatory (LSST Ivezic et al. 2019), the Roman Space Telescope and *Euclid*, the high-quality data and the large data volume from these upcoming lensing surveys is expected to reduce the statistical uncertainty in weak lensing measurements compared to current surveys by an

order of magnitude. As the precision of the measurements increases, low-level sources of systematic biases could become significant compared to the statistical noise. Besides the dominant systematic effects, such as photometric redshifts and intrinsic alignment, more efforts should be put on systematic biases that are presently ignored in current weak lensing analyses. In the following three chapters, we will investigate three different sources of relatively weak systematics biases in the context of future weak lensing surveys.

### 1.3 Structure of the Thesis

Besides the introduction in the first chapter, the main result of this thesis consists the results of three research projects.. All together they improve our understanding of systematics in the field of weak gravitational lensing. In Chapters 2, we focus on the assumption of Gaussian likelihood function and the resulting systematic bias for weak lensing. Chapters 3 and 4 concentrates on systematics due to the optical and detector effects specifically for the upcoming Roman Space Telescope. Below I provide a brief summary on what these projects are about.

In chapter 2, I present our investigation on the non-Gaussian weak lensing likelihood functions and its impact on the cosmological parameters. We study the significance of non-Gaussianity in the likelihood of weak lensing shear two-point correlation functions using simulated weak lensing data (Harnois-Déraps et al., 2018). We find significant non-Gaussianity in the marginal distributions of the correlation functions, indicating that the multivariate likelihood distributions are non-Gaussian. In order to estimate the bias in cosmological parameters due to carrying out the analysis while assuming an incorrect likelihood distribution, we construct likelihood models with both parametric and non-parametric methods and then infer the posteriors of  $\Omega_m$  and  $\sigma_8$  using the MCMC stochastic sampling method. To properly assess likelihood models and avoid overfitting, we adopt cross-validation and perform non-parametric two-samples tests, including energy distance (Székely & Rizzo, 2004) and maximum mean discrepancy (Gretton et al., 2012). Based on the results on simulated data and estimates of how the non-Gaussianity scales with survey area, we demonstrate that neglecting the non-Gaussianity of the likelihood is not a significant source of bias for wide-field galaxy surveys such as LSST. This work is accepted for publication in MNRAS.

Chapter 3 presents our study of the light polarization effects on Weak Lensing. We investigated two polarization-induced systematic biases: the selection bias in favor of galaxies with specific orientations and the polarization-dependent PSF uncertainty. The selection bias arises from the dependence of the transmitted intensity of polarized light on the angle of polarization. The polarization-dependent PSF, on the other hand, is a result of the polarization-dependent optical aberrations. We build toy models to obtain for the first time an estimate for both of these polarization-induced biases and show that both biases are comparable to the Roman Space Telescope systematics tolerance level, indicating the need for more detailed studies. This work is published in Lin et al. (2020), MNRAS, 496, 532.

Chapter 4 presents our results on the study of the persistence detector effect and the resulting systematic effect on weak lensing. The Roman Space Telescope will survey the sky in the near infrared (NIR) band using its specially designed HgCdTe photodiode arrays called H4RG. For the NIR arrays, charges that are trapped in the photodiodes during earlier exposure are gradually released into the current exposure, leading to contamination of the images and errors of the galaxy shapes. This memory effect is called persistence. We use image simulations that incorporate the persistence to study its impact on galaxy shapes and weak lensing signal. We analyze the shape errors due to persistence. No significant correlation between the shape shift by persistence is detected, indicating that persistence does not introduce coherent shape distortions on galaxies. In the scales of interest, the effect of persistence is about two orders of magnitude lower than the Roman Space Telescope additive shear error budget, indicating that the persistence is expected to be a weak effect on weak lensing for Roman Space Telescope given the current design. The work presented in this chapter, at the time of writing, is in preparation for submission. The content is subject to further changes before it is submitted for publication.

Finally, a summary of the conclusions and the outlook is presented in Chapter 5.

## Chapter 2

# Non-Gaussianity in the Weak Lensing Correlation Function Likelihood

### 2.1 Abstract

We study the significance of non-Gaussianity in the likelihood of weak lensing shear two-point correlation functions, detecting significantly non-zero skewness and kurtosis in one-dimensional marginal distributions of shear two-point correlation functions in simulated weak lensing data. We examine the implications in the context of future surveys, in particular LSST, with derivations of how the non-Gaussianity scales with survey area. We show that there is no significant bias in one-dimensional posteriors of  $\Omega_m$  and  $\sigma_8$  due to the non-Gaussian likelihood distributions of shear correlations functions using the mock data (100 deg<sup>2</sup>). We also present a systematic approach to constructing approximate multivariate likelihoods with one-dimensional parametric functions by assuming independence or more flexible non-parametric multivariate methods after decorrelating the data points using principal component analysis (PCA). While the use of PCA does not modify the non-Gaussianity of the multivariate likelihood, we find empirically that the one-dimensional marginal sampling distributions of the PCA components exhibit less skewness and kurtosis than the original shear correlation functions. Modeling the likelihood with marginal parametric functions based on the assumption of independence between PCA components thus gives a lower limit for the biases. We further demonstrate that the difference in cosmological parameter constraints between the multivariate Gaussian likelihood model and more complex non-Gaussian likelihood models would be even smaller for an LSST-like survey. In addition, the PCA approach automatically serves as a data compression method, enabling the retention of the majority of the cosmological information while reducing the dimensionality of the data vector by a factor of  $\sim 5$ .

## 2.2 Introduction

Weak gravitational lensing is the deflection of light by the gravitational field of large-scale structure, which leads to minute distortions of the observed galaxy images compared to their original shapes in the galaxy source plane. Measuring the correlation functions of the galaxy shapes is therefore a way to measure the growth of structure and the geometry of the Universe (e.g., Bartelmann & Schneider, 2001; Hoekstra & Jain, 2008; Kilbinger, 2015; Mandelbaum, 2018) and hence a promising avenue to constrain cosmology (Huff et al., 2014; Jee et al., 2016; Hildebrandt et al., 2018; Troxel et al., 2018a; Hikage et al., 2019b).

With the next generation of weak lensing surveys, such as the Large Synoptic Survey Telescope (LSST<sup>1</sup>; Ivezić et al. 2019), the Wide-Field Infrared Survey Telescope (WFIRST<sup>2</sup>) and *Euclid*<sup>3</sup>, we expect data sets that are both wider and deeper compared to current surveys (e.g. KiDS<sup>4</sup>, DES<sup>5</sup>, HSC<sup>6</sup>) in the near future. The high-quality data from upcoming lensing surveys is expected to reduce the statistical uncertainty in weak lensing measurements compared to current surveys by an order of magnitude. In order to fully exploit the cosmological constraining power of weak lensing surveys, much effort has been committed to understand observational systematic effects, such as uncertainties in the measurement of galaxy shapes (e.g., Massey et al., 2013; Mandelbaum et al., 2015) and photometric redshifts. On the astrophysical side, major progress has been made in modeling systematics that affect the interpretation of the weak lensing signal, such as the intrinsic alignment of galaxies (e.g., Mandelbaum et al., 2011; Joachimi et al., 2015; Troxel & Ishak, 2015; Krause et al., 2016), the nonlinear evolution of the dark matter density field (e.g., Takahashi et al., 2012; Heitmann et al., 2014), and baryonic effects that modify the latter (e.g., Eifler et al., 2015; Mead et al., 2015; Chisari et al., 2018).

However, inaccuracies in the last step of the analysis, the inference of cosmological parameters from measurements of observables such as  $\xi_{\pm}$  and some model for the likelihood function, are less well-explored in the weak lensing community. Uncertainties related to parameter space sampling, discrepancy metrics, and the likelihood of the summary statistics itself are important aspects of the cosmological interpretation that can lead to potential biases in the analysis.

While likelihood-free approaches such as Approximate Bayesian Computation (e.g., Akeret et al., 2015; Peel et al., 2017) are beginning to emerge as a tool for cosmological inference, most analyses still assume a likelihood function to transition from observations to cosmological parameters. Among the possible choices, the multivariate Gaussian likelihood function is the simplest and most commonly used.

Even though the Cosmic Microwave Background (CMB) temperature field is close to

---

<sup>1</sup><http://www.lsst.org/lstt>

<sup>2</sup><https://wfirst.gsfc.nasa.gov>

<sup>3</sup><http://sci.esa.int/euclid/>

<sup>4</sup><http://kids.strw.leidenuniv.nl>

<sup>5</sup><https://www.darkenergysurvey.org>

<sup>6</sup><http://hsc.mtk.nao.ac.jp/ssp/>

a Gaussian and the non-Gaussian features in CMB power spectra are small, the CMB data analyses (Hinshaw et al., 2013; Planck Collaboration et al., 2014a, 2016) have made some progress beyond this simple Gaussian assumption.

They use second-order statistics to capture the cosmological information in the underlying temperature and polarization field, which are (very close to) Gaussian. For idealized CMB observations (full-sky, isotropic beam, spatially uniform noise) the empirical power spectra of the underlying Gaussian temperature and polarization fields has a Wishart distribution given the model power spectra. At sufficiently high  $\ell$ , the likelihood function of power spectra approaches a multivariate Gaussian following the Central Limit Theorem. When going beyond the idealized case, the inclusion of potentially non-Gaussian foreground distributions such as galactic dust emission, Cosmic Infrared Background, and radio point sources breaks the initial assumption that the measured field is Gaussian and consequently breaks the conclusion that the likelihood of the power spectra beyond a certain  $\ell$  is well-approximated by a multivariate Gaussian. As further detailed in Planck Collaboration et al. (2016), the Planck analysis masks foreground contaminants and assumes that the non-Gaussian features are subdominant outside of the masked regions.

The situation is different for weak lensing. Due to non-linear structure evolution at late times, the shear field itself is non-Gaussian, which invalidates the premise of the CMB argument. Nevertheless, weak lensing analyses have assumed a multivariate Gaussian likelihood function to be the underlying distribution of shear two-point statistics (Fu et al., 2008; Semboloni et al., 2011; Huff et al., 2014; Hildebrandt et al., 2016; Troxel et al., 2018a). In these analyses, the non-Gaussianity of the shear field enters only via non-linear matter density power spectra (Takahashi et al., 2012; Heitmann et al., 2014) that are used to model the observed two-point statistics and via so-called non-Gaussian covariances, which indicate that a non-vanishing four-point function is included in the covariance computation.

Non-Gaussian shear covariances due to non-linear clustering have been studied in Takada & Jain (2009); Sato et al. (2009, 2011b); Harnois-Déraps & van Waerbeke (2015). These studies use the multivariate Gaussian shear likelihood with the contribution of covariance from non-Gaussian fields included. The impact on cosmological constraints depends on the scales considered and on depth and area of the survey. Computing non-Gaussian covariances is essential in cosmic shear analyses, and most recent cosmic shear measurements have opted for different strategies, such as analytical computation of Gaussian and non-Gaussian covariances (e.g., Jee et al., 2016; Hildebrandt et al., 2017; Krause et al., 2017), or for covariance estimated through numerical simulations (e.g. Heymans et al., 2013). For future surveys, several studies indicate the high computational costs of a brute-force numerical simulations approach (Dodelson & Schneider, 2013; Taylor & Joachimi, 2014), which led to the development of new covariance estimators (Joachimi, 2017; Friedrich & Eifler, 2018). For cosmic shear, Barreira et al. (2018) have shown that the Gaussian covariance plus Super Sample Covariance terms are sufficient, both of which can be easily implemented analytically.

When going beyond the Gaussian likelihood function for the convergence power spectrum, previous studies have considered a lognormal distribution and the copula method

for describing the non-Gaussian distribution (Taruya et al., 2002; Hilbert et al., 2011; Sato et al., 2011a, 2010). In configuration space, Hartlap et al. (2009) revisited the assumption of Gaussianity for the two-point correlation function and tested the non-Gaussianity with independent component analysis (ICA). The authors measure the distribution of the correlation function in 9600 realizations of ray-tracing simulations resembling the Chandra Deep Field South lensing analysis and perform three full likelihood analyses: using ICA, a standard multivariate Gaussian with a ray-tracing covariance, and with a Gaussian covariance. This paper triggered several attempts to build analytical expressions for the likelihood function of the shear two-point correlation function that improve over the standard Multivariate Gaussian approximation. For example, Schneider & Hartlap (2009) have shown that two-point correlation functions of Gaussian fields cannot take arbitrary values since this would violate the constraint of non-negativity of power spectrum. As a consequence the sampling distribution of the correlation functions cannot be an exact multivariate Gaussian. Keitel & Schneider (2011) employ Fourier mode expansion and characteristic functions of a Gaussian random field to derive an analytical expression for the likelihood function of its uni- and bi-variate correlation functions. In Wilking & Schneider (2013); Wilking et al. (2015), the authors transform the correlation functions such that a quasi-Gaussian approximation of the likelihood function is justified and tested its performance with simulations. A recent paper by Sellentin & Heavens (2018) explored the high-order correlations between the data points of the CFHTLenS cosmic shear correlation functions in search for non-Gaussianity. In Sellentin et al. (2018), the authors measure the skewed distributions of weak lensing shear correlation functions in simulations and follow the CMB literature in developing an analytical expression for the correlation function likelihood.

It is well-established in the literature that the sampling distribution of shear two-point correlation functions is not strictly Gaussian. Despite this, the Gaussian likelihood model is still the standard in weak lensing likelihood analysis for current surveys. If the likelihood is not Gaussian, then analysing the data with a Gaussian assumption could bias the cosmological parameter constraints. It is therefore important to quantify potential cosmological parameter biases that may arise due to un-modeled aspects of the likelihood function, so as to determine whether an alternative way of modeling the likelihood function is needed. This effect was discussed in the literature (e.g. Hartlap et al. 2009; Sellentin et al. 2018; Taylor et al. 2019), but agreement on whether the Gaussian approximation cause a significant bias for cosmological weak lensing analyses has not yet been reached. It is difficult to fully demonstrate the impact of the non-Gaussian likelihood on cosmological parameter constraints. Part of the difficulty comes from the dimensionality of the problem, with reconstructing in detail the full high-dimensional non-Gaussian likelihood currently being an unsolved problem. In this work, we do not address the full problem, but make some advance upon previous work by modelling the non-Gaussian likelihood with a large ensemble of weak lensing simulations, and further estimate the biases on cosmological parameters.

This paper is structured as follows: In Sect. 4.4, we describe the details of our simulated weak lensing data. In Sect. 2.4, after showing how the data vectors are modeled theoretically, we describe the likelihood analysis and likelihood models. Section 2.5 expands on

the systematic approach of assessing the performance of likelihood models and data compression. In Sect. 2.6, we show the results for non-Gaussianity of weak lensing observables and the impact on cosmological parameter estimates. Section 2.7 contains our discussion and conclusions.

## 2.3 Simulations

The simulated (mock) weak lensing data that are used in this paper are based on the Scinet Light Cone Simulations<sup>7</sup> (Harnois-Déraps & van Waerbeke, 2015; Harnois-Déraps et al., 2018, SLICS hereafter), which are specifically tailored for statistical studies of weak lensing analyses. They consist of a series of lines-of-sight (LOS) of 100 deg<sup>2</sup> each, constructed by ray-tracing in their own independent realization. In the simulations, no masks are applied, and hence the patches are 10 by 10 deg<sup>2</sup>. The underlying  $N$ -body simulations evolved 1536<sup>3</sup> dark matter particles in a box length of 505  $h^{-1}$ Mpc, and produced 18 mass planes between redshift 0.0 and 3.0, which are then converted into shear maps using the Born approximation. We used 932 such independent realizations in which the initial random seeds changed prior to the  $N$ -body run, with the assumed cosmology fixed to that of WMAP9+SN+BAO flat  $\Lambda$ CDM cosmology (Hinshaw et al., 2013):  $\Omega_m = 0.2905$ ,  $\Omega_\Lambda = 0.7095$ ,  $\Omega_b = 0.0473$ ,  $h = 0.6898$ ,  $\sigma_8 = 0.826$  and  $n_s = 0.969$ .

The mock galaxy catalogues are then created in a way that is meant to reproduce the redshift distributions of weak lensing source galaxies in LSST (Chang et al., 2013):

$$n(z) \propto z^\alpha \exp \left[ - \left( \frac{z}{z_0} \right)^\beta \right], \quad (2.1)$$

with  $\{\alpha, \beta, z_0\} = \{1.21, 1.0, 0.5\}$ , assuming a source number density of 26 gal/arcmin<sup>2</sup>. The mocks are split in 10 tomographic redshift bins  $n_i(z)$ , each containing the same number of galaxies. These distributions are further smoothed by a Gaussian kernel of width  $\sigma = (1+z)\sigma_z$  and  $\sigma_z = 0.02$ . For the detailed redshift distributions of the 10 LSST-like source bins, see Fig. A1 in Harnois-Déraps et al. (2018).

Besides the cosmological shear  $\gamma$  (see Sect. 2.4.1), the observed ellipticity  $\epsilon^{\text{obs}}$  includes the intrinsic shapes of galaxies  $\epsilon_{\text{int}}$  through the shear addition formula:

$$\epsilon^{\text{obs}} = \frac{\gamma + \epsilon_{\text{int}}}{1 + \gamma \epsilon_{\text{int}}^*}. \quad (2.2)$$

In the above expression, shear and ellipticities are written as complex variables, and the  $\epsilon_{1/2}^{\text{obs}}$  components are recovered from the real and imaginary parts respectively. The two components of the intrinsic galaxy shapes are each drawn from a Gaussian distribution with zero mean, a standard deviation of 0.29 inspired by the KiDS-450 (Hildebrandt et al., 2016) data, and the constraint that  $|\epsilon_{\text{int}}|^2 \leq 1$ .

---

<sup>7</sup><https://slics.roe.ac.uk/>

In our analysis, we measure the non-Gaussian shapes of the likelihood function of shear correlation functions with the SLICS simulation, and then use them to estimate the parameter biases for large weak lensing surveys such as LSST. Due to the differences in survey areas, different sources of uncertainty dominate: the 100 deg<sup>2</sup> SLICS simulations are shape-noise dominated while LSST data will be cosmic-variance dominated. Therefore, in this work we sometimes switch off the intrinsic shape noise in order to separately understand the contributions of cosmic variance and shape noise to the shape of the likelihood function. For all results that are presented, we refer to the results as ‘without shape noise’ or ‘with shape noise’.

## 2.4 Method

### 2.4.1 Cosmic shear correlation function data vector

The weak lensing effect is mathematically approximated as a linear transformation that maps the unlensed location to the lensed location. The transformation matrix  $\mathcal{A}$ , which connects the shape of a source with the observed images, can be written as

$$\mathcal{A} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}, \quad (2.3)$$

where  $\kappa$  is the convergence and  $\gamma_1, \gamma_2$  are the two components of the spin-two shear  $\gamma = \gamma_1 + i\gamma_2$  in Cartesian coordinates. Conventionally the coordinate system is rotated so that the separation vector is parallel to the  $x$ -axis. The shear components are decomposed into the tangential direction ( $\gamma_t$ ) and the cross direction ( $\gamma_\times$ ) in the rotated coordinates:

$$\gamma_t = -\mathbf{Re}(\gamma e^{-2i\phi}), \gamma_\times = -\mathbf{Im}(\gamma e^{-2i\phi}), \quad (2.4)$$

where  $\phi$  is the polar angle of the separation vector of the two galaxies. From the shear components we can write the two shear correlation functions as a function of angular separation

$$\xi_{\pm}^{ij}(\theta) = \langle \gamma_t^i \gamma_t^j \rangle(\theta) \pm \langle \gamma_\times^i \gamma_\times^j \rangle(\theta). \quad (2.5)$$

Here  $i$  and  $j$  are indices of tomographic redshift bins, and the angle brackets refer to ensemble average. The correlation functions are computed from the SLICS simulations with the package TREECORR<sup>8</sup> (Jarvis et al., 2004).

The angular bins for measuring shear correlation functions initially divide the scales from 0.32 to 400 arcmin into logarithmically spaced bins of width  $\Delta(\ln\theta) = 0.23$ ; however, we further apply angular selections to minimize the impact of known limitations in the simulations (Harnois-Déraps & van Waerbeke, 2015). We require  $\theta > 0.8$  arcmin for  $\xi_+$ , to avoid resolution effects in the simulations at small scales. Since  $\xi_-$  is more sensitive to small-scale uncertainties, we apply a more aggressive constraint:  $\theta > 6.5$  arcmin. We

---

<sup>8</sup><https://github.com/rmjarvis/TreeCorr>

also require  $\theta < 160$  arcmin for  $\xi_+$  to avoid scales with significant power loss due to the box size. On the other hand,  $\xi_-$  is less affected by the box size effect within 400 arcmin; therefore we do not introduce extra constraints on the large scale for  $\xi_-$ .

These scales are similar to those used in the recent KiDS-450 cosmic shear analysis by Hildebrandt et al. (2017), and presume that the analysis has separately accounted for uncertainties at small scales due to baryon feedback, modeling of the non-linear power spectrum and of the intrinsic alignments of galaxies<sup>9</sup>. Modeling these uncertainties at the precision required by LSST will be challenging. In the recent DES analysis (Troxel et al., 2017) conservative scale cuts were implemented in order to avoid biases due to imperfect modeling of said astrophysical effects. For this paper, we assume an optimistic scenario in which the LSST weak lensing measurement pipeline includes these small angular scales.

The complete data vector is the concatenation of all  $\xi_+$  and  $\xi_-$  values across all tomographic bins and  $\theta$  bins. Each simulated realization has 55 correlation functions, each with 42 angular bins (24 for  $\xi_+$  and 18 for  $\xi_-$ ).

The estimator of the data covariance from the many simulation ensemble is defined as

$$\hat{C}^{ij} = \frac{1}{\nu} \sum_k^{N_s} (\xi_k^{ij} - \bar{\xi}^{ij})(\xi_k^{ij} - \bar{\xi}^{ij})^T, \quad (2.6)$$

where  $i$  and  $j$  indicate the tomographic redshift bins,  $\xi_k$  is the data vector of the  $k$ -th realization,  $\bar{\xi}$  is the mean data vector across all simulated realizations, and  $\nu = N_s - 1$  is the number of degrees of freedom given that the mean is estimated from the data. If the number of data-points,  $N_d$ , exceeds the number of realizations,  $N_s$ , we can neither ensure that the data covariance matrix is positive definite nor control the error in the data covariance matrix and its inverse. Therefore, we rebinned the data vector to reduce the number of points from 2310 to 770 by combining the angular bins in groups of three. Our final  $\theta$  binning was chosen such that  $N_s = 932$ ,  $N_d = 770$  and hence  $N_d < N_s$ .

An illustration of the data vector for a particular set of tomographic bins, the diagonal covariance matrix elements, and a comparison with analytic theory predictions is given in Fig. 2.1. The theoretical predictions are based on the HALOFIT method (Takahashi et al., 2012), which models the nonlinear power spectrum with a fitting function. The HALOFIT model agrees well in general with the measurements from the simulations on the scales of interest, but there are still percent-level errors compared to the simulations (Harnois-Déraps & van Waerbeke, 2015). The finite box effect of the simulations also introduces power drops with respect to the theoretical predictions (Harnois-Déraps & van Waerbeke, 2015). The mismatch between the simulations and the theory will be compensated by a correction factor

$$\xi_{\text{theory}}^{ij}(\theta) / \langle \xi_{\text{mock}}^{ij}(\theta) \rangle. \quad (2.7)$$

The correction factor above is not intended to fix these or other limitations of the simulations (which have some simplifications compared to reality, e.g., lack of baryonic physics).

---

<sup>9</sup>Note that these effects could contribute to the level of non-Gaussianity in the data, but that is beyond the scope of this paper.

Rather, the correction is applied in order to assure that any bias we observe in our likelihood analysis comes from the covariance instead of the very small but nonzero intrinsic mismatch between the simulations and the theory. Note that the correction factor is applied before the computation of the covariance, skewness and kurtosis, and is therefore included in these estimators.

## 2.4.2 Modeling of observables

For the simulated likelihood analyses in Sect. 2.6 we employ the COSMOLIKE analysis and forecasting software package<sup>10</sup>. COSMOLIKE has been used in several forecasts exploring joint analyses of multiple cosmological probes (Eifler et al., 2014; Krause & Eifler, 2017; Schaan et al., 2017) and systematics mitigation strategies, such as the impact of baryons and intrinsic alignment (Eifler et al., 2015; Krause et al., 2016). On the observational side, the code was used in a weak lensing analysis of Sloan Digital Sky Data (Huff et al., 2014), the analysis of DES science verification data (Becker et al., 2016) and the recent DES Year 1 analysis (Krause et al., 2017; Troxel et al., 2018a; Abbott et al., 2018).

We compute the linear power spectrum of the best fitting flat  $\Lambda$ CDM cosmology for WMAP9 + BAO + SN (Hinshaw et al., 2013) using the Eisenstein & Hu (1999) transfer function and model the non-linear evolution of the density field as described in Takahashi et al. (2012). From the density power spectrum  $P_\delta(k, z)$ , we compute the shear power spectrum using the Limber and the flat sky approximations as

$$C_{\kappa\kappa}^{ij}(l) = \frac{9H_0^{42}}{4c^4} \int_0^{\chi_h} d\chi \frac{g^i(\chi)g^j(\chi)}{a^2(\chi)} \left( \frac{l}{f_K(\chi)}, \chi \right), \quad (2.8)$$

with  $l$  being the 2D wave vector perpendicular to the line of sight,  $\chi$  denoting the comoving coordinate,  $\chi_h$  is the comoving distance to the horizon,  $a(\chi)$  is the scale factor, and  $f_K(\chi)$  the comoving angular diameter distance (throughout set to  $\chi$  since we assume a flat Universe). The lens efficiency  $g^i$  is defined as an integral over the redshift distribution of source galaxies  $n(\chi(z))$  in the  $i^{\text{th}}$  tomographic interval

$$g^i(\chi) = \int_\chi^{\chi_h} d\chi' n^i(\chi') \frac{f_K(\chi' - \chi)}{f_K(\chi')}. \quad (2.9)$$

We compute the cosmic shear two-point functions  $\xi_\pm$  using the flat-sky approximation

$$\xi_\pm^{ij}(\theta) = \int \frac{dl l}{2\pi} J_{0/4}(l\theta) C_{\kappa\kappa}^{ij}(l), \quad (2.10)$$

with  $J_n(x)$  the  $n$ -th order Bessel function of the first kind.

---

<sup>10</sup>[www.cosmolike.info](http://www.cosmolike.info)

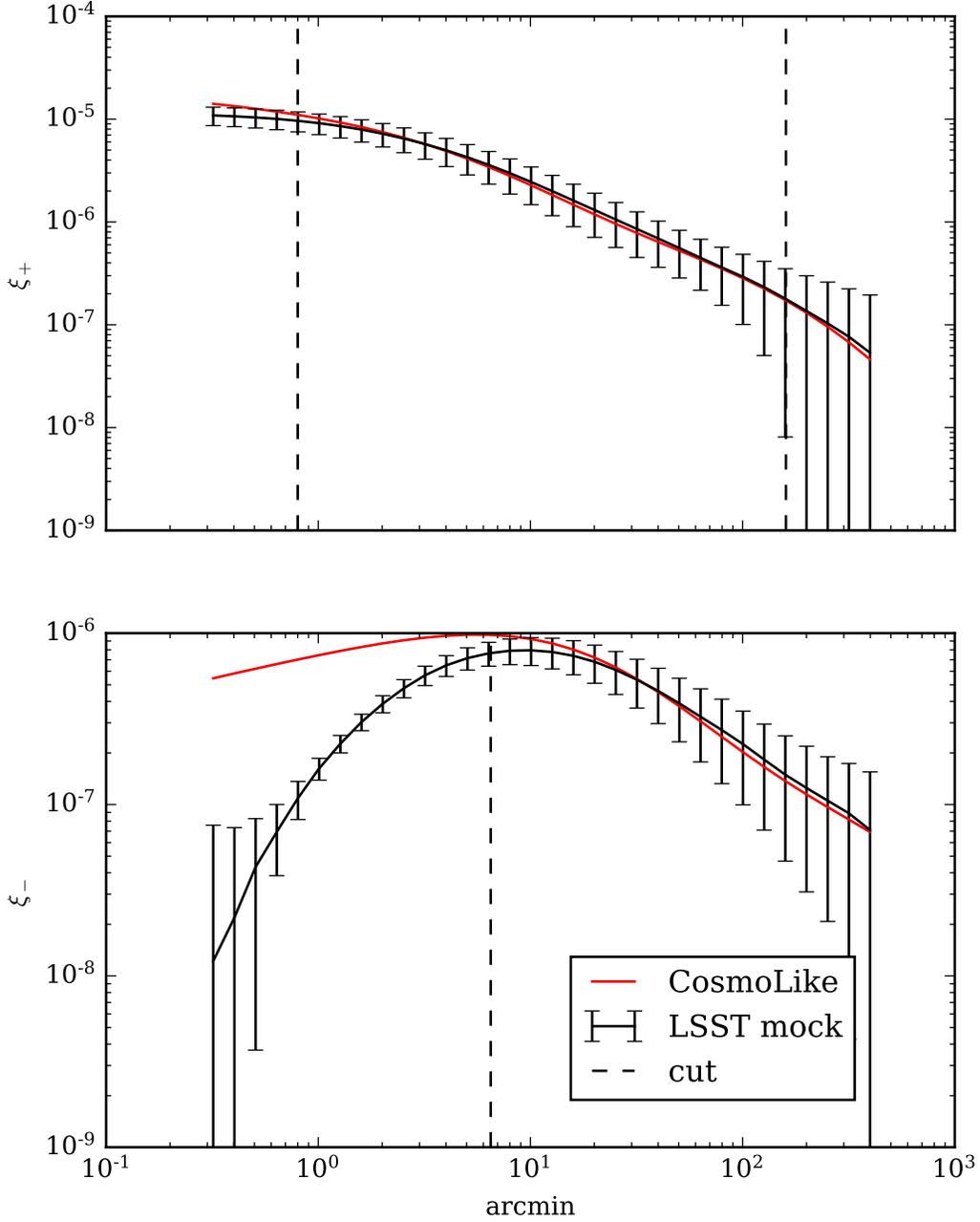


Figure 2.1: Correlation functions  $\xi_+$  (top) and  $\xi_-$  (bottom) for the auto-correlation of the first tomographic bin  $(z_1, z_2) = (1, 1)$  of our mocks with shape noise and theoretical predictions from COSMOLIKE. The solid black curve shows the average and the 16% and 84% percentiles of the 932 mock realizations due to cosmic variance only. The adopted  $\theta$  ranges for  $\xi_{\pm}$  are shown by the vertical dashed lines, which were chosen due to limitations in the simulations as described in the text. The mismatch between the two curves is captured by the ratio  $\xi_{\text{theory}}^{ij}(\theta) / \langle \xi_{\text{smock}}^{ij}(\theta) \rangle$ .

### 2.4.3 Likelihood functions and data covariance matrix

From Bayes' theorem, we can compute the posterior distribution of the cosmological parameters. In the standard likelihood analysis, the likelihood function is parametrized as a multivariate Gaussian function

$$L(\vec{\xi} | \vec{\pi}) \propto \exp\left[-\frac{1}{2}(\vec{\xi} - \vec{\xi}_\pi)^T \mathbf{C}^{-1}(\vec{\xi} - \vec{\xi}_\pi)\right], \quad (2.11)$$

where  $\vec{\xi}$ ,  $\vec{\pi}$  and  $\mathbf{C}$  denote the data vector, cosmological parameters and the covariance matrix respectively. The covariance matrix is fixed throughout the analysis.

Analytically computed covariance matrices are noise-free and can be factorized or inverted without complications. However, our covariance matrix estimated from numerical simulations is inherently noisy, and the noise level is affected by the number of realizations ( $N_s = 932$  in our case) and the size of the data vector ( $N_d = 770$ ).

To compute the likelihood function in the form of Eq. (2.11), we need the inverse covariance matrix  $\Psi$ , also called precision matrix. An unbiased estimator of the precision matrix is given by (Anderson, 2003; Hartlap et al., 2007; Taylor et al., 2013)

$$\hat{\Psi} = \frac{\nu - N_d - 1}{\nu} (\hat{\mathbf{C}})^{-1} \quad (2.12)$$

in the case that the noise is Gaussian-distributed and the data points are statistically independent with  $\nu = N_s - 1$ .

If the observed data is drawn from a multivariate Gaussian, Sellentin & Heavens (2016) show that after marginalizing over the noisy covariance matrix, the likelihood measured from the simulation realizations follows a multivariate t-distribution rather than a multivariate Gaussian. An earlier correction proposed by Hartlap et al. (2007) uses the unbiased inverse covariance matrix and keeps the multivariate Gaussian distribution to construct the likelihood function when the underlying distribution is Gaussian and the covariance is noisy. Thus, the Hartlap's correction method leads to

$$L(\vec{\xi} | \vec{\pi}) \propto \exp\left[-\frac{1}{2}(\vec{\xi} - \vec{\xi}_\pi)^T \hat{\Psi}(\vec{\xi} - \vec{\xi}_\pi)\right]. \quad (2.13)$$

The Sellentin-Heavens likelihood shows improvement over Hartlap's correction in terms of parameter inference with the marginalization over noise in covariance matrices. But it is still unclear how to extend the t-distributed likelihood function to cases where the underlying distribution of the data is non-Gaussian. In this paper, we adopt Eq. (2.12) instead, since it is easier to apply to near-Gaussian distributions. Since we quantify the difference between the Gaussian likelihood and the non-Gaussian likelihood through bias in cosmological parameter space, the bias depends more on the asymmetry of the likelihood distributions and thus is less sensitive to the difference between these two methods.

## 2.4.4 PCA transformation

For the shear correlation functions  $\xi_{\pm}$ , the multivariate Gaussian likelihood function with the form described in Eq. (2.11) is the most commonly used likelihood model in the literature. Since it is not trivial to build robust multivariate non-Gaussian likelihood functions, we perform the principal component analysis (PCA) transformations first on the data vector to remove the correlation between the data points. PCA is an orthogonal transformation that transforms data points into coordinates without the linear correlations. It can be described as:

$$\hat{\Lambda} = Q^T \hat{C} Q \quad (2.14)$$

with columns of the transformation matrix  $Q$  containing the eigenvectors of the covariance matrix  $\hat{C}$  estimated from the simulations. After the PCA transformation, the matrix  $\hat{\Lambda}$  is diagonal and the diagonal elements are the eigenvalues of the covariance matrix  $\hat{C}$ . The new coordinates are usually referred to as principal components, and the components are sorted according to decreasing variance. In addition, PCA automatically concentrates the information into a smaller number of principal components. For instance, Harnois-Déraps & Pen (2013) studied the noisy covariance matrix of the matter power spectrum and found that the information content of the leading principal components remains stable as they raised the noise level by reducing the number of measurements from 200 to only 4.

Note that in Eq. (2.14) we use the covariance instead of its inverse in the PCA. It may sound counterintuitive that the data points with the highest variance, i.e. the highest uncertainty, contain more information. However, the principal components with highest variance are also those with the largest signal-to-noise ratio in our data (this was also found in Harnois-Déraps & Pen (2013)), and thus contain more information. To quantify the information content on the training data contained in the first  $N$  components, we perform a Markov Chain Monte Carlo (MCMC) analysis on the mock data with different numbers of principal components, and define the information content as the square root of the determinant of the inverse parameter covariance (Fisher matrix) of  $\Omega_m$  and  $\sigma_8$ ,  $\sqrt{\det(\mathbf{F})}$ . Other parameters are fixed throughout the MCMC and a flat prior that limits parameters in the ranges  $0.05 < \Omega_m < 0.6$  and  $0.5 < \sigma_8 < 1.1$  is assumed. Figure 2.2 demonstrates the relation between the information content retained and the number of principal components used. It shows that if we apply data compression such that we have 80% fewer data points, we lose only 23% of the information content.

## 2.4.5 Multivariate likelihood models

After the PCA transformation, the data points are linearly uncorrelated. We next continue to use parametric and non-parametric functions to describe the likelihood distributions in the PCA space.

In this paper we consider the following likelihood models for the PCA coordinates:

- Gaussian function

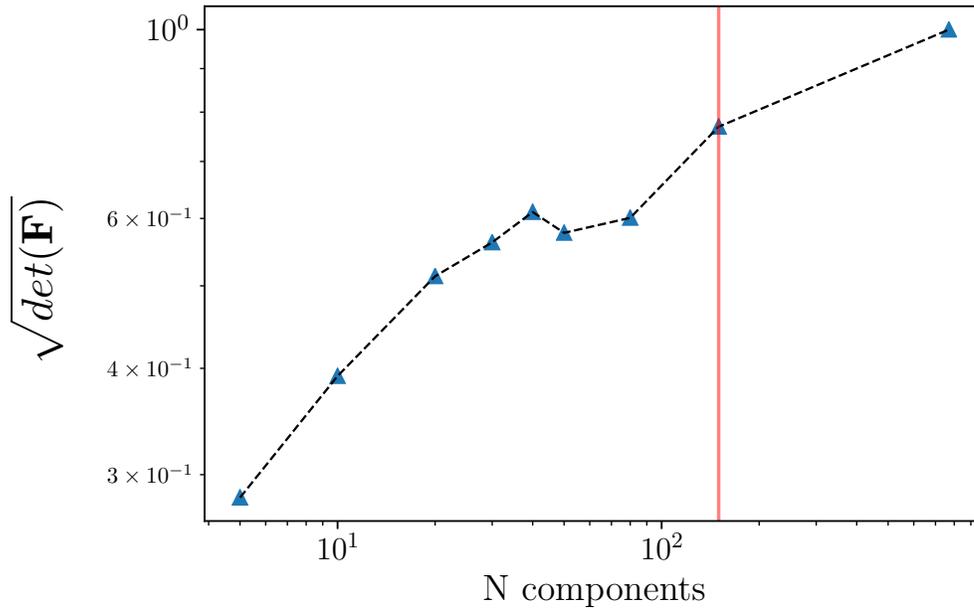


Figure 2.2: Information content of the training data versus number of principal components used in data analysis: the square root of the determinant of the inverse parameter covariance ( $\Omega_m$  and  $\sigma_8$  only), i.e. the Fisher information matrix, quantifies the amount of retained information about the cosmological parameters. In the plot, the metric is normalized by the total information content (770 components). By keeping only 20% of the data points (the red vertical line), we lose 23% of information.

- non-Gaussian Edgeworth function
- $k$ -nearest neighbors
- spectral series

Here we describe the two parametric and the two non-parametric models listed above. The performance and results of the models are covered in more details in Sects. 2.5 and 2.6.

Under the assumption that the underlying likelihood function measured in simulations is close to a multivariate Gaussian function, we approximate the multivariate likelihood function in the PCA coordinates as a product of parametric “marginal distribution functions”. The independence of PCA components is a strong assumption, but it is the assumption that the standard multivariate Gaussian likelihood makes. The product of one-dimensional Gaussian distributions of PCA components is identical to the multivariate Gaussian likelihood.

The Edgeworth function is a Gaussian function multiplied by correction terms constructed by its cumulants. It serves as an improvement upon the Gaussian likelihood. In this paper, we adopt Petrov’s formula of the Edgeworth expansion (Blinnikov & Moessner, 1998; Petrov, 1962) and the coefficients in the expansion are fixed by the cumulants of the simulation data. In the case where the standard deviation  $\sigma=1$ , the first four terms in the expansion are

$$\begin{aligned} \text{Edgeworth}(x) = & \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(\Delta x)^2}{2}\right] \\ & \cdot \left[1 + \frac{\kappa_3}{3!} He_3(\Delta x) + \frac{\kappa_4}{4!} He_4(\Delta x) \right. \\ & \left. + \frac{10\kappa_3^2}{6!} He_6(\Delta x) + \dots\right], \end{aligned} \quad (2.15)$$

where  $He_n(x)$  are Hermite polynomials given by

$$He_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}, \quad (2.16)$$

$\Delta x = x - \mu$  and  $\kappa_n$  are cumulants. Moments and cumulants are two different sets of quantities that can summarize a distribution. Cumulants arise naturally from Fourier transformation. In the Fourier transformation, the probability density function  $f(x)$  is transformed into

$$\tilde{f}(k) = \int_{-\infty}^{\infty} e^{ikx} f(x) dx. \quad (2.17)$$

The cumulants are then defined as the coefficients of the power series expansion

$$\ln \tilde{f}(k) = \sum_{n=1}^{\infty} \kappa_n \frac{(ik)^n}{n!}. \quad (2.18)$$

Since the Edgeworth function is not guaranteed to be positive and could have oscillatory behavior, one should be careful with anomalies (negative probabilities) and avoid its use in strongly non-Gaussian cases. In our case, we do not find negative possibilities when modeling the marginal distributions of PCs with the Edgeworth expansion. We show in Fig. 2.3 an example of the marginal one-dimensional distributions of  $\xi_+$  and the two parametric models.

In addition to these parametric models, we also model the likelihoods more generally by estimating the high-dimensional density ratio  $\beta(x) = f(x)/g(x)$  non-parametrically. Unlike the parametric methods that approximate the multivariate distributions as products of marginal distributions, our non-parametric methods do not assume independence. In our case, we take  $g(x)$  to be the Gaussian model. Once  $\beta(x)$  is fitted, we can sample from the estimated density  $f(x)$  by importance sampling with  $g(x)$  as the proposal distribution. In this paper, the density ratio is estimated by non-parametric methods based on the k-nearest-neighbors kernel density estimator (Lincheng & Zhijun, 1985) and the Spectral Series estimators. The k-nearest-neighbors estimator (knn) approximates the density at a point by a kernel smoothing applied to the k nearest neighbors of that point, and the Spectral Series estimator (Izbicki et al., 2014) combines orthogonal series expansion and adaptively chosen bases to construct non-parametric likelihood functions. Let  $\{\psi_i\}$  be an orthonormal basis with respect to the data distribution; then the density estimator for spectral analysis has the form

$$\beta(x) = \sum \alpha_i \psi_i(x). \quad (2.19)$$

In low-dimensional non-parametric curve estimation, the basis is fixed to the usual choices, such as a Fourier basis. In the Spectral Series method, the basis is driven by data so as to capture the intrinsic dimensionality of the data (Izbicki et al., 2014). With the density ratio, we are improving the Gaussian likelihood model using non-parametric methods. This is different from measuring the high-dimensional density with non-parametric models directly. Since the non-parametric models have more degrees of freedom compared to parametric models and make no assumption on the likelihood distributions, including the non-parametric methods make our list of models more complete.

## 2.4.6 Skewness and Kurtosis

We examine particular departures from Gaussianity of the shear two-point correlation function by calculating higher moments of the distributions: skewness and kurtosis. The skewness can be quantified as the normalized expectation value of the third central moment

$$\text{Skew}[X] = \frac{\mathbb{E}[(X - \mu)^3]}{\sigma^3} \quad (2.20)$$

and it measures the asymmetry of the distribution. Increasing the number of samples  $N_s$  does not reduce the skewness, but rather reduces the uncertainty of the skewness estimate, so repeated measurements would not remove the non-Gaussianity. Gaussian functions are symmetric; their skewness is zero. The distribution of shear correlation functions, however,

is not perfectly symmetric. In Appendix 2.8.1 we derive for Gaussian fields the general expressions for the third moment of the likelihood of the shear correlation functions, from which the skewness can be predicted. Following the same derivation in Appendix 2.8.1, the  $n^{\text{th}}$  moment scales as  $1/f_{\text{sky}}^{n-1}$  in general. Besides cosmic variance, the effect of shape noise can also be included in this framework. We show that for the scales much smaller than the survey size, the third moment decreases with the survey size as  $f_{\text{sky}}^{-2}$  and hence the skewness as defined in Eq. (2.20) decreases as  $f_{\text{sky}}^{-1/2}$  ( $\sigma \propto f_{\text{sky}}^{-1/2}$ ). As the scale  $\theta$  approaches the survey window size, the third moment rises faster than  $\sigma^3$  and thus the skewness will increase. This trend is consistent with expectations from the Central Limit Theorem, which explains a decreasing skewness from an increase in survey area through the fact that the number of modes that are averaged over within given bin increases.

Besides the skewness, the asymmetry in the distribution is also captured by the mean-mode difference. In Appendix 2.9, we limit the possible range of the mean-mode difference by assuming a unimodal distribution. For larger surveys, the mean-mode difference in terms of  $\sigma$  also follows the same scaling relation as the skewness,  $(\tilde{\xi} - \bar{\xi})/\sigma \propto f_{\text{sky}}^{-1/2}$ .

Additionally, we measure the kurtosis of the likelihood function as a metric for the level of non-Gaussianity. The kurtosis is defined as:

$$\text{Kurt}[X] = \frac{\mathbb{E}[(X - \mu)^4]}{\sigma^4} - 3 \quad (2.21)$$

measures the symmetric outliers of the distribution. Since the fourth moment of the standard normal distribution equals 3, the kurtosis (or more precisely, the excess kurtosis) is defined as the normalized fourth moment minus 3.

## 2.5 Likelihood model assessment

The goal of this section is to introduce the statistical tools that we use to construct and assess the one-dimensional and multidimensional likelihood models in PCA coordinates. To properly compare different likelihood models and avoid overfitting, we use 10-fold cross-validation. That is, the 932 realizations are partitioned into 10 non-overlapping subsets. For each experiment, 9 of the subsets are used to compute the mean and covariance of the Gaussian model or used to train the non-parametric models, while the remaining subset is used for testing. After the models are built, we draw samples from the models (90 realizations for each sample) and compare these samples to the test samples using different two-sample tests and distance metrics. This process is then repeated 10 times with each of the 10 data sets used once as the testing data.

In order to quantify the performance of the one-dimensional models for the likelihood of the data in PCA space, we perform cross-validation as described above with univariate Kolmogorov-Smirnov (KS) tests. The two-sample KS test statistic is the maximum distance between two empirical cumulative distribution functions. The p-value of the two-sample KS test statistic is defined under the null hypothesis that the two samples are drawn

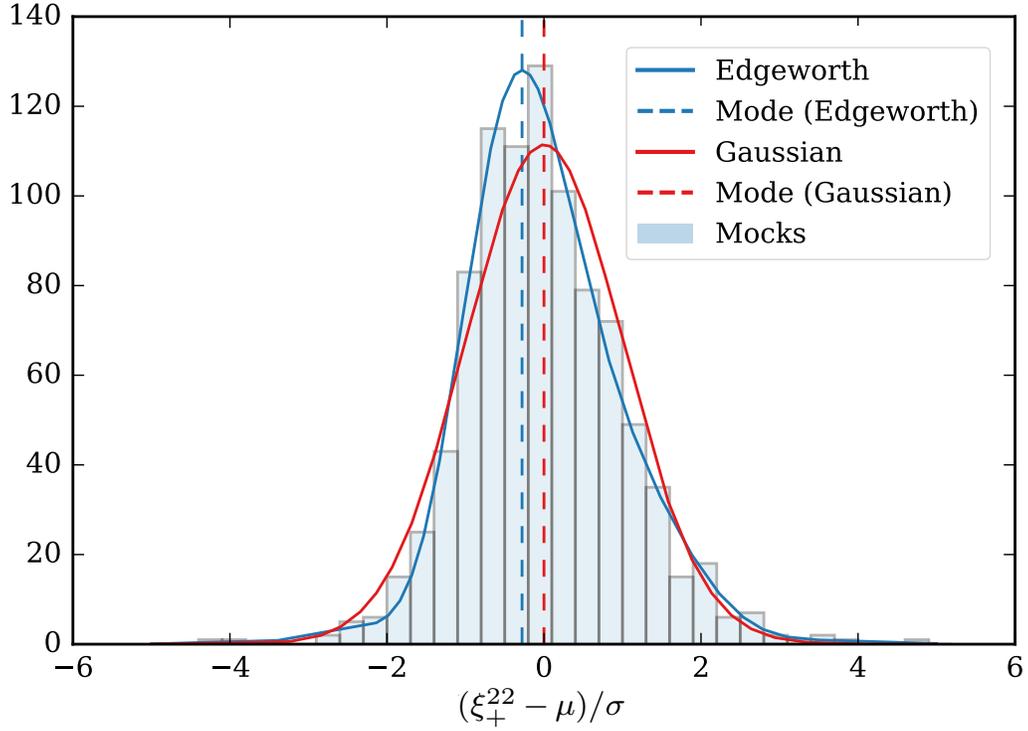


Figure 2.3: Example of the non-Gaussian distribution of  $\xi_+^{22}$  at  $\theta = 159$  arcmin in the mock weak lensing data with shape noise. This figure exhibits the low-level non-Gaussianity, and in particular the nonzero skewness, in the mock weak lensing data. If the distribution is skewed, the mean of the distribution deviates from the peak of the distribution (mode), which could lead to parameter biases if this feature of the likelihood is not adequately modeled.

from the same distribution. We take  $p < 0.05$  of this KS statistic as an indication that the null hypothesis is rejected and that the two sets of samples are not likely to have been drawn from the same distribution.

Besides tests on univariate distributions, we assess how close samples from the multivariate models are to test samples, using two (non-parametric) multivariate test statistics: the maximum mean discrepancy (MMD; Gretton et al. 2012) and the energy distance (ED; Székely & Rizzo 2004; Baringhaus & Franz 2004). The energy distance is a statistical distance between two probability distributions. It is defined as the square root of

$$D^2(p, q) = 2\mathbb{E}\|X - Y\| - \mathbb{E}\|X - X'\| - \mathbb{E}\|Y - Y'\| \quad (2.22)$$

where  $\mathbb{E}$  is the expectation value,  $X$ ,  $Y$ ,  $X'$ , and  $Y'$  are independent random vectors, the distribution of  $X$  and  $X'$  is  $p$ , and the distribution of  $Y$  and  $Y'$  is  $q$ . Here we use the Euclidean metric (and a sample estimate of the above expression). The maximum mean discrepancy can then be seen as a generalization of the energy distance to reproducing kernel Hilbert spaces. More specifically, we use a Gaussian kernel  $K_h(x, y)$  with bandwidth  $h$  to measure the similarity between two vectors  $x$  and  $y$ , and we define our MMD test statistic as the MMD sample estimate:

$$T = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^n K_h(X_i, X_j) - \frac{2}{mn} \sum_{i=1}^m \sum_{j=1}^n K_h(X_i, Y_j) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n K_h(Y_i, Y_j). \quad (2.23)$$

where  $X_1, \dots, X_m \sim p$  and  $Y_1, \dots, Y_n \sim q$ .

The two test statistics above provide quantitative distance measures between samples drawn from two sampling distributions of weak lensing correlations functions  $p$  and  $q$ . In the case of weak lensing analysis,  $X$  and  $Y$  in the above notations are multivariate shear correlation function data vectors drawn from the hold-out simulation data or the trained models. These test statistics or distance metrics<sup>11</sup> are invariant to orthogonal transformations (such as a PCA rotation) and often used for comparing higher-dimensional data.

## 2.6 Results

### 2.6.1 Biases due to non-Gaussian distributions of $\xi_{\pm}$

In the previous sections, we introduced the approach of building and assessing the likelihood distributions. Here we apply the tools to examine the biases due to non-Gaussianity in distributions of the shear correlation functions  $\xi_{\pm}$ .

---

<sup>11</sup>We use the publicly available R packages “kernlab” and “energy”. We implement MMD with the Gaussian kernel and determine the bandwidth by the median heuristic method, i.e.  $h^2 = \text{median}(\|X_i - X_j\|^2)$ .

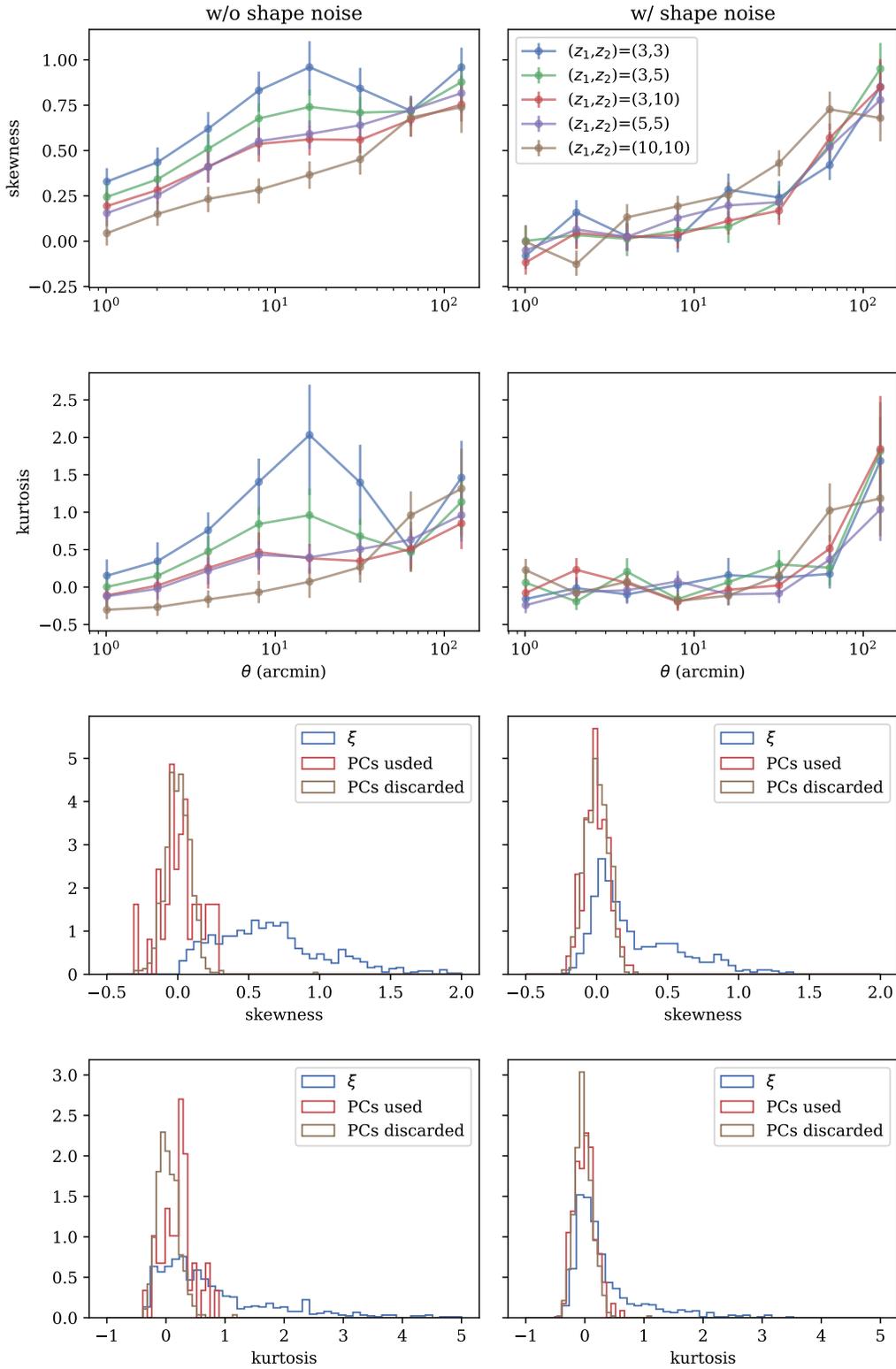


Figure 2.4: The top two rows show the skewness and kurtosis of  $\xi_+$  in tomographic bins  $(z_1, z_2) = (3,3), (3,5), (3,10), (5,5)$  and  $(10,10)$  for data without shape noise (left column) and data with shape noise (right column). The curves for data without shape noise exhibit strong non-zero skewness and kurtosis and demonstrate particular departures from Gaussianity in the marginal distributions. At lower redshift, the skewness and kurtosis are more significant. The non-Gaussianity is not as strongly detected in data with shape noise. That is because the mock data is dominated by shape noise in most of the scales that we

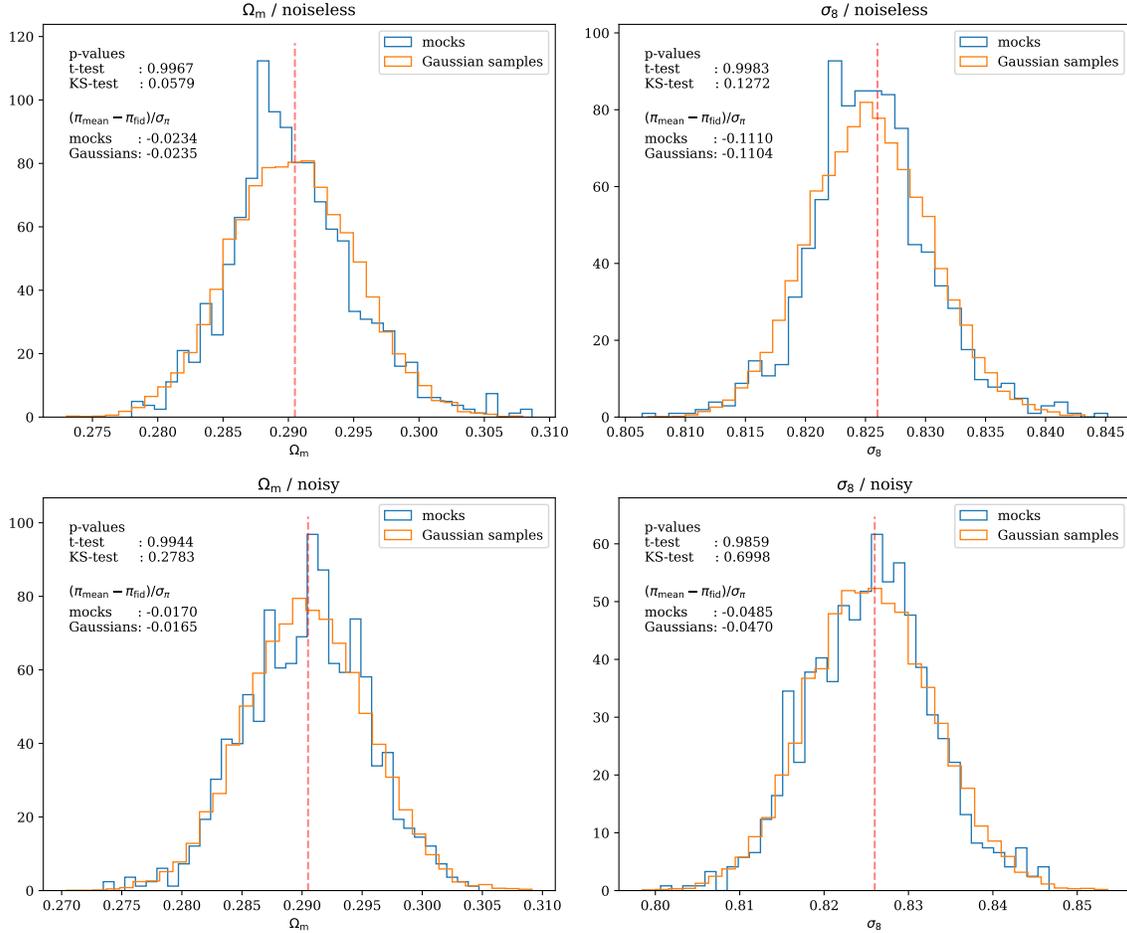


Figure 2.5: One-dimensional maximum likelihood estimates of  $\Omega_m$  and  $\sigma_8$  for data without shape noise (upper panels) and data with shape noise (lower panels). Besides the estimates based on the 932 realizations of correlation functions in the different simulation realizations (blue), the line in orange shows the parameter estimates of ten thousand Gaussian-distributed samples with the same mean and covariance matrix as the simulation realizations. By comparing the two histograms, we can estimate the impact of the Gaussian likelihood assumption with respect to parameter inference. The p-values of two-sample statistics, including the independent T-test and the KS-test, and the deviations of the mean values from the true point (red dotted line) are listed at the upper left corner of each panel. None of the statistics is below the threshold p-value of 5%, meaning that, for all cases, there is no significant difference between the two distributions and that there is no significant parameter bias due to the Gaussian likelihood assumption.

If the likelihood function of  $\xi_{\pm}$  is a multivariate Gaussian function, then its marginal distributions are Gaussian by construction. However, we detect non-Gaussianity in the marginal distributions of  $\xi_{\pm}$  for many values of  $\theta$  and tomographic bins. The finding that the likelihood of  $\xi_{\pm}$  is skewed was previously discussed in Sellentin & Heavens (2018). They use CFHTLenS Clone simulations for the tomographic analysis of the CFHTLenS data that provide 1656 semi-independent simulations for the 210 data points of CFHTLenS. For a specific data point at  $\theta = 35$  arcmin, they found that the most likely lensing amplitude is about 5% below the mean, so the distribution is ‘left-skewed’.

In Fig. 2.4 we show the non-zero skewness and kurtosis of the 1-D distributions of  $\xi_{\pm}$  in selected tomographic bins, and of PCA coordinates, for data with and without shape noise. For the shape noise-free data, the skewness and kurtosis both decrease as redshift increases. The magnitude of non-Gaussianity shown in the skewness and kurtosis is statistically significant and peaks roughly at  $\theta = 20$  arcmin. It is difficult to gain insight into the  $\theta$ -dependence of the skewness of the shear 2PCF, since the latter is an integral over the  $C(l)$  values with highly oscillating filter functions  $J_{0/4}$ .

Most of the scales that we consider are dominated by the shape noise, which strongly suppresses the skewness and kurtosis. At scales around 100 arcmin, the skewness and the kurtosis start to reach a comparable level as in the shape noise-free case, since the shape noise is relatively less important at large scales. Note that these results for skewness and kurtosis of the marginal 1D likelihoods of  $\xi_{\pm}$  values do not fully represent the level of non-Gaussianity in the multivariate observable space, since the  $\xi_{\pm}$  values are highly correlated across  $\theta$  and redshift bins.

Compared to  $\xi_{\pm}$ , the sampling distribution of the data in PCA space has far less significant skewness and kurtosis. The PCA components are linear combinations of  $\xi_{\pm}$  values. During the transformation, a large number of the  $\xi_{\pm}$  data points are added with positive and negative weights. As a result the linear combination then has reduced skewness and is closer to Gaussian. This is shown in the last two rows of Fig. 2.4, which compares the level of skewness and kurtosis in the marginal 1D distribution of  $\xi$  and in PCA components. We again note that the multivariate sampling distribution is not affected by the PCA transformation, the marginal distributions can change due to rotations and the resulting marginal likelihoods depend on the specific forms of rotations. After the PCA rotation, we find that the 1-D likelihoods empirically become more Gaussian, as illustrated in the last two rows of Fig. 2.4.

To estimate the biases in cosmological parameter due to the non-Gaussianity in distributions of  $\xi_{\pm}$ , we apply the maximum likelihood method. Maximum likelihood estimation provides an intuitive way of estimating the biases in cosmological parameters due to a failure to model the non-Gaussian distributions of weak lensing two point functions. In this method, we compute the maximum likelihood estimate of cosmological parameters of each realization in the mocks with the Gaussian covariance. Given our 932 simulations, this yields 932 MLEs in cosmological parameter space. Limited by the number of realizations, we perform the maximum likelihood estimation for individual parameters, either  $\Omega_m$  or  $\sigma_8$  (separately), with the other parameters fixed to their true values in the simulations.

To avoid a bias due to a small mismatch between the mocks and the theoretical prediction (see Fig. 2.1), the elements of data vectors for each simulation realization are rescaled by ratios  $\xi_{\text{theory-fid}}^{ij}(\theta)/\langle\xi_{\text{mock}}^{ij}(\theta)\rangle$ .

In order to estimate the impact of incorrect likelihood models, we also perform the maximum likelihood estimation on 10,000 Gaussian-distributed samples of two-point correlation functions. We first model the likelihood distributions of the simulation realizations with a multivariate Gaussian, and then draw samples from them. The samples share the same mean value and covariance matrix as the mocks but follow a multivariate Gaussian distribution. Hence the difference between the 1-D parameter estimates of the two sets of samples (mocks and Gaussian samples) comes solely from the incorrect assumption of likelihood functions.

We show in Fig. 2.5 the 1-D parameter estimates of  $\Omega_m$  and  $\sigma_8$  for the mock data and the 10,000 Gaussian samples with and without shape noise. In finding the maximum likelihood estimates, only one parameter ( $\Omega_m$  or  $\sigma_8$ ) is explored at a time. To compare the 1-D distributions, we consider two statistical tests: the independent T-test and the KS-test. The T-test determines whether there is a significant difference between the mean values of two samples. Judging from the high p-values of the T-test for all the four cases in Fig. 2.5, we do not see significant shifts between the average values of the two sets of samples. This indicates that the mean values of 1-D parameter posteriors are not affected by the non-Gaussian marginal distributions of shear correlation functions that we observe in the mock data. In addition, we find that the mean values in the four panels are located around the fiducial values of parameters ( $\Omega_m = 0.2905$  and  $\sigma_8 = 0.826$ ). The deviations from the true value with respect to the uncertainties are listed in Fig. 2.5.

Besides the T-test, we also report the p-values of the two-sample KS-test. The p-values of the KS statistics are all above the 5% threshold, meaning that we do not detect significant difference between the 1-D distributions of parameters of the two samples. With the T-test we learn that the average cosmological parameters are not shifted by the non-Gaussian features of likelihoods in the mocks. The KS-test results suggest that the 1-D distributions of parameters are neither skewed nor distorted in other ways significantly by failure to model the left-skewed marginal distributions of shear correlation functions.

## 2.6.2 Modeling the distributions of principal components

In this subsection, we show results of attempts to model the likelihood function of PCA components.

When we perform KS tests of the univariate Gaussian and Edgeworth models on the training data set, all of the PCA components have  $p$ -values that are approximately uniformly distributed between 0 and 1 as shown in the bottom row of Fig. 2.6. This means that both of the parametric models have good univariate fitting performance on the 90% training data for data with and without shape noise. When we perform KS tests for these models against the testing data (top row of Fig. 2.6), however, the  $p$ -values concentrate around 0 for higher principal components for all the parametric models we consider, in-

cluding Gaussian, Edgeworth function to the second order and Edgeworth function to the fourth order. This indicates that neither one of the models generalizes well to hold-out data for the computed low-variance principal components as a result of the PCA decomposition. Including the components that are mostly dominated by noise would overfit the data. In addition, Fig. 2.6 shows that the Gaussian model fits decently to the leading principal components of the current data. We could not noticeably improve the fitting performance by using more complicated models such as the 1-D Edgeworth function, but we can avoid overfitting by adopting the PCA framework and discarding the high-order principal components with additional benefit of data compression. Note that the KS test statistic is the largest difference between the two empirical CDFs. The KS-statistics take on discrete values, since the empirical CDFs are discrete due to the finite number of samples.

An analysis of variances also shows that the first 40 components capture over 99.9% of the variance of the training data without intrinsic shape noise, whereas the corresponding number of components for data with shape noise is 400.

Figure 2.7 shows the multivariate performance of different likelihood models including parametric models (Gaussian, Edgeworth to the 2nd order, Edgeworth to the 4th order) and non-parametric models ( $k$ -nearest-neighbor and Spectral Series) for data with shape noise. The lower the statistics (that is, the smaller the distance between samples from the model and samples from the test sets), the higher the performance. The last column of Fig. 2.7 shows the null distribution of the test statistics, and tells us what distance values to expect if two samples or sets (with 93 realizations each) were drawn from the true distribution of the data. We estimated this distribution by repeatedly resampling the unmodeled data without replacement, and then in each fold computing distances between these samples and the hold-out sample.

For the data without shape noise, models based on 40 components give good performance on hold-out data, whereas we need around 150 components in the presence of shape noise as seen in Fig. 2.7 for similar performance. We do not see significant differences in performance between the multivariate Gaussian model and the other parametric or non-parametric methods. The estimated null distribution in the last column of Fig. 2.7 indicates that the MMD statistic is powerful enough to discriminate between models for our data set. There is a significant difference between the MMD statistics of models and the real data, implying that there is still some room for improvement. But improving the models would require an order-of-magnitude increase in the number of realizations, which is currently beyond our reach. Hence, our conclusion is that given the current number of realizations and inherent data noise, more complex multi-dimensional likelihood models do not seem to improve upon a simple Gaussian likelihood model typical of an LSST-like survey.

The number of components that retains 99.9% of the variance is just an estimate. In order to find the optimal number of components to use, we apply our multivariate testing framework to a  $k$ -dimensional Gaussian likelihood model for different candidate values of  $k$ . Figure 2.8 indicates that we can further reduce the dimension of the model to 150 for data with shape noise without oversmoothing or “underfitting” the data. For data without shape noise, we can reduce the dimensionality to 40, the same as for our previous 99.9%

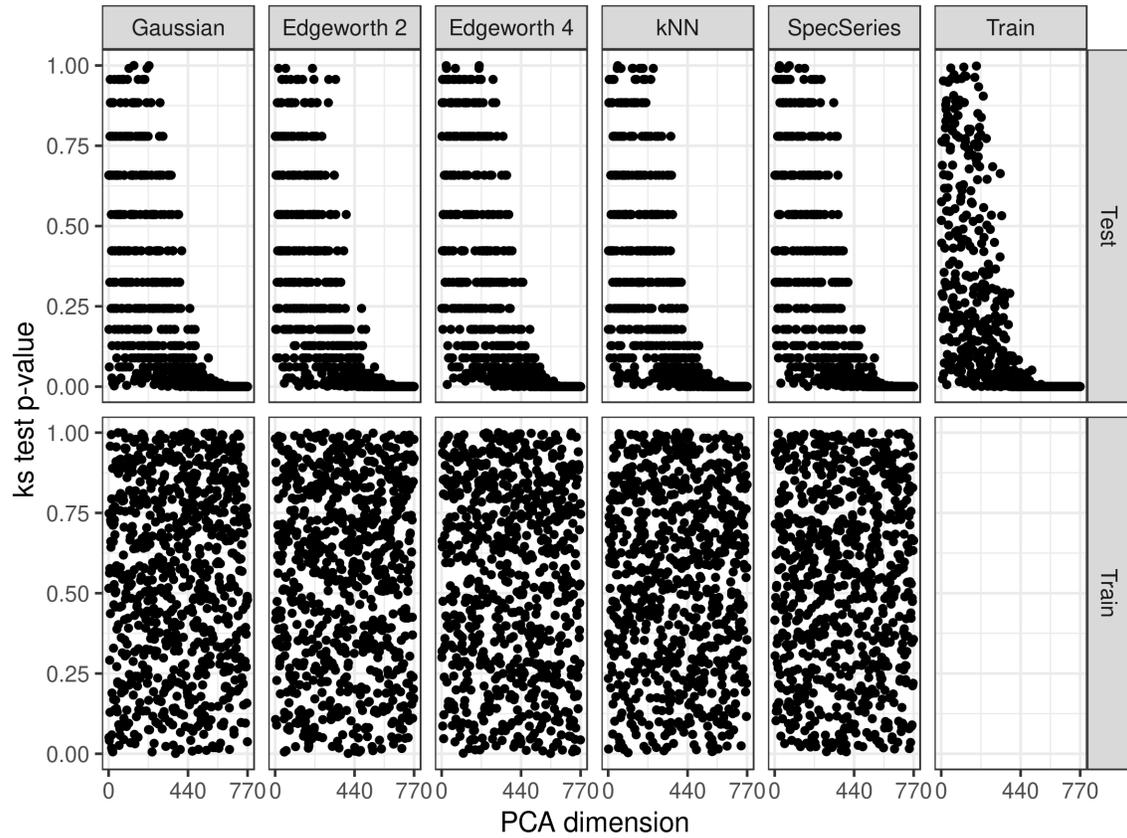


Figure 2.6: KS-test statistics for each PCA dimension on testing and training data. The first five columns show the p-value against the data drawn from the trained models and the upper panel of the last column shows the p-value of training data against testing data. All models have uniformly distributed p-value for training data, showing good univariate fitting performance. The models fail to generalize to testing data for higher-order PCA dimensions. The coarseness of the points is due to the smaller number of data points in the test set.

variance estimate.

### 2.6.3 Biases due to non-Gaussian distributions of principal components

In Sect. 2.6.1, the maximum likelihood method was used to directly estimate cosmological parameter biases due to the non-Gaussian likelihood function of weak lensing shear correlation functions. This method of estimating the level of parameter biases when using the Gaussian likelihood approximation suggested that those biases are subdominant to statistical uncertainties in the mocks. In this subsection, we consider the multivariate PCA models that we have built and assessed in Section 2.6.2 and estimate parameter biases by comparing the MCMC chains of different models with EMCEE <sup>12</sup> (Foreman-Mackey et al., 2013) MCMC sampler. The MCMC chains only explore  $\Omega_m$  and  $\sigma_8$  with the other parameters fixed, and we use flat priors for  $\Omega_m$  and  $\sigma_8$  with range  $0.05 < \Omega_m < 0.6$  and  $0.5 < \sigma_8 < 1.1$ . In the MCMC sampling, we assume a constant covariance estimated from the simulations at the fiducial cosmology, regardless of the changes in  $\Omega_m$  and  $\sigma_8$ .

Figures 2.9 and 2.10 show contour plots of posteriors for the cosmological parameter constraints derived from the mock data with shape noise. The posteriors are re-centered to the true value as a result of the rescaling by the ratio in Eq. (2.7). Fig. 2.9 compares the performance of models with different numbers of principal components and shows the effect of data compression. The 20% difference in the contour areas is due to information loss in data compression. Note that the model with all 770 PCA components and 1-D Gaussian distributions is strictly identical to the standard multivariate Gaussian likelihood of  $\xi_{\pm}$ . Hence the contour labeled as G770 in Figure 2.9 is also what we expect from the standard Gaussian likelihood analysis. Figure 2.10, on the other hand, compares Gaussian and non-Gaussian Edgeworth models in the PCA coordinates to demonstrate any biases due to non-Gaussian likelihood functions. Consistent with the results of the maximum likelihood estimation, no significant difference is found between the contours of the Gaussian model and the Edgeworth model for the mock data with shape noise.

## 2.7 Conclusions and Discussions

It is well known that the multivariate likelihood function of weak lensing shear correlation functions is not a Gaussian. Approximating the likelihood with a Gaussian should introduce associated biases in recovered cosmological parameters; however, whether this would cause significant biases for weak lensing shear correlation functions for upcoming lensing surveys and invalidate the mainstream multivariate Gaussian likelihood assumption has not yet been established. In this paper, we make further advances towards answering this question by modeling and testing non-Gaussian likelihood functions with simulated weak lensing data, and estimating the resulting biases on cosmological parameters.

---

<sup>12</sup><https://github.com/dfm/emcee>

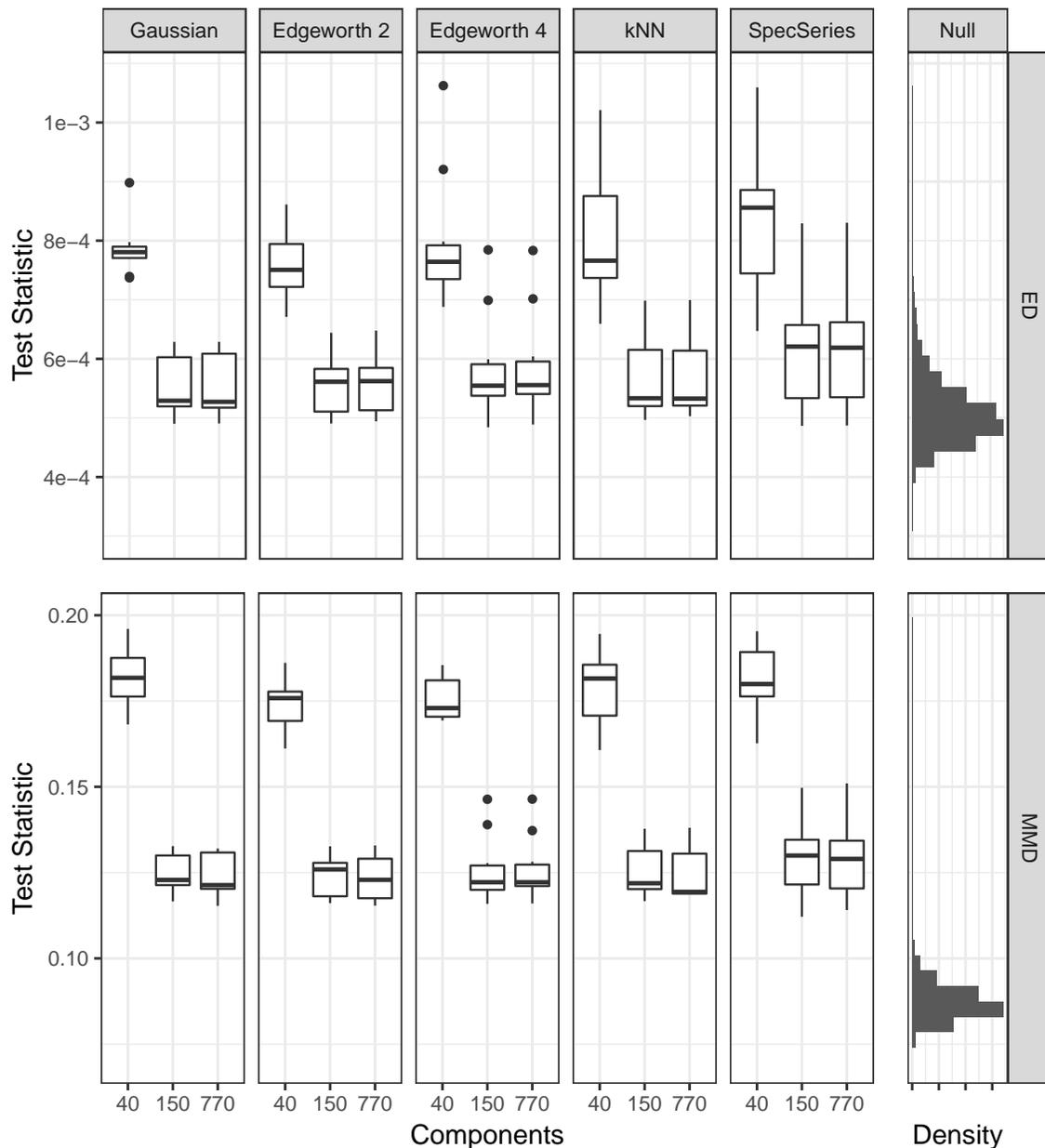


Figure 2.7: Box plots of test statistics for multivariate two-sample tests of different models on hold-out data (10 folds) with shape noise. The statistics are distance metrics of the two multivariate two sample tests (ED and MMD). Smaller statistics mean better agreement between the samples drawn from the trained model and the testing data set. All models are roughly equivalent in generalization performance, but we need to include many more dimensions (150 components are needed for data with shape noise) than in the noiseless case where 40 components are sufficient. The boxes show the range between the first and the third quartile, with the median labeled by the central bar in the boxes, the variability outside the upper and lower quartiles labeled by the vertical lines, and outliers as individual dots. The estimated null distribution is depicted in the final column; this represents the distribution of test statistics under the assumption that the estimated model is correct. For MMD we see that our observed statistics are outside the range of the null distribution. This means that the MMD statistic is discriminatory enough for our data set. It also implies that improving the likelihood models is possible, but leveraging the advantages of more flexible models may require much larger amounts of data.

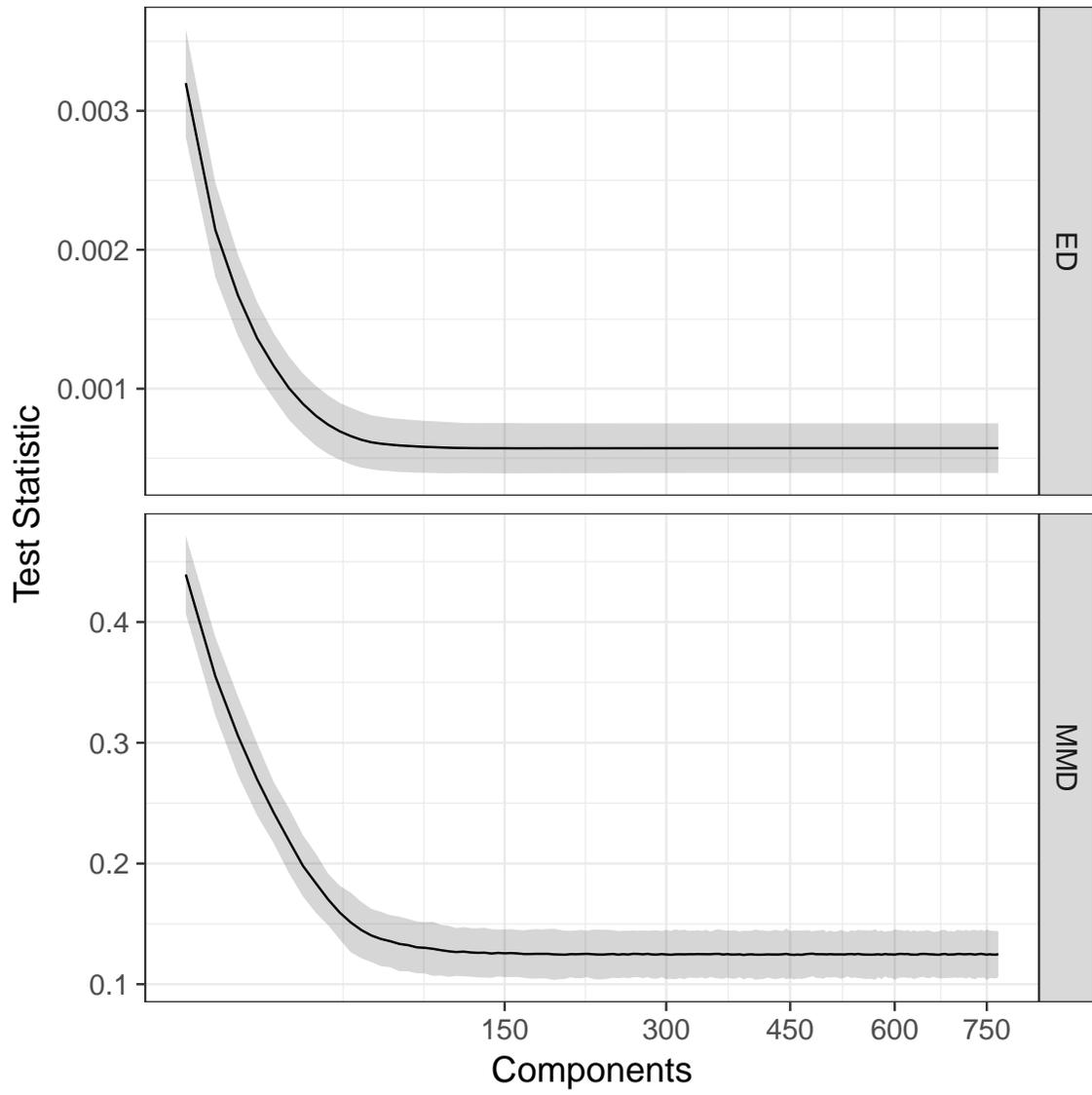


Figure 2.8: Test statistics versus number of components for multivariate two-sample tests on hold-out data with shape noise; the solid line shows the median value and the shaded area represents the interquartile range. Using 150 components gives similar generalization performance as using more components.

## data vector: PCA coordinates

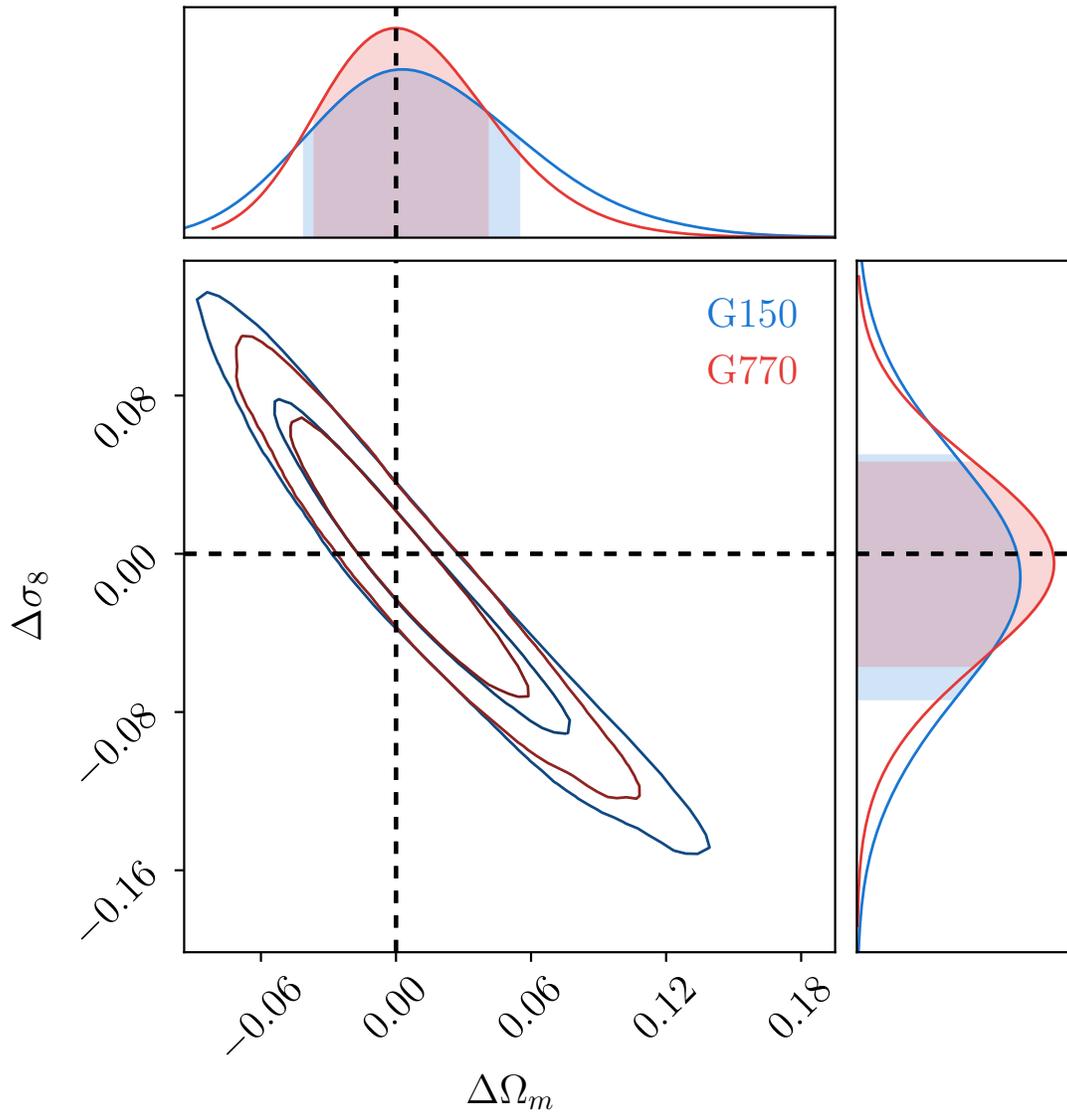


Figure 2.9: Parameter constraints for the Gaussian likelihood models with 150 and 770 principal components using mock data with shape noise. This figure compares the same Gaussian model with 150 and 770 principal components. There is a  $\sim 20\%$  difference in the contour area between 150 and 770 components due to the loss of information when using fewer components. Note that the G770 curve also represents the standard multivariate Gaussian likelihood of  $\xi_{\pm}$ .

## data vector: PCA coordinates

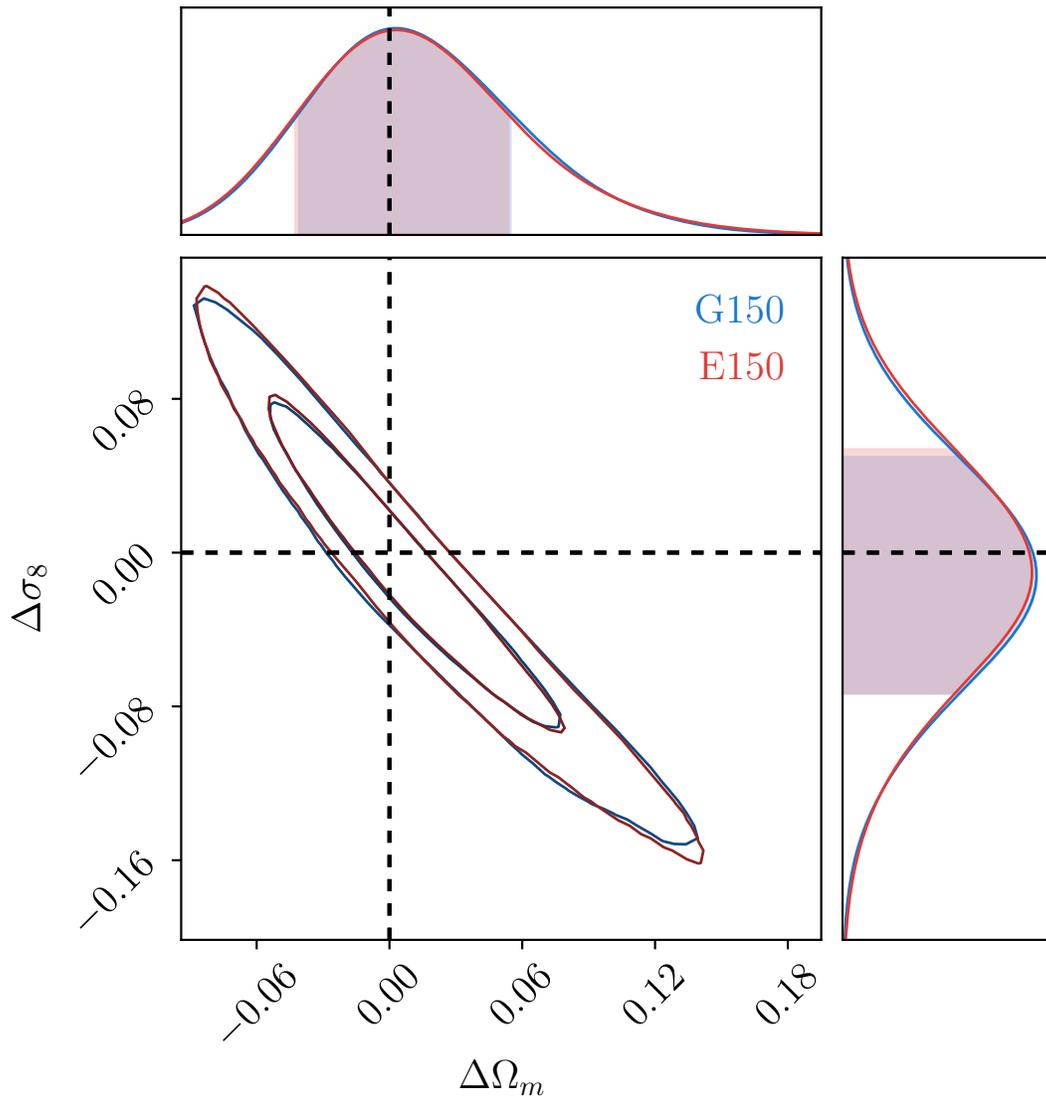


Figure 2.10: Parameter constraints for the Gaussian and the Edgeworth model with 150 components on the shape-noisy mock data. This figure compares the Gaussian model and the Edgeworth model with the same number of principal components (150 components). No significant difference is found between the posteriors of the cosmological parameters when using a Gaussian or Edgeworth model for the likelihood.

We have measured the level of non-Gaussianity of shear correlation function likelihoods in weak lensing simulations and suggest avenues to improve corresponding cosmic shear likelihood analyses. The simulations are based on 100 deg<sup>2</sup> lines-of-sight with the same source redshift distribution and number density as expected for LSST. A systematic approach to constructing the likelihood models was used, and biases in the parameter space ( $\Omega_m, \sigma_8$ ) due to the assumption of a multivariate Gaussian likelihood model were assessed.

We explored the non-Gaussianity in univariate distributions of our data vectors and find detectable non-Gaussianity in the marginal 1D likelihoods of shear correlation functions in  $\theta$  and redshift bins. Even though the marginal distributions of  $\xi_{\pm}$  provide hints on the non-Gaussianity, the linear transformation of PCA reduces the level of skewness and kurtosis in the principal components significantly.

Besides the tests on marginal 1D likelihoods, we use maximum likelihood estimation and 2-D MCMC likelihood analyses with multivariate likelihood models to estimate the parameter biases due to the skewed likelihood distributions of shear correlation functions. In both analyses, we do not observe significant parameter biases in terms of  $\Omega_m$  and  $\sigma_8$  due to the Gaussian likelihood assumption in our data. Note that the results of the Edgeworth likelihood model provide a lower limit on the bias due to non-Gaussian likelihood functions for the 100 deg<sup>2</sup> simulations since this model is built from marginal Edgeworth corrections of simple Gaussian functions in the PCA space based on the assumption of independence of PCs and the marginal distributions of PCA components exhibit small skewness and kurtosis. The multivariate non-parametric models that are not sensitive to coordinate rotations provide more general constraints on the bias, but their performance in terms of predicting hold-out data does not exceed that of the Edgeworth model given our data. We are not asserting that this is the final conclusion for the parameter biases due to the uncertainty of likelihoods. Measuring the high-dimensional likelihoods is extremely difficult. It is likely that our likelihood models, especially the non-parametric models, would benefit from orders of magnitude of increase in the number of simulation realizations. However, given the currently available number of simulation realizations, our results show no significant biases using the standard multivariate Gaussian likelihood.

In Appendices 2.8.1 and 2.9 we show that the skewness and the  $(\text{mode} - \text{mean})/\sigma$  of shear correlation functions decrease with the survey area as  $f_{\text{sky}}^{-1/2}$ . The non-Gaussianity of the likelihood and the resulting cosmological parameter biases would therefore be even smaller for large surveys such as LSST. The scaling relation is derived by assuming that the shear fields are Gaussian, corresponding to larger scales. Future work should explore the relative contributions of the non-linear growth (non-Gaussian field) to the non-Gaussianity in likelihoods.

Therefore, given the small bias measured from the current simulations, we do not expect noticeable biases due to the non-Gaussian distribution of the weak lensing shear correlation function for future generation of large-scale structure observations. A multivariate Gaussian likelihood will continue to be a valid approximation in cosmic shear analyses.

In addition to our findings on the Gaussian approximation to the likelihood function, PCA poses a straightforward avenue to solving some of the practical problems related to

covariance matrices. Several practical challenges for debiasing the covariance matrix of the weak lensing observables from simulations have been featured in the literature (Schneider et al., 2011; Dodelson & Schneider, 2013; Taylor & Joachimi, 2014; Blot et al., 2016), all of which are connected to the large number of data points expected in future cosmic shear surveys.

One way to address the problem is to identify a more efficient observable and hence compress the cosmological information into fewer data points. We have shown that the assumption of a multivariate Gaussian in PCA coordinates is valid for the mock data and that PCA ranks the data points efficiently as a function of signal-to-noise. This allows us to reduce the dimensionality of the data vector by throwing out modes with low signal-to-noise, which in turn alleviates some of the pressing problems related to covariance inversion.

Data compression approaches have been suggested in the past, such as MOPED (Heavens et al., 2000, 2017) and COSEBIs (Schneider et al., 2010); it remains to be seen if the likelihood function of these summary statistics is closer to a Gaussian than, e.g. for the shear two-point correlation function. Using PCA, we have been able to reduce the number of dimensions from 770 to 40 for data without shape noise and from 770 to 150 for data with shape noise with acceptable loss in cosmological information.

In brief, our conclusions are itemized as follows:

1. We find significant non-Gaussianity in the marginal distributions of the correlation functions, indicating that the multivariate likelihood distributions are non-Gaussian.
2. We estimate the bias in cosmological parameters induced by ignoring the non-Gaussianity of the likelihood with maximum likelihood estimation assuming a Gaussian likelihood, and do not detect strong biases in  $\Omega_m$  and  $\sigma_8$  in our simulated data.
3. Since the  $(\text{mode} - \text{mean})/\sigma$  of shear correlation functions scales with the survey area as  $f_{\text{sky}}^{-1/2}$  under the assumption that the fields are Gaussian and that the angular power spectra follow a gamma distribution, we expect the bias due to assuming a Gaussian likelihood to be smaller for LSST than for these small-area mock catalogs.
4. Fully reconstructing high-dimensional distributions directly from simulations is very difficult, requiring simulation volumes well beyond those presently available for this study. Our results based on the described set of simulated data suggest that neglecting the non-Gaussianity of the likelihood for shear-shear correlations is not a significant source of bias for ongoing surveys or even future ones such as LSST.

While the simulated weak lensing data implies that the impact of non-Gaussianity is negligible in current and future cosmic shear surveys, it is an open question if this result is stable for different redshift distributions, different survey parameters, and different cosmologies. In Eifler et al. (2009) they show that the covariance and hence the likelihood is cosmology-dependent and that the strength of this effect depends on the specific properties of the survey. It is also unclear if some of the systematics that affect shear observables can introduce non-Gaussianity (e.g., as foregrounds do in case of the CMB),

which could warrant further mitigation strategies. Finally, it will be interesting to explore the non-Gaussianity of the likelihood function for the multi-probe analysis case, e.g., when including galaxy clustering, galaxy-galaxy lensing, and perhaps even higher-order statistics of auto and cross observables of clustering and shear. A multi-probe simulation with even more realizations is required for the joint data vector.

## Acknowledgments

This paper has undergone internal review in the LSST Dark Energy Science Collaboration. The authors would like to give thanks to the DESC Publication Board and the internal reviewers A. Heavens, R. Rosenfeld and J. Zuntz for the feedback and comments that improved our manuscript. CHL, JHD, TE and RM developed the concept and designed the analysis framework; CHL carried out the analysis and prepared the manuscript with the help of JHD, TE, RM, ABL, TP and SS; JHD conducted weak-lensing simulations; TP and ABL designed and performed the statistical model assessment, and fitted the nonparametric likelihood models; SS derived expressions of the third moment of shear and mean-mode difference. JHD acknowledges support from the European Commission under a Marie-Skłodowska-Curie European Fellowship (EU project 656869) and from the European Research Council under grant number 647112. Computations for the  $N$ -body simulations were performed in part on the Orcinus supercomputer at the WestGrid HPC consortium ([www.westgrid.ca](http://www.westgrid.ca)), in part on the GPC supercomputer at the SciNet HPC Consortium. SciNet is funded by: the Canada Foundation for Innovation under the auspices of Compute Canada; the Government of Ontario; Ontario Research Fund - Research Excellence; and the University of Toronto. This work was partially supported by NSF DMS-1520786, DOE grant DESC0011114, and Department of Energy Cosmic Frontier program. Part of the research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration and is supported by NASA ROSES ATP 16-ATP16-0084 grant and by NASA ROSES 16-ADAP16-0116.

The DESC acknowledges ongoing support from the Institut National de Physique Nucléaire et de Physique des Particules in France; the Science & Technology Facilities Council in the United Kingdom; and the Department of Energy, the National Science Foundation, and the LSST Corporation in the United States. DESC uses resources of the IN2P3 Computing Center (CC-IN2P3–Lyon/Villeurbanne - France) funded by the Centre National de la Recherche Scientifique; the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231; STFC DiRAC HPC Facilities, funded by UK BIS National E-infrastructure capital grants; and the UK particle physics grid, supported by the GridPP Collaboration. This work was performed in part under DOE Contract DE-AC02-76SF00515.

## 2.8 Third Central Moment

### 2.8.1 Convergence

Here we derive the expression for the third central moment of the (tomographic) cross-correlation function of two convergence maps denoted by  $X$  and  $Y$ . Similar calculations have also been done by Keitel & Schneider (2011). We use a different notation that is easier to generalize for cross-correlations and restrict ourselves to computing the third central moment, though calculations can be extended straightforwardly but tediously to higher moments. We show here how skewness is affected by  $f_{sky}$ , a relation we used in Sect. 2.4.6 in the main text.

For simplicity, we will work in the flat-sky approximation. We will also assume that the fields are Gaussian and ignore the non-Gaussian terms, though we will indicate where the non-Gaussian terms should enter. Using simulations, we have already shown that at the small scales, where the non-Gaussian terms are most important, the impact of non-Gaussian likelihood is small.

The cross-correlation function of two fields can be written as

$$x_Y(\vec{\theta}) = \frac{1}{\mathcal{A}_W(\vec{\theta})} \int d^2\vec{\theta}' W(\vec{\theta}' + \vec{\theta}) W(\vec{\theta}')_1(\vec{\theta}')_2(\vec{\theta}' + \vec{\theta}) \quad (2.24)$$

where  $\kappa_1$  belongs to field  $X$  and  $\kappa_2$  to field  $Y$ .  $W(\vec{\theta})$  is the survey window function. We have assumed that the noise and the  $\kappa_i$  have zero mean and are also uncorrelated with each other on all scales. The normalization factor is the integral over window functions

$$\mathcal{A}_W(\vec{\theta}) = \int d^2\vec{\theta}' W(\vec{\theta}' + \vec{\theta}) W(\vec{\theta}') = \int \frac{d^2\vec{\ell}}{(2\pi)^2} e^{-i\vec{\ell}\cdot\vec{\theta}} \tilde{W}(\vec{\ell}) \tilde{W}(-\vec{\ell}) = \int \frac{d\ell}{2\pi} J_0(\ell\theta) \tilde{W}(\ell) \tilde{W}(-\ell) \quad (2.25)$$

The third central moment ( $S_3$ ) of the correlation function is given as

$$\begin{aligned} \mathbf{S}_3(\widehat{\xi}_{\kappa_1\kappa_2}(\vec{\theta}_i) \widehat{\xi}_{\kappa_3\kappa_4}(\vec{\theta}_j) \widehat{\xi}_{\kappa_5\kappa_6}(\vec{\theta}_l)) &= \left( \widehat{\xi}_{\kappa_1\kappa_2}(\vec{\theta}_i) - \widehat{\xi}_{\kappa_1\kappa_2}(\vec{\theta}_i) \right) \left( \widehat{\xi}_{\kappa_3\kappa_4}(\vec{\theta}_j) - \widehat{\xi}_{\kappa_3\kappa_4}(\vec{\theta}_j) \right) \left( \widehat{\xi}_{\kappa_5\kappa_6}(\vec{\theta}_l) - \widehat{\xi}_{\kappa_5\kappa_6}(\vec{\theta}_l) \right) \\ &= \kappa_1\kappa_2(\vec{\theta}_i)_{\kappa_3\kappa_4}(\vec{\theta}_j)_{\kappa_5\kappa_6}(\vec{\theta}_l) + 2\kappa_1\kappa_2(\vec{\theta}_i)_{\kappa_3\kappa_4}(\vec{\theta}_j)_{\kappa_5\kappa_6}(\vec{\theta}_l) \\ &\quad - \left\{ \kappa_1\kappa_2(\vec{\theta}_i) \text{Cov}(\kappa_3\kappa_4(\vec{\theta}_j), \kappa_5\kappa_6(\vec{\theta}_l)) + \text{perms} \right\} \end{aligned} \quad (2.26)$$

where  $\kappa_1, \kappa_3, \kappa_5$  belong to field  $X$  and  $\kappa_2, \kappa_4, \kappa_6$  belong to  $Y$ .  $S_3$  is a function of three  $\theta$  variables and is therefore a third-order tensor. The ‘perms’ denotes permutations over combinations of (12, 34, 56), where one combination gives  $\xi$  and other two give covariance. So there are 6 permutations in the last term of Eq. (2.27) (or 3 terms if the symmetry of covariance is considered). We use  $\kappa_1\kappa_2\kappa_3\kappa_4\kappa_5\kappa_6_{ijl}$  as short-hand for the first term on the

right hand side in Eq. (2.27), which we would like to simplify:

$$\kappa_1 \kappa_2 \kappa_3 \kappa_4 \kappa_5 \kappa_6_{ij} = \frac{1}{\mathcal{A}_W(\vec{\theta}_i) \mathcal{A}_W(\vec{\theta}_j) \mathcal{A}_W(\vec{\theta}_l)} \int d^2 \vec{\theta} \int d^2 \vec{\theta}' \int d^2 \vec{\theta}'' \kappa_1(\vec{\theta}) \kappa_2(\vec{\theta} + \vec{\theta}_i) \kappa_3(\vec{\theta}') \kappa_4(\vec{\theta}' + \vec{\theta}_j) \kappa_5(\vec{\theta}'') \kappa_6(\vec{\theta}'' + \vec{\theta}_l) W(\vec{\theta}) W(\vec{\theta}') W(\vec{\theta} + \vec{\theta}_i) W(\vec{\theta}' + \vec{\theta}_j) W(\vec{\theta}'') W(\vec{\theta}'' + \vec{\theta}_l) \quad (2.28)$$

Writing the  $\kappa_i$  in terms of its Fourier space counterpart  $\tilde{\kappa}_i$ , we get

$$\kappa_1 \kappa_2 \kappa_3 \kappa_4 \kappa_5 \kappa_6_{ij} = \frac{1}{\mathcal{A}_W(\vec{\theta}_i) \mathcal{A}_W(\vec{\theta}_j) \mathcal{A}_W(\vec{\theta}_l)} \int d^2 \vec{\theta} \int d^2 \vec{\theta}' \int d^2 \vec{\theta}'' \iiint \prod_{n=1}^6 \left[ \frac{d^2 \vec{\ell}_n}{(2\pi)^2} \right] \iiint \prod_{m=1}^6 \left[ \frac{d^2 \vec{q}_m}{(2\pi)^2} \tilde{W}(\vec{q}_m) \right] \times e^{i(\vec{\ell}_1 - \vec{q}_1) \cdot \vec{\theta}} e^{i(\vec{\ell}_2 - \vec{q}_2) \cdot (\vec{\theta} + \vec{r}_i)} e^{i(\vec{\ell}_3 - \vec{q}_3) \cdot \vec{\theta}'} e^{i(\vec{\ell}_4 - \vec{q}_4) \cdot (\vec{\theta}' + \vec{\theta}_j)} e^{i(\vec{\ell}_5 - \vec{q}_5) \cdot \vec{\theta}''} e^{i(\vec{\ell}_6 - \vec{q}_6) \cdot (\vec{\theta}'' + \vec{\theta}_l)} \quad (2.29)$$

$$\times \tilde{\kappa}_1(\vec{\ell}_1) \tilde{\kappa}_2(\vec{\ell}_2) \tilde{\kappa}_3(\vec{\ell}_3) \tilde{\kappa}_4(\vec{\ell}_4) \tilde{\kappa}_5(\vec{\ell}_5) \tilde{\kappa}_6(\vec{\ell}_6) \\ \kappa_1 \kappa_2 \kappa_3 \kappa_4 \kappa_5 \kappa_6_{ij} = \frac{1}{\mathcal{A}_W(\vec{\theta}_i) \mathcal{A}_W(\vec{\theta}_j) \mathcal{A}_W(\vec{\theta}_l)} \iint \frac{d^2 \vec{\ell}_1}{(2\pi)^2} \frac{d^2 \vec{\ell}_3}{(2\pi)^2} \frac{d^2 \vec{\ell}_5}{(2\pi)^2} \iiint \prod_{m=1}^4 \left[ \frac{d^2 \vec{q}_m}{(2\pi)^2} \tilde{W}(\vec{q}_m) \right] e^{-i(\vec{\ell}_1 - \vec{q}_1) \cdot \vec{\theta}_i} e^{-i(\vec{\ell}_3 - \vec{q}_3) \cdot \vec{\theta}_j} \times \tilde{\kappa}_1(\vec{\ell}_1) \tilde{\kappa}_2(-\vec{\ell}_1 + \vec{q}_1 + \vec{q}_2) \tilde{\kappa}_3(\vec{\ell}_3) \tilde{\kappa}_4(-\vec{\ell}_3 + \vec{q}_3 + \vec{q}_4) \kappa_5(\vec{\ell}_5) \tilde{\kappa}_6(-\vec{\ell}_5 + \vec{q}_5 + \vec{q}_6). \quad (2.30)$$

We have integrated over  $d^2 \vec{\theta}$ ,  $d^2 \vec{\theta}'$ ,  $d^2 \vec{\theta}''$  and then over  $d^2 \vec{\ell}_2$ ,  $d^2 \vec{\ell}_4$ ,  $d^2 \vec{\ell}_6$  to obtain the last expression.

We now expand the six-point function into two separable parts: the connected or non-Gaussian component  $\tilde{\kappa}_1 \tilde{\kappa}_2 \tilde{\kappa}_3 \tilde{\kappa}_4 \kappa_5 \kappa_6'$  and the Gaussian component, which using Wick's theorem can be expanded as the sum of the product of two-point functions:

$$\kappa_1 \kappa_2 \kappa_3 \kappa_4 \kappa_5 \kappa_6_{ij} = \frac{1}{\mathcal{A}_W(\vec{\theta}_i) \mathcal{A}_W(\vec{\theta}_j) \mathcal{A}_W(\vec{\theta}_l)} \iint \frac{d^2 \vec{\ell}_1}{(2\pi)^2} \frac{d^2 \vec{\ell}_3}{(2\pi)^2} \frac{d^2 \vec{\ell}_5}{(2\pi)^2} \iiint \prod_{m=1}^6 \left[ \frac{d^2 \vec{q}_m}{(2\pi)^2} \tilde{W}(\vec{q}_m) \right] e^{-i(\vec{\ell}_1 - \vec{q}_1) \cdot \vec{\theta}_i} e^{-i(\vec{\ell}_3 - \vec{q}_3) \cdot \vec{\theta}_j} \times [\tilde{\kappa}_1 \tilde{\kappa}_2 \tilde{\kappa}_3 \tilde{\kappa}_4 \kappa_5 \kappa_6' + \tilde{\kappa}_1 \tilde{\kappa}_2 \tilde{\kappa}_3 \tilde{\kappa}_4 \kappa_5 \kappa_6 + \text{all perms}] \quad (2.31)$$

where the 'all perms' now denote all possible combinations over  $(1, 2, 3, 4, 5, 6)$ . Note that a similar expansion is also required for four-point functions in the covariance terms in Eq. (2.27) (see for example Singh et al. 2017 for a covariance expansion in similar notation). Simplifying, the auto-correlation terms from the same theta bin, e.g.  $\kappa_1 \kappa_2$ , cancel out and we are left with the permutations that contain only cross-correlation terms involving at most one  $\kappa_i$  from each  $\vec{\theta}_j$  bins (there are  ${}^4 C_1 \times {}^2 C_1 = 8$  such permutations). Thus we have

$$\mathbf{S}_3 = \frac{1}{\mathcal{A}_W(\vec{\theta}_i) \mathcal{A}_W(\vec{\theta}_j) \mathcal{A}_W(\vec{\theta}_l)} \iint \frac{d^2 \vec{\ell}_1}{(2\pi)^2} \frac{d^2 \vec{\ell}_3}{(2\pi)^2} \frac{d^2 \vec{\ell}_5}{(2\pi)^2} \iiint \prod_{m=1}^6 \left[ \frac{d^2 \vec{q}_m}{(2\pi)^2} \tilde{W}(\vec{q}_m) \right] e^{-i(\vec{\ell}_1 - \vec{q}_1) \cdot \vec{\theta}_i} e^{-i(\vec{\ell}_3 - \vec{q}_3) \cdot \vec{\theta}_j} e^{-i(\vec{\ell}_5 - \vec{q}_5) \cdot \vec{\theta}_l} \times [\tilde{\kappa}_1 \tilde{\kappa}_2 \tilde{\kappa}_3 \tilde{\kappa}_4 \tilde{\kappa}_5 \tilde{\kappa}_6' + \tilde{\kappa}_1 \tilde{\kappa}_3 \tilde{\kappa}_2 \tilde{\kappa}_5 \tilde{\kappa}_4 \tilde{\kappa}_6 + \text{all cross perms}] \quad (2.32)$$

To further simplify the expressions, we will assume that the scales of interest are smaller than the survey window size and ignore the coupling between the window function and the power spectra. Within this assumption, the window function integrals can be carried out without the power spectra to finally give

$$S_3 = \frac{\mathcal{A}_W(\vec{\theta}_i - \vec{\theta}_j + \vec{\theta}_k)}{\mathcal{A}_W(\vec{\theta}_i)\mathcal{A}_W(\vec{\theta}_j)\mathcal{A}_W(\vec{\theta}_l)} \iint \frac{d^2\vec{\ell}}{(2\pi)^2} e^{-i\vec{\ell}(\vec{\theta}_i - \vec{\theta}_j + \vec{\theta}_l)} [C_{13}(\ell)C_{25}(\ell)C_{46}(\ell) + \text{all cross perms}] \quad (2.33)$$

$$S_3 = \frac{\mathcal{A}_W(\vec{\theta}_i - \vec{\theta}_j + \vec{\theta}_k)}{\mathcal{A}_W(\vec{\theta}_i)\mathcal{A}_W(\vec{\theta}_j)\mathcal{A}_W(\vec{\theta}_l)} \int \frac{d\ell \ell}{2\pi} J_0(\ell\theta_i)J_0(\ell\theta_j)J_0(\ell\theta_l) [2C_{XY}(\ell)^3 + 6C_{XX}(\ell)C_{YY}(\ell)C_{XY}(\ell)] \quad (2.34)$$

where in the second equation we have used the fact that odd-odd combinations such as ‘13’ give the auto-correlation of field  $X$ ,  $C_{XX}$ , even-even combinations give the auto-correlation of  $Y$ ,  $C_{YY}$  and the even-odd combinations give the cross-correlation  $C_{XY}$ . Since  $A_W$  scales with the fraction of sky covered by a survey,  $f_{\text{sky}}$ , Eq. (2.34) suggests that the third central moment scales as  $f_{\text{sky}}^{-2}$ . Thus for any given scale that is well within the survey window size, the third central moment of the convergence correlation function decreases faster than the covariance as the survey area increases. However, as the scale  $\theta$  approaches the size of survey, the area factor in the normalization approaches zero, i.e.  $A_W(\theta) \rightarrow 0$  for large  $\theta$  and  $S_3$  will increase.  $S_3$  rises faster than covariance,  $S_2$  and thus the skewness increases at large scales. This is consistent with the expectation from the central limit theorem as large scales have fewer modes within the survey and we expect them to be more skewed.

We also note that we have ignored the connected terms, which can be important if the fields are non-Gaussian. The connected terms have three factors of area,  $A_W$  in the denominator and a factor of window functions in the numerators. Under the assumption that the window and the connected terms can be decoupled, the window term in the numerator will cancel one factor of  $A_W$  and the connected term will also scale similarly to the Gaussian term. The decoupling of the window is not well justified in the case of coupling between large-scale (super-sample) and small-scale modes, but such coupling terms have also been shown to scale with similar factors of area in the context of super-sample covariance. We have not tested such scaling for the six-point function in this work.

## 2.8.2 Shear

For shear we begin by noting that  $\xi_+ \propto \langle \gamma_X \gamma_Y^* \rangle$  and  $\xi_- \propto \langle \gamma_X \gamma_Y \rangle$ . Using these relations and the expressions from the previous subsection, we get

$$S_3(\xi_-) = \frac{\mathcal{A}_W(\vec{\theta}_i - \vec{\theta}_j + \vec{\theta}_k)}{\mathcal{A}_W(\vec{\theta}_i)\mathcal{A}_W(\vec{\theta}_j)\mathcal{A}_W(\vec{\theta}_l)} \int \frac{d\vec{\ell} \ell}{2\pi} J_4(\ell\theta_i)J_4(\ell\theta_j)J_4(\ell\theta_k) [2C_{XY}(\ell)^3 + 6C_{XX}(\ell)C_{YY}(\ell)C_{XY}(\ell)] \quad (2.35)$$

$$S_3(\xi_+) = \frac{\mathcal{A}_W(\vec{\theta}_i - \vec{\theta}_j + \vec{\theta}_k)}{\mathcal{A}_W(\vec{\theta}_i)\mathcal{A}_W(\vec{\theta}_j)\mathcal{A}_W(\vec{\theta}_l)} \int \frac{d\vec{\ell} \ell}{2\pi} [2C_{XY}(\ell)^3 J_0(\ell\theta_i)J_0(\ell\theta_j)J_0(\ell\theta_k)] \\ + [2C_{XX}(\ell)C_{YY}(\ell)C_{XY}(\ell)J_4(\ell\theta_i)J_4(\ell\theta_j)J_0(\ell\theta_k) + \text{cyc}(i - k) + \text{cyc}(j - k)] \quad (2.36)$$

where ‘cyc’ denotes the interchanging of order of the Bessel functions for different  $\theta$ . We conclude from the equation above that the same scaling with  $f_{\text{sky}}$  applies to shear as well.

## 2.9 Difference between mean and mode

Assuming a unimodal distribution for a random variable  $X$ , the difference between the most likely value, mode  $\tilde{X}$ , and the mean  $\bar{X}$  in terms of the standard deviation  $\sigma$  has an upper bound (Johnson & Rogers, 1951)

$$\frac{|\tilde{X} - \bar{X}|}{\sigma} \leq \sqrt{3} \quad (2.37)$$

In our main analysis using 100 deg<sup>2</sup> simulations, we have already shown that the difference between the mean and mode is much smaller than this bound.

In order to address the question of how this difference scales with the survey area, we begin by noting that for angular power spectra  $C_\ell$ , which follow a gamma distribution with  $\nu \approx (2\ell + 1)f_{\text{sky}}$  degrees of the freedom (see e.g. Percival & Brown, 2006, for a detailed discussion of the  $C_\ell$  likelihood for CMB), the difference between the mean and the mode is

$$\tilde{C}_\ell - \bar{C}_\ell \approx -\frac{2}{\nu}\bar{C}_\ell \quad (2.38)$$

With the variance  $\sigma_{C_\ell}^2 \approx 2(C_\ell^2/\nu)$ , the difference in terms of  $\sigma$  is

$$\frac{|\tilde{C}_\ell - \bar{C}_\ell|^2}{\sigma_{C_\ell}^2} \approx \frac{2}{\nu}. \quad (2.39)$$

Here we assume the cosmic variance-dominated regime, but the scaling in the shot noise-dominated regime is similar since the shot noise also scales as  $1/f_{\text{sky}}$  for a fixed number density of tracers.

Since the correlation functions,  $\xi_{\pm}$ , are the Hankel transform of  $C_{\ell}$  (in the flat sky approximation), the difference between the mean and the mode for the correlation functions is

$$\tilde{\xi} - \bar{\xi} = \int \frac{d\ell \ell}{2\pi} J_n(\ell\theta) (\tilde{C}_{\ell} - \bar{C}_{\ell}) \approx -\frac{2}{f_{\text{sky}}} \int \frac{d\ell \ell}{2\pi} J_n(\ell\theta) \frac{\bar{C}_{\ell}}{2\ell + 1} \quad (2.40)$$

where  $n = 0$  for  $\xi_+$  and  $n = 4$  for  $\xi_-$ .

Since the covariance,  $\mathbf{C}$ , of  $\xi$  scales as  $1/f_{\text{sky}}$ , the difference between the mean and the mode in terms of signal-to-noise ratio scales as

$$(\tilde{\xi} - \bar{\xi})^T \mathbf{C}^{-1} (\tilde{\xi} - \bar{\xi}) \propto \frac{1}{f_{\text{sky}}} \quad (2.41)$$

Eq. (2.41) also gives the scaling of the bias in the log likelihood (Gaussian) and the parameter covariance with  $f_{\text{sky}}$ .

# Chapter 3

## The Impact of Light Polarization Effects on Weak Lensing Systematics

### 3.1 Abstract

A fraction of the light observed from edge-on disk galaxies is polarized due to two physical effects: selective extinction by dust grains aligned with the magnetic field, and scattering of the anisotropic starlight field. Since the reflection and transmission coefficients of the reflecting and refracting surfaces in an optical system depend on the polarization of incoming rays, this optical polarization produces both (a) a selection bias in favor of galaxies with specific orientations and (b) a polarization-dependent PSF. In this work we build toy models to obtain for the first time an estimate for the impact of polarization on PSF shapes and the impact of the selection bias due to the polarization effect on the measurement of the ellipticity used in shear measurements. In particular, we are interested in determining if this effect will be significant for WFIRST. We show that the systematic uncertainties in the ellipticity components are  $8 \times 10^{-5}$  and  $1.1 \times 10^{-4}$  due to the selection bias and PSF errors respectively. Compared to the overall requirements on knowledge of the WFIRST PSF ellipticity ( $4.7 \times 10^{-4}$  per component), both of these systematic uncertainties are sufficiently close to the WFIRST tolerance level that more detailed studies of the polarization effects or more stringent requirements on polarization-sensitive instrumentation for WFIRST are required.

### 3.2 Introduction

Weak gravitational lensing arises due to deflection of light by the gravitational fields of large-scale structure, leading to tangential shear distortions in galaxy shapes. Measuring the correlation functions of galaxy shapes is therefore a method by which we can measure the growth of structure in the Universe (Bartelmann & Schneider, 2001; Massey et al., 2007; Hoekstra & Jain, 2008; Kilbinger, 2015; Mandelbaum, 2018) and hence a powerful method

for constraining cosmological parameters (Huff et al., 2014; Jee et al., 2016; Hildebrandt et al., 2018; Troxel et al., 2018b; Hikage et al., 2019a).

Since weak gravitational shear is only a percent-level signal, weak lensing measurements rely on the use of large galaxy samples to reduce statistical uncertainties. In the upcoming Stage-IV surveys, including Euclid<sup>1</sup> (Laureijs et al., 2011), LSST<sup>2</sup> (Ivezić et al., 2008; LSST Science Collaboration et al., 2009), and WFIRST<sup>3</sup> (Spergel et al., 2015), the statistical uncertainties of the weak lensing measurements are expected to reach sub-percent level precision. With such small statistical uncertainties, the future of weak lensing analysis requires a better understanding and more careful control of systematics to avoid systematic uncertainties dominating over statistical uncertainties. Low-level sources of systematic uncertainty that are presently ignored in existing weak lensing analyses will become significant as the precision of the measurements increases.

Optical gravitational lensing measurements are based on how the light intensity profile of galaxies is affected by the matter distribution in the Universe. Besides intensity, the observed radiation also includes polarization information (Radhakrishnan, 1989). However, to the best of our knowledge, the information in polarization has not been considered in optical weak lensing. While the light intensity profile contains the weak lensing information that we hope to measure, it is possible that the polarization of light could introduce weak lensing systematic errors if not accounted for.

Many polarization measurements of light from galaxies have been made at radio frequencies (Beck et al., 2002; Beck, 2007; Stil et al., 2009; Akahori et al., 2018). At the radio frequency, the radiation is dominated by synchrotron emissions. Synchrotron emission is intrinsically linearly polarized perpendicular to the magnetic field; at long wavelengths, Faraday rotation changes the polarization angle and provides an additional probe of the magnetic field structure.

Weak lensing surveys are conducted, however, in the optical and near-infrared (NIR) wavelengths. The physics behind radio polarization is different from the optical/NIR. Unlike the radio band which is dominated by synchrotron emissions, the light at the optical and near infrared frequencies is not itself linearly polarized. Some polarization effects may be induced through the interactions with dust grains (Mathis, 1990). The optical galactic polarization has been studied and measured in Scarrott et al. (1990, 1991); Fendt et al. (1996); Scarrott et al. (1987); Fendt et al. (1998); Scarrott (1996); Jones (2000). Even though the linear polarization of light from galaxies has been observed, polarization-induced systematic errors have not generally been considered for optical weak lensing analyses.

The overall weak lensing analysis covers many steps from image processing to the inference of cosmological parameter constraints. Different systematic uncertainties enter in each step. The sheared galaxy image that contains weak lensing information first propagates through the atmosphere (for ground-based observations) and the telescope optics, and is affected by the Point Spread Function (PSF) of both. The final images are further

---

<sup>1</sup><http://sci.esa.int/euclid/>

<sup>2</sup><http://www.lsst.org/lsst>

<sup>3</sup><https://wfirst.gsfc.nasa.gov>

affected by detector non-idealities. The subsequent processes in the weak lensing analysis involve using the final image to construct the shear catalogues and summary statistics such as two-point correlation functions, and finally inferring the cosmological parameters using a likelihood analysis. Among all of these steps, the polarization effect influences how the light rays propagate in the optical system before the final images are realized on the detectors. Since the s and p-polarized components of light have different reflection and refraction coefficients at optical surfaces, including the mirrors and the anti-reflection coating on the detectors, the throughput of the optics and telescope PSF depend on the polarization of the input light (Breckinridge et al., 2015; Chipman et al., 2015). If the light from galaxies is linearly polarized, the existing analysis processes that ignore the polarization information could lead to biases in the weak lensing results. In this paper, we consider two potential polarization-related systematic errors: galaxy selection biases, and the PSF modeling errors due to the polarization-dependent PSF. The selection bias arises from the dependence of the transmitted intensity of polarized light on the angle of polarization. The polarization-dependent PSF, on the other hand, is a result of the polarization-dependent optical aberrations.

In Section 3.3, we briefly review polarization effects in optical/NIR galaxy images. Sections 3.5 and 3.6 describe the assumptions we make and how we construct our toy models for polarization-related selection and PSF effects respectively. We present our results in Section 3.7 and conclude in Section ??.

### 3.3 Optical/NIR Polarization of galaxies

As shown in previous studies on optical polarization, light from distant stars and galaxies can be linearly polarized in the optical wavelengths (Fendt et al., 1996; Scarrott et al., 1987; Fendt et al., 1998). At optical and near infrared frequencies, the linear polarization of light arises from physical mechanisms including the anisotropic scattering by spherical dust grains and the selective extinction by aligned dust grains. The alignment was once proposed to be caused by magnetic dissipation – the Davis-Greenstein mechanism (Davis & Greenstein, 1951) – but is now attributed to precession around the magnetic field combined with torques from scattering of starlight, which also usually results in alignment of the long axis of the grains perpendicular to the magnetic field (Andersson et al., 2015). Both of these mechanisms can have complicated wavelength dependence. The dust scattering cross section increases toward the blue and peaks in the ultraviolet (Draine, 2003), but the polarization fraction of the scattered light exhibits a non-monotonic behaviour (Weingartner & Draine, 2001). Polarization by selective extinction by aligned dust grains has long been known to peak in  $\sim V$  band (e.g. Spitzer, 1978). By observing in four bands across the 0.9–2.0  $\mu\text{m}$  observer-frame wavelength range, and studying sources across a range of redshifts, WFIRST will be sensitive to both sources of polarization in multiple regimes.

### 3.3.1 Toy Model for Galaxy Polarizations

To simplify the calculation and make an order of magnitude estimation of the polarization effects, we begin by assuming that only edge-on disk galaxies are observed and that each edge-on galaxy has fixed polarized fraction  $p$ , which specifies the fraction of light from galaxy that is polarized. Although these assumptions will be relaxed by correction factors discussed in Sect. 3.3.2, this simplified model provides an easy and intuitive way to describe the galaxy polarization.

In the case of face-on disc galaxies, the polarization orientation varies with location in the galaxy, leading to the cancellation of optical polarizations in spiral and circular patterns (Scarrott et al., 1990; Simmons & Audit, 2000). Thus, for a simple order of magnitude estimate of the polarization-dependent selection bias and PSF, we consider only edge-on disc galaxies in this work. Fendt et al. (1996) reported the polarization of 3 edge-on galaxies, the optical polarization orientations of which are perpendicular to the major axes. However, it is also possible that the polarization orientation of an edge-on galaxy is parallel to the disk plane, as shown in Scarrott et al. (1990) and Scarrott (1996). The diversity of polarization orientation arises from two competing effects: the selective extinction of non-spherical dust grains aligned with the galactic magnetic field and the anisotropic scattering. The toroidal magnetic field on the disc plane would generate polarizations parallel to the major axis for edge-on galaxies (since the grains align perpendicular to the disc and preferentially absorb that polarization), while the polarization by anisotropic scattering is perpendicular to the major axes (since the grains are in a radiation field where more light is coming from directions parallel to than perpendicular to the disc). In Jones et al. (2012), the authors presented the integrated polarization survey of 70 nearby galaxies and suggested that the dust scattering is the dominant source of optical polarization. The magnitude of the polarization-dependence of the PSF is only affected by the polarization level rather than the polarization orientation, but the polarization orientation has a great impact on the selection bias. In our toy model, we consider only the polarization orientation that is perpendicular to the major axis of the edge-on galaxies. This will provide the upper limit for the polarization-induced selection bias.

The level of polarization is typically of order of several percent (Sofue et al., 1986; Scarrott et al., 1990; Draper et al., 1995; Scarrott, 1996; Jones et al., 2012). For edge-on galaxies, the anisotropic scattering (Jura, 1982) and alignment with the magnetic field (Fendt et al., 1996; Scarrott et al., 1990) under several  $\mu\text{G}$  of interstellar magnetic field both demonstrate polarization of order of 5%. The integrated polarization level would be less than the level in polarization maps in Fendt et al. (1996) and Scarrott (1996) due to partial cancellation of polarization vectors over the galaxy. The integrated polarization levels of 70 galaxies reported by Jones et al. (2012) are mostly below 1% with several above 2%. Hence, we choose the fraction of polarized light emitted by an edge-on disk galaxy to be 2%.

### 3.3.2 Corrections to the toy model

Here we fold in additional factors that were left out in our initial assumptions. We stated earlier the assumption that our sample consists solely of edge-on spiral galaxies. However, if we consider randomly-oriented disks, only a subset would be completely edge-on. We can parameterize the inclination  $i$  of the disk galaxies with the convention that  $i = 0$  corresponds to face-on galaxies while  $i = 90$  corresponds to edge-on disk galaxies. The probability of observing a galaxy within  $di$  of an inclination  $i$  is

$$P(i) di = \sin i di \quad (3.1)$$

for  $i$  between 0 and 90. This is already normalized, and hence we can calculate the expectation value of the inclination  $\langle i \rangle$ , which gives 1 rad. For face-on disk galaxies, the light we observe on a whole is unpolarized. If we use the  $\sin^2 i$  law for the dependence of polarization effect on inclination angle in Simmons & Audit (2000) and Jones et al. (2012), we can then say that the random inclinations of the disk galaxies results in  $\langle e_1 \rangle$  being multiplied by a factor of  $f = 0.42$ .

In addition, only about 50% of galaxies are spiral galaxies, as inferred from the visual classifications of galaxy morphology in the Great Observatories Origins Deep Survey (GOODS) fields (Giavalisco et al., 2004; Bundy et al., 2005) into three broad morphology classes (Ellipticals, Spirals, Peculiar/Irregulars). The 2978 galaxies used in the morphology classification by Bundy et al. (2005) are selected from the observations made by HST and the Advanced Camera for Surveys (ACS) with a magnitude limit in the  $z_{850}$  band  $z_{AB} = 22.5$ .

We take the spiral ratio of 50% as a rough estimate since the morphological composition is subject to the cosmic variance due to the small size of the GOODS field ( $0.1 \text{ deg}^2$ ). In addition, WFIRST will cover different wavelength bands, magnitudes and redshifts compared with ACS. However, this number will suffice for an order of magnitude estimation.

Hence, we multiply all results by a correction factor of  $f = 0.21$ .

## 3.4 Polarization dependent Optical response and PSF

In this section, we describe the mechanisms that induce the dependence of final image on polarizations.

### 3.4.1 Polarization Repsonce Difference

We start by introducing the polarization response difference  $\Delta_p$ . The polarization response difference  $\Delta_p$  reflects the difference in the detected flux on pixels due to the polarization-dependent response of the optical system and the detection efficiency along different polarization axes. The choice of coordinates and the axes is arbitrary. Without loss of generality, we define the  $x$ -axis as the axis along which the flux of polarized light is most attenuated. The polarization angle  $\theta$  is defined against this axis.

The polarization response difference  $\Delta_p$  due to the polarization-dependent throughput has two major sources: the anti-reflective coating of the detector and the fold mirrors. Since the actual coating model of the detector is not publicly available, we estimate the plausible levels of polarization-dependent transmission of it by considering the standard single layer anti-reflective coating. We build the coating model by setting the refractive index of the substrate to be 3.2 and tuning the refractive index of the coating to reproduce the ratio of the peak to the valley in the quantum efficiency curve of the WFIRST Wide-Field Instrument<sup>4</sup>. This model predicts  $\Delta_p = 0.03$  at the corner of the field. WFIRST’s mirrors use a protected silver coating. A bare silver fold flat would induce fractional polarization  $\Delta_p = 0.0014$  ( $0.9 \mu\text{m}$ ) or  $0.0017$  ( $1.93 \mu\text{m}$ )<sup>5</sup>, but coatings can significantly modify the reflection coefficients (see, e.g., Harrington & Sueoka 2017 for an analysis in the case of the Daniel K. Inouye Solar Telescope). The coatings for WFIRST at  $45^\circ$  incidence have  $\Delta_p$  ranging from 0.006 at  $0.5 \mu\text{m}$ , decreasing to near zero at  $1.1 \mu\text{m}$ , and then rising again to 0.005 at  $2.4 \mu\text{m}$ .<sup>6</sup> The first fold mirror (F1) is at a  $45^\circ$  angle and is expected to dominate the polarization, since there are much smaller angles of incidence for the second fold and the tertiary mirror. The filter may contribute too (it is at normal incidence at the center of the field, but the angle of incidence increases toward the edge), and so we plan to include it in a more complete model in the future. Our choice of  $\Delta_p = 0.05$  in Table 4.1 is the sum of the effects of the anti-reflection coating (0.03) and these three off-axis mirrors ( $3 \times 0.006$ , with the recognition that the true effect is smaller at most wavelengths) in the optical design of the WFIRST Wide-Field Instrument.

We note that linear retardance at the fold flat will result in some conversion of Stokes  $U \rightarrow V$ ; this phenomenon is an issue for instruments that aim to measure circular polarization. It is less of a problem here since the mechanism of polarization-dependent sensitivity (off-axis anti-reflection coating) depends on linear, not circular, polarization.

### 3.4.2 Polarization-Induced Astigmatism

When light is reflected by an optical device inside a telescope, such as the primary mirror, the reflected ray gains an extra reflectivity coefficient relative to the ray of incidence:

$$r = |r| e^{i\phi}.$$

The real part of the coefficient describes the relative amplitude of the reflected light, and the imaginary part describes the phase change. Both the real part and the imaginary part depend on the mirror coatings, wavelength, angle of incidence, and the polarization. The

<sup>4</sup>[https://wfirst.gsfc.nasa.gov/science/WFIRST\\_Reference\\_Information.html](https://wfirst.gsfc.nasa.gov/science/WFIRST_Reference_Information.html)

<sup>5</sup>This uses the index of refraction model of Johnson & Christy (1972) and the standard Fresnel coefficients for reflection at an interface (e.g., Born & Wolf, 1999), and we note that  $\Delta_p = |R_s - R_p|/(R_s + R_p)$ , where  $R_s$  and  $R_p$  are the power reflection coefficients for the two polarizations.

<sup>6</sup>We thank J. Kruk (private communication) for providing this information, and Harris Corporation for providing permission to publicly release the models.

phase change by the reflection is different for s- and p-polarizations. This polarization-dependent phase change induces the polarization variation as well as wavefront aberrations to the system. We define the phase shift between the two polarizations as retardance,

$$\delta = \delta_s - \delta_p.$$

We can estimate the magnitude of the polarization astigmatism by adding the retardance of each reflection fold.

The WFIRST technology report on polarization<sup>7</sup>, which is reviewed by the Technology Assessment Committee, provides the optical modeling and linear retardance at several wavelengths. Astigmatism has two spatial modes, vertical astigmatism (Zernike polynomial  $Z_2^2$ ) and oblique astigmatism ( $Z_2^{-2}$ ), which follow  $\rho^2 \cos(2\phi)$  and  $\rho^2 \sin(2\phi)$  patterns on the polar plane respectively with polar coordinate  $\rho$  and angular coordinate  $\phi$ . At 950nm, the linear retardance is 0.005 waves (rms) for vertical astigmatism and 0.006 waves for oblique astigmatism. For our toy model, we choose 0.005 waves as a conservative estimate for both vertical and oblique astigmatisms due to the linear retardance of the fast primary mirror in the Y band of WFIRST. The WFIRST Wide-Field Instrument contains 18 Sensor Chip Assemblies (SCA). The PSF varies as a function of position in the focal plane, including across each SCA. We use the central PSF for each SCA. In addition to spatial variation, the PSF is also dependent on polarization. The difference in linear retardance of 0.005 waves between the two polarization directions introduces extra aberrations on top of the default PSF.

## 3.5 Selection Bias

In this section, we describe how we use the ingredients in our toy model to estimate how polarization-related selection biases can affect weak lensing.

### 3.5.1 Defining a Galaxy Population

We begin by assuming that the galaxy angles are intrinsically uniformly distributed, and that galaxy magnitudes follow a distribution  $P(m)$ . To define  $P(m)$ , we use a simulated WFIRST photometry catalog based on CANDELS<sup>8</sup>. Fig. 3.1 shows the distribution of magnitudes in the J band. We then fit a rising exponential  $N \propto 10^{am}$  using magnitude cuts at 18 and 24 for each band to avoid the fall-off of counts near the depth limit.

The fitting curve is also shown in Fig. 3.1. The normalization of the number counts is arbitrary. It is the slope  $a$  that affects the selection bias, since it determines what fraction of the sample is sufficiently close to the boundary that it can be affected by polarization-dependent selection effects. For the magnitude distribution in the J band, the fitting within  $18 < m < 24$  yields  $a = 0.36$ .

<sup>7</sup>[https://wfirst.gsfc.nasa.gov/science/sdt\\_public/wps/references/PIAACMCstatus20161116.pdf](https://wfirst.gsfc.nasa.gov/science/sdt_public/wps/references/PIAACMCstatus20161116.pdf)

<sup>8</sup><https://github.com/WFIRST-HLS-Cosmology/Docs/wiki/Home-Wiki>

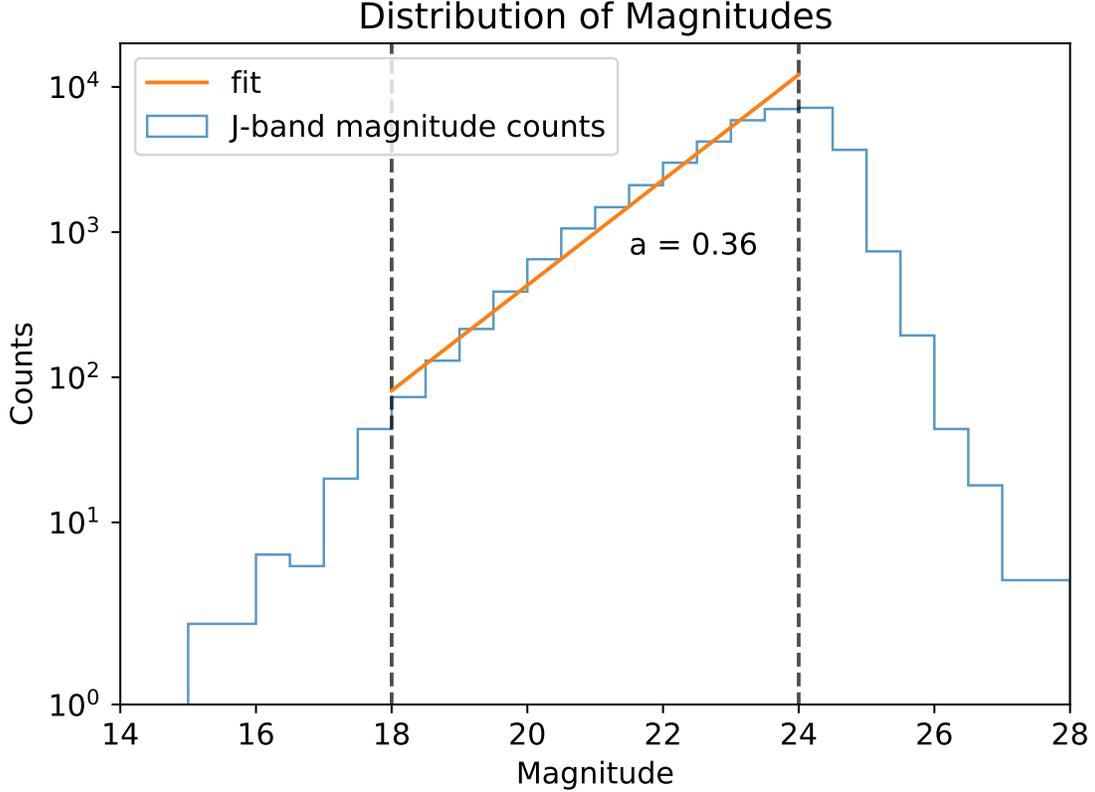


Figure 3.1: This figure shows the histogram of WFIRST magnitudes  $P(m)$  in J band (blue) and fitting curve (orange) of the simulated magnitude distribution to the exponential model  $N \propto 10^{am}$  with  $a = 0.36$ . The fits are valid within the magnitude range  $18 < m < 24$ . The vertical normalization of the number counts is arbitrary. The WFIRST photometry magnitudes for the four bands are simulated based on CANDELS catalogues. The dashed lines show magnitude cuts for the fits.

Additional corrections to the simplified toy model, for example to account for the distribution of galaxy inclination angles and the fraction of spiral galaxies, were introduced in Sect. 3.3.2.

### 3.5.2 Effect of Polarization on Magnitude

Following the definitions of  $p$  and  $\Delta_p$  in Sect. 3.4, we express the modification of the magnitude  $m$  of a galaxy due to polarization effect via the following equation:

$$m' = m - 2.5 \log_{10}[(1 - p) + p\sqrt{1 + (b^2 - 1) \cos^2 \theta}], \quad (3.2)$$

where the bias factor  $b$  is defined as  $b = (1 - \Delta_p)/(1 + \Delta_p)$ .

Since magnitude is logarithmically related to flux, the bias factor results in a linear shift in magnitude. Hence, the term inside the logarithm represents the factor by which the flux is reduced, and has a maximal value of 1. Since  $p$  represents the fraction of light that is polarized,  $(1 - p)$  corresponds to the unpolarized light. The square root term corresponds to the reduction in flux after the component of the flux of light polarized along the  $x$ -axis is multiplied by a factor of  $b$ , which ranges from 0 to 1. The logarithm term is negative, and hence  $m'$  is greater than  $m$ , i.e. the object appears fainter due to the polarization-dependent attenuation of light.

### 3.5.3 Magnitude Cut

We assume that galaxies are selected by imposing a magnitude cut  $m_{\text{cut}}$  on the galaxies such that galaxies fainter than the cut are filtered away. While this is not correct in detail (Zuntz et al., 2018; Mandelbaum et al., 2018), it is sufficient for an order-of-magnitude estimate. The consequence of modifying galaxy magnitudes  $m$  to the polarization angle-dependent  $m'$  is that galaxies with magnitudes near this cut could be moved to the other side of the cut. Since  $m'$  is a function of angle, the resultant population of galaxies that remain after the cut will no longer have uniformly distributed angles, violating a basic assumption in weak lensing and inducing a selection bias since an unlensed population would not have some nonzero mean ensemble shear estimate (Hirata et al., 2004; Mandelbaum et al., 2005).

We first consider the case where there is no selection bias, so  $b = 1$ . If we consider galaxies with a fixed polarization angle  $\theta$ , their magnitudes  $m$  should follow the distribution  $P(m)$ . Hence, when we apply the magnitude cuts to the galaxy population, we can write the probability of a remaining galaxy having the polarization angle  $\theta$  as:

$$P(\theta) \propto P(m_{\text{low}} < m < m_{\text{high}}) = \int_{m_{\text{low}}}^{m_{\text{high}}} P(m) dm. \quad (3.3)$$

The polarization-dependent response makes  $P(\theta)$  more complicated because  $m'$  is a function of  $\theta$ :

$$P(\theta) \propto P(m_{\text{low}} < m' < m_{\text{high}}) = \int_{m_{\text{low}} + 2.5 \log_{10}(c)}^{m_{\text{high}} + 2.5 \log_{10}(c)} P(m) dm, \quad (3.4)$$

where  $c = (1 - p) + p\sqrt{1 + (b^2 - 1)\cos^2\theta}$ .

### 3.5.4 Selection bias estimation process

We can write the expectation value of the galaxy orientation after making the cut as:

$$\langle \theta \rangle = \frac{\int_0^\pi \theta P(\theta) d\theta}{\int_0^\pi P(\theta) d\theta}. \quad (3.5)$$

Parameter	$p$	$\Delta_p$	$ \mathbf{e} $	$a$
Value	0.02	0.05	0.92	0.36

Table 3.1: Choice of variables for the toy model. The variables  $p$ ,  $\Delta_p$ ,  $|\mathbf{e}|$ , and  $a$  are the polarization fraction, the difference of response of polarization orientations, the absolute value of ellipticity and the slope of the fitted magnitude distribution. The polarization fraction is relevant to all polarization effects, including the selection bias and the PSF errors. The other variables,  $\Delta_p$ ,  $|\mathbf{e}|$  and  $a$ , are specific to the selection bias.

Note that  $\pi < \theta < 2\pi$  is degenerate with  $0 < \theta < \pi$  for the polarization angle. Substituting Eq. (3.4) gives us:

$$\langle \theta \rangle = \frac{1}{A} \int_0^\pi \theta \int_{m_{\text{low}}+2.5 \log_{10}(c)}^{m_{\text{high}}+2.5 \log_{10}(c)} P(m) dm d\theta, \quad (3.6)$$

where  $A$  is the normalization factor,

$$A = \int_0^\pi \int_{m_{\text{low}}+2.5 \log_{10}(c)}^{m_{\text{high}}+2.5 \log_{10}(c)} P(m) dm d\theta. \quad (3.7)$$

As a sanity check, for the case when  $c = 1$ , i.e. the case with no polarization-dependent selection effect, the inner integral evaluates to a constant  $k$  and the normalization factor is  $k\pi$ . Hence, we obtain  $\langle \theta \rangle = \frac{\pi}{2}$  as expected for a uniform distribution of angles from 0 to  $\pi$ .

To quantify the effect of the selection bias on weak lensing measurements, we relate it to the two-component ellipticity  $\mathbf{e} = (e_1, e_2)$ , which is an observable quantity used to construct ensemble shear estimates. Making an arbitrary choice of axis with no loss of generality, we present results for the first component  $\langle e_1 \rangle$ .

$$\begin{aligned} \langle e_1 \rangle &= \langle |\mathbf{e}| \cos(2\theta) \rangle \\ &= \frac{|\mathbf{e}|}{A} \int_0^\pi \cos(2\theta) \int_{m_{\text{low}}+2.5 \log_{10}(c)}^{m_{\text{high}}+2.5 \log_{10}(c)} P(m) dm d\theta. \end{aligned} \quad (3.8)$$

### 3.5.5 Parameters in the simplified toy model

We obtain an upper bound on the selection bias by choosing pessimistic conditions. The parameters that need to be specified and our values used are listed in Table 4.1.

In our analysis, we choose the polarization polarization to be 2% and estimate  $\Delta_p$  to be 5% along the attenuated polarization axis. We choose the absolute value of ellipticity  $|\mathbf{e}|$  of edge-on disk galaxies to be 0.92 based on the edge-on disk galaxy samples from the SDSS dataset (Kautsch et al., 2006). The ellipticity is defined based on the second moments of galaxy image light profile throughout the paper (3.11). The magnitude of ellipticity is

therefore  $|e| = (1 - b^2/a^2)/(1 + b^2/a^2)$  with major and minor axes  $a$  and  $b$  instead of  $|e| = 1 - b/a$  used in Kautsch et al. (2006). The quoted value above of the ellipticity of edge-on disk galaxies has taken into account the change of definition. As mentioned in Sect. 3.5.1,  $P(m)$  is obtained from the WFIRST simulated photometry catalog based on CANDELS. To do the experiment, we select galaxies with signal-to-noise ratio within  $18 < S/N < 200$ . This leads to the magnitude cuts  $(m_{\text{low}}, m_{\text{high}}) = (21.6, 24.2)$  in the J-band (Spergel et al., 2015), where we obtained a fit to the exponential magnitude distribution  $N \propto 10^{am}$  with  $a = 0.36$ .

## 3.6 PSF Errors

In this section, we describe how we use the polarization-induced aberrations introduced in Sect. 3.4.2 to simulate the PSF errors for WFIRST.

### 3.6.1 Image Simulation with Galsim

In this work, we use GALSIM<sup>9</sup> (Rowe et al., 2015) version 1.6.1 and the WFIRST module (Kannawadi et al., 2016) of GALSIM to simulate the polarization-dependent PSF effect for WFIRST. GALSIM is a package for simulating images of stars and galaxies. It can simulate galaxies from different galaxy models and also generate optical PSFs from parametric models. In particular, GALSIM has a module especially designed for the image simulations for WFIRST. With GALSIM, we simulate the PSFs for WFIRST observations for different bandpasses and Sensor Chip Assemblies (SCAs). Besides the default WFIRST aberrations, additional aberration (astigmatism) can be applied to test the polarization effect on PSFs.

### 3.6.2 PSF shape measurement

In order to estimate the impact of polarization on PSF errors, we measure the sizes and shapes of WFIRST PSFs with both default and polarization-dependent aberrations by considering the image of a point star. The moments of the star image are measured using the `hsm.FindAdaptiveMom` routine in GALSIM, which calculates the adaptive moments introduced by Bernstein & Jarvis (2002) and Hirata & Seljak (2003). This routine fits the 2-D image  $I(x)$  with an elliptical Gaussian by iteratively minimizing the squared error

$$SE = \int \left\| I(x) - A e^{-\frac{1}{2}(x-\mu)^T M^{-1}(x-\mu)} \right\|^2 d^2x. \quad (3.9)$$

---

<sup>9</sup><https://github.com/GalSim-developers/GalSim>

The best-fit  $\mu$  and  $\mathbf{M}$  are the weighted coordinates of the centroid and moments

$$\begin{aligned}\mu &= \frac{\int x I(x) \omega(x) d^2x}{\int I(x) \omega(x) d^2x} \text{ and} \\ \mathbf{M}_{ij} &= \frac{2 \int (x - \mu)_i (x - \mu)_j I(x) \omega(x) d^2x}{\int I(x) \omega(x) d^2x},\end{aligned}\tag{3.10}$$

with the elliptic Gaussian weight function  $\omega(x)$  that minimizes the squared error.

The ellipticity components and the adaptive size  $\sigma$  of the PSF are then defined in terms of the second moments:

$$\begin{aligned}e_1 &= \frac{\mathbf{M}_{xx} - \mathbf{M}_{yy}}{\mathbf{M}_{xx} + \mathbf{M}_{yy}}, \\ e_2 &= \frac{2\mathbf{M}_{xy}}{\mathbf{M}_{xx} + \mathbf{M}_{yy}}, \text{ and} \\ \sigma &= \sqrt[4]{\det \mathbf{M}}.\end{aligned}\tag{3.11}$$

Please note that the ensemble average of this definition of ellipticity based on the second moments of the light profile, when applied to galaxies, is not an unbiased estimator of the weak lensing shear. For nearly round objects,  $\langle e \rangle \approx 2\langle \gamma \rangle$ . The requirements on our knowledge of PSF model ellipticity include the proper conversion factors to account for how systematic uncertainty in PSF model ellipticity (for this ellipticity definition) affects estimates of ensemble weak lensing shears for galaxies.

In the image simulation, we oversample the image below the WFIRST pixel scale (0.11 arcsec/pix) and down to 0.01 arcsec/pix to obtain convergent adaptive moments and adaptive sizes. The detector non-idealities, including the reciprocity failure, non-linearity and interpixel capacitance etc., are not included. We only consider the extra aberrations of PSFs due to polarization.

## 3.7 Results

In this section, we present the estimated polarization-induced systematic biases using the toy model described in Sect. 3.5 and Sect. 3.6. We show the results for both the selection bias and the PSF errors.

### 3.7.1 Selection bias

Assuming an exponential magnitude distribution, we can calculate the expected mean selection bias in the shear. Given our fit for the exponential magnitude distribution in Sect. 3.5.1, we can substitute into Eq. (3.8) and simplify the expression. We first evaluate

the inner integral,

$$\begin{aligned}
P(\theta) &\propto \int_{m_{\text{low}}+2.5\log_{10}(c)}^{m_{\text{high}}+2.5\log_{10}(c)} P(m) dm \\
&\propto [(1-p) + p\sqrt{1+(b^2-1)\cos^2\theta}]^{2.5a}.
\end{aligned}
\tag{3.12}$$

Going back to Eq. (3.8),

$$\begin{aligned}
\langle e_1 \rangle &= \langle |\mathbf{e}| \cos(2\theta) \rangle \\
&= \frac{|\mathbf{e}| \int_0^\pi (\cos 2\theta) [(1-p) + p\sqrt{1+(b^2-1)\cos^2\theta}]^{2.5a} d\theta}{\int_0^\pi [(1-p) + p\sqrt{1+(b^2-1)\cos^2\theta}]^{2.5a} d\theta}.
\end{aligned}
\tag{3.13}$$

This gives  $\langle e_1 \rangle = 0$  when the bias parameter is 1, as expected. For a given set of variables  $p$  and  $\Delta_p$ , we compute this integral numerically to determine the selection bias due to polarized light profiles of edge-on disk galaxies. We note that the prefactor involving the magnitude cuts in Eq. (3.12) gets cancelled in Eq. (3.13), indicating that the expected shear component of the toy model does not depend on the magnitude cuts. In addition, the bright end limit of the magnitude cut contributes less than 10% to the total selection effect. Setting the lower limit of the magnitude cut to negative infinity is not expected to make significant difference for the order of magnitude test.

Using Eq. (3.13) and  $|\mathbf{e}| = 0.92$ , we vary  $p$  and  $\Delta_p$  and numerically evaluate the integral, giving us the results in Fig. 3.2. The additional correction factors, including the galactic inclination angle and the fraction of spiral galaxies, discussed in Sect. 3.3.2, are taken into account in this figure. For the values listed in Table 4.1 of  $p$  and  $\Delta_p$ , we obtain the estimate of  $\langle e_1 \rangle = 8 \times 10^{-5}$ . This is represented by the white point in Fig. 3.2.

### 3.7.2 PSF size and shape errors

To estimate the effect of light polarizations on PSF, we run the image simulation we describe in Sect. 3.6. In Fig. 3.3 we show the changes in  $e_1^{\text{PSF}}$ ,  $e_2^{\text{PSF}}$ , and the adaptive size  $\sigma$  due to the extra polarization-induced aberration measured across the 18 SCAs.

Table 3.2 shows the PSF errors in the Y band across the SCAs with all the factors above taken into consideration. The effect of the extra aberrations depends on the PSF of the SCAs, which is related to the positions of SCAs in the field of view.<sup>10</sup> For the WFIRST HLS weak lensing survey, the tolerance on the relative error of trace of the second moments of the PSFs ( $\delta\text{tr}(M)/\text{tr}(M) = 2(\delta\sigma)/\sigma$ ) is  $9.3 \times 10^{-4}$  and the required knowledge of PSF ellipticity is  $4.7 \times 10^{-4}$  per component (Spergel et al., 2015; Kannawadi et al., 2016). Even though the estimated PSF errors, including the error of the trace and the error of PSF ellipticity components, are within the WFIRST tolerance, the polarization-induced PSF errors alone could account for around 20% of the error budget, indicating that more investigations into polarization-related PSFs beyond this simple toy model are required.

<sup>10</sup>Please see the WFIRST observatory reference information at <https://wfirst.gsfc.nasa.gov/>

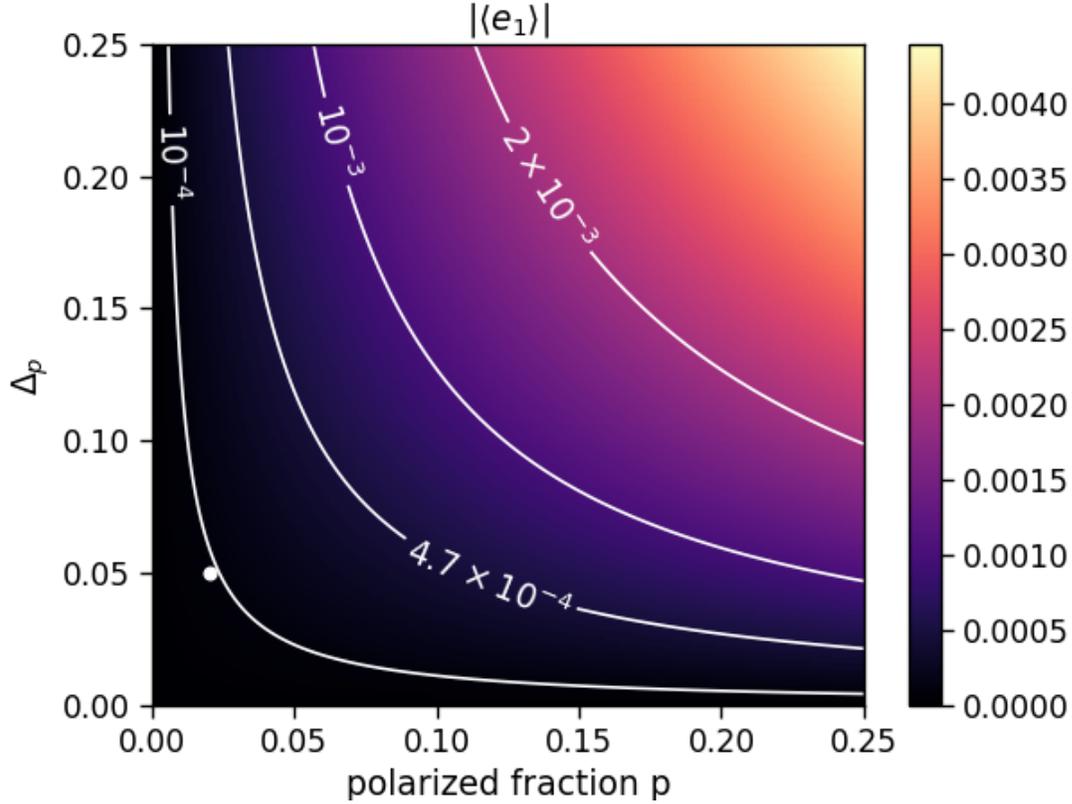


Figure 3.2: The expected ensemble mean shear  $\langle e_1 \rangle$  component due to the selection bias for various choices of variables  $p$  and  $\Delta_p$ . In the ideal case without a polarization-dependent optical response,  $\langle e_1 \rangle$  should be zero. At our choice of parameter in Table 4.1 (the white dot at the bottom left corner), the bias in  $\langle e_1 \rangle$  is  $\sim 8 \times 10^{-5}$ .

		PSF error ( $\times 10^{-4}$ )			Selection bias ( $\times 10^{-4}$ )	WFIRST tolerance ( $\times$ )
		worst	mean	best		
Size Bias	$\delta \text{tr}(M)/\text{tr}(M)$	2.0	-0.75	-0.067	-	9.3
Ellipticity Bias	$\delta e_1$	-1.1	-0.48	0.024	0.8	4.7

Table 3.2: This table summarizes the estimates of systematic errors due to polarization effects in our toy model. The columns show the (worst, mean, best) PSF errors due to extra polarization-induced vertical astigmatism among SCAs in Y band, the selection bias, compared to the WFIRST tolerance in the last column.

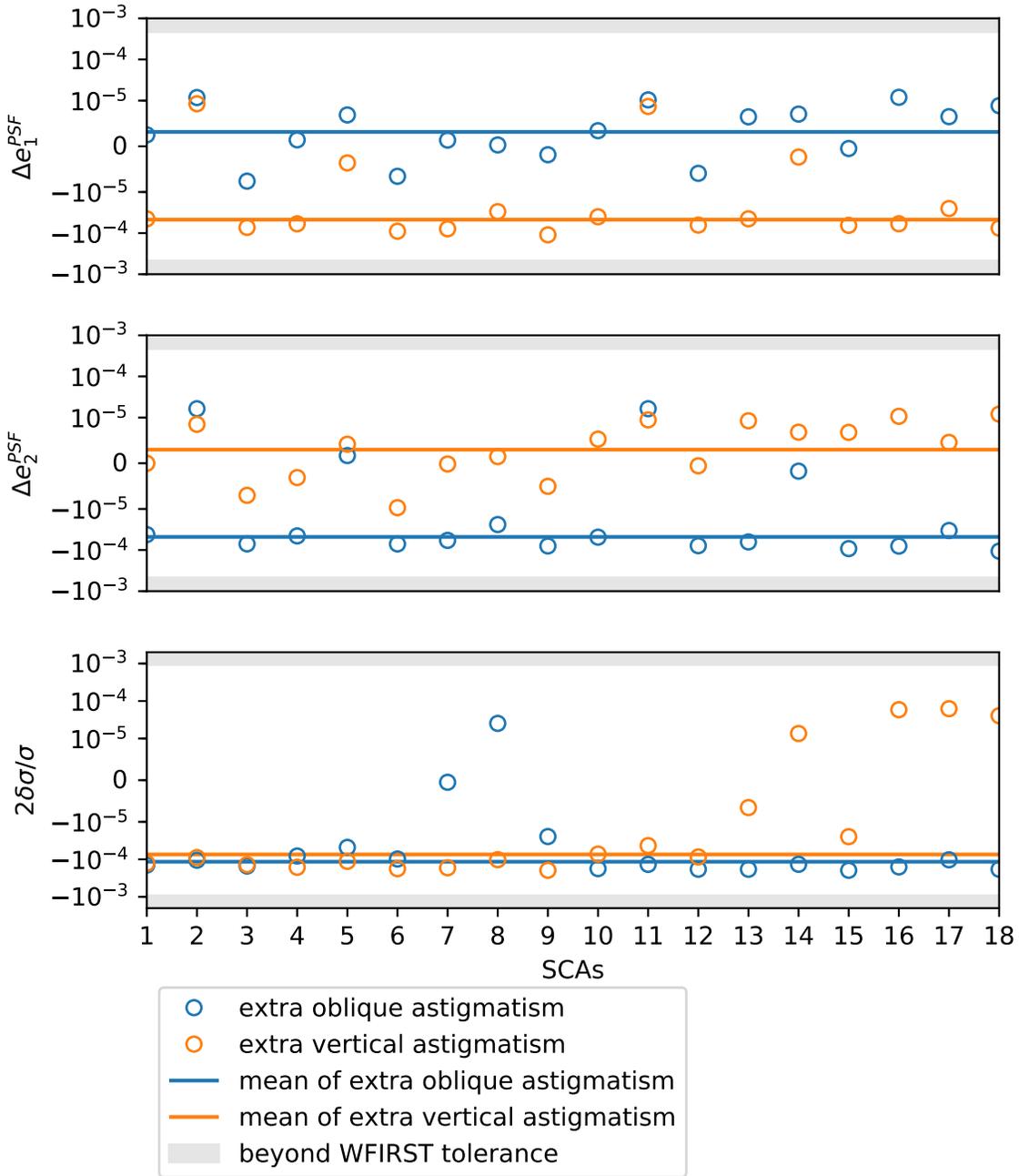


Figure 3.3: Changes in PSF ellipticities  $e_1^{\text{PSF}}$  and  $e_2^{\text{PSF}}$  and the fractional change in the adaptive size  $2(\delta\sigma)/\sigma$  due to the extra 0.005 oblique and vertical astigmatism on top of WFIRST PSF models in the Y band. The dots show the changes in each SCA, and the average offset across 18 SCAs is labelled by the solid horizontal lines. The 2% linear polarization and the additional factors in Sect. 3.3.2 are taken into account. The shaded regions are beyond the WFIRST tolerance on the ellipticity components (first two rows) and the tolerance on the relative error of trace of the second moments of the PSFs (the third row). The polarization-induced PSF errors could account for around 20% of the error budget.

### 3.8 Discussion

In this work, we constructed a toy model to estimate the magnitude of polarization-driven systematic errors in the context of optical/NIR weak lensing surveys for the first time in the literature. We propose a simple toy model to make order of magnitude estimates for both selection bias and PSF errors due to the polarization effects. The selection bias due to the polarization-dependent optical efficiency can lead to a spurious mean shear estimate of  $|\langle e_1 \rangle| \approx 8 \times 10^{-5}$ . For the PSF model error, our image simulations with GALSIM in the WFIRST Y band give PSF model size and shape biases of  $\delta \text{tr}(M)/\text{tr}(M) = 2.0 \times 10^{-4}$  and  $\delta e_1^{\text{PSF}} = 1.1 \times 10^{-4}$ , compared to the WFIRST requirements on the entire PSF  $\delta \text{tr}(M)/\text{tr}(M) = 9.3 \times 10^{-4}$  and  $\delta e_{1/2}^{\text{PSF}} = 4.7 \times 10^{-4}$ .

There are several caveats to our toy model. The purpose of this work is to give a rough estimate of the polarization effects on weak lensing measurements, including the selection bias and PSF errors. Hence we adopt a simplified galaxy selection by magnitude cuts only, and use a power-law model for the galaxy number counts. We also make assumptions about the light polarization of galaxies, the dependence on galaxy inclination angle, the orientation of polarizations and the dependence of polarization fraction on rest frame wavelength. Some assumptions help to simplify the estimation, and some are unfortunately necessary due to the limited number of optical/NIR observations of polarizations available and/or due to what information about the WFIRST instrumentation vendors have released.

For WFIRST, the key priority will be to obtain enough measurements on both the telescope/instrument and the source galaxies to remove biases associated with polarization. The polarized detector response at non-normal incidence can be measured in the laboratory with a representative detector, and we are in discussions with the WFIRST Project Office about the best way to implement this measurement. We are also having discussions about what relevant information on the optics can be released. Given how faint the galaxies are, a promising way to obtain the required information on the galaxies is from WFIRST itself. In the WFIRST supernova deep fields, the same set of galaxies in a  $5 \text{ deg}^2$  area are observed at  $\sim 144$  epochs over 2 years (Spergel et al., 2015). In the Spergel et al. (2015) deep field strategy, the aggregate S/N for a typical lensing source (AB=24.5) is 250, and at 3 galaxies per arcmin<sup>2</sup> per redshift bin, there would be 54,000 galaxies per bin in a supernova deep field.<sup>11</sup> The supernova fields are located near the Ecliptic poles. When observing the supernova fields, the telescope roll angle is adjusted to point its solar array toward the Sun for power. The roll angle would therefore go through a cycle every year. If there is a polarization-dependent throughput and the galaxies are polarized then we expect to see their magnitudes vary sinusoidally with a period of 6 months and a phase that corresponds

---

[science/WFIRST\\_Reference\\_Information.html](#) for the arrangement of SCAs and the field of view layout.

<sup>11</sup>This calculation uses the WFIRST Exposure Time Calculator (Hirata et al., 2012), following Hounsell et al. (2018), but updating with the current WFIRST Phase B throughput table, although this makes only a small difference.

to their position angle. The amplitude of this variation<sup>12</sup> is  $\approx \Delta_p \times p$ . WFIRST will not measure this for any one galaxy, but if  $f\Delta_p p$  is really  $\sim 0.00021$  (where  $f = 0.21$  is the suppression factor in §3.3.2), then the idealized aggregate S/N ratio on the measurement of the polarization effect is

$$\frac{1}{\sqrt{2}} \times 0.00021 \times 250 \times \sqrt{5.4 \times 10^4} = 8.6. \quad (3.14)$$

An advantage of this is that it directly measures the combination  $\Delta_p \times p$  that is of direct interest, even without depending in the detailed model for instrument polarization  $\Delta_p$ . Future work should consider how to make this measurement practical, particularly when incorporating large-scale flat field uncertainties and dithering strategies.

In this paper, the biases due to polarization-dependent throughput and PSF are both estimated with WFIRST-specific parameters. The polarization biases for surveys with different optical design, targeted wavelength and point spread function would be different. In addition, most weak lensing surveys to date are ground-based and thus subject to atmospheric turbulence. It is nontrivial to infer the impact of polarization effect for other surveys without tailored analysis. But for WFIRST, both of the polarization-induced errors are comparable to the WFIRST tolerance. We thus recommend more complete studies on polarization-induced systematic uncertainties of weak lensing measurements.

---

<sup>12</sup>Defined as half of the peak-to-valley difference; the root-mean-square is a factor of  $1/\sqrt{2}$  smaller.

# Chapter 4

## Image Persistence Effect on Weak Lensing

### 4.1 Abstract

The High Latitude Survey of the Roman Space Telescope is expected to measure the positions and shapes of hundreds of millions of galaxies on an area of  $2220 \text{ deg}^2$ . This will push the weak lensing science further with its high quality data and unprecedented systematics control. The Roman Space Telescope will survey the sky in the near infrared (NIR) band using its specially designed HgCdTe photodiode arrays called H4RG. For the NIR arrays, charges that are trapped in the photodiodes during earlier exposure are gradually released into the current exposure, leading to contamination of the images and errors of the galaxy shapes. This memory effect is called persistence. In this work, we use image simulations that incorporate the persistence to study its impact on galaxy shapes and weak lensing signal. We analyze the shape errors due to persistence. No significant correlation between the shape shift by persistence is detected, indicating that persistence does not introduce coherent shape distortions on galaxies. In the scales of interest, the effect of persistence is about two orders of magnitude lower than the Roman Space Telescope additive shear error budget, indicating that the persistence is expected to be a weak effect on weak lensing for Roman Space Telescope given the current design.

### 4.2 Introduction

Weak gravitational lensing arises due to deflection of light by the gravitational fields of large-scale structure, leading to tangential shear distortions in galaxy shapes. Measuring the correlation functions of galaxy shapes is therefore a method by which we can measure the growth of structure in the Universe (Bartelmann & Schneider, 2001; Massey et al., 2007; Hoekstra & Jain, 2008; Kilbinger, 2015; Mandelbaum, 2018) and hence a powerful method for constraining cosmological parameters (Huff et al., 2014; Jee et al., 2016; Hildebrandt

et al., 2018; Troxel et al., 2018b; Hikage et al., 2019a).

Since weak gravitational shear is only a percent-level signal, weak lensing measurements rely on the use of large galaxy samples to reduce statistical uncertainties. In the upcoming Stage-IV surveys, including Euclid<sup>1</sup> (Laureijs et al., 2011), LSST<sup>2</sup> (Ivezić et al., 2008; LSST Science Collaboration et al., 2009), and Roman Space Telescope<sup>3</sup> (Spergel et al., 2015), the statistical uncertainties of the weak lensing measurements are expected to reach sub-percent level precision. With such small statistical uncertainties, the future of weak lensing analysis requires a better understanding and more careful control of systematics to avoid systematic uncertainties dominating over statistical uncertainties.

In this work, we investigate the new source of weak lensing systematics due to persistence effect of infrared detectors. The Roman Space Telescope will survey the sky in the near infrared (NIR) band using its specially designed HgCdTe photodiode arrays called H4RG. For the NIR arrays, charges that are trapped in the photodiodes during earlier exposure are gradually released into the current exposure, leading to contamination of the images and errors of the galaxy shapes. This persistence signal from previous exposures would then contaminate the images of galaxies and impact the galaxy shapes.

Smith and Zavodny (Smith et al., 2008a,b) were the first to hypothesize that image persistence is due to traps in the depletion region of the diode. Since that time, the model has been largely adopted by the community, and much work has been done on characterizing traps in infrared detectors. Long et al. (2012) provided an initial description of persistence in the WFC3 detector based on several calibration programs. Persistence is a strong function of the fluence of the source in the original image. For pixels that were saturated, the current due to persistence exceeds the dark current for several exposures. Anderson and Regan (Anderson et al., 2014) used electrical stimulation of the detector to map out the physical location of traps within each pixel. Illumination tests done by Regan et al. (2012) provided data supporting the theory that these traps are also at least partially responsible for reciprocity failure (also known as count rate nonlinearity), though Biesiadzinski et al. (2011) showed that the traps cannot entirely explain reciprocity failure.

Modeled behavior of image persistence has been used to plan observations and place operational restrictions on what astronomical sources may be observed, and the ultimate goal is to model persistence accurately enough so that it may be corrected for in astronomical images. The correction for image persistence of WFC3cam on Hubble (Long et al., 2012) is based on the power-law decay model with a limited range of time validity, and a user who suspects image persistence is contaminating their data must request that a persistence map be specifically created for their data, requiring human intervention for each dataset that must be corrected. WFC3cam users also have access to a MAST tool that warns of any preceding images that may have left persistence signatures.

Although some efforts have been done on modeling the image persistence effect and also the correction scheme, no published work so far is dedicated to study the impact of

---

<sup>1</sup><http://sci.esa.int/euclid/>

<sup>2</sup><http://www.lsst.org/lsst>

<sup>3</sup><https://wfirst.gsfc.nasa.gov>

image persistence in the context of weak lensing. In this work, we will use image simulations that incorporate the image persistence model to study its impact on galaxy shapes and weak lensing signal. In Section 4.3, we outline the theory behind image persistence and the model to be used in the simulations. Section 4.4 describe the key components in the image simulations. We present our results in Section 4.5 and conclude in Section 4.6.

## 4.3 Persistence Effect

### 4.3.1 Theory of trapped charge carriers

Persistence is the effect of an increase in the dark current due to prior illuminations. For the near-infrared HgCdTe photodiode detector, the hypothesized mechanism of image persistence involves charge capture and charge emission in the PN junctions (Smith et al., 2008a; Tulloch et al., 2019; Anderson et al., 2014). As the PN junctions are exposed to radiation illuminations, the photo-generated charges, including electrons and holes, start to accumulate, leading to the reduction of the width of the depletion region. Resetting the device to a resetting voltage recovers the depletion region, while some charges are still trapped in defects and are left behind. The trapped charges will be continuously release. The delayed charge emission would thus contaminate the following exposures. The magnitude of the persistence effect, usually quantified in terms of charges released per second (e-/s), correlates strongly with the illumination level. As the illumination approaches the saturation level of the detector, more defects are filled and this effect gets stronger.

### 4.3.2 Persistence model

Even though the hypothesized mechanism of charge capture is widely accepted to explain persistence, a proper physical theory to predict the behavior of persistence has not yet been developed. At the current stage, we rely on empirical models to describe the effect. An empirical exposure time dependent Fermi model was adopted to model the Hubble WFC3 IR curve. The persistence effect of H4RG detectors is approximately one order of magnitude weaker than that of WFC3 around the saturation level, but the Fermi model still serves as a good fitting law for H4RG persistence. This model follows the mathematical form:

$$P(x, t) = A \left( \frac{1}{e^{\frac{-(x-x_0)}{\delta x}} + 1} \right) \left( \frac{x}{x_0} \right)^\alpha \left( \frac{t}{1000} \right)^{-\gamma}, \quad (4.1)$$

with illumination level  $x$ , time after reset  $t$  and other fitting parameters.

$A$	$x_0$	$\delta x$	$\alpha$	$\gamma$
0.017	$6.0 \times 10^4$	$5.0 \times 10^4$	0.045	1

Table 4.1: The fitting parameters of the exposure time dependent Fermi model as described in Eq. 4.1.

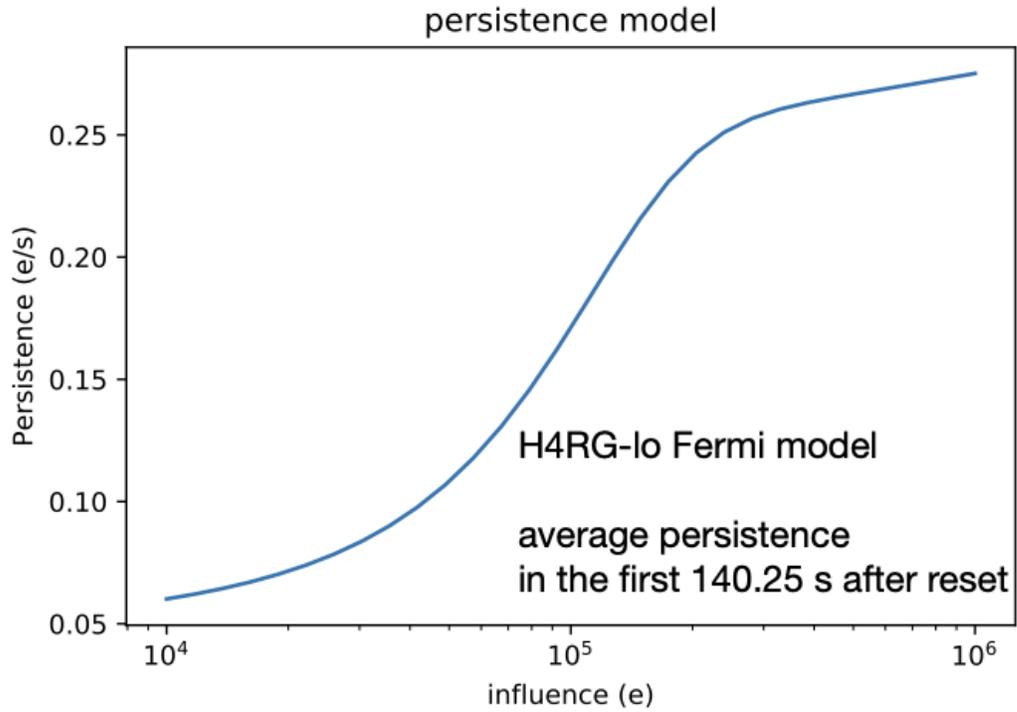


Figure 4.1: The curve of the H4RG-lo persistence model with parameters specified in table 4.1. This curve shows the average of persistence current during the entire first 140.25 seconds after reset for various influence levels.

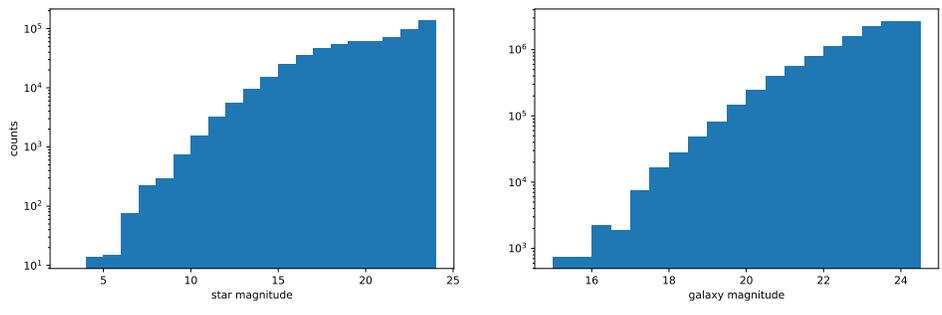


Figure 4.2: Magnitude distribution of stars and galaxies in an area of 100 deg<sup>2</sup>. Left: the magnitude distribution of the simulated Galaxia stars in J band. Right: the simulated J-band Photometry of Roman Space Telescope used in the galaxy catalog.

## 4.4 Simulation

In this work, we use the simulation suite developed by Troxel et al. (2019). The persistence effect is applied along with other detector effects to test the impact of persistence on galaxy shapes. We focus on the persistence due to bright stars. Here we highlight the key elements of the simulation. We refer readers to Troxel et al. (2019) for more details.

### 4.4.1 Galsim

The image simulations in the work are carried out using the GALSIM<sup>4</sup> package Rowe et al. (2015). GALSIM is a package for simulating images of stars and galaxies. It can simulate galaxies from different galaxy models and also generate optical PSFs from parametric models. In particular, GALSIM has a module especially designed for the image simulations for Roman Space Telescope (Kannawadi et al., 2016). With GALSIM, we simulate the images of star and galaxies for different bandpasses and Sensor Chip Assemblies (SCAs).

### 4.4.2 Galaxy catalog

The input galaxy catalog was generated by using random spatial locations in the  $100\text{deg}^2$  sky area with a galaxy number density of  $40\text{arcmin}^{-2}$ . The photometric properties of the galaxies are drawn from a simulated Roman photometry catalog based on the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) survey<sup>5</sup> (Hemmati et al., 2019). In figure. 4.2, we illustrate the distribution of photometries of the simulated photometry catalog in the J-band.

### 4.4.3 Implementation of persistence effect

We adopted the implementation of the persistence effect in the Roman Space Telescope module of GALSIM, which follows the model introduced in Sect. 4.3.2. The parameters in the Fermi model are fitted based on lab characterization of sample H4RG detectors ("FIX ME", cite source). For pixels with influence below half saturation well, which is equivalent to  $5 \times 10^{-4}e^-$ , we choose a linear model instead due to lack of complete measurement of persistence at low fluence.

### 4.4.4 simulation steps

#### Building truth catalogs

Before images are simulated, truth catalogs of objects to be simulated, including stars and galaxies, are constructed. The position and magnitudes of stars follow the input catalog.

---

<sup>4</sup><https://github.com/GalSim-developers/GalSim>

<sup>5</sup><https://github.com/WFIRST-HLS-Cosmology/Docs/wiki/Home-Wiki>

For the galaxy truth catalog, besides the position and photometric properties, each galaxy is assigned an intrinsic ellipticity drawn from a Gaussian distribution  $G(0, 0.27)$  with a random orientation. No additional gravitational shear is applied on top of intrinsic ellipticities.

## Image simulation

The image simulation is implemented using the simulation suite developed by Troxel et al. The simulation suite simulated the star and galaxies in the input catalog and generated SCA images on  $4088 \times 4088$  pixels stamps. Effects including sky background, reciprocity failure, electron quantization, dark current, nonlinearity, interpixel capacitance and readout noise are included in the image generation, following the order that each effect physically occurs. Besides these effects, we apply persistence from previous exposures using the model introduced in Sect. 2.3. Since the persistence decays with the resetting time, in the simulation process we limit the number of past exposures considered for persistence to 10 to facilitate the simulation process. For each galaxy, images with and without the persistence contamination are simulated to identify the impact.

## Shape measurement

The shape measurement is done by fitting the galaxy light profile to an exponential model. The fitting is carried out using the Gaussian mixture fitting module NGMIX<sup>6</sup> with 6 fitting parameters, including the positions  $x, y$ , the ellipticities  $e_1, e_2$ , the size and the flux. Two sets of simulations are run with persistence effect included in one and switched off in the other. The impact on galaxy shapes due to persistence is then captured in the difference of ellipticities of the two sets of simulations.

# 4.5 Results

## 4.5.1 Persistence effect in simulations

To build the intuition about how persistence is going to affect the shapes of galaxy, we start with showing the image of persistence of bright stars which are the major source of persistence in the sky. Fig. 4.3 shows the persistence of bright stars with difference magnitudes at the time of 140s after reset using the model and parameter specified in Eq. 4.1 and Tab. 4.1. At saturating pixels, the persistence current plateaus at around  $0.25 e^-/s$  (Fig. 4.1), leading to persistence signals of several dozens of electrons in a 140s exposure as at centers of the four panels of Fig. 4.3. The image persistence due to bright stars is comparable to images of  $m = 22 - 24$  galaxies (Fig. 4.3) in terms of brightness. When the two blend, the shapes of galaxies could be greatly affected. These persistence images also illustrate the obscuration and the number of struts of the pupil plane. If a galaxy happens to overlap on top of the radial spike of the persistence, it would introduce directional changes

---

<sup>6</sup><https://github.com/esheldon/ngmix>

to the shape, as shown in Fig. 4.4. In Fig. 4.4, we show an image of a magnitude 22 galaxy and the persistence of a magnitude 7 star to demonstrate the relative contribution of the two components.

### 4.5.2 Correlation between shape errors of galaxies

To test whether the persistence effect introduces coherent distortions to galaxy shapes, we measure the correlation of the shape shift  $\Delta e$  due to persistence. In the simulation, a total number of 2.2 million galaxies are simulated, with around 1% of them having  $\Delta|e| \geq 1.0^{-8}$  and around 0.5% having  $\Delta|e| \geq 1.0^{-4}$ . We use TREECORR<sup>7</sup> to measure the two point correlation functions of ellipticities  $\Delta e_1$  and  $\Delta e_2$  (Jarvis et al., 2004). In measuring the correlation, all the galaxies, no matter affected by persistence or not, are taken into consideration. The result is shown in Fig. 4.5. The error bars are estimated using the jackknife method with 20 patches over the area. We do not detect significant coherent shifts in ellipticities due to persistence of bright stars. Even though there are cases where strong persistence greatly affect the shapes of galaxies, the total averaging effect is low. Hence, it is possible to treat the impact of persistence on galaxy shapes as an uncorrelated noise. To quantify the significance of persistence effect on galaxy shapes, we compute the correlation functions of the Roman total additive shear systematic error budget (Doré et al., 2018; Troxel et al., 2019) in the same scale as the comparison baseline. The level of  $\xi_+$  of persistence is at least one order of magnitude lower than the Roman additive shear bias in the scales considered.

## 4.6 Conclusions

In this work, we study the impact of image persistence, especially the image persistence of bright stars, on weak lensing signals in the context of Roman High Latitude Survey. The images are simulated using the Roman image simulation suite in Troxel et al. (2019) and the image simulation software GALSIM (Rowe et al., 2015). A total number of 2.2 million galaxies over an area of  $100\text{deg}^2$  are simulated in hundreds of dither positions. In the simulations, detector effects relevant to Roman Space Telescope are incorporated, but we put our focus on the persistence effect. An empirical persistence model of the H4RG near-infrared photodiode detector is applied to model the image persistence of earlier exposures.

The Roman Project has put a lot of efforts into the design of H4RG flight detector in order to minimize the image persistence. For weak lensing, even though no mitigation method of persistence is used in our simulations to remove the image persistence effect from earlier exposures, we do not expect significant impact on galaxy shapes for weak lensing sciences given the current persistence performance of the detectors. The correlation functions of  $\Delta|e|$  due to persistence do not show strong correlation between the galaxies.

---

<sup>7</sup><https://github.com/rmjarvis/TreeCorr>

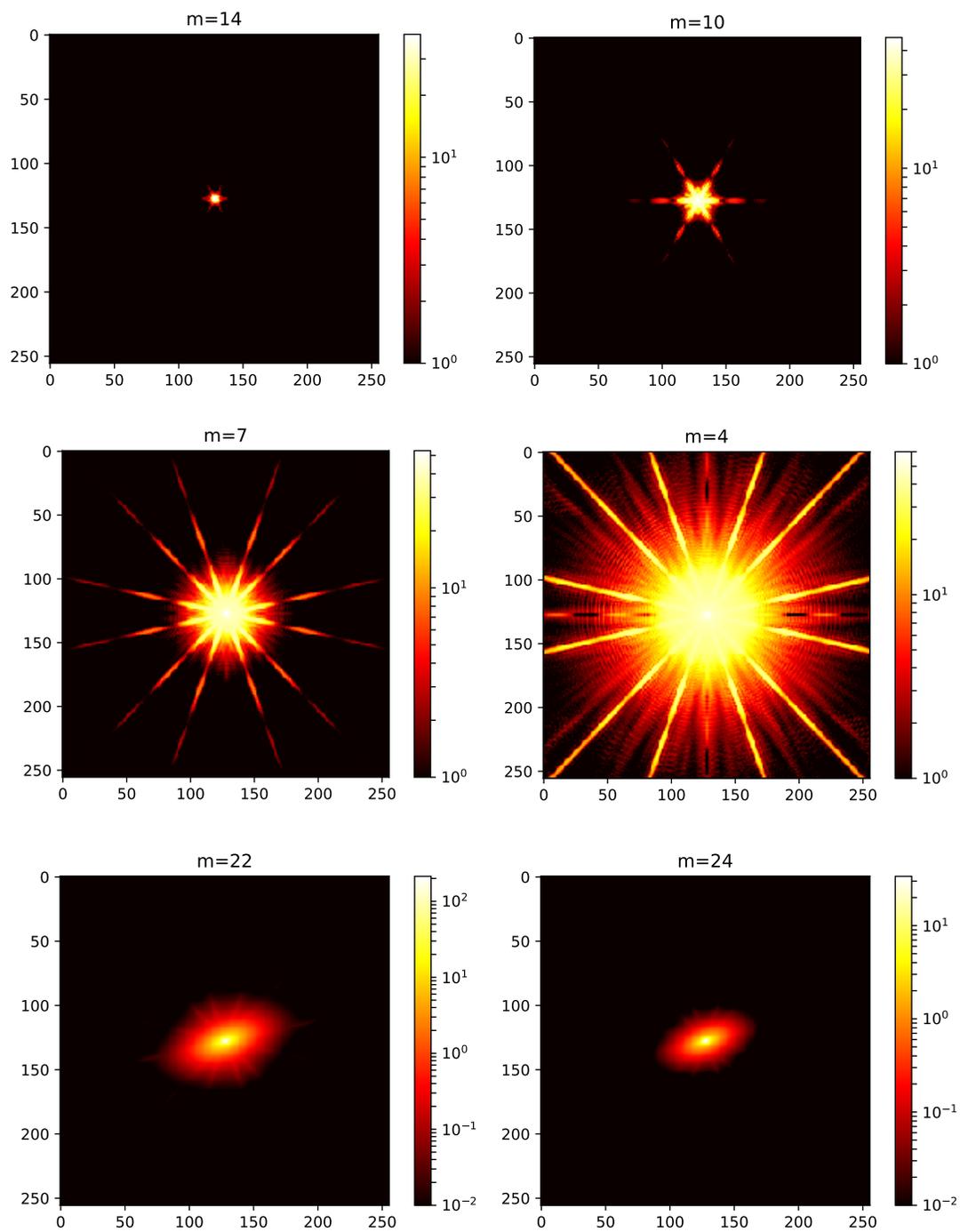


Figure 4.3: Top two rows: images of persistence of stars with different magnitudes taken in 140s exposures after reset. Bottom row: simulated images of  $m = 22$  and  $m = 24$  galaxies captured by WFIRST in 140s exposures. The images are drawn on 256x256 images stamps with x and y axes indicating the pixel numbers using the Roman pixel scale. The colors are in units of electrons with color maps on the right of each panel.

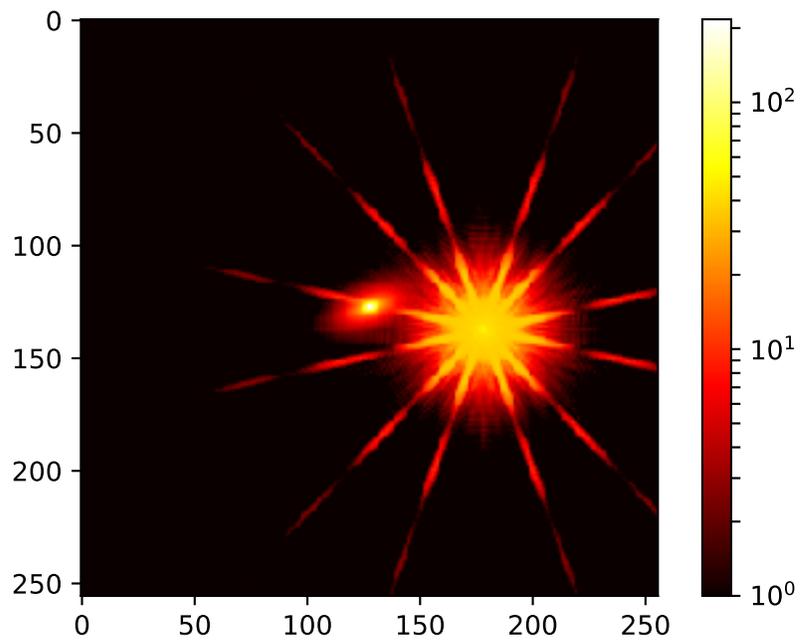


Figure 4.4: Illustration of the impact of star persistence on galaxy shapes. In this figure, a magnitude 22 galaxy and the persistence of a magnitude 7 star are plotted. In this case, the two objects are located 51 pixels away from each other to illustrate the relative flux and size of the two components.

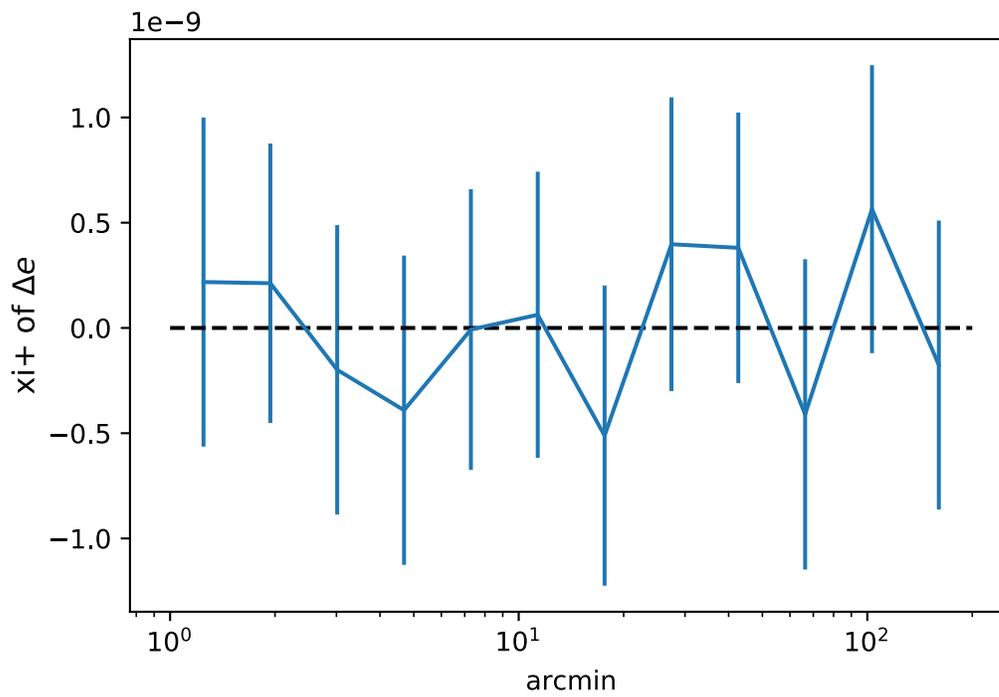


Figure 4.5: The two-point correlation function  $\xi_+$  of the change in galaxy ellipticity due to persistence effect from previous exposures.

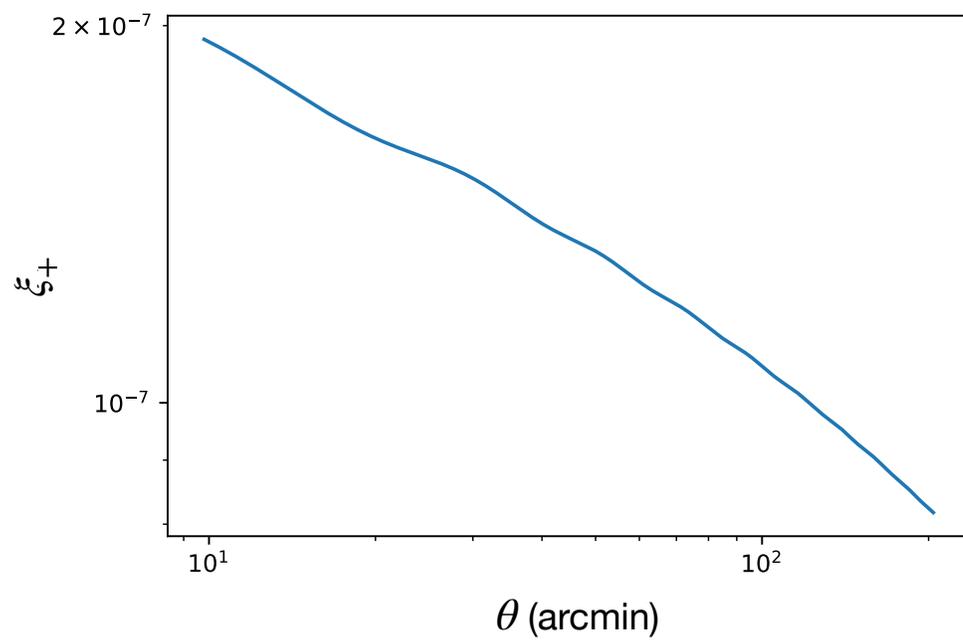


Figure 4.6: The two-point correlaion function  $\xi_+$  of the shear additive bias.

A comparison against the Roman total additive shear error budget further indicates that the persistence is a weak effect on weak lensing.

We would like to note that in this work we focus only on the persistence effect of bright stars. Future work may determine the coherent shape shift due to the persistence during the telescope slews.

# Chapter 5

## Conclusion and Outlook

Weak gravitational lensing is a powerful way to probe the large-scale structure and expansion history of the universe, and thus provides hints to the mysteries of dark matter and dark energy. Despite the potential, the systematic uncertainty of the weak lensing measurements must be well controlled so as to extract the lensing effect and to draw credible conclusions about the nature of dark energy and dark matter. Chapters 2-4 deal with three different sources of systematics uncertainties. The systematic errors considered in chapters 2-4 are subdominant sources to the error budget. This dissertation focuses on testing the significance of the subdominant systematic uncertainties in the context of weak lensing surveys in 2020s. In chapter 2, we present our study on the non-Gaussian weak lensing likelihood functions and its impact on the cosmological parameters. Chapter 3 presents our study of the light polarization effects on Weak Lensing. Chapter 4 presents our results on the study of the persistence detector effect and the resulting systematic effect on weak lensing. Here in this chapter we present our final conclusions and discuss the future outlook.

In chapter 2, we study the impact of non-Gaussianity in the likelihood of weak-lensing two-point statistics on the inference of two cosmological parameters. By analyzing simulated data derived from N-body simulations, we show that the standard assumption of a Gaussian likelihood, despite the true likelihood being non-Gaussian, does not cause significant biases in the resulting parameters. This finding suggests that given the currently available simulation using the simpler Gaussian likelihood for these statistics should suffice for current surveys as well as for upcoming larger-area surveys like LSST. A multivariate Gaussian likelihood will continue to be a valid approximation in cosmic shear analyses. However, this may not be the final conclusion for the parameter biases due to the uncertainty of likelihoods. Measuring the high-dimensional likelihoods is extremely difficult. It is likely that our likelihood models, especially the non-parametric models, would benefit from orders of magnitude of increase in the number of simulation realizations. In addition, it will be interesting to extend the analysis of likelihood non-Gaussianity from the shear-only analysis to a multi-probe analysis. However, it will become even more difficult to fully determine the multivariate likelihood distribution as the number of dimensions increases, and a multi-probe simulation with even more realizations is required to model the likeli-

hood of the joint data vector. The high computational cost of N-body simulations limits the number of realizations available. Even in the case of cosmic shear only tomographic analysis, reconstructing in detail the full high-dimensional non-Gaussian likelihood is still an unsolved problem. A potential way to address this problem is through fast simulation of the 3x2pt correlation functions. This will be achieved either by generating realisations of random Gaussian fields given the correlated theoretical 3x2pt power spectra or by using approximated N-body gravity solver to include the non-linear clustering. Both methods yield larger simulated data sets so as to allow us to precisely characterize and model the multivariate likelihood distribution of the joint data vector.

In chapter 3, we constructed a toy model to estimate the magnitude of polarization-driven systematic errors in the context of optical/NIR weak lensing surveys. The light polarization induces two types of systematic biases: the selection bias in favor of galaxies with specific orientations and the polarization-dependent PSF uncertainty. We build toy models to obtain an estimate for both of these polarization-induced biases and show that both biases are comparable to the Roman systematics tolerance level, indicating the need for more detailed studies. For future works, we can develop a practical analysis pipeline beyond the our toy model to remove the biases due to polarization effects. In the toy model, we make assumptions about the light polarization of galaxies and adopt a simplified galaxy selection method. It will be interesting to construct a pipeline that considers realistic galaxy selection criteria, dithering strategies, camera rotation and the wavelength dependence of the polarization effect. This is going to improve our ability to control the associated systematic biases in weak lensing measurements. In addition, it is promising to test the polarization effect in the Roman Space Telescope supernova deep fields. In the supernova fields, the same set of galaxies are observed repeatedly over a period of time. If there is a polarization-dependent throughput and the galaxies are polarized then we expect to see their magnitudes vary sinusoidally with a period of 6 months. This measurement of the magnitude oscillation will provide us with necessary parameters in modeling the polarization-induced biases.

In chapter 4, we investigate the new source of weak lensing systematics due to persistence effect of infrared detectors. For the NIR arrays, charges that are trapped in the photodiodes during earlier exposure are gradually released into the current exposure, leading to contamination of the images and errors of the galaxy shapes. This persistence signal from previous exposures would then contaminate the images of galaxies and impact the galaxy shapes. We use image simulations that incorporate the image persistence model to study its impact on galaxy shapes and weak lensing signal. In the simulations, detector effects relevant to Roman Space Telescope are incorporated, but we put our focus on the persistence effect. An empirical persistence model of the H4RG near-infrared photodiode detector is applied to model the image persistence of earlier exposures. Even though no mitigation method of persistence is used in our simulations to remove the image persistence effect from earlier exposures, we do not expect significant impact on galaxy shapes for weak lensing sciences given the current persistence performance of the detectors. The correlation functions of  $\Delta|e|$  due to persistence do not show strong correlation between the

galaxies. A comparison against the Roman total additive shear error budget further indicates that the persistence is a weak effect on weak lensing. Besides persistence, two other detector artifacts, including burn-in and count-rate non-linearity, are also directly related to the charge capture and charge release process of the charge traps in the semiconductor. Future works can extend the analysis to cover all the effects caused by charge release and capture. This is going to require detailed understanding of the interaction and correlation between these three effects and an image simulation scheme that incorporate the physical effects that happen in parallel.

# Bibliography

- Abbott T. M. C., et al., 2018, *Phys. Rev. D*, 98, 043526
- Akahori T., et al., 2018, , 70, R2
- Akeret J., Refregier A., Amara A., Seehars S., Hasner C., 2015, *Journal of Cosmology and Astro-Particle Physics*, 2015, 043
- Anderson T. W., 2003, Wiley
- Anderson R. E., Regan M., Valenti J., Bergeron E., 2014, arXiv e-prints, p. arXiv:1402.4181
- Andersson B. G., Lazarian A., Vaillancourt J. E., 2015, , 53, 501
- Bacon D. J., Refregier A. R., Ellis R. S., 2000, , 318, 625
- Baringhaus L., Franz L., 2004, *Journal of multivariate analysis*, 88, 190
- Barreira A., Krause E., Schmidt F., 2018, , 2018, 053
- Bartelmann M., Schneider P., 2001, , 340, 291
- Beck R., 2007, in Miville-Deschênes M. A., Boulanger F., eds, *EAS Publications Series Vol. 23*, *EAS Publications Series*. pp 19–36 (arXiv:astro-ph/0603531), doi:10.1051/eas:2007003
- Beck R., Shoutenkov V., Ehle M., Harnett J. I., Haynes R. F., Shukurov A., Sokoloff D. D., Thierbach M., 2002, , 391, 83
- Becker M. R., et al., 2016, , 94, 022002
- Bernal J. L., Verde L., Riess A. G., 2016, , 2016, 019
- Bernstein G. M., Jarvis M., 2002, , 123, 583
- Biesiadzinski T., Lorenzon W., Newman R., Schubnell M., Tarlé G., Weaverdyck C., 2011, , 123, 958

Blinnikov S., Moessner R., 1998, , 130, 193

Blot L., Corasaniti P. S., Amendola L., Kitching T. D., 2016, , 458, 4462

Born M., Wolf E., 1999, Principles of Optics

Boveia A., Doglioni C., 2018, Annual Review of Nuclear and Particle Science, 68, 429

Breckinridge J. B., Lam W. S. T., Chipman R. A., 2015, , 127, 445

Bundy K., Ellis R. S., Conselice C. J., 2005, , 625, 621

Chang C., et al., 2013, , 434, 2121

Chipman R. A., Lam W. S. T., Breckinridge J., 2015, Proc. SPIE

Chisari N. E., et al., 2018, , 480, 3962

Clowe D., Randall S. W., Markevitch M., 2007, Nuclear Physics B Proceedings Supplements, 173, 28

Conley A., et al., 2011, , 192, 1

DESI Collaboration et al., 2016, arXiv e-prints, p. arXiv:1611.00036

Davis Jr. L., Greenstein J. L., 1951, , 114, 206

Dodelson S., 2017, Gravitational Lensing. Cambridge University Press, doi:10.1017/9781316424254

Dodelson S., Schneider M. D., 2013, , 88, 063537

Doré O., et al., 2018, arXiv e-prints, p. arXiv:1804.03628

Draine B. T., 2003, , 598, 1017

Draper P. W., Done C., Scarrott S. M., Stockdale D. P., 1995, , 277, 1430

Eifler T., Schneider P., Hartlap J., 2009, , 502, 721

Eifler T., Krause E., Schneider P., Honscheid K., 2014, , 440, 1379

Eifler T., Krause E., Dodelson S., Zentner A. R., Hearin A. P., Gnedin N. Y., 2015, , 454, 2451

Einstein A., 1917, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin, pp 142–152

Eisenstein D. J., Hu W., 1999, , 511, 5

Eisenstein D. J., et al., 2005, , 633, 560

Fendt C., Beck R., Lesch H., Neininger N., 1996, , 308, 713

Fendt C., Beck R., Neininger N., 1998, , 335, 123

Feng J. L., 2010, , 48, 495

Feroz F., Hobson M. P., Bridges M., 2009, , 398, 1601

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, , 125, 306

Freedman W. L., Madore B. F., Scowcroft V., Burns C., Monson A., Persson S. E., Seibert M., Rigby J., 2012, , 758, 24

Friedrich O., Eifler T., 2018, , 473, 4150

Fu L., et al., 2008, , 479, 9

Giavalisco M., et al., 2004, , 600, L93

Gretton A., Borgwardt K. M., Rasch M. J., Schölkopf B., Smola A., 2012, Journal of Machine Learning Research, 13, 723

Harnois-Déraps J., Pen U.-L., 2013, , 431, 3349

Harnois-Déraps J., van Waerbeke L., 2015, , 450, 2857

Harnois-Déraps J., et al., 2018, , 481, 1337

Harrington D. M., Sueoka S. R., 2017, Journal of Astronomical Telescopes, Instruments, and Systems, 3, 018002

Hartlap J., Simon P., Schneider P., 2007, , 464, 399

Hartlap J., Schrabback T., Simon P., Schneider P., 2009, , 504, 689

Hastings W. K., 1970, Biometrika, 57, 97

Heavens A. F., Jimenez R., Lahav O., 2000, , 317, 965

Heavens A. F., Sellentin E., de Mijolla D., Vianello A., 2017, , 472, 4244

Heitmann K., Lawrence E., Kwan J., Habib S., Higdon D., 2014, , 780, 111

Hemmati S., et al., 2019, , 877, 117

Heymans C., et al., 2013, , 432, 2433

Hikage C., et al., 2019a, Publications of the Astronomical Society of Japan, p. 22

Hikage C., et al., 2019b, , 71, 43

Hilbert S., Hartlap J., Schneider P., 2011, , 536, A85

Hildebrandt H., et al., 2016, , p. stw2805

Hildebrandt H., et al., 2017, , 465, 1454

Hildebrandt H., et al., 2018 (arXiv:1812.06076)

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

Hirata C., Seljak U., 2003, , 343, 459

Hirata C. M., et al., 2004, , 353, 529

Hirata C. M., Gehrels N., Kneib J.-P., Kruk J., Rhodes J., Wang Y., Zoubian J., 2012, arXiv e-prints, p. arXiv:1204.5151

Hoekstra H., Jain B., 2008, Annual Review of Nuclear and Particle Science, 58, 99

Hounsell R., et al., 2018, , 867, 23

Huff E. M., Eifler T., Hirata C. M., Mandelbaum R., Schlegel D., Seljak U., 2014, , 440, 1322

Ivezić Ž., et al., 2008, arXiv e-prints, p. arXiv:0805.2366

Ivezić Ž., et al., 2019, , 873, 111

Izbicki R., Lee A. B., Schafer C. M., 2014, Journal of Machine Learning Research (AIS-TATS Track), pp 420–429

Jarvis M., Bernstein G., Jain B., 2004, , 352, 338

Jee M. J., Tyson J. A., Hilbert S., Schneider M. D., Schmidt S., Wittman D., 2016, , 824, 77

Joachimi B., 2017, , 466, L83

Joachimi B., et al., 2015, , 193, 1

Johnson P. B., Christy R. W., 1972, , 6, 4370

Johnson N. L., Rogers C. A., 1951, The Annals of Mathematical Statistics, 22, 433

Jones T. J., 2000, , 120, 2920

Jones A., Wang L., Krisciunas K., Freeland E., 2012, The Astrophysical Journal, 748, 17

Jura M., 1982, , 258, 59

Kannawadi A., Shapiro C. A., Mandelbaum R., Hirata C. M., Kruk J. W., Rhodes J. D., 2016, , 128, 095001

Kautsch S. J., Grebel E. K., Barazza F. D., Gallagher J. S. I., 2006, , 445, 765

Keitel D., Schneider P., 2011, , 534, A76

Kilbinger M., 2015, Reports on Progress in Physics, 78, 086901

Krause E., Eifler T., 2017, , 470, 2100

Krause E., Eifler T., Blazek J., 2016, , 456, 207

Krause E., et al., 2017 (arXiv:1706.09359)

Kristian J., 1967, , 147, 864

LSST Science Collaboration et al., 2009, preprint, (arXiv:0912.0201)

Larson D., et al., 2011, , 192, 16

Laureijs R., et al., 2011, preprint, (arXiv:1110.3193)

Lewis A., Bridle S., 2002, , 66, 103511

Lin C.-H., Tan B., Mandelbaum R., Hirata C. M., 2020, , 496, 532

Lincheng Z., Zhijun L., 1985, Acta Mathematica Sinica, 1, 314

Long K. S., Baggett S. M., MacKenty J. W., Riess A. G., 2012, in . p. 84421W, doi:10.1117/12.926778

Mandelbaum R., 2018, Annual Review of Astronomy and Astrophysics, 56, 393

Mandelbaum R., et al., 2005, , 361, 1287

Mandelbaum R., et al., 2011, , 410, 844

Mandelbaum R., et al., 2015, , 450, 2963

Mandelbaum R., et al., 2018, Publications of the Astronomical Society of Japan, 70, S25

Massey R., et al., 2007, The Astrophysical Journal Supplement Series, 172, 239

Massey R., et al., 2013, MNRAS, 429, 661

Mathis J. S., 1990, , 28, 37

Mead A. J., Peacock J. A., Heymans C., Joudaki S., Heavens A. F., 2015, , 454, 1958

Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E., 1953, The Journal of Chemical Physics, 21, 1087

Padilla L. E., Tellez L. O., Escamilla L. A., Vazquez J. A., 2019, arXiv e-prints, p. arXiv:1903.11127

Peel A., Lin C.-A., Lanusse F., Leonard A., Starck J.-L., Kilbinger M., 2017, , 599, A79

Percival W. J., Brown M. L., 2006, , 372, 1104

Percival W. J., et al., 2010, , 401, 2148

Perlmutter S., et al., 1999, , 517, 565

Petrov V. V., 1962, Vestnik Leningrad. Univ., 17, 150

Planck Collaboration et al., 2014a, , 571, A15

Planck Collaboration et al., 2014b, , 571, A16

Planck Collaboration et al., 2016, , 594, A13

Planck Collaboration et al., 2018, arXiv e-prints, p. arXiv:1807.06209

Radhakrishnan V., 1989, , 30, 181

Regan M., Bergeron E., Lindsay K., Anderson R., 2012, in . p. 84424W, doi:10.1117/12.927033

Riess A. G., et al., 1998, , 116, 1009

Riess A. G., et al., 2009, , 699, 539

Rowe B. T. P., et al., 2015, Astronomy and Computing, 10, 121

Rubin V. C., Ford W. Kent J., 1970, , 159, 379

Sato M., Hamana T., Takahashi R., Takada M., Yoshida N., Matsubara T., Sugiyama N., 2009, , 701, 945

Sato M., Ichiki K., Takeuchi T. T., 2010, Physical Review Letters, 105, 251301

Sato M., Ichiki K., Takeuchi T. T., 2011a, , 83, 023501

Sato M., Takada M., Hamana T., Matsubara T., 2011b, , 734, 76

Scarrott S. M., 1996, , 37, 297

- Scarrott S. M., Ward-Thompson D., Warren-Smith R. F., 1987, , 224, 299
- Scarrott S. M., Rolph C. D., Semple D. P., 1990, in Beck R., Kronberg P. P., Wielebinski R., eds, IAU Symposium Vol. 140, Galactic and Intergalactic Magnetic Fields. pp 245–251
- Scarrott S. M., Eaton N., Axon D. J., 1991, , 252, 12P
- Schaan E., Krause E., Eifler T., Doré O., Miyatake H., Rhodes J., Spergel D. N., 2017, , 95, 123512
- Schneider P., Hartlap J., 2009, , 504, 705
- Schneider P., Eifler T., Krause E., 2010, , 520, A116
- Schneider M. D., Cole S., Frenk C. S., Szapudi I., 2011, , 737, 11
- Sellentin E., Heavens A. F., 2016, , 456, L132
- Sellentin E., Heavens A. F., 2018, , 473, 2355
- Sellentin E., Heymans C., Harnois-Déraps J., 2018, , 477, 4879
- Semboloni E., Schrabback T., van Waerbeke L., Vafaei S., Hartlap J., Hilbert S., 2011, , 410, 143
- Simmons J. F. L., Audit E., 2000, , 319, 497
- Singh S., Mandelbaum R., Seljak U., Slosar A., Vazquez Gonzalez J., 2017, , 471, 3827
- Skilling J., 2006, Bayesian Anal., 1, 833
- Smith R. M., Zavodny M., Rahmer G., Bonati M., 2008a, in . p. 70210J, doi:10.1117/12.789372
- Smith R. M., Zavodny M., Rahmer G., Bonati M., 2008b, in . p. 70210K, doi:10.1117/12.789619
- Sofue Y., Fujimoto M., Wielebinski R., 1986, , 24, 459
- Spergel D., et al., 2015, preprint, (arXiv:1503.03757)
- Spitzer L., 1978, Physical processes in the interstellar medium, doi:10.1002/9783527617722.
- Stil J. M., Krause M., Beck R., Taylor A. R., 2009, The Astrophysical Journal, 693, 1392
- Székely G. J., Rizzo M. L., 2004, InterStat, 5, 1
- Takada M., Jain B., 2009, , 395, 2065

Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, , 761, 152

Taruya A., Takada M., Hamana T., Kayo I., Futamase T., 2002, , 571, 638

Taylor A., Joachimi B., 2014, , 442, 2728

Taylor A., Joachimi B., Kitching T., 2013, , 432, 1928

Taylor P. L., Kitching T. D., Alsing J., Wandelt B. D., Feeney S. M., McEwen J. D., 2019, , 100, 023519

Trotta R., 2008, Contemporary Physics, 49, 71

Troxel M. A., Ishak M., 2015, , 558, 1

Troxel M. A., et al., 2017, preprint, (arXiv:1708.01538)

Troxel M. A., et al., 2018a, Phys. Rev. D, 98, 043528

Troxel M. A., et al., 2018b, , 98, 043528

Troxel M. A., et al., 2019, arXiv e-prints, p. arXiv:1912.09481

Tulloch S., George E., ESO Detector Systems Group 2019, Journal of Astronomical Telescopes, Instruments, and Systems, 5, 036004

Van Waerbeke L., et al., 2000, , 358, 30

Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, , 530, 87

Weingartner J. C., Draine B. T., 2001, , 548, 296

Wilking P., Schneider P., 2013, , 556, A70

Wilking P., Röseler R., Schneider P., 2015, , 582, A107

Wittman D. M., Tyson J. A., Kirkman D., Dell'Antonio I., Bernstein G., 2000, , 405, 143

Wood-Vasey W. M., et al., 2007, , 666, 694

Zuntz J., et al., 2018, , 481, 1149

Zwicky F., 1937, , 86, 217